Normal Integration via Inverse Plane Fitting with Minimum Point-to-Plane Distance

Xu Cao[1] Boxin Shi[2,3] Fumio Okura[1] Yasuyuki Matsushita[1]
[1]Osaka University [2]Peking University [3]Peng Cheng Laboratory

Abstract

This paper presents a surface normal integration method that solves an inverse problem of local plane fitting. Surface reconstruction from normal maps is essential in photometric shape reconstruction. To this end, we formulate normal integration in the camera coordinates and jointly solve for 3D point positions and local plane displacements. Unlike existing methods that consider the vertical distances between 3D points, we minimize the sum of squared point-to-plane distances. Our method can deal with both orthographic or perspective normal maps with arbitrary boundaries. Compared to existing normal integration methods, our method avoids the checkerboard artifact and performs more robustly against natural boundaries, sharp features, and outliers. We further provide a geometric analysis of the source of artifacts that appear in previous methods based on our plane fitting formulation. Experimental results on analytically computed, synthetic, and real-world surfaces show that our method yields accurate and stable reconstruction for both orthographic and perspective normal maps[1].

1. Introduction

Reconstructing a surface from its surface normal map, a problem called normal integration [28], is essential for photometric 3D reconstruction, such as shape from shading [19], photometric stereo [34], and shape from polarization [20]. Among all approaches to normal integration, variational-based methods have been studied for many years [14, 16, 28], which are based on functional optimization. The objectives are turned into linear systems by either discretizing the functional itself [14] or its optimal condition [28, 29] in the form of discrete Poisson's equation. Recently, a non-variational approach based on discrete geometry processing (DGP) [37] was proposed for normal integration. DGP also solves a linear system to estimate the surface. For all the three linear systems, however, their least squares approximate solutions suffer from several problems. For discrete functional, a checkerboard artifact was reported in [38], as shown in Fig. 1 top-left. Discrete Poisson's equation and DGP are sensitive to sharp features or outliers, as shown in Fig. 1 top-middle and top-right.

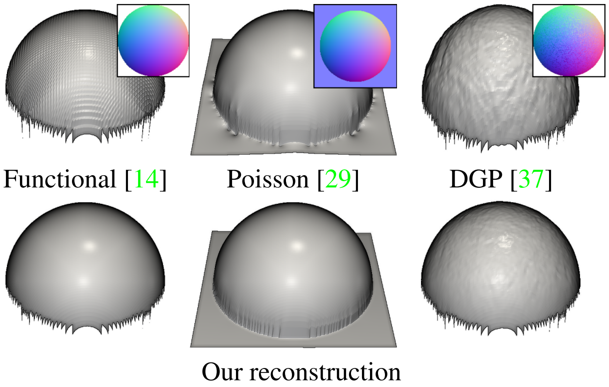

Figure 1. **(Top)** Typical problems in solving the linear systems for normal integration. From left to right: The checkerboard artifact, Gibbs phenomenon near sharp features, and sensitivity to outliers. **(Bottom)** Our results from the same normal maps as above. There is no checkerboard artifact, Gibbs phenomenon almost disappears, and outliers hardly distort the surface.

Various remedies have been studied to improve the robustness against these artifacts. First, regularization [13, 14] or smoothness term [38] was used to refine the reconstruction quality. However, introducing a regularization term requires tuning a weighting factor, increasing the effort to obtain a satisfactory result. Second, 2-norm of the residual vector was replaced by p-norm [6] to increase the robustness; however, it comes with heavier computation. Third, normal vectors with small elevation angles was treated as outliers [37], but the heuristic thresholding does not ensure the robustness to outliers.

Unlike the existing approaches, we focus on the residual vector of the linear systems. Discretizing the functional or Poisson's equation by finite difference leads to measuring the difference between the height values of 3D points. The difference is therefore the *vertical distance* (*i.e.*, the

[1] Source code is available at https://github.com/hoshino042/NormalIntegration.

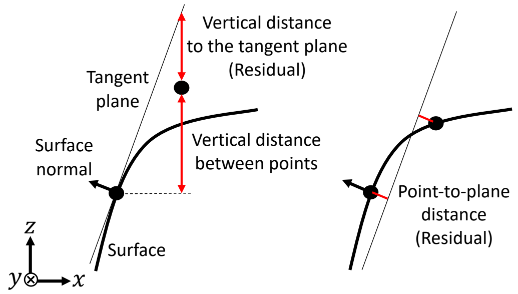

Figure 2. **(Left)** Vertical distance, *i.e*., distance along the viewing direction and **(Right)** point-to-plane distance, *i.e*., distance along the plane normal (Ours).

distance along the viewing direction) between 3D points, as illustrated in Fig. 2 left. DGP also explicitly computes the vertical distance in its local shaping step[2]. These systems then equate the vertical distance between points to the measured gradient. Geometrically, the residual therefore represents the vertical distance from the estimated point to the tangent plane at its neighboring point.

This paper introduces the perpendicular *point-to-plane distance* (*i.e*., the distance along the plane normal direction) to the problem of normal integration, as illustrated in Fig. 2 right. We formulate normal integration as an inverse problem of local plane fitting in the camera coordinates. To measure point-to-plane distances, we introduce plane displacements as unknowns, in addition to the points' height values. From a geometric perspective, our method simultaneously moves points on camera rays and planes along their normal directions to minimize the sum of squared point-to-plane distances. Thanks to the formulation's flexibility, it is straightforward to apply our method to orthographic and perspective normal maps with arbitrary boundaries, or to estimate height values on a different domain, such as the one used by DGP [37]. As a result, our method avoids the checkerboard artifact and performs robustly against sharp features or outliers, as exemplified in the bottom row of Fig. 1.

Our contributions are summarized as follows

- We propose a new method for the problem of normal integration by jointly optimizing point positions and local plane displacements in the camera coordinates. To our knowledge, this is the first approach performing normal integration based on point-to-plane distances.
- We show the source of the checkerboard artifact in the central-difference discretized functional by evolving our plane fitting formulation.
- We point out that discrete Poisson's equation and DGP share the same intrinsic principle thus perform similarly, and demonstrate that our method is more robust to outliers and sharp features than these methods.

2. Related Work

Non-variational methods. Early works solve the problem of normal integration by path integration [10, 30, 35]. Another approach is to enforce surface integrability by projecting the gradient field onto the space spanned by a set of basis functions, *e.g*., Fourier basis functions [9], cosine functions [11], orthonormal vectors [21], or shapelets [23]. Recently, Xie *et al*. [37] proposed a discrete geometry processing based normal integration method, which recovers a discrete polygonal mesh directly from normal maps. Their follow-up works dealt with discontinuity preserving [36] and unconnected normal maps [33].

