sensors

MDPI

Article

A Review of Depth and Normal Fusion Algorithms

Doris Antensteiner [1,*], Svorad Štolc [1] and Thomas Pock [1,2]

[1] Center for Vision, Automation and Control, Austrian Institute of Technology, Vienna 1210, Austria; svorad.stolc@ait.ac.at (S.Š.); pock@icg.tugraz.at (T.P.)
[2] Institute of Computer Graphics and Vision, Graz University of Technology, Graz 8010, Austria
* Correspondence: doris.antensteiner@ait.ac.at; Tel.: +43-664-8825-6103

Received: 11 November 2017; Accepted: 26 January 2018; Published: 1 February 2018

Abstract: Geometric surface information such as depth maps and surface normals can be acquired by various methods such as stereo light fields, shape from shading and photometric stereo techniques. We compare several algorithms which deal with the combination of depth with surface normal information in order to reconstruct a refined depth map. The reasons for performance differences are examined from the perspective of alternative formulations of surface normals for depth reconstruction. We review and analyze methods in a systematic way. Based on our findings, we introduce a new generalized fusion method, which is formulated as a least squares problem and outperforms previous methods in the depth error domain by introducing a novel normal weighting that performs closer to the geodesic distance measure. Furthermore, a novel method is introduced based on Total Generalized Variation (TGV) which further outperforms previous approaches in terms of the geodesic normal distance error and maintains comparable quality in the depth error domain.

Keywords: depth reconstruction; surface normals; optimization; Total Generalized Variation; primal-dual algorithm; computational imaging; least squares

1. Introduction

Measuring the depth of a scene accurately is essential for many tasks including applications in industrial environments, object recognition and security assurance. Usually the depth is measured by stereo cameras, structure from motion (SfM), time of flight (ToF) sensors, or light field cameras. These methods show accurate absolute depth maps but lack detail in high frequency depth structures. The reason for this lies in the dependency on the presence of structural information in the image as well as in the analysis routine, which is usually done by hypothesis testing and therefore limited in range and step sizes. Stereo matching methods include correlation based techniques [1], semi global matching [2,3], block based matching [4,5] and stereo matching for micro array cameras [6]. Several methods were introduced to retrieve depth information from light field data using epipolar image slope analysis [7], structure tensors [8], fine-to-coarse approaches [9,10] and by line consistency metrics [11]. Structure from motion and time of flight techniques were presented in [12–14], respectively. In contrast to depth based approaches, methods that recover surface normals, such as photometric stereo [15], show high frequency details but lack an absolute depth reference. Shape from shading was used to retrieve surface normals by [16]. A robust normal reconstruction using photometric stereo information with a Markov Random Field (MRF) was introduced in [17]. Previous methods have been presented, which retrieve surface normals from a calibrated stereo setup. This can be achieved e.g., by estimating the homography between two matched patches [18,19] or by using the affine transform data between two projections [20] additionally to the reconstruction of a sparse depth map retrieved from a stereo correspondence analysis. A learning-based method using a tandem of convolutional neural networks to estimate depth and surface normals from image data simultaneously was introduced in [21].

Combining depth and surface normal data allows precise depth reconstructions for low- as well as high-frequency components in the depth map.

Depth maps and surface normals were previously combined in various ways. Shape from shading was used under general illumination in [22], photometric stereo normals were incorporated in [23–25]. Another approach was presented in [26], where the tangent plane of the given normals was projected into the measured normal field. This normal constraint was previously used in several algorithms (e.g., [27–30]). The method described in [31] uses a standard depth constraint and forces the Laplacian of the optimal solution to be in the proximity of the derivative of the given normals. In [32] a depth and photometric stereo fusion algorithm was introduced, which uses additional Laplacian smoothing term and adaptive pixel-wise weighting parameters to preserve surface discontinuities. The Laplacian smoothing term was also added in [33]. A extended penalty is chosen in [34], where the normal is enforced to be close to the normal from the initial depth map, while 2nd-order spherical harmonics are used to constrain the normals according to the observed shading in the input image, a smoothness function enforces the similarity of 1st-order neighbors and an additional term constrains the normals to unit length. A method to refine depth by photometric stereo information using RGB-D cameras was introduced in [35], where an energy function is optimizing for the depth, smoothness, shading and temporal aliasing of a scene. Surface normals from polarization cues were used to enhance the depth map in [36], in an iterative process the depth is refined with corrected surface normal information and a depth fidelity constraint, which enforces consistency between the surface from normals and accurate regions in the depth map. An original approach for inferring about the surface normals from light field data as well as a hybrid setup combining depth maps with surface normals using a block coordinate descent algorithm was demonstrated in [37].

Even though several methods to combine surface depth with surface orientation data previously emerged, a thorough analysis and classification of the properties of those approaches was missing. In this paper, our first main contribution comprises an in-depth comparison of several variational methods using depth maps and surface orientation data, as well as a classification and evaluation of weighting terms for surface orientations using gradients or surface normals. We analyze orientation weighting terms of common methods and explain their differences in respect to the geodesic distance weighting. We show that methods which behave closer to this natural surface normal weighting term show a better performance, especially in regions with steep depth edges. Based on the findings we introduce our second main contribution, a new generalized formulation of a previously introduced method [26], which outperforms other methods regarding the error in the depth domain. Our third main contribution is a novel gradient-based approach, which is using TGV and outperforms other methods in the domain of the geodesic error of the resulting normals.

2. Depth and Surface Normal Cues

At present, 3D models are used for a wide range of analysis tasks. Depth models are being constructed by acquisition devices using stereo systems, light field cameras, time of flight (ToF), or other range scanning techniques. Common methods show a high precision in the absolute depth measurement, but a low quality in fine relative details. These errors are major obstructions for tasks such as finding defects in objects. Measuring the normal fields of objects by using methods such as photometric stereo [15] or shape from shading [38] will allow the reconstruction of surfaces with highly precise local details. On the downside, those methods show errors in the low frequency domain and therefore result in a low absolute depth accuracy.

Combining depth maps with surface normal information allows an exact 3D reconstruction both in absolute depth and fine surface details. This can be achieved by optimizing energy functions by variational methods, where the solution is penalized for deviating from the depth model and from the surface normals. In state-of-the-art techniques, the surface normal component is either used directly or by converting it to gradient information, where the x- and y-component can be treated independently.

Such an independent treatment can be beneficial for applications where data components are missing, as for example line-scanners [39].

3. Notations and Preliminaries

In this section, we introduce the essential notations used across this paper. By default we assume discretized surface structures of the size of $M \times N$ pixels. In order to access the image location, we define the following index set $\mathcal{I} = \{(i,j) : 1 \leq i \leq M, 1 \leq j \leq N\}$.

The discrete depth map of our scene is scalar valued in each pixel and defined as follows:

$$Z = (Z_{i,j})_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N}. \tag{1}$$

Variables with a bold font refer to surface structures where each pixel is vector valued. Hence, the surface gradient field $\boldsymbol{G}$ in x- and y-direction is defined as:

$$\boldsymbol{G} = \left(\boldsymbol{G}_{i,j}\right)_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N \times 2}, \quad \text{where} \quad \boldsymbol{G}_{i,j} = (G_{i,j,x}, G_{i,j,y}). \tag{2}$$

The gradient of the depth map is computed using standard finite differences:

$$\nabla Z = ((\nabla Z)_{i,j})_{i,j \in \mathcal{I}}, \quad \text{where} \quad (\nabla Z)_{i,j} = ((\nabla_x Z)_{i,j}, (\nabla_y Z)_{i,j}), \tag{3}$$

and the gradient operator in x- and y-direction $\nabla : \mathbb{R}^{M \times N} \to \mathbb{R}^{M \times N \times 2}$ is given by:

$$\begin{aligned} (\nabla_x Z)_{i,j} &= \begin{cases} Z_{i+1,j} - Z_{i,j} & \text{if } 1 \leq i < M, \\ 0, & \text{otherwise,} \end{cases} \\ (\nabla_y Z)_{i,j} &= \begin{cases} Z_{i,j+1} - Z_{i,j} & \text{if } 1 \leq j < N, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{4}$$

The surface normal field is defined as follows:

$$\boldsymbol{N} = \left(\boldsymbol{N}_{i,j}\right)_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N \times 3}, \quad \text{where} \quad \boldsymbol{N}_{i,j} = (N_{i,j,x}, N_{i,j,x}, N_{i,j,z}) \in \mathbb{R}^3. \tag{5}$$

By definition, we have $|\boldsymbol{N}_{i,j}|_2 = 1$. The relation between the surface gradient estimation and the surface normals is defined for all $i, j \in \mathcal{I}$ as follows:

$$N_{i,j,x} = \frac{-(\nabla_x Z)_{i,j}}{|(-(\nabla Z)_{i,j}, 1)|_2}, \quad N_{i,j,y} = \frac{-(\nabla_y Z)_{i,j}}{|(-(\nabla Z)_{i,j}, 1)|_2}, \quad \text{and} \quad N_{i,j,z} = \frac{1}{|(-(\nabla Z)_{i,j}, 1)|_2}. \tag{6}$$

Furthermore, we are using two specific surface tangent vectors, which are aligned with the x- and y-vector respectively and defined as follows:

$$\boldsymbol{T} = (\boldsymbol{T}_{i,j})_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N \times 2 \times 3}, \quad \text{where} \tag{7}$$

$$\boldsymbol{T}_{i,j} = (\boldsymbol{T}_{i,j,x}, \boldsymbol{T}_{i,j,y}) \quad \text{and} \tag{8}$$

$$\boldsymbol{T}_{i,j,x} = (1, 0, (\nabla_x Z)_{i,j}) \quad \text{and} \quad \boldsymbol{T}_{i,j,y} = (0, 1, (\nabla_y Z)_{i,j}). \tag{9}$$

4. Depth and Surface Normal Fusion Algorithms

In this paper, we analyze state-of-the-art methods in a systematic way and introduce two novel approaches. The described hybrid depth and surface normal methods are categorized in terms of their penalty functions. State-of-the-art approaches used similar depth penalty terms and differed in the surface orientation weighting and regularization. We organize the described methods in two categories: (i) gradient-based and (ii) normal-based approaches as well as with respect to their

treatment of flat and steep surface regions. While the presented methods show similar behaviors in flat areas, they differ in the penalization of steep regions. We show the quadratic penalty functions of the methods presented in Figures 1 and 2 for lateral and polar deviations, which are illustrated in Figure 3. We will show that the behavior of the energy function for different inclination angles correlates with the quantitative and qualitative depth reconstruction performance of each individual method. We argue that the geodesic distance function shows the most natural behavior with the favorable property of penalizing the distance of the normal orientation independent of the steepness of the edges. Due to this ideal behavior, we use the function to evaluate other distance measures in Section 5. Without loss of generality, we assume a dense depth and normal map for the algorithms described in this paper. In case of sparse input data, we suggest an extension based on Poisson surface reconstruction [40] to deal with sparse data.

An overview of the presented methods is given in Table 1. The first method we present in Section 4.2 is the construction of depth from only the gradient surface orientation information (i.e., no prior information about the absolute depth is being used). Using only the surface orientation for the depth reconstruction results in large-scale low-frequency errors (and therefore depth offsets). Later we overcome this problem by introducing an additional depth constraint in all following methods.

