Supplementary Material

The Supplementary Material is organized as follows. In Appendix A, we derive the mathematical formulation at the core of our method. In Appendix B, we provide a novel analysis of the multiplicative factor $\gamma_{b\to a}$ used by BiNI [7] and extended in our method, and provide important insights on its effect on convergence. Appendix C provides additional insights on the positivity of the log term in our formulation ((15) in the main paper), including a mathematical proof that this property is preserved throughout the optimization, and discusses corner cases. In Appendix D, we study the impact of the choice of the ray direction vector $\boldsymbol{\tau_m}$, that controls our local planarity assumption. In Appendix E, we study the effect of the discontinuity activation term $\beta^{(t)}_{b\to a}$ in our formulation. Appendix F presents an ablation on different pixel connectivity. Appendix G presents an evaluation of the formulation accuracy with metrics in addition to the one introduced in Sec. 4.2. Appendix H provides results of our method under noisy input normals. Appendix I provides an evaluation on the DiLiGenT-MV dataset [25], which extends the DiLiGenT dataset. Finally, Appendix J discusses the limitations of our method.

A. Derivation of our formulation

In the following Section, we provide a derivation of the coefficients (3) of our formulation (2). Rearranging the equations in the system (1) emerging from our local planarity assumption and using $x_b = \tau_{x_b} z_b$, $y_b = \tau_{y_b} z_b$ (by definition of τ_{x_b}, τ_{y_b}) yields the following linear system in the variables $dx_{ma}, dy_{ma}, dz_{ma}, dx_{mb}, dy_{mb}, dz_{mb}$:

$$\mathbf{C} \cdot \begin{bmatrix} dx_{ma} \\ dy_{ma} \\ dz_{ma} \\ dx_{mb} \\ dy_{mb} \\ dz_{mb} \end{bmatrix} = \mathbf{d}, \tag{17}$$

where

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & -\tau_{x_m} \\ 0 & 0 & 0 & 0 & 1 & -\tau_{y_m} \\ -1 & 0 & \tau_{x_a} & 1 & 0 & -\tau_{x_a} \\ 0 & -1 & \tau_{y_a} & 0 & 1 & -\tau_{y_a} \\ 0 & 0 & 0 & n_{bx} & n_{by} & n_{bz} \\ n_{ax} & n_{ay} & n_{az} & 0 & 0 & 0 \end{bmatrix}, \text{ and}$$
$$\mathbf{d} = \begin{bmatrix} (\tau_{x_m} - \tau_{x_b}) z_b \\ (\tau_{y_m} - \tau_{y_b}) z_b \\ (\tau_{x_a} - \tau_{x_b}) z_b \\ (\tau_{y_a} - \tau_{y_b}) z_b \\ 0 \\ -n_{az} \varepsilon_{b\to a} \end{bmatrix}. \tag{18}$$

Solving (17) yields the following expressions for dz_{ma} and dz_{mb}:

$$\begin{aligned} dz_{ma} =& \frac{-n_{az}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a}} \cdot \varepsilon_{b\to a} + \\ & \frac{(n_{ax}\tau_{x_a} + n_{ay}\tau_{y_a} - n_{ax}\tau_{x_m} - n_{ay}\tau_{y_m}) \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_b}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b \\ =& \frac{-n_{az}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a}} \cdot \varepsilon_{b\to a} + \frac{(\boldsymbol{n_a}^\top \boldsymbol{\tau_a} - \boldsymbol{n_a}^\top \boldsymbol{\tau_m}) \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_b}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b, \\ dz_{mb} =& \frac{n_{bx}\tau_{x_b} + n_{by}\tau_{y_b} - n_{bx}\tau_{x_m} - n_{by}\tau_{y_m}}{\boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b \\ =& \frac{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot (\boldsymbol{n_b}^\top \boldsymbol{\tau_b} - \boldsymbol{n_b}^\top \boldsymbol{\tau_m})}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b. \end{aligned} \tag{19}$$

The final step to obtain our formulation (2), (3) follows from writing:

$$\begin{aligned} z_a =& z_b + dz_{mb} - dz_{ma} \\ =& \frac{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b + \\ & \frac{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot (\boldsymbol{n_b}^\top \boldsymbol{\tau_b} - \boldsymbol{n_b}^\top \boldsymbol{\tau_m})}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b + \\ & \frac{-(\boldsymbol{n_a}^\top \boldsymbol{\tau_a} - \boldsymbol{n_a}^\top \boldsymbol{\tau_m}) \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_b}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b + \\ & \frac{n_{az}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a}} \cdot \varepsilon_{b\to a} \\ =& \frac{n_{az}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a}} \cdot \varepsilon_{b\to a} + \frac{\boldsymbol{n_a}^\top \boldsymbol{\tau_m} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_b}}{\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m}} \cdot z_b. \end{aligned} \tag{20}$$

Alternative derivation. An alternative, more concise derivation[1] can be obtained by noting that the perpendicularity constraints encoded by the last two equations in (1) can be more compactly expressed as

$$\boldsymbol{n_a}^\top (\boldsymbol{p_m} + \boldsymbol{\varepsilon_z} - \boldsymbol{p_a}) = 0 \tag{21}$$

$$\boldsymbol{n_b}^\top (\boldsymbol{p_m} - \boldsymbol{p_b}) = 0, \tag{22}$$

where $\boldsymbol{\varepsilon_z} := (0, 0, \varepsilon_{b\to a})^\top$. From (22) it follows that

$$\frac{\boldsymbol{n_b}^\top \boldsymbol{p_b}}{\boldsymbol{n_b}^\top \boldsymbol{p_m}} = 1. \tag{23}$$

Expanding (21) and multiplying its first term by 1 using the equivalence (23) yields

$$\frac{\boldsymbol{n_a}^\top \boldsymbol{p_m} \cdot \boldsymbol{n_b}^\top \boldsymbol{p_b}}{\boldsymbol{n_b}^\top \boldsymbol{p_m}} + \boldsymbol{n_a}^\top \boldsymbol{\varepsilon_z} - \boldsymbol{n_a}^\top \boldsymbol{p_a} = 0. \tag{24}$$

Using $\boldsymbol{p_i} = z_i \boldsymbol{\tau_i}, i \in \{a, b, m\}$ (by definition) and the fact that $\boldsymbol{n_a}^\top \boldsymbol{\varepsilon_z} = n_{az} \cdot \varepsilon_{b\to a}$, (24) can be rewritten as

$$\frac{\boldsymbol{n_a}^\top \boldsymbol{\tau_m} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_b}}{\boldsymbol{n_b}^\top \boldsymbol{\tau_m}} z_b + n_{az} \cdot \varepsilon_{b\to a} - (\boldsymbol{n_a}^\top \boldsymbol{\tau_a}) z_a = 0. \tag{25}$$

Dividing all terms in (25) by $\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$ and rearranging yields our formulation (2), (3).

[1] We thank the anonymous reviewer NayZ for suggesting this alternative derivation.

B. Influence of the multiplicative factor $\gamma_{b\to a}$

As noted in Sec. 2 of the Supplementary Material of BiNI [9], the coefficient $\gamma_{b\to a}$[2], which we extend in our formulation, is crucial to achieving optimal convergence during optimization. In particular, their formulation based on the functional $\gamma_{b\to a}(\tilde{z}_a - \tilde{z}_b) = \delta_{b\to a}$ ((8) in the main paper) performs significantly better than the one derived from the equivalent equation $\tilde{z}_a - \tilde{z}_b = \delta_{b\to a}/\gamma_{b\to a}$. Similarly, we find that our formulation $\gamma_{b\to a}(\tilde{z}_a - \tilde{z}_b) = \gamma_{b\to a}\log(\omega_{b\to a} + \omega_{\varepsilon_a}\cdot\alpha_{b\to a})$ ((11) in the main paper) achieves significantly better convergence than the equivalent $\tilde{z}_a - \tilde{z}_b = \log(\omega_{b\to a} + \omega_{\varepsilon_a}\cdot\alpha_{b\to a})$.

