In J.-F. Aujol, M. Nikolova, N. Papadakis (Eds.):
Scale Space and Variational Methods in Computer Vision.
Lecture Notes in Computer Science, Vol. 9087, pp. 551–562, Springer, Berlin, 2015.
The final publication is available at link.springer.com.

Multiview Depth Parameterisation with Second Order Regularisation

Christopher Schroers, David Hafner, and Joachim Weickert

Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Campus E1.7, Saarland University, 66041 Saarbrücken, Germany
{schroers,hafner,weickert}@mia.uni-saarland.de

Abstract. In this paper we consider the problem of estimating depth maps from multiple views within a variational framework. Previous work has demonstrated that multiple views improve the depth reconstruction, and that higher order regularisers model a good prior for typical real-world 3D scenes. We build on these findings and stress an important aspect that has not been considered in variational multiview depth estimation so far: We investigate several parameterisations of the unknown depth. This allows us to show, both analytically and experimentally, that directly work with depth values introduces an undesirable bias. As a remedy, we reveal that an inverse depth parameterisation is generally preferable. Our analysis clearly points out its benefits w.r.t. the data and the smoothness term. We verify these theoretical findings by means of experiments.

1 Introduction

The task of reconstructing 3D scenes from a number of images along with corresponding camera poses is commonly referred to as *multiview stereo*. It is important for a variety of applications, and thus has received a huge amount of attention over the last decades. One can approach the multiview stereo problem by dividing it into the following two steps: First, one computes depth maps for a number of input images. Second, these depth maps are merged with a volumetric approach, see e.g. [1–3]. In this way, the multiview stereo problem constitutes a common example, where one is interested in obtaining a depth map given multiple views. This is the problem we focus on in our paper.

Related Work. Ignoring the fact that multiple views are available, variational stereo algorithms that consider image pairs can be regarded as related work, see e.g. [4–9]. While these variational formulations compute disparities relying on a first order regularisation, higher order regularisation has proven to be a very successful strategy for many applications [10–12]. Often, coupled formulations are used instead of directly implementing a higher order regulariser. Popular variants for this are total generalised variation [13] or an approach as in [11]. Also infimal convolution is a much related alternative, where first ideas of this can be found in [14]. Recently, Ranftl et al. demonstrated the benefits of second order regularisation in the context of optic flow [12] and stereo [10].

However, considering only two of the multiple views discards a lot of the available information. Unfortunately, it is not convenient to extend the concept of computing disparities to a general multiview setting. Hence, there are a number of variational formulations that directly estimate depth from multiple views. Such methods have shown the benefits of using multiple images in the process of depth map estimation. To the best of our knowledge, the basic idea of considering multiple views to estimate a single depth map within a variational formulation is almost two decades old and goes back to Robert and Deriche [15]. They employed a quadratic data term along with a nonquadratic regulariser that is able to preserve depth discontinuities. More recently, Stühmer et al. [16] presented a similar formulation with a robust penaliser for the smoothness term as well as the data term. Instead of the brightness constancy, assumed by [15] and [16], Semerijan [17] uses a gradient constancy assumption and a finite element discretisation.

All of the aforementioned approaches are directly parameterised by the unknown depth. However, in related problems such as monocular SLAM, an inverse depth parameterisation of point features has been shown to be beneficial [18]. Also the dense tracking and mapping approach of Newcombe et al. uses inverse depth to compute cost values in a discrete cost volume [19] and the recently developed LSD-SLAM estimates probabilistic semi-dense inverse depth maps [20].

Contributions. While existing variational multiview formulations [15–17] directly compute the unknown depth from a number of arbitrarily placed cameras, we generalise them by introducing a depth parameterisation. This allows us to efficiently analyse advantages and drawbacks of different parameterisations such as a direct depth parameterisation and an inverse depth parameterisation. More specifically, we analyse two important aspects: On the one hand, the choice of parameterisation is important when considering the linearisation of the data term in a variational framework. Here, we show that for common camera setups, the inverse depth parameterisation is preferable. On the other hand, the choice of parameterisation is also important in the smoothness term, especially in the presence of second order regularisation. Here, we show that for an inverse depth parameterisation, piecewise affine functions correspond to piecewise planar surfaces. This is not the case for a direct depth parameterisation. We give deep insights into the introduced bias by analysing the shape operator of the corresponding 3D surface.

Paper Organisation. In Section 2, we present a variational formulation for the estimation of depth maps from multiple views with an arbitrary parameterisation. Subsequently, we analyse different parameterisations in detail (Section 3). In Section 4, we discuss the minimisation. Finally, we show experimental results (Section 5) before we conclude our work (Section 6).