Second, we introduce the gradient-based approach with a depth constraint formulated as a least squares problem in Section 4.3. The respective contour plot of the orientation distance measure shows a strong penalization of steep edges in contrast to flat surface regions. It is easy to see (Figure 1), that the error from the same angular deviation due to noisy normals may generate from small up to infinity error in the gradient domain, depending on the inclination angle. For demonstrative purposes, we additionally show two extensions of the gradient-based method with regularization terms. One forces gradient-based smoothing. The other one enforces smoothness with a Laplacian term and can be used for the reconstruction with sparse depth and surface normal data.

Third, the method of Heber [41] for combining depth with surface orientations is shown in Section 4.4, which scales the given normal by the length of the optimized gradient.

Forth, a review of the method of Nehab [26] is given in Section 4.5, which reprojects the tangents of the optimized surface onto the given normal.

Fifth, as one of our main contributions in this paper, we introduce a new penalty function in form of a generalized method of Nehab in Section 4.6. Using a novel parametrization moves the penalty function closer to the geodesic normal energy and hence penalizes deviations in steep edges and flat regions more equally.

Last, we introduce our second main contribution, a novel Total Generalized Variation (TGV) model in Section 4.7, which penalizes the distance of the gradients of the surface orientation and gives significantly improved reconstruction results. The reason for this lies in the decoupling of the gradient through the TGV term.

4.1. Geodesic Distance

In the 3D space, the geodesic distance is the most natural surface normal penalty as distances between surface normals are weighted equally, independent of the surface slope angle with respect to the observer. Therefore it is used in this paper as a comparison measure for the evaluation in Section 5.

The geodesic distance $d_{i,j}$ is defined as the inverse cosine of the point-wise dot product between the given normal $\hat{\boldsymbol{N}} \in \mathbb{R}^{M \times N \times 3}$ and the estimated solution normal field $\boldsymbol{N} \in \mathbb{R}^{M \times N \times 3}$ as follows:

$$d_{i,j} = acos(\langle \hat{\boldsymbol{N}}_{i,j}, \boldsymbol{N}_{i,j} \rangle), \quad \forall i, j \in \mathcal{I}, \tag{10}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard dot product, which is defined as $\langle a, b \rangle = \sum_{i=1}^{n} a_i \, b_i = ||a|| \, ||b|| \, cos(\phi)$ and ϕ describes the angle between a and b. We can formulate the distance in Equation (10) by utilizing

the surface gradient estimation $\nabla Z \in \mathbb{R}^{M \times N \times 2}$ for the surface normal $\boldsymbol{N}$ with the relations shown in Equation (6) as follows:

$$d_{i,j} = acos\left(\frac{\langle \hat{\boldsymbol{N}}_{i,j}, (-(\nabla Z)_{i,j}, 1)\rangle}{|(-(\nabla Z)_{i,j}, 1)|_2}\right). \tag{11}$$

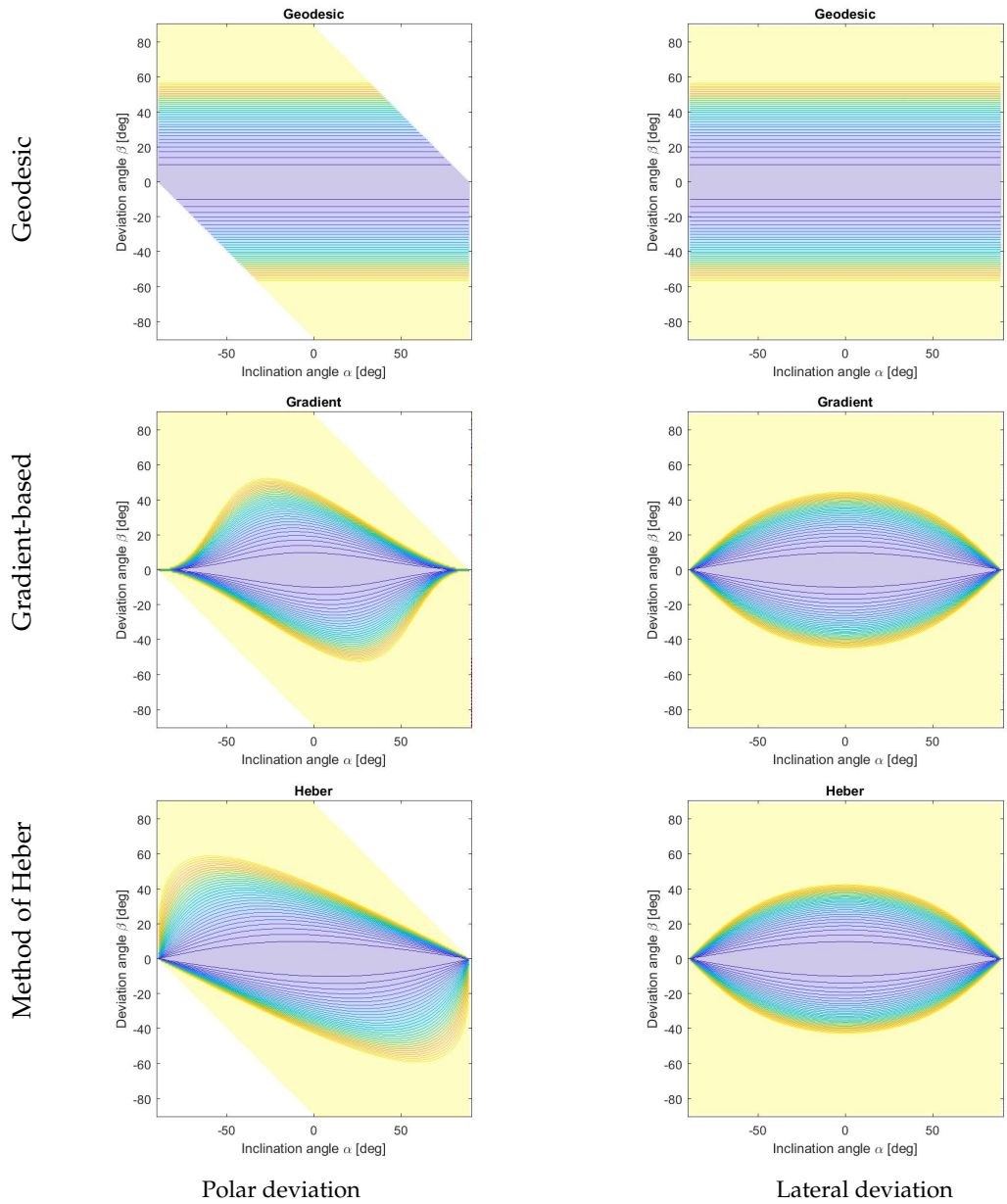


Figure 1. Quadratic penalty functions of the surface orientation constraint visualized for deviations in polar (left column) and lateral (right column) directions. Shown are the geodesic energy (top row), the gradient-based method (middle row) and the method of Heber (bottom row). Colors of the contours indicate the respective penalty values obtained for different combinations of the inclination and deviation angles. Note that the range is clipped between 0 (blue) and 1 (yellow). See Figure 3 for the explanation of the parametrization used.

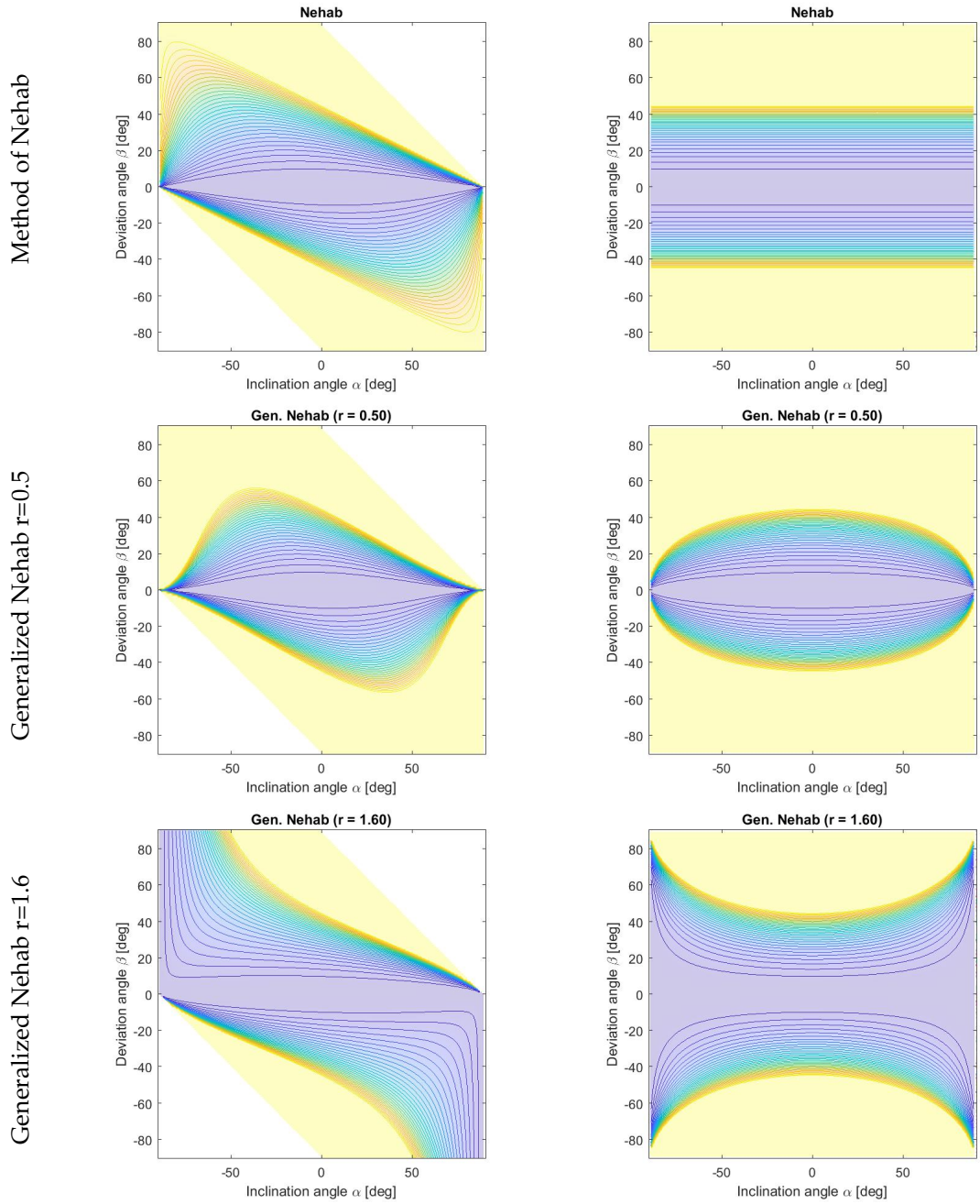


Figure 2. Quadratic penalty functions of the surface orientation constraint visualized for deviations in polar (left column) and lateral (right column) directions. Shown are the method of Nehab (top row), the generalized method of Nehab with $r = 0.5$ (middle row) as well as $r = 1.6$ (bottom row). A more natural weighting, which is in the proximity of the geodesic distance, can be achieved by the adaption of the parameter r. A well balanced distance measure such as with $r = 1.6$ has the potential of performing better in steep regions (high inclination angle) while preserving performance in the flat regions. Colors of the contours indicate the respective penalty values obtained for different combinations of the inclination and deviation angles. Note that the range is clipped between 0 (blue) and 1 (yellow). See Figure 3 for the explanation of the parametrization used.