Section 5 demonstrates that the residual vector of DGP's linear system has the same geometric meaning as that of discrete Poisson's equation; therefore, DGP and discrete Poisson's equation perform similarly.

Variational methods. Pioneered by Ikeuchi [18] and detailed by Horn and Brooks [16], the problem of normal integration has been formulated in a variational manner by optimizing a functional. Agrawal *et al*. [1] proposed a general framework to construct a range of solutions to the functional. Instead of minimizing the least squares functional, different kinds of regularizations were studied to increase robustness, like L1 norm [6, 27]. Depth measurements were also used as prior knowledge of the surface [2, 17, 25]. How to better discretize the functional is also an essential issue in practice [3, 28]. Further, boundary conditions should be carefully handled [8]. A typical method for the discretization is to vectorize the height map and to approximate the partial derivative by finite difference. Zhu and Smith [38] used 2D Savitzky-Golay filters to approximate partial derivative. Harker and O'Leary [12, 14] discretized the functional into a Sylvester equation without vectorizing the height map. Our method's critical difference from the variational methods is that our linear system's residual vector measures the point-to-plane distance.

Two methods have been studied to reconstruct surfaces from perspective normal maps. First, Durou *et al*. [7, 8, 28] integrated the perspective normal map by taking the logarithm of height values. They then solved the discrete Poisson's equation similar to the orthographic case, and exponentiate the result to recover the desired height map. Second, Nehab *et al*. [25], as well as Zhu and Smith [38], derived the relation between the perspective normal map and the height map based on that the normal vector should be perpendicular to the surface tangent vectors. To our knowledge, our inverse plane fitting formulation provides a third alternative to solve the problem of normal integration in perspective case.

[2] See Fig. 3 of [37].

3. Problem Statement

We first briefly review the definition of and conventional methods for the problem of normal integration. A normal map $\mathbf{n} : \Omega_n \to \mathcal{S}^2 \subset \mathbb{R}^3$ records the surface orientation at a discrete pixel coordinate $\mathbf{u} = [u, v]^\top \in \Omega_n$, where the domain $\Omega_n \subset \{1, \dots, H\} \times \{1, \dots, W\}$ is defined by the image of dimension $H \times W$. Let $\mathbf{p}(\mathbf{u}) = [x(\mathbf{u}), y(\mathbf{u}), z(\mathbf{u})]^\top$ be the 3D surface point and $\mathbf{n}(\mathbf{u}) = [n_x(\mathbf{u}), n_y(\mathbf{u}), n_z(\mathbf{u})]^\top \in \mathcal{S}^2$ be the unit surface normal vector corresponding to the pixel coordinate $\mathbf{u}$. The problem of normal integration is to reconstruct the height map $z(\cdot)$ from the normal map $\mathbf{n}(\cdot)$.

Variational methods formulate normal integration as *height/shape from gradient*, *i.e.*, recovering a height map from a measured gradient field $[p, q]^\top$. Under orthographic projection,

$$p = -\frac{n_x}{n_z}, \quad q = -\frac{n_y}{n_z}. \tag{1}$$

We refer the reader to [28] for detailed derivation of $[p, q]^\top$ under different types of projections. Variational methods then find the optimal surface by minimizing the following functional

$$\mathcal{J}(z) = \iint_{\Omega_n} (\partial_u z - p)^2 + (\partial_v z - q)^2 \, du \, dv, \tag{2}$$

where $\partial_u \cdot$ and $\partial_v \cdot$ denote partial derivatives of the function $z : \Omega_n \to \mathbb{R}$ along u and v axes on the image plane. Applying the Euler-Lagrange equation to the functional Eq. (2) leads to the necessary optimal condition

$$\Delta z = \mathrm{div}(p, q) = \partial_u p + \partial_v q, \tag{3}$$

which is in the form of Poisson's equation. Equations (2) and (3) consider the problem of normal integration on a continuous domain. As the normal map is measured on a discrete domain, previous methods either discretize the functional [14] or Poisson's equation [28].

We show in Section 5 that these conventional discretization methods minimize a residual by measuring the *vertical distance* along viewing directions. On the other hand, our method minimizes the *point-to-plane distance* along surface normal directions.

4. Normal Integration via Inverse Plane Fitting

Our main idea is to approximate the surface locally by planes perpendicular to normal vectors. We consider the form of a plane equation in a 3D space as

$$\mathbf{p}^\top \mathbf{n} + d = 0, \tag{4}$$

where $\mathbf{n} \in \mathcal{S}^2$ is the unit normal vector perpendicular to the plane, and d is the distance from the coordinate origin to the plane along the plane's normal direction. We will call d the *plane displacement* throughout this paper. All points $\mathbf{p} \in \mathbb{R}^3$ on the plane should satisfy Eq. (4).

Suppose we are given a set of normal vectors $\{\mathbf{n}(\mathbf{u})\}$ over the pixel coordinates. For each normal vector $\mathbf{n}(\mathbf{u})$, there is a family of planes parallel to the tangent plane, parameterized by its displacement $d(\mathbf{u})$ as

$$\mathbf{p}^\top \mathbf{n}(\mathbf{u}) + d(\mathbf{u}) = 0. \tag{5}$$

As the surface is locally approximated by planes, we enforce nearby points of the point $\mathbf{p}(\mathbf{u})$ that correspond to pixel $\mathbf{u}$ to lie on a plane as

$$\min \sum_{\mathbf{p} \in \mathcal{N}(\mathbf{p}(\mathbf{u}))} (\mathbf{p}^\top \mathbf{n}(\mathbf{u}) + d(\mathbf{u}))^2, \tag{6}$$

where $\mathcal{N}(\mathbf{p}(\mathbf{u}))$ denotes the set of nearest neighbor points to $\mathbf{p}(\mathbf{u})$. Equation (6) is exactly the same as PlaneSVD [22], a method for estimating normals from known 3D points by fitting planes. In the problem of normal integration, normal vectors $\mathbf{n}$ are given, and the 3D points $\mathbf{p}$ are to be estimated, which is why we term our method as "inverse plane fitting". To recover the 3D points, we need to jointly consider plane equations at all points as

$$\min_{\mathbf{p}, d} \sum_{\mathbf{u} \in \Omega_n} \sum_{\mathbf{p} \in \mathcal{N}(\mathbf{p}(\mathbf{u}))} (\mathbf{p}^\top \mathbf{n}(\mathbf{u}) + d(\mathbf{u}))^2. \tag{7}$$