2 Variational Multiview Depth Estimation

In this section, we describe a variational framework that allows the estimation of a depth map d from multiple views under an arbitrary parameterisation. To this end, we express d as the composition of an unknown $\rho : \Omega \to \mathbb{R}_+$ and a parameterisation $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $d = \phi \circ \rho$. Then our energy functional has the form

$$E(\rho, \boldsymbol{w}) = \int_\Omega \Big(D(\phi \circ \rho) \, + \, \alpha \; S(\rho, \boldsymbol{w}) \Big) \, \mathrm{d}\boldsymbol{x}, \tag{1}$$

with a data term $D(\phi\circ\rho)$, a smoothness term $S(\rho, \boldsymbol{w})$, and a positive smoothness weight α. Since we apply second order regularisation in terms of a coupling model, we require the additional coupling variable $\boldsymbol{w}$. In the following sections, we explain our model components in more detail.

Data Term. Let us assume we are given n colour images $\boldsymbol{f}_1, \dots, \boldsymbol{f}_n$ and a reference image $\boldsymbol{f}_0$. The task of the data term $D(\phi\circ\rho)$ is to enforce photoconsistency between all available views. To this end, we first introduce a function $\boldsymbol{g}_i(\boldsymbol{x}, \phi\circ\rho)$ that maps a location $\boldsymbol{x} \in \Omega$ in the reference frame $\boldsymbol{f}_0$ with its depth $(\phi \circ \rho)(\boldsymbol{x})$ to the corresponding location in another image $\boldsymbol{f}_i$. This allows to model the assumption that corresponding points $\boldsymbol{x}$ and $\boldsymbol{g}_i(\boldsymbol{x}, \phi\circ\rho)$ have similar colour values as follows:

$$D(\phi \circ \rho) = \frac{1}{n} \sum_{i=1}^{n} \Psi\Big(\| \boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho)) - \boldsymbol{f}_0(\boldsymbol{x}) \|^2 \Big) \,, \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm and the function $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ provides a robust penalisation. A common choice is $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$, which approximates an L_1 data term for $\epsilon \to 0$.

Smoothness Term. Higher order regularisation has shown its potential in several applications. Essentially, there are two possibilities to design such regularisers: Either by a direct penalisation of higher order derivatives or by introducing a coupling variable. We opt for the second choice that results in the smoothness term

$$S(\rho, \boldsymbol{w}) = \Psi\big(\| \boldsymbol{\nabla} \rho - \boldsymbol{w} \|^2 \big) + \beta \, \Psi\big(\| \boldsymbol{\mathcal{J}} \boldsymbol{w} \|_F^2 \big) \,, \tag{3}$$

where $\boldsymbol{\nabla}$ is the spatial gradient, $\boldsymbol{\mathcal{J}}$ the Jacobian, and $\|\cdot\|_F$ the Frobenius norm.

Since our main focus is the analysis of different parameterisations, we restrict ourselves to the discussed model assumptions. Once the parameterisations are well understood, they can be incorporated in more sophisticated methods that rank favourably in public benchmark systems.

3 Depth Parameterisations

Before analysing possible parameterisations ϕ in (1), we briefly explain the pinhole camera model as it builds the basis for our analysis. Subsequently, we consider the backprojection of constant and affine patches for each parameterisation. This yields important insights on the effect of the regularisation on the resulting surface. Finally, we treat the effects of different paramterisations on the data term.

3.1 Pinhole Camera Model

With homogeneous coordinates, the projection by a pinhole camera model can be described by the linear map $\boldsymbol{P} \in \mathbb{R}^{3\times 4}$:

$$\boldsymbol{P} = \boldsymbol{K} \begin{pmatrix} \boldsymbol{R}\, \boldsymbol{t} \end{pmatrix}, \tag{4}$$

where $\boldsymbol{R} \in SO(3)$ is a rotation matrix and $\boldsymbol{t} \in \mathbb{R}^3$ is a translation, such that the blockmatrix $\begin{pmatrix} \boldsymbol{R}\, \boldsymbol{t} \end{pmatrix}$ describes the extrinsic camera parameters. On the other hand, the matrix

$$\boldsymbol{K} = \begin{pmatrix} k_x & 0 & u \\ 0 & k_y & v \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

contains the intrinsic camera parameters k_x and k_y, which specify the focal length, and the principal point $(u, v)^\top$. With this notation, we express the projection of a 3D point $\boldsymbol{X} \in \mathbb{R}^3$ to a point $\boldsymbol{x} \in \mathbb{R}^2$ in the image plane by

$$\boldsymbol{x} = \boldsymbol{\pi}\big(\boldsymbol{P}\tilde{\boldsymbol{X}}\big), \tag{6}$$

where $\tilde{\boldsymbol{X}} = (\boldsymbol{X}^\top, 1)^\top$ is the homogeneous version of $\boldsymbol{X}$. We use this notation to denote homogeneous coordinates throughout the whole paper. The function $\boldsymbol{\pi}(a, b, c) = ({}^{a}\!/\!{}_{c},\ {}^{b}\!/\!{}_{c})^\top$ maps a homogeneous coordinate to its Euclidean counterpart.

