Desktop-based High-Quality Facial Capture for Everyone

Alexandros Lattas
a.lattas@imperial.ac.uk
Imperial College London
London, UK

Yiming Lin
yiming.lin@lumirithmic.com
Lumirithmic
London, UK

Jayanth Kannan
jay@lumirithmic.com
Lumirithmic
London, UK

Luca Filipi
luca@lumirithmic.com
Lumirithmic
London, UK

Ekin Ozturk
ekin.ozturk17@imperial.ac.uk
Imperial College London,
Lumirithmic
London, UK

Claudio Guarnera
claudio.guarnera@york.ac.uk
University of York,
Lumirithmic
York, UK

Gaurav Chawla
gaurav@lumirithmic.com
Lumirithmic
London, UK

Abhijeet Ghosh
ghosh@imperial.ac.uk
Imperial College London,
Lumirithmic
London, UK

(a) Tablet-based Setup (b) Rendering (c) Monitor-based Setup (d) Rendering

Figure 1: Two novel desktop-based setups (a, c) for high-quality facial capture (b, d). (a) Setup consisting of a set of portable mobile devices for static facial capture. (c) Setup consisting of a set LCD displays for static and dynamic facial capture.

ACM Reference Format:
Alexandros Lattas, Yiming Lin, Jayanth Kannan, Luca Filipi, Ekin Ozturk, Claudio Guarnera, Gaurav Chawla, and Abhijeet Ghosh. 2022. Desktop-based High-Quality Facial Capture for Everyone . In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '22 Talks), August 07-11, 2022.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532836.3536258

1 INTRODUCTION

Realistically rendered human faces have wide ranging applications in computer graphics, in entertainment, advertising, and virtual presence in AR/VR, and in the envisioned *metaverse*. Realistic modeling of facial shape and appearance has been revolutionized with the development of acquisition techniques for high-quality 3D facial capture. However, realistic facial appearance capture typically requires use of custom designed and complex apparatus such as the Lightstage [Ghosh et al. 2011; Kampouris et al. 2018].

Instead, we present two novel desktop-based systems for high-quality facial capture including geometry and facial appearance. The proposed acquisition systems are highly practical, consisting purely of commodity components, enabling a significant reduction in cost, along with increased portability and scalability compared to alternative capturing systems. We present a novel set of illumination patterns for efficient acquisition of reflectance using our setup, with diffuse-specular separation. Moreover, we utilize incomplete hemispherical illumination to compute accurate photometric normals. We demonstrate high-quality results with two capture setups – one entirely consisting of portable mobile devices targeting static facial capture, and the other consisting of desktop LCD displays targeting both static and dynamic facial capture (see Fig. 1).

SIGGRAPH '22 Talks, August 07-11, 2022, Vancouver, BC, Canada

ACM ISBN 978-1-4503-9371-3/22/08.
https://doi.org/10.1145/3532836.3536258

2 DESKTOP-BASED CAPTURE SYSTEM

2.1 Tablet-based Setup

Our first setup consists purely of a set of mobile devices – eight tablets and five smartphones, that are mounted on a desk as shown in Fig. 1 (a). The tablets (iPad Air 4th generation) are arranged in two rows (latitudes) of four devices and oriented longitudinally so as to cover a significant zone of the frontal hemisphere around a subject's face. The screens of the tablets are oriented towards the subject and provide controlled piece-wise continuous illumination. We also mount five smartphones (iPhone 12 Pro) in the setup along the equatorial plane and employ their high-resolution back cameras (zoom lens) for acquiring facial reflectance and photometric normals. The devices are all controlled in synchronization during capture process where one device acts as the master and wirelessly communicates capture command and timings to other devices.

2.2 Monitor-based Setup

Our second setup consists of four desktop LCD monitors (27" Asus ProArt PA279CV) that we mount together on a desk in portrait mode as shown in Fig. 1 (c). The monitors are arranged longitudinally,

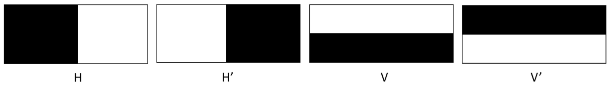

Figure 2: Horizontal and vertical binary patterns and their complements employed with form-factor modulation.

facing the subject to cover a similar frontal hemispherical zone of directions with screen illumination. The workstation includes a high-end graphics card (Nvidia RTX 3070 Ti) which is employed to drive the four monitors via HDMI and display ports. We additionally mount eight digital cameras with small form-factor (Canon mirrorless EOS M200) within the vertical gaps between the monitors for multiview capture. During acquisition, the workstation controls the monitors and the set of cameras in synchronization to rapidly capture a sequence of images under a set of controlled illumination patterns for 3D shape and reflectance capture. This setup achieves fine-grained synchronization between multiple displays, even supporting video rate dynamic facial appearance capture.

