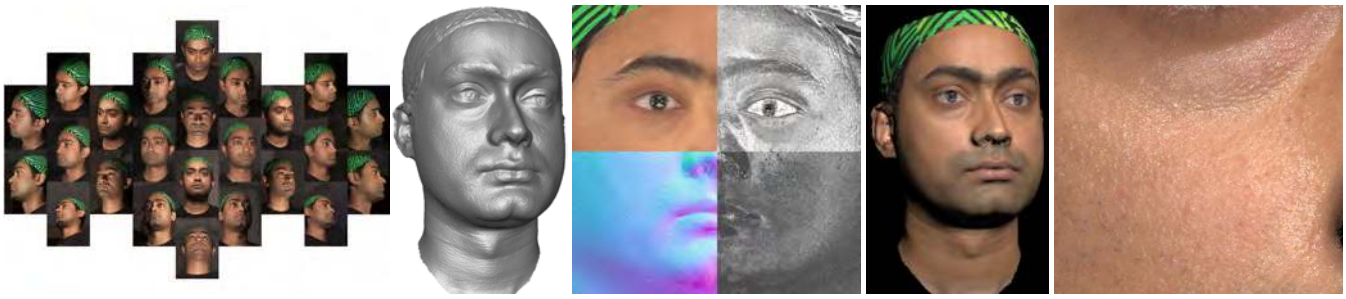
EUROGRAPHICS 2016 / J. Jorge and M. Lin
(Guest Editors)

Volume 35 (2016), Number 2

Near-Instant Capture of High-Resolution Facial Geometry and Reflectance

G. Fyffe[1] P. Graham[1], B. Tunwattanapong[1], A. Ghosh[2] and P. Debevec[1]

[1]USC Institute for Creative Technologies, USA
[2]Imperial College London, UK



(a) (b) (c) (d) (e)

Figure 1: *(a) Multi-view images shot under rapidly varying flash directions. (b) Refined geometry. (c) Clockwise from top left: diffuse albedo, specular albedo, specular exponent × 0.02, surface normal map. (d) Rendering. (e) Zoom of cheek rendering.*

Abstract

We present a near-instant method for acquiring facial geometry and reflectance using a set of commodity DSLR cameras and flashes. Our setup consists of twenty-four cameras and six flashes which are fired in rapid succession with subsets of the cameras. Each camera records only a single photograph and the total capture time is less than the 67ms blink reflex. The cameras and flashes are specially arranged to produce an even distribution of specular highlights on the face. We employ this set of acquired images to estimate diffuse color, specular intensity, specular exponent, and surface orientation at each point on the face. We further refine the facial base geometry obtained from multi-view stereo using estimated diffuse and specular photometric information. This allows final submillimeter surface mesostructure detail to be obtained via shape-from-specularity. The final system uses commodity components and produces models suitable for authoring high-quality digital human characters.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and scanning I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

1. Introduction

Modeling realistic human characters is frequently done using 3D recordings of the shape and appearance of real people across a set of facial expressions [PHL*98, ARL*10] to build blendshape facial models. To cross the "Uncanny Valley", faces require high-quality geometry, texture maps, reflectance properties, and surface detail at the level of skin pores and fine wrinkles. Unfortunately, there has not yet been a technique for recording such datasets that is near-instantaneous and relatively low-cost. While some facial capture techniques are instantaneous and inexpensive [BBB*10, BHPS10], these do not generally provide lighting-independent texture maps, specular reflectance information, or high-resolution surface normal detail for relighting. In contrast, techniques using multiple photographs and spherical lighting setups [WMP*06, GFT*11] do capture such reflectance properties, but this comes at the expense of longer capture times and complicated custom equipment.

