Curvature Fields from Shading Fields

Xinran Han XINRANHAN@G.HARVARD.EDU
Todd Zickler ZICKLER@SEAS.HARVARD.EDU
Harvard University

Editors: Sophia Sanborn, Christian Shewmake, Simone Azeglio, Nina Miolane

Abstract

We re-examine the estimation of 3D shape from images that are caused by shading of diffuse Lambertian surfaces. We propose a neural model that is motivated by the well-documented perceptual effect in which shape is perceived from shading without a precise perception of lighting. Our model operates independently in each receptive field and produces a scalar statistic of surface curvature for that field. The model's architecture builds on previous mathematical analyses of lighting-invariant shape constraints, and it leverages geometric structure to provide equivariance under image rotations and translations. Applying our model in parallel across a dense set of receptive fields produces a curvature field that we show is quite stable under changes to a surface's albedo pattern (texture) and also to changes in lighting, even when lighting varies spatially across the surface.

Keywords: shape perception, neural distance field, shape from shading, low-level vision

1. Introduction

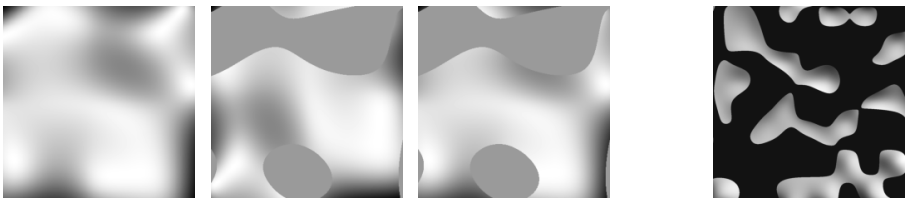
There are many unanswered questions about the computational mechanisms that underlie the perception of three-dimensional shape from diffuse shading in an image. One question relates to the perceptual ambiguity that exists between perceiving an image as being caused by shading and perceiving it as being caused by an alternative effect like defocus or translucency (Mooney et al., 2019; Marlow and Anderson, 2021). Another relates to the human ability to perceive shape without inferring a precise representation of incident lighting (Cavanagh, 2005). Figure 1 contains some examples.

Computational models for "shape from shading" have been explored extensively in computer vision since the 1970s (Horn, 1970), but they do not provide answers to these questions because they deviate from human perception in two fundamental ways.

The first relates to representation. Whereas computational models represent output shape as a single depth map or field of surface normals (Horn and Brooks, 1989; Zhang et al., 1999; Barron and Malik, 2014; Bednarik et al., 2018), perception produces shape representations that are seemingly topological or qualitative (Zucker, 2020; Todd and Petrov, 2022; Linton et al., 2023). Moreover, most computational models include an explicit representation of lighting that is inferred in conjunction with shape (Zoran et al., 2014; Barron and Malik, 2014; Wimbauer et al., 2022). This is at odds with the proven perceptual abilities to report shape much faster than lighting and to report shape without noticing artificial manipulations of lighting (Cavanagh, 2005; Ostrovsky et al., 2005).

The second difference relates to assumptions about the causes of input pixel values. Almost all computational models assume their input pixel values were caused by diffuse shading, as opposed to being caused by some other effect. This assumption often manifests

© 2023 X. Han & T. Zickler.



(a) (b) (c) (d)

Figure 1: (a) A diffusely shaded image without high-curvature ridges can be perceived as defocus instead. Introducing boundaries can either (b) weaken or (c) strengthen the perception of shading and 3D shape. (d) When the perception of shading and 3D shape is strong, it is often unaccompanied by a precise perception of lighting. People can see the shape easily, while it takes some time to find which one has different lighting. Here, one of the segments is lit differently than the others—can you see which one? (Figures are inspired by Anderson and Marlow (2022) and Cavanagh (2005).)

as an oracle that pre-segments the image and says "these pixels were caused by shading". The human observer, in contrast, does not have this luxury. She must decide for herself which pixels were caused by shading and which ones were not. Perceptual experiments show this determination can be challenging and ambiguous, even in very simple scenarios (Marlow et al., 2019), and previous models provide no means for exploring why or how.

As a step toward bridging the gap between traditional machine vision models for shape from shading and the insights gained from psychophysical studies, we aim to address the following open questions under the Lambertian shading assumption: What information about shape should be extracted from shading cues alone, and without inferred lighting? How can we computationally model this process?

We take a neural-geometric approach to answering these questions. We begin with the ansatz that from shading alone, a vision system should aim to infer a particular scalar curvature statistic—the logarithmic of Casorati curvature—instead of explicit surface normals or depth. Then, we build on previous mathematical characterizations of Lambertion shading constraints (Kunsberg and Zucker, 2014; Xiong et al., 2014; Heal et al., 2020) to design a neural architecture that infers this curvature statistic within each local receptive field. Crucially, the architecture is designed to have a desirable mix of invariances and equivariances: There is invariance to both the surface albedo ("paint color") and the scene lighting; and there is equivariance under 2D similarity transformations (rotation, translation and scale) of the image plane.

We examine the behavior of our system using synthetic images like the one in Figure 1(a), but with albedo and lighting patterns that vary from image to image, and that vary spatially within each image. We find that our model produces curvature fields that are quite stable under these changes to albedo and lighting patterns, and that are much more accurate and stable than comparable curvature fields obtained from existing shape from shading or inverse rendering models.

We note that our approach is complementary to a recent trend of training large black-box neural networks to perform monocular depth estimation (e.g., Ranftl et al. (2020, 2021)) and inverse rendering (e.g., Li et al. (2018); Yu and Smith (2019); Wimbauer et al. (2022)). While the black-box approach has the advantage of capacity to exploit all sorts of cues in a training set, it lacks interpretability and often struggles to generalize to images that are radically new. Our work suggests an alternative front-end architecture that could help future systems be more efficient and adaptable.

2. Methodology

According to a shadow-less Lambertian shading model, a graph-like surface with height function $h(\bar{x}, \bar{y})$ induces image $I(\bar{x}, \bar{y})$ as

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

where $\rho \in \mathbb{R}_+$ and $\mathbf{L} \in \mathbb{R}^3$ are the albedo and incident lighting at each location, and $\mathbf{N}(\bar{x}, \bar{y}) = (-\frac{\partial h}{\partial \bar{x}}(\bar{x}, \bar{y}), -\frac{\partial h}{\partial \bar{y}}(\bar{x}, \bar{y}), 1)$ is the field of un-normalized surface normals.

As depicted in Fig. 2, our system processes each receptive field P independently. So hereon, we use a patch-centered coordinate system $(x, y) \in P$, with coordinate $(0, 0)$ at the receptive field's center, and $I(x, y)$ and $f(x, y)$ the image and surface height functions, respectively, within the receptive field. Critically, we assume the receptive field size is such that the local surface $f(x, y)$ is well approximated by a quadratic function

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

with third and higher-order derivatives being negligible.[1] Our task is to infer the local shape $\mathbf{f} = (f_x, f_y, f_{xx}, f_{xy}, f_{yy})$, or in fact a statistic of that shape, from local intensity $I(x, y)$. Note that when convenient, we express quadratic shape as $\mathbf{f} = (n, \kappa)$ and refer to $n = (f_x, f_y)$ as *orientation* and $\kappa = (f_{xx}, f_{xy}, f_{yy})$ as *curvature*.

Our system has two levels of processing. First, at each image point we use derivative filters to compute the local 2-jet, meaning the derivatives up to second order, $\mathbf{I} = (I, I_x, I_y, I_{xx}, I_{xy}, I_{yy})$. This is fed to a (pre-trained) neural unsigned distance function $g_\theta(\mathbf{I}, \mathbf{f})$ whose zero-set encodes the manifold of local shapes $\mathbf{f}$ that are consistent with $\mathbf{I}$. One of our contributions is a design for g_θ that provides equivariance under similarity transformations of (x, y) and invariance to the albedo and lighting at the point in question.

In the second stage of processing, we aggregate information from many points in P using spatial consistency. This stage has an iterative form, akin to test-time optimization, and it is simple with no trainable parameters. We design it to output a particular scalar statistic of the local shape $\mathbf{f}$ in patch P, namely the logarithm of the Casorati curvature. As we describe next, our choice of output is motivated by both mathematics and vision science.

1. In the wild, the "right" receptive field size will vary across the visual field and cannot be known beforehand, so it would be necessary to apply and fuse our patch-based processing across multiple scales. In this paper, we simply assume the surface $h(\bar{x}, \bar{y})$ has spatial frequencies that permit one receptive field size everywhere, and that this size is known *a priori*.