3.2 Backprojection of Parameterised Depth Maps

The previous section showed how to project a 3D point to the image plane. Here we are interested in the other way around, i.e. the backprojection. We analyse the following parameterisations:

$$\text{(i) direct depth:} \quad \phi(r) = r, \tag{7}$$

$$\text{(ii) inverse depth:} \quad \phi(r) = {}^{1}\!/\!{}_{r}. \tag{8}$$

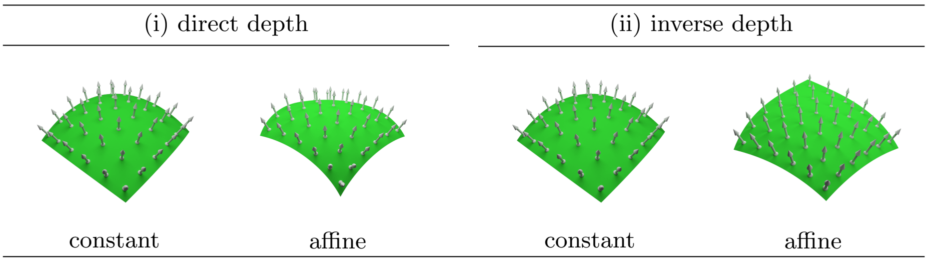

Fig. 1. Resulting surfaces when backprojecting along the line of sight.

In each case, there is a further design choice that we want to analyse, namely the choice of the distance, in which we measure. Basically, there are two meaningful possibilities to compute a backprojection:

(a) along the line of sight: $$\boldsymbol{\ell}\big(\boldsymbol{x}, \phi \circ \rho\big) = \left\|\boldsymbol{K}^{-1}\tilde{\boldsymbol{x}}\right\|^{-1} \boldsymbol{K}^{-1}\tilde{\boldsymbol{x}} \cdot (\phi \circ \rho)(\boldsymbol{x}), \tag{9}$$

(b) along the optical axis: $$\boldsymbol{s}\big(\boldsymbol{x}, \phi \circ \rho\big) = \boldsymbol{K}^{-1}\tilde{\boldsymbol{x}} \cdot (\phi \circ \rho)(\boldsymbol{x}). \tag{10}$$

Figure 1 shows the resulting surfaces when backprojecting a constant and an affine function along the line of sight. Note that both parameterisations (i) and (ii) map constant and affine functions to curved surfaces. This means that a first order regulariser would already introduce an unwanted bias because it favours a (piecewise) constant ρ and thus curved surfaces. Therefore, we will not further consider parameterisations along the line of sight in our context.

In contrast, Figure 2 shows that both parameterisations along the optical axis map constant depth functions to surfaces with constant depth, and thus seem to be reasonable choices when employing a first order regularisation. However, considering an affine function (with a nonzero slope), we see that the depth parameterisation does not create a planar surface whereas the inverse depth parameterisation does. In the following sections, we analyse this in detail to get a better understanding of both choices.

3.3 Analysis of Backprojected Depth Maps

Let us consider (10) as a mapping from some parameter space Ω to a surface M, i.e. $\boldsymbol{s} : \Omega \subset \mathbb{R}^2 \to M \subset \mathbb{R}^3$. Generally, the tangent plane of a regular parameterised surface corresponding to a point $(x_0, y_0)^\top$ is spanned by the two tangent vectors

$$\boldsymbol{s}_x = \frac{\partial \boldsymbol{s}}{\partial x} \quad \text{and} \quad \boldsymbol{s}_y = \frac{\partial \boldsymbol{s}}{\partial y}, \tag{11}$$

evaluated at $(x_0, y_0)^\top$. The *first fundamental form* describes the inner product of two tangent vectors. It can be represented by the symmetric matrix

$$\boldsymbol{I} = \begin{pmatrix} \langle \boldsymbol{s}_x, \boldsymbol{s}_x \rangle & \langle \boldsymbol{s}_x, \boldsymbol{s}_y \rangle \\ \langle \boldsymbol{s}_y, \boldsymbol{s}_x \rangle & \langle \boldsymbol{s}_y, \boldsymbol{s}_y \rangle \end{pmatrix}, \tag{12}$$