In this paper, we present a facial capture technique that combines benefits of single-shot techniques (simple setup of low-cost cameras where each camera takes precisely one photo) and multi-shot techniques (specular reflectance information and high-resolution photometric normals). The inventive step is firing the cameras at nearly the same instant, but slightly offset, to image several dif-

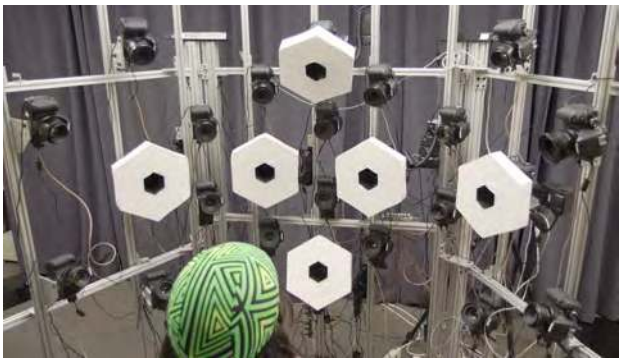


Figure 2: *Facial capture setup, consisting of 24 entry-level DSLR cameras and six diffused ring flashes, all one meter from the face. A set of images taken with this arrangement can be seen in Fig. 1.*

ferent lighting conditions before the subject blinks, rather than a single lighting condition. This novel framework likely has applications beyond our specific setup. We use a 24-camera entry-level DSLR photogrammetry setup similar to common commercial systems[†] and use six ring flashes to light the face. However, instead of firing all the flashes and cameras at once, each flash is fired sequentially with a subset of the cameras, with the exposures packed milliseconds apart for a total capture time of 66ms, which is faster than the blink reflex [BBL67]. This arrangement produces 24 independent specular reflection angles evenly distributed across the face, allowing a shape-from-specularity approach to obtain high-frequency surface detail. But, unlike other shape-from-specularity techniques, our images are not taken from the same viewpoint. Hence, we compute an initial estimate of the facial geometry using passive stereo, and then refine the geometry using specular photometric detail. The resulting system produces accurate, high-resolution facial geometry and reflectance with near-instant capture in a relatively low-cost setup. The principal contributions of this work are:

1. A near-instantaneous photometric capture setup for measuring the geometry and diffuse and specular reflectance of faces.
2. A camera-flash arrangement pattern which produces evenly-distributed specular reflections over the face with a single photo per camera and fewer lighting conditions than cameras.
3. A novel per-pixel separation of diffuse and specular reflectance using multi-view color-space analysis and novel photometric estimation of specular surface normals for geometry refinement.

2. Related work

Passive Multi-View Stereo Our facial base geometry reconstruction step is similar to the work of Furukawa and Ponce [FP09] who proposed multi-view stereopsis as a match-expand-filter procedure that produces dense patch reconstruction from an initial set of sparse correspondences. However, since subsurface scattering typically blurs surface detail [RR08] for semi-translucent materials such as skin, the resolution which can be recovered for faces is limited. Passive multi-view stereo has been employed by Beeler et al. [BBB*10] and Bradley et al. [BHPS10] to reconstruct high quality facial geometry under diffuse illumination. Beeler et al. apply mesoscopic augmentation as in [LZ94, GWM*08] to hallucinate detailed geometry, which, while not metrically accurate, increases the perceived realism of the models by adding the appearance of skin detail. Valgaerts et al. [VWB*12] present a passive facial capture system which achieves high quality facial geometry reconstruction under arbitrary uncontrolled illumination. They reconstruct base geometry from stereo correspondence and incorporate high frequency surface detail using shape from shading and incident illumination estimation as in Wu et al. [WVL*11]. The technique achieves impressive results for uncontrolled lighting, but does not take full advantage of specular surface reflections to estimate detailed facial geometry and reflectance.

Structured Lighting Systems Numerous successful techniques using structured light projection have addressed 3D facial scanning, including applications to dynamic facial capture [RHHL02, ZSCS04, DNRR05, ZH06]. However, these techniques generally operate at lower resolution than is required to record high resolution facial detail and do not specifically address reflectance capture.