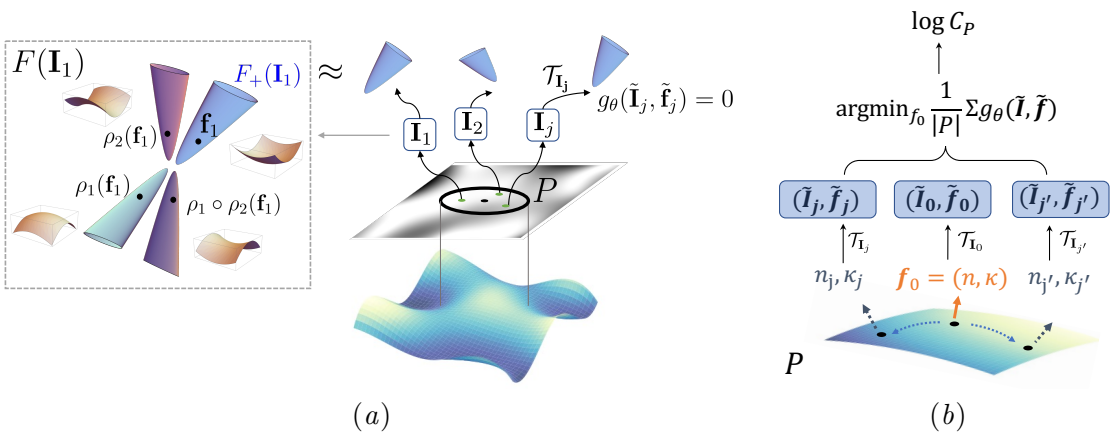


Figure 2: Two stages of processing to produce $\log C_P$ from image patch P. (a) Pre-trained neural UDF g_θ operates in parallel on image 2-jets $\mathbf{I}_j$ at every point. The UDF's zero-set at j approximates a canonical piece $F_+(\mathbf{I}_j)$ of the two-dimensional manifold of quadratic shapes $\mathbf{f} \in F(\mathbf{I}_j)$ that are consistent with $\mathbf{I}_j$. Geometric transforms $\mathcal{T}_{\mathbf{I}_j}$ precede each UDF to provide equivariance. (b) We iteratively estimate the patch's central normal and curvature, $\mathbf{f}_0 = (n, \kappa) \in \mathbb{R}^5$, and thus its $\log C_P$, by quadratically transporting the central (n, κ) to get values (n_j, κ_j) at each j and minimizing the sum of neural distances that they imply.

2.1. Why Casorati Curvature?

Mathematically, when the lighting and albedo fields are unknown, there is an entire three-parameter family of surfaces, $a\bar{x} + b\bar{y} + ch(\bar{x}, \bar{y})$ with $a, b, c \in \mathbb{R}$, that can exactly explain a Lambertian image $I(\bar{x}, \bar{y})$. This is true even when lighting $\mathbf{L}(\bar{x}, \bar{y}) = \mathbf{L}$ is uniform over the image (Belhumeur et al., 1999), and it has been dubbed the generalized bas-relief ambiguity. Perceptual experiments by Koenderink et al. (2001) are consistent with this idea: They suggest that when subjects are forced to make choices about first-order shape $(\partial h/\partial \bar{x}, \partial h/\partial \bar{y})$ they choose surfaces from among the three-parameter family somewhat arbitrarily. For a second order Taylor approximation to the surface in patch P, this ambiguity implies an equivalence class of local shapes $(a + cf_x, b + cf_y, cf_{xx}, cf_{xy}, cf_{yy})$, which in turn implies that in the absence of other cues, a quadratic shape from shading processor for patch P should produce, at best, an estimate of curvature up to scale: $c\kappa = c(f_{xx}, f_{xy}, f_{yy})$.

A separate analysis of shading, this time within small patches, has shown that when lighting and albedo are unknown and uniform over a patch such as P, an exactly-quadratic surface (Equation (2)) can, at best, be determined up to a four-way choice that comprises a convex-concave-saddle ambiguity (Xiong et al., 2014). This implies that a reasonable expectation for local shading processors like ours is to produce a statistic of shape that is invariant to this ambiguity. Next, we show that Casorati curvature is one such statistic.

Casorati curvature was introduced by Koenderink (1990) as a convenient reparametrization of local curvature $\kappa = (f_{xx}, f_{xy}, f_{yy})$. The reparameterization is $(C, \sigma, 2\phi)$ with Casorati

curvature C, shape index σ and orientation ϕ; and it relates to κ via

$$C = \sqrt{(k_1^2 + k_2^2)/2}; \quad \sigma = \arctan \frac{k_2 + k_1}{k_2 - k_1}; \quad \phi = \frac{1}{2} \arctan \frac{2 f_{xy}}{f_{xx} - f_{yy}},$$

with $k_1 \geq k_2 \in \mathbb{R}$ the principal curvatures (i.e., eigenvalues of the Hessian formed from κ). The Casorati curvature measures the overall deviation from local planarity, while orientation $2\phi \in [-\pi, \pi)$ indicates the rotation of the principal curvature frame relative to the (x, y) frame, and shape index $\sigma \in [-\pi/2, +\pi/2]$ smoothly transitions through convex, saddle-like and concave quadratics that have equal deviations from planarity.

To show that Casorati curvature is invariant under the four-way ambiguity of Xiong et al. (2014), we apply the mathematical expressions of the ambiguity's two defining automorphisms, ρ_1, ρ_2, to the reparameterized local shape and find (for $\sigma > 0$)

$$\begin{aligned} \rho_1 &: (f_x, f_y, C, \sigma, 2\phi) \to (-f_x, -f_y, C, -\sigma, \pi - 2\phi) \\ \rho_2 &: (f_x, f_y, C, \sigma, 2\phi) \to (f_x \cos(2\phi) + f_y \sin(2\phi), f_x \sin(2\phi) - f_y \cos(2\phi), C, \pi/2 - \sigma, 2\phi). \end{aligned} \tag{3}$$

This shows that Casorati curvature is indeed invariant, and it also provides an intuitive description of the ambiguity: A convex shape $\mathbf{f}$ transforms to a concave shape via ρ_1 and to saddles via ρ_2 and $\rho_1 \circ \rho_2$.

Combining these observations leads to the logarithm of Casorati curvature as a desirable output for any locally-quadratic shading processor, because it is invariant to both the local four-way ambiguity and the global multiplier c from the generalized bas relief ambiguity. The next sections propose and evaluate an architecture for producing this output.

3. Point-wise Shape with a Neural Unsigned Distance Function

Our first stage is the parallel application of a neural unsigned distance function (UDF) g_θ to every point in P. Our architecture draws on the differential geometry results of Kunsberg and Zucker (2014) and their subsequent algebraic geometry counterparts (Heal et al., 2020; Heal, 2021). These provide mathematical characterizations of the set of second-order shapes $\mathbf{f} \in \mathbb{R}^5$ that are consistent with an image 2-jet $\mathbf{I} = (I, I_x, I_y, I_{xx}, I_{xy}, I_{yy})$ when the point's albedo and lighting are unknown. One result says that given a point's 2-jet, such as $\mathbf{I}_1$ in Figure 2(*a*), there is a two-dimensional manifold of consistent shapes $\mathbf{f}$. Another result says the manifold comprises four isomorphic pieces, so neural representations like ours need only represent one piece. We state the results here, and details are in the appendix.

Theorem 1 *(Kunsberg and Zucker (2014), Heal et al. (2020)) Assume shadow-less Lambertian shading (Equation (1)) of a quadratic surface (Equation (2)) and the image 2-jet* $\mathbf{I}$ *at a point* (x, y) *that has locally-constant albedo and lighting, i.e.,* $\forall k > 0, D^k_{x,y}\rho(x, y) = D^k_{x,y}\mathbf{L}(x, y) = 0$. *A non-degenerate shape* $\boldsymbol{f} \in \mathbb{R}^5$ *that is consistent with* $\mathbf{I}$ *satisfies the three fourth-order polynomials,* $C_i(\mathbf{f}; \mathbf{I}) = 0, i = 1, 2, 3$, *in Appendix A.*

This result applies to shapes $\mathbf{f}$ that are non-degenerate, meaning they satisfy $(f_{xx} + f_{yy})(f_{xx}f_{yy} - f_{xy}^2)(4f_{xy}^2 + (f_{xx} - f_{yy})^2) \neq 0$, which excludes perfect planes, cylinders and surfaces with equal principal curvature magnitudes. We call the set of non-degenerate shapes that form the zero-set of $C_i(\mathbf{f}; \mathbf{I})$ the *consistent shape-set* and denote it as $F(\mathbf{I})$, a two-dimensional algebraic variety.

Remark 2 *(Heal et al. (2020)) The automorphisms of Equation (3) divide $F(\mathbf{I})$ into four isomorphic pieces, allowing the consistent shape-set to be represented by only one piece. We select the piece that corresponds to convex shapes and call it the* positive shape set*, defined as $F_+(\boldsymbol{I}) = \{\mathbf{f} \in F(\mathbf{I}) \mid f_{xx} + f_{yy} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0\}$.*

The left of Figure 2(*a*) depicts the projection of $F(\mathbf{I}_1)$ onto the three curvature dimensions (f_{xx}, f_{xy}, f_{yy}), with colors showing its four pieces. Knowing a consistent shape $\mathbf{f}_1$ on the positive shape set allows computing three other consistent shapes via Equation (3).[2]

3.1. Equivariance

There are no known symbolic expressions for the zero-set of $C_i(\mathbf{f}; \mathbf{I})$ given $\mathbf{I}$ (Heal et al., 2020; Heal, 2021), so we propose representing $F_+(\mathbf{I})$ implicitly using a neural UDF $g_\theta(\mathbf{I}, \mathbf{f})$ that evaluates to zero for all $\mathbf{f} \in F_+(\mathbf{I})$. Here, we show that our UDF can be made equivariant under 2D similarity transformations of the image plane and invariant to overall brightness by preceding it with a measurement-dependent transformation $\mathcal{T}_\mathbf{I} : (\mathbf{I}, \mathbf{f}) \mapsto (\tilde{\mathbf{I}}, \tilde{\mathbf{f}})$. We also show this transformation reduces the dimension of the UDF domain from eleven to eight.

Our transformation $\mathcal{T}_\mathbf{I}$ essentially invokes a point-wise change of the image coordinate system to one that aligns with the eigenvectors of the image Hessian constructed from (I_{xx}, I_{xy}, I_{yy}) and that is scaled according to the Hessian eigenvalues. We describe it here as a composition of four components.

