

# Supplementary Material: Combining Shape from Shading and Stereo: A Variational Approach for the Joint Estimation of Depth, Illumination and Albedo

Daniel Maurer  
maurer@vis.uni-stuttgart.de

Yong Chul Ju  
ju@vis.uni-stuttgart.de

Michael Breuß  
breuss@tu-cottbus.de

Andrés Bruhn  
bruhn@vis.uni-stuttgart.de

Institute for Visualization and Interactive Systems  
University of Stuttgart, Germany

Institute for Visualization and Interactive Systems  
University of Stuttgart, Germany

Institute for Applied Mathematics and Scientific  
Computing, BTU Cottbus-Senftenberg, Germany

Institute for Visualization and Interactive Systems  
University of Stuttgart, Germany

In the following we provide additional details regarding the differential formulation, the parameter settings for the different methods as well as additional figures that illustrate the results of our experiments in more detail.

## 1 Differential Formulation

**Stereo Data Term.** The differential formulation of the stereo data term is obtained by performing a linearisation w.r.t. the depth increment  $dz^k$ . Let us introduce  $\varphi_i^{k,c} := I_0^c(\mathbf{x}_0) - I_i^c(\mathbf{x}_i^k)$  for the brightness constancy assumption for each colour channel  $c \in \{1, 2, 3\}$ , where  $\mathbf{x}_i^k$  is the position of the projected surface point with known depth  $z^k$  corresponding to the position  $\mathbf{x}_0$  in the reference frame. If we furthermore summarise the assumptions on the three channels in a single vector

$$\varphi_i^k := \begin{pmatrix} \varphi_i^{k,1} \\ \varphi_i^{k,2} \\ \varphi_i^{k,3} \end{pmatrix} \quad (1)$$

we obtain the following linearised constraints for the brightness and gradient constancy assumption, respectively:

$$\bar{\varphi}_{i,0}^k := \varphi_i^k(\mathbf{x}_0) + \frac{\partial \varphi_i^k(\mathbf{x}_0)}{\partial z^k(\mathbf{x}_0)} \cdot dz^k(\mathbf{x}_0), \quad (2)$$

$$\bar{\varphi}_{i,x}^k := \frac{\partial}{\partial x_0} \varphi_i^k(\mathbf{x}_0) + \frac{\partial^2 \varphi_i^k(\mathbf{x}_0)}{\partial z^k(\mathbf{x}_0) \partial x_0} \cdot dz^k(\mathbf{x}_0), \quad (3)$$

$$\bar{\varphi}_{i,y}^k := \frac{\partial}{\partial y_0} \varphi_i^k(\mathbf{x}_0) + \frac{\partial^2 \varphi_i^k(\mathbf{x}_0)}{\partial z^k(\mathbf{x}_0) \partial y_0} \cdot dz^k(\mathbf{x}_0). \quad (4)$$

Further we follow [4, 6] and normalise the linearised constraints. To this end, we introduce the following normalisation factors

$$\theta_0^{k,c} := \left( \left( \frac{\partial \phi_i^{k,c}}{\partial z^k(\mathbf{x}_0)} \right)^2 + \zeta^2 \right)^{-1}, \quad (5)$$

$$\theta_x^{k,c} := \left( \left( \frac{\partial^2 \phi_i^{k,c}}{\partial z^k(\mathbf{x}_0) \partial y_0} \right)^2 + \zeta^2 \right)^{-1}, \quad \theta_y^{k,c} := \left( \left( \frac{\partial^2 \phi_i^{k,c}}{\partial z^k(\mathbf{x}_0) \partial x_0} \right)^2 + \zeta^2 \right)^{-1}, \quad (6)$$

where  $\zeta$  is a small parameter to prevent division by zero. Combining linearisation and normalisation, we finally obtain the following differential formulation for the stereo term

$$D_{\text{Stereo}}(dz^k) = \int_{\Omega_0} \frac{1}{n} \sum_{i=1}^n \Psi_L \left( \left( \bar{\varphi}_{i,0}^k \right)^\top \text{diag} \left( \theta_0^{k,1}, \theta_0^{k,2}, \theta_0^{k,3} \right) \bar{\varphi}_{i,0}^k \right) \quad (7)$$

$$+ \Psi_L \left( \left( \bar{\varphi}_{i,x}^k \right)^\top \text{diag} \left( \theta_x^{k,1}, \theta_x^{k,2}, \theta_x^{k,3} \right) \bar{\varphi}_{i,x}^k \right) \quad (8)$$

$$+ \Psi_L \left( \left( \bar{\varphi}_{i,y}^k \right)^\top \text{diag} \left( \theta_y^{k,1}, \theta_y^{k,2}, \theta_y^{k,3} \right) \bar{\varphi}_{i,y}^k \right) d\mathbf{x}_0. \quad (9)$$

