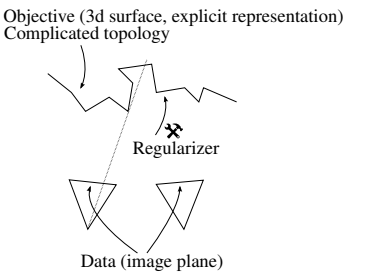
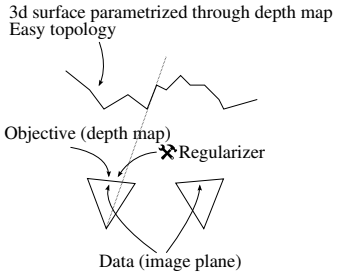
Efficient Minimal-Surface Regularization of Perspective Depth Maps in Variational Stereo

Gottfried Graber[1], Jonathan Balzer[2], Stefano Soatto[2], Thomas Pock[1,3]
[1]Graz University of Technology [2]UCLA [3]AIT Austrian Institute of Technology GmbH



(a) Explicit representation of 3d surfaces independent of images.



(b) Surface parametrized by a depthmap, strongly linked to reference camera frame

Figure 1: In shape-based 3d reconstruction (a), the surface is given explicitly (e.g., by a triangle mesh), and the regularizer acts on this explicit representation. In depth-map-based stereo (b), the surface is parametrized by a range map. Typically, the regularizer acts on the *parametrization*, i.e., the depth map.

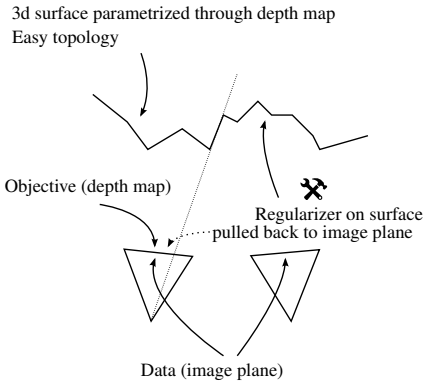


Figure 2: In our approach, we combine the advantages of shape-based methods with those of range maps. Our regularizer, while defined on the image plane, respects the inner geometry of the surface.

Inferring the shape of a 3d surface from multiple images (3d-reconstruction) is an intrinsically ill-posed problem, which is typically solved by means of optimization. We can distinguish fundamentally different approaches from the way the surface is represented: Shape-based methods (Fig. 1(a)) employ an explicit representation of the 3d surface, e.g., through a triangle mesh. This representation is independent of the images, which on the one hand allows for flexibility, on the other hand requires computation of topology and visibility during optimization. Depth-map-based stereo (Fig. 1(b)) *parametrizes* the 3d surface by assigning a distance value to every pixel of the reference image, which bypasses the difficulties of computing topology/visibility. Unfortunately, the image plane is not the natural place to enforce regularization of the surface. This work proposes a novel regularizer for depth-map-based stereo, which is defined on the image plane, but respects the intrinsic geometry of the surface (see Fig. 2).

Starting from a parametrization of a 3d surface $\mathbf{X}(x,y)$ by a perspective depth map $z(x,y)$

$$\mathbf{X}(\mathbf{x}) = zK^{-1}\mathbf{x}, \tag{1}$$

where $\mathbf{x} = (x,y)$ are pixel coordinates, K is the intrinsic calibration matrix of the camera, the *first fundamental form* is given by

$$\mathbf{I} = \begin{pmatrix} \langle \mathbf{X}_x, \mathbf{X}_x \rangle & \langle \mathbf{X}_x, \mathbf{X}_y \rangle \\ \langle \mathbf{X}_x, \mathbf{X}_y \rangle & \langle \mathbf{X}_y, \mathbf{X}_y \rangle \end{pmatrix}, \tag{2}$$

where subscripts denote partial derivatives. The first fundamental form is used to calculate lengths and angles on the surface. In particular, the infinitesimal area element is given by

$$\mathrm{d}A = \sqrt{\det \mathbf{I}} = \frac{z}{f_1 f_2} \sqrt{\|\nabla_{\boldsymbol{f}} z\|^2 + (\langle \nabla_{\boldsymbol{f}} z, \boldsymbol{x} \rangle + z)^2}, \tag{3}$$

where $\nabla_{\boldsymbol{f}}$ denotes the gradient weighted by the focal lengths, i.e. $\nabla_{\boldsymbol{f}} z = (f_1 z_x, f_2 z_y)^T$. The total area can be obtained by integrating (3) over the image domain.

In our variational stereo algorithm, we use (3) as regularization term. Compared to the widely used Total Variation (TV) regularizer, our regularization term does not suffer from staircasing artifacts. The latter stem from the fact that the space of functions with minimal TV is spanned by a piecewise constant basis. Total Generalized Variation (TGV) enriches this basis with polynomials of higher order, however, due to computational difficulties, only second order TGV is relevant in practice. Our regularizer on the other hand is not restricted to a certain class of functions at all since it is defined on the surface itself.

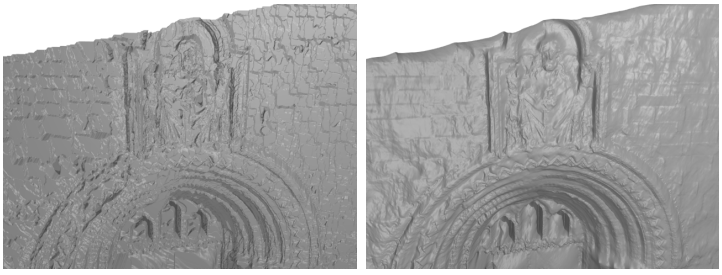
The area element (3) is non-convex, which makes optimization challenging. By exploiting gauge freedom (i.e. the fact that there are infinitely many equivalent ways to parametrize a surface through a depth map), we propose to re-parametrize the depth map by $z = \phi(\zeta) = \sqrt{2\zeta}$. Remarkably, this makes the area element a linear function of ζ, and thus compatible with highly efficient primal-dual algorithms for large scale problems.

Using a GPU-based parallel implementation, the runtime of our algorithm is approximately 150 ms for a pair of 640×480 images, which makes the method attractive for (near) real-time applications.

Fig. 3 and 4 show 3d renderings of depth maps from the Strecha dataset for dense multiview stereo. In the close-ups, the staricaising of TV (resulting in fronto-parallel surfaces) is clearly visible, whereas our minimal area regularizer has no difficulties in representing slanted or curved surfaces.

(a) TV closeup (b) Ours closeup

Figure 3: Results for the *Fountain-P11* scene



(a) TV closeup (b) Ours closeup

Figure 4: Results for the *Herz-Jesu-P25* scene