Photometric Stereo Photometric stereo [Woo78] has been applied to recover dynamic facial performances using simultaneous illumination from a set of red, green and blue lights [HVB*07, KHE10]. However, these techniques are either data intensive or do not recover reflectance information. An exception is Georghiades [Geo03], who recovers shape and both diffuse and specular reflectance information for a face lit by multiple unknown point lights. The problem is formulated as uncalibrated photometric stereo and a constant specular roughness parameter is estimated over the face, achieving a medium scale reconstruction of the facial geometry. Zickler et al. [ZMKB08] showed that photometric invariants allow photometric stereo to operate on specular surfaces when the illuminant color is known. The practicality of photometric surface orientations in computer graphics has been demonstrated by Rushmeier et al. [RTG97] for creating bump maps, and Nehab et al. [NRDR05] for embossing such surface orientations for improved 3D geometric models. Hertzmann and Seitz [HS05] showed that with exemplar reflectance properties, photometric stereo can be applied accurately to materials with complex BRDF's, and Goldman et al. [GCHS05] presented simultaneous estimation of normals and a set of material BRDFs. However, all of these require multiple lighting conditions per viewpoint, which is prohibitive to acquire using near-instant capture with commodity DSLRs.

Specular Photometric Stereo Most of the above techniques have exploited diffuse surface reflectance for surface shape recovery. This is because typically specular highlights are not view-independent and shift across the subject as the location of the light and camera changes. Zickler et al. [ZBK02] exploits Helmholtz reciprocity to overcome this limitation for pairs of cameras and light sources. Significant work [CGS06, WMP*06, DHT*00] analyzes specular reflections to provide higher-resolution surface orientations for translucent surfaces. Ma et al. [MHP*07] and Ghosh

[†] DSLR facial photogrammetry setups can be found at The Capture Lab (http://www.capturelab.com/), Autodesk (used in [LLR13]), Ten24 (http://www.ten24.info/), and Infinite Realities (http://ir-ltd.net/)

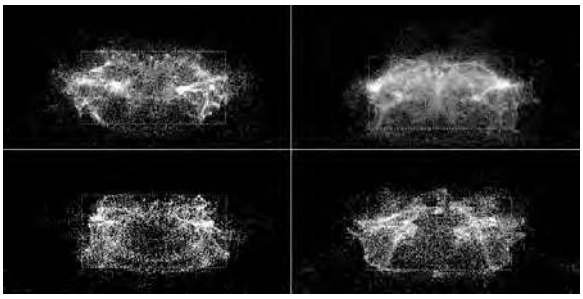


Figure 3: *Surface normal distributions for four faces, covering ears, forehead, and the front of the neck. The extents of the dotted rectangles are* $\pm 90^\circ$ *horizontally by* $\pm 45^\circ$ *vertically, each containing more than* 90% *of the normals.*

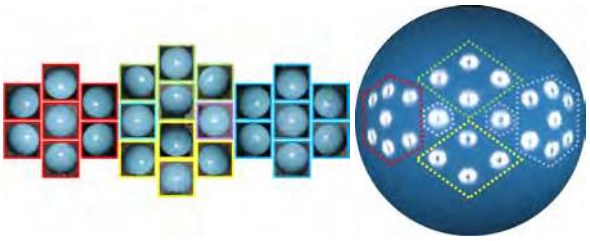


Figure 4: *(Left) 24 images of a shiny blue plastic ball shot with the apparatus. (Right) All 24 images reprojected onto a frontal view and summed, showing 24 evenly-spaced specular reflections from the six flash lighting conditions. The colored lines indicate which images correspond to each flash.*

et al. [GFT*11] perform photometric stereo using spherical gradient illumination and polarization difference imaging to isolate specular reflections, recording specular surface detail from a small number of images. While these techniques can produce high quality facial geometry, they require a complex acquisition setup such as an LED sphere and many photographs. In our work, we aim to record comparable facial geometry and reflectance with off-the-shelf components and near-instant capture.

Diffuse-Specular Separation Both polarization and color space analysis can be used for diffuse-specular separation [NFB97]. Mallick et al. [MZKB05] use a linear transform from RGB color space to an SUV color space where S is the monochromatic reflectance and UV is the chroma of the diffuse reflectance. Our diffuse-specular separation of flash-lit facial data leverages this approach, but exploits multi-view data to separate the S component.