In the following, we provide below a novel analysis of this phenomenon in light of our generic formulation based on ray direction vectors, which allows rewriting $\gamma_{b\to a}$ as

$$\gamma_{b\to a} = f\cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}, \tag{26}$$

where f is the (fixed) focal length, which we generalize to the (pixel-pair specific) factor $\|\boldsymbol{u_b} - \boldsymbol{u_a}\| \,/\, \|\boldsymbol{\tau_b} - \boldsymbol{\tau_a}\|$[3]. All the supporting experiments in this Section are run on the DiLiGenT benchmark, for 1200 iterations and for simplicity using our version without $\alpha_{b\to a}$ computation.

We start by noting that, for each pixel pair (a, b), the coefficient $\gamma_{b\to a}$ has two effects on the optimization:

- **Effect 1 (weighting):** On one side, it introduces a quadratic factor $\gamma^2_{b\to a}$ in the corresponding term of the optimization cost function $(\tilde{\mathbf{A}}\tilde{\mathbf{z}}-\tilde{\mathbf{b}})^\top\tilde{\mathbf{W}}(\tilde{\mathbf{A}}\tilde{\mathbf{z}}-\tilde{\mathbf{b}})$ (*cf*. (5) in the main paper), or equivalently in its associated normal equation $\tilde{\mathbf{A}}^\top\tilde{\mathbf{W}}\tilde{\mathbf{A}}\tilde{\mathbf{z}} = \tilde{\mathbf{A}}^\top\tilde{\mathbf{W}}\tilde{\mathbf{b}}$, since both the rows of $\tilde{\mathbf{A}}$ and the corresponding elements of $\tilde{\mathbf{b}}$ are scaled by a factor $\gamma_{b\to a}$ (*cf*. (8) and (11) in the main paper). In other words, the optimization cost function reads as

$$(\tilde{\mathbf{A}}\tilde{\mathbf{z}}-\tilde{\mathbf{b}})^\top\tilde{\mathbf{W}}(\tilde{\mathbf{A}}\tilde{\mathbf{z}}-\tilde{\mathbf{b}}) = \sum_{(a,b)} w^{\text{BiNI}}_{b\to a}\cdot\gamma^2_{b\to a}\cdot(\tilde{z}_a-\tilde{z}_b-\text{RHS})^2, \tag{27}$$

  where RHS is $\delta_{b\to a}/\gamma_{b\to a}$ for BiNI and $\log(\omega_{b\to a} + \omega_{\varepsilon_a}\cdot\alpha_{b\to a})$ for Ours. Therefore, each residual is effectively scaled by $w^{\text{BiNI}}_{b\to a}\cdot\gamma^2_{b\to a}$ rather than only by $w^{\text{BiNI}}_{b\to a}$.
- **Effect 2 (sharpness of the bilateral weights):** On the other side, it impacts the magnitude of the bilateral weights $w^{\text{BiNI}}_{b\to a} = \sigma_k(\text{res}^2_{-b\to a} - \text{res}^2_{b\to a})$, where $\text{res}_{b\to a} \coloneqq \gamma_{b\to a}(\tilde{z}_a - \tilde{z}_b)$ (see also (10) in the main paper). Since from (26) $\gamma_{b\to a} \approx \gamma_{-b\to a}$, with exact equality when f is constant, it follows that

$$\begin{aligned} w^{\text{BiNI}}_{b\to a} &= \sigma_k(\gamma^2_{b\to a}\cdot((\tilde{z}_a-\tilde{z}_b)^2 - (\tilde{z}_a-\tilde{z}_{-b})^2)) \\ &= \sigma_{k\cdot\gamma^2_{b\to a}}((\tilde{z}_a-\tilde{z}_b)^2 - (\tilde{z}_a-\tilde{z}_{-b})^2), \end{aligned} \tag{28}$$

  i.e., $\gamma^2_{b\to a}$ can be subsumed into the parameter k of the sigmoid σ_k. As a consequence, $\gamma_{b\to a}$ controls the convergence of the bilateral weights, so that for fixed $\tilde{z}_a$ and $\tilde{z}_b$, a larger $\gamma^2_{b\to a}$ causes smaller depth differences between the two sides to be detected as a one-sided discontinuity, and smaller values result in a less sharp convergence.

Crucially, we observe that the effects of the two terms f and $\boldsymbol{n_a}^\top\boldsymbol{\tau_a}$ in (26) can be decoupled and summarized in the following two Propositions:

Proposition 1: Effect of the term f

The term f acts as a constant (or near constant, in the case of $f = \|\boldsymbol{u_b} - \boldsymbol{u_a}\| \,/\, \|\boldsymbol{\tau_b} - \boldsymbol{\tau_a}\|$) that controls the sharpness of the bilateral weights $w^{\text{BiNI}}_{b\to a}$.

Proposition 2: Effect of the term $n_a{}^\top\tau_a$

The term $\boldsymbol{n_a}^\top\boldsymbol{\tau_a}$ introduces an active weighting mechanism (in addition to $w^{\text{BiNI}}_{b\to a}$) based on the collinearity between surface normals and ray directions, reducing the influence of pixel pairs close to a discontinuity.

We provide below arguments and empirical verifications supporting the above Propositions.

Argument for Proposition 1. Since f is constant (or approximately constant), it can be factored out of each term $\gamma^2_{b\to a}$ in the optimization cost function (27). Since multiplying the cost function by a constant factor does not affect its minimizing solution, it follows that the term f is not an influencing factor for Effect 1 (weighting). We verify this by running our method using $\gamma_{b\to a} = \boldsymbol{n_a}^\top\boldsymbol{\tau_a}$ in our cost function (27) and $\gamma_{b\to a} = f\cdot\boldsymbol{n_a}^\top\boldsymbol{\tau_a}$ in the bilateral weights (28). As expected, up to minimal differences that we attribute to machine precision, the results match those obtained when using the full factor $\gamma_{b\to a} = f\cdot\boldsymbol{n_a}^\top\boldsymbol{\tau_a}$ in the cost function (*cf*. first and second row in Tab. 3).

We verify that instead the term f does indeed contribute to Effect 2 (sharpness of the bilateral weights) by varying its value in the $\gamma_{b\to a}$ factor of the bilateral weights, while maintaining a fixed $\gamma_{b\to a} = \boldsymbol{n_a}^\top\boldsymbol{\tau_a}$ in our cost function. Comparing rows 2 to 5 in Tab. 3 shows that indeed different values of f result in different convergence; while the change is object-specific, the main emerging trend appears to indicate that worse convergence is obtained for lower values of f, which correspond to a less sharp sigmoid.

Argument for Proposition 2. Since unlike f the term $\boldsymbol{n_a}^\top\boldsymbol{\tau_a}$ is highly pixel specific, it is not possible to find a single constant that can be absorbed into the parameter k of the sigmoid. It is therefore not straightforward to draw conclusions about its contribution to Effect 2 (sharpness of the bilateral weights). We can however verify that the term

[2] Denoted as $\tilde{n}_z$ in [9].

[3] Note that for an ideal pinhole camera with $f = f_x = f_y$ one has $\|\boldsymbol{u_b} - \boldsymbol{u_a}\| = \|(u_b - u_a, v_b - v_a)\|$ and $\|\boldsymbol{\tau_b} - \boldsymbol{\tau_a}\| = \|((u_b - u_a)/f, (v_b - v_a)/f, 0)\| = \|\boldsymbol{u_b} - \boldsymbol{u_a}\|/f$, from which one recovers $\|\boldsymbol{u_b} - \boldsymbol{u_a}\| \,/\, \|\boldsymbol{\tau_b} - \boldsymbol{\tau_a}\| = f$.