1. Contrast normalization. Since polynomials C_i are linear in intensity I, we can achieve invariance to overall brightness simply by normalizing a 2-jet by its intensity and leaving the shape-vector unchanged: $(\mathbf{I}, \mathbf{f}) \mapsto (\mathbf{I}/I, \mathbf{f})$.

2. Rotation to image Hessian frame. We rotate the local (x, y)-frame to align with the eigenvectors of the image Hessian. Suppose matrix $R_\mathbf{I}^T H_\mathbf{I} R_\mathbf{I} = \mathrm{diag}(\lambda_1, \lambda_2)$ with eigenvalues $\lambda_1 \geq \lambda_2$, then the transformation is: $(I_x, I_y) \mapsto (I_x, I_y)R_\mathbf{I}$; $(f_x, f_y) \mapsto (f_x, f_y)R_\mathbf{I}$;

$$\begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix} \mapsto R_\mathbf{I}^T \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix} R_\mathbf{I} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}; \text{ and } \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \mapsto R_\mathbf{I}^T \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} R_\mathbf{I}.$$

3. Isotropic scale. We scale the (x, y)-frame isotropically based on the sum of squared image Hessian eigenvalues. Defining $s = \sqrt{1/(\lambda_1^2 + \lambda_2^2)}$ this implies:

$$\mathbf{I} \mapsto (I, sI_x, sI_y, s^2 I_{xx}, s^2 I_{xy}, s^2 I_{yy}); \quad \mathbf{f} \mapsto (f_x, f_y, sf_{xx}, sf_{xy}, sf_{yy}).$$

4. Reflection, if needed. Finally, if the transformed value of I_x is negative, we reflect the orientation via $\mathbf{I} \mapsto (I, -I_x, -I_y, I_{xx}, I_{xy}, I_{yy})$ and $\mathbf{f} \mapsto (-f_x, -f_y, f_{xx}, f_{xy}, f_{yy})$.

One can show (Heal, 2021) that the composition of these transformations does not affect consistency, meaning $\forall i \; C_i(\mathbf{f}; \mathbf{I}) = 0 \Leftrightarrow C_i(\tilde{\mathbf{f}}; \tilde{\mathbf{I}}) = 0$. It is also clear that the transformed 2-jet has the form $\tilde{\mathbf{I}} = (1, \tilde{I}_x, \tilde{I}_y, \tilde{I}_{xx}, 0, \tilde{I}_{yy})$ with $\tilde{I}_x \geq 0$; and since $\lambda_1 \geq \lambda_2$, we know $(\tilde{I}_{xx}, \tilde{I}_{yy})$ is contained in the unit half-circle, which we parameterize by, say, $\alpha \in [0, \pi]$. Thus, the transformation maps 2-jets in $\mathbb{R}^6$ to $\mathbb{R}^+ \times \mathbb{R} \times \mathbb{S}^+$, and we can write $\mathcal{T}_\mathbf{I} : (\mathbf{I}, \mathbf{f}) \mapsto \Big((\tilde{I}_x, \tilde{I}_y, \alpha), (\tilde{f}_x, \tilde{f}_y, \tilde{f}_{xx}, \tilde{f}_{xy}, \tilde{f}_{yy})\Big)$.

2. Appendix A includes the expressions for the automorphisms in the $\{f_{xx}, f_{xy}, f_{yy}\}$ parameterization.

By preceding our UDF with transformation $\mathcal{T}_{\mathbf{I}}$ and following it with the inverse transformation $\mathcal{T}_{\mathbf{I}}^{-1}$ applied to each zero-distance shape $\tilde{\mathbf{f}}$, we obtain a system for making inferences about point-wise shape $\mathbf{f}$ that is invariant to image brightness and equivariant to 2D similarities. Moreover, we substantially decrease the dimension of the UDF domain, which reduces network complexity and training.

We note that our neural representation and transformation $\mathcal{T}_{\mathbf{I}}$ are inspired by those in (Heal et al., 2020) and offer three improvements. By aligning our local (x, y)-frame with the image Hessian instead of with the gradient (I_x, I_y) used by (Heal et al., 2020), we accommodate all non-degenerate points, including those with vanishing I_x, I_y. Also, we expand upon that work's equivariance under rigid transformations by also including scale s. Finally and most importantly, we avoid Heal et al. (2020)'s assumption that $F_+(\mathbf{I})$ has the graph structure $\kappa = h_\theta(n; \mathbf{I})$, which is not valid as we show in Appendix C.

3.2. UDF Architecture and Training

Our neural approximation to the positive shape set is a parameterized unsigned distance function $g_\theta : \mathbb{R}^+ \times \mathbb{R} \times \mathbb{S}^+ \times \mathbb{R}^5 \to \mathbb{R}^+$, $(\tilde{\mathbf{I}}, \tilde{\mathbf{f}}) \mapsto d$ that evaluates to zero when (transformed) point-wise shape $\tilde{\mathbf{f}}$ is consistent with (transformed) 2-jet $\tilde{\mathbf{I}}$. For convenience we use notation $v \triangleq (\tilde{\mathbf{I}}, \tilde{\mathbf{f}})$, an 8-vector. Our training scheme is inspired by Gropp et al. (2020). Namely, we learn parameters θ by numerically generating samples that are consistent, $v_i = (\tilde{\mathbf{I}}_i, \tilde{\mathbf{f}}_i)$ with $\tilde{\mathbf{f}}_i \in F_+(\tilde{\mathbf{I}}_i)$, and random samples that are not, $v' \in \mathbb{R}^8$; and then we minimize

$$l(\theta) = \frac{1}{|\{v_i\}|} \sum_{v_i} |g_\theta(v_i)| + \lambda \mathbb{E}_{\{v'\}} (||\nabla_{v'} g_\theta(v')|| - 1)^2 \tag{4}$$

with some $\lambda > 0$. The first term encourages zero distance for each v_i, and the second "Eikonal" term encourages the general behavior of a distance function (Gropp et al., 2020), which has unit gradient magnitude almost everywhere. To improve convergence, we find it helpful to create a third set of samples v'' that are distance ϵ from $F_+(\tilde{\mathbf{I}}_i)$ and add the additional loss $\mathbb{E}_{\{v''\}} \left(|g_\theta(v'') - \epsilon| + (||\nabla_{v''} g_\theta(v'')|| - 1)^2 \right)$. Details are in Appendix D.

We create 440k training samples v_i using a numerical polynomial solver. The network has 8 fully-connected layers each with 512 hidden units and a softplus layer, and a skip connection from the input to the middle layer. We use geometric weight initialization from Atzmon et al. (2019) and train for 7k epochs using Adam optimizer and learning rate 0.001.

4. Spatial Consistency within Receptive Field P

While there is an entire manifold of shapes that are consistent with a single 2-jet $\mathbf{I}$, the results from (Xiong et al., 2014; Heal, 2021) suggest that combining multiple 2-jets within a patch P can reduce the ambiguity to the discrete four-way choice, and thus can uniquely identify the Casorati curvature, which is invariant to this four-way choice.

As depicted in Figure 2(*b*), our strategy is to define and iteratively optimize over a patch-level orientation and curvature $(n, \kappa) \in \mathbb{R}^5$ that equals these values at the receptive field's center $(0, 0)$. Since the surface is quadratic over P by assumption, choosing a value for these patch-level variables implies values for point-wise shapes $\mathbf{f}_j = (n_j, \kappa_j)$ at all points

$(x_j, y_j) \in P$ via quadratic transport:

$$n_j = A_j \kappa^T + n \text{ with } A_j = \begin{pmatrix} x_j - x & y_j - y & 0 \\ 0 & x_j - x & y_j - y \end{pmatrix}. \tag{5}$$

Thus, we can measure the consistency of patch-level shape (n, κ) with all of the 2-jets in patch P by summing the consistencies of the point-wise shapes they imply. That is, we use iterative optimization to solve

$$\underset{n,\kappa}{\operatorname{argmin}} \frac{1}{|P|} \sum_{j \in P} g_\theta \left(\tilde{\mathbf{I}}_j ; \mathcal{T}_{\mathbf{I_j}}(n_j, \kappa_j) \right), \tag{6}$$

with each (n_j, κ_j) determined by (n, κ) via Equation (5). We use 400 steps of the Adam optimizer with learning rate 0.01.

By applying this process in parallel to the overlapping receptive fields in an image (stride 1) we obtain estimates of (n, κ) for every pixel. Each pixel is independently subject to the four-way ambiguity, but the log-Casorati curvature $\log C$ is uniquely determined. This is what we visualize as the output in our experiments.

We emphasize that even though we use the log-Casorati curvature field for visualization, our system's output is richer than that. It comprises the full values (n, κ) at each pixel, which via the automorphisms of Equation (3), represent an equivalence class of four local shapes for that pixel. Appendix F contains a preliminary experiment that combines this richer information with external boundary cues to reconstruct a field of surface normals.

5. Experiments

For our simulation experiments, we focus on shapes whose images are inherently ambiguous. By this we mean they are void of other shape cues like boundaries, periodic textures or specular highlights. They are also void of high-curvature ridges that recent results have shown can also induce "anchor curves" for shape (Kunsberg and Zucker, 2021).

