NPLMV-PS: Neural Point-Light Multi-View Photometric Stereo

Fotios Logothetis
Toshiba Europe
Cambridge UK
fotios.logothetis@toshiba.eu

Ignas Budvytis
Independent researcher
Cambridge UK
ignas.budvytis@gmail.com

Roberto Cipolla
University of Cambridge
Cambridge UK
rc10001@cam.ac.uk

Abstract

This document contains qualitative results on DiLiGenT-MV [5] benchmark in Section 1 and a brief discussion of DiLiGenT-MV data anomalies in Section 2.

1. Qualitative results on sparse DiLiGenT-MV [5]

Qualitative results on *sparse* DiLiGenT-MV [5] benchmark are shown in Figure 1. We include all 3 cases namely [6 views, 6 lights], [6 views, 30 lights] and [20 views, 6 lights]. For us, we include best version (N+I) and compare with SpN [2], and Brahimi et al [1], which was computed with 6 views and 75 lights which is comparable to our 6 views 30 lights case. We observe that our methods achieves consistent low error in most regions of most objects and it is thus the overall best competitor. It is notable, that Brahimi et al [1] does not model cast shadows and thus achives high error in concavities like between the legs of *Bear* and the inside of *Reading*.

2. DiLiGenT-MV data anomalies

This section gives additional information about identified data anomalies in DiLiGenT-MV data. First of all, we report that the rotation matrices for the Reading object do not have determinant 1 (as valid rotation matrices should). For example, for view 1 this is shown in Equation 1:

$$R_1 = \begin{bmatrix} 0.0238 & 1.0031 & -0.0137 \\ 0.4530 & -0.0230 & -0.8912 \\ -0.8912 & 0.0150 & -0.4533 \end{bmatrix} \text{ with } det(R1) = 1.0035 \tag{1}$$

In addition, we note that the intrinsic matrix is different for the Reading object than the rest with the difference being in the x axis focal length as well as the principal point as shown in Equation 3:

$$K_{Reading} = \begin{bmatrix} 3759.1 & 0 & 305.5 \\ 0 & 3759 & 255.5 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$$K_{rest} = \begin{bmatrix} 3772.1 & 0 & 305.875 \\ 0 & 3759 & 255.875 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

As we show in Table 2 of the main submission, the best performance and compatibility with the supplied GT normal maps was achieved with using the Reading intrinsics for all objects, as well as fixing the scaling in rotation matrices with SVD.

In addition, we also note that on the Bear object various images appear to be corrupted as shown in Figure 2. This has been a known issue for the first view (firstly noted by Ikehata in [4]) but we found corrupted images in other views. As most of the images are very dark, this is not easy to notice unless the brightness is adjusted.

Finally, we note that the background seems to be occluding part of the bottom for some objects as shown in Figure 3. This justifies our choise of removing the bottom 6mm of all objects for all methods in evaluating reconstruction accuracy.