Figure 8. **Visualization of the terms $|\boldsymbol{n_a}^\top \boldsymbol{\tau_a}|$, DiLiGenT dataset [31].** The terms encode the degree of collinearity between the surface normals and the ray direction vectors. Low values are attained at pixels where the ray direction vector is perpendicular to the surface normal, a necessary condition for the corresponding point to lie on the object boundary.

Value of $\gamma_{b\to a}$		bear	buddha	cat	cow	harvest	pot1	pot2	reading	goblet
Cost function (27)	$w^{\text{BiNI}}_{b\to a}$ (28)									
$f \cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	$f \cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	0.07	0.26	0.06	0.08	5.54	0.49	0.13	0.11	6.33
$\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	$f \cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	0.07	0.25	0.06	0.08	5.33	0.49	0.13	0.12	6.60
$\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	$3000 \cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	0.09	0.27	0.11	0.09	3.89	0.47	0.15	0.12	7.96
$\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	$2000 \cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	0.06	0.98	0.17	0.18	1.71	0.48	0.25	0.27	8.63
$\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	$1000 \cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	0.04	1.41	0.08	0.30	2.51	0.72	0.28	1.19	9.46
f	$f \cdot \boldsymbol{n_a}^\top \boldsymbol{\tau_a}$	0.48	2.53	0.69	0.39	4.84	14.40	0.42	3.16	10.28

Table 3. **Ablation on the terms in $\gamma_{b\to a}$, DiLiGenT dataset [31].** For each experiment, we report the mean absolute depth error (MADE) [mm]. All experiments are without $\alpha_{b\to a}$ computation, $k = 2$ for $w^{(t)}_{b\to a}$ (as default), and are run for 1200 iterations. Where used, f denotes $\|\boldsymbol{u_b} - \boldsymbol{u_a}\| \,/\, \|\boldsymbol{\tau_b} - \boldsymbol{\tau_a}\|$. For reference, the values of f_x and f_y in the dataset are $f_x \approx 3772.1$ [px] and $f_y \approx 3759.0$ [px].

$\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$ has a strong influence on Effect 1 (weighting), by removing it from the $\gamma_{b\to a}$ factor of the cost function (which is therefore set to f), while maintaining it in $\gamma_{b\to a}$ in the bilateral weights. Comparing the last and the first row of Tab. 3 confirms that the accuracy of the reconstruction dramatically decreases when the term does not contribute to the cost function, which indicates that it plays an active role in determining the convergence of the optimization, by introducing equation-specific weights. Interestingly, as we previously observed in Sec. 3.1, the term $\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$ strongly correlates with surface discontinuities, with pixels close to object boundaries or local discontinuities attaining a small value for this term. More generally, as evident from its dot-product definition, the term $\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$ encodes the degree of collinearity between surface normal and the ray direction vector at each pixel (*cf*. Fig. 8 for a visualization). As a consequence, its effect over the optimization is to balance the influence of the residuals, decreasing the weight of errors close to discontinuities, while increasing the influence of residuals at points where the camera rays hit the surface at a close-to-right angle.

C. Analysis of the positivity of the log term

In this Section we provide further insights on the positivity of the log term in our formulation ((15) in the main paper).

We start by empirically verifying that, for our choice $\boldsymbol{\tau_m} = (\boldsymbol{\tau_a} + \boldsymbol{\tau_b})/2$, the terms $\boldsymbol{n_a}^\top \boldsymbol{\tau_m}$ and $\boldsymbol{n_b}^\top \boldsymbol{\tau_m}$ are both strictly positive for all but a single pixel (object `pot1`) across all the objects in the DiLiGenT dataset, used for our main experiments. Furthermore, also for this outlier pixel, the effects of the two pixels cancel out and the corresponding term $\omega_{b\to a} = (\boldsymbol{n_a}^\top \boldsymbol{\tau_m} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_b})/(\boldsymbol{n_a}^\top \boldsymbol{\tau_a} \cdot \boldsymbol{n_b}^\top \boldsymbol{\tau_m})$ is strictly positive, leading to a positive log term at all pixels in the first iteration of our optimization.

We now briefly analyze under which conditions we can expect an outlier, negative $\omega_{b\to a}$ term. Since, as noted in Sec. 3.1, for physically meaningful normals (*i.e.*, corresponding to observable surface points) the positivity of $\omega_{b\to a}$ reduces to the positivity of $\boldsymbol{n_a}^\top \boldsymbol{\tau_m}$ and $\boldsymbol{n_b}^\top \boldsymbol{\tau_m}$, we can focus on the case where the latter two terms have opposite signs. Figure 9 provides an illustration of an instance in which such a corner case may arise. In the depicted setting, the surface has low inclination relative to the camera on the side of point $\boldsymbol{p_a}$, but large inclination on the side of point $\boldsymbol{p_b}$. As consequence, on the side of $\boldsymbol{p_a}$ both the angles between $\boldsymbol{n_a}$ and $\boldsymbol{\tau_a}$ and between $\boldsymbol{n_a}$ and $\boldsymbol{\tau_m}$ are significantly larger than $90°$, *i.e.* $\boldsymbol{n_a}^\top \boldsymbol{\tau_a} < 0$ and $\boldsymbol{n_a}^\top \boldsymbol{\tau_m} < 0$. On the opposite side, however, the angle between $\boldsymbol{n_b}$ and $\boldsymbol{\tau_b}$ is only slightly larger than $90°$ (hence $\boldsymbol{n_b}^\top \boldsymbol{\tau_b} \approx 0$, but still negative), while the angle between $\boldsymbol{n_b}$ and $\boldsymbol{\tau_m}$ is smaller than $90°$, causing $\boldsymbol{n_b}^\top \boldsymbol{\tau_m}$ to be positive and therefore $\omega_{b\to a}$ to be negative. While such outlier cases might indeed arise, it is possible to detect and handle them, for instance by excluding the corresponding equation from the optimization or by choosing a different value of $\boldsymbol{\tau_m}$ (*cf*. Appendix D). Furthermore, their occurrence is unlikely in practice, since the sign flipping between $\boldsymbol{n_b}^\top \boldsymbol{\tau_b}$ and $\boldsymbol{n_b}^\top \boldsymbol{\tau_m}$ would need to occur within a very limited angular space: as a reference, using $\boldsymbol{\tau_m} = (\boldsymbol{\tau_a} + \boldsymbol{\tau_b})/2$, the angle between $\boldsymbol{\tau_m}$ and $\boldsymbol{\tau_b}$ is approximately $\frac{1}{2} \arctan\left(\frac{1\text{ px}}{3700\text{ px}}\right) \approx 0.008°$ in the DiLiGenT

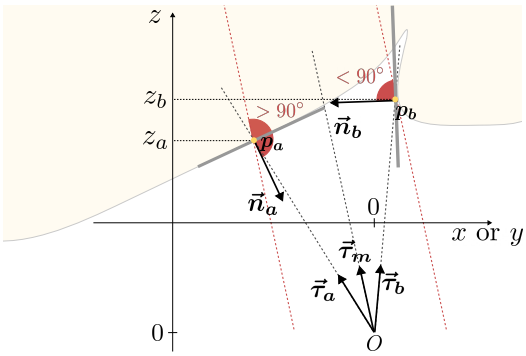

Figure 9. **Visualization of a corner case in our local planarity assumption in 3D.** For the chosen configuration, the ray direction vector $\boldsymbol{\tau_m}$ forms an angle smaller than $90°$ with $\boldsymbol{n_b}$ and larger than $90°$ with $\boldsymbol{n_a}$, resulting in $\boldsymbol{n_a}^\top \boldsymbol{\tau_m} < 0$ and $\boldsymbol{n_b}^\top \boldsymbol{\tau_m} > 0$.