We create surfaces $h(\bar{x}, \bar{y})$ that are cubic splines with a grid of 6×6 knots having random height values, and we render images of them $I(\bar{x}, \bar{y})$ using Equation (1) with various albedo and lighting fields. The shapes are all sufficiently low-frequency for the quadratic assumption (Equation (2)) to be valid in every 7×7 patch. Each input is a 41×41 grid of image derivatives (the 2-jet at each pixel) and our patch processor operates in parallel in the overlapping stride-one 7×7 patches. We visualize the output as a log-Casorati curvature field over the 41×41 image domain. Figure 3 shows two examples.

For each surface, we consider a variety of albedo and lighting conditions. In each panel of Figure 3, each of the first three columns has uniform albedo $\rho(\bar{x}, \bar{y}) = 1$ and spatially-uniform lighting $\mathbf{L}(\bar{x}, \bar{y}) = \mathbf{L}$ from a randomly-chosen direction. The next two columns each have uniform albedo $\rho(\bar{x}, \bar{y}) = 1$ and a lighting field $\mathbf{L}(\bar{x}, \bar{y})$ whose directions change spatially within an angular anulus of $25°$ to $50°$ of the view direction. The final two columns are rendered with similar varying lighting fields but with albedo fields $\rho(\bar{x}, \bar{y})$ that are piecewise constant. Specifically, Column 6 has circular regions with $\rho = (0.7, 0.85)$, and Column 7 has circular regions with $\rho = (0.1, 0.2)$. All of the images in Figure 3 are rendered without noise. Appendix E contains additional results, including some with additive Gaussian noise.

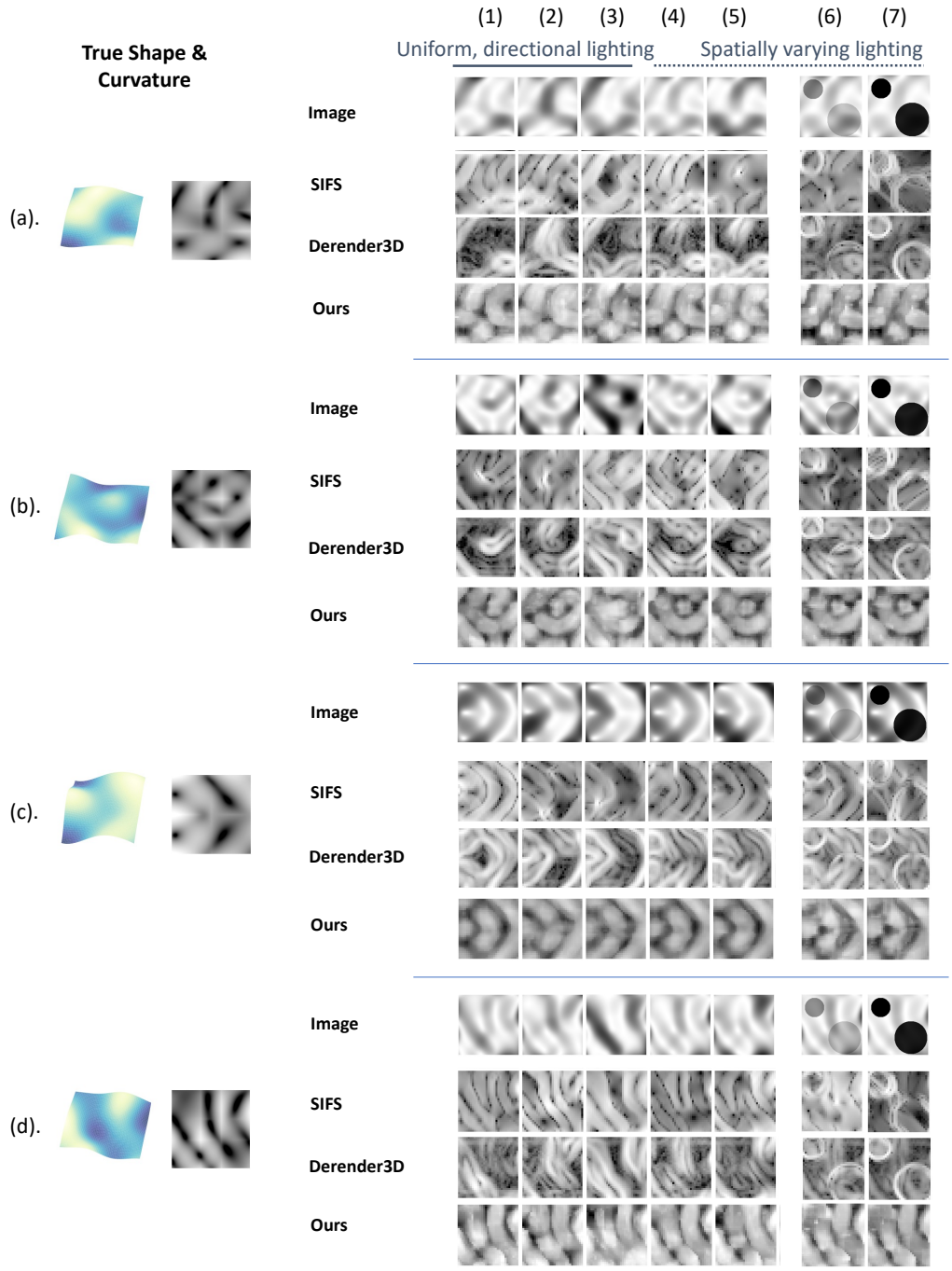


Figure 3: True and estimated curvature fields for four surfaces with various lighting and albedo patterns. The first three columns have uniform albedo and are lit uniformly but from different directions, while the last four columns have spatially varying lighting, and either uniform or spatially-varying albedo. We compare with SIFS (a boundary-less variant of SIRFS (Barron and Malik, 2014)) and Derender3D (Wimbauer et al., 2022) by computing log-Casorati curvatures from their predicted normal fields.

As baselines, we show comparable results derived from two previous computer vision models: Derender3D (Wimbauer et al., 2022) and a boundary-less and uniform-albedo variant of SIRFS (Barron and Malik, 2014), known as SIFS. Like most existing computer vision models for shape reconstruction, these are inverse-rendering systems that aim to infer precise lighting (and reflectance) in addition to shape. Both models operate on entire images instead of patches so they can exploit the global statistics of the surface shapes that occur in their training sets. We choose these two models because they are rare examples that do not rely critically on the presence of boundary contours or other shape cues. The boundary-less and uniform-albedo variant of SIRFS is an interpretable model with an engineered prior for $h(\bar{x}, \bar{y})$, and derender3D is an autoencoder trained on more than 20k images. Each of the two models produces a surface normal field, and from these we compute curvature and then log-Casorati curvature fields.

We find that our model's curvature fields are more accurate and more stable under variations in lighting and albedo, especially when lighting and albedo fields vary within an image, and when some albedo values are very small. We attribute our model's stability to its architecture, which by Theorem 1 and the structure of Equation (6), allows the albedo and lighting to vary (as piecewise constant functions) even inside of P. We attribute the baseline models' weaker performance to the fact that our images are different from their training images, which highlights the challenge of creating models that can generalize.

Limitations. While they are randomly generated, our test surfaces are all sufficiently low-frequency for the quadratic surface approximation to be accurate using the same receptive field size at every location. In the wild this will almost never occur, so it would be necessary to execute our processing using many receptive field sizes (and possibly many Gaussian derivative scales), and to design fusion or selection schemes that can exploit the appropriate sizes where needed. Along these lines, we have also yet to explore how to incorporate a point-wise UDF front-end architecture like ours into a drastically-larger model that has capacity to leverage other visual cues and non-local context. Another limitation of our model is its reliance on Lambertian rendering for UDF pre-training. This is very different from how biological systems must learn, and it is not well-aligned with perceptual evidence that humans can infer shape information for a variety of non-Lambertian diffuse rendering functions (Zucker, 2020). Perhaps there are unsupervised schemes that can train UDF-based architectures like ours directly from images instead of from algebraic equations.

6. Conclusion

Experiments with human perception suggest there are useful computational "shortcuts" for inferring 3D shape from diffuse shading without having to infer precise lighting or execute complete inverse rendering. We have introduced some ideas for what such a shortcut could be and how it could be implemented, with particular attention to achieving some desirable invariance and equivariance properties. Our simple prototype system, which operates patch-wise and has no trainable parameters beyond those of the pre-trained UDF g_θ, exhibits the ability in simulation to produce curvature fields that are more accurate and stable than alternative inverse rendering models, in particular when the albedo field or incident lighting field are varying across a surface. This may inspire new directions for research in learned representations that are more efficient or that align more closely with biological systems.

References

Barton L Anderson and Phillip J Marlow. Perceiving the shape and material properties of 3d surfaces. *Trends in Cognitive Sciences*, 2022.

Matan Atzmon, Niv Haim, Lior Yariv, Ofer Israelov, Haggai Maron, and Yaron Lipman. Controlling neural level sets. *Advances in Neural Information Processing Systems*, 32, 2019.

Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2014.

Jan Bednarik, Pascal Fua, and Mathieu Salzmann. Learning to reconstruct texture-less deformable surfaces from a single view. In *2018 international conference on 3d vision (3DV)*, pages 606–615. IEEE, 2018.

Peter N Belhumeur, David J Kriegman, and Alan L Yuille. The bas-relief ambiguity. *International journal of computer vision*, 35(1):33–44, 1999.

Patrick Cavanagh. The artist as neuroscientist. *Nature*, 434(7031):301–307, 2005.

Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *Proceedings of the 37th International Conference on Machine Learning*, pages 3789–3799, 2020.

Kathryn Heal. *A Lighting-Invariant Approach to Local Shape from Shading*. PhD thesis, Harvard University, 2021.