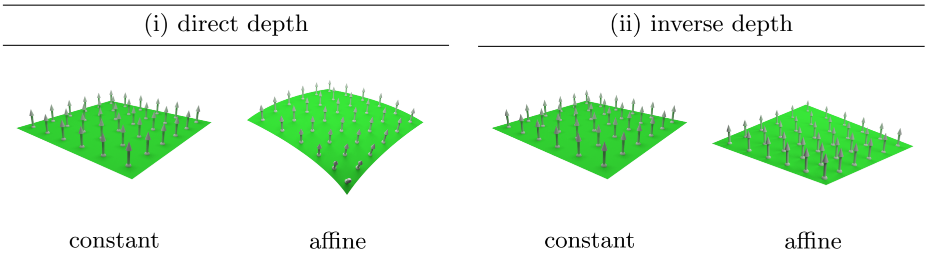

Fig. 2. Resulting surfaces when backprojecting along the optical axis.

and allows the evaluation of metric properties such as the surface area. Similarly, the *second fundamental form* is important for describing curvatures. It can be represented by the symmetric matrix

$$\boldsymbol{II} = \begin{pmatrix} \langle \boldsymbol{n}, \boldsymbol{s}_{xx} \rangle & \langle \boldsymbol{n}, \boldsymbol{s}_{xy} \rangle \\ \langle \boldsymbol{n}, \boldsymbol{s}_{yx} \rangle & \langle \boldsymbol{n}, \boldsymbol{s}_{yy} \rangle \end{pmatrix}, \tag{13}$$

where $\boldsymbol{n}$ is the unit surface normal

$$\boldsymbol{n} = \frac{\boldsymbol{s}_x \times \boldsymbol{s}_y}{\|\boldsymbol{s}_x \times \boldsymbol{s}_y\|}. \tag{14}$$

The composition of the first and second fundamental form defines the *shape operator*

$$\boldsymbol{S} = \boldsymbol{I}^{-1}\, \boldsymbol{II}. \tag{15}$$

It allows to evaluate the Gaussian curvature K and the mean curvature H, which are given by $\det(\boldsymbol{S})$ and $1/2 \cdot \operatorname{tr}(\boldsymbol{S})$, respectively.

Direct Depth. Let us first consider the direct depth parameterisation where the unknown ρ corresponds to the sought depth. With this, we analyse the resulting surface in the case that the *depth is affine*: $\rho(\boldsymbol{x}) = \langle \boldsymbol{a}, \tilde{\boldsymbol{x}} \rangle$ with $\boldsymbol{a} = (a, b, c)^\top$. This is a reasonable and interesting case because a second order regulariser favours (piecewise) affine functions. For this case we obtain the two tangent vectors

$$\boldsymbol{s}_x = \boldsymbol{K}^{-1} \begin{pmatrix} \langle \boldsymbol{a}, \tilde{\boldsymbol{x}} \rangle + ax \\ ay \\ a \end{pmatrix} \quad \text{and} \quad \boldsymbol{s}_y = \boldsymbol{K}^{-1} \begin{pmatrix} bx \\ \langle \boldsymbol{a}, \tilde{\boldsymbol{x}} \rangle + by \\ b \end{pmatrix}, \tag{16}$$

such that

$$\hat{\boldsymbol{n}} = \boldsymbol{K}^\top \begin{pmatrix} -a \\ -b \\ 2ax + 2by + c \end{pmatrix} \tag{17}$$

points along the surface normal (14), i.e. $\hat{\boldsymbol{n}} = \|\hat{\boldsymbol{n}}\| \cdot \boldsymbol{n}$. Equation 17 shows that the normal direction depends on the location $\boldsymbol{x} = (x, y)^\top$ when backprojecting an affine depth function. To get deeper insights on how the surface normals vary, let us consider the surface curvature by means of the shape operator. With (13), the second fundamental form for this example reads

$$\boldsymbol{II} = -\frac{2}{\|\hat{\boldsymbol{n}}\|} \begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix} . \tag{18}$$

Since this matrix is singular, we can directly conclude that the determinant of the shape operator (15) and consequently the Gaussian curvature K is zero. This further implies that at least one of the principal curvatures is zero. To check if both principal curvatures are zero, let us additionally consider the mean curvature

$$H = \frac{1}{2}\,\mathrm{tr}(\boldsymbol{S}) = -\frac{(ak_x)^2 + (bk_y)^2}{\det\left(\boldsymbol{K}^{-1}\right)\|\hat{\boldsymbol{n}}\|^3} . \tag{19}$$

This shows that the mean curvature is in general not equal to zero, i.e. the surface is bent in one direction. Only for constant functions, i.e. with a and b equal to zero, we also obtain a vanishing mean curvature and thus, a planar surface.

