

This is a repository copy of *Least squares surface reconstruction on arbitrary domains*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/164094/>

Version: Accepted Version

Proceedings Paper:

Zhu, Dizhong and Smith, William Alfred Peter orcid.org/0000-0002-6047-0413 (Accepted: 2020) Least squares surface reconstruction on arbitrary domains. In: 16th European Conference on Computer Vision (ECCV 2020) Proceedings. 16th European Conference on Computer Vision (ECCV 2020), 23-28 Aug 2020 Lecture Notes in Computer Science. Springer. (In Press)

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Least squares surface reconstruction on arbitrary domains

Dizhong Zhu¹[0000–0003–4086–7293] and William A. P.
Smith¹[0000–0002–6047–0413]

¹ University of York, York, UK

² {dizhong.zhu,william.smith}@york.ac.uk

Abstract. Almost universally in computer vision, when surface derivatives are required, they are computed using only first order accurate finite difference approximations. We propose a new method for computing numerical derivatives based on 2D Savitzky-Golay filters and K-nearest neighbour kernels. The resulting derivative matrices can be used for least squares surface reconstruction over arbitrary (even disconnected) domains in the presence of large noise and allowing for higher order polynomial local surface approximations. They are useful for a range of tasks including normal-from-depth (i.e. surface differentiation), height-from-normals (i.e. surface integration) and shape-from-x. We show how to write both orthographic or perspective height-from-normals as a linear least squares problem using the same formulation and avoiding a nonlinear change of variables in the perspective case. We demonstrate improved performance relative to state-of-the-art across these tasks on both synthetic and real data and make available an open source implementation of our method.

Keywords: height-from-gradient, surface integration, Savitzky-Golay filter, surface reconstruction, least squares

1 Introduction

Estimating derivatives of a noisy measured signal is a basic problem in signal processing and finds application in areas ranging from spectroscopy to finance. The inverse of this problem arises when reconstructing a function from noisy measurements of its derivatives. This is a common problem in computer vision when estimating a surface (either an orthographic relative height map or a perspective absolute depth map) from noisy measurements of the surface normals or 2D surface gradient. This problem is usually known as surface integration or height-from-gradient [21]. More generally, shape-from-x methods that use a surface orientation cue to directly reconstruct a discrete 3D surface representation also require numerical approximations of the surface derivative. These approaches include shape-from-shading [5], polarisation [30], texture [3] and photometric stereo [1]. In addition, merging depth and surface normal estimates [19] requires a derivative operator to relate the two. Finally, recent work on deep

depth estimation computes surface gradients in-network so that either surface gradient supervision can be used [17] or to compute surface normals from depth maps [6]. Hence, numerical surface derivative approximations are of fundamental importance in computer vision.

It is therefore surprising that almost universally in the surface reconstruction literature, numerical derivative approximations that are only first order accurate (forward or backward finite difference) are used that make an implicit assumption of surface planarity and are highly susceptible to noise. Occasionally, central difference (second order accurate) [22] or smoothed central difference (increased robustness to noise) [19] kernels have been used but the only work to consider kernels accurate to arbitrary order is that of Harker and O’Leary [8–11]. However, their formulation works only on a rectangular domain meaning it cannot be applied to objects with arbitrary foreground masks. In addition, they use 1D kernels which cannot gain robustness by using a local neighbourhood spanning different rows and columns.

In this paper, we extend the idea of least squares surface integration in a number of ways. Like Harker and O’Leary we use kernels that are higher order accurate but, differently, we allow for arbitrary, even disconnected, domains. To the best of our knowledge, we are the first to use 2D Savitzky-Golay filters over an arbitrary domain for surface reconstruction problems (height-from-gradient and shape-from-x). Second, we propose to also use Savitzky-Golay filters as a smoothness regulariser. Unlike planar regularisers, such as a Laplacian filter [28] or zero surface prior [22], we are able to use a high regularisation weight to cope with very significant noise, yet still recover smooth curved surfaces without over flattening. Third, our least squares surface reconstruction approach is very general, allowing both orthographic and perspective projection (without requiring a nonlinear change of variables [21]), and an optional depth prior. Fourth, we propose an alternate formulation for height-from-normals that uses surface normal components rather than implied surface gradients and is numerically more stable. Finally, we make available an open source implementation of the methods that can easily be integrated into a surface reconstruction pipeline.

1.1 Related work

Computing differential surface properties from discrete surface representations is a large topic within computer graphics. Of particular interest in this work is the task of computing surface normals from potentially noisy depth maps. Mitra et al. [18] describe a classical approach in which surfaces are locally approximated by a plane fitted by least squares to nearest neighbour points. Klassing et al. [14] compare a variety of approaches and conclude that the straightforward plane PCA method performs well. Comino et al. [4] incorporate knowledge of the sensing device in order to develop an adaptive algorithm.

Classical approaches to recovering surface height from the surface normal or gradient field are based on the line integral [15, 23, 29]. They optimise local least squares cost functions and differ in their path selection strategy. Global methods were pioneered by Horn and Brooks [12], who posed the problem in

the continuous domain as a least squares optimisation problem. Although not convergent or practical, this approach led the way to many more modern approaches. Frankot and Chellappa [7] solved the same problem but formulated on the Fourier basis with a fast algorithm based on the DFT. Kovesi [16] uses a shapelet basis instead. Both methods assume periodic boundary conditions that introduces bias.

