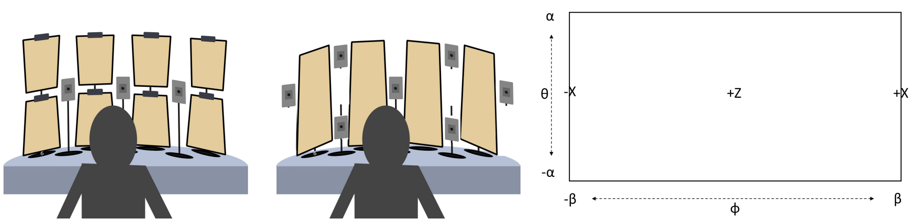
Supplemental Materials: Practical and Scalable Desktop-based High-Quality Facial Capture

Alexandros Lattas[2], Yiming Lin[1], Jayanth Kannan[1], Ekin Ozturk[1,2], Luca Filipi[1], Giuseppe Claudio Guarnera[1,3], Gaurav Chawla[1], and Abhijeet Ghosh[1,2]

[1] Lumirithmic Ltd, London, UK
{yiming.lin,jay,luca,gaurav.chawla}@lumirithmic.com
[2] Imperial College London, London, UK
{a.lattas,ekin.ozturk17,ghosh}@imperial.ac.uk
[3] University of York, York, UK claudio.guarnera@york.ac.uk

1 Desktop Capture Setups



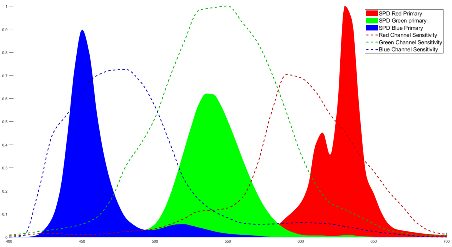
(a) Tablet-based setup (b) Monitor-based setup (c) Illumination Coverage

Fig. 1: Schematics of the two proposed desktop-based facial capture systems. (a): Tablet-based setup consisting purely of mobile devices. (b): Monitor-based setup consisting of four 27" LCD monitors mounted in portrait mode. (c): Frontal hemispherical zone (latitude-longitude parameterization) of directions covered by the display illumination. The zone spans $\beta < \pm 90^\circ$ along the longitudinal direction, while spanning $\alpha < \pm 45^\circ$ along the latitudinal direction.

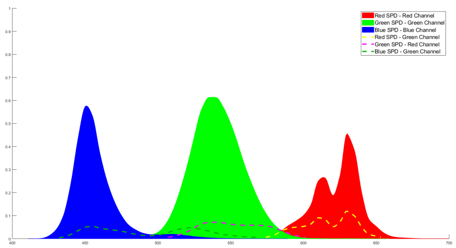
In Fig. 1, we present schematics of the two proposed desktop capture systems – the first one employing portable mobile devices (tablets and smartphones), and the second one employing larger LCD monitors for piece-wise continuous illumination on a subject's face. The displays are arranged to illuminate a subject's face from a frontal hemispherical zone of directions (Fig. 1), spanning $\beta < \pm 90^\circ$ along the longitudinal direction, while spanning $\alpha < \pm 45^\circ$ along the latitudinal direction.

2 Spectral Cross-Talk

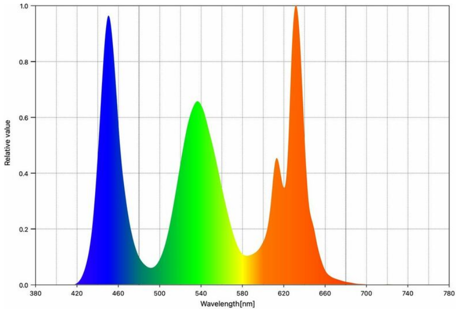
The spectral power distribution (SPD) of the DCI P3 gamut color LCD panel displays that we employ in both the proposed setups is shown in Fig. 2. The color display SPD has three clear peaks corresponding to blue, green and red, although the red has a smaller amber peak as well. Fig. 2 plots the spectral sensitivity of one of the color cameras (iPhone 12 Pro) we employ for measurements overlaid on the SPD of the display illumination. We recover the camera sensitivity using measurements of a color chart under known display illumination spectrum using the procedure described in [3]. Fig. 2 also plots the magnitude of cross-talk in terms of a camera color channel being sensitive to illumination from a neighboring color channel. As can be seen, for our given pairs of display illumination SPD and camera sensitivity, there is significant spectral cross-talk between the green and red color channels.



(a) Display SPD and camera sensitivity



(b) Display SPD and spectral cross-talk



(c) Display SPD of DCI P3

Fig. 2: (a) SPD of display illumination overlayed with the spectral sensitivity of one of the color cameras (iPhone 12 Pro) employed for acquisition. (b) SPD of display illumination overlayed with spectral cross-talk in camera due to neighboring wavelength of spectral illuminations showing significant cross-talk between the green and red color channels. (c) Spectral power distribution (SPD) of DCI P3 gamut color LCD displays employed for both desktop capture setups.

3 Additional Results

(a) Photo (b) Diff. Alb. (c) Sp. Albedo (d) Sp. Norm. (e) Roughness

Fig. 3: Side-view high-resolution results of two subjects. All results were acquired using our 2-shot spectral multiplexing method and the tablet-based setup.