Inverse Depth. Let us now consider the alternative parameterisation $\phi(r) = {}^1\!/r$. Then the unknown ρ corresponds to the inverse depth. Again we assume that the unknown, in this case the *inverse depth*, is affine. Accordingly, we obtain the two tangent vectors

$$\boldsymbol{s}_x = \frac{\boldsymbol{K}^{-1}}{\langle \boldsymbol{a}, \tilde{\boldsymbol{x}} \rangle^2} \begin{pmatrix} by + c \\ -ay \\ -a \end{pmatrix} \quad \text{and} \quad \boldsymbol{s}_y = \frac{\boldsymbol{K}^{-1}}{\langle \boldsymbol{a}, \tilde{\boldsymbol{x}} \rangle^2} \begin{pmatrix} -bx \\ ax + c \\ -b \end{pmatrix} , \tag{20}$$

such that

$$\hat{\boldsymbol{n}} = \boldsymbol{K}^\top \boldsymbol{a} \tag{21}$$

points along the surface normal (14). Thus the surface normal is constant in all considered cases for the inverse depth parameterisation. In other words, back-projecting an affine inverse depth always results in a planar surface. In the same way as before, one can verify that both the Gaussian and the mean curvature of the surface are zero.

Summary. Table 1 summarises the discussed findings for all four parameterisations. In conclusion, this shows that the inverse depth parameterisation along the optical axis is preferable when using a second order regularisation.

3.4 Linearity Analysis of the Data Term

Previously, we analysed the influence of parameterisations w.r.t. the smoothness term. Now we analyse its effects on the data term. Since the unknown ρ appears as argument of $\boldsymbol{f}_i$, the presented energy (1) is non-convex. To cope with this, most minimisation strategies perform a linearisation. In this regard, we analyse how the different depth parameterisations affect this linearisation. In particular, we are interested in the deviation from linearity of $\boldsymbol{g}_i$ in ρ because this quantity depends on the chosen parameterisation. As introduced in Section 2, the function $\boldsymbol{g}_i$ maps a location $\boldsymbol{x} \in \Omega$ in the reference frame $\boldsymbol{f}_0$ with its depth $(\phi \circ \rho)(\boldsymbol{x})$ to the corresponding location in another image $\boldsymbol{f}_i$. This mapping can be described as a composition of a backprojection (10) and a projection (6):

$$\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho) = \boldsymbol{\pi}\big(\boldsymbol{P}_i \cdot \tilde{\boldsymbol{s}}_i(\boldsymbol{x}, \phi \circ \rho)\big). \tag{22}$$

Since scaled homogeneous coordinates are equivalent, it is possible to multiply $\tilde{\boldsymbol{s}}_i$ by $(\phi \circ \rho)(\boldsymbol{x})^{-1}$ and rewrite Equation 22 as

$$\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho) = \boldsymbol{\pi}\big(\boldsymbol{K}_i \boldsymbol{R}_i \boldsymbol{K}^{-1}\, \tilde{\boldsymbol{x}} + \boldsymbol{K}_i \boldsymbol{t}_i\, (\phi \circ \rho)(\boldsymbol{x})^{-1}\big)\,. \tag{23}$$

Direct Depth. For common setups, the camera offsets in z-direction are much smaller than the occurring depth values. This is because one typically walks around an object mainly with lateral motion while roughly keeping the distance with only small rotations between views. This causes converging camera setups that keep the object in the middle of the view. Hence, we assume in the following analysis that the z-component of $\boldsymbol{t}_i$ is zero. Please note the relation $\boldsymbol{t} = -\boldsymbol{R}\boldsymbol{c}_i$ between $\boldsymbol{t}_i$ and the camera centre $\boldsymbol{c}_i$ and that setting the z-component of $\boldsymbol{t}_i$ to zero does not restrict us to camera motions in the x-y-plane. This allows to simplify (23) to

$$r_3^{-1} \cdot \left(\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} (\phi \circ \rho)(\boldsymbol{x})^{-1} \right), \tag{24}$$

with the abbreviations $\boldsymbol{r} = \boldsymbol{K}_i \boldsymbol{R}_i \boldsymbol{K}^{-1} \tilde{\boldsymbol{x}}$ and $\boldsymbol{z} = \boldsymbol{K}_i \boldsymbol{t}_i$ that do not depend on $\rho(\boldsymbol{x})$. With the direct depth parameterisation $(\phi \circ \rho)(\boldsymbol{x})^{-1} = \rho(\boldsymbol{x})^{-1}$, we obtain a hyperbola and thus expect an additional linearisation error.