Agrawal et al. [2] construct a discrete Poisson equation and solve it efficiently. However, they use forward/backward finite difference approximations and a zero gradient boundary assumptions that biases the reconstruction. Simchony et al. [26] solve a Poisson equation using the Discrete Cosine Transform but require a rectangular domain. Harker and O’Leary [8–11] proposed a least squares approach and a subsequent series of refinements including a variety of regularisers. Their formulation is based on a rectangular matrix representation for the unknown height field. In this case, the numerical derivatives can be obtained by pre and post multiplication with an appropriate derivative matrix $\mathbf{D}_v \mathbf{Z}$ and $\mathbf{Z} \mathbf{D}_u$ where surface heights are stored in a matrix the same size as the image, $\mathbf{Z} \in \mathbb{R}^{W \times H}$. The differentiation matrices are both square with size equal to the height and width of the image respectively, $\mathbf{D}_u \in \mathbb{R}^{H \times H}$ and $\mathbf{D}_v \in \mathbb{R}^{W \times W}$. Hence, their size is $O(n)$ for $n = WH$ pixels. Harker and O’Leary show that the least squares problem can be written as a Sylvester equation and solved extremely efficiently. For robustness in the presence of noise, it is important to use local context to assist in the computation of the derivatives. The drawback of these $O(n)$ derivative matrices is that they can only use a neighbourhood of pixels in the same row (for horizontal derivatives) or column (for vertical). But the most significant drawback of their approach is the requirement for a rectangular domain. This rarely holds when either dealing with objects with a foreground mask or noisy sensor data with holes.

Recently, Quéau et al. [22] proposed the state-of-the-art method based on solving a least squares system formulated using sparse differentiation matrices $\mathbf{D}_u \mathbf{z}$ and $\mathbf{D}_v \mathbf{z}$ where foreground surface heights are stored in a vector with arbitrary ordering, $\mathbf{z} \in \mathbb{R}^n$. The differentiation matrices are both of the same size, $\mathbf{D}_u, \mathbf{D}_v \in \mathbb{R}^{n \times n}$. In the case where all pixels are foreground, $n = HW$ and these matrices are very large. However, they are sparse since each row has non-zero values only in columns corresponding to pixels in the local region of the pixel being differentiated. In the minimal case (forward, backward or central difference), each row has only two non-zero values. In contrast to Harker and O’Leary, this approach can deal with arbitrary domains. However, unlike Harker and O’Leary, it uses an average of forward and backward finite difference (central difference) which are not exact for higher order surfaces.

All of these methods make the assumption of orthographic projection. Quéau et al. [21] point out that any orthographic algorithm can be used for perspective surface integration by a nonlinear change of variables by solving in the log domain. The drawback of this transformation is that the solution is only least squares optimal in the transformed domain. When exponentiating to recover the perspective surface, large spikes can occur. The only method formulated in the

perspective domain that we are aware of is that of Nehab et al. [19], though in a slightly different context of merging depth and normals. Moreover, they use a derivative approximation based on smoothed central difference.

In this paper we bring together the best of both of these formulations and propose an approach that can handle arbitrary domains and uses arbitrary order numerical derivative approximations. Moreover, we reformulate the least squares height from normals problem such that it can handle both perspective and orthographic projection models.

2 Linear least squares height-from-normals

We denote a 3D point in world units as $\mathbf{p} = (x, y, z)$ and an image location in camera units (pixels) as (u, v) such that $\mathbf{u} = (u, v)$ is a pixel location in the image. We parameterise the surface by the height or depth function $z(\mathbf{u})$. In normals-from-depth we are given a noisy observed depth map and wish to estimate the surface normal map $\mathbf{n}(\mathbf{u}) = [n_x(\mathbf{u}), n_y(\mathbf{u}), n_z(\mathbf{u})]^T$ with $\|\mathbf{n}(\mathbf{u})\| = 1$. In surface integration we are given $\mathbf{n}(\mathbf{u})$ and wish to estimate $z(\mathbf{u})$.

To the best of our knowledge, all existing methods compute height-from-*gradient*, i.e. they transform the given surface normals into the surface gradient and solve the following pair of PDEs, usually in a least squares sense:

$$\frac{\partial z(\mathbf{u})}{\partial u} = \frac{-n_x(\mathbf{u})}{n_z(\mathbf{u})}, \quad \frac{\partial z(\mathbf{u})}{\partial v} = \frac{-n_y(\mathbf{u})}{n_z(\mathbf{u})}. \quad (1)$$

The problem with this approach is that close to the occluding boundary, n_z gets very small making the gradient very large. The squared errors in these pixels then dominate the least squares solution. We propose an alternative formulation that is more natural, works with both orthographic and perspective projections and, since it uses the components of the normals directly, is best referred to as *height-from-normals*. The idea is that the surface normal should be perpendicular to the tangent vectors. This leads to a pair of PDEs:

$$\frac{\partial \mathbf{p}(\mathbf{u})}{\partial u} \cdot \mathbf{n}(\mathbf{u}) = 0, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial v} \cdot \mathbf{n}(\mathbf{u}) = 0, \quad (2)$$

where $\mathbf{p}(\mathbf{u})$ denotes the 3D position corresponding to pixel position \mathbf{u} and $\frac{\partial \mathbf{p}(\mathbf{u})}{\partial u}$, $\frac{\partial \mathbf{p}(\mathbf{u})}{\partial v}$ are the image plane derivatives (i.e. partial derivatives with respect to pixel coordinates) of the 3D point position. We now consider how to formulate equations of this form in two different cases: orthographic and perspective projection.

2.1 Linear equations

Orthographic case The 3D position, $\mathbf{p}(\mathbf{u})$, of the point on the surface that projects to pixel position \mathbf{u} and its derivatives are given by:

$$\mathbf{p}(\mathbf{u}) = \begin{bmatrix} u \\ v \\ z(\mathbf{u}) \end{bmatrix}, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial u} = \begin{bmatrix} 1 \\ 0 \\ \frac{\partial z(\mathbf{u})}{\partial u} \end{bmatrix}, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial v} = \begin{bmatrix} 0 \\ 1 \\ \frac{\partial z(\mathbf{u})}{\partial v} \end{bmatrix}. \quad (3)$$

Substituting these derivatives into (2) we obtain:

$$\frac{\partial z(\mathbf{u})}{\partial u} n_z(\mathbf{u}) = -n_x(\mathbf{u}), \quad \frac{\partial z(\mathbf{u})}{\partial v} n_z(\mathbf{u}) = -n_y(\mathbf{u}). \quad (4)$$

Note that this is a simple rearrangement of (1) but which avoids division by n_z .