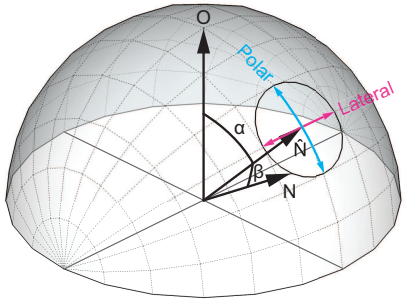


Figure 3. The distance between a two surface normals $\hat{\boldsymbol{N}}$ and $\boldsymbol{N}$ is expressed by means of the horizontal polar coordinate system. The distances in Figures 1 and 2 are measured in lateral and polar directions using angles α and β. The former describes the angle between the given normal $\hat{\boldsymbol{N}}$ and the the upright vector O, the latter defines the angular deviation between the normals $\hat{\boldsymbol{N}}$ and $\boldsymbol{N}$.

Table 1. Overview of the presented methods. The behavior of the surface orientation constraint in regions with different orientations is visualized for each method in Figures 1 and 2.

Method	Depth Penalty	Orientation Penalty	Balance at Flat Regions	Balance at Steep Regions
Surface orientation only	✗	Gradient-based	✓	✗
Gradient-based	✓	Gradient-based	✓	✗
Gradient-based	✓	Gradient-based + regularization with zero Laplace (Equation (23))	✓	✗
Gradient-based	✓	Gradient-based + regularization with gradient (Equation (21))	✓	✗
Method of Heber	✓	Scaled normal	✓	✗
Method of Nehab	✓	Projection of surface tangents to the given normal field	✓	∽
Generalized Nehab (ours)	✓	Projection of surface tangents to the given normal field with additional N_z weighting	✓	✓
TGV (ours)	✓	Gradient-based + TGV	✓	✓

The surface orientation weighting of the geodesic distance is illustrated in the contour plot in the first row of Figure 1. The first column shows the polar deviation of the coordinates and the second column shows the lateral deviation, parameterized by the inclination angle α and the deviation angle β, as illustrated in Figure 3. For the following methods a balanced weighting over all inclination angles both for the polar and lateral deviation is favored, as provided by the geodesic distance.

4.2. Gradient-Based Method with Surface Orientation Constraint Only

As typical for photometric stereo methods, depth can be partly recovered from surface orientations only. In order to provide a complete context for the method considered in this paper, we show here a method that is using solely gradient-based data. Gradient-based methods have previously been frequently utilized for depth reconstruction (e.g., [16,42–47]). Given a gradient field $\hat{\boldsymbol{G}} \in \mathbb{R}^{M\times N\times 2}$, we calculate the surface gradients for the estimated depth map Z in x- ($\nabla_x Z$) and y-direction ($\nabla_y Z$) respectively. Combining relations from Equation (6), the relations of surface normals and gradients are as follows:

$$|(-(\nabla Z)_{i,j}, 1)|_2 = -\frac{(\nabla_x Z)_{i,j}}{N_{i,j,x}} = -\frac{(\nabla_y Z)_{i,j}}{N_{i,j,y}} = \frac{1}{N_{i,j,z}}, \quad \forall i,j \in \mathcal{I}, \tag{12}$$

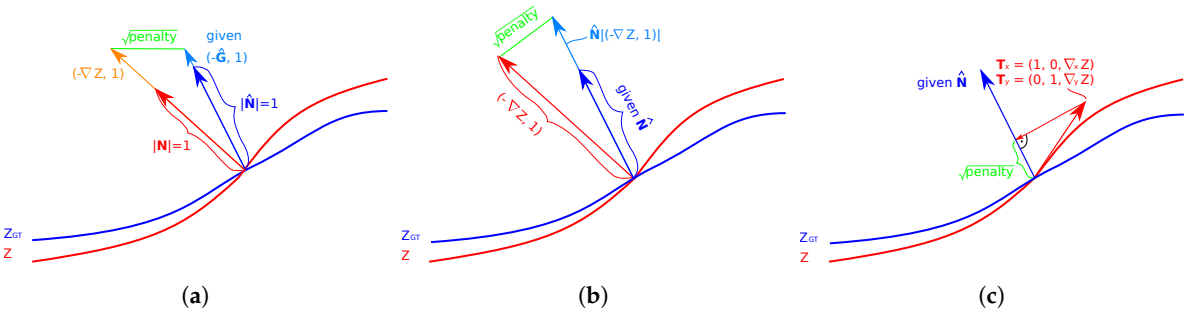


Figure 4. Comparison of the penalty of (**a**) the gradient-based method; (**b**) the method of Heber and (**c**) the method of Nehab. Gradient-based method: comparison of the measured gradient field components $(-\nabla_x Z, -\nabla_y Z)$ with the given gradient field $(-\hat{G}_x, -\hat{G}_y)$. Method of Heber: scaling of the normal vector $\hat{\boldsymbol{N}}$ by the length of $(-\nabla_x Z, -\nabla_y Z, 1)$. Method of Nehab: distance by projection of the tangents $\boldsymbol{T}_x$ and $\boldsymbol{T}_y$ of the optimized surface onto the given normal field $\hat{\boldsymbol{N}}$.

hence the surface gradients are given as:

$$(\nabla_x Z)_{i,j} = -\frac{N_{i,j,x}}{N_{i,j,z}} \quad \text{and} \quad (\nabla_y Z)_{i,j} = -\frac{N_{i,j,y}}{N_{i,j,z}}, \quad \forall i,j \in \mathcal{I}, \tag{13}$$

and the given gradient fields correspond to the surface normals by:

$$\hat{G}_{i,j,x} = -\frac{\hat{N}_{i,j,x}}{\hat{N}_{i,j,z}} \quad \text{and} \quad \hat{G}_{i,j,y} = -\frac{\hat{N}_{i,j,y}}{\hat{N}_{i,j,z}}, \quad \forall i,j \in \mathcal{I}, \tag{14}$$

in x- and y- direction respectively. Our goal is to compute a depth map Z such that

$$(\nabla Z)_{i,j} \approx \hat{\boldsymbol{G}}_{i,j}, \quad \forall i,j \in \mathcal{I}. \tag{15}$$

The comparison of the resulting penalty between the measured and the given gradients is illustrated in Figure 4a. Since Equation (15) is an overdetermined system of linear equations, it can be solved as the following least squares problem:

$$\min_{Z} \frac{1}{2} ||\nabla Z - \hat{\boldsymbol{G}}||^2, \tag{16}$$

whose global minimizer Z_{min} satisfies the first order sufficient optimality condition:

$$\nabla^*(\nabla Z_{min} - \hat{\boldsymbol{G}}) = 0, \tag{17}$$

where $\nabla^* : \mathbb{R}^{M \times N \times 2} \rightarrow \mathbb{R}^{M \times N}$ denotes the adjoint of the ∇ operator, with $\nabla^* = (\nabla_x^*, \nabla_y^*)$. We compute the minimizer using a standard conjugate gradient method.

It is well known that reconstructing the depth using only surface normal data usually results in errors in the low frequency domain. In the past, this has been improved by different approaches, such as introducing additional boundary conditions [44]. We resolve this problem by hybrid depth and surface normal formulations which proved very efficient in finding an accurate surface reconstruction.

4.3. Gradient-Based Method

In this section, we discuss a hybrid gradient-based method which reconstructs depth using a gradient-based algorithm (similar to the one from Section 4.2) extended by the use of the given initial depth $\hat{Z}$. Also here the gradient of the estimated depth map ∇Z is forced to be in the proximity of a

measured gradient $\hat{\boldsymbol{G}}$ in x- and in y-direction. Therefore, we formulate an overdetermined system of equations as follows:

$$Z \approx \hat{Z} \quad \text{and} \quad (\nabla Z)_{i,j} \approx \hat{\boldsymbol{G}}_{i,j}, \quad \forall i,j \in \mathcal{I}. \tag{18}$$

The corresponding least squares problem is given as:

$$\min_{Z} \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}||\nabla Z - \hat{\boldsymbol{G}}||^2, \tag{19}$$

where $\lambda > 0$ is used to balance between the depth and the orientation constraints. The global optimizer Z_{min} is found by a standard conjugate gradient method, with the optimality condition given as follows:

$$Z_{min} - \hat{Z} + \lambda \nabla^* (\nabla Z_{min} - \hat{\boldsymbol{G}}) = 0. \tag{20}$$

The contour plots of the surface orientation penalty function corresponding to the gradient-based method are shown in Figure 1. Note that with this method, the penalty is notably stronger for deviations in steep regions than in flat regions.

For demonstrative purposes, we introduce the extension of the gradient-based method in Equation (19) with a regularization term. We use a Laplace 2nd-order method which is driven by the derivative of the given gradient field. We formulate the following least squares problem:

$$\min_{Z} \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}||\nabla Z - \hat{\boldsymbol{G}}||^2 + \frac{\lambda_R}{2}||\Delta Z - (-\nabla^* \hat{G})||^2, \tag{21}$$

where $\lambda_R > 0$ balances the regularization, $\nabla^* \hat{G}$ can be decomposed into $(\nabla_x^* \hat{G}_x, \nabla_y^* \hat{G}_y)$ and Δ denotes the Laplace operator which is defined as follows:

$$\Delta_x = -\nabla_x^* \nabla_x \quad \text{and} \quad \Delta_y = -\nabla_y^* \nabla_y. \tag{22}$$

Such a gradient-based regularization can be applied to all presented methods.

The presented gradient-based approaches require dense depth and surface orientation data. A possibility to cope with sparse data is an additional smoothness assumption coupled with pixel-wise weighting parameters. Hence, we add another term to Equation (19) with a Laplacian smoothness assumption as follows:

$$\min_{Z} \frac{1}{2}||\lambda_1 \odot (Z - \hat{Z})||^2 + \frac{1}{2}||\boldsymbol{\lambda}_2 \odot (\nabla Z - \hat{\boldsymbol{G}})||^2 + \frac{1}{2}||\boldsymbol{\lambda}_3 \odot (\Delta Z)||^2. \tag{23}$$

The weighting parameters λ^1 and $\boldsymbol{\lambda^2}$ can be given a priori by the confidence of a data point and $\boldsymbol{\lambda^3}$ by the inverse confidence and are defined as follows:

$$\lambda_1 = (\lambda_{1,i,j})_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N}, \tag{24}$$

$$\boldsymbol{\lambda}_2 = (\boldsymbol{\lambda}_{2,i,j})_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N \times 2}, \quad \text{where} \quad \boldsymbol{\lambda}_{2,i,j} = (\lambda_{2,i,j,x}, \lambda_{2,i,j,y}), \quad \text{and} \tag{25}$$

$$\boldsymbol{\lambda}_3 = (\boldsymbol{\lambda}_{3,i,j})_{i,j \in \mathcal{I}} \in \mathbb{R}^{M \times N \times 2}, \quad \text{where} \quad \boldsymbol{\lambda}_{3,i,j} = (\lambda_{3,i,j,x}, \lambda_{3,i,j,y}). \tag{26}$$

In case of stereo or light-field methods for depth reconstruction the parameters can be assessed base on the matching confidence. Unknown points would have a confidence of zero. The same extension for sparse data is applicable for all following methods. Therefore, without loss of generality we discuss the weighting of the surface orientation term with a focus on dense depth and surface orientation data without an additional smoothness assumption.