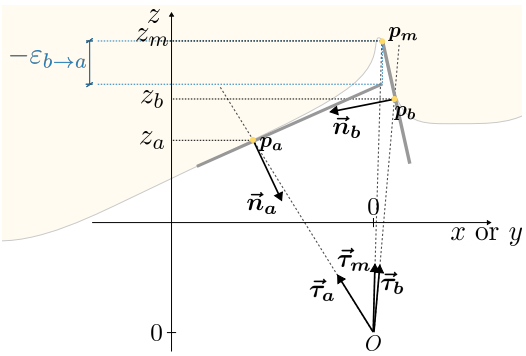

Figure 10. **Visualization of an adaptive strategy for $\boldsymbol{\tau_m}$.** If the surface has a large inclination relative to the camera on one of the two sides (here the side of $\boldsymbol{p_b}$, hence $|\boldsymbol{n_b}^\top \boldsymbol{\tau_b}| \ll |\boldsymbol{n_a}^\top \boldsymbol{\tau_a}|$), orienting $\boldsymbol{\tau_m}$ closer to the latter side yields a smaller $|\varepsilon_{b\to a}|$.

dataset, for which $f_x \approx 3772.1$ px and $f_y \approx 3759.0$ px.

Assuming $\omega_{b\to a} > 0$, hence that the argument of the log term in (15) is positive in the first optimization iteration, it is straightforward to show that the argument also stays positive throughout the optimization, as we prove below.

From (14) in the main paper, $\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t+1)} = \exp\left(\tilde{z}_a^{(t)} - \tilde{z}_b^{(t)}\right)$. Since the exponential function is bijective and defined anywhere in $\mathbb{R}$, it follows that for any value of $\tilde{z}_a^{(t)}$ and $\tilde{z}_b^{(t)}$ a corresponding value for the term $\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t+1)}$ can be found and thereby of $\alpha_{b\to a}^{(t+1)}$ (provided that $\omega_{\varepsilon_a} \neq 0$, *i.e.*, from (3) $n_{a_z} \neq 0$, which is always the case because $n_{a_z} = 0$ corresponds to a surface perpendicular to the image plane). Since the exponential function has strictly positive codomain, it also follows that for all t's:

$$\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t+1)} > 0. \tag{29}$$

From $\omega_{b\to a} > 0$ and (29) and since $\beta_{b\to a}^{(t)} \in [0,1]$ by design, it follows that $\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \cdot \beta_{b\to a}^{(t)} > 0$, which proves the hypothesis. Indeed:

- If $\omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \geq 0$, one has

$$\begin{aligned}
&\omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \cdot \beta_{b\to a}^{(t)} \geq 0 && \left(\beta_{b\to a}^{(t)} \geq 0\right)\\
\Rightarrow\ &\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \cdot \beta_{b\to a}^{(t)} \geq \omega_{b\to a} && (\omega_{b\to a} \in \mathbb{R})\\
\Rightarrow\ &\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \cdot \beta_{b\to a}^{(t)} > 0; && (\omega_{b\to a} > 0)
\end{aligned}$$

- If $\omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} < 0$, it follows that

$$\begin{aligned}
&\omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \cdot \beta_{b\to a}^{(t)} \geq \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} && \left(\beta_{b\to a}^{(t)} \in [0,1]\right)\\
\Rightarrow\ &\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \cdot \beta_{b\to a}^{(t)} \geq\\
&\qquad \omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} && (\omega_{b\to a} > 0)\\
\Rightarrow\ &\omega_{b\to a} + \omega_{\varepsilon_a} \cdot \alpha_{b\to a}^{(t)} \cdot \beta_{b\to a}^{(t)} > 0. && (\text{from (29)})
\end{aligned}$$

D. Impact of the choice of $\boldsymbol{\tau_m}$

In the following Section, we provide an ablation on the choice of $\boldsymbol{\tau_m}$, which controls the planar assumption of our method (*cf*. Fig. 9 and Fig. 2 in the main paper).

As mentioned in Sec. 3.1 in the main paper, $\boldsymbol{\tau_m}$ can be parametrized as interpolating between $\boldsymbol{\tau_a}$ and $\boldsymbol{\tau_b}$, *i.e.*, $\boldsymbol{\tau_m} = \boldsymbol{\tau_a} + \lambda_m(\boldsymbol{\tau_b} - \boldsymbol{\tau_a})$, with $\lambda_m \in [0,1]$. A natural choice, which we adopt in our main experiments, is to orient $\boldsymbol{\tau_m}$ at an equal angular distance from $\boldsymbol{\tau_a}$ and $\boldsymbol{\tau_b}$, *i.e.* setting $\lambda_m = 0.5$ uniformly for all pixels. However, we note that in certain settings a pixel-pair-specific choice $\lambda_{m,b\to a} : \boldsymbol{\tau_m} = \boldsymbol{\tau_a} + \lambda_{m,b\to a}(\boldsymbol{\tau_b} - \boldsymbol{\tau_a})$ might be desirable. An argument in favor of this point is for instance shown through a corner case similar to that considered in Appendix C (Fig. 10), in which on one of the two sides (the side of $\boldsymbol{p_b}$ in Fig. 10) the surface has a significantly larger inclination relative to the camera. As a consequence, as exemplified by Fig. 10, our planar assumption holds more accurately if $\boldsymbol{\tau_m}$ is oriented closer to the side with the larger inclination, in which case a smaller discontinuity term $|\varepsilon_{b\to a}|$ is obtained. Since, as mentioned in Appendix B, the quantity $\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$ naturally encodes surface orientation with respect to the camera, the condition of unbalanced inclination between the two sides can also be expressed as $|\boldsymbol{n_b}^\top \boldsymbol{\tau_b}| \ll |\boldsymbol{n_a}^\top \boldsymbol{\tau_a}|$. In this ablation, we additionally consider the quantity n_{az}, which similarly to $\boldsymbol{n_a}^\top \boldsymbol{\tau_a}$ attains a low value in proximity to discontinuities.

We note that the interpolating function $\lambda_{m,b\to a}$ needs to be such that $\boldsymbol{\tau_m}$ intersects the same surface point $\boldsymbol{p_m}$ both in the direction $b \to a$ (*i.e.*, when considering b a neighbor of a) and in the direction $a \to b$ (*i.e.*, when considering a a neighbor of b). This can be expressed mathematically by the condition $\lambda_{m,b\to a} = 1 - \lambda_{m,a\to b}$. We note that the sigmoid function naturally fulfills this condition when composed with an even function, and we therefore set in this ablation $\lambda_{m,b\to a} = \sigma_{k_m}(f(a,b))$, with different val-

ues for k_m, and with $f(a,b)$ either $(\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a})^2 - (\boldsymbol{n_b}^\mathsf{T}\boldsymbol{\tau_b})^2$, $n_{az}^2 - n_{bz}^2$, or $(n_{az}\cdot\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a})^2 - (n_{bz}\cdot\boldsymbol{n_b}^\mathsf{T}\boldsymbol{\tau_b})^2$.

Table 4 shows the results of this ablation, which we perform on the DiLiGenT dataset. For most objects, introducing a pixel-specific λ_m results generally in lower reconstruction accuracy using any of the functions $f(a,b)$ listed above; larger values of k_m (hence more sharply weighting inclination differences between the two sides) further decrease the performance. A noticeable exception is represented by the two objects with larger discontinuities (`harvest` and `goblet`), for which specific choices of parameters can lead to improved reconstruction accuracy.

Finally, we highlight that pixel-specific values of λ_m find an additional, critical application in handling potential outliers in the input normal map. We discuss this important aspect in detail in Appendix H.