Perspective case In the perspective case, the 3D coordinate corresponding to the surface point at \mathbf{u} and its derivatives are given by:

$$\mathbf{p}(\mathbf{u}) = \begin{bmatrix} \frac{u-c_u}{f} z(\mathbf{u}) \\ \frac{v-c_v}{f} z(\mathbf{u}) \\ z(\mathbf{u}) \end{bmatrix}, \quad (5)$$

where f is the focal length of the camera and (c_u, c_v) is the principal point. The derivatives are given by:

$$\frac{\partial \mathbf{p}(\mathbf{u})}{\partial u} = \begin{bmatrix} \frac{1}{f} \left((u-c_u) \frac{\partial z(\mathbf{u})}{\partial u} + z(\mathbf{u}) \right) \\ \frac{1}{f} (v-c_v) \frac{\partial z(\mathbf{u})}{\partial u} \\ \frac{\partial z(\mathbf{u})}{\partial u} \end{bmatrix}, \quad \frac{\partial \mathbf{p}(\mathbf{u})}{\partial v} = \begin{bmatrix} \frac{1}{f} (u-c_u) \frac{\partial z(\mathbf{u})}{\partial v} \\ \frac{1}{f} \left((v-c_v) \frac{\partial z(\mathbf{u})}{\partial v} + z(\mathbf{u}) \right) \\ \frac{\partial z(\mathbf{u})}{\partial v} \end{bmatrix}. \quad (6)$$

Again, these can be substituted into (2) to relate the derivatives of z to the surface normal direction.

2.2 Discrete formulation

Assume that we are given a foreground mask comprising some subset of the discretised image domain, $\mathcal{F} \subseteq \{1, \dots, W\} \times \{1, \dots, H\}$ with $|\mathcal{F}| = n$. The depth values for the n foreground pixels are stored in a vector $\mathbf{z} \in \mathbb{R}^n$ with arbitrary ordering. We make use of a pair of matrices, $\mathbf{D}_u, \mathbf{D}_v \in \mathbb{R}^{n \times n}$, that compute discrete approximations to the partial derivative in the horizontal and vertical directions respectively. The exact form of these matrices is discussed in the next section. Once these discrete approximations are used, the PDEs in (2) become linear systems of equations in \mathbf{z} . This leads to a linear least squares formulation for the height-from-normals problem.

Orthographic case In the orthographic case, we stack equations of the form (4):

$$\begin{bmatrix} \text{diag}(\mathbf{n}_z) \mathbf{D}_u \\ \text{diag}(\mathbf{n}_z) \mathbf{D}_v \end{bmatrix} \mathbf{z} = \begin{bmatrix} -\mathbf{n}_x \\ -\mathbf{n}_y \end{bmatrix} \quad (7)$$

where

$$\mathbf{n}_x = \begin{bmatrix} n_x(\mathbf{u}_1) \\ \vdots \\ n_x(\mathbf{u}_n) \end{bmatrix}, \quad \mathbf{n}_y = \begin{bmatrix} n_y(\mathbf{u}_1) \\ \vdots \\ n_y(\mathbf{u}_n) \end{bmatrix}, \quad \mathbf{n}_z = \begin{bmatrix} n_z(\mathbf{u}_1) \\ \vdots \\ n_z(\mathbf{u}_n) \end{bmatrix}. \quad (8)$$

Note that (7) is satisfied by any offset of the true \mathbf{z} , corresponding to the unknown constant of integration. This is reflected in the fact that:

$$\text{rank} \left(\begin{bmatrix} \mathbf{D}_u \\ \mathbf{D}_v \end{bmatrix} \right) = n - 1. \quad (9)$$

So, in the orthographic case, we can only recover \mathbf{z} up to an unknown offset.

Perspective case In the perspective case, we stack equations obtained by substituting (6) in (2) to obtain:

$$\begin{bmatrix} \mathbf{N}\mathbf{T}_x \\ \mathbf{N}\mathbf{T}_y \end{bmatrix} \mathbf{z} = \mathbf{0}_{2n \times 1}, \quad (10)$$

where

$$\mathbf{T}_x = \begin{bmatrix} \frac{1}{f}\mathbf{U} & \frac{1}{f}\mathbf{I} \\ \frac{1}{f}\mathbf{V} & \mathbf{0}_{n \times n} \\ \mathbf{I} & \mathbf{0}_{n \times n} \end{bmatrix} \begin{bmatrix} \mathbf{D}_u \\ \mathbf{I} \end{bmatrix}, \quad \mathbf{T}_y = \begin{bmatrix} \frac{1}{f}\mathbf{U} & \mathbf{0}_{n \times n} \\ \frac{1}{f}\mathbf{V} & \frac{1}{f}\mathbf{I} \\ \mathbf{I} & \mathbf{0}_{n \times n} \end{bmatrix} \begin{bmatrix} \mathbf{D}_v \\ \mathbf{I} \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} \text{diag}(\mathbf{n}_x) \\ \text{diag}(\mathbf{n}_y) \\ \text{diag}(\mathbf{n}_z) \end{bmatrix}^T, \quad (11)$$

$\mathbf{U} = \text{diag}(u_1 - c_u, \dots, u_n - c_u)$ and $\mathbf{V} = \text{diag}(v_1 - c_v, \dots, v_n - c_v)$. Note that (10) is a homogeneous linear system. This means that it is also satisfied by any scaling of the true \mathbf{z} . So, in the perspective case, we can only recover \mathbf{z} up to an unknown scaling.

3 Numerical differentiation kernels

We now consider the precise form of \mathbf{D}_u and \mathbf{D}_v and propose a novel alternative with attractive properties. Since the derivative matrices act linearly on \mathbf{z} they can be viewed as 2D convolutions over $z(u, v)$. Note however that each row of \mathbf{D}_u or \mathbf{D}_v can be different - i.e. different convolution kernels can be used at different spatial locations.