4.4. The Method of Heber

A hybrid variational refinement model was described by Heber [41], where an initial rough depth $\hat{Z}$ is given and refined with surface normal information:

$$\min_{Z} \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}\left|\left|(-\nabla Z, 1) - \hat{\boldsymbol{N}} \odot |(-\nabla Z, 1)|_2\right|\right|_2^2, \quad \text{where} \tag{27}$$

$$|(-\nabla Z, 1)|_2 = \left(|(-(\nabla Z)_{1,1}, 1)|_2, |(-(\nabla Z)_{1,2}, 1)|_2, ..., |(-(\nabla Z)_{M,N}, 1)|_2\right) \tag{28}$$

defines the vector of pointwise 2-norms and the symbol $\odot$ denotes the operator for the point-wise product, also known as Hadamard product. This method is conceptually similar to the gradient-based method described in Section 4.3, also here the same depth constraint ensures a result in the proximity of an initial depth solution $\hat{Z}$. However, the surface orientation constraint used in Heber's method exploits the given normal field $\hat{\boldsymbol{N}}$ directly instead of the gradient field $\hat{\boldsymbol{G}}$. Here the normalization of ∇Z by division by the length of the vector is overcome by multiplying the term on one side by $|(-\nabla Z, 1)|_2$, which leads to a convex problem [41].

An illustration of the comparison of the measured ∇Z and the given normal $\hat{\boldsymbol{N}}$ is shown in Figure 4b. The given normal is scaled to the length of the measured gradients ∇Z. The contour map in Figure 1 demonstrates the normal weighting of the Heber's energy term. Here, similar to the gradient-based method, a deviation in steep edges is penalized more than a deviation in flat regions. Also, a different weighting applies whether the estimated gradients are steeper or flatter than the given values.

As the energy function from Equation (27) is convex and differentiable, this algorithm can be solved by an (accelerated) gradient descent method or a (fast) proximal gradient approach. For our evaluation, we used a plain gradient descent approach.Nesterov [48] proposed an accelerated gradient descent method with a simple weighted gradient step, followed by additional sliding, based on the last estimation. A fast proximal gradient method has an additional extrapolation step compared to the proximal gradient method. An example is the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [49] (see Appendix A.1).

4.5. The Method of Nehab

The method of Nehab [26] combines depth and surface normals by solving a system of linear equations, consisting of depth and surface orientation constraints. This method is similar to the gradient-based method described in Section 4.3, only the surface normal information is leveraged in a different way, making different trade-offs between flat and steep gradients.

In this method, the surface normal constraint optimizes the sum of squared projections of a set of surface tangent vectors $\boldsymbol{T}$, as defined in Equation (7) through Equation (9), of the reconstructed surface Z onto the given normal field $\hat{\boldsymbol{N}}$. This surface normal penalty has been adopted previously in several approaches (e.g., [21,27–30]). The projection is illustrated in Figure 4c. Note that the lowest penalty 0 is reached when the tangent vector is precisely orthogonal to the given normal vector $\hat{\boldsymbol{N}}$. We first consider the formulation as described by Nehab [26] by formulating an overdetermined linear system of sparse equations:

$$\begin{aligned} Z &\approx \hat{Z}, \\ \langle \hat{\boldsymbol{N}}_{i,j}, \boldsymbol{T}_{i,j,x} \rangle = \hat{N}_{i,j,x} + \hat{N}_{i,j,z}(\nabla_x Z)_{i,j} &\approx 0, \quad \text{and} \\ \langle \hat{\boldsymbol{N}}_{i,j}, \boldsymbol{T}_{i,j,y} \rangle = \hat{N}_{i,j,y} + \hat{N}_{i,j,z}(\nabla_y Z)_{i,j} &\approx 0, \quad \forall i,j \in \mathcal{I}, \end{aligned} \tag{29}$$

which leads to the following least squares problem:

$$\min_{Z} \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}\left|\left|\hat{N}_z \odot \nabla_x Z + \hat{N}_x\right|\right|^2 + \frac{\lambda}{2}\left|\left|\hat{N}_z \odot \nabla_y Z + \hat{N}_y\right|\right|^2, \tag{30}$$

where the parameter $\lambda \in [0,1]$ weights the influence of the initial depth and the given normals. The global optimizer Z_{min} is calculated with a standard conjugate gradient method, with the optimality condition given as follows:

$$Z_{min} - \hat{Z} + \lambda\Big(\nabla_x^*(\hat{N}_z \odot (\hat{N}_z \odot \nabla_x Z_{min} + \hat{N}_x)) + \nabla_y^*(\hat{N}_z \odot (\hat{N}_z \odot \nabla_y Z_{min} + \hat{N}_y))\Big) = 0. \tag{31}$$

The weighting of the surface normal information is explained in more detail in Figure 4c and further illustrated in a contour plot in Figure 2. The polar deviation to the given surface orientation in steep and flat regions is penalized similarly to the method of Heber (see Figure 1, bottom row, left). However, the lateral deviations are weighted equally which is the same behavior as the ideal geodesic distance function (see Figure 1, top row, right).

Note that the approach of Nehab, as described in Equation (29), corresponds to the gradient-based method with an additional local N_z-weighting applied to both sides of the surface normal constraint of Equation (18). This can be shown by utilizing the equivalences defined in Equations (13) and (14) as follows:

$$\hat{N}_{i,j,x} = \hat{N}_{i,j,z}\,(\nabla_x Z)_{i,j} \quad \text{and} \quad \hat{N}_{i,j,y} = \hat{N}_{i,j,z}\,(\nabla_y Z)_{i,j}, \quad \forall i,j \in \mathcal{I}, \tag{32}$$

and formulating a corresponding overdetermined linear system of equations:

$$\begin{aligned} Z &\approx \hat{Z}, \\ \hat{N}_{i,j,z}\,(\nabla_x Z)_{i,j} &\approx \hat{N}_{i,j,z}\,\hat{\boldsymbol{G}}_{i,j,x}, \quad \text{and} \\ \hat{N}_{i,j,z}\,(\nabla_y Z)_{i,j} &\approx \hat{N}_{i,j,z}\,\hat{\boldsymbol{G}}_{i,j,y}, \quad \forall i,j \in \mathcal{I}. \end{aligned} \tag{33}$$

Compared with the gradient-based method, the N_z-weighting inherent to Nehab's method improves the behavior of the penalty function by weakening the influence of regions with steep edges. Rationale behind the N_z-weighting follows from the fact that flat regions exhibit a higher value of N_z (close to 1) while steep regions receive low N_z values (close to 0). Such a local weighting helps preventing the over-penalization in steep regions as observed in the gradient-based method.

4.6. Generalized Nehab

In order to allow the normal weighting to reach a closer proximity to the geodesic distance, we propose a generalization of the method of Nehab. We extend on the concept of the gradient N_z-weighting as explained in Equation (33) by introducing an additional exponent r that controls influence of the N_z-weighting such that:

$$\begin{aligned} Z &\approx \hat{Z}, \\ (\hat{N}_{i,j,z})^r\,(\nabla_x Z)_{i,j} &\approx (\hat{N}_{i,j,z})^r\,\hat{\boldsymbol{G}}_{i,j,x}, \quad \text{and} \\ (\hat{N}_{i,j,z})^r\,(\nabla_y Z)_{i,j} &\approx (\hat{N}_{i,j,z})^r\,\hat{\boldsymbol{G}}_{i,j,y}, \quad \forall i,j \in \mathcal{I}. \end{aligned} \tag{34}$$

We optimize the corresponding least squares problem with a standard conjugate gradient method:

$$\min_{Z} \frac{1}{2}||Z - \hat{Z}||^2 + \frac{\lambda}{2}\left|\left|(\hat{N}_z)^r \odot \nabla_x Z - (\hat{N}_z)^r \odot \hat{\boldsymbol{G}}_x\right|\right|^2 + \frac{\lambda}{2}\left|\left|(\hat{N}_z)^r \odot \nabla_y Z - (\hat{N}_z)^r \odot \hat{\boldsymbol{G}}_y\right|\right|^2, \tag{35}$$

where we find a global optimizer Z_{min} that satisfies the following optimality condition:

$$Z_{min} - \hat{Z} + \lambda \left(\nabla_x^* \left((\hat{N}_z)^r \odot \left((\hat{N}_z)^r \odot \nabla_x Z_{min} - (\hat{N}_z)^r \odot \hat{\boldsymbol{G}}_x \right) \right) + \right.$$
$$\left. \nabla_y^* \left((\hat{N}_z)^r \odot \left((\hat{N}_z)^r \odot \nabla_y Z_{min} - (\hat{N}_z)^r \odot \hat{\boldsymbol{G}}_y \right) \right) \right) = 0. \tag{36}$$

The influence of varying r on the behavior of the surface orientation normal penalty function is shown in Figure 2. It can be seen that with $r = 0.5$ the generalized method of Nehab penalizes steeper slopes stronger than the original method of Nehab (equivalent to $r = 1$) and weaker than the gradient-based method (equivalent to $r = 0$). On the other hand, with $r = 1.6$ the proposed method exhibits a more tolerant behavior towards steeper inclination angles than the original method of Nehab, resulting in a global behavior that is reasonably close to the geodesic penalty in both the polar and lateral deviation directions. The value of the parameter r might still be further optimized.

Regarding the optimal choice of r, we have recognized that $r = 1.6$ behaves in the vicinity to the geodesic distance. If the structure of the data shows strong noise on edges steeper than approximately 50 degrees, a lower choice of r can be considered.

4.7. Total Generalized Variation

In this section, we introduce another novel method to reconstruct a refined surface with a hybrid depth and surface normal approach that is based on a Total Generalized Variation (TGV) approach as introduced in [50]. The TGV is an extension of the Total Variation (TV), which is a very popular and efficient regularization technique used currently in many image processing applications, however, it is known for producing staircasing artifacts in slope regions of the solution. In contrast, the TGV overcomes the staircasing problem by allowing solutions of higher order. Our formulation is a gradient-based approach, which restricts the surface to be close with a quadratic penalty to an initial depth solution $\hat{Z}$, while enforcing the auxiliary gradient field $\boldsymbol{G}$ to be in the proximity of the given gradient field $\hat{\boldsymbol{G}}$, propagated through the TGV penalty. Thereby, we simultaneously reconstruct the surface Z and the auxiliary gradient field $\boldsymbol{G}$, such that our discrete model is given as follows:

$$\min_{Z,\boldsymbol{G}} \alpha_1 ||\nabla Z - \boldsymbol{G}||_{2,1} + \alpha_0 ||\boldsymbol{\nabla G}||_{2,1} + \frac{\alpha}{2}||Z - \hat{Z}||^2 + \frac{\beta}{2}||\boldsymbol{G} - \hat{\boldsymbol{G}}||^2, \tag{37}$$

where the gradient operator $\boldsymbol{\nabla} : \mathbb{R}^{M \times N \times 2} \rightarrow \mathbb{R}^{M \times N \times 4}$ computes finite differences and $\boldsymbol{\nabla G}$ can be decomposed into $(\nabla G_x, \nabla G_y)$, where ∇ was defined in Equation (4). Therefore, the first and second order components of our TGV regularization have the following form:

$$||\nabla Z - \boldsymbol{G}||_{2,1} = \sum_{i,j} \sqrt{(\nabla_x Z)_{i,j}^2 + (G_{i,j,x})^2 + (\nabla_y Z)_{i,j}^2 + (G_{i,j,y})^2}, \quad \text{and} \tag{38}$$

$$||\boldsymbol{\nabla G}||_{2,1} = \sum_{i,j} \sqrt{(\nabla_x G_x)_{i,j}^2 + (\nabla_y G_x)_{i,j}^2 + (\nabla_x G_y)_{i,j}^2 + (\nabla_y G_y)_{i,j}^2}. \tag{39}$$

The function in Equation (37) consists of a depth and an orientation constraint as well as the TGV regularization terms. The depth constraint (i.e., the third term of Equation (37)) enforces the depth solution Z to stay in the vicinity of our noisy initial depth map $\hat{Z}$. The orientation constraint (i.e., the fourth term of Equation (37)) enforces $\boldsymbol{G}$ to stay in the proximity of the given surface gradient estimation $\hat{\boldsymbol{G}}$. The TGV regularization part comprises a first and a second order term (first two terms of Equation (37)). The latter represents the TV component and penalizes the l_1-norm of the second order gradient field, which forces the gradient field $\boldsymbol{G}$ to be piecewise constant. The former enforces

the approximated gradient of the depth map ∇Z to stay in the proximity of the auxiliary gradient field $\boldsymbol{G}$ also through an l_1 penalty.