**SfS Data Term.** Regarding the more difficult linearisation of the SfS term we follow the recent work of Maurer *et al.* [3]. Due to the hyperbolic nature of this term, we first replace the partial depth derivatives in  $\mathbf{R}$  with an appropriate upwind scheme approximation and obtain  $\bar{\mathbf{R}}$ . In a second step we linearise  $\bar{\mathbf{R}}^k$  w.r.t. the relevant increments. By introducing  $\phi^k := \mathbf{I}_0(\mathbf{x}_0) - \bar{\mathbf{R}}^k$ , where  $\bar{\mathbf{R}}^k$  is computed with the known values of level  $k$ , the differential SfS data term can then be written as

$$D_{\text{SfS}}(dz^k, d\mathbf{l}^k, d\rho^k) = \int_{\Omega_0} \left( \left\| \phi^k + \sum_{\mathbf{h} \in H} \frac{\partial \phi^k(\mathbf{x}_0)}{\partial z^k(\mathbf{x}_0 + \mathbf{h})} dz(\mathbf{x}_0 + \mathbf{h})^k + \frac{\partial \phi^k(\mathbf{x}_0)}{\partial \mathbf{l}^k(\mathbf{x}_0)} d\mathbf{l}^k(\mathbf{x}_0)^k + \frac{\partial \phi^k(\mathbf{x}_0)}{\partial \rho^k(\mathbf{x}_0)} d\rho^k(\mathbf{x}_0)^k \right\|_2^2 \right) d\mathbf{x}_0, \quad (10)$$

where  $H = \{-\mathbf{h}_y, -\mathbf{h}_x, \mathbf{0}, +\mathbf{h}_x, +\mathbf{h}_y\}$  is the neighbourhood involved in the upwind approximation of  $\mathbf{R}$  and  $\mathbf{h}_x = (h_x, 0)^\top$ ,  $\mathbf{h}_y = (0, h_y)^\top$  are the grid spacings  $h_x$  and  $h_y$  in  $x$ - and  $y$ -direction, respectively. For additional details on the linearisation we refer the reader to [3].

**Regularisation Terms.** In contrast to the two data terms, the differential formulation of the three smoothness terms is rather straightforward. It is given by

$$R_{\text{illum}}(d\mathbf{l}^k) = \int_{\Omega_0} \Psi_I \left( \left\| \mathcal{J}(\mathbf{l}^k + d\mathbf{l}^k) \right\|_F^2 \right) d\mathbf{x}_0, \quad (11)$$

$$R_{\text{albedo}}(d\rho^k) = \int_{\Omega_0} g \left( \left\| \mathcal{J}(\mathbf{ch}(\mathbf{I}_0)) \right\|_F^2 \right) \cdot \left\| \mathcal{J}(\rho^k + d\rho^k) \right\|_F^2 d\mathbf{x}_0, \quad (12)$$

$$R_{\text{depth}}(dz^k) = \inf_{d\mathbf{u}^k} \int_{\Omega_0} \left( C(dz^k, d\mathbf{u}^k) + \alpha_{\mathbf{u}} \cdot S(d\mathbf{u}^k) \right) d\mathbf{x}_0, \quad (13)$$

with

$$C(dz^k, \mathbf{d}\mathbf{u}^k) = \sum_{d=1}^2 \Psi_C^d \left( \left( \mathbf{r}_d^\top \left( \nabla(z^k + dz^k) - (\mathbf{u}^k + \mathbf{d}\mathbf{u}^k) \right) \right)^2 \right), \quad (14)$$

$$S(\mathbf{d}\mathbf{u}^k) = \sum_{d=1}^2 \Psi_S^d \left( \sum_{m=1}^2 \left( \mathbf{r}_m^\top \mathcal{J}(\mathbf{u}^k + \mathbf{d}\mathbf{u}^k) \mathbf{r}_d \right)^2 \right). \quad (15)$$

**Increment Regularisation.** Since the differential formulation of the energy is based on an incremental linearisation of the data terms, one has to ensure robustness w.r.t. large erroneous increments (e.g. if the linearisation is locally not valid). To this end, we penalise the length of the increments with respective weighting factors  $\alpha_{dz}$ ,  $\alpha_{\mathbf{d}\mathbf{l}}$ , and  $\alpha_{\mathbf{d}\mathbf{p}}$ , which yields

$$R_{\text{inc}} = \int_{\Omega_0} \alpha_{dz} \cdot \|dz^k\|_2^2 + \alpha_{\mathbf{d}\mathbf{l}} \cdot \|\mathbf{d}\mathbf{l}^k\|_2^2 + \alpha_{\mathbf{d}\mathbf{p}} \cdot \|\mathbf{d}\mathbf{p}^k\|_2^2 \, d\mathbf{x}_0. \quad (16)$$

Please note that, as the incremental coarse-to-fine fixed point iteration converges, the influence of the additional length regularisation vanishes, since the term only penalises the increments and not the actual values. This particularly holds if multiple fixed point steps are performed per resolution level, since increments in later steps tend to zero.