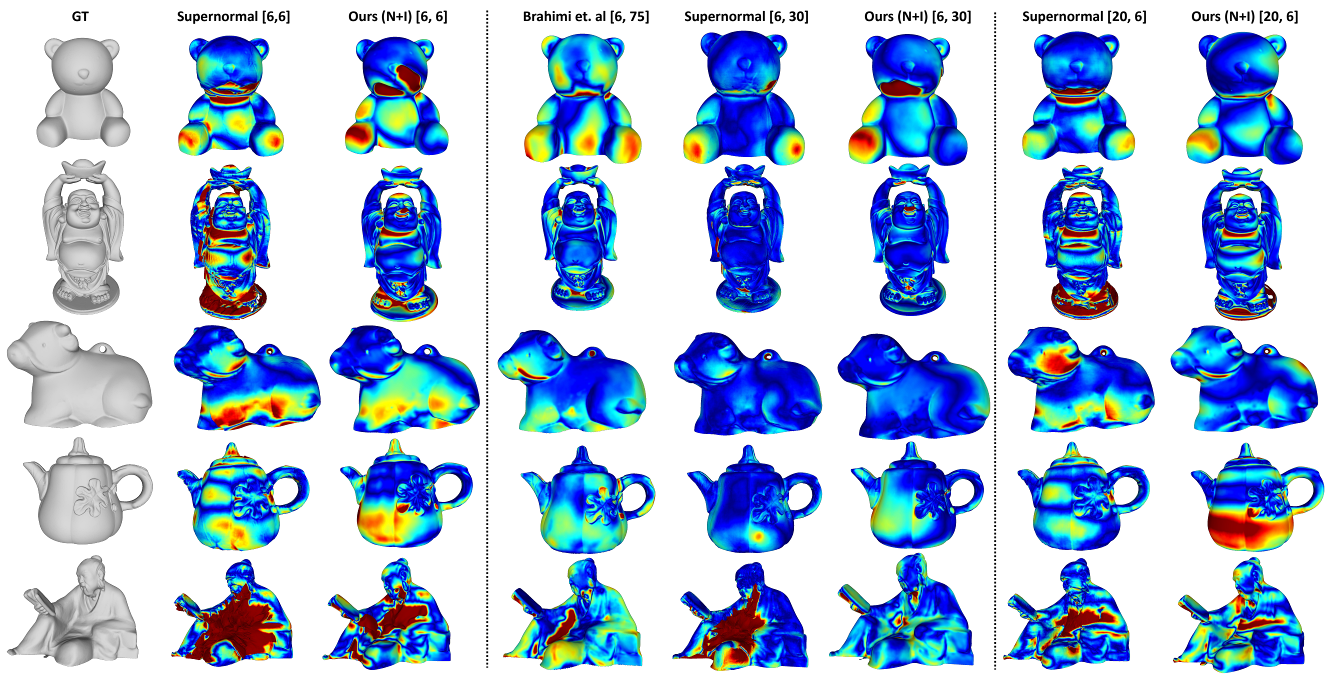

Figure 1. Qualitative results on the sparse version of real DiLiGenT-MV [5] benchmark. The square bracket for each case denote number of [views, lights]. For each mesh vertex, the minimum distance to the GT mesh is shown with the error bars set to red corresponding to 1mm.

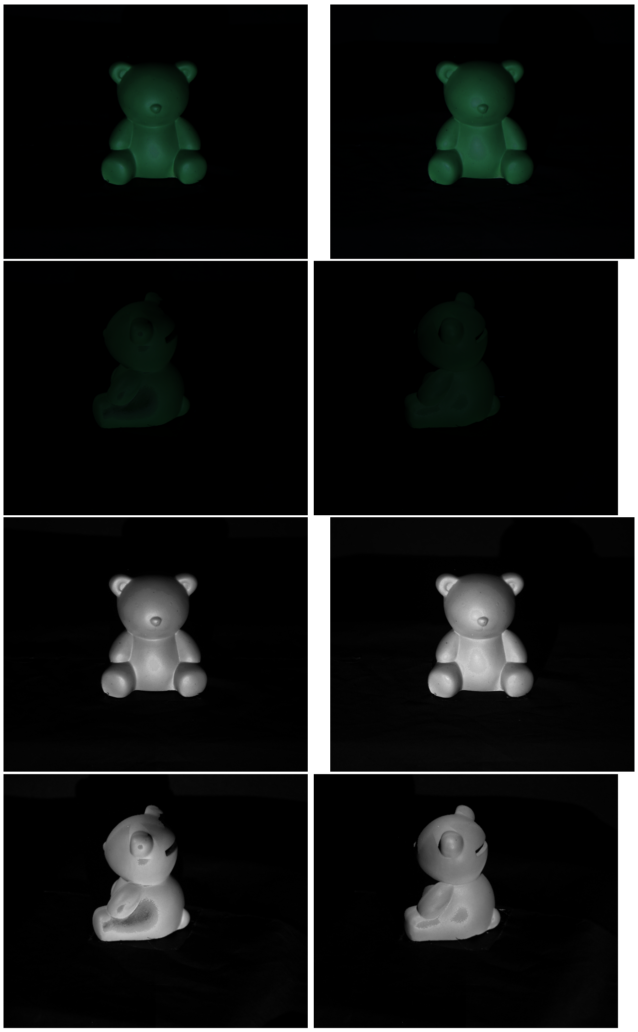

Figure 2. Example of corrupted images on DiLiGenT-MV data. From left to right view 1 lights 1 and 10, view 15 lights 48 and 64 (for the Bear object). Top row contains the original images, bottom row contains brightened up grayscale versions that make visualisation easier. It is clear that there is some data corruption around the specular highlights.

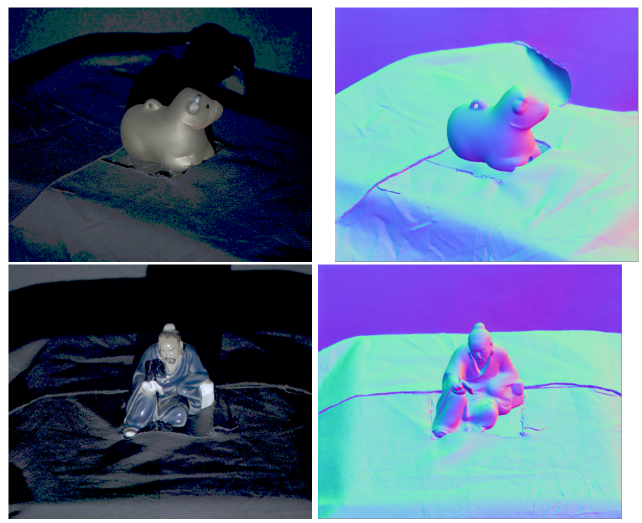

Figure 3. Example of background occluding the bottom part of Cow (left) and Reading (right) objects. We show brightened average RGB image as well as full image normal maps (computed with Uni MS-PS [3]) to better visualise this issue.

References

[1] Mohammed Brahimi, Bjoern Haefner, Zhenzhang Ye, Bastian Goldluecke, and Daniel Cremers. Sparse views near light: A practical paradigm for uncalibrated point-light photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11862–11872, June 2024. 1

[2] Xu Cao and Takafumi Taketomi. Supernormal: Neural surface reconstruction via multi-view normal integration. *CVPR*, 2024. 1

[3] Clément Hardy, Yvain Quéau, and David Tschumperlé. Uni MS-PS: a Multi-Scale Encoder Decoder Transformer for Universal Photometric Stereo. working paper or preprint, Feb. 2024. 3

[4] S. Ikehata. Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces. In *ECCV*, 2018. 1

[5] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE Trans. Image Process.*, 2020. 1, 2