3. Hardware Setup and Capture Process

Our capture setup is designed to record accurate 3D geometry with both diffuse and specular reflectance information per pixel while minimizing cost and complexity and maximizing the speed of capture. In all, we use 24 entry-level DSLR cameras and a set of six ring flashes arranged on a gantry seen in Fig. 2. The capture rig consists of 24 Canon EOS 600D entry-level consumer DSLR cameras,

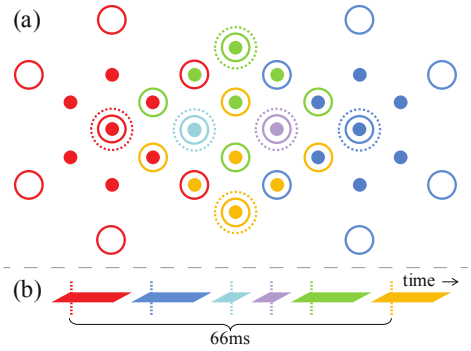


Figure 5: *(a) Location of the flashes (dotted circles), cameras (solid circles), and associated specular highlight half-angles (filled dots). The subject faces the center. The colors are for illustration; all flashes are the same white color. (b) The firing sequence for the flashes (dotted lines) and camera exposures (solid strips).*

which record RAW mode digital images at 5202 x 3565 pixel resolution. Using consumer cameras instead of machine vision video cameras dramatically reduces cost, as machine vision cameras of this resolution are very expensive and require high-bandwidth connections to dedicated capture computers. But to keep the capture near-instantaneous, we can only capture a *single* image with each camera, as these entry-level cameras require at least 1/4 second before taking a second photograph. Since our processing algorithm determines fine-scale surface detail from specular reflections, we wish to observe a specular highlight from the majority of the surface orientations of the face. We tabulated the surface orientations for four scanned facial models and found, not surprisingly, that over 90% of the orientations fell between $\pm 90^\circ$ horizontally and $\pm 45^\circ$ vertically of straight forward (Fig. 3). Thus, we arrange the flashes and cameras to create specular highlights for an even distribution of normal directions within this space as seen in Fig. 4. One way to achieve this distribution would be to place a ring flash on the lens of every camera and position the cameras over the ideal distribution of angles. Then, if each camera fires with its own ring flash, a specular highlight will be observed back in the direction of each camera. However, this requires shooting each camera with its own flash in succession, lengthening the capture process and requiring many flash units. Instead, we leverage the fact that position of a specular highlight depends not just on the lighting direction but also on the viewing direction, so that multiple cameras fired at once with a flash see different specular highlights according to the half-angles between the flash and the cameras. Using this fact, we arrange the 24 cameras and six diffused Sigma EM-140 ring flashes as seen in Fig. 5 to observe 24 specular highlights evenly distributed across the face. The colors indicate which cameras (solid circles) fire with which of the six flashes (dotted circles) to create observations of the specular highlights on surfaces (solid discs). For example, six cameras to the subject's left shoot with the "red" flash, four cameras shoot with the "green" flash, and a single camera shoots when the "purple" flash fires. In this arrangement, most of the cameras are not immediately adjacent to the flash they fire with, but they create specular reflections along a half-angle which does point toward a camera which is adjacent to the flash as shown in Fig. 6.

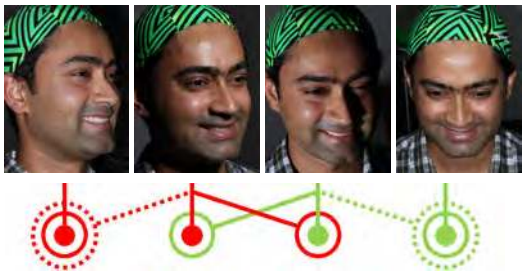


Figure 6: *Interleaved cameras and highlights: a subset of four images taken with the apparatus. The first and third cameras fire with the "red" flash, producing specular highlights at surface normals pointing toward the first and* second *cameras. Likewise, the second and fourth cameras fire with the "green" flash, producing highlights at surface normals pointing toward the third and fourth cameras. Left-to-right, the highlights progress across the face.*