By far the most commonly used numerical differentiation kernels are forward (fw) and backward (bw) difference, shown here for both the horizontal (h) and vertical (v) directions:

$$\mathbf{K}_{fw}^h = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_{fw}^v = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{K}_{bw}^h = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_{bw}^v = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (12)$$

As resolution increases and the effective step size decreases, forward and backward differences tend towards the exact derivatives. However, for finite step size they are only exact for order one (planar) surfaces and highly sensitive to noise. Averaging forward and backward yields the central difference (c) approximation, used for example by Quéau et al. [22]:

$$\mathbf{K}_c^h = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_c^v = \frac{1}{2} \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (13)$$

This is order two accurate but still only uses two pixels per derivative and so is sensitive to noise. One way to address this is to first smooth the z values with a smoothing kernel \mathbf{S} and then compute a finite difference approximation. By associativity of the convolution operator we can combine the smoothing and finite difference kernels into a single kernel. For example, the smoothed central difference (sc) approximation, as used by Nehab et al. [19] is given by:

$$\mathbf{K}_{\text{sc}}^{\text{h}} = \mathbf{K}_{\text{c}}^{\text{h}} * \mathbf{S} = \frac{1}{12} \begin{bmatrix} -1 & 0 & 1 \\ -4 & 0 & 4 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{K}_{\text{sc}}^{\text{v}} = \mathbf{K}_{\text{c}}^{\text{v}} * \mathbf{S} = \frac{1}{12} \begin{bmatrix} -1 & -4 & -1 \\ 0 & 0 & 0 \\ 1 & 4 & 1 \end{bmatrix}, \quad (14)$$

where in this case \mathbf{S} is a rounded approximation of a 3×3 Gaussian filter with standard deviation 0.6. A problem with both smoothed and unsmoothed central difference is that the derivatives and therefore the linear equations for a given pixel do not depend on the height of that pixel. This lack of dependence between adjacent pixels causes a severe “checkerboard” effect that necessitates the use of an additional regulariser, often smoothness. Commonly, this is the discrete Laplacian [28]. However, a smoothness penalty based on this filter is minimised by a planar surface. So, as the regularisation weight is increased, the surface becomes increasingly flattened until it approaches a plane.

With all of these methods alternative kernels must be used at the boundary of the foreground domain. For example, switching from central to backward differences. This means that the numerical derivatives are not based on a consistent assumption.

3.1 2D Savitzky-Golay filters

We now show how to overcome the limitations of the common numerical differentiation and smoothing kernels using 2D Savitzky-Golay filters.

The idea of a Savitzky-Golay filter [24] is to approximate a function in a local neighbourhood by a polynomial of chosen order. This polynomial is fitted to the observed (noisy) function values in the local neighbourhood by linear least squares. Although the polynomial may be of arbitrarily high order, the fit residuals are linear in the polynomial coefficients and so a closed form solution can be found. This solution depends only on the relative coordinates of the pixels in the local neighbourhood. So, it can be applied (linearly) to any data values meaning that reconstruction with the arbitrary order polynomial can be accomplished with a straightforward (linear) convolution.

The surface around a point (u_0, v_0) is approximated by the order k polynomial $z_{u_0, v_0}(u, v) : \mathbb{R}^2 \mapsto \mathbb{R}$ with coefficients a_{ij} :

$$z_{u_0, v_0}(u, v) = \sum_{i=0}^k \sum_{j=0}^{k-i} a_{ij} (u - u_0)^i (v - v_0)^j. \quad (15)$$

Assume we are given a set of pixel locations, $\mathcal{N}_{u_0, v_0} = \{(u_1, v_1), \dots, (u_m, v_m)\}$, forming a neighbourhood around (u_0, v_0) and the corresponding z values for

those pixels. We can form a set of linear equations

$$\begin{bmatrix} 1, v_1 - v_0, (v_1 - v_0)^2, \dots, (u_1 - u_0)^k \\ \vdots \\ 1, v_m - v_0, (v_m - v_0)^2, \dots, (u_m - u_0)^k \end{bmatrix} \mathbf{a} = \mathbf{C}_{\mathcal{N}_{u_0, v_0}} \mathbf{a} = \mathbf{z}_{\mathcal{N}_{u_0, v_0}}, \quad (16)$$

where $\mathbf{a} = [a_{00}, a_{01}, a_{02}, \dots, a_{k0}]^T$ and $\mathbf{z}_{\mathcal{N}_{u_0, v_0}} = [z(u_1, v_1), \dots, z(u_m, v_m)]^T$. The least squares solution for \mathbf{a} is given by $\mathbf{C}_{\mathcal{N}_{u_0, v_0}}^+ \mathbf{z}_{\mathcal{N}_{u_0, v_0}}$ where $\mathbf{C}_{\mathcal{N}_{u_0, v_0}}^+$ is the pseudoinverse of $\mathbf{C}_{\mathcal{N}_{u_0, v_0}}$. Note that $\mathbf{C}_{\mathcal{N}_{u_0, v_0}}^+$ depends only on the relative coordinates of the pixels chosen to lie in the neighbourhood of the (u_0, v_0) . Also note that $z_{u_0, v_0}(0, 0)$ is given simply by a_{00} which is the convolution between the first row of $\mathbf{C}_{\mathcal{N}_{u_0, v_0}}^+$ and the z values. This is a smoothed version of $z(u_0, v_0)$ in which the original surface is locally approximated by a best fit, order k polynomial. Similarly, the first derivative of the fitted polynomial in the horizontal direction is given by a_{10} and in the vertical direction by a_{01} , corresponding to two other rows of $\mathbf{C}_{\mathcal{N}_{u_0, v_0}}^+$. Note that the order k is limited by the size of the neighbourhood. Specifically, we require at least as many pixels as coefficients, i.e. $k \leq m$.

When \mathcal{N}_{u_0, v_0} is a square neighbourhood centred on (u_0, v_0) then the appropriate row of $\mathbf{C}_{\mathcal{N}_{u_0, v_0}}^+$ can be reshaped into a square convolution kernel. Convolving this with a $z(u, v)$ map with rectangular domain \mathcal{F} amounts to locally fitting a polynomial of order k and either evaluating the polynomial at the central position, acting as a smoothing kernel, or evaluating the derivative of the polynomial in either vertical or horizontal direction.