Given our primal problem in Equation (37) we formulate a primal-dual (PD) problem, which belongs to the class of saddle-point problems, as follows:

$$\min_{\boldsymbol{x}} \max_{\boldsymbol{y}} (K\boldsymbol{x})^T \boldsymbol{y} - F^*(\boldsymbol{y}) + H(\boldsymbol{x}), \tag{40}$$

where H describes the depth and orientation constraints, F defines the TGV component, with its convex conjugate F^* and the linear operator K is defined as:

$$K = \begin{pmatrix} \nabla & -I \\ 0 & \boldsymbol{\nabla} \end{pmatrix}. \tag{41}$$

Our primal variable is denoted as $\boldsymbol{x} \in \mathbb{R}^{M \times N \times 3}$ and $\boldsymbol{y} \in \mathbb{R}^{M \times N \times 6}$ represents the dual variable:

$$\boldsymbol{x} = \begin{bmatrix} Z \\ G_x \\ G_y \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}, \quad \boldsymbol{y}_1 = \begin{bmatrix} y_{1,1} \\ y_{1,2} \end{bmatrix}, \quad \boldsymbol{y}_2 = \begin{bmatrix} y_{2,1} \\ y_{2,2} \\ y_{2,3} \\ y_{2,4} \end{bmatrix}. \tag{42}$$

where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ hold the dual variables of the first and the second term in Equation (37) respectively, with $\boldsymbol{y}_1$ holding two terms for each ∇_x and ∇_y, while $\boldsymbol{y}_2$ holds four terms, each for ∇_x and ∇_y in G_x and G_y. For $F(\boldsymbol{x}) = \alpha ||\boldsymbol{x}||_p$, where $|| \cdot ||_p$ is an l_p-norm, we calculate the convex conjugate as follows:

$$F^*(\boldsymbol{y}) = \delta_{||\cdot||_{2,\infty} \leq \alpha}(\boldsymbol{y}) = \begin{cases} 0, & \text{if } |\boldsymbol{y}_{i,j}|_2 \leq \alpha, \forall i, j \\ \infty, & \text{otherwise,} \end{cases} \tag{43}$$

which is the indicator function of the polar ball. Therefore, F^* and H are defined as follows:

$$F^*(\boldsymbol{y}) = \delta_{||\cdot||_{2,\infty} \leq \alpha_1}(\boldsymbol{y}_1) + \delta_{||\cdot||_{2,\infty} \leq \alpha_0}(\boldsymbol{y}_2), \quad \text{and} \tag{44}$$

$$H(\boldsymbol{x}) = \frac{\alpha}{2} ||Z - \hat{Z}||^2 + \frac{\beta}{2} ||\boldsymbol{G} - \hat{\boldsymbol{G}}||^2. \tag{45}$$

An optimal solution to our hybrid formulation is found with the PD algorithm, as described in Appendix A.2. For a more detailed description of the PD algorithm, see [51].

5. Evaluation

To evaluate all considered algorithms we need two data structures, namely an initial depth map $\hat{Z}$ and surface orientation information such as surface normals $\hat{\boldsymbol{N}}$ or gradients $\hat{\boldsymbol{G}}$. Initial depth maps can be provided by stereo or light field correspondence analysis or other depth scanning methods. An estimate of the surface orientation is calculated using e.g., photometric stereo or polarization imaging. Our discussed hybrid algorithms differ mainly in how the surface orientation information is taken into account. In particular, the penalty functions associated with the surface orientation constraints vary significantly for different methods, as illustrated in Figure 1 through Figure 4c. We proposed two novel methods in Sections 4.6 and 4.7, namely the generalized method of Nehab and the TGV approach. The former balances the surface orientation information with an improved weighting function withing the least squares framework. The latter uses a TGV regularization term combined with a gradient-based approach.

Qualitative and Quantitative Evaluation

In order to perform a comprehensive quantitative evaluation of the above mentioned methods, we considered using synthetic evaluation data that comprises full ground truth (GT) information. The initial depth $\hat{Z}$ for the synthetic evaluation data is given by a ground truth depth map $\hat{Z}_{GT}$, which is thresholded after adding noise:

$$\hat{Z} = \frac{1}{k}\Big[k \cdot (\hat{Z}_{GT} + noise)\Big]. \tag{46}$$

In this study, we considered a normally distributed additive noise as this type of noise well simulates the behavior of matching errors of stereo or light field methods combined with image sensor noise. The noise used in our evaluations has a maximum amplitude of 7% of the depth range. The constant k defines the number of discretization steps and $[\cdot]$ rounds to the nearest integer number $n \in \mathbb{Z}$. These discretization artifacts attempt to simulate the output of the discrete regularized correspondence analysis, which is often applied in real-world scenarios (e.g., as described in [37]).

For the orientation constraints, we assume surface normals $\hat{\boldsymbol{N}}$ derived from the ground truth depth model $\hat{\boldsymbol{N}}_{\boldsymbol{GT}}$ by adding a normally distributed noise with a maximum amplitude of 23% of the normal range in the spherical coordinate system:

$$\hat{\boldsymbol{N}} = \frac{\hat{\boldsymbol{N}}_{\boldsymbol{GT}} + noise}{||\hat{\boldsymbol{N}}_{\boldsymbol{GT}} + noise||} \tag{47}$$

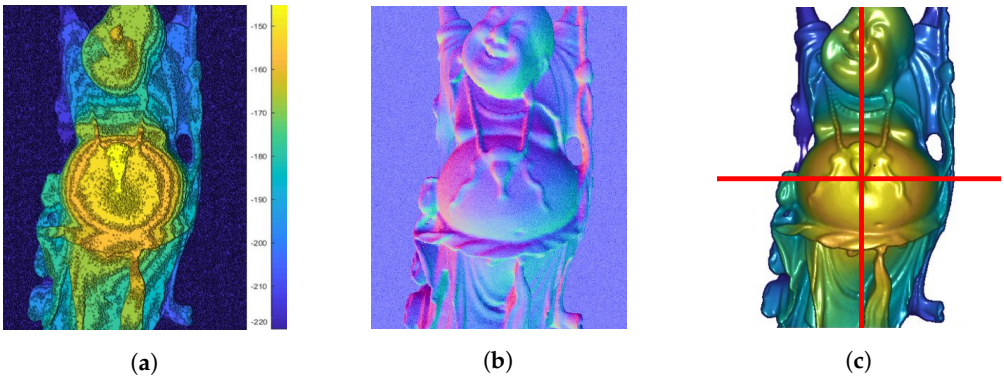
In Figure 5, examples of the evaluation data $\hat{Z}$ and $\hat{\boldsymbol{N}}$ as well as the corresponding ground truth depth $\hat{Z}_{GT}$ are shown. All 3D datasets used for evaluations were taken from the Stanford 3D scanning repository [52] and rendered with POV-Ray [53].

Qualitative comparisons of the methods described in Section 4 are presented in Figures 6 and 7. The left column shows the color coded depth reconstructions as delivered by different methods. The middle two columns show the vertical and horizontal depth profiles as marked in Figure 5c. In each of these plots, the red, gray and black lines indicate the reconstructed depth Z, the initial depth $\hat{Z}$ and the ground truth depth $\hat{Z}_{GT}$, respectively. The error maps showing signed distances to the ground truth depth are provided in the right column. The corresponding geodesic distances from the ground truth surface normals $\hat{N}_{GT}$ for each method are displayed in Figure 8. Results of the quantitative evaluations are shown in Table 2 for gradient-based methods and in Table 3 for normal based methods. For the evaluation we used three datasets from [52]: Buddha (shown in qualitative analysis), Dragon, and Armadillo. The tables hold fractional precisions of two digits in the depth evaluation and a higher fractional precision of four digits in the normal evaluation due to different ranges. The average is calculated using 4 digits after the comma, rounding errors can cause differences in the last digit. The MSE distance in the depth domain is calculated by the quadratic distance between the ground truth depth and the depth result:

$$MSE_Z = d_Z = \frac{1}{MN}||\hat{Z}_{GT} - Z||^2 \tag{48}$$

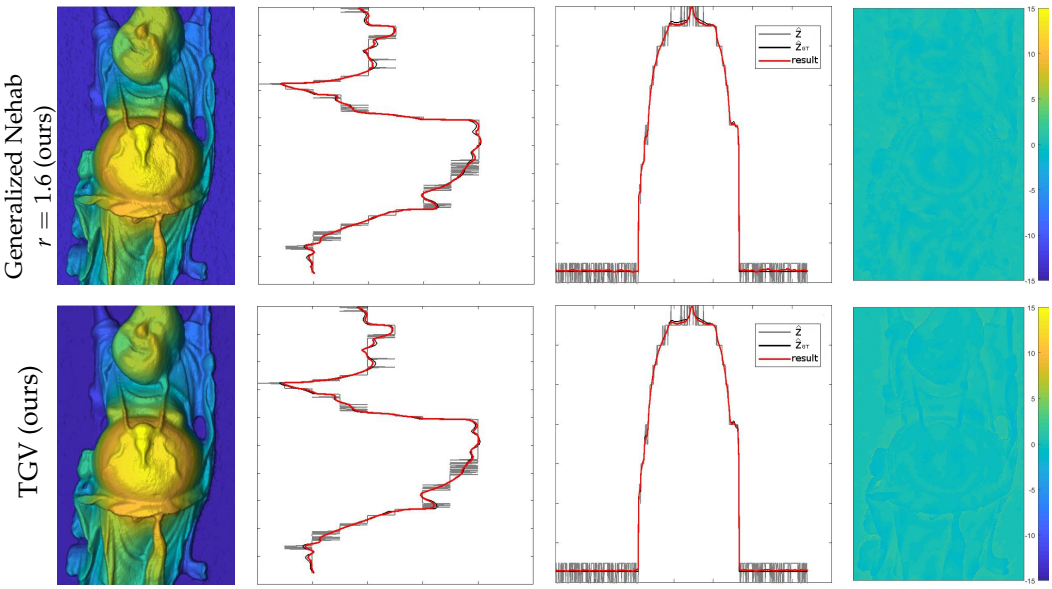
The geodesic distance is calculated as described in Section 4.1 by using the following equation:

$$GEO_N = d_N = \frac{1}{MN}\sum_{i,j}^{MN} acos\left(\frac{\langle \hat{\boldsymbol{N}}_{i,j}, (-(\nabla_x Z_{GT})_{i,j}, -(\nabla_y Z_{GT})_{i,j}, 1)\rangle}{|(-(\nabla_x Z_{GT})_{i,j}, -(\nabla_y Z_{GT})_{i,j}, 1)|_2}\right) \tag{49}$$