Kathryn Heal, Jialiang Wang, Steven J Gortler, and Todd Zickler. A lighting-invariant point processor for shading. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 94–102, 2020.

Berthold KP Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. 1970.

Berthold KP Horn and Michael J Brooks. *Shape from shading*. MIT press, 1989.

Jan Koenderink and Andrea van Doorn. Shape, shading, brain and awareness. In *Neuromathematics of vision*, pages 87–106. Springer, 2014.

Jan J Koenderink. *Solid shape*, volume 2. MIT press Cambridge, MA, 1990.

Jan J Koenderink and Andrea J Van Doorn. Surface shape and curvature scales. *Image and vision computing*, 10(8):557–564, 1992.

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.

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.

Benjamin Kunsberg and Steven W Zucker. From boundaries to bumps: when closed (extremal) contours are critical. *Journal of Vision*, 21(13):7–7, 2021.

Zhengqin Li, Zexiang Xu, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics (TOG)*, 37(6):1–11, 2018.

Paul Linton, Michael J Morgan, Jenny CA Read, Dhanraj Vishwanath, Sarah H Creem-Regehr, and Fulvio Domini. New approaches to 3d vision, 2023.

Phillip J Marlow and Barton L Anderson. The cospecification of the shape and material properties of light permeable materials. *Proceedings of the National Academy of Sciences*, 118(14):e2024798118, 2021.

Phillip J Marlow, Scott WJ Mooney, and Barton L Anderson. Photogeometric cues to perceived surface shading. *Current Biology*, 29(2):306–311, 2019.

Scott WJ Mooney, Phillip J Marlow, and Barton L Anderson. The perception and misperception of optical defocus, shading, and shape. *Elife*, 8:e48214, 2019.

Yuri Ostrovsky, Patrick Cavanagh, and Pawan Sinha. Perceiving illumination inconsistencies in scenes. *Perception*, 34(11):1301–1314, 2005.

René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE transactions on pattern analysis and machine intelligence*, 44(3):1623–1637, 2020.

René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12179–12188, 2021.

James T Todd and Alexander A Petrov. The many facets of shape. *Journal of Vision*, 22 (1):1–1, 2022.

Felix Wimbauer, Shangzhe Wu, and Christian Rupprecht. De-rendering 3d objects in the wild. In *CVPR*, 2022.

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*, 37(1):67–79, 2014.

Ye Yu and William AP Smith. Inverserendernet: Learning single image inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2019.

Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape-from-shading: a survey. *IEEE transactions on pattern analysis and machine intelligence*, 21(8):690–706, 1999.

Daniel Zoran, Dilip Krishnan, Jose Bento, and Bill Freeman. Shape and illumination from shading using the generic viewpoint assumption. *Advances in Neural Information Processing Systems*, 27, 2014.

Steven W Zucker. On qualitative shape inferences: a journey from geometry to topology. *arXiv preprint arXiv:2008.08622*, 2020.

Appendix A. Polynomials of Consistent Shape

The polynomials that implicitly describe the consistent shapes $\mathbf{f}$ for image measurement $\mathbf{I}$ are (from Heal et al. (2020)):

$$\begin{aligned} C_1(\mathbf{f};\mathbf{I}) = {} & f_x^4 I_{xx} + 2f_x^3 f_{xx} I_x + f_x^2 f_{xy}^2 I + 2f_x^2 f_{xy} f_y I_x + 2f_x^2 f_y^2 I_{xx} \\ & + 2f_x^2 I_{xx} - 2f_x f_{xx} f_{xy} f_y I + 2f_x f_{xx} f y^2 I_x + 2f_x f_{xx} I_x \\ & + f_{xx}^2 f_y^2 I + f_{xx}^2 I + f_{xy}^2 I + 2f_{xy} f_y^3 I_x + 2f_{xy} f_y I_x \\ & + f_y^4 I_{xx} + 2f_y^2 I_{xx} + I_{xx} \end{aligned} \tag{7}$$

$$\begin{aligned} C_2(\mathbf{f};\mathbf{I}) = {} & f_x^4 I_{yy} + 2f_x^3 f_{xy} I_y + 2f_x^2 f_y^2 I_{yy} + 2f_x^2 f_y f_{yy} I_y + f_x^2 f_{yy}^2 I \\ & + 2f_x^2 I_{yy} + 2f_x f_{xy} f_y^2 I_y - 2f_x f_{xy} f_y f_{yy} I + 2f_x f_{xy} I_y \\ & + f_{xy}^2 f_y^2 I + f_{xy}^2 I + f_y^4 I_{yy} + 2f_y^3 f_{yy} I_y + 2f_y^2 I_{yy} \\ & + 2f_y f_{yy} I_y + f_{yy}^2 I + I_{yy} \end{aligned} \tag{8}$$

$$\begin{aligned} C_3(\mathbf{f};\mathbf{I}) = {} & f_x^4 I_{xy} + f_x^3 f_{xx} I_y + f_x^3 f_{xy} I_x + f_x^2 f_{xy} f_y I_y + f_x^2 f_{xy} f_{yy} I \\ & + 2f_x^2 f_y^2 I_{xy} + f_x^2 f_y f_{yy} I_x + 2f_x^2 I_{xy} + f_x f_{xx} f_y^2 I_y \\ & - f_x f_{xx} f_y f_{yy} I + f_x f_{xx} I_y - f_x f_{xy}^2 f_y I \\ & + f_x f_{xy} f_y^2 I_x + f_x f_{xy} I_x + f_{xx} f_{xy} f_y^2 I + f_{xx} f_{xy} I \\ & + f_{xy} f_y^3 I_y + f_{xy} f_y I_y + f_{xy} f_{yy} I + f_y^4 I_{xy} + f_y^3 f_{yy} I_x \\ & + 2f_y^2 I_{xy} + f_y f_{yy} I_x + I_{xy} \end{aligned} \tag{9}$$

As noted in Heal et al. (2020), the real solutions to the polynomial system $\{C_i = 0, \forall i\}$ are identical to those for the rational equations in Corollary 4.2 of Kunsberg and Zucker (2014).

The automorphism in Equation (3) can be expressed in the Euclidean frame as:

$$\rho_1 : \mathbf{f} \to -(f_x, f_y, f_{xx}, f_{xy}, f_{yy}), \quad \rho_2 : \mathbf{f} \to \frac{1}{\sqrt{4f_{xy}^2 + (f_{xx} - f_{yy})^2}} \begin{pmatrix} f_x f_{xx} - f_x f_{yy} + 2f_y f_{xy} \\ 2f_x f_{xy} + f_y f_{yy} - f_y f_{xx} \\ f_{xx}^2 - f_{xx} f_{yy} + 2f_{xy}^2 \\ f_{xx} f_{xy} + f_{xy} f_{yy} \\ f_{yy}^2 - f_{xx} f_{yy} + 2f_{xy}^2 \end{pmatrix} \tag{10}$$

Appendix B. Curvature Landscape

Define the following transformation: $r = \frac{1}{2}(f_{xx} - f_{yy}); s = f_{xy}; t = \frac{1}{2}(f_{xx} + f_{yy})$. Denote the principal curvatures of the local shape as $k_1 \geq k_2 \in \mathbb{R}$. The definitions of curvature landscape descriptors $(C, \sigma, 2\phi)$ can be written as

$$\begin{aligned}
&\textbf{Casorati curvature } C = \sqrt{r^2 + s^2 + t^2} = \sqrt{(k_1^2 + k_2^2)/2};\\
&\textbf{Shape Index } \sigma = \arctan \frac{t}{\sqrt{r^2 + s^2}} = \arctan \frac{k_2 + k_1}{k_2 - k_1};\\
&\textbf{Orientation } \phi = \frac{1}{2} \arctan \frac{s}{r}
\end{aligned}$$

Since the principal curvature magnitude $|k_1|, |k_2|$ remains unchanged for the four-way ambiguity (Xiong et al., 2014), it is expected that the Casorati curvature is an invariant statistics under the convex-concave-saddle ambiguity.

We can also express r, s, t as

$$\begin{aligned}
t &= C \sin \sigma;\\
r &= C \cos \sigma \cos 2\phi;\\
s &= C \cos \sigma \sin 2\phi.
\end{aligned}$$

Then (r, s, t) and $(C, \sigma, 2\phi)$ are the Cartesian and polar coordinats of the quadratic shape space, according to Koenderink and Van Doorn (1992); Koenderink and Doorn (2014). We visualize this space in Fig. 4, where the radius of the sphere corresponds to the casorati curvature. Note that the orientation ϕ is not defined for $\sigma = \pm\frac{\pi}{2}$ since the shape is rotationally symmetric.

Appendix C. Representing F_+: Explicit Graph or Implicit Unsigned Distance Function?

Intuition for representing the positive shape set F_+ implicitly by a neural unsigned distance function $(f_x, f_y, f_{xx}, f_{xy}, f_{yy}) \mapsto \mathbb{R}_+$ (as done in this paper) instead of explicitly representing it by a neural graph $(f_x, f_y) \mapsto (f_{xx}, f_{xy}, f_{yy})$ (as done in Heal et al. (2020)) can be gleaned from the following one-dimensional case.