To solve this optimization, we introduce two assumptions. First, we will assume that the 2D projections of all 3D points on the image plane are known. This constraint reduces the degree of freedom of points from 3 to 1, *i.e.*, all points can only move along known camera rays. We can then parameterize a point $\mathbf{p}$ as $\mathbf{p}(z; \mathbf{u}_\mathbf{p})$, where $\mathbf{u}_\mathbf{p}$ is the known projection of the 3D point on the image plane, and z is the 3D point's unknown position on the camera ray. Second, we will assume that the closeness relationship between 3D points remains the same for their projections on the image plane. That is,

$$\mathcal{N}(\mathbf{p}(\mathbf{u})) = \{\mathbf{p}(z; \mathbf{u}_\mathbf{p}) \mid \mathbf{u}_\mathbf{p} \in \mathcal{N}(\mathbf{u})\}. \tag{8}$$

This assumption holds for continuous surfaces. With the two assumptions, the objective Eq. (7) becomes

$$\min_{z, d} \sum_{\mathbf{u} \in \Omega_n} \sum_{\mathbf{u}_\mathbf{p} \in \mathcal{N}(\mathbf{u})} (\mathbf{p}(z; \mathbf{u}_\mathbf{p})^\top \mathbf{n}(\mathbf{u}) + d(\mathbf{u}))^2. \tag{9}$$

There is still one issue left for optimizing Eq. (9), namely, definition of the set of $\mathbf{u}_\mathbf{p}$, which reflects the points of interest we will solve. We show two design options in this paper. The first obvious choice is to let $\{\mathbf{u}_\mathbf{p}\} = \Omega_n$, which is the same as conventional variational methods. The second choice is to place $\mathbf{u}_\mathbf{p}$ a half-a-pixel step from $\mathbf{u}$ in

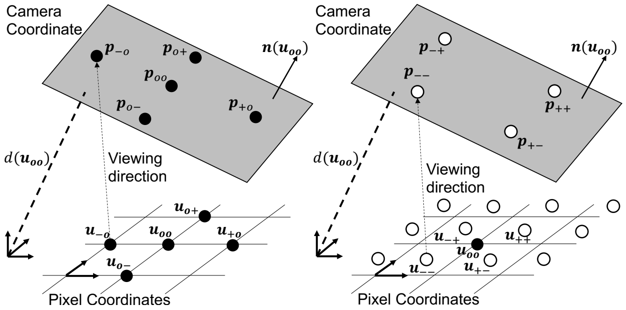

Figure 3. **(Left)** A local five-point stencil in domain Ω_n on the image plane. We enforce all points projected onto this stencil to lie on a plane with the center normal vector and an unknown displacement. **(Right)** The domain Ω_z used by DGP [37], denoted as circles. Each pixel in Ω_n has four neighbor pixels in Ω_z.

Ω_n, as illustrated in Fig. 3 right. The second one is the same as that used by DGP [37], and we denote it as Ω_z.

To retrieve the nearest neighbor projections of $\mathbf{u}$, we use ball query with radius 1. In the case of $\{\mathbf{u}_\mathbf{p}\} = \Omega_n$, this query results in a five-point stencil for pixels whose four neighbor pixels are also in Ω_n, as shown in Fig. 3 left. We use a subscript "oo" to indicate the center pixel $\mathbf{u}$, its corresponding 3D point $\mathbf{p}$, and its height value z. Subscripts "-o", "+o", "o+", "o-" are used for its left / right / upper / lower counterparts in the local five-point stencil. For boundary or corner pixels, $\mathcal{N}(\mathbf{u})$ has four or three elements, respectively. The inner summation in Eq. (9) now becomes

$$\begin{aligned}&\sum_{\mathbf{u}_\mathbf{p}\in\mathcal{N}(\mathbf{u})}(\mathbf{p}(z;\mathbf{u}_\mathbf{p})^\top\mathbf{n}(\mathbf{u})+d(\mathbf{u}))^2\\&=\|[\mathbf{p}_{\text{oo}}\quad\mathbf{p}_{\text{o+}}\quad\mathbf{p}_{\text{o-}}\quad\mathbf{p}_{\text{+o}}\quad\mathbf{p}_{\text{-o}}]^\top\mathbf{n}(\mathbf{u})+d(\mathbf{u})\mathbf{1}\|_2^2,\end{aligned}\tag{10}$$

where $\mathbf{1}$ is an all-ones vector.

In the case of $\{\mathbf{u}_\mathbf{p}\} = \Omega_z$, $\mathcal{N}(\mathbf{u})$ of all pixels has four elements, no matter whether it is a boundary pixel or not. Similarly, the inner summation in Eq. (9) becomes

$$\begin{aligned}&\sum_{\mathbf{u}_\mathbf{p}\in\mathcal{N}(\mathbf{u})}(\mathbf{p}(z;\mathbf{u}_\mathbf{p})^\top\mathbf{n}(\mathbf{u})+d(\mathbf{u}))^2\\&=\|[\mathbf{p}_{\text{++}}\quad\mathbf{p}_{\text{+-}}\quad\mathbf{p}_{\text{-+}}\quad\mathbf{p}_{\text{--}}]^\top\mathbf{n}(\mathbf{u})+d(\mathbf{u})\mathbf{1}\|_2^2,\end{aligned}\tag{11}$$

From now on, we will call five-point and four-point plane fitting for the case $\{\mathbf{u}_\mathbf{p}\} = \Omega_n$ and $\{\mathbf{u}_\mathbf{p}\} = \Omega_z$, respectively.

4.1. Solution Method

Orthographic case. Under a (scaled) orthographic projection or weak perspective projection, a surface point $\mathbf{p}(z;\mathbf{u}_\mathbf{p})$ is projected onto the image plane with a scaling along xy-axes, *i.e.*, $\mathbf{p}(z;\mathbf{u}_\mathbf{p}) = [h\mathbf{u}_\mathbf{p}^\top, z]^\top$. The scale factor h transfers the camera's xy-coordinates to the pixel coordinates; in the context of finite difference, h is called the step size. Without loss of generality, we will assume $h = 1$ in this paper for notational simplicity. Inserting the point representation into plane equation of Eq. (10) (inner part of the norm) and moving known terms to the right-hand side, we obtain the system of linear equations at one pixel as

$$\begin{bmatrix}n_z&&&&&1\\&n_z&&&&1\\&&n_z&&&1\\&&&n_z&&1\\&&&&n_z&1\end{bmatrix}\begin{bmatrix}z_{\text{oo}}\\z_{\text{o+}}\\z_{\text{o-}}\\z_{\text{+o}}\\z_{\text{-o}}\\d\end{bmatrix}=-\begin{bmatrix}\mathbf{u}_{\text{oo}}^\top\\\mathbf{u}_{\text{o+}}^\top\\\mathbf{u}_{\text{o-}}^\top\\\mathbf{u}_{\text{+o}}^\top\\\mathbf{u}_{\text{-o}}^\top\end{bmatrix}\begin{bmatrix}n_x\\n_y\end{bmatrix}.\tag{12}$$