The pattern of specular reflection angles observed can be seen on a blue plastic ball in Fig. 4. While the flashes themselves release their light in less than 1ms, the camera shutters can only synchronize to 1/200th of a second (5ms). When multiple cameras are fired along with a flash, a time window of 15ms is required since there is some variability in when the cameras take a photograph. In all, with the six flashes, four of which fire with multiple cameras, a total recording time of 66ms (1/15th sec) is achieved as in Fig. 5(b). By design, this is a shorter interval than the human blink reflex [BBL67].

3.1. Implementation Details

The one custom component in our system is a USB-programmable 80MHz Microchip PIC32 micro-controller for triggering the cameras via the remote shutter release input. The flashes are set to manual mode, full power, and are triggered by their corresponding cameras via the "hot shoe". The camera centers lie on a 1m radius sphere, framing the face using inexpensive Canon EF 50mm f/1.8 II lenses. A checkerboard calibration object is used to focus the cameras and to geometrically calibrate the camera's intrinsic, extrinsic, and distortion parameters. We also photograph an X-Rite ColorChecker Passport to calibrate the flash color and intensity. With the flash illumination, we can achieve a deep depth of field at an aperture of f/16 with the camera at its minimal gain of ISO 100 to provide well-focused images with minimal noise. While the cameras have built-in flashes, these could not used due to an Electronic Through-The-Lens (ETTL) metering process involving short bursts of light before the main flash. Our ring flashes are brighter and their locations are easily derived from the camera calibrations. By design, there is no flash in the subject's line of sight, and subjects reported no discomfort from the capture process.

3.2. Alternate Designs

We considered other design elements for the system including cross- and parallel-polarized lights and flashes, polarizing beamsplitters for diffuse/specular separation, camera/flash arrangements exploiting Helmholtz reciprocity for stereo correspondence [ZBK02], or a floodlit lighting condition with diffuse light from everywhere as employed in passive capture systems. While these techniques offer specific advantages for reflectance component separation, robust stereo correspondence, and/or deriving a diffuse albedo map (from flood lit illumination), we did not use them since each would either require additional cameras and/or lights for reflectance acquisition, or not achieve reflectance separation/estimation when employing flood lighting.

4. Deriving Geometry and Reflectance

We process the photographs into an accurate 3D model plus diffuse and specular reflectance maps as follows (refer to Fig. 7): We first leverage passive stereo reconstruction to build an approximate geometric mesh of the face from the photographs (Sec. 4.1). We then separate the diffuse and specular reflectance components of the photographs using a novel multi-view color-space analysis (Sec. 4.2). We further employ color-subspace photometric stereo for estimating diffuse (chroma) normals and albedo, and specular photometric stereo for estimating specular normals (Sec. 4.3). We also estimate a per-pixel specular exponent, producing a complete set of maps for rendering with high resolution surface details and skin reflectance. All maps are computed in (u,v) texture space for rendering purposes. Finally, we refine the geometric base mesh of the face using the estimated specular photometric information for the final high resolution facial geometry reconstruction (Sec. 4.4).

4.1. Constructing the Base Mesh

We begin by building a base mesh using AGISoft's PhotoScan[‡] software. Similar base mesh results can be obtained with PMVS2 [FP10] or Autodesk's 123D Catch[§] software. The base mesh is reconstructed with passive multi-view stereo and the 24 flash-lit photographs. Note that our images do not all have the same lighting and contain specular reflections and shadows, none of which is ideal for passive stereo reconstruction. However, we have sufficiently dense views under similar-enough lighting for the algorithm to find enough matching points between the images to construct a geometric model of the face accurate to within a few millimeters and aligning skin details to within one pixel; we call this the *base mesh*. We manually trim away extraneous surfaces (Fig. 8(a)), and create a minimally-distorted 4,096 × 4,096 pixel (u,v) texture map space using the commercial software UnFold 3D[¶].