3.2 K-nearest pixels kernel

In general, the foreground domain will not be rectangular. Often, it corresponds to an object mask or semantic segmentation of a scene. In this case, we need a strategy to deal with pixels that do not have the neighbours required to use the square kernel. 2D Savitzky-Golay filters are ideal for this because the method described above for constructing them can be used for arbitrary local neighbourhoods. We propose to use the K-nearest pixels in \mathcal{F} to a given pixel. In practice, we compute the square $d \times d$ kernel once and use this for all pixels where the required neighbours lie in \mathcal{F} . For those that do not, we find the d^2 nearest neighbours in \mathcal{F} (one of which will be the pixel itself). Where tie-breaks are needed, we do so randomly, though we observed no significant difference in performance if all tied pixels are included. In Figure 1 we show an example of a standard and non-standard case. All non-white pixels lie in \mathcal{F} . Pixel A has the available neighbours to use the square kernel while B does not and uses a custom kernel.

Each element in a kernel for a pixel is copied to the appropriate entries in a row of \mathbf{D}_u or \mathbf{D}_v . We similarly construct a matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ containing the a_{00} kernels, i.e. the smoothing kernel. Each row of these three matrices has d^2 non-zero entries.

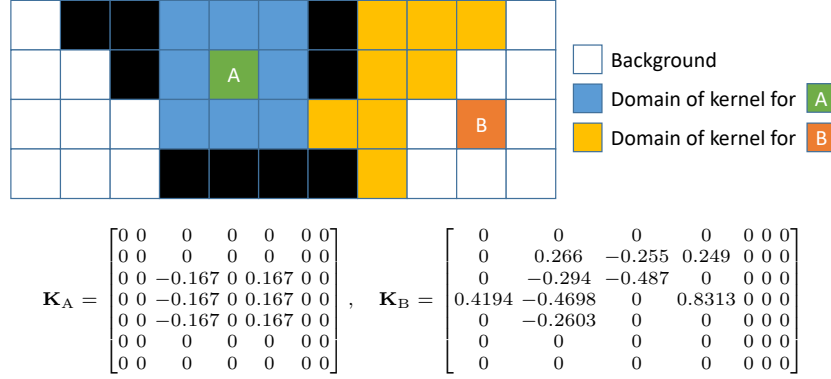


Fig. 1. An example of computing 2D Savitzky-Golay filters on an arbitrary domain. In this example, we use a 3×3 kernel. For point A we can use the default square kernel. The order two Savitzky-Golay filter for the horizontal derivative is shown below as \mathbf{K}_A . For point B we use the 3^2 nearest pixels and build a custom order two Savitzky-Golay filter shown below as \mathbf{K}_B . In practice, higher order kernels provide better performance.

3.3 3D K-nearest neighbours kernel

For normals-from-depth where a noisy depth map is provided, the K-nearest neighbours kernel idea can be extended to 3D. The idea is to use the depth map with (5) to transform pixels to 3D locations, then to perform the KNN search in 3D. The advantage of this is that kernels will avoid sampling across depth discontinuities where the large change in depth will result in adjacent pixels being far apart in 3D distance. This allows us to create large, robust kernels but without smoothing over depth discontinuities.

4 Implementation

For an efficient implementation, all pixel coordinates from \mathcal{F} are placed in a KNN search tree so that local neighbourhoods can be found quickly and pixels that can use the square mask are identified by convolution of the mask with a square filter of ones.

To compute normals-from-depth, we use our proposed derivative matrices (with 3D KNN search) to compute the partial derivatives of z , take the cross product between horizontal and vertical derivatives (6) and normalise to give the unit surface normal.

To compute height-from-normals, we solve a system of the form of (7) (orthographic) or (10) (perspective). We augment the system of equations with a smoothness penalty of the form $\lambda(\mathbf{S} - \mathbf{I})\mathbf{z} = \mathbf{0}$, where λ is the regularisation weight. This encourages the difference between the smoothed and reconstructed z values to be zero. For the orthographic system, we resolve the unknown offset by solving for the minimum norm solution - equivalent to forcing the mean z

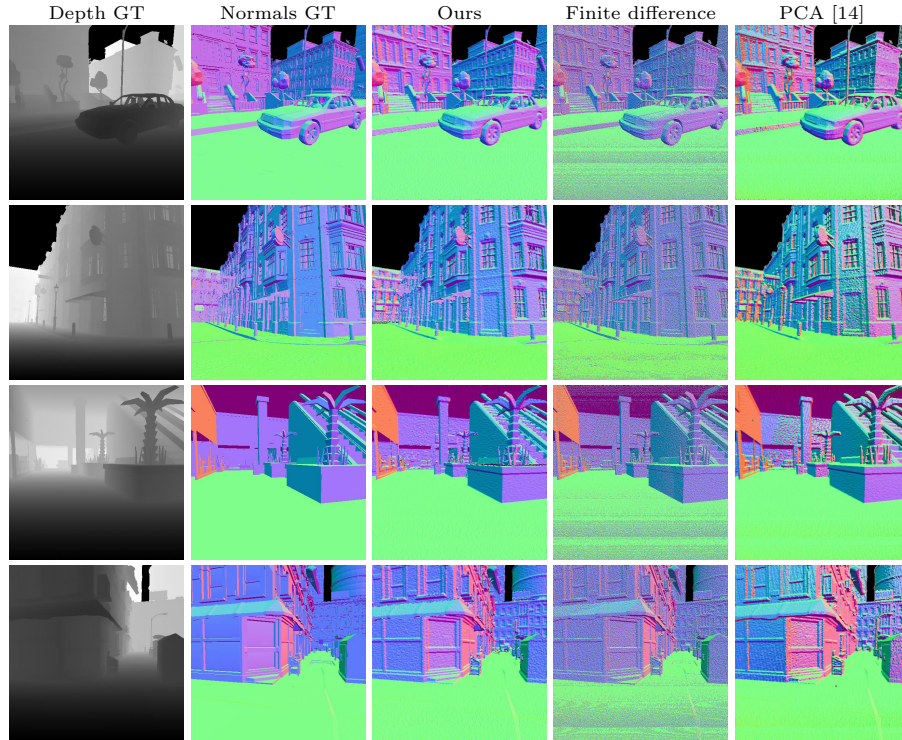


Fig. 2. Normals-from-depth results on the 3D Ken Burns dataset [20]. Zoom for detail.

value to zero. For the perspective case, since the system is homogeneous in theory we could solve for the $\|\mathbf{z}\| = 1$ solution by solving a minimum direction problem using the sparse SVD. In practice, we find it is faster to add an additional equation forcing the solution at one pixel to unity. Finally, we can optionally include a depth prior simply by adding the linear equation $\omega \mathbf{I} \mathbf{z} = \omega \mathbf{z}_{\text{prior}}$, where ω is the prior weight.