Stacking equations at all pixels results in a sparse, overdetermined linear system $\mathbf{A}\left[\begin{smallmatrix}\mathbf{z}\\\mathbf{d}\end{smallmatrix}\right] = \mathbf{b}$. The number of rows of $\mathbf{A}$ or the number of all plane equations is $\sum_{\mathbf{u}\in\Omega_n}|\mathcal{N}(\mathbf{u})|$, which is close to $5|\Omega_n|$; the number of columns of $\mathbf{A}$ is $2|\Omega_n|$, as $\mathbf{z}$ and $\mathbf{d}$ are stacks of $|\Omega_n|$ height values and plane displacements, respectively. We can derive a least squares approximate solution by minimizing the following residual

$$\min_{\mathbf{z},\mathbf{d}}\|\mathbf{e}\|_2^2\quad\text{s.t.}\quad\mathbf{e}=\mathbf{A}\left[\begin{smallmatrix}\mathbf{z}\\\mathbf{d}\end{smallmatrix}\right]-\mathbf{b}.\tag{13}$$

Similarly, if we work on the domain Ω_z, we can derive the same plane equation as Eq. (12) but with four equations at each pixel. The coefficient matrix $\mathbf{A}$'s dimension is $4|\Omega_n| \times (|\Omega_n| + |\Omega_z|)$, as there are $|\Omega_n|$ displacements and $|\Omega_z|$ height values.

Perspective case. Suppose a perspective pinhole camera with intrinsics $\mathbf{K}$ projects a surface point $\mathbf{p}(z;\mathbf{u}_\mathbf{p})$ onto the image plane at $\mathbf{u}_\mathbf{p}$. Its back-projection is a camera ray containing $\mathbf{p}(\mathbf{u})$ with an unknown scale z as

$$\mathbf{p}(z;\mathbf{u}_\mathbf{p}) = z\mathbf{K}^{-1}\tilde{\mathbf{u}}_\mathbf{p} \equiv z\tilde{\mathbf{p}},\tag{14}$$

where $\tilde{\mathbf{u}}_\mathbf{p} = [u, v, 1]^\top$ is $\mathbf{u}_\mathbf{p}$ in homogeneous coordinates, $\tilde{\mathbf{p}}$ is the back-projected point $\mathbf{K}^{-1}\tilde{\mathbf{u}}_\mathbf{p}$. In the case of $\{\mathbf{u}_\mathbf{p}\} = \Omega_n$, inserting Eq. (14) into the plane equation of Eq. (10) yields five equations at one pixel as

$$\begin{bmatrix}\mathbf{n}^\top\tilde{\mathbf{p}}_{\text{oo}}&&&&&1\\&\mathbf{n}^\top\tilde{\mathbf{p}}_{\text{o+}}&&&&1\\&&\mathbf{n}^\top\tilde{\mathbf{p}}_{\text{o-}}&&&1\\&&&\mathbf{n}^\top\tilde{\mathbf{p}}_{\text{+o}}&&1\\&&&&\mathbf{n}^\top\tilde{\mathbf{p}}_{\text{-o}}&1\end{bmatrix}\begin{bmatrix}z_{\text{oo}}\\z_{\text{o+}}\\z_{\text{o-}}\\z_{\text{+o}}\\z_{\text{-o}}\\d\end{bmatrix}=\mathbf{0}.\tag{15}$$

In the case of $\{\mathbf{u}_\mathbf{p}\} = \Omega_z$, there are four similar equations at each pixel. Stacking all equations yields a sparse, homogeneous system $\mathbf{A}\left[\begin{smallmatrix}\mathbf{z}\\\mathbf{d}\end{smallmatrix}\right] = \mathbf{0}$. The dimension of $\mathbf{A}$ is the same as that in the corresponding orthographic case. Then our objective using least squares becomes

$$\min_{\mathbf{z},\mathbf{d}}\|\mathbf{e}\|_2^2\quad\text{s.t.}\quad\mathbf{e}=\mathbf{A}\left[\begin{smallmatrix}\mathbf{z}\\\mathbf{d}\end{smallmatrix}\right].\tag{16}$$

The non-trivial solution to Eq. (16) is the right-singular vector corresponding to the smallest singular value of $\mathbf{A}$[3].

[3] See, for example, A5.3 of [15] for a detailed proof.

Rank, ambiguity, and solver. The system matrix $\mathbf{A}$ is rank deficient in the orthographic case and has full-rank in the perspective case. In orthographic case, $\text{nullity}(\mathbf{A}) = 1$; the null vector is $\begin{bmatrix} \mathbf{1} \\ -\mathbf{n}_z^{\circ -1} \end{bmatrix}$, where $\mathbf{n}_z^{\circ -1}$ is Hadamard inverse of the stacking vector of all n_z. The null vector can be verified in Eq. (12), and implies an offset ambiguity in the solution. Geometrically understanding, if all points are moved by 1 unit along the camera rays, the best fit planes are moved by $-n_z^{-1}$ unit along planes' normal directions. We used LSQR [26] to solve Eq. (13) in this paper. When the size of the system matrix $\mathbf{A}$ becomes large, solving the normal equation of Eq. (13) by multigrid methods [4] can be more efficient.

In perspective case, the solution is up to a scale ambiguity as we are solving a homogeneous system Eq. (16). The system matrix $\mathbf{A}$ can be full-rank because the residual of plane fitting never achieves 0 unless the surface is a plane. This intuition is consistent with our observation that $\mathbf{A}$ is only rank-deficient when the surface is a plane.

To efficiently find the solution to Eq. (16), we can compute the eigenvector corresponding to the smallest eigenvalue of $\mathbf{A}^\top \mathbf{A}$, as $\mathbf{A}$'s right singular vectors are $\mathbf{A}^\top \mathbf{A}$'s eigenvectors. We used ARPACK's implementation [24] of implicitly restarted Lanczos method [32], which is efficient at finding a few eigenvectors of a large sparse matrix.

5. Robustness Analysis and Comparison

This section analyzes three previous orthographic normal integration methods in the language of our plane fitting formulation: Discrete functional [14], discrete Poisson's equation [29], and DGP [37]. We begin by describing the source of the *checkerboard artifact* that appears with central-difference-discretized functional. We then show the residual vectors of the discrete Poisson's equation and DGP share the same geometric meaning.

