A Dark Flash Normal Camera: Supplemental Materials

Zhihao Xia[1*] Jason Lawrence[2] Supreeth Achar[2]
[1]Washington University in St. Louis [2]Google Research

A. Visible Lighting Conditions Used in Our Evaluation

We evaluate our method using five different visible lighting conditions that range from favorable ("well lit") to challenging ("shadows", "mixed colors", "overexposure", and "low light"). Figure 9 illustrates and details how these different lighting conditions are simulated from the RGB OLAT training images.

(a) Well lit (b) Shadows (c) Mixed colors (d) Overexposure (e) Low light

Figure 9: The five visible lighting conditions used in our evaluations. (a) A well lit image is created by averaging equally the four OLAT RGB images, using weights that avoid any under- or over-saturated pixels. (b) Picking a single OLAT image at random produces an RGB input with strong cast shadows. (c) We simulate mixing different colored lights by picking two OLAT images at random and remapping the color of each to a random color temperature in the range $[1900K, 2900K]$ and $[7000K, 20000K]$, respectively, and then averaging them together. (d) Scaling the intensities of one of the OLAT images with a random scale factor in the range $[1.8, 2.3]$ and then clipping the result yields an image with harsh lighting and saturated intensities. (e) Adding Gaussian white noise to each pixel ($\sigma = 25$ 8-bit gray levels) simulates images captured under low light conditions.

B. Additional Ablation Studies

The graphs in Figure 10 expand on the results shown in Figure 6 in the paper. They plot the mean angular error of the estimated normals against the baseline for three of our five lighting scenarios, and for three different techniques: our network modified to take only a single NIR input image ("NIR Only"), our network modified to take only a single RGB input image ("RGB Only"), and our full network, which considers both. In each graph the magnitude of the lighting issue increases from left to right.

Variations in illuminant color have a relatively small effect on the normals estimated by the "RGB only" network, but overexposed pixels and image noise cause large errors. As one would expect the "NIR only" network is invariant to these lighting changes (flat line in these graphs), and our proposed approach consistently outperforms both.

C. Comparisons to Prior Single (RGB) Image Normal Estimation Methods

We compare our method to two recent RGB-only techniques:

*Work done while Zhihao Xia was an intern at Google.

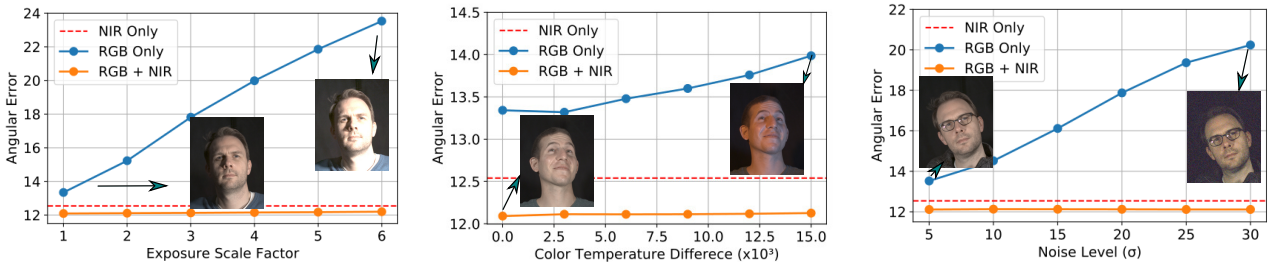


Figure 10: Mean absolute angular errors in degrees of normal maps estimated by three different techniques in three challenging lighting conditions as the magnitude of each lighting issue increases from left to right. A similar network that uses only a single RGB image to estimate normals (RGB Only) and one that uses only a single NIR image (NIR Only) both perform worse that our method (RGB + NIR), with the RGB Only network deteriorating rapidly alongside the lighting issue. *Left:* harsh lighting that produces cast shadows with increasing image exposures. *Center:* mixing lights with different color temperatures. *Right:* increasing levels of noise, which occurs in low-light conditions.

SfSNet [3]. For a fair comparison, we retrain SfSNet on our dataset. We use the same basic network architecture as our method, but with only a single RGB image as input and with an additional lighting estimation branch. This lighting estimation branch takes the output of the encoder network and generates an estimate of the lighting in the input RGB image. SFSNet uses second order spherical harmonics to represent the scene lighting, which isn't well suited for approximating the type of point lighting in our dataset. Therefore we instead represent the lighting as a blending weight vector over the four visible OLATs. The lighting estimation branch is supervised using the ground truth OLAT mixing weights of the input RGB image. The normal branch of SfSNet is supervised using the same stereo normals and same image reconstruction loss that we use for our method.