## 1.1 Parameter Scaling

To account for the fact that the zero initialisation of the illumination vector does not allow the SfS data term to provide any useful information at coarser levels, a sigmoid weighting function is introduced that increases the SfS weight  $v$  at finer levels. The corresponding weight is given by

$$v := s_v \cdot v, \quad \text{with} \quad s_v := \frac{1}{1 + e^{\frac{-(k/k_{\text{max}}) + b}{a}}}, \quad (17)$$

where the slope and the shift can be adjusted with the parameters  $a$  and  $b$ , respectively. Throughout our experiments we set  $a = 0.1$  and  $b = 0.5$ . Further we apply the same scaling to the albedo and illumination regularisation weights, to ensure that the relative weighting is not affected. Finally, we employ a resolution level depended scaling of  $\alpha_z$ ,  $\alpha_{\mathbf{l}}$ , and  $\alpha_{\mathbf{p}}$  using

$$\alpha_z := s_\alpha \cdot \alpha_z, \quad \alpha_{\mathbf{l}} := s_\alpha \cdot \alpha_{\mathbf{l}}, \quad \alpha_{\mathbf{p}} := s_\alpha \cdot \alpha_{\mathbf{p}}, \quad \text{with} \quad s_\alpha := \left( \frac{1}{h_x^k \cdot h_y^k} \right)^{1/2}, \quad (18)$$

where  $h_x^k$  and  $h_y^k$  denote the grid size of the current level in  $x$ - and  $y$ -direction, respectively.

## 2 Parameters

In all our experiments we used a fixed set of solver related parameters: 20 iterations per resolution level, 2 non-linear fixed-point iterations, 20 SOR iterations, a downsampling factor of  $\eta = 0.8$ , an over-relaxation parameter of  $\omega = 1.8$  and  $h_z = h_{\mathbf{l}} = h_{\mathbf{p}} = 10^{-12}$  for the numerically computed derivatives (cf. [3]). In addition, we chose the regularisation parameter for the data term normalisation  $\zeta = 0.01$  and the gradient constancy weighting factor  $\gamma = 1.0$ . In case of the Charbonnier penaliser we set  $\lambda = 0.01$  and for the robust function  $\Psi_L$  we used

Combined approach	$v$	$\alpha_z$	$\alpha_u$	$\alpha_l$	$\alpha_p$	$\alpha_{dz}$	$\alpha_{dl}$	$\alpha_{dp}$	$\lambda^{(1)}$
Blunderbuss Pete	0.05	100	0.1	50	100	0	0.5	0.5	0.005
Angel	3.0	200	0.1	30	600	0	3.0	3.0	0.005
Fountain-P11	1.0	350	0.1	15	300	0	0.5	0.5	0.001
Herz-Jesu-P8	1.0	400	0.1	15	300	0	0.5	0.5	0.001
Pure stereo approach	$v$	$\alpha_z$	$\alpha_u$	$\alpha_l$	$\alpha_p$	$\alpha_{dz}$	$\alpha_{dl}$	$\alpha_{dp}$	$\lambda^{(1)}$
Blunderbuss Pete	-	100	0.1	-	-	0	-	-	0.005
Angel	-	400	0.1	-	-	0	-	-	0.005
Fountain-P11	-	300	0.1	-	-	0	-	-	0.001
Herz-Jesu-P8	-	400	0.1	-	-	0	-	-	0.001
Galliani <i>et al.</i> [1]	$w^{(2)}$	$n_d^{(3)}$	$\gamma$	$\alpha$	$\tau_{col}$	$\tau_{grad}$	$z_{min}$	$z_{max}$	
Fountain-P11	35×35	100	10	0.9	10	2	3.5	12	-
Herz-Jesu-P8	31×31	100	10	0.9	10	2	5	19	-
Graber <i>et al.</i> [2]	$n_p^{(4)}$	$n_w^{(5)}$	$n_s^{(6)}$	$\eta$	$\lambda$	$\varepsilon$	$z_{min}$	$z_{max}$	
Fountain-P11	30	30	60	0.9	0.0114	$10^{-3}$	3.5	12	-
Herz-Jesu-P8	30	30	60	0.9	0.045	$10^{-3}$	5	19	-

<sup>(1)</sup>  $\lambda$  for the Perona-Malik penaliser     
<sup>(2)</sup> window size     
<sup>(3)</sup> iterations for the diffusion-like scheme  
<sup>(4)</sup> number of pyramid levels     
<sup>(5)</sup> warps per pyramid level     
<sup>(6)</sup> iterations per warp

Table 1: Parameter settings for the different datasets and approaches.