Discrete functional. Central difference is typically used to discretize the functional Eq. (2), yielding two equations at one pixel location as

$$\begin{cases} \frac{1}{2}(z_{\text{+o}} - z_{\text{-o}}) = p = -\frac{n_x}{n_z} \\ \frac{1}{2}(z_{\text{o+}} - z_{\text{o-}}) = q = -\frac{n_y}{n_z} \end{cases} . \tag{17}$$

Solving the linear system based on Eq. (17) causes the checkerboard artifact, as shown in Fig. 4 (e). One speculation about the source of the artifact is because Eq. (17) is independent of the center point z_{oo} [38]. We argue that it is in fact due to the loss of constraints between the left / right and upper / lower point pairs. And as long as we specify their relations, the checkerboard artifact can be avoided even if there is no constraint using the center point.

To validate our argument, we gradually manipulate our five-point plane fitting equations Eq. (12) to the discrete functional Eq. (17). We first discard the plane constraint on the center point $\mathbf{p}_{\text{oo}}$ in Eq. (12), yielding four equations at one pixel as

$$\begin{cases} (u-1)n_x + vn_y + z_{\text{-o}}n_z + d = 0 & \text{(left)} \\ (u+1)n_x + vn_y + z_{\text{+o}}n_z + d = 0 & \text{(right)} \\ un_x + (v+1)n_y + z_{\text{o+}}n_z + d = 0 & \text{(upper)} \\ un_x + (v-1)n_y + z_{\text{o-}}n_z + d = 0 & \text{(lower)} \end{cases} . \tag{18}$$

Equation (18) enforces four points on one plane, but now let us relax the constraint as

$$\begin{cases} (u-1)n_x + vn_y + z_{\text{-o}}n_z + d = 0 & \text{(left)} \\ (u+1)n_x + vn_y + z_{\text{+o}}n_z + d = 0 & \text{(right)} \\ un_x + (v+1)n_y + z_{\text{o+}}n_z + d' = 0 & \text{(upper)} \\ un_x + (v-1)n_y + z_{\text{o-}}n_z + d' = 0 & \text{(lower)} \end{cases} , \tag{19}$$

which means we fit two parallel planes for the four points. In other words, there could be a local offset (*i.e.*, $d \neq d'$) between the left / right and upper / lower point pairs. However, we can derive the same equations from Eqs. (18) and (19) if we take the difference between the left and right, upper and lower point [4]

$$\begin{cases} \frac{n_z}{2}(z_{\text{+o}} - z_{\text{-o}}) = -n_x \\ \frac{n_z}{2}(z_{\text{o+}} - z_{\text{o-}}) = -n_y \end{cases} . \tag{20}$$

Dividing both sides by n_z, we then arrive at the discrete functional Eq. (17).

Figure 4 displays the estimated surfaces by the linear systems based on the equations in this evolution. From Fig. 4 (a) to (b), we discard the constraint on the center point. But it does not create the checkerboard artifact, and the reason is that the four neighbor points are still enforced on the same plane. From Fig. 4 (b) to (c), the checkerboard artifact occurs, implying that relaxing the constraint from Eq. (18) to Eq. (19) introduces the

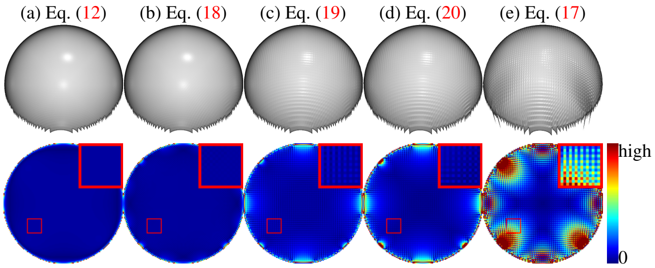

Figure 4. An evolution from plane fitting to central-difference discretized functional. Each column shows the estimated surface and the absolute difference map. Central difference implicitly allows a local offset in each five-point stencil (*i.e.*, fit two parallel planes at one pixel), introducing the checkerboard artifact.

[4] This formulation is the same as Zhu and Smith's method [38] in the orthographic case with central-difference-based discretization.

checkerboard artifact. Further, the checkerboard artifact between Fig. 4 (c) and (d) is indistinguishable, validating that central difference implicitly allows a local offset between the left / right and upper / lower point pairs in each five-point stencil. The checkerboard artifact in Fig. 4 (d) is weaker than that in (e) because Eq. (20) transfers the vertical distance to the distance along plane normal direction (multiplied by n_z). This phenomenon further validates that measuring distance along normal directions (Eq. (20)) is better than that along viewing directions (Eq. (17)).

Discrete Poisson. Poisson's equation (3) is typically discretized at a five-point stencil [1, 8, 28] as

$$z_{\text{+o}} + z_{\text{-o}} + z_{\text{o+}} + z_{\text{o-}} - 4z_{\text{oo}} = \frac{p_{\text{+o}} - p_{\text{-o}}}{2} + \frac{q_{\text{o+}} - q_{\text{o-}}}{2}. \tag{21}$$

Quéau *et al.* [29] point out it is equivalent to consider all forward / backward differences at one pixel

$$\begin{cases} z_{\text{+o}} - z_{\text{oo}} = p_{\text{oo}} \\ z_{\text{oo}} - z_{\text{-o}} = p_{\text{oo}} \end{cases} \quad \begin{cases} z_{\text{o+}} - z_{\text{oo}} = q_{\text{oo}} \\ z_{\text{oo}} - z_{\text{o-}} = q_{\text{oo}} \end{cases}. \tag{22}$$

The least squares approximate solutions of two linear systems constructed by stacking Eq. (21) or (22) are the same[5]. We will use Eq. (22) to demonstrate the geometric meaning of the residual of discrete Poisson's equation.

Consider a center and its right-hand point lying on the tangent plane of the center point

$$\begin{cases} un_x + vn_y + \hat{z}_{\text{oo}}n_z + d = 0 & \text{(center)} \\ (u+1)n_x + vn_y + \hat{z}_{\text{+o}}n_z + d = 0 & \text{(right)} \end{cases}. \tag{23}$$

We denote by $\hat{z}$ to stress they are the height values of the intersections of camera rays and the tangent plane, not the ones by solving the linear system. Taking the difference between the two equations yields

$$\hat{z}_{\text{+o}} - \hat{z}_{\text{oo}} = -\frac{n_x}{n_z} = p_{\text{oo}}, \tag{24}$$

which reveals that the gradient p_{oo} is equal to the vertical distance between two neighbor intersections. We can derive similar equivalence for other pixel pairs in the five-point stencil and plug them into Eq. (22) as

$$\begin{cases} z_{\text{+o}} - z_{\text{oo}} = \hat{z}_{\text{+o}} - \hat{z}_{\text{oo}} \\ z_{\text{oo}} - z_{\text{-o}} = \hat{z}_{\text{oo}} - \hat{z}_{\text{-o}} \end{cases} \quad \begin{cases} z_{\text{o+}} - z_{\text{oo}} = \hat{z}_{\text{o+}} - \hat{z}_{\text{oo}} \\ z_{\text{oo}} - z_{\text{o-}} = \hat{z}_{\text{oo}} - \hat{z}_{\text{o-}} \end{cases}. \tag{25}$$

These equations imply that discrete Poisson's equation enforces the equality between the vertical distance of estimated neighbor points and neighbor tangent intersections.