Directional Face Relighting Network [2]. We also compare our method to the intrinsic component estimation stage of Nestmeyer et al. [2], which uses a UNet to predict a normal and albedo map from a single RGB image. Since we assume that the lighting conditions in the input RGB image are unknown, we do not provide the source lighting direction to the network. We retrain their network on our dataset by supervising the normal estimation network path with the same stereo normals used to train our method, and by using one of the RGB OLAT images chosen at random as the relighting target image. Note that their instrinsic component estimation stage does not consider cast shadows.

Figure 11 shows examples of the outputs generated by all three techniques for different lighting conditions. Our method outperforms both techniques even in the well lit condition, which we attribute to our novel training strategy that combines shape information from complementary stereo and photometric signals, and the additional information provided by the NIR input. In challenging lighting conditions, the benefit of our method becomes more significant. Table 1 in the main paper shows mean absolute errors for these two RGB-only techniques and our proposed method.

D. Expanded Stereo Refinement Results

Figure 12 expands on Figure 7 in the paper. Specifically it shows the input images and the normal map estimated by our method, along with the raw stereo depths prior to any smoothing or refinement.

E. Expanded Lighting Adjustment Results

Figure 13 expands on Figure 8 in the paper. Specifically it shows the supplemental rendered image that is combined with the input RGB image in order to brighten shadowed regions along the face. We also compare the difference between using a strictly Lambertian image formation model and our full Lambertian-plus-specular model in generating the supplemental image that is combined with the input. Note that our full model does a better job at reproducing specular highlights along the cheek and tip of the nose.

F. Additional Video Results

Please see our project page darkflashnormalpaper.github.io for results and comparisons on image sequences.

Figure 11: Estimated surface normals using three different techniques under five lighting conditions. Existing RGB-only methods perform poorly under non-ideal lighting conditions. Since our method uses a well-lit NIR image in addition to the RGB input, it is able to generate good normal estimates even when lighting conditions degrade.