A locally quadratic "surface" has height $f(x) = ax + \frac{1}{2}bx^2$. The normal at the center pixel $x = 0$ is thus $\mathbf{N}(x) = (-b, 1)$. With light direction $\mathbf{L} = (L_x, 1)$, albedo ρ and view direction $(0, 1)$, the intensity at image point x is $I(x) = \rho\mathbf{L} \cdot \frac{\mathbf{N})(x)}{||\mathbf{N}(x)||}$. We acquire image measurement $\mathbf{I} = (I, I_x, I_{xx})$ by calculating spatial image derivatives at $x = 0$. Using similar derivations as in Heal (2021), we obtain the polynomial defining the consistent shape set at $x = 0$:

$$D_1(\mathbf{f}; \mathbf{I}) = D_1(a, b; \mathbf{I}) = b^2 I + (1 + a^2)2abI_x + (1 + a^2)^2 I_{xx}. \tag{11}$$

Any consistent shape that is non-degenerate ($b \neq 0$) satisfies $D_1(a, b; \mathbf{I}) = 0$.

As an aside, here is the analog to our paper's canonicalization $\mathcal{T}_{\mathbf{I}}$ in this one-dimensional case:

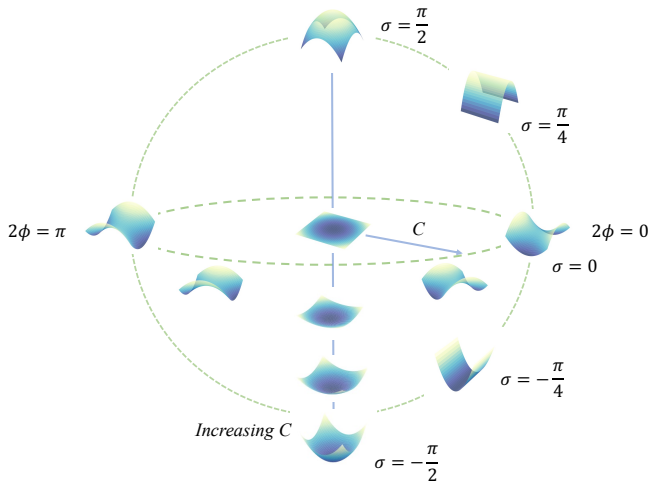


Figure 4: Visualization of curvature Landscape descriptors $(C, \sigma, 2\phi)$. Casorati curvature C represents the deviation from planarity. Shape index σ is scale-invariant and characterizes the quadratic shape into spherical cups/caps, saddles, ridges, etc. Orientation ϕ describes the relative alignment of the principal curvatures w.r.t. the Cartesian frame.

- Constant scaling: division by I such that $(I, I_x, I_{xx}) \to (1, \frac{I_x}{I}, \frac{I_{xx}}{I})$. Curve parameters (a, b) remain unchanged.
- Isotropic scaling by $\sqrt{|I_{xx}|}$: $(1, I_x, I_{xx}) \to (1, \frac{I_x}{\sqrt{|I_{xx}|}}, \pm 1)$ and $(a, b) \to (a, \frac{b}{\sqrt{|I_{xx}|}})$.
- Reflection: When $I_x < 0$, we can transform it to positive value by reflection $I_x \to -I_x$ and $a \to -a$. The second order terms I_{xx}, b are not affected.

C.1. 1D Case: F_+ is not a graph

In the one-dimensional case, given an image measurement (I, I_x, I_{xx}) and a surface orientation a, one can solve for the corresponding consistent curvatures b analytically:

$$b = \frac{-2aI_x(1+a^2) \pm \sqrt{4a^2(1+a^2)^2 I_x^2 - 4I(1+a^2)^2 I_{xx}}}{2I}. \tag{12}$$

We can also plot the consistent shape set $F(\mathbf{I}) = \{(a, b) | D_1(a, b; \mathbf{I}) = 0\}$ on the (a, b)-plane, for any measurement $\mathbf{I}$. Figure 5 shows a few examples.

Analogous to the four-way choice in the main paper, in this one-dimensional scenario there is a two-way concave/convex ambiguity. For instance, shapes $(a, b) = (-1, 1)$ and $(a', b') = (1, -1)$ generate the same image under lighting $\mathbf{L} = (0, 1)$. Thus, we can define the positive shape set $F_+(\mathbf{I})$ as the subset of consistent shapes for which $b > 0$.

But even when we consider just the positive shape set (those with $b > 0$), we see that in two of the three examples, there are two distinct positive curvature values that form a

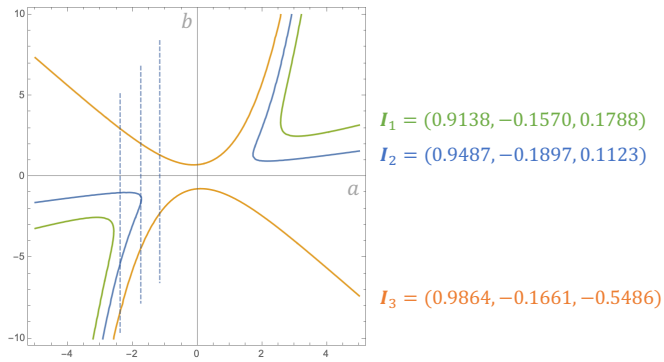


Figure 5: Solution set for three 1D curve image measurements $\mathbf{I}_i$. The dotted lines show that for given a, we can have none, one or two solutions with positive/negative curvature b.

consistent shape together with a. That is, the positive shape set $F_+(\mathbf{I})$ cannot be represented as a function of a. This is made precise by the discriminant:

$$\Delta = 4a^2(1+a^2)^2 I_x^2 - 4I(1+a^2)^2 I_{xx} = 4(1+a^2)^2(a^2 I_x^2 - I I_{xx}). \tag{13}$$

When $I_{xx} < 0$ we always have the discriminant > 0. Since the value Δ under the square root is also larger than the first term in the numerator, for any $a \in \mathbb{R}$, we can find a solution $b_1 > 0$ and another solution $b_2 < 0$, such as the orange line in Fig 5. When $I_{xx} > 0$, there are cases when a has no solution, one solution or two solutions of $b \in \mathbb{R}^+$. Examples are shown by the green and blue lines.

C.2. 2D Case: Numerical evidence

In the two-dimensional case, we have not yet found examples where $F_+(\mathbf{I})$ can be symbolically proven to not be a graph, but it is easy to acquire very strong evidence from a numerical polynomial solver. An example is shown in Table 1 and Figure 6, where we fix measurement $\mathbf{I}$ and orientation (f_x, f_y) and ask a numerical polynomial solver [3] for the real-valued roots (f_{xx}, f_{xy}, f_{yy}) that satisfy $C_i(f_{xx}, f_{xy}, f_{yy}; f_x, f_y, \mathbf{I}) = 0\ \forall i$, from which we select the ones with positive curvature. The table and figure report these roots for some choices of orientation (f_x, f_y), along with their associated Casorati curvature values.

First, we observe that as in the 1D case, there are multiple (two) positive curvatures that form a consistent shape along with orientation (f_x, f_y), suggesting that the positive shape set is not a graph over orientation. Moreover, for image measurements where a chosen orientation (f_x, f_y) has multiple positive curvatures (f_{xx}, f_{xy}, f_{yy}) that are consistent, we observe behavior that mimics the one-dimensional case: When $|f_x|, |f_y|$ are small (i.e., the normal is close to the view direction), the curvatures κ, κ' and their Casorati curvatures are close. When $|f_x|, |f_y|$ are larger (i.e., as the normal turns away from the view direction), the

3. Solved using https://reference.wolfram.com/language/ref/NSolve.html

Table 1: Given surface first order terms f_x, f_y, there can be multiple positive shape solutions. Numerically solved examples for $\mathbf{I} = (0.6315, -0.3403, 0.3672, -1.2077, 0.5387, -0.0505)$. We use / to denote the case when there is no real solution output from the solver.

f_x	f_y	f_{xx}	f_{xy}	f_{yy}	Casorati curvature
-0.1	-0.2	/	/	/	/
-0.5	-1	/	/	/	/
-1	-2	3.0835	-0.9895	3.8359	3.6180
		3.0722	-1.1204	2.3267	2.9464
-2	-3	4.7432	-0.4776	8.5723	6.9440
		4.6068	-1.3903	2.4684	3.9485
-3	-4	6.6049	0.0904	12.7974	10.1837
		6.2794	-1.7003	2.9464	5.1911

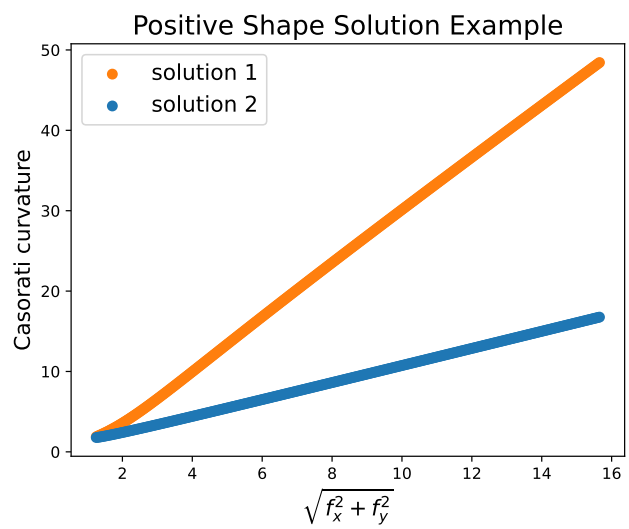


Figure 6: Positive shape solution from a specific image measurement $\mathbf{I}$ along the line $f_y = 2f_x$. We compute the Casorati curvature from each numerically solved (f_{xx}, f_{xy}, f_{yy}) solution. We see from the figure that a given (f_x, f_y) can have 2 solutions, one with higher curvature (orange) and the other with lower curvature (blue) and their difference diverge as $|f_x|, |f_y|$ becomes larger.

difference between κ, κ' also increases. This is analogous to the 1D visualization where, for the blue and green cases, the two consistent positive curvatures b diverge as $|a|$ gets larger.