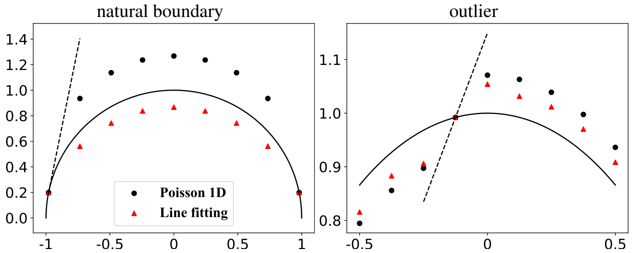

Figure 5. Poisson equation v.s. line fitting on a 2D semi-circle. **(Left)** The boundary point has a large tangent slope (dashed line), deviating estimated points to the tangent along viewing direction. **(Right)** We rotated one normal vector as an outlier, whose tangent line is shown as the dashed line. Poisson's equation performs more sensitive to the outlier than our line fitting.

The residuals are then the vertical distances from the estimated point to the tangent at its neighbor points.

To demonstrate why this residual is sensitive to natural boundary or outliers, we compared Poisson's equation and plane fitting on a 1D function in Fig. 5. Plane fitting degrades to line fitting with the line equation $xn_x + yn_y + d = 0$. We used a circle function $y = \sqrt{1-x^2}$ for $x \in [-1, 1]$ as the ground truth surface. In Fig. 5 left, we sampled 9 normal vectors at points evenly distributed in $[-0.98, 0.98]$ and estimate the height values of the 9 points. The two boundary points are close to the surface's natural boundary, so the corresponding tangent has a large slope. The points estimated by discrete Poisson's equation deviate largely toward the tangent along viewing direction. The points estimated by line fitting, on the other hand, are close to the surface since the point-to-plane distance would not change largely by moving them on the camera rays.

In Fig. 5 right, we evenly sampled 9 points in $[-0.5, 0.5]$, *i.e.*, there is no effect of the natural boundary. However, we rotated one normal vector as an outlier, whose tangent line is shown as the dashed line. Due to the large vertical distance introduced by the outlier normal vector, the points estimated by discrete Poisson's equation deviate more towards the tangent than line fitting.

Discrete Geometry Processing (DGP). We will show that the residual of DGP [37] is essentially the same as that of discrete Poisson's equation, but on domain Ω_z. DGP consists of two steps: (1) local shaping and (2) global blending. Local shaping defines a local 3D coordinate centered at $\mathbf{u}_{\text{oo}}$ on the image plane, and then computes the *vertical distance* from a neighbor pixel to the plane with normal vector $\mathbf{n}_{\text{oo}}$ passing through $\mathbf{u}_{\text{oo}}$ (*i.e.*, $d = 0$). For example, the vertical

[5] Equation (21) should be modified for boundary pixels [3].

distance $\hat{z}_{++}^{(\text{local})}$ of the upper-right pixel $\mathbf{u}_{++}$ is computed as

$$\begin{aligned} &0.5n_x + 0.5n_y + n_z \hat{z}_{++}^{(\text{local})} = 0 \\ \rightarrow\ &\hat{z}_{++}^{(\text{local})} = -\frac{0.5n_x + 0.5n_y}{n_z}. \end{aligned} \tag{26}$$

Global blending then equates the difference between a point and the average of its four neighbor points to the local vertical distance as

$$z_{++} - \frac{z_{++} + z_{+-} + z_{-+} + z_{--}}{4} = \hat{z}_{++}^{(\text{local})}. \tag{27}$$

The vertical distance in the local coordinates is the same as the vertical distance in the camera coordinates, *i.e.*, $\hat{z}_{++}^{(\text{local})} = \hat{z}_{++} - \hat{z}_{\text{oo}}$. Let z_{oo} be the average of four neighbor height values, then Eq. (27) can be simplified and all four equations at one pixel location now read

$$\begin{cases} z_{++} - z_{\text{oo}} = \hat{z}_{++} - \hat{z}_{\text{oo}} \\ z_{+-} - z_{\text{oo}} = \hat{z}_{+-} - \hat{z}_{\text{oo}} \end{cases} \quad \begin{cases} z_{-+} - z_{\text{oo}} = \hat{z}_{-+} - \hat{z}_{\text{oo}} \\ z_{--} - z_{\text{oo}} = \hat{z}_{--} - \hat{z}_{\text{oo}} \end{cases}. \tag{28}$$

This form is similar to discrete Poisson's equation in Eq. (25), except that the height values are defined on the domain Ω_z. Therefore the residual of DGP has the same geometric meaning as that of discrete Poisson's equation.

Figure 6 compares discrete Poisson's equation, DGP, and our methods on two analytically computed normal maps; see our supplementary material for the strategy of computing the surface height/normal maps and adding noise/outliers. As there is a global offset ambiguity in the estimated height maps, we first found an optimal offset between the estimated and the ground truth height maps. We then computed the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as evaluation metrics. Discrete Poisson's equation and DGP show an indistinguishable error pattern and achieve close RMSE / MAE. This phenomenon validates our understanding that the residual vectors of DGP and discrete Poisson's equation have the same geometric meaning. On the other hand, our method performs more robustly to outliers and sharp features.

6. Comparison of Perspective Methods

Figure 7 compares state-of-the-art perspective normal integration methods on an analytically computed, synthetic, and real-world normal map. We computed the ground truth normal and height map of a SPHERE by a perspective camera model. We used Mitsuba[6] to render a normal map and the corresponding height map of Stanford BUNNY. We then added Gaussian noise to and randomly selected 1% normal vectors as outliers in SPHERE and BUNNY's normal map.