We provide a complete implementation of our method in Matlab³.

5 Evaluation

Normals-from-depth To evaluate normals-from-depth we use the synthetic dataset of Niklaus et al. [20]. This comprises realistic scene renderings and includes depth and normal maps. The normal maps are obtained by rasterising the ground truth mesh and so correctly handle depth discontinuities. The depth maps contain noise due to quantisation. We compute normals from this noisy depth and compare against the ground truth rasterised normal map. The combination of quantisation noise and depth discontinuities make the task surprisingly difficult.

³ <https://github.com/waps101/LSQSurfaceReconstruction>

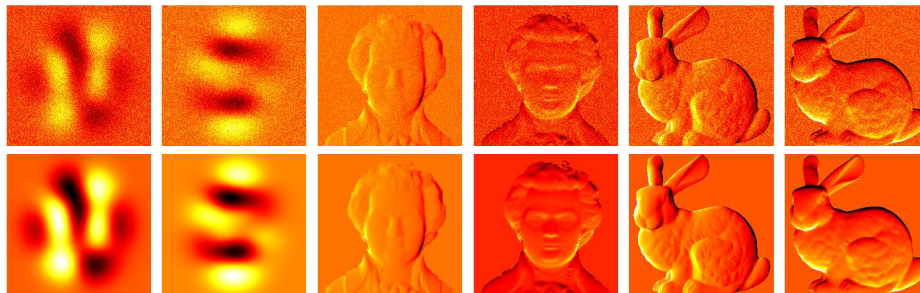


Fig. 3. First row presents the input noisy x and y gradient maps corresponding to the surfaces in Figure 4. The second row shows the ground truth without noise.

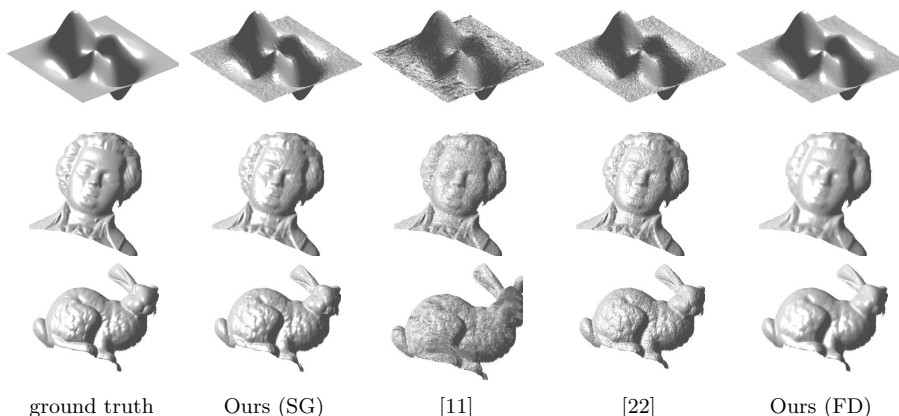


Fig. 4. Qualitative results on synthetic noisy data.

Qualitative results are shown in Figure 2. Note that the finite difference (using forward difference) normal map is extremely noisy. Plane PCA [14] reduces noise but introduces planar discontinuities across depth boundaries while our result is smooth but preserves depth discontinuities. This is due to the 3D nearest neighbour filters. We show quantitative results for two scenes in Table 1. Here we also include smoothed central finite difference. The proposed approach reduces error relative to the next best performing method by over 25%.

Method	Scene	
	city-walking	victorian-walking
Ours	15.73	19.84
FD (sc)	23.60	26.38
FD (fw)	38.11	39.86
[14]	25.37	30.06

Table 1. Median angular error of estimated surface normals on two scenes from 3D Ken Burns dataset [20].

Height-from-normals We evaluate our height-from-normals method on both synthetic and real data. We compare against two state-of-the-art methods [11] and

the best performing variant (Total Variation minimisation) of [22]. When applying these methods to perspective data we use the transformation proposed in [21] to reconstruct in the log domain before exponentiating to recover perspective depth. We compare against our approach using backward finite difference (FD) and the proposed Savotzky-Golay filters (SG).

We begin by evaluating on synthetic data (peaks, Mozart and Stanford bunny) with Gaussian noise added to the input surface gradients (see Figure 4). We show qualitative results in Figure 4. Note that [11] introduces a checkerboard effect in the presence of noise. This is due to the independence of adjacent pixel height estimates caused by approximating derivatives only along a single row or column. [22] is noisy due to having no explicit smoothness prior. Finite difference is smooth due to the strong Laplacian filter regularisation but this causes the surface to flatten. Our

result preserves the global shape while also retaining local smoothness. We show the influence of varying the standard deviation of the noise in Figure 5. [11] degrades very quickly with noise. We slightly outperform [22] and note that our method is much more straightforward, requiring only the solution of a sparse linear system.

Next, we evaluate on real data. In this case, it is a perspective reconstruction task. We use the surface normals estimated by [13] on the DiLiGenT benchmark [25]. We use the camera parameters, foreground masks and ground truth depth provided with the benchmark. We integrate the estimated normals and compare the resulting depth to ground truth. Since the scale of the depth is unknown, we first compute the optimal scale between reconstruction and ground truth prior to computing the RMS error. We show qualitative results in Figure 6. Note that the nonlinear transformation causes spikes in the comparison methods. Our approach yields the visually best results, apart from smoothing across the discontinuities in the first example. We show quantitative results in Table 3. Note that, in the cases where we are outperformed by [22], their result often contains severe spike artefacts that create a poor visual reconstruction.