Appendix D. Details on Model Training and Inference

D.1. Training Data Generation

To obtain the training samples, we first generate image vectors $\mathbf{I}$ by sampling some light source directions $\mathbf{L}$ and surface vectors $\mathbf{f}$. Then for each $\mathbf{I}$ we densely sample surface normals and solve for the roots of the polynomial system using Mathematica. We filter for solutions in the positive shape set F_+ and set a upper bound for curvature values to obtain around 440k training data points.

We train the network g_θ on canonicalized $\tilde{\mathbf{f}}, \tilde{\mathbf{I}}$ with the loss in Equation (4) for 7k iterations using the Adam optimizer and learning rate 0.001. The loss hyper-parameter λ is set to 1.

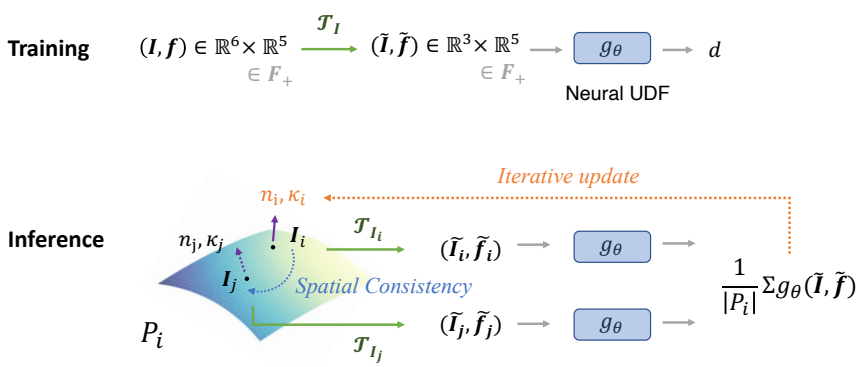


Figure 7: During training, we optimize UDF weights θ by minimizing distance d at ground truth points $(\mathbf{I}, \mathbf{f})$ together with loss $\mathcal{L}_{aug}$ at augmented points and the Eikonal loss at random points. At inference time, given patch P_i centered at i, the surface vector of every neighboring pixel j is computed with spatial consistency constraint (Equation (5)) using the normal n_i and curvature κ_i of point i. After applying canonicalization $\mathcal{T}_{\mathbf{I_j}}$ pixel-wise, we use gradient descent to minimize the mean distance loss over the center pixel's shape.

In Fig. 7 we show how the canonicalization step is applied in both training and inference. When training the UDF, consistent $\mathbf{f}, \mathbf{I}$ pairs from the ground truth manifold F_+ are densely sampled. They are then canonicalized by $\mathcal{T}_\mathbf{I}$ to train the UDF weights θ. We also incorporate an augmentation loss and an Eikonal loss on points near the ground truth manifold as described in the main paper.

During inference, the transformation $\mathcal{T}_\mathbf{I}$ is analytic and invertible, so we can use auto-differentiation to optimize for the shape.

D.2. Data Augmentation for UDF Training

The method of Gropp et al. (2020) learns a *signed* distance function for an $(n-1)$-dimensional manifold embedded in $\mathbb{R}^n$. In those cases, the normal direction is unique (up to sign) and one can also regularize the loss based on normals if provided by data. However, their sampling scheme becomes problematic for manifolds that have lower dimension. To learn an implicit neural representation for such lower-dimensional objects, we need to modify the network architecture to output *unsigned* distance instead of *signed* distance, because signed distance is ill-defined when there is no notion of an interior.

After modifying the network to output unsigned distance, we test it on a toy problem comprising a space curve in 3D. We find that the sampling scheme from Gropp et al. (2020) is not sufficient to learn a good representation of the space curve, converging instead to a higher dimensional surface. Examples are shown in Fig.8. This is because the method relies on implicit geometric regularization and only sample sparsely in space. Since the logarithmic spiral is a subset of the cone and a circle is a subset of the sphere, without further dense sampling to disambiguate the two shape, the UDF for the curve converges to the higher dimensional shape instead.

We use the following data augmentation to improve the sampling scheme. For k-dimensional differentiable manifold in $\mathbb{R}^n$ with explicit parametrization, we can compute its tangent space basis vectors $\{\mathbf{t}_i \in \mathbb{R}^n, i \in [1, 2, \ldots, k]\}$ by taking partial derivatives. Then we can derive the normal space basis $\{\mathbf{n}_i \in \mathbb{R}^n, i \in [1, 2, \ldots, n-k]\}$ by finding vectors that are orthogonal to $\{\mathbf{t}_i\}$. Then we randomly sample in the normal space to get a unit-length direction vector $\mathbf{d} = \sum \beta_i \mathbf{n}_i(\mathbf{x})$ that is locally orthogonal to the manifold at $\mathbf{x}$. Taking a distance of ϵ away from the manifold gives an augmented data point $\mathbf{x}''$. By construction, it has unsigned distance ϵ to the ground truth manifold. Repeating this step on multiple points $\{\mathbf{x}\}$ sampled from the ground truth manifold forms a 'shell' around the shape. We can apply the distance loss and Eikonal loss on the augmented data points:

$$\mathcal{L}_{aug} = \mathbb{E}_{\mathbf{x}''} \left(\tau_1 |g_\theta(\mathbf{x}'') - \epsilon| + \tau_2 (||\nabla_{\mathbf{x}''} g_\theta(\mathbf{x}'')|| - 1)^2 \right)$$

We then train the network along with the regular training loss described in the main paper. The original data and augmented data for the space curves is visualized in Fig. 8.

For manifold with implicit parametrizations, the tangent space at point x is spanned by the null space of the Jacobian matrix. We can then find an orthogonal direction by computing vectors that are perpendicular to all the tangent vectors at a point. In the case of the consistent shape polynomial for F_+, the canonicalized shape space has 8 dimensions. We compute its Jacobian (3 by 8) by taking partial derivatives of C_i with respect to the canonicalized parameters $\tilde{I}_x, \tilde{I}_y, \alpha, \tilde{f}_x, \tilde{f}_y, \tilde{f}_{xx}, \tilde{f}_{xy}, \tilde{f}_{yy}$. Then for each sampled data point $\mathbf{v} \in \mathbb{R}^8$, we sample along some orthogonal direction to the shape manifold for ϵ distance and obtain an augmented data point $\mathbf{v}''$.

This approach relies on having an analytic expression for the manifold. In practice, we also need to ensure that the distance ϵ is chosen such that the augmented points do not intersect with the original manifold. For training the consistent shape UDF of our model, we choose $\epsilon = 0.1$ and $\tau_1 = \tau_2 = 1$.

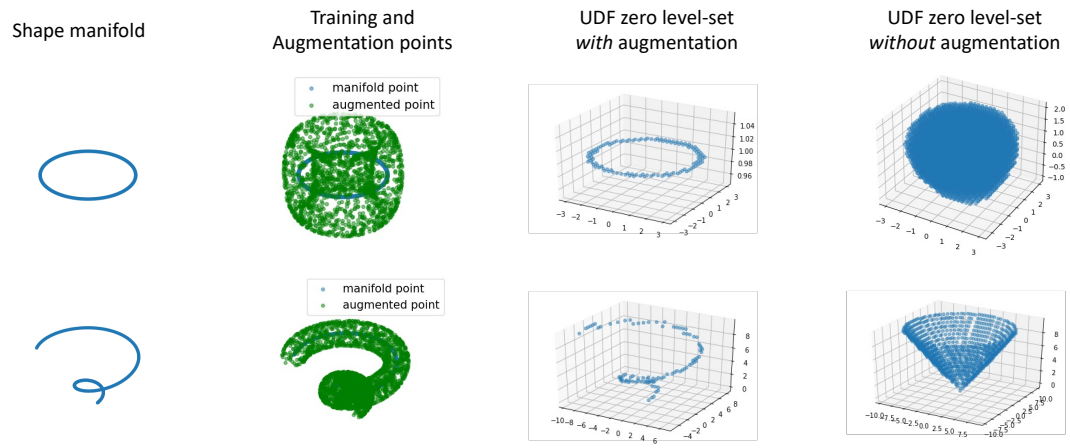


Figure 8: Examples for training the UDF with and without data augmentation. The zero level-set is computed on a discrete grid for points with distance < 0.1. We see that without augmentation, the network tends to converge to a surface instead of the space curve.

Appendix E. More Results on Curvature Field Estimation

Here we present more experiment results for inferring the log-Casorati curvature field on synthetic surfaces.

Fig. 9 summarizes our models' predicted curvature field on multiple spline surfaces. The last column (f) has additive Gaussian noise with variance $= 4e^{-4}$. For all images we use overlapping patch of size 7 by 7 for optimization.

In Fig. 10 we present additional results where the algorithm uses different patch sizes. For the first three surfaces, we use overlapping patches of size 7 by 7. For the fourth surface, we use patch of size 9 by 9. For the last test surface we use patch size of 11 by 11. The patch size is visualized by the blue box on the images.

Appendix F. Planar cut and normal reconstruction

Inspired by previous psychophysical studies, we hypothesize that without additional priors or boundary cues, a perceptually consistent normal field cannot be uniquely determined. When such cue is present and the resulting image induces 3D perception, we can incorporate additional shape prior, such as a convexity prior to reconstruct the normal field partially. We examine this hypothesis using the planar cut experiment inspired by Anderson and Marlow (2022), where a surface is cut with a horizontal plane at certain height.