E. Impact of the discontinuity activation term

In this Section, we provide an ablation analysis on the impact of our discontinuity activation term $\beta_{b\to a}^{(t)}$ on the reconstruction accuracy. Table 5 reports the mean absolute depth error on the DiLiGenT dataset as we vary the hyperparameters q and ρ (*cf*. (16) in the main paper), the effect of which can be visualized in Fig. 11. For $\rho = 0.25$, the results show object-specific trends, with some objects achieving higher accuracy for sharper changes of $\beta_{b\to a}^{(t)}$ (larger q, for instance `harvest`, `pot1`, `reading`) and others favoring a smoother discontinuity activation term (smaller q, for instance `bear`, `pot2`). For $\rho = 0.5$, the method achieves worse accuracy, in most instances also lower than the version without computation of $\alpha_{b\to a}$ (*cf*. Tab. 2 in the main paper). This performance drop is expected, since for $\rho = 0.5$ the discontinuity term significantly deviates from its designed objective, namely that it should tend smoothly to zero as $w_{b\to a}^{\mathrm{BiNI}^{(t-1)}} \to 0.5^-$ and smoothly to one as $w_{b\to a}^{\mathrm{BiNI}^{(t-1)}} \to 0^+$ (*cf*. Sec. 3.3 in the main paper for a detailed explanation of this design choice).

F. Impact of the connectivity

Since our method allows using pixel connectivities not limited to standard 4-connectivity, in this Section we investigate whether using alternative connectivities can yield improved reconstruction accuracy. Table 6 shows the results of this ablation, where we test our method on the DiLiGenT dataset using standard 4-connectivity (as in the main paper), 4-connectivity defined along the diagonals rather than the horizontal and vertical direction, and full 8-connectivity. While 4-connectivity along the diagonals, with very limited exceptions, generally results in significantly worse performance, we note that, interestingly, full 8-connectivity produces comparable or slightly better reconstructions than standard 4-connectivity on some objects

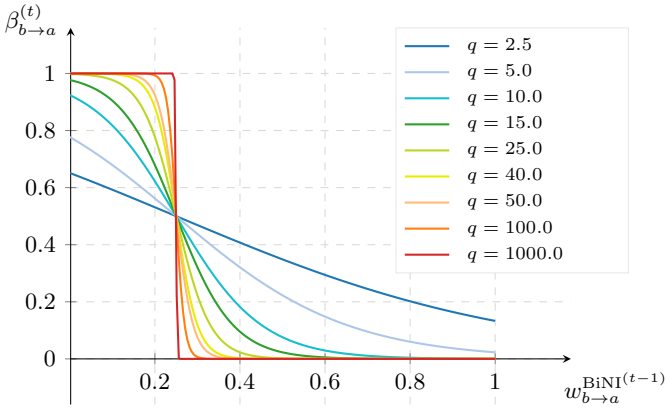

Figure 11. **Discontinuity activation term (16) for $\rho = 0.25$ and different values of q.** For $\rho = 0.5$, the plots are shifted to the right by 0.25 units along the $w_{b\to a}^{\mathrm{BiNI}^{(t-1)}}$ axis. *Cf*. Tab. 5 for a quantitative evaluation on the effect of the parameters ρ and q.

(*e.g.* `cow`, `pot1`, `pot2`). However, this improvement is contrasted by reduced accuracy on other objects (*e.g.* `buddha`, `cat`, `reading`) and reduced effect of the $\alpha_{b\to a}$ computation, leaving standard 4-connectivity as the most robust and balanced option.

G. Additional evaluations of the formulation accuracy

In Tables 7 and Tab. 8, similarly to Tab. 1 in the main paper, we provide metrics to evaluate how accurately our formulation approximates the ground-truth relation between depth and surface normals compared to previous methods. In particular, to complement the evaluation of the *absolute* accuracy from the main paper, we report here *relative* metrics, specifically the residual $|(\tilde{z}_a - \tilde{z}_b - \mathrm{RHS} \,/\, \gamma_{b\to a}) \,/\, \tilde{z}_a|$ computed on the ground-truth log-depth map (Tab. 7) and the residual $|(z_a - \exp(\mathrm{RHS} \,/\, \gamma_{b\to a}) \cdot z_b) \,/\, z_a|$ computed on the ground-truth depth map (Tab. 8), where RHS denotes the right-hand side of (8) for BiNI and (11) for Ours.

The results confirm the findings from the main paper. Namely, while for two objects our method has larger residual standard deviation than BiNI [7] (`buddha` and `pot1`), it achieves lower mean residual error by one or two orders of magnitude and lower standard deviation for most objects.

H. Results for noisy inputs

In this Section, we investigate the robustness of our method to noise in the input normal map.

Similarly to previous methods [9], we simulate the presence of outlier normals by replacing the original normals with randomly sampled unit vectors, with different percentages of sampled pixels. Figure 12 shows that without preprocessing the normal maps, our method can reconstruct

λ_m	k_m	bear	buddha	cat	cow	harvest	pot1	pot2	reading	goblet
0.5	N/A	0.07	0.26	0.06	0.08	4.83	0.50	0.13	0.12	6.56
$\sigma_{k_m}\left((\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a})^2-(\boldsymbol{n_b}^\mathsf{T}\boldsymbol{\tau_b})^2\right)$	1	0.15	0.33	0.09	0.12	5.12	0.52	0.17	0.19	5.73
	2	0.22	0.72	0.13	0.16	2.45	0.53	0.22	0.29	6.23
	3	0.29	1.40	0.16	0.19	3.66	0.56	0.30	0.38	6.11
$\sigma_{k_m}\left(n_{az}^2-n_{bz}^2\right)$	1	0.15	0.33	0.09	0.12	4.65	0.50	0.17	0.19	5.69
	2	0.22	0.71	0.13	0.16	2.51	0.53	0.22	0.28	6.12
	3	0.29	1.42	0.16	0.19	5.49	0.56	0.30	0.38	6.06
$\sigma_{k_m}\left((n_{az}\cdot\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a})^2-(n_{bz}\cdot\boldsymbol{n_b}^\mathsf{T}\boldsymbol{\tau_b})^2\right)$	1	0.11	0.25	0.08	0.11	4.95	0.51	0.16	0.15	5.38
	2	0.15	0.45	0.10	0.13	2.73	0.52	0.20	0.20	5.40
	3	0.19	1.03	0.13	0.16	2.74	0.55	0.24	0.39	5.61

Table 4. **Mean absolute depth error (MADE) [mm] on the DiLiGenT benchmark [31] for different choices of λ_m, where $\boldsymbol{\tau_m} = \boldsymbol{\tau_a} + \lambda_m(\boldsymbol{\tau_b} - \boldsymbol{\tau_a})$.** All the experiments are run for 1200 iterations with $\alpha_{b\to a} = 0$. σ_{k_m} denotes the sigmoid function $\sigma_{k_m}(x) = 1/(1+\exp(-k_m \cdot x))$.

ρ	q	bear	buddha	cat	cow	harvest	pot1	pot2	reading	goblet
0.25	2.5	0.04	0.28	**0.06**	0.11	4.35	0.57	0.13	0.17	5.86
	5.0	**0.02**	**0.22**	0.22	0.09	1.11	0.53	**0.12**	0.14	2.43
	10.0	**0.02**	0.25	**0.06**	**0.08**	0.93	0.54	**0.12**	0.16	1.63
	15.0	0.03	0.24	**0.06**	**0.08**	0.78	0.55	**0.12**	0.16	**1.52**
	25.0	0.03	0.25	**0.06**	0.10	0.83	0.55	0.13	0.13	5.78
	40.0	0.03	0.23	**0.06**	**0.08**	**0.60**	0.51	0.13	0.18	6.22
	50.0	0.03	0.24	**0.06**	**0.08**	0.73	0.49	0.13	0.17	4.72
	100.0	0.03	0.23	**0.06**	**0.08**	4.01	**0.48**	0.14	0.17	6.21
	1000.0	0.03	0.23	**0.08**	**0.08**	0.64	**0.48**	0.14	**0.10**	6.10
0.50	2.5	0.08	0.39	**0.06**	0.12	2.20	0.62	0.14	0.20	5.98
	5.0	0.09	0.47	0.09	0.12	3.40	0.64	0.13	0.52	6.25
	10.0	0.09	0.52	0.09	0.12	1.88	0.58	0.13	0.54	6.18
	15.0	0.09	0.57	0.08	0.12	2.52	0.64	0.18	0.55	6.14
	25.0	0.09	0.67	0.08	0.12	1.10	0.63	0.17	0.73	4.62
	40.0	0.09	0.40	0.11	0.12	2.13	0.69	0.16	0.59	6.96
	50.0	0.09	0.70	0.12	0.12	2.21	0.61	0.17	0.45	7.23
	100.0	0.10	0.83	0.11	0.11	2.03	0.60	0.16	0.46	7.26
	1000.0	0.10	0.70	0.14	0.11	2.58	0.87	0.16	0.51	6.94

Table 5. **Mean absolute depth error (MADE) [mm] on the DiLiGenT dataset [31] for $\rho \in \{0.25, 0.50\}$ and different values of q.** For each object, **bold** denotes the best result across the experiments. All the experiments are run for 1200 iterations.