Inverse Depth. This is not the case for the inverse depth parameterisation with $(\phi \circ \rho)(\boldsymbol{x})^{-1} = \rho(\boldsymbol{x})$. In fact, Equation 24 reveals that $\boldsymbol{g}_i$ is linear in $\rho(\boldsymbol{x})$ in this case. Thus, no error is introduced when linearising $\boldsymbol{g}_i$ w.r.t. the inverse depth. To summarise, also the linearisation analysis shows that an inverse depth parameterisation turns out to be more appropriate for multiview depth estimation than the standard direct depth parameterisation.

4 Minimisation

To solve the energy (1), we perform a first order Taylor linearisation around ρ_0 in the data term (2):

$$\boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho)) \approx \boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho_0)) + (\rho - \rho_0) \cdot \partial_\rho \boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho)|_{\rho = \rho_0}. \tag{25}$$

Applying the chain rule gives

$$\partial_\rho \boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho)) = \boldsymbol{\mathcal{J}} \boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho)) \cdot \boldsymbol{\mathcal{J}} \boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho), \tag{26}$$

Table 1. Preservation of planarity.

	(i) direct depth		(ii) inverse depth	
	(a) line of sight	(b) optical axis	(a) line of sight	(b) optical axis
constant	no	yes	no	yes
affine	no	no	no	yes

where the image derivatives $\mathcal{J}\boldsymbol{f}_i$ are independent of the parameterisation. The second term in (26) is given by

$$\mathcal{J}\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho) = \mathcal{J}\boldsymbol{\pi}\left(\boldsymbol{P}_i \begin{pmatrix} \boldsymbol{K}^{-1}\tilde{\boldsymbol{x}} \\ (\phi \circ \rho)(\boldsymbol{x})^{-1} \end{pmatrix}\right) \boldsymbol{K}_i \boldsymbol{t}_i \; \partial_\rho (\phi \circ \rho)(\boldsymbol{x})^{-1}, \tag{27}$$

where for the direct depth parameterisation $\partial_\rho(\phi \circ \rho)(\boldsymbol{x})^{-1} = -\rho(\boldsymbol{x})^{-2}$, and for the inverse depth parameterisation $\partial_\rho(\phi \circ \rho)(\boldsymbol{x})^{-1} = 1$. With the abbreviations

$$\boldsymbol{m}_i = \partial_\rho \boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho))\big|_{\rho=\rho_0} \quad \text{and} \quad \boldsymbol{b}_i = \boldsymbol{m}_i\, \rho_0 + \boldsymbol{f}_0(\boldsymbol{x}) - \boldsymbol{f}_i(\boldsymbol{g}_i(\boldsymbol{x}, \phi \circ \rho_0)) \tag{28}$$

the energy (1) with the linearised data term reads

$$E(\rho, \boldsymbol{w}) = \int_\Omega \frac{1}{n} \sum_{i=1}^{n} \Psi\Big(\|\boldsymbol{m}_i\, \rho - \boldsymbol{b}_i\|^2\Big) \; + \alpha \Big(\Psi\big(\|\boldsymbol{\nabla}\rho - \boldsymbol{w}\|^2\big) + \beta\, \Psi\big(\|\mathcal{J}\boldsymbol{w}\|_F^2\big)\Big) \mathrm{d}\boldsymbol{x}.$$

Euler-Lagrange Equations. The minimiser of the linearised energy functional fulfils the corresponding Euler-Lagrange equations w.r.t. ρ and $\boldsymbol{w}$. With

$$\Psi'_{Di} = \Psi'\big(\|\boldsymbol{m}_i \rho - \boldsymbol{b}_i\|^2\big), \quad \Psi'_C = \Psi'\big(\|\boldsymbol{\nabla}\rho - \boldsymbol{w}\|^2\big), \quad \Psi'_S = \Psi'\big(\|\mathcal{J}\boldsymbol{w}\|_F^2\big) \tag{29}$$

they are given by

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \Psi'_{Di} \cdot \langle \boldsymbol{m}_i, \boldsymbol{m}_i\, \rho + \boldsymbol{b}_i \rangle - \alpha\ \operatorname{div}(\Psi'_C \cdot (\boldsymbol{\nabla}\rho - \boldsymbol{w})) &= 0, \\ \Psi'_C \cdot (w_1 - p_x) - \beta\ \operatorname{div}(\Psi'_S \cdot \boldsymbol{\nabla} w_1) &= 0, \\ \Psi'_C \cdot (w_2 - p_y) - \beta\ \operatorname{div}(\Psi'_S \cdot \boldsymbol{\nabla} w_2) &= 0 \end{aligned} \tag{30}$$

with boundary conditions $(\boldsymbol{\nabla}\rho - \boldsymbol{w})^\top \boldsymbol{n} = 0$ and $\mathcal{J}\boldsymbol{w}\, \boldsymbol{n} = \boldsymbol{0}$, where $\boldsymbol{n}$ is the 2D outer normal here.