We hypothesize that a planar cut will reinforce the perception of 3D only when its boundaries are aligned with low curvature regions. An example is shown in Fig. 11 left. In the middle column, when the log-Casorati curvature minimum (dark regions) align or has small angles with the planar cut contour, the planar cut is perceived as consistent. In the right column, large areas of curvature minimum occur in the interior of the shape instead

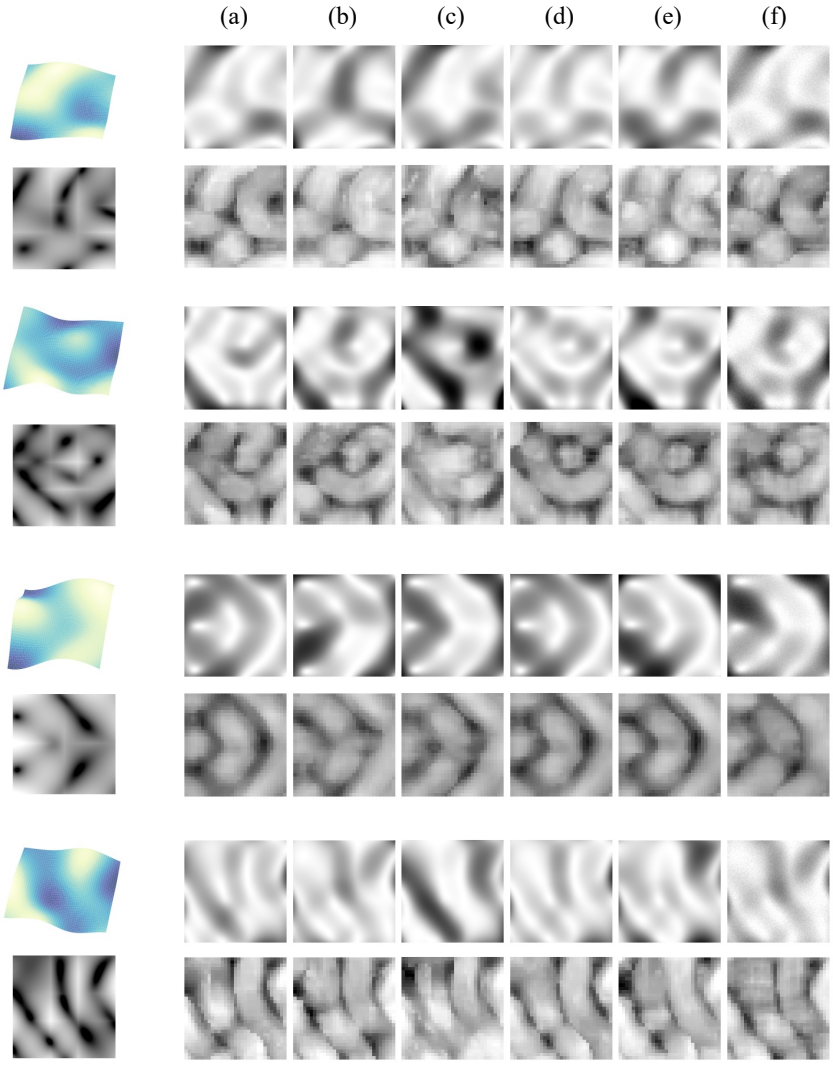


Figure 9: Ground truth and estimated log-Casorati curvature field visualizations. Column (a)-(c) are single directional light source. Column (d)-(e) are spatially varying light sources. Column (f) has added Gaussian noise with variance = $4e^{-4}$.

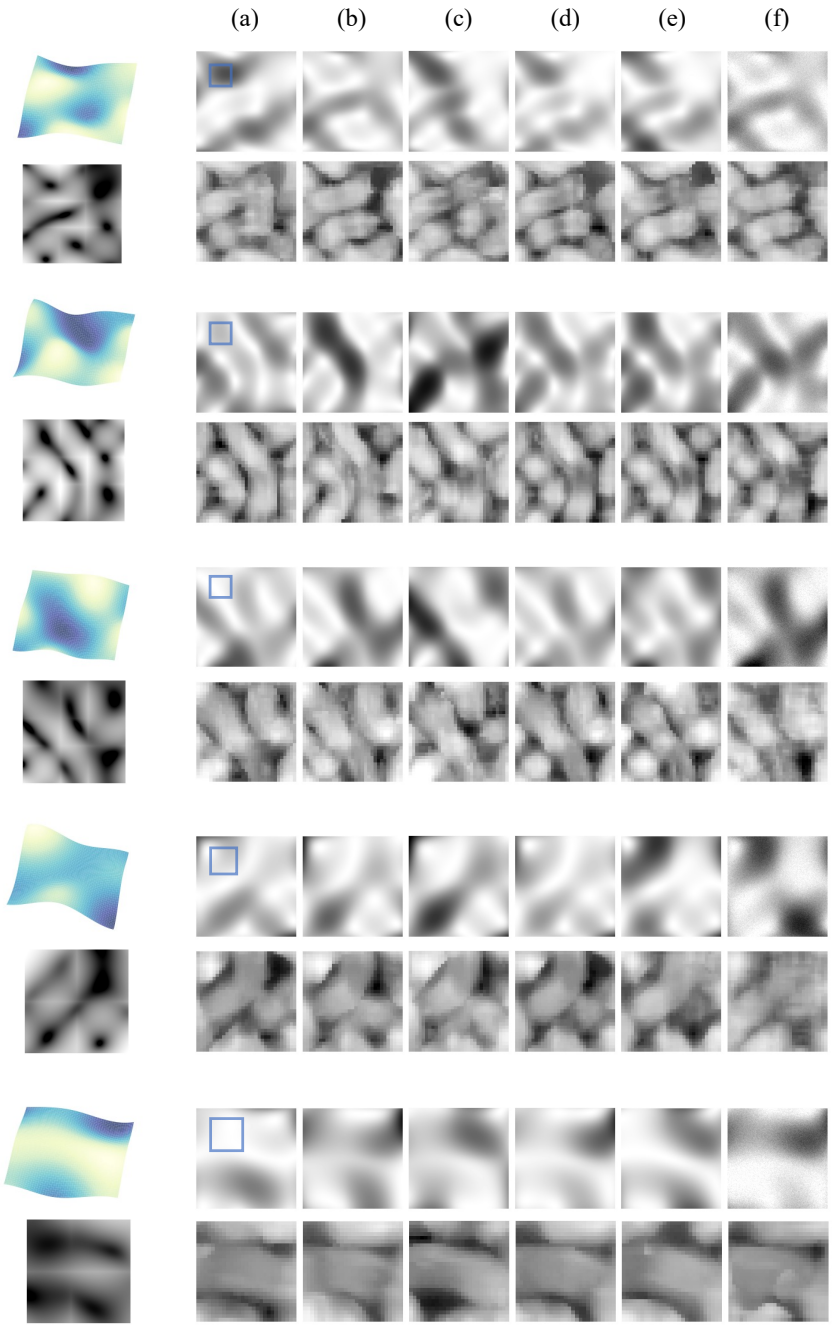


Figure 10: Ground truth and estimated log-Casorati curvature field visualizations. Column (a)-(c) are single directional light source. Column (d)-(e) are spatially varying light sources. Column (f) has added Gaussian noise with variance = $4e^{-4}$. The blue box indicates the patch size used for each row of test images.

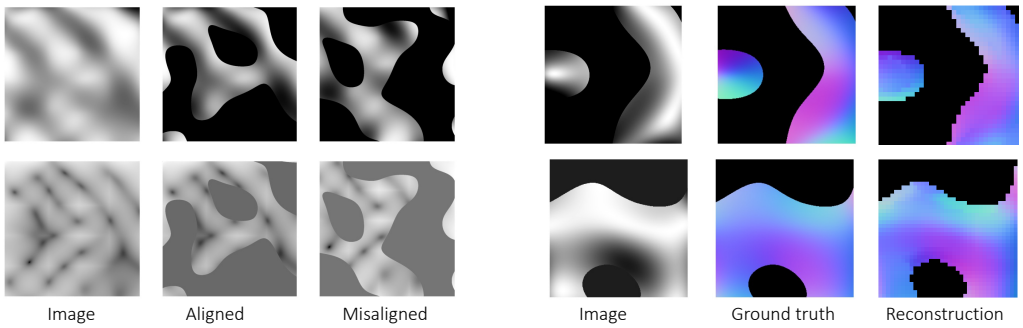


Figure 11: Left: The planar cut applied to the shading image and underlying surface's curvature field. Right: Normal reconstruction of two shaded images given consistent planar cut boundary.

of along the contour. Therefore the inconsistent shading cue is overridden by the contour information and thus we no longer perceive a diffusely-shaded 3D surface.

When the planar cut is indeed consistent with shading, our model can output a reconstruction of the normal field using a simple heuristic. We first initialize the normal field with the output from the overlapping patches by applying 6. Using the estimated curvature field and previous observations, we apply the following steps. For high curvature regions, we select the 'convex' shape explanation from the 4-way choice of the normal field output. For low curvature regions, we first aggregate the boundary information in its neighboring region. Then we determine from the 4-way choice described in equation 10 the normals that has smallest angle with the boundary normals. An interpolated normal from the two is assigned to those pixels. At last we apply spatial smoothing along the low curvature regions. Examples of the reconstruction outcome are shown in Fig.11 right. The top surface's estimated normal field has mean angular error (MAE) = 17.1° and the bottom result has MAE = 9.7°.