Photometric stereo Finally, we evaluate the effect of using our proposed filters in a photometric stereo experiment. We choose the photometric stereo method of Smith and Fang [27]. This is not state-of-the-art (using only the Lambertian reflectance model) but it solves directly for an orthographic height map from a set of calibrated photometric stereo images and makes use of derivative matrices in this solution. Hence, it makes a good test case for our alternative deriva-

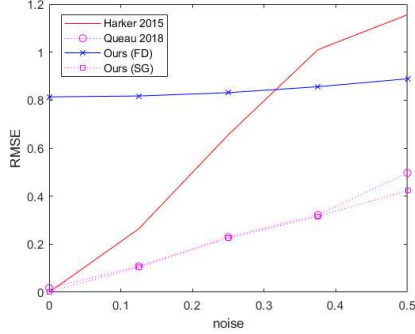


Fig. 5. Root Mean Square Error between the recovered depth and the ground truth versus the standard deviation of the additive Gaussian noise. We use the synthetic image of the peaks with increasing Gaussian noise.

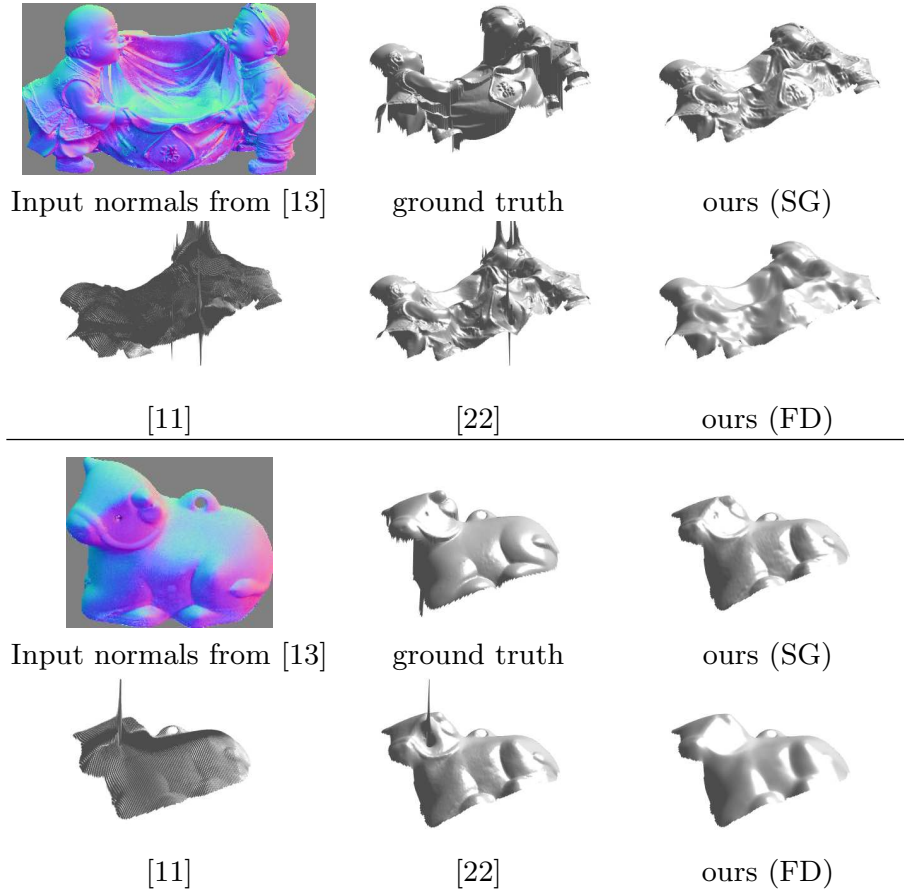


Fig. 6. Qualitative perspective height-from-normals results on real data.

tive matrices. It works by taking ratios between pairs of intensity observations yielding linear equations in the surface gradient. Then, substituting numerical derivative approximations, solves for the least squares optimal height map. We use the authors original implementation which uses smoothed central difference derivatives and Laplacian smoothing filter. We compare this implementation with one in which the only modification we make is to replace the derivative and smoothing matrices with our K-nearest pixel, 2D Savitzky-Golay filters. We keep all other parameters fixed. We again use the DiLiGenT dataset [25], this time running the photometric stereo algorithm on the input images, computing an orthographic height map, computing normals from this and calculating the mean angular error to ground truth. We show quantitative results in Figure 3. Simply replacing the derivative and smoothing matrices significantly reduces error, sometimes by over 50%.

	Method	bear	buddha	cat	cow	goblet	harvest	pot 1	pot 2	reading
Depth	[11]	6.11	6.26	7.29	4.82	19.03	11.72	3.51	2.69	11.68
	[22]	5.55	4.35	5.90	2.49	15.49	11.45	2.78	1.75	9.66
	Ours (SG)	5.42	3.97	5.53	2.13	14.90	11.55	3.07	1.93	10.29
Normal	[11]	20.30	24.06	19.27	21.10	24.09	55.50	20.88	20.42	39.00
	[22]	20.06	22.64	18.60	15.20	19.84	29.64	20.26	19.34	24.18
	Ours (SG)	19.97	22.02	18.38	14.78	19.21	26.22	20.12	19.17	23.57

Table 2. Perspective surface integration errors for depth (Root Mean Square Error in millimetres) and surface normal (median angular error in degrees) of recovered depth and surface normals of recovered depth relative to ground truth.

Method	bear	buddha	cat	cow	goblet	harvest	pot 1	pot 2	reading
[27]	12.7	26.3	15.1	25.4	20.0	33.2	20.0	24.2	25.5
[27] + SG filters	7.33	14.8	8.43	23.7	16.4	29.2	9.88	14.7	14.0

Table 3. Photometric stereo evaluation. We compare the surface normal median angular error between the two methods and ground truth.