Implementation. We discretise (30) with finite differences on a regular grid. This results in a nonlinear system of equations, which we solve with two nested loops. While we update the nonlinear terms Ψ_D, Ψ_C, and Ψ_S (29) in the outer loop, we solve the linear system in the inner loop with the Fast-Jacobi algorithm [21]. Furthermore, we employ a coarse-to-fine approach to overcome linearisation errors.

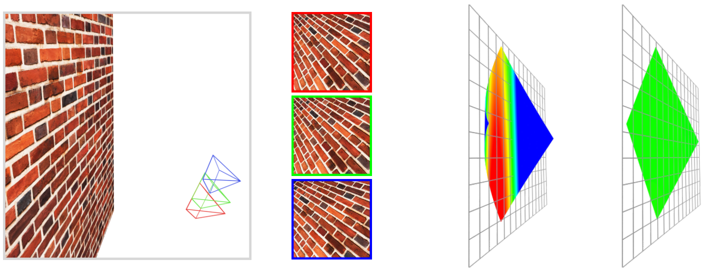

Fig. 3. *From left to right*: **(a)** Camera setup and geometry. **(b)** Corresponding input images. **(c)** Reconstruction with direct depth parameterisation. **(d)** Reconstruction with inverse depth parameterisation. See text for details.

5 Experiments

Our evaluation consists of two main parts. First, we underpin our theoretical findings from Section 3 by means of experiments with synthetic data. Figure 3(a) shows a 3D scene with a planar surface and three cameras, and Figure 3(b) depicts the images captured with the corresponding cameras. Figure 3 (c) and (d) show the computed reconstructions with a direct depth and an inverse depth parameterisation, respectively. We clearly see that performing second order regularisation on the *depth* introduces a bias towards curved surfaces as discussed in Section 3. In contrast, performing second order regularisation on the *inverse depth* does not introduce such as bias and thus yields a significantly better reconstruction. In Figure 3 (c) and (d), we apply the following colour code to visualise the reconstruction errors: Green represents an error of zero, whereas red and blue correspond to behind and in front of the ground truth surface, respectively.

In the second part of our evaluation, we run tests on six publically available multiview data sets from the Middlebury benchmark [22] to obtain a quantitative comparison between both parameterisations. More specifically, we use five images for each 3D scene to compute the depth map. We have optimised the smoothness parameters α and β for each parameterisation, but kept them fixed over the individual scenes. Here, we measure for both parameterisation the reconstruction quality in terms of the root mean square error in depth. Table 2 shows that the inverse depth parameterisation provides a significantly better reconstruction quality than the direct depth parameterisation in all cases. These quantitative experiments confirm our findings from Section 3. The inverse depth parameterisation does not only have advantages in theory, but also practically achieves superior reconstructions. Besides the discussed benefits for the linearisation and second order regularisation, there is another advantage that we have not stressed so far: It is a natural choice to initialise the inverse depth with zero. This corresponds to a depth of infinity. Thus, an initialisation of the direct depth with a large constant seems desirable but turns out to be problematic.

Table 2. Root mean square errors for six data sets from the Middlebury benchmark [22].

	direct depth	inverse depth
Barn 1	0.67	**0.25**
Barn 2	1.48	**0.51**
Bull	0.50	**0.23**
Poster	0.48	**0.22**
Sawtooth	1.09	**0.43**
Venus	0.58	**0.29**

6 Conclusions

In this work, we have analysed different depth parameterisations within the context of multiview depth estimation with higher order regularisation. Our first finding is that parameterisations along the line of sight are not suitable for such a scenario. In fact, we show that parameterisations along the optical axis are much more reasonable. For them, we present a detailed analysis of a direct depth and an inverse depth parameterisation. We point out several advantages of the inverse depth parameterisation: First, it is compatible with second order regularisation. Piecewise affine inverse depth leads to piecewise planar 3D surfaces. On the contrary, this is not the case for the direct depth parameterisation. It introduces a bias which we quantify both theoretically by means of the shape operator as well as by experiments. Second, we show that an inverse depth parameterisation is not only advantageous for the smoothness term. It is also preferable for the linearisation required in the data term compared to a direct depth parameterisation. Last but not least, the inverse depth approach additionally admits a more meaningful initialisation.

Based on our findings, we recommend the inverse depth parameterisation along the optical axis as the parameterisation of choice for variational multiview depth estimation. We believe that this insight can improve the performance of existing methods. In future work, we plan to extend our model with more sophisticated assumptions on the data term and regulariser.