Method	Connectivity	bear	buddha	cat	cow	harvest	pot1	pot2	reading	goblet
Ours w/o $\alpha_{b\to a}$ computation	4-connectivity	0.07	**0.26**	**0.06**	**0.08**	4.83	0.50	0.13	**0.12**	6.56
	4-connectivity (diagonal)	0.26	0.39	0.30	0.09	**1.68**	0.47	0.15	0.26	7.31
	8-connectivity	**0.06**	0.35	0.29	0.09	2.56	**0.36**	**0.12**	0.39	**4.44**
Ours	4-connectivity	**0.03**	**0.24**	**0.06**	**0.08**	**0.73**	0.49	**0.13**	**0.17**	**4.72**
	4-connectivity (diagonal)	0.12	0.69	0.28	0.09	1.76	0.50	0.14	0.42	5.56
	8-connectivity	0.15	0.35	0.32	**0.08**	3.82	**0.37**	**0.13**	0.50	5.14

Table 6. **Mean absolute depth error (MADE) [mm] on the DiLiGenT dataset [31] for different connectivities.** For each object and method, **bold** denotes the best result across the connectivities. All the experiments are run for 1200 iterations with $\boldsymbol{\tau_m} = (\boldsymbol{\tau_a} + \boldsymbol{\tau_b})/2$. Ours corresponds to the hyperparameter setting of our main experiments ($q = 50.0$ and $\rho = 0.25$ in (16)).

most of the underlying surface, but suffers from the presence of spike artifacts and non-smooth effects on the surface (second block from the top in Fig. 12). We note, however, that a large part of the outliers can and should be detected, because they correspond to physically impossible normals. In particular, as previously observed both in the main paper and in Appendix B, a necessary condition for the surface to be observable at one point $\boldsymbol{p_a}$ is that the dot product $\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a}$ at the corresponding pixel a is negative. We observe that enforcing this condition by applying an averaging filter to the normals at pixels where $\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a} > 0$ results in a reduction of the amount of spike artifacts (third block from the top in Fig. 12). We additionally note that the presence of outliers can also be detected by inspecting the distribution of $\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a}$ or of its absolute value: while in a natural surface these quantities vary continuously across the surface with the exception of boundary regions, for the perturbed normal maps salt-and-pepper noise can be observed in correspondence to

the outliers (*cf*. second row in the top block of Fig. 12). We verify that applying average filtering also to pixels where $|\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a}|$ deviates significantly from the mean value in its neighborhood further mitigates the effect of the outliers, removing spike artifacts and recovering the smoothness of the surface (*cf*. lowermost block in Fig. 12).

While the above test effectively highlights the impact of outliers on the reconstruction, we argue that it does not fully accurately reflect the statistical characteristics of noise emerging in real-world normal maps, in particular those predicted by learning-based methods. To provide an additional evaluation of the robustness of our method under noise in the input normals, we perturb the surface normals by rotating them around an axis that we randomly sample for each pixel, with an angle of rotation that we sample from a Gaussian distribution. Figure 13 shows the results of this ablation, where we vary the standard deviation of the Gaussian distribution between 1 and 10 degrees. Similarly to the experiment with outliers, providing the raw normal map as input to our method results in spike artifacts (second block from the top in Fig. 13). Noticeably, however, most of these artifacts can be corrected by average filtering of the pixels with invalid normals alone (third block from the top in Fig. 13), showing that physically impossible normals constitute the main factor behind these artifacts. As in the case with outliers, additionally filtering pixels where $|\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a}|$ deviates largely from the mean value in the pixels' neighborhood allows further reducing artifacts and removing spikes (lowermost block in Fig. 13).

Outlier filtering through $\boldsymbol{\tau_m}$. The spike artifacts resulting from the outlier normals have been identified in the literature as consequences of a type of *Gibbs phenomenon* [6, 17]. A closer analysis of the terms of our formulation reveals that such artifacts arise at outlier pixels where the terms $\boldsymbol{n_i}^\mathsf{T}\boldsymbol{\tau_j}$, for $(i,j) \in \{(a,a),(a,m),(b,b),(b,m)\}$, are either greater than 0 or have small magnitude, *i.e*., $\boldsymbol{n_i}^\mathsf{T}\boldsymbol{\tau_j} > 0$ or $|\boldsymbol{n_i}^\mathsf{T}\boldsymbol{\tau_j}| \approx 0$. In the latter case, in particular, the term $\omega_{b\to a}$, which depends on the multiplication of two such terms both in its numerator and its denominator, can significantly deviate from 1. This, in turn, results in $z_a \gg z_b$ or $z_a \ll z_b$ through (2) and thus introduces very large discontinuities that imbalance the optimization.

Crucially, our method offers a natural way to handle these outliers by controlling the ray direction $\boldsymbol{\tau_m} = \boldsymbol{\tau_a} + \lambda_m \cdot (\boldsymbol{\tau_b} - \boldsymbol{\tau_a})$ associated to the mid-point m (see Appendix D). We find that a simple strategy that results in an effective reduction of the influence of the outliers is to: *(i)* detect $\omega_{b\to a}$ terms that are outliers when $\lambda_m = 0.5$, evaluated as $|\log(\omega_{b\to a})| > \log(1+\epsilon_{\text{out}})$, where ϵ_{out} is a hyperparameter (for instance $\epsilon = 0.1$, corresponds to a depth variation larger than 10% between z_a and z_b, *cf*. (2)); *(ii)* uniformly sample multiple values of $\lambda_m \in [0,1]$ for these pixels and select the value of λ_m that yields the $\omega_{b\to a}$ term closest to 1. As shown in the last row of Fig. 12 and Fig. 13, applying this strategy (here with $\epsilon_{\text{out}} = 0.01$) results in a significant reduction of the spike artifacts, with complete removal of the artifacts in the case of rotational noise (Fig. 13).

I. Additional evaluations

In this Section, we provide additional evaluations of our method and of the baseline of BiNI [7] on the DiLiGenT-MV dataset [31], which extends the DiLiGenT dataset for a subset of 5 of its objects (`bear`, `buddha`, `cow`, `pot2`, `reading`) by rendering a total of 20 views per object. The dataset contains both ground-truth normals and normals from photometric stereo, which therefore allows us to quantitatively evaluate the methods also on real normal maps. We run all methods with the same settings as the main experiments, using 1200 iterations, and apply the outlier filtering strategy described in Appendix H for our method, setting $\epsilon_{\text{out}} = 0.1$.

Table 9 reports the mean absolute error (averaged across the 20 object views) against ground-truth depth, which we render with BlenderProc [10] using ground-truth meshes and camera parameters. The results confirm that our method performs better than BiNI also on normals from photometric stereo, with discontinuity estimation further increasing our accuracy.