6 Conclusions

In this paper we have explored alternatives to the widely used numerical derivative approximations. This often overlooked choice turns out to be significant in the performance of normals-from-depth, height-from-normals and shape-from-x. While we propose specific methods for these problems, the main takeaway from this paper is that any algorithm that uses sparse derivative matrices could plug in our matrices based on 2D Savitzky-Golay filters and see benefit. There are many possible extensions. Our approach does not consider or deal with discontinuities (apart from the 3D nearest neighbour extension for normals-from-depth). A hybrid between our approach and the discontinuity aware approach of [22] may be possible. Another interesting avenue is integrating better differentiation kernels into deep learning frameworks. Where a segmentation mask is available or estimated, it should be possible to apply appropriate kernels to each segment avoiding smoothing across depth discontinuities. The challenge here would be making it sufficiently efficient for use in deep learning as well as making the kernel selection differentiable. Finally, it would be interesting to see whether normals-from-depth can be learnt as a black box process and whether this outperforms our handcrafted kernels.

Acknowledgements

W. Smith is supported by a Royal Academy of Engineering/The Leverhulme Trust Senior Research Fellowship.

References

1. Ackermann, J., Goesele, M.: A survey of photometric stereo techniques. *Foundations and Trends® in Computer Graphics and Vision* **9**(3-4), 149–254 (2015)
2. Agrawal, A., Raskar, R., Chellappa, R.: What is the range of surface reconstructions from a gradient field? In: *Proc. ECCV*. pp. 578–591 (2006)
3. Clerc, M., Mallat, S.: The texture gradient equation for recovering shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(4), 536–549 (2002)
4. Comino, M., Andujar, C., Chica, A., Brunet, P.: Sensor-aware normal estimation for point clouds from 3d range scans. In: *Computer Graphics Forum*. vol. 37, pp. 233–243. Wiley Online Library (2018)
5. Durou, J.D., Falcone, M., Sagona, M.: Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding* **109**(1), 22–43 (2008)
6. Eigen, D., Fergus, R.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2650–2658 (2015)
7. Frankot, R.T., Chellappa, R.: A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on pattern analysis and machine intelligence* **10**(4), 439–451 (1988)
8. Harker, M., O’Leary, P.: Least squares surface reconstruction from measured gradient fields. In: *Proc. CVPR* (2008)
9. Harker, M., O’Leary, P.: Least squares surface reconstruction from gradients: Direct algebraic methods with spectral, tikhonov, and constrained regularization. In: *Proc. CVPR*. pp. 2529–2536 (2011)
10. Harker, M., O’Leary, P.: Direct regularized surface reconstruction from gradients for industrial photometric stereo. *Computers in Industry* **64**(9), 1221–1228 (2013)
11. Harker, M., O’Leary, P.: Regularized reconstruction of a surface from its measured gradient field. *Journal of Mathematical Imaging and Vision* **51**(1), 46–70 (2015)
12. Horn, B.K., Brooks, M.J.: The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing* **33**(2), 174–208 (1986)
13. Ikehata, S., Aizawa, K.: Photometric stereo using constrained bivariate regression for general isotropic surfaces. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2179–2186 (2014)
14. Klasing, K., Althoff, D., Wollherr, D., Buss, M.: Comparison of surface normal estimation methods for range sensing applications. In: *2009 IEEE International Conference on Robotics and Automation*. pp. 3206–3211. IEEE (2009)
15. Klette, R., Koschan, A., Schluns, K.: *Three-dimensional data from images*. Springer-Verlag Singapore Pte. Ltd., Singapore (1998)
16. Kovesi, P.: Shapelets correlated with surface normals produce surfaces. In: *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*. vol. 2, pp. 994–1001. IEEE (2005)
17. Li, Z., Snavely, N.: Megadepth: Learning single-view depth prediction from internet photos. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2041–2050 (2018)
18. Mitra, N.J., Nguyen, A.: Estimating surface normals in noisy point cloud data. In: *Proceedings of the nineteenth annual symposium on Computational geometry*. pp. 322–328 (2003)

19. Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3d geometry. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)* **24**(3), 536–543 (2005)
20. Niklaus, S., Mai, L., Yang, J., Liu, F.: 3d ken burns effect from a single image. *ACM Transactions on Graphics* **38**(6), 184:1–184:15 (2019)
21. Quéau, Y., Durou, J.D., Aujol, J.F.: Normal integration: a survey. *Journal of Mathematical Imaging and Vision* **60**(4), 576–593 (2018)
22. Quéau, Y., Durou, J.D., Aujol, J.F.: Variational methods for normal integration. *Journal of Mathematical Imaging and Vision* **60**(4), 609–632 (2018)
23. Robles-Kelly, A., Hancock, E.R.: A graph-spectral method for surface height recovery. *Pattern recognition* **38**(8), 1167–1186 (2005)
24. Savitzky, A., Golay, M.J.: Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **36**(8), 1627–1639 (1964)
25. Shi, B., Mo, Z., Wu, Z., Duan, D., Yeung, S.K., Tan, P.: A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(2), 271–284 (2019)
26. Simchony, T., Chellappa, R., Shao, M.: Direct analytical methods for solving poisson equations in computer vision problems. *IEEE transactions on pattern analysis and machine intelligence* **12**(5), 435–446 (1990)
27. Smith, W.A.P., Fang, F.: Height from photometric ratio with model-based light source selection. *Computer Vision and Image Understanding* **145**, 128–138 (2016)
28. Smith, W.A.P., Ramamoorthi, R., Tozza, S.: Height-from-polarisation with unknown lighting or albedo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(12), 2875–2888 (2019)
29. Wu, Z., Li, L.: A line-integration based method for depth recovery from surface normals. *Computer Vision, Graphics, and Image Processing* **43**(1), 53–66 (1988)
30. Zhu, D., Smith, W.A.: Depth from a polarisation+ rgb stereo pair. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 7586–7595 (2019)