Acknowledgements. Funding by the Cluster of Excellence *Multimodal Computing and Interaction* is gratefully acknowledged.

References

1. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proc. SIGGRAPH 96. Volume 3., New Orleans, LA (August 1996) 303–312
2. Zach, C., Pock, T., Bischof, H.: A globally optimal algorithm for robust TV-L^1 range image integration. In: Proc. Ninth International Conference on Computer Vision, Rio de Janeiro, Brazil (October 2007) 1–8

3. Zach, C.: Fast and high quality fusion of depth maps. In: Proc. Fourth International Symposium on 3D Data Processing, Visualization and Transmission, Atlanta, GA (June 2008) 1–8
4. March, R.: Computation of stereo disparity using regularization. Pattern Recognition Letters **8**(3) (1988) 181 – 187
5. Mansouri, A.R., Mitiche, A., Konrad, J.: Selective image diffusion: Application to disparity estimation. In: Proc. IEEE International Conference on Image Processing, Chicago, IL (October 1998) 284–288
6. Scharstein, D., Szeliski, R.: Stereo matching with non-linear diffusion. International Journal of Computer Vision **28** (June 1998) 155–174
7. Slesareva, N., Bruhn, A., Weickert, J.: Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps. In Kropatsch, W.G., Sablatnig, R., Hanbury, A., eds.: Pattern Recongnition. Volume 3663 of Lecture Notes in Computer Science., Berlin, Springer (2005) 33–40
8. Zimmer, H., Bruhn, A., Valgaerts, L., Breuß, M., Weickert, J., Rosenhahn, B., Seidel, H.P.: PDE-based anisotropic disparity-driven stereo vision. In: Proc. Vision, Modeling, and Visualization, Konstanz, Germany, Akademische Verlagsgesellschaft Aka (October 2008) 263–272
9. Valgaerts, L., Bruhn, A., Weickert, J.: A variational model for the joint recovery of the fundamental matrix and the optical flow. In Rigoll, G., ed.: Pattern Recognition. Volume 5096 of Lecture Notes in Computer Science., Springer, Berlin (2008) 314–324
10. Ranftl, R., Gehrig, S., Pock, T., Bischof, H.: Pushing the limits of stereo using variational stereo estimation. In: Proc. IEEE Intelligent Vehicles Symposium, Alcala de Henares, Spain (2012) 401–407
11. Hewer, A., Weickert, J., Seibert, H., Scheffer, T., Diebels, S.: Lagrangian strain tensor computation with higher order variational models. In: Proc. British Machine Vision Conference, Bristol, UK, BMVA Press (September 2013)
12. Ranftl, R., Bredies, K., Pock, T.: Non-local total generalized variation for optical flow estimation. In Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., eds.: Computer Vision – ECCV 2014. Volume 8689 of Lecture Notes in Computer Science. Springer, Berlin (2014) 439–454
13. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences **3**(3) (September 2010) 492–526
14. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numerische Mathematik **76**(2) (April 1997) 167–188
15. Robert, L., Deriche, R.: Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities. In Buxton, B., Cipolla, R., eds.: Computer Vision – ECCV 96. Volume 1064 of Lecture Notes in Computer Science. Springer, Berlin (1996) 439–451
16. Stühmer, J., Gumhold, S., Cremers, D.: Real-time dense geometry from a handheld camera. In Goesele, M., Roth, S., Kuijper, A., Schiele, B., Schindler, K., eds.: Pattern Recognition. Volume 6376 of Lecture Notes in Computer Science. Springer, Berlin (September 2010) 11–20
17. Semerjian, B.: A new variational framework for multiview surface reconstruction. In Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., eds.: Computer Vision – ECCV 2014. Volume 8694 of Lecture Notes in Computer Science. Springer, Berlin (September 2014) 719–734
18. Civera, J., Davison, A.J., Montiel, J.: Inverse depth parametrization for monocular SLAM. IEEE Transactions on Robotics **24**(5) (October 2008) 932–945

19. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: Dense tracking and mapping in real-time. In: Proc. IEEE International Conference on Computer Vision, Barcelona, Spain (November 2011) 2320–2327
20. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: Large-scale direct monocular slam. In Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., eds.: Computer Vision – ECCV 2014. Volume 8690 of Lecture Notes in Computer Science. Springer, Berlin (September 2014) 834–849
21. Grewenig, S., Weickert, J., Schroers, C., Bruhn, A.: Cyclic schemes for PDE-based image analysis. Technical Report 327, Saarland University, Saarbrücken, Germany (March 2013)
22. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision **47** (April 2002) 7–42