J. Limitations

Requirement for physically meaningful normals. While effective strategies for the mitigation of outliers can be designed, as described in Appendix H, our method requires that the input normals are physically meaningful, *i.e*., $\boldsymbol{n_a}^\mathsf{T}\boldsymbol{\tau_a} < 0$. As a consequence, an additional preprocessing step on the input normals (*cf*. Appendix H for example strategies) is required in the presence of outliers, to ensure that the above condition is fulfilled.

Non-central camera models. Since it is based on ray direction vectors, our formulation does not allow handling camera models that are non-central, *i.e*., that do not assume all camera rays to originate from a single point (such as axial cameras [30]). A particular case of non-central cameras are orthographic cameras, which assume the center of projection to be at an infinite distance from the scene. As a consequence, in this model all ray direction vectors are parallel to each other and perpendicular to the image plane, *i.e*., $\boldsymbol{\tau_a} = \boldsymbol{\tau_b} = \boldsymbol{\tau_m} = (0,0,1)^\mathsf{T}$ for all a, b, m. We note that in this case our formulation (2) reduces to $z_a = \varepsilon_{b\to a} + z_b$, which, while correct, does not depend on the surface normals and is thus not applicable to normal integration.

Run time and input size. Similarly to previous optimization-based approaches [7, 24, 28], our method is not compatible with real-time deployment, with optimiza-

Method	bear	buddha	cat	cow	harvest	pot1	pot2	reading	goblet
BiNI [7]	$(2.37 \pm 3.15) \times 10^{-5}$	$(3.18 \pm 8.12) \times 10^{-5}$	$(0.35 \pm 2.28) \times 10^{-4}$	$(2.65 \pm 4.32) \times 10^{-5}$	$(0.38 \pm 1.86) \times 10^{-4}$	$(2.89 \pm 6.75) \times 10^{-5}$	$(2.59 \pm 4.07) \times 10^{-5}$	$(0.36 \pm 1.03) \times 10^{-4}$	$(0.32 \pm 1.01) \times 10^{-4}$
Ours	$(0.08 \pm 1.25) \times 10^{-5}$	$(0.09 \pm 1.47) \times 10^{-4}$	$(0.04 \pm 2.48) \times 10^{-4}$	$(0.18 \pm 2.64) \times 10^{-5}$	$(0.18 \pm 1.77) \times 10^{-4}$	$(0.09 \pm 6.52) \times 10^{-4}$	$(0.33 \pm 3.03) \times 10^{-5}$	$(0.78 \pm 8.88) \times 10^{-5}$	$(0.61 \pm 9.10) \times 10^{-5}$

Table 7. **Relative formulation accuracy on the ground-truth log-depth map, DiLiGenT dataset [31]**. For both methods, we report mean and standard deviation across the pixels of the relative residual $|(\tilde{z}_a - \tilde{z}_b - \text{RHS} \;/\; \gamma_{b\to a}) \;/\; \tilde{z}_a|$ computed on the ground-truth log-depth map, where RHS denotes the right-hand side of (8) for BiNI and (11) for Ours. We use $\boldsymbol{\tau_m} = (\boldsymbol{\tau_a} + \boldsymbol{\tau_b})/2$ and $\alpha_{b\to a} = 0$ for Ours.

Method	bear	buddha	cat	cow	harvest	pot1	pot2	reading	goblet
BiNI [7]	$(2.46 \pm 2.39) \times 10^{-4}$	$(2.45 \pm 5.39) \times 10^{-4}$	$(3.30 \pm 6.30) \times 10^{-4}$	$(2.35 \pm 2.89) \times 10^{-4}$	$(0.29 \pm 1.33) \times 10^{-3}$	$(2.17 \pm 4.46) \times 10^{-4}$	$(1.89 \pm 3.01) \times 10^{-4}$	$(3.59 \pm 7.21) \times 10^{-4}$	$(2.37 \pm 7.45) \times 10^{-4}$
Ours	$(0.60 \pm 9.13) \times 10^{-5}$	$(0.06 \pm 1.10) \times 10^{-3}$	$(0.03 \pm 1.87) \times 10^{-3}$	$(0.13 \pm 1.94) \times 10^{-4}$	$(0.13 \pm 1.30) \times 10^{-3}$	$(0.07 \pm 8.46) \times 10^{-3}$	$(0.25 \pm 2.22) \times 10^{-4}$	$(0.57 \pm 6.51) \times 10^{-4}$	$(0.45 \pm 6.66) \times 10^{-4}$

Table 8. **Relative formulation accuracy on the ground-truth depth map, DiLiGenT dataset [31]**. For both methods, we report mean and standard deviation across the pixels of the relative residual $|(z_a - \exp(\text{RHS} \;/\; \gamma_{b\to a}) \cdot z_b) \;/\; z_a|$ computed on the ground-truth depth map, where RHS denotes the right-hand side of (8) for BiNI and (11) for Ours. We use $\boldsymbol{\tau_m} = (\boldsymbol{\tau_a} + \boldsymbol{\tau_b})/2$ and $\alpha_{b\to a} = 0$ for Ours.

Method	bear		buddha		cow		pot2		reading	
	GT	PS	GT	PS	GT	PS	GT	PS	GT	PS
BiNI [7]	0.30	0.45	2.33	1.14	0.26	0.29	0.72	0.90	0.89	1.30
Ours w/o $\alpha_{b\to a}$	**0.24**	0.45	1.89	1.04	0.23	0.29	0.73	**0.83**	0.86	**1.14**
Ours	**0.24**	**0.44**	**1.64**	**1.02**	**0.21**	**0.28**	**0.66**	**0.83**	**0.80**	1.24

Table 9. **Mean absolute depth error (MADE) [mm] on the DiLiGenT-MV dataset [25], averaged across the 20 object views.** GT: ground-truth normals, PS: normals from photometric stereo. All tests use 1200 iterations.

tion converging in a time frame in the order of several seconds (50 to 120 seconds for input normal maps of size 512×612). Additionally, like for previous approaches, our system matrix $\mathbf{A}$ (*cf*. (1) in the main paper), albeit sparse, has both a number of rows and a number of columns that scale linearly with the number of valid pixels in the input normal map. This leads to larger processing time and memory usage for large input sizes, making it currently unsuitable for high-resolution maps and highly complex scenes. More optimized implementations could reduce runtime and memory usage. Investigating more substantial modifications that could move away completely from the drawbacks of optimization-based integration is an interesting direction, but falls outside the scope of this study.

Hyperparameters. Our method depends on a number of hyperparameters, namely the parameters q and ρ of our discontinuity activation term $\beta^{(t)}_{b\to a}$ (*cf*. (16) in the main paper), the parameter k controlling the sharpness of the bilateral weights $w^{\text{BiNI}}_{b\to a}$ (*cf*. (10) in the main paper), and the ray directions $\boldsymbol{\tau_m}$ that control our planarity assumption (*cf*. Sec. 3.1 in the main paper). While the default choices $k = 2$ and $\boldsymbol{\tau_m} = (\boldsymbol{\tau_a} + \boldsymbol{\tau_b})/2$ consistently result in optimal results (*cf*. Tab. 3 and Appendix D), a certain degree of object specificity can be observed in $\beta^{(t)}_{b\to a}$, particularly in its parameter q (*cf*. Appendix E). Therefore, tuning the latter parameter might be desirable to achieve slight improvements in performance.