$\varepsilon = 0.001$ . The remaining parameters were chosen individually for each dataset as listed in Table 1. The parameter settings for the method of Galliani *et al.* [1] and Graber *et al.* [2] are also given in Table 1. Regarding the method of Graber *et al.* we even optimised the level depended parameter scaling function in the python code from

$$\lambda_l = \lambda \cdot \left(\frac{1}{\eta}\right)^l \quad \text{to} \quad \lambda_l = \lambda \cdot \left(\frac{1}{\eta}\right)^{1.5 \cdot l}, \quad (19)$$

which led to a noticeable improvement of the results. In this context  $l$  denotes the current pyramid level (the finest resolution is given on level  $l = 0$ ).

**Parameter selection.** In order to obtain good reconstruction results in general parameters have to be adjusted properly. While the solver related parameters mainly effect the trade off between reconstruction quality and runtime performance and thus can be set fixed in advance (see above), the model parameters have direct impact on the quality and thus must be selected more carefully. In this context, a convenient strategy is to first adjust the parameters of the pure stereo model and then extend the parameter set to the parameters of the combined model. From our experience the optimal model parameters may vary depending on the intrinsic camera parameters as well as the distance of the scene to the camera. However, even some of the model parameters turned out to be suitable for a wider range of images and thus can also be set fixed in advance (see above). What remains to be adjusted are essentially the hyper parameters for the different terms of the differential energy (see Table 1).



### 3 Experiments

Apart from providing additional results for the three experiments in the main paper, we also performed a fourth experiment that analyses the influence of the normalisation of the stereo data term. To this end, we focussed on the pure stereo method and computed results for the *Blunderbuss Pete* dataset with and without normalisation. The corresponding results are depicted in Figure 1. They reveal that without normalisation artefacts arise at image edges – even if one increases the amount of regularisation. This finding is in accordance with [6].

Let us now turn towards the additional results. Figure 2, Figure 3 and Figure 4 show the input images as well as the computed albedo and illumination direction. Moreover, Figure 5 and Figure 6 depict larger versions of the results of the corresponding experiments in the main paper. Finally, Figure 7 and 8 show the reconstructed scenes from a different perspective.

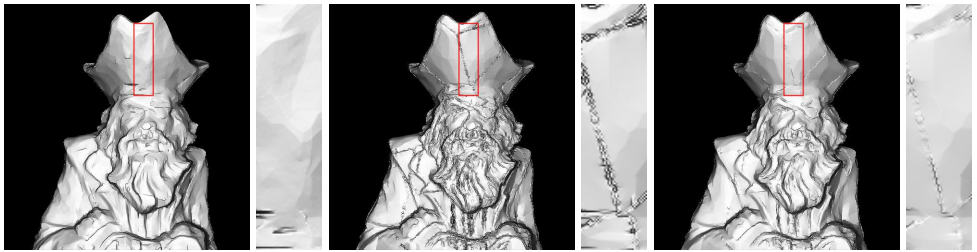


Figure 1: Influence of the normalisation. **From left to right:** Pure stereo with normalisation. Pure stereo without normalisation ( $\alpha_z = 5 \cdot 10^3$ ,  $\alpha_z = 10^4$ ).