4.2. Diffuse-Specular Separation

From our color calibration and since skin is dielectric, we can assume that the RGB specular color $\vec{s}$ in all images is (1,1,1). If we knew the diffuse color $\vec{d}$, i.e., the RGB color of the subsurface scattered light at a given pixel, then it would be trivial to decompose the pixel's RGB color into its diffuse and specular components. Mallick et al. [MZBK06] proposed a color-space separation using pixel neighborhoods to infer the diffuse color which works well for

‡ http://www.agisoft.ru/products/photoscan/
§ http://www.123dapp.com/catch
¶ http://www.polygonal-design.fr/

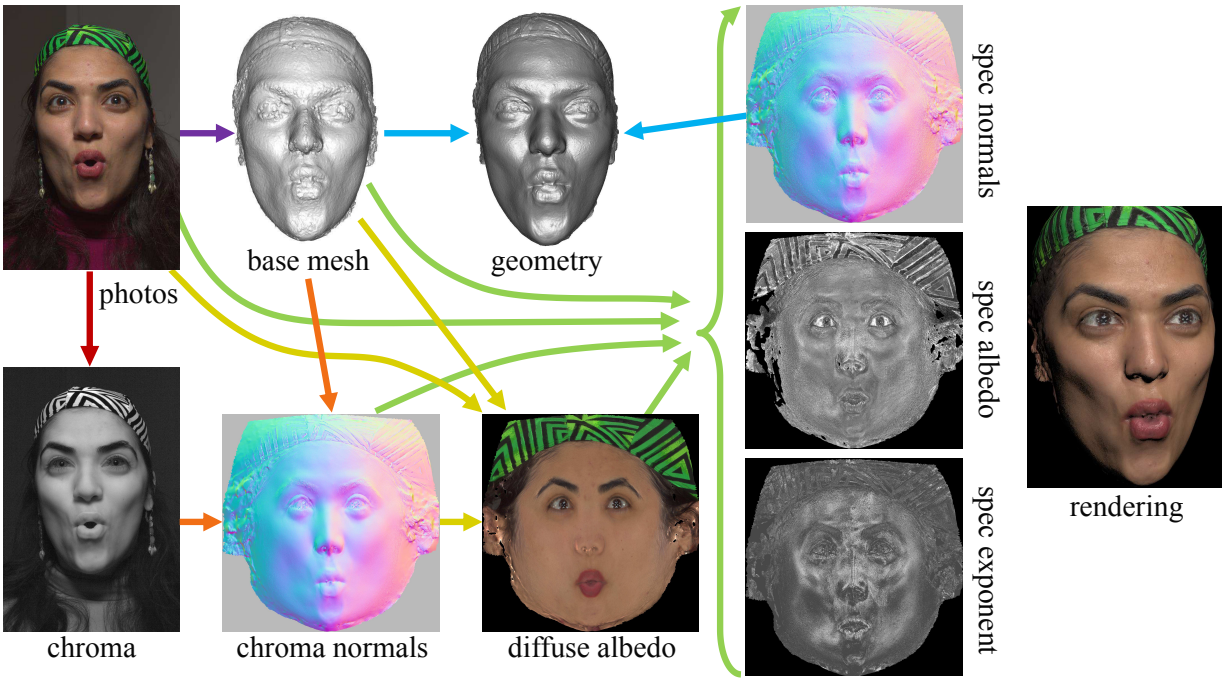


Figure 7: *Data flow diagram. 24 input photographs are used to produce a base mesh. Pixel correspondences obtained via the base mesh are employed to compute chroma normals, diffuse albedo, and finally specular reflectance parameters. The base mesh is combined with the specular normals to produce refined geometry. Synthetic renderings can be produced from the refined geometry and reflectance maps.*