(a) Unpolarized (b) Cross Polarized (c) Diff. Albedo, Ours

Fig. 4: Comparison between (a) an unpolarized full-on capture of a subject, (b) a diffuse albedo acquired by employing a cross polarized filter on the camera [2] and (c) a diffuse albedo acquired by our method.

In Fig. 3, we present additional results in high-resolution. Specifically, we show side-views of the captured photo, diffuse albedo, specular albedo, diffuse normals, specular normals and specular roughness, from two subjects from the main paper. These are acquired with our 2-shot method using the tablet-based setup.

Moreover, in Fig. 4 we present a high-resolution comparison of our method and Ghosh et al. [2]. We acquire a diffuse albedo by fitting a cross-polarized filter on our cameras, and present it side-to-side with the mixed albedo and our diffuse albedo. We calculate a mean squared error of 0.0028 and a Peak Signal-to-Noise Ratio (PSNR) of 25.41. Our method is able to successfully to separate the diffuse signal, without having to rely on extra hardware, such as a polarized filter.

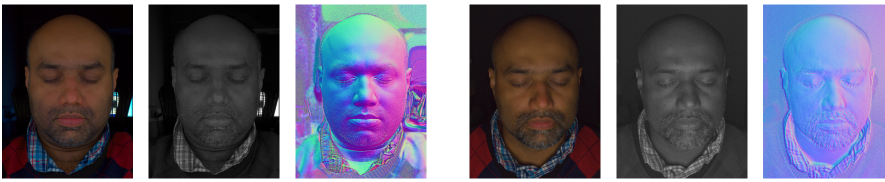
4 Reconstruction Pipeline

The reconstruction pipeline employed in this work follows a standard multi-view reconstruction approach, which we describe below. We use a full-on illuminated image per camera for the reconstruction, which can be acquired by adding a) a white binary pattern and its complement pattern (e.g. H and H') or b) the two color patterns, which also add to full white illumination. We acquire a base geometry using the COLMAP structure-for-motion implementation [7][6]. The mesh is then cropped using an elliptical bounding box and smoothed with a Laplacian filter. Then, the per-camera reflectance maps are projected on the mesh using the camera parameters acquired from COLMAP. A template mesh with a predefined UV coordinate parameterization can also be fit to the reconstructed mesh, using NICP [1]. Finally, our acquired normal maps can be employed to add additional geometric details to the mesh [5].

5 Additional Comparisons

We include additional comparative experiments that were presented in the main manuscript, at higher resolution and further discussion. Fig. 5 shows a comparison of our results with the 2-shot tablet-based capturing system and method, and the same method implemented on a display and camera, similar to the apparatus of [8]. More specifically, we mount a single DSLR camera on the center of the top of a display, and partition the single display to show our patterns. Our method is still able to perform diffuse-specular separation and normal estimation, however the face is not uniformly illuminated, as seen in the diffuse albedo $\mathbf{A}_D$ and specular albedo $\mathbf{A}_S$, while the calculated normals $\mathbf{N}_S$ are biased towards the center of the face.

Moreover, we include a comparison with a recent single-image deep-learning method for facial reflectance acquisition (AvatarMe++ [4]), which is fed the fully illuminated image from our capture. As shown in Fig. 6, our 2-shot method is able to reconstruct the accurate shape and reflectance properties of the subject.



(a) $\mathbf{A}_D$ (b) $\mathbf{A}_S$ (c) $\mathbf{N}_S$ (d) $\mathbf{A}_D$ (e) $\mathbf{A}_S$ (f) $\mathbf{N}_S$

Fig. 5: Comparison of diffuse albedo $\mathbf{A}_D$, specular albedo $\mathbf{A}_S$, and specular normals $\mathbf{N}_S$ separation, between our 2-shot method (left), versus our method implemented on a single screen and camera (right).

(a) Input (b) Ours, 2-shot (c) AvatarMe++ [4]

Fig. 6: Our results compared with AvatarMe^{++}, a deep-learning method [4]. From left to right: fully-illuminated image from our capture set and input to AvatarMe^{++}, rendering and geometry of our method, rendering and geometry of AvatarMe^{++}.

References

1. Amberg, B., Romdhani, S., Vetter, T.: Optimal step nonrigid icp algorithms for surface registration. In: 2007 IEEE conference on computer vision and pattern recognition. pp. 1–8. IEEE (2007)
2. Ghosh, A., Fyffe, G., Tunwattanapong, B., Busch, J., Yu, X., Debevec, P.: Multiview face capture using polarized spherical gradient illumination. ACM TOG **30**(6) (2011)
3. Jiang, J., Liu, D., Gu, J., Süsstrunk, S.: What is the space of spectral sensitivity functions for digital color cameras? In: 2013 IEEE Workshop on Applications of Computer Vision (WACV). pp. 168–179. IEEE (2013)
4. Lattas, A., Moschoglou, S., Ploumpis, S., Gecer, B., Ghosh, A., Zafeiriou, S.P.: Avatarme++: Facial shape and BRDF inference with rendering-aware GANs. TPAMI (2021)
5. Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3d geometry. ACM TOG (Proc. SIGGRAPH) (3), 536–543 (2005)
6. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
7. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV) (2016)
8. Sengupta, S., Curless, B., Kemelmacher-Shlizerman, I., Seitz, S.M.: A light stage on every desk. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2420–2429 (2021)