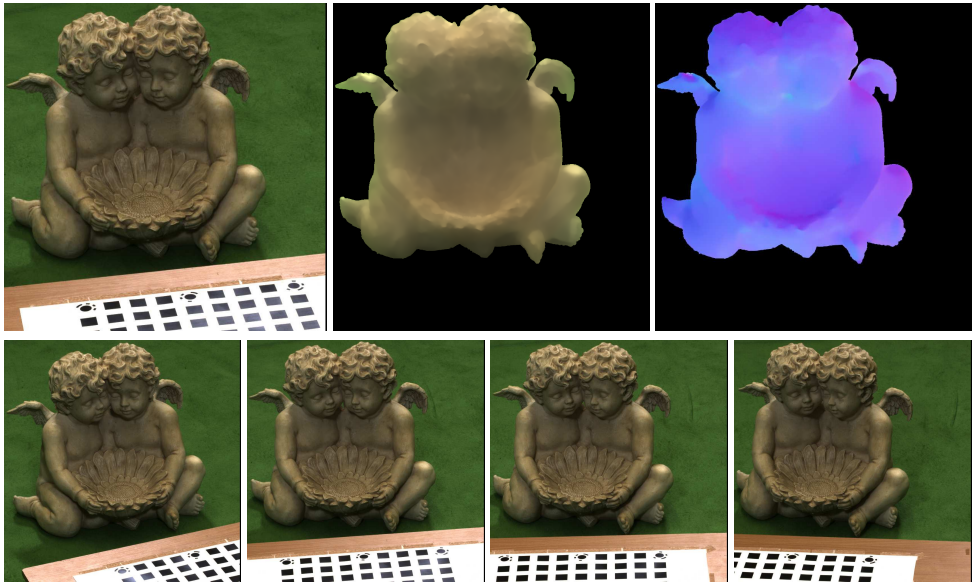


Figure 2: Real-world *Angel* dataset. **Top row:** Reference view (0), computed albedo, computed illumination direction. **Bottom row:** Match views (1,27,26,25).



Figure 3: Real-world *Fountain-P11* dataset [5]. **Top row:** Reference view (6), match view (7). **Bottom row:** Computed albedo, Computed illumination direction.



Figure 4: Real-world *Herz-Jesu-P8* dataset [5]. **Top row:** Reference view (5), match view (6). **Bottom row:** Computed albedo, Computed illumination direction.