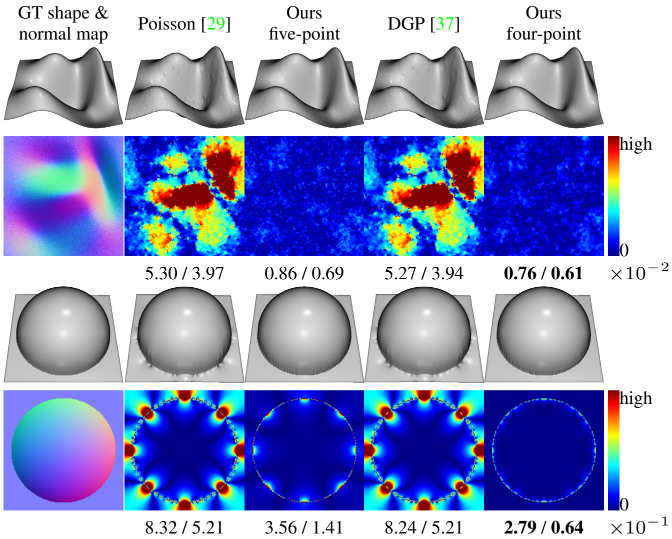

Figure 6. Reconstructed shape and its absolute difference map from **(Top)** the normal map with Gaussian noise and 1% outliers of anisotropic Gaussian surface [14] and **(Bottom)** a noise-free normal map of sphere with background padded. Discrete Poisson's equation and DGP shows an indistinguishable error pattern and close RMSE / MAE (numbers below the error maps). The shapes from discrete Poisson's equation or DGP are distorted by outliers, and Gibbs phenomenon appears near sharp features.

For the real-world surface, we used the normal map estimated by the state-of-the-art photometric stereo method [5] on the object HARVEST in DiLiGenT benchmark [31]. See our supplementary material for the comparison on all objects in DiLiGenT.

To better reflect the linear systems' features, we only looked at the least squares approximate solutions of compared methods without any regularization or smoothness term. As there is a global scale ambiguity in the estimated height maps, we found an optimal scale between the estimated and ground truth height map before computing RMSE and MAE. We applied the exponentiation [28] to the discrete functional [14] and discrete Poisson's equation [29]. It is observed that the exponentiation amplifies the checkerboard artifact of discrete functional and sensitivity to outliers of discrete Poisson's equation. Spikes occur in both methods. For Zhu and Smith's method [38], we used two variants as the discretization of partial derivative: Central difference and the third-order Savitzky-Golay (SG) filter.Zhu and Smith's method [38] performs robustly to outliers. However, a weaker checkerboard artifact exists in the central difference case. Moreover, a higher-order SG filter does not refine the reconstruction quality compared to central difference. Instead, it introduces severer high-frequency noise[7]. On the other hand, our methods perform stably on

[6] Mitsuba Renderer. https://www.mitsuba-renderer.org/index_old.html, last accessed on November 16, 2020.

[7] We empirically found it is the SG smoothing filter that plays a crucial role in refining the visual quality in Zhu and Smith's method [38].

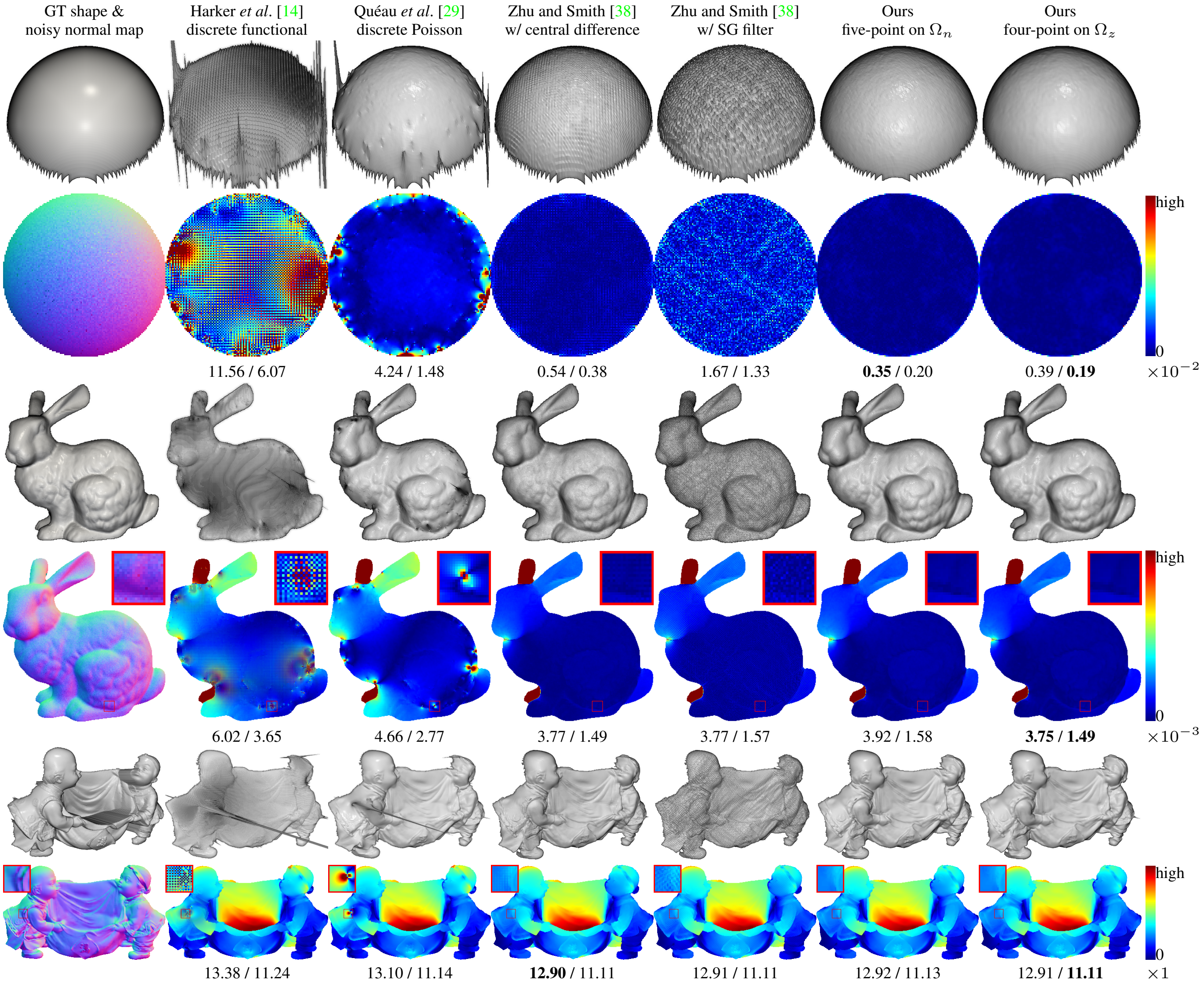

Figure 7. Visual comparison of perspective normal integration methods on normal maps of analytically computed SPHERE, synthetic BUNNY, and real-world HARVEST. Every two rows show the ground truth & estimated surface, the normal map as well as the absolute difference map (between ground truth and estimated height map). The two numbers underneath error maps are RMSE and MAE, respectively. We add Gaussian noise to and randomly select 1% normal vectors as outliers in SPHERE and BUNNY's normal maps.

all objects. The effect of outliers is hardly observed, and systematic high-frequency noise never occurs.