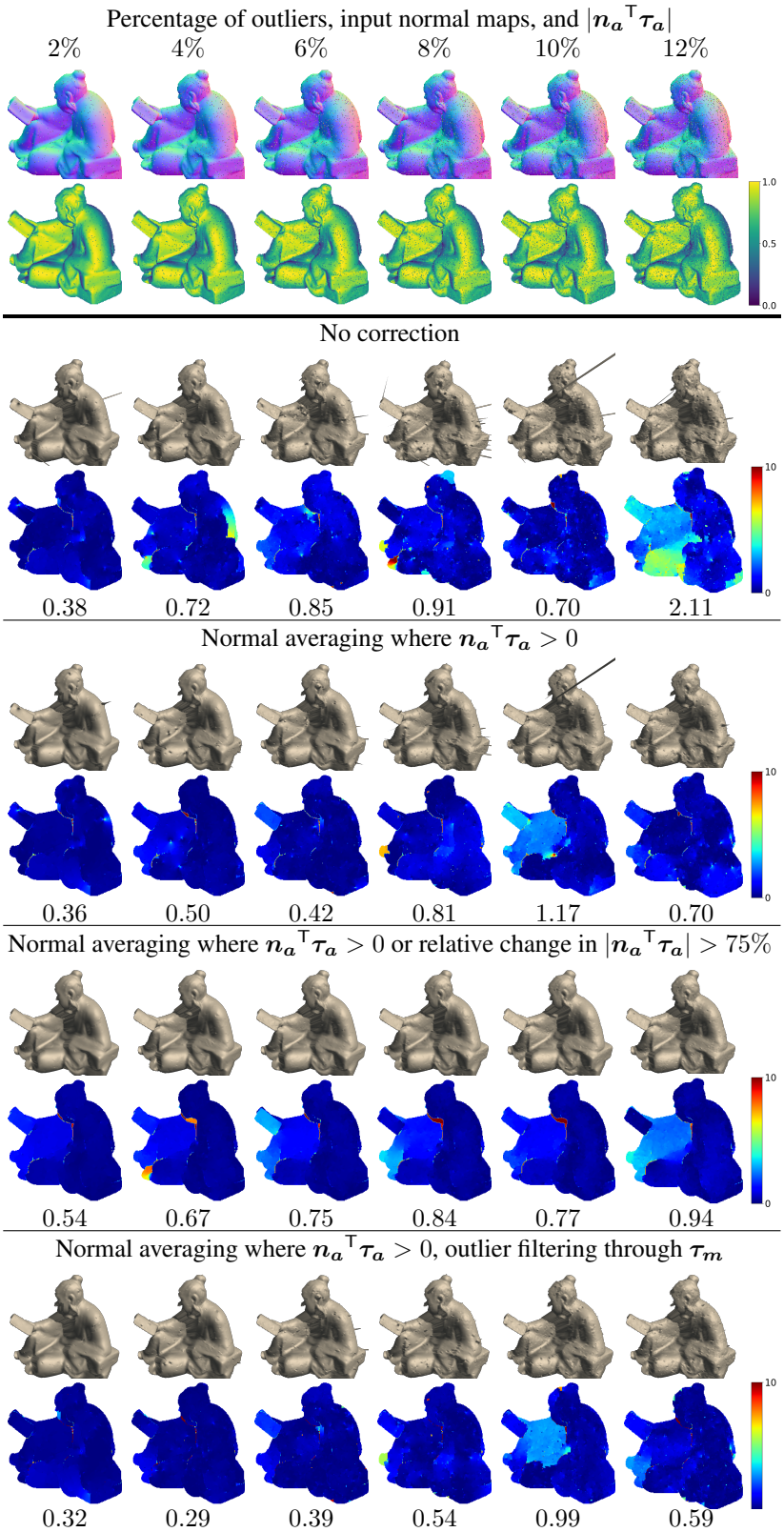

Figure 12. **Ablation on the effect of outliers, object `harvest` from the DiLiGenT [31] dataset.** We introduce increasing amounts of outliers, for which we replace the surface normal with a randomly sampled unit-norm vector. For each variant, we show the reconstructed surface, the corresponding absolute depth error map, and its mean value (MADE, in mm).

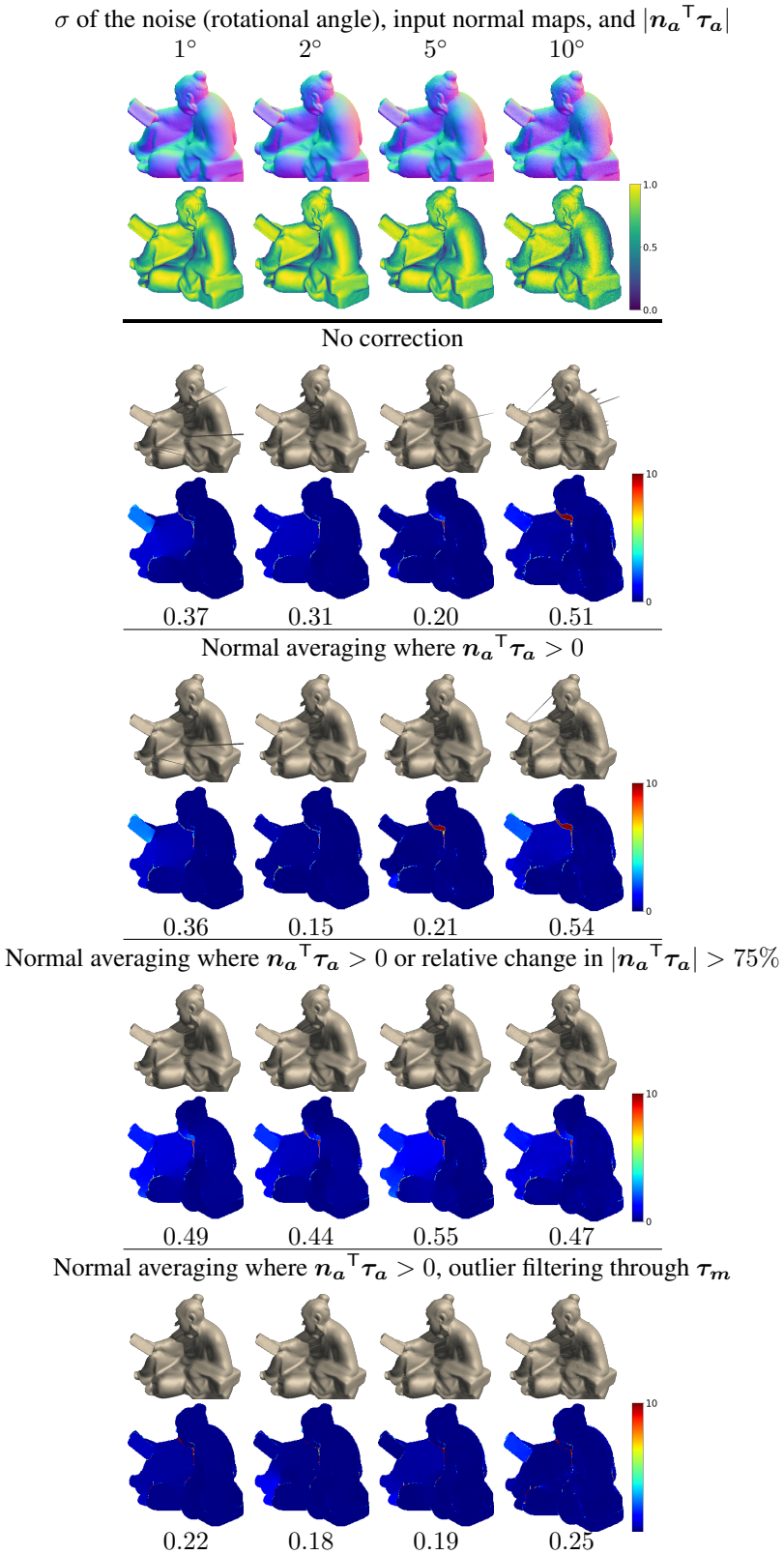

Figure 13. **Ablation on the effect of rotational noise, object `harvest` from the DiLiGenT [31] dataset.** We perturb the surface normals at each pixel, rotating them around randomly sampled axes by angles sampled from Gaussian distributions with increasingly larger standard deviations. For each variant, we show the reconstructed surface, the corresponding absolute depth error map, and its mean value (MADE, in mm).