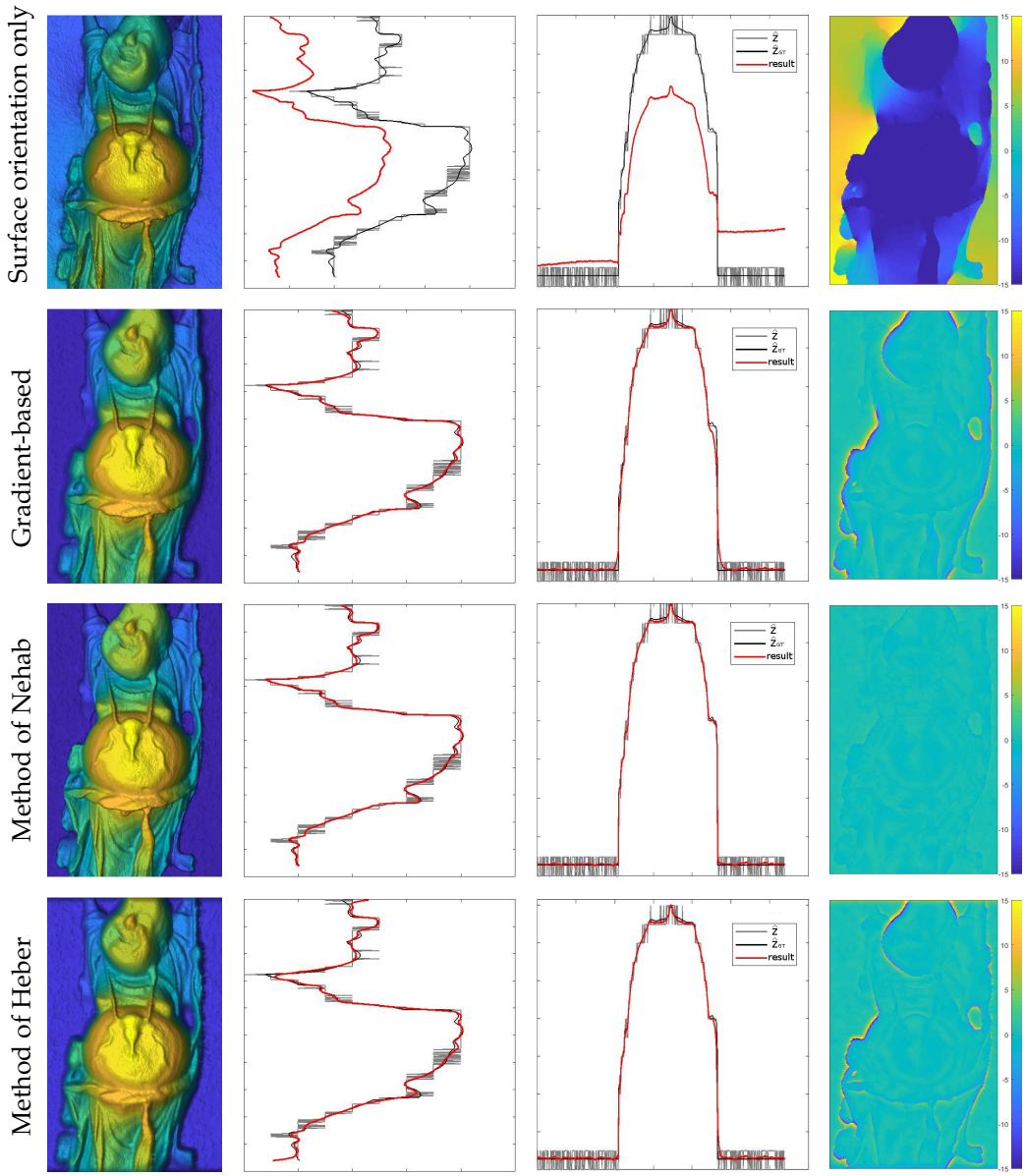
(**a**) (**b**) (**c**)

Figure 5. Example evaluation data, consisting of a noisy quantized depth $\hat{Z}$ and noisy surface normals $\hat{N}$. The ground truth depth $\hat{Z}_{GT}$ is shown with a horizontal and vertical cross-section line, which indicates the positions of evaluations in the following Figure 6. (**a**) Initial depth $\hat{Z}$; (**b**) Given surface normals $\hat{N}$; (**c**) Ground truth depth $\hat{Z}_{GT}$.



(**a**) Depth result (**b**) Vertical line (**c**) Horizontal line (**d**) Error map

Figure 6. Qualitative evaluation of our novel approaches, as introduced in Sections 4.6 and 4.7. The following methods are demonstrated: (1st row) the generalized method of Nehab with $r = 1.6$, (2nd row) the TGV approach. The left column shows the color coded depth reconstructions as delivered by different methods. The middle two columns show the vertical and horizontal depth profiles as marked in Figure 5c. In each of these plots, the red, gray and black lines indicate the reconstructed depth Z, the initial depth $\hat{Z}$ and the ground truth depth $\hat{Z}_{GT}$, respectively. The error maps showing signed distances to the ground truth depth are provided in the right column.



(**a**) Depth result (**b**) Vertical line (**c**) Horizontal line (**d**) Error map

Figure 7. Qualitative evaluation of the state-of-the-art methods described in Section 4. The following methods are demonstrated: (1st row) the gradient-based method using surface orientations only, (2nd row) the combined gradient-based method, (3rd row) the method of Nehab, and (4th row) the method of Heber. The left column shows the color coded depth reconstructions as delivered by different methods. The middle two columns show the vertical and horizontal depth profiles as marked in Figure 5c. In each of these plots, the red, gray and black lines indicate the reconstructed depth Z, the initial depth $\hat{Z}$ and the ground truth depth $\hat{Z}_{GT}$, respectively. The error maps showing signed distances to the ground truth depth are provided in the right column.

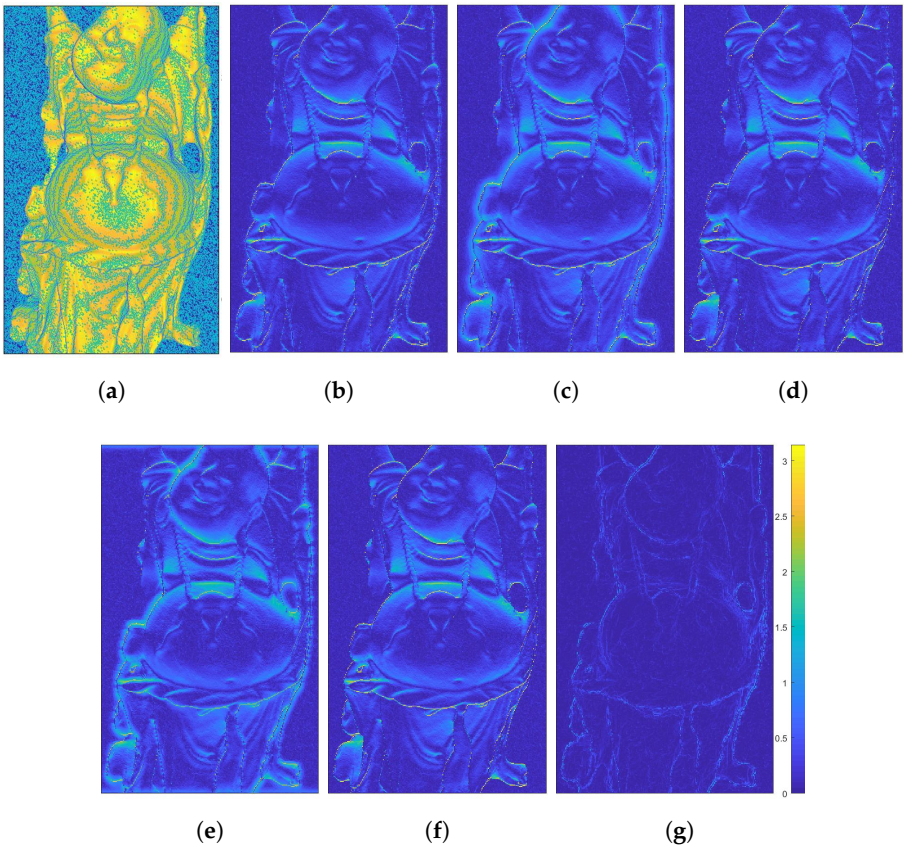


Figure 8. Geodesic distances of the ground truth normals to the normals from the initial depth map $\hat{Z}$ as well as the distances to the normal results of the presented methods (in the same order as displayed in Figures 6 and 7). (**a**) Initial depth $\hat{Z}$; (**b**) Surface orientation only; (**c**) Gradient-based; (**d**) Method of Nehab; (**e**) Method of Heber; (**f**) Generalized Nehab $r = 1.6$ (ours); (**g**) TGV (ours).

As can be seen in Figures 7(1st row) and 8b, using surface orientation information alone in the gradient-based formulation provides visually pleasing detail reconstructions ($d_N = 0.2365$), though it is performing worst in terms of the absolute distance to the ground truth depth with an average MSE of $d_Z = 66.68$. In Figures 7(2nd row) and 8c, one can see that adding a depth constraint to the same gradient-based formulation improves the result drastically ($d_Z = 2.04$ and $d_N = 0.2951$). Nevertheless, this method still shows somewhat low performance around steep edges. Note that each of the demonstrated methods could be used to reconstruct surfaces from surface normals only solely by dropping the depth term. The method of Nehab shown in Figures 7(3rd row) and 8d exhibits the capability to improve the result over the previous methods exploiting a better surface orientation weighting strategy ($d_Z = 0.14$, $d_N = 0.2532$). These methods are optimized using a least squares solver. The evaluation of Heber's method is shown in Figures 7(4th row) and 8e. The method is optimized using gradient descent and reaches average results of $d_Z = 1.79$ and $d_N = 0.3049$, which is significantly worse than the method of Nehab. The results of our generalized method of Nehab with a parametrization $r = 1.6$ are shown in in Figures 6(1st row) and 8f. This method improves robustness over the standard method of Nehab against noise in surface normals and outperforms all other evaluated methods in the absolute depth error domain with $d_Z = 0.11$. As the normal weighting here is in closer vicinity to the geodesic penalty than the method of Nehab, this approach reaches an improved normal error of $d_N = 0.2442$. This method is optimized using a least squares solver. Our novel TGV method, shown in Figures 6(2nd row) and 8g, provides by far the best normal accuracy of $d_N = 0.0666$ and performs among the best in the depth domain ($d_Z = 0.20$). It is optimized with a primal-dual algorithm.