7. Conclusion

In this paper we presented, analyzed, and evaluated an inverse plane fitting approach to the problem of normal integration. We formed a linear system that minimizes the sum of squared point-to-plane distances instead of conventional vertical distances. We showed the source of the checkerboard artifact in the discrete functional-based approach. We also provided a geometric interpretation of the sensitivity to outliers or the sharp features of discrete Poisson's equation and DGP. Our method is straightforward, robust against outliers and sharp features, hyper-parameter free, yet based on least squares. We are interested in further studying discontinuity preserving normal integration and depth normal fusion based on our formulation.

8. Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19H01123, National Natural Science Foundation of China under Grant No. 61872012, 62088102, and Beijing Academy of Artificial Intelligence (BAAI). We would like to thank Zhuoyu Yang for fruitful discussions.

References

[1] Amit Agrawal, Ramesh Raskar, and Rama Chellappa. What is the range of surface reconstructions from a gradient field? In *Proc. of European Conference on Computer Vision (ECCV)*, 2006.

[2] Doris Antensteiner, Svorad Stolc, and Thomas Pock. A review of depth and normal fusion algorithms. *Sensors*, 2018.

[3] Martin Bahr, Michael Breus, Yvain Quéau, Ali Sharifi Boroujerdi, and Jean-Denis Durou. Fast and accurate surface normal integration on non-rectangular domains. *Computational Visual Media*, 2017.

[4] William L Briggs, Van Emden Henson, and Steve F McCormick. *A multigrid tutorial*. SIAM, 2000.

[5] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K Wong, and Yasuyuki Matsushita. What is learned in deep uncalibrated photometric stereo? In *Proc. of European Conference on Computer Vision (ECCV)*, 2020.

[6] Zhouyu Du, Antonio Robles-Kelly, and Fangfang Lu. Robust surface reconstruction from gradient field using the L1 norm. In *Proc. of Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA)*, 2007.

[7] Jean-Denis Durou, Jean-François Aujol, and Frédéric Courteille. Integrating the normal field of a surface in the presence of discontinuities. In *Workshops of the Computer Vision and Pattern Recognition (CVPRW)*, 2009.

[8] Jean-Denis Durou and Frederic Courteille. Integration of a normal field without boundary condition. In *Workshops of the International Conference on Computer Vision (ICCVW)*, 2007.

[9] Robert T. Frankot and Rama Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 1988.

[10] Silvano Galliani, Michael Breus, and Yong Chul Ju. Fast and robust surface normal integration by a discrete eikonal equation. In *Proc. of British Machine Vision Conference (BMVC)*, 2012.

[11] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2001.

[12] Matthew Harker and Paul O'Leary. Least squares surface reconstruction from measured gradient fields. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2008.

[13] Matthew Harker and Paul O'Leary. Least squares surface reconstruction from gradients: Direct algebraic methods with spectral, Tikhonov, and constrained regularization. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2011.

[14] Matthew Harker and Paul O'leary. Regularized reconstruction of a surface from its measured gradient field. *Journal of Mathematical Imaging and Vision*, 2015.

[15] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[16] Berthold KP Horn and Michael J Brooks. The variational approach to shape from shading. *Comput. Vis. Graph. Image Process.*, 1986.

[17] Itsik Horovitz and Nahum Kiryati. Depth from gradient fields and control points: Bias correction in photometric stereo. *Image and Vision Computing*, 2004.

[18] Katsushi Ikeuchi. Constructing a depth map from images. Technical report, Massachusetts Inst of Tech Cambridge Artificial Intelligence Lab, 1983.

[19] Katsushi Ikeuchi and Berthold KP Horn. Numerical shape from shading and occluding boundaries. *Artificial intelligence*, 1981.

[20] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3D: High-quality depth sensing with polarization cues. In *Proc. of International Conference on Computer Vision (ICCV)*, 2015.

[21] Bilge Karacali and Wesley Snyder. Reconstructing discontinuous surfaces from a given gradient field using partial integrability. *Computer Vision and Image Understanding*, 2003.

[22] Klaas Klasing, Daniel Althoff, Dirk Wollherr, and Martin Buss. Comparison of surface normal estimation methods for range sensing applications. In *Proc. of International Conference on Robotics and Automation (ICRA))*, 2009.

[23] Peter Kovesi. Shapelets correlated with surface normals produce surfaces. In *Proc. of International Conference on Computer Vision (ICCV)*, 2005.

[24] Richard B Lehoucq, Danny C Sorensen, and Chao Yang. *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM, 1998.

[25] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 2005.

[26] Christopher C Paige and Michael A Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)*, 1982.

[27] Yvain Quéau and Jean-Denis Durou. Edge-preserving integration of a normal field: Weighted least-squares, TV and L1 approaches. In *Proc. of International Conference on Scale Space and Variational Methods in Computer Vision*, 2015.

[28] Yvain Quéau, Jean-Denis Durou, and Jean-François Aujol. Normal integration: a survey. *Journal of Mathematical Imaging and Vision*, 2018.

[29] Yvain Quéau, Jean-Denis Durou, and Jean-François Aujol. Variational methods for normal integration. *Journal of Mathematical Imaging and Vision*, 2018.

[30] Antonio Robles-Kelly and Edwin R Hancock. A graph-spectral method for surface height recovery from needle-maps. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2001.

[31] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2019.

[32] Danny C Sorensen. Implicit application of polynomial filters in ak-step arnoldi method. *Siam journal on matrix analysis and applications*, 1992.

[33] Miaohui Wang, Wuyuan Xie, and Maolin Cui. Surface reconstruction with unconnected normal maps: An efficient mesh-based approach. In *Proc. of ACM International Conference on Multimedia (ACMMM)*, 2020.

[34] Robert J Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Proc. of Image Understanding Systems and Industrial Applications I*, 1979.

[35] Zhongquan Wu and Lingxiao Li. A line-integration based method for depth recovery from surface normals. *Computer Vision, Graphics, and Image Processing*, 1988.

[36] Wuyuan Xie, Miaohui Wang, Mingqiang Wei, Jianmin Jiang, and Jing Qin. Surface reconstruction from normals: A robust dgp-based discontinuity preservation approach. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2019.

[37] Wuyuan Xie, Yunbo Zhang, Charlie CL Wang, and Ronald C-K Chung. Surface-from-gradients: An approach based on discrete geometry processing. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2014.

[38] Dizhong Zhu and William A. P. Smith. Least squares surface reconstruction on arbitrary domains. In *Proc. of European Conference on Computer Vision (ECCV)*, 2020.