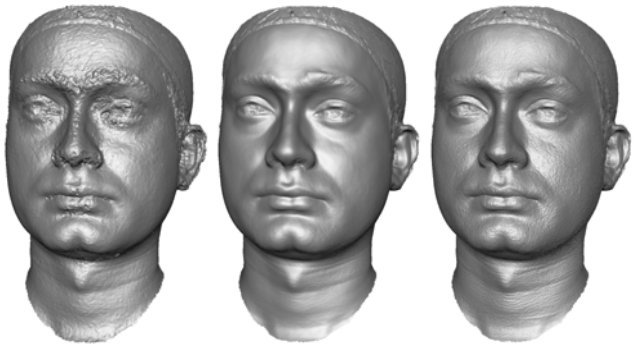


Figure 8: *(a) Base mesh from multi-view stereo. (b) Refined mesh using the diffuse (chroma) normals (c) Refined mesh using the specular reflectance analysis, which exhibits skin mesostructure details.*

relatively homogeneous dielectric materials. However, due to variation in melanin and hemoglobin, the diffuse color varies across the face. So, like [DHT*00, WMP*06], we leverage multiple images of a surface point in our dataset for diffuse-specular separation. The novelty of our method is the use of multiple views of a surface point under each illumination direction, which varies the specular highlight position, allowing effective separation with fewer illumination directions. We begin with photometric stereo on the chroma signal, following [ZMKB08]. Suppose we are examining a point on the face that projects into the different views in our dataset onto pixel values $\vec{p}_i = [p_r^i \; p_g^i \; p_b^i]^\top$, $i \in (1 \ldots k)$. (We omit views where the point is occluded or in shadow, using the base mesh to compute depth maps for visibility testing.) We rotate the RGB colors into the so-called *suv* color space via a matrix transform such that the s component aligns with $\vec{s}$, yielding $[p_s^i \; p_u^i \; p_v^i]^\top$. Then the *chroma intensities* $p_{uv}^i = \sqrt{p_u^i + p_v^i}$ are employed to compute a *chroma normal* $\vec{n}_{uv}$ using Lambertian photometric stereo (detailed in Sec. 4.3), as the u and v channels contain no specular highlight. As chroma information comes from light that has scattered deeply into the skin, the chroma normal map of a face has an extremely soft quality to it. We employ the chroma normal map for rendering the translucent appearance of skin similar to the hybrid normal rendering approach of Ma et al. [MHP*07]. However, the chroma normal is unsuitable for constructing detailed surface geometry. We therefore desire a normal map constructed from *specular* information, motivating us to separate the s channel into diffuse and specular components. As our dataset contains multiple illumination directions, we might use the most saturated pixel to establish a ratio of diffuse s to uv, allowing all p_s^i to be separated. However, this leaves a significant amount of single scattering reflection in the specular component, which would confound our specular analysis. Thus we instead compute the $s:uv$ ratio based on a blend of all the pixel values weighted by $(1-(\vec{n}_{uv}\cdot\vec{h}_i)^{10})^2$, where $\vec{h}_i$ is the halfway vector between the view vector and lighting direction for $\vec{p}_i$. This weighting is handcrafted to liberally suppress half-angles that might exhibit specular highlights (near $\vec{n}\cdot\vec{h}$) while not entirely suppressing the diffuse component. With the ratio $s:uv$ and the chroma surface normal in hand, it is trivial to establish the RGB diffuse albedo and to remove the diffuse component from all pixel values, leaving only specu-

lar highlights. Fig. 9 shows separation results for some example views. Note that single scattering is excluded from the specular estimate, contrary to polarization-based separation. This allows us to employ *Blinn-Phong* photometric stereo to extract detailed specular surface normals from the specular highlight intensities (detailed in Sec. 4.3), as Blinn-Phong has been previously shown to well model specular reflectance in human skin [WMP*06].