For completion, we additionally show the results of two regularized gradient methods in Table 3. First, we regularize with smoothness as shown in Equation (23), which can be used for sparse data. In our dense case, where we used scalars for the weighting parameters, the normal accuracy shows a minor improvement due to smoothing in return of a weaker depth accuracy. Second, we regularize with the gradient, as shown in Equation (21). With this we can reach an improvement both in the depth and normal accuracy. These regularization terms could be used for all presented methods, which would exceed the scope of this paper. Note that our novel TGV method shows a significantly better performance in both accuracy measurements.

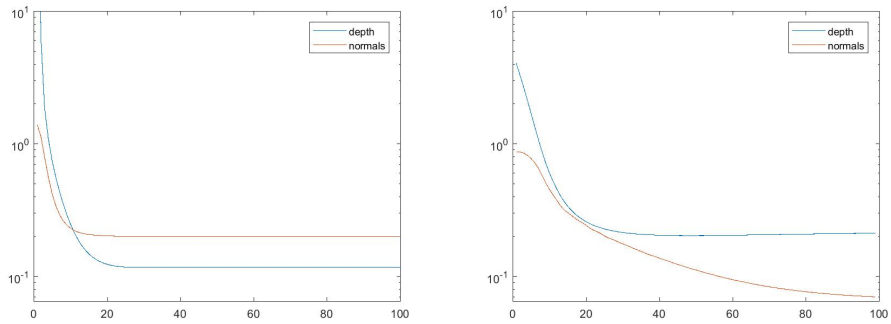
As can be seen in Tables 2 and 3, the two best performing methods are our generalized method of Nehab with $r = 1.6$ (normal based) and our novel TGV method (gradient-based). We show the convergence of both in Figure 9 with respect to the depth error defined in Equation (48) and the normal error as defined in Equation (49). While the generalized Nehab converges in both terms after approximately 25 iterations, the TGV settles at around the same iteration step with a depth error which is performing slightly worse than the other method, but continues to highly optimize the normal error. Note that a gradient-based formulation was chosen to demonstrate graceful properties of our novel TGV approach. It significantly improved the results compared with the standard gradient-based method. Alternatively, for even better performance, other penalty functions such as the generalized method of Nehab, can be chosen and will be a matter of future research.

Table 2. gradient-based: quantitative evaluation of the distance of the optimized depth values to the ground truth depth. The evaluations were performed on objects from the Stanford database [52], which were rendered using POV-Ray [53]. Evaluated are the MSE to the ground truth depth and the geodesic distance to the ground truth normals.

Gradient-Based	Dataset Method	$\hat{Z}$	Surface Orientation Only	Gradient Based	Gradient Based + Reg. with Laplacian Smoothness (Equation (23))	Gradient Based + Reg. with Gradient (Equation (21))	TGV (Ours)
Depth $[MSE_Z]$	Dragon	4.23	34.05	2.04	2.15	1.85	**0.19**
	Buddha	4.85	117.29	2.12	2.25	2.01	**0.22**
	Armadillo	4.60	48.71	1.95	2.06	1.83	**0.18**
	Average	4.53	66.68	2.04	2.15	1.90	**0.20**
Normals $[GEO_N]$	Dragon	0.8226	0.2776	0.3344	0.3200	0.3017	**0.0664**
	Buddha	0.8767	0.1922	0.2535	0.2339	0.2125	**0.0668**
	Armadillo	0.8611	0.2397	0.2973	0.2797	0.2599	**0.0666**
	Average	0.8535	0.2365	0.2951	0.2779	0.2580	**0.0666**

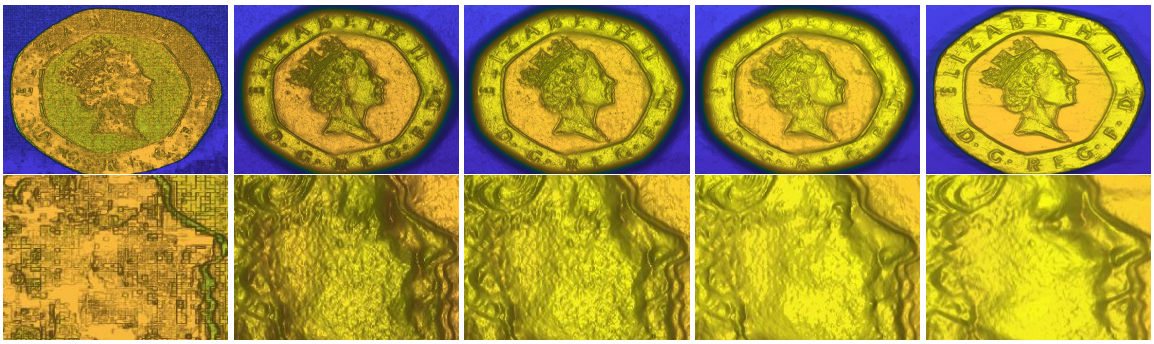
Table 3. Normal based: quantitative evaluation of the distance of the optimized depth values to the ground truth depth. The evaluations were performed on objects from the Stanford database [52], which were rendered using POV-Ray [53]. Evaluated are the MSE to the ground truth depth and the geodesic distance to the ground truth normals.

Normal Based	Dataset Method	$\hat{Z}$	Method of Heber	Method of Nehab	Generalized Nehab $r = 1.6$ (Ours)
Depth $[MSE_Z]$	Dragon	4.23	2.01	0.13	**0.10**
	Buddha	4.85	1.60	0.15	**0.12**
	Armadillo	4.60	1.75	0.13	**0.10**
	Average	4.53	1.79	0.14	**0.11**
Normals $[GEO_N]$	Dragon	0.8226	0.3474	0.2941	**0.2849**
	Buddha	0.8767	0.2579	0.2102	**0.2013**
	Armadillo	0.8611	0.3094	0.2553	**0.2464**
	Average	0.8535	0.3049	0.2532	**0.2442**



(**a**) Generalized Nehab $r = 1.6$ (ours) (**b**) TGV (ours)

Figure 9. Convergence analysis of the two winning methods: the generalized Nehab with $r = 1.6$ (**a**) and the TGV (**b**). Distances to the ground truth depth and surface normals are shown after each iteration, plotted with a logarithmic y-axis for better visibility.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 10. Real world evaluation of a coin object acquired by a multi-line scanner. The acquisition setup was previously described in detail in [37]. (**a**) $\hat{Z}$; (**b**) Gradient-based; (**c**) Method of Nehab; (**d**) Generalized Nehab (ours); (**e**) TGV (ours).

We demonstrated the described algorithms with different surface orientation weighting on a real world example, as shown in Figure 10. This object was acquired with a multi-line scanner, the hybrid light field - photometric stereo acquisition framework described in [37]. In this very specific case we only have one normal direction. An example how to deal with this data in the gradient-based approach is weighting the gradient vector in Equation (23) in the missing direction with zero. Specifics of the multi-line-scanning environment are out of the scope of this paper and are a matter of future research.

6. Conclusions

We presented a review and classification of methods combining depth and surface orientation data (normals or gradients), in order to reach an improved surface depth estimation. State-of-the-art methods differ mostly in the formulation of the surface orientation constraint (see Section 4) and capabilities of the method-specific solvers.

We illustrated the differences between various formulations of the surface orientation constraint and explained performance discrepancies. Furthermore, we used our findings to introduce a generalization of the method of Nehab (see Section 4.6) that significantly outperforms other methods in terms of absolute depth accuracy. Additionally, we introduced a novel method based on TGV (see Section 4.7), which outperforms all other methods in the surface normal domain and shows a competitive performance in the depth accuracy. While our generalized Nehab method converges faster (see Figure 9) and gives the most accurate result in the depth domain, our TGV based approach

refines the surface orientation further and converges at the most accurate orientation result with a high accuracy in the depth domain.

Further research will include the specialization on line-scanning algorithms, TGV weighting adaption and computational acceleration. With specialized hybrid algorithms that fit data from line scanning sensors we will determine a solution with incomplete surface orientation data. The surface orientation constraint of our TGV formulation is currently gradient-based. Plugging in another formulation with a better balanced normal weighting could improve the results even further and will be a matter of future research. Furthermore we will focus on computational acceleration of the proposed algorithms, where we will exploit their inherent structure to achieve efficient parallelization.

Acknowledgments: This work is supported by the research initiative 'Intelligent Vision Austria' with funding from the Austrian Federal Ministry of Science, Research and Economy and the Austrian Institute of Technology.

Author Contributions: D. Antensteiner and S. Štolc conceived and designed the experiments; D. Antensteiner performed the experiments; D. Antensteiner and S. Štol analyzed the data; T. Pock contributed reagents/materials/analysis tools; D. Antensteiner wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CRF	Conditional Random Field
GT	Ground Truth
FISTA	Fast Iterative Shrinkage Thresholding Algorithm
ISTA	Iterative Shrinkage Thresholding Algorithm
MRF	Markov Random Field
MSE	Mean Squared Error
PD	Primal-Dual
SfF	Shape from Focus
SfM	Structure from Motion
TGV	Total Generalized Variation
ToF	Time of Flight
TV	Total Variation

Appendix A. Optimization Theory

In this section we show two advanced optimization methods used to fuse depth and surface orientation as well as the type of functions they can optimize.

Appendix A.1. Accelerated Proximal Gradient Method

The fast iterative shrinkage thresholding algorithm (FISTA) minimizes an objective function by using proximal operators (see [54]). It works on general conditions as non-smooth convex functions and can be very fast, as there exist simple proximal operators for various energy functions. Proximal gradient methods consider optimization problems with the objective split in two components:

$$\min_x F(x) + H(x), \tag{A1}$$

where $F : \mathbb{R}^N \to \mathbb{R}$ and $H : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex functions and where one of the functions (F) is additionally differentiable. The proximal gradient problem uses a step size $\lambda^k > 0$ and is denoted as follows:

$$x^{k+1} \leftarrow prox_{\lambda^k H}(x^k - \lambda^k \nabla F(x^k)). \tag{A2}$$

If ∇F is Lipschitz continuous with a constant L and a fixed step size $\lambda^k = \lambda \in (0, 1/L]$ is used, the method converges in $\mathcal{O}(1/k)$. Accelerated proximal gradient methods have an additional extrapolation step in the algorithm and are denoted as follows:

$$y^{k+1} \leftarrow x^k + w^k(x^k - x^{k-1}), \tag{A3}$$
$$x^{k+1} \leftarrow prox_{\lambda^k H}(y^{k+1} - \lambda^k \nabla F(y^{k+1})), \tag{A4}$$

where λ^k denotes the step size and w^k an extrapolation parameter which is described as:

$$w^k = \frac{k-1}{k+2}. \tag{A5}$$

FISTA is one example of such an accelerated proximal gradient method and described in detail in [49]. In [41], the FISTA algorithm is used to optimize the method of Heber, which we discuss in Section 4.4.