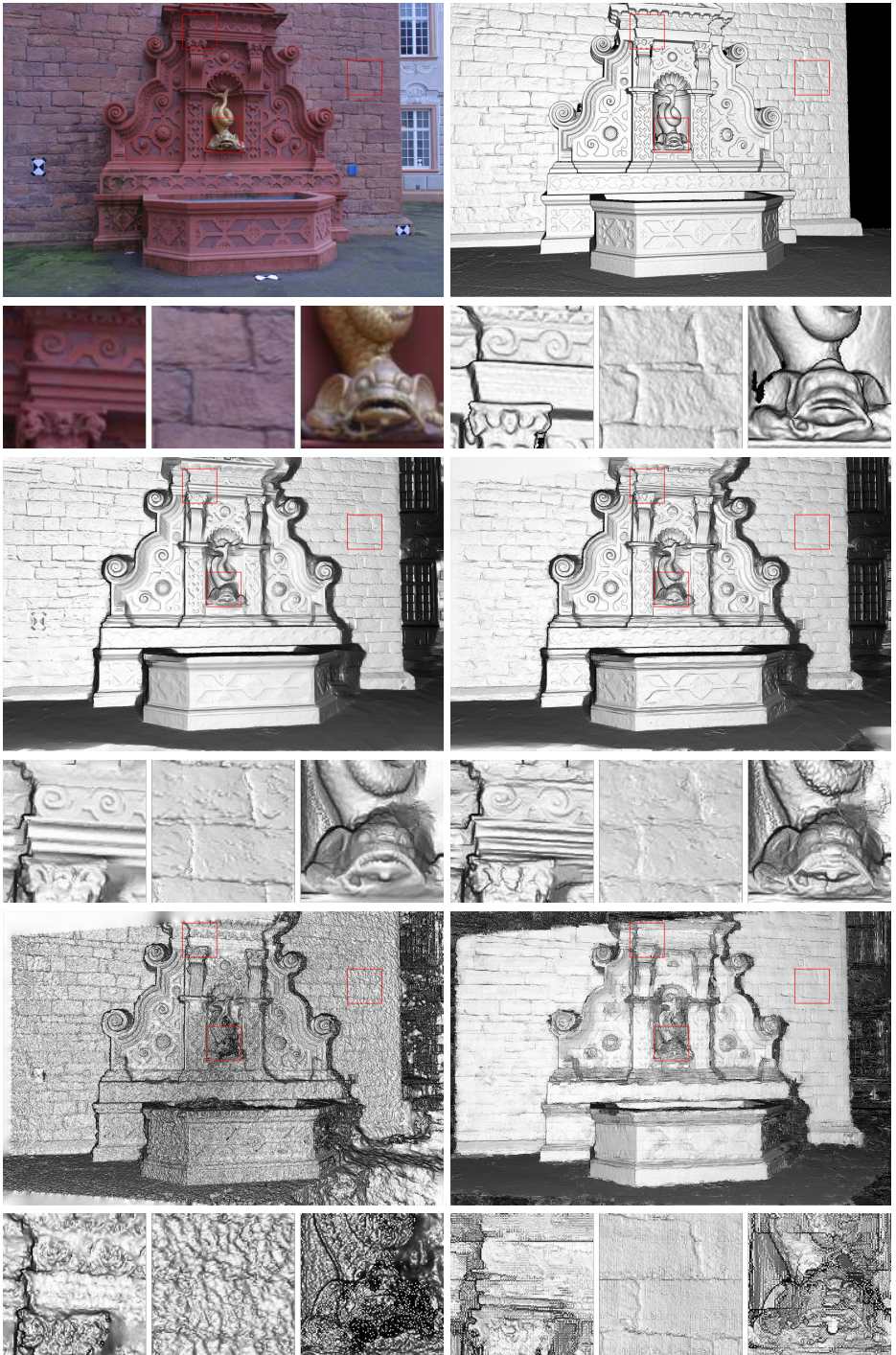


Figure 5: Real-world *Fountain-P11* dataset [5]. **First row:** Reference image, Ground truth. **Second row:** Our combined approach, our pure stereo approach. **Third row:** Graber *et al.* [2], Galliani *et al.* [1].

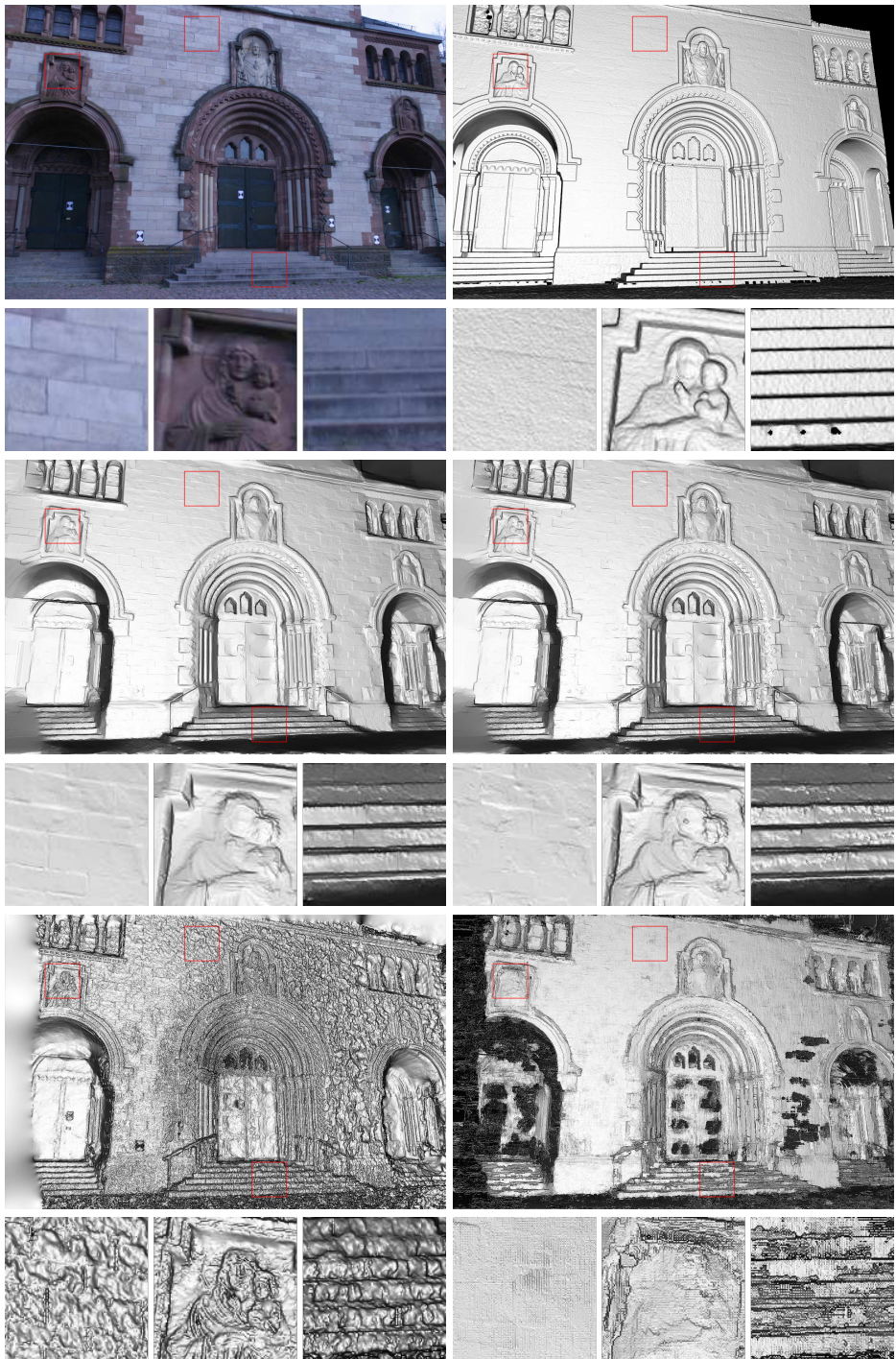


Figure 6: Real-world *Herz-Jesu-P8* dataset [5]. **First row:** Reference image, Ground truth. **Second row:** Our combined approach, our pure stereo approach. **Third row:** Graber *et al.* [2], Galliani *et al.* [1].