4.3. Diffuse and Specular Photometric Stereo

Given multiple observed pixel values p_i of a surface point under differing illumination directions $\vec{l}_i$, it is possible to recover the surface normal $\vec{n}$ and albedo ρ by leveraging certain assumptions about the reflectance properties of the surface using. This process is known as *photometric stereo* [Woo78]. The photometric stereo equations are presented with a distant light assumption, and light intensity π. If the actual distances r_i to the light sources are known, and the intensities I_i are known, then the pixel values can be adjusted to conform to the assumptions by multiplying them by $\pi r_i^2/I_i$ before proceeding with photometric stereo. We review the photometric stereo equations for exposition. In the Lambertian case, the lighting equation is $L\vec{\beta} = P$, where $L = \left[\vec{l}_1\ \vec{l}_2\ \cdots\ \vec{l}_k\ \right]^\top$, $\vec{\beta} = \rho\vec{n}$, and $P = \left[p_1\ p_2\ \cdots\ p_k\right]^\top$. Any i with $p_i = 0$ are omitted, as the lighting equation does not hold. The solution via pseudoinverse is:

$$\vec{\beta} = (L^\top L)^{-1} L^\top P. \tag{1}$$

In the Blinn-Phong case, the lighting equation is expressed in terms of halfway vectors $\vec{h}_i$ instead of lighting directions, and includes an exponent α with an associated normalization factor to conserve energy, leading to the following:

$$H\vec{\gamma} = P, \text{ where } H = \left[\vec{h}'_1\ \vec{h}'_2\ \cdots\ \vec{h}'_k\right]^\top,$$
$$\vec{h}'_i = p_i^{1-1/\alpha}\left(f_i\tfrac{\alpha+8}{8}\right)^{1/\alpha}\frac{\vec{v}_i+\vec{l}_i}{\|\vec{v}_i+\vec{l}_i\|}, \text{ and } \vec{\gamma} = \rho^{1/\alpha}\vec{n}, \tag{2}$$

with v_i the direction towards the viewer, α the Blinn-Phong exponent, and f_i is a Fresnel term. Note that (2) may be rearranged to produce the familiar Blinn-Phong lighting model. Provided α and f_i are known, the solution via pseudoinverse has the same form:

$$\vec{\gamma} = (H^\top H)^{-1} H^\top P. \tag{3}$$

As there are fewer non-zero values in the specular signal than the chroma signal, the specular surface normal is noisier than the chroma surface normal, especially as the normal bends away from all of the halfway vectors. Therefore we introduce a regularization term based on the chroma signal:

$$\vec{\gamma} = (H^\top H + \lambda L^\top L)^{-1}(H^\top P_s + \lambda L^\top P_{uv}/\rho_{uv}), \tag{4}$$

where P_s indicates the use of the specular signal values in (2), P_{uv} represents the chroma signal values, ρ_{uv} is the chroma albedo (already computed), and λ is the regularization strength. Note that dividing the chroma signal by the chroma albedo scales the solution of the regularization term to exactly $\vec{n}_{uv}$, rendering it compatible with the solution of the specular term (since $\rho_s^{1/\alpha}\vec{n}_s \approx \vec{n}_{uv}$). We found that setting $\lambda = 0.015$ retained the high-frequency detail in the specular surface normals that was not present in the chroma surface normals, while significantly reducing noise. We also estimate the per surface point specular exponent α for rendering purposes. As the specular exponent and Fresnel term are not known in advance, we employ an iterative inverse rendering process. At iteration t (with $t = 1\ldots8$) we let $\alpha = 2^{t+1}$, and use the Schlick approximation for the Fresnel term with the normal from the previous iteration: $f_i = 0.1 + 0.9(1 - \vec{v}_i \cdot \vec{n}^{\,t-1})^5$, starting with $\vec{n}^{\,0} = \vec{n}_{uv}$. While the index of refraction of skin suggests that 0.03 should be used for reflectance at normal incidence in place of 0.1, we found that small errors in the diffuse-specular separation produced artifacts which are mitigated by allowing the Fresnel term to take a slightly higher value at normal incidence. After each iteration, we relight the estimated specular components for each lighting condition using the corresponding specular exponent, albedo and normal and retain the values from the iteration that produces the maximum photoconsistency to the specular highlights.