Appendix A.2. Primal-Dual

Primal-dual methods can be used to solve a large number of optimization problems. The considered saddle-point problem, which is a primal-dual formulation has the following primal form:

$$\min_{x \in X} F(Kx) + H(x). \tag{A6}$$

The corresponding saddle point problem is described as:

$$\min_{x \in X} \max_{y \in Y} \langle Kx, y \rangle + H(x) - F^*(y). \tag{A7}$$

We are using the Chambolle-Pock [55] primal-dual method and compute the proximal maps for H and F^*. A proximal descent is computed in the primal variable x and, in an alternating manner, a proximal ascent in the dual variable y:

$$\begin{cases} x^{k+1} \leftarrow \text{prox}_{\tau H}(x^k - \tau K^T y^k) \\ y^{k+1} \leftarrow \text{prox}_{\sigma F^*}(y^k + \sigma K(2x^{k+1} - x^k)) \end{cases} \tag{A8}$$

The convergence rate is $\mathcal{O}(1/N)$, and can be improved if H or F^* are uniformly convex to $\mathcal{O}(1/N^2)$. Both, primal and the dual being uniformly convex the algorithm shows a linear convergence $\mathcal{O}(e^N)$ [55–57]. We discuss the primal-dual optimization on our hybrid TGV formulation in Section 4.7.

References

1. Hirschmüller, H.; Innocent, P.R.; Garibaldi, J. Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. *Int. J. Comput. Vis.* **2002**, *47*, 229–246.
2. Hirschmuller, H. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341.
3. Haller, I.; Pantilie, C.; Oniga, F.; Nedevschi, S. Real-Time Semi-Global Dense Stereo Solution with Improved Sub-Pixel Accuracy. In Proceedings of the 2010 IEEE Intelligent Vehicles Symposium, San Diego, CA, USA, 21–24 June 2010; pp. 369–376.
4. Je, C.; Park, H.M. Optimized Hierarchical Block Matching for Fast and Accurate Image Registration. *Signal Process. Image Commun.* **2013**, *28*, 779–791.
5. Yang, X.; Chen, X.; Xi, J. Block Based Dense Stereo Matching Using Adaptive Cost Aggregation and Limited Disparity Estimation. In Proceedings of the 2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Turin, Italy, 22–25 May 2017; pp. 1–6.

6. Chen, X.; Li, D.; Zou, J. Depth Estimation of Stereo Matching Based on Microarray Camera. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; pp. 108–112.
7. Bolles, R.C.; Baker, H.H.; Marimont, D.H. *Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion*; Springer: Berlin, Germany, 1987; Volume 1, pp. 7–55.
8. Wanner, S.; Straehle, C.; Goldluecke, B. Globally Consistent Multi-Label Assignment on the Ray Space of 4D Light Fields. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013; pp. 1011–1018.
9. Kim, C.; Zimmer, H.; Pritch, Y.; Sorkine-Hornung, A.; Gross, M. Scene Reconstruction from High Spatio-Angular Resolution Light Fields. *ACM Trans. Graph.* **2013**, *32*, 73:1–73:12, doi:10.1145/2461912.2461926.
10. Tosic, I.; Berkner, K. Light Field Scale-Depth Space Transform for Dense Depth Estimation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Columbus, OH, USA, 23–28 June 2014; pp. 435–442.
11. Tao, M.W.; Wang, T.C.; Malik, J.; Ramamoorthi, R. Depth Estimation for Glossy Surfaces with Light-Field Cameras. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 1–14.
12. Longuet-Higgins, H.C. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1987; pp. 61–62.
13. Snavely, N.; Seitz, S.M.; Szeliski, R. *Photo Tourism: Exploring Photo Collections in 3D*; ACM: New York, NY, USA, 2006; Volume 25, pp. 835–846.
14. Park, J.; Kim, H.; Tai, Y.W.; Brown, M.S.; Kweon, I. High Quality Depth Map Upsampling for 3D-ToF Cameras. In Proceedings of the 2011 IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 1623–1630.
15. Woodham, R.J. Photometric Method for Determining Surface Orientation from Multiple Images. *Int. Soc. Opt. Photonics* **1980**, *19*, 191139, doi:10.1117/12.7972479.
16. Frankot, R.T.; Chellappa, R. A Method for Enforcing Integrability in Shape From Shading Algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **1988**, *10*, 439–451.
17. Wu, T.P.; Tang, C.K. Dense Photometric Stereo Using a Mirror Sphere and Graph Cut. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 140–147.
18. Malis, E.; Vargas, M. *Deeper Understanding of the Homography Decomposition for Vision-Based Control*; Research Report RR-6303; INRIA: Rocquencourt, France, 25 September 2007.
19. Köser, K. Geometric Estimation with Local Affine Frames and Free-form Surfaces. Ph.D. Thesis, Christian-Albrechts-Universität zu Kiel, Kiel, Germany, 2008.
20. Barath, D.; Molnar, J.; Hajder, L. Novel Methods for Estimating Surface Normals from Affine Transformations. In *Communications in Computer and Information Science*; Braz, J., Pettré, J., Richard, P., Kerren, A., Linsen, L., Battiato, S., Imai, F., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 316–337.
21. Eigen, D.; Fergus, R. Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
22. Wu, C.; Wilburn, B.; Matsushita, Y.; Theobalt, C. High-Quality Shape from Multi-View Stereo and Shading Under General Illumination. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 969–976.
23. Zhang, L.; Curless, B.; Hertzmann, A.; Seitz, S.M. Shape and Motion under Varying Illumination: Unifying Structure from Motion, Photometric Stereo, and Multi-view Stereo. In Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; pp. 618–625.
24. Joshi, N.; Kriegman, D.J. Shape from Varying Illumination and Viewpoint. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–7.
25. Esteban, C.H.; Vogiatzis, G.; Cipolla, R. Multiview Photometric Stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 548–554.
26. Nehab, D.; Rusinkiewicz, S.; Davis, J.; Ramamoorthi, R. Efficiently Combining Positions and Normals for Precise 3D Geometry. *ACM Trans. Graph.* **2005**, *24*, 536–543.

27. Haque, S.M.; Chatterjee, A.; Govindu, V.M. High Quality Photometric Reconstruction Using a Depth Camera. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2283–2290.
28. Park, J.; Sinha, S.N.; Matsushita, Y.; Tai, Y.W.; Kweon, I.S. Robust Multiview Photometric Stereo Using Planar Mesh Parameterization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1591–1604.
29. Han, Y.; Lee, J.Y.; Kweon, I.S. High Quality Shape from a Single RGB-D Image under Uncalibrated Natural Illumination. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 1617–1624.
30. Chatterjee, A.; Govindu, V.M. Photometric refinement of depth maps for multi-albedo objects. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 933–941.
31. Shi, B.; Inose, K.; Matsushita, Y.; Tan, P.; Yeung, S.K.; Ikeuchi, K. Photometric Stereo Using Internet Images. In Proceedings of the 2014 2nd International Conference on 3D Vision, Tokyo, Japan, 8–11 December 2014.
32. Yu, H. Edge-preserving Photometric Stereo via Depth Fusion. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CVPR '12, Providence, RI, USA, 16–21 June 2012; IEEE Computer Society: Washington, DC, USA, 2012; pp. 2472–2479.
33. Ti, C.; Yang, R.; Davis, J.; Pan, Z. Simultaneous Time-of-Flight sensing and photometric stereo with a single ToF sensor. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 4334–4342.
34. Yu, L.F.; Yeung, S.K.; Tai, Y.W.; Lin, S. Shading-Based Shape Refinement of RGB-D Images. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, CVPR'13, Portland, OR, USA, 23–28 June 2013; IEEE Computer Society: Washington, DC, USA, 2013; pp. 1415–1422.
35. Wu, C.; Zollhöfer, M.; Nießner, M.; Stamminger, M.; Izadi, S.; Theobalt, C. Real-Time Shading-Based Refinement for Consumer Depth Cameras. *ACM Trans. Graph.* **2014**, *33*, doi:10.1145/2661229.2661232.
36. Kadambi, A.; Taamazyan, V.; Shi, B.; Raskar, R. Polarized 3D: High-Quality Depth Sensing with Polarization Cues. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 3370–3378.
37. Antensteiner, D.; Štolc, S.; Valentin, K.; Blaschitz, B.; Huber-Mörk, R.; Pock, T. High-Precision 3D Sensing with Hybrid Light Field and Photometric Stereo Approach in Multi-Line Scan Framework. In Proceedings of the IS&T International Symposium on Electronic Imaging: Intelligent Robotics and Industrial Applications using Computer Vision 2017, Burlingame, CA, USA, 29 January–2 February 2017.
38. Horn, B.K.P.; Brooks, M.J. (Eds.) *Shape from Shading*; MIT Press: Cambridge, MA, USA, 1989.
39. Štolc, S.; Soukup, D.; Holländer, B.; Huber-Mörk, R. Depth and All-In-Focus Imaging by a Multi-Line-Scan Light-Field Camera. *J. Electron. Imaging* **2014**, *23*, 19–23.
40. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson Surface Reconstruction. In Proceedings of the Fourth Eurographics Symposium on Geometry Processing, SGP'06, Cagliari, Sardinia, 26–28 June 2006; Eurographics Association: Aire-la-Ville, Switzerland, 2006; pp. 61–70.
41. Heber, S. Variational Shape from Unconventional Imaging Devices. Ph.D. Thesis, Graz University of Technology, Graz, Austria, 18 May 2015.
42. Scherr, T. Gradient-Based Surface Reconstruction and the Application to Wind Waves. Master's Thesis, Ruprecht-Karls-University Heidelberg, Heidelberg, Germany, 3 November 2017.
43. Rostami, M.; Michailovich, O.; Wang, Z. Gradient-Based Surface Reconstruction Using Compressed Sensing. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; pp. 913–916.
44. Harker, M.; O'Leary, P. Least Squares Surface Reconstruction from Measured Gradient Fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008, Anchorage, AK, USA, 23–28 June 2008.
45. Du, Z.; Robles-Kelly, A.; Lu, F. Robust Surface Reconstruction from Gradient Field Using the L1 Norm. In Proceedings of the 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007), Glenelg, Australia, 3–5 December 2007; pp. 203–209.
46. Agrawal, A.; Chellappa, R.; Raskar, R. An Algebraic Approach to Surface Reconstruction from Gradient Fields. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005; Volume 1, pp. 174–181.

47. Horn, B.K.P.; Brooks, M.J. The Variational Approach to Shape from Shading. *Comput. Vis. Graph. Image Process.* **1986**, *33*, 174–208.
48. Nesterov, Y.E. *Introductory Lectures on Convex Optimization: A Basic Course*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2003.
49. Beck, A.; Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202.
50. Bredies, K.; Kunisch, K.; Pock, T. Total Generalized Variation. *SIAM J. Imaging Sci.* **2010**, *3*, 492–526.
51. Chambolle, A.; Pock, T. *An Introduction to Continuous Optimization for Imaging*; Cambridge University Press: Cambridge, UK, 2016; Volume 25, pp. 161–319.
52. The Stanford 3D Scanning Repository. Available online: http://graphics.stanford.edu/data/3Dscanrep/ (acccessed on 10 November 2017).
53. Persistence of Vision Pty. Ltd. (2004) [Computer Software]. Available online: http://www.povray.org/download/ (acccessed on 10 November 2017).
54. Parikh, N.; Boyd, S. *Proximal Algorithms*; Now Publishers Inc.: Breda, The Netherlands, 2013.
55. Chambolle, A.; Pock, T. *A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging*; Springer: Dordrecht, The Netherlands, 2011; Volume 40, pp. 120–145.
56. Nesterov, Y. Smooth Minimization of Non-Smooth Functions. *Math. Program.* **2005**, *103*, 127–152.
57. Beck, A.; Teboulle, M. *Gradient-Based Algorithms with Applications to Signal Recovery Problems*; Cambridge University Press: Cambridge, UK, 2010; pp. 42–88.

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).