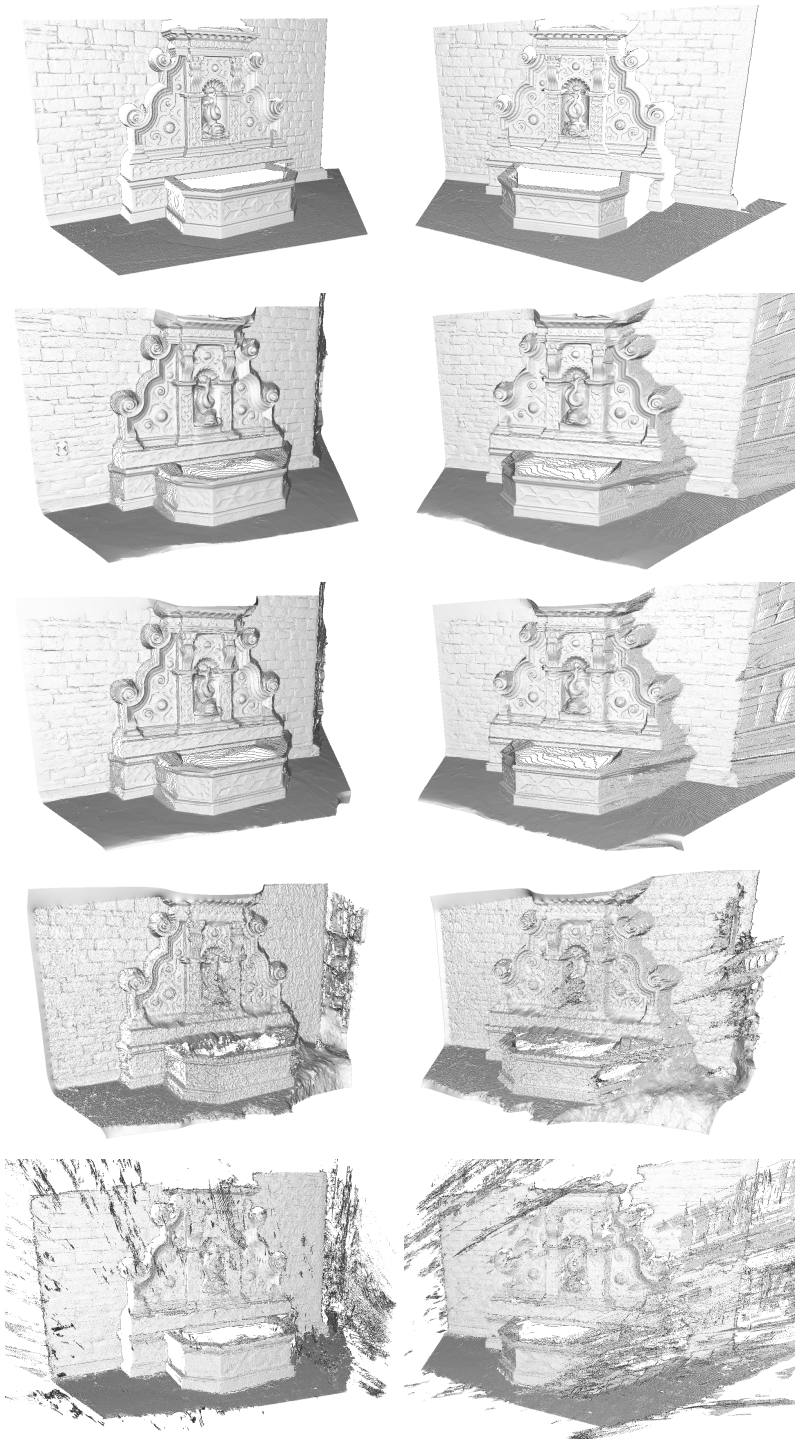


Figure 7: Real-world *Fountain-P11* dataset [5]. **From top to bottom:** Ground truth, our combined approach, our pure stereo approach, Graber *et al.* [2], Galliani *et al.* [1].