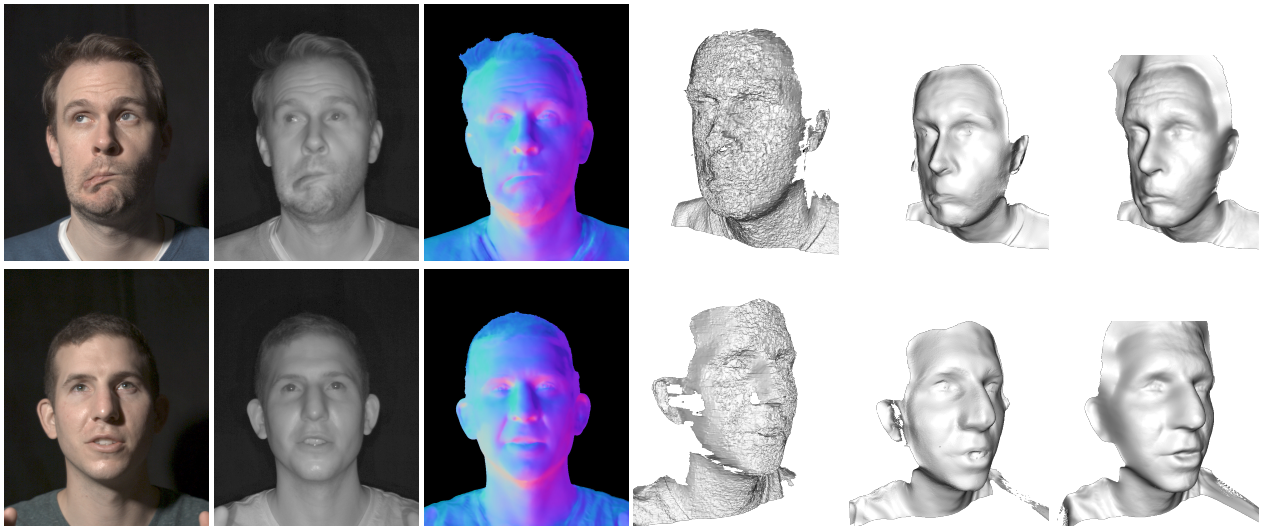
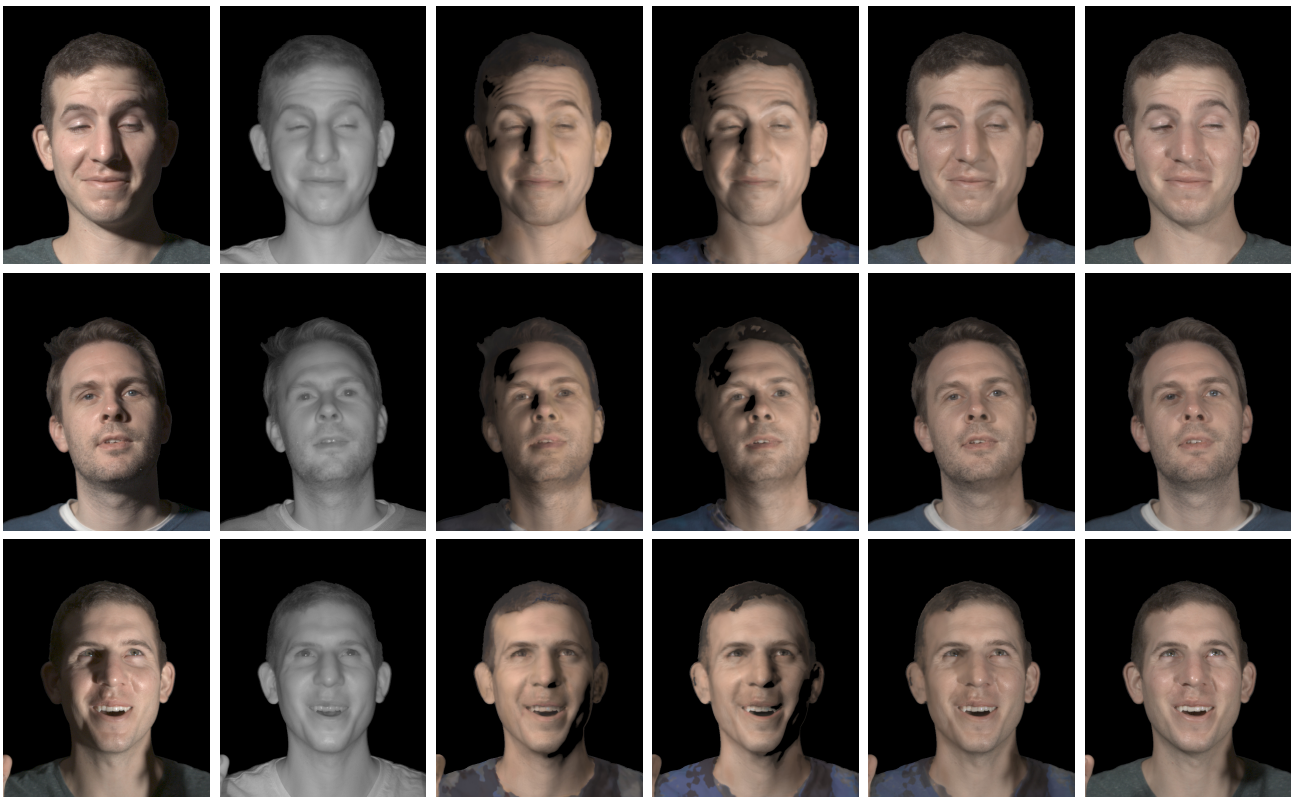


Figure 12: Stereo methods often struggle to recover fine-scale surface details. Applying a guided bilateral filter to raw stereo depths yields a smoother surface but with distorted features (e.g. the nose appears pinched and reduced). We use the method of Nehab et al. [1] to compute a refined surface according to normals estimated with our method. Note how this better preserves details around the eyes, nose, and mouth, along with fine wrinkles and creases.



RGB Input | NIR Input | Virtual Light (w/o Blinn-Phong) | Virtual Light (w/ Blinn-Phong) | Relit | Ground Truth

Figure 13: Our method can be used to simulate adding lights to a scene to fill in shadows. We show this virtual light image generated using a Lambertian reflectance model alone and using our full Lambertian + Blinn-Phong model, which produces more realistic highlights. When combined with the input RGB image (Relit) this approach compares favorably to ground truth.

References

[1] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Transactions on Graphics (TOG)*, 24(3):536–543, 2005. 4

[2] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas M Lehrmann. Learning physics-guided face relighting under directional light. In *Proc. CVPR*, 2020. 2

[3] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D. Castillo, and David W. Jacobs. Sfsnet: Learning shape, refectance and illuminance of faces in the wild. In *Proc. CVPR*, 2018. 2