2.3 Modulated Binary Illumination

We employ horizontally and vertically aligned binary illumination patterns over the hemispherical zone of illumination of the proposed capture setups for acquiring albedo and photometric normals with diffuse-specular separation. While this is similar to the approach of Kampouris et al. [2018], our patterns require additional form-factor modulation due to being near-field with limited angular extent. Furthermore, we make the crucial observation that over the zone of hemisphere covered by the displays in our setup, we do *not* need all three axis aligned (X, Y, Z) binary illumination conditions and their complements employed by [Kampouris et al. 2018]. Specifically, when we illuminate a subject with the horizontally aligned binary pattern and its complement (H and H' respectively), the lit portion of the zone does not have its centroid aligned with X-axis but is actually centered around $+45°$ and $-45°$ directions respectively in the XZ plane (with $0°$ corresponding to the +Z axis). This results in the complementary pair of horizontal binary patterns H and H' being sufficient for determining the x and z components of a photometric normal. We additionally only need measurement of the vertically aligned binary pattern V and its complement V' to determine the y component of the photometric normal. This reduces the number of measurements to only four photographs under form-factor modulated horizontal and vertical binary patterns (see Fig. 2). We further reduce the measurements to just *two photographs* using spectral multiplexing of these patterns into R, G, and B color channels of display illumination.

2.4 Photometric Processing

2.4.1 Photometric Normals. The set of acquired binary patterns can be used to compute photometric normals as described above. However, due to incomplete hemispherical illumination (limited to $< \pm 45°$ latitudinal zone) used in our setups, this would result in skewed normals. Instead, we employ Singular Value Decomposition (SVD) to extract the normals with more correct global orientation.

2.4.2 Diffuse-Specular Separation. For separating the reflectance albedo into diffuse $\mathbf{A_D}$ and specular $\mathbf{A_S}$ components, we employ a linear system (Eq. 1), similar to that proposed by [Kampouris et al. 2018] for binary spherical gradient illumination. To do this, we rotate the H, H' and V, V' binary patterns to be axis-aligned along X and Y respectively. Assuming the binary observation along X is brighter than its complement X', the linear system is expressed as:

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{X'} \end{bmatrix} = \begin{bmatrix} \mathbf{N.x} & 1 \\ (1-\mathbf{N.x}) & 0 \end{bmatrix} \begin{bmatrix} A_D \\ A_S \end{bmatrix}, \tag{1}$$

where $\mathbf{N.x}$ is the x component of the photometric normal scaled to $[0, 1]$. We average the two estimates for $\mathbf{A_S}$ obtained from the X and Y binary pairs respectively to obtain the final estimate of specular albedo. This in turn also determines the diffuse albedo. Fig. 3 presents results of separated diffuse and specular albedo and photometric normals for a subject shown in Fig. 1.

(a) Diff. Alb. (b) Spec. Alb. (c) Diff. Norm. (d) Spec. Norm. (e) Photo (f) Render. (2D)

Figure 3: Photometric acquisition using our binary patterns.

2.5 Dynamic Capture

The reflectance capture approach with color-multiplexed binary patterns (2-shots) is well-suited for dynamic capture. As proof-of-concept, we employ our monitor-based setup for dynamic capture and synchronize a machine vision camera (FLIR Grasshopper 3) with the refresh rate of the monitors. We capture alternating color-multiplexed patterns at 60 fps. This provides an effective capture rate of 30 fps for the alternating patterns which we employ for acquisition of diffuse-specular separated albedo and photomeric normals of a dynamic facial performance (see Fig. 4). Slight facial motion can be accommodated with optical flow alignment.

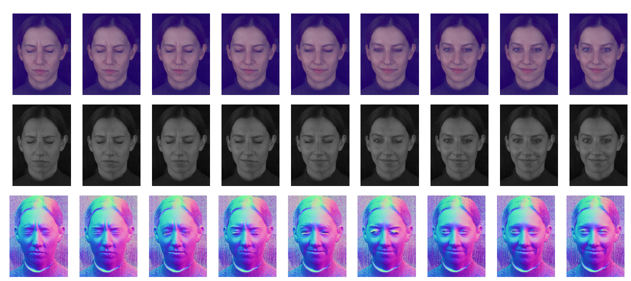

Figure 4: Dynamic capture using our monitor-based setup, acquired at 60 FPS for results at 30 FPS. From top-to-bottom: diffuse albedo, specular albedo, surface normals.

3 CONCLUSION

We presented two novel desktop-based setups for high quality facial appearance capture . The setups are highly practical and scalable, consisting entirely of commodity components, and can make high-quality static and dynamic facial capture widely accessible.

REFERENCES

Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM TOG* 30, 6 (2011).

Christos Kampouris, Stefanos Zafeiriou, and Abhijeet Ghosh. 2018. Diffuse-Specular Separation using Binary Spherical Gradient Illumination. In *EGSR (EI&I)*. 1–10.