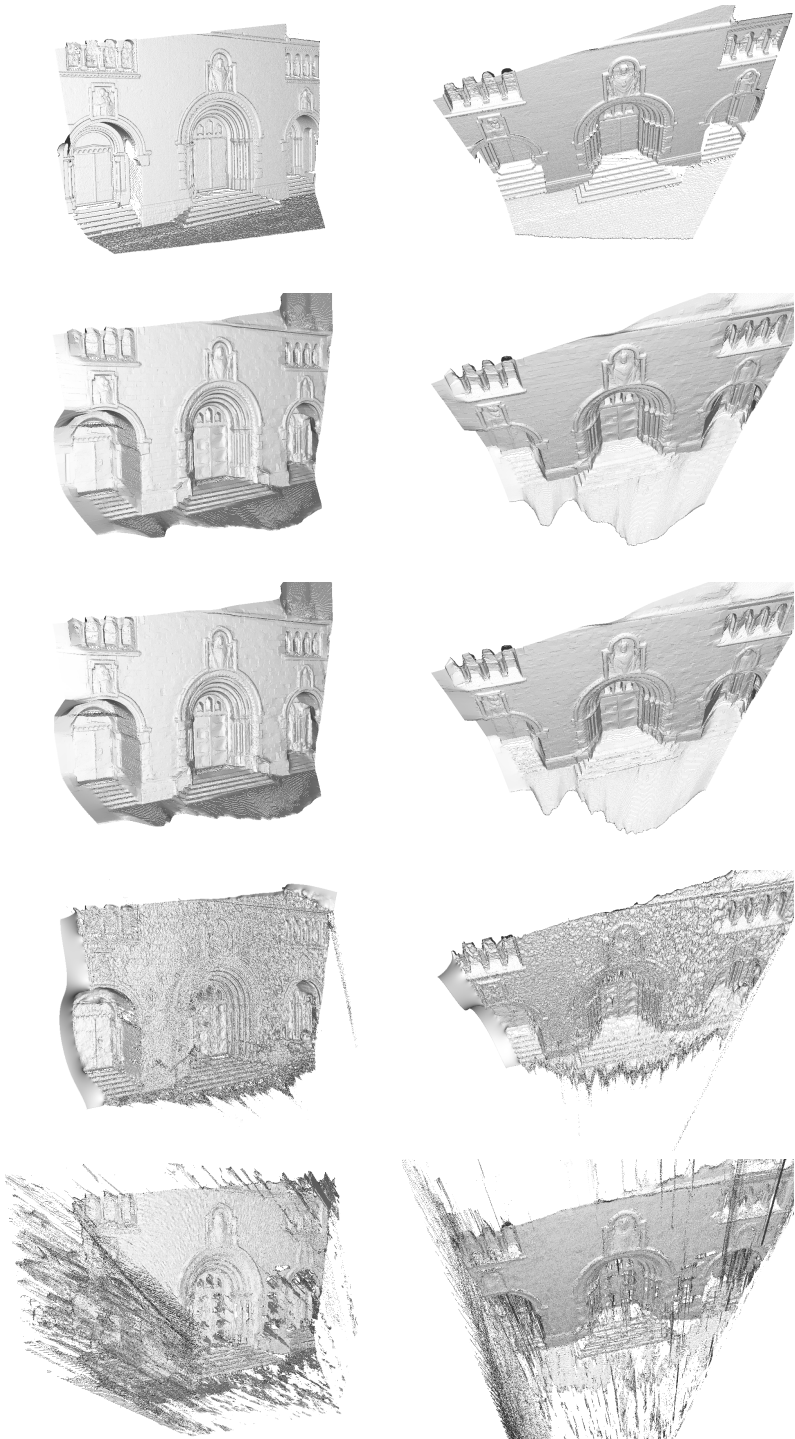


Figure 8: Real-world *Herz-Jesu-P8* dataset [5]. **From top to bottom:** Ground truth, our combined approach, our pure stereo approach, Graber *et al.* [2], Galliani *et al.* [1].

## References

- [1] S. Galliani, K. Lasinger, and K. Schindler. Massively parallel multiview stereopsis by surface normal diffusion. In *Proc. IEEE International Conference on Computer Vision*, pages 873–881, 2015.
- [2] G. Graber, J. Balzer, S. Soatto, and T. Pock. Efficient minimal-surface regularization of perspective depth maps in variational stereo. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 511–520, 2015.
- [3] D. Maurer, Y. C. Ju, M. Breuß, and A. Bruhn. An efficient linearisation approach for variational perspective shape from shading. In *Proc. German Conference on Pattern Recognition*, pages 249–261, 2015.
- [4] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 310–315, 1991.
- [5] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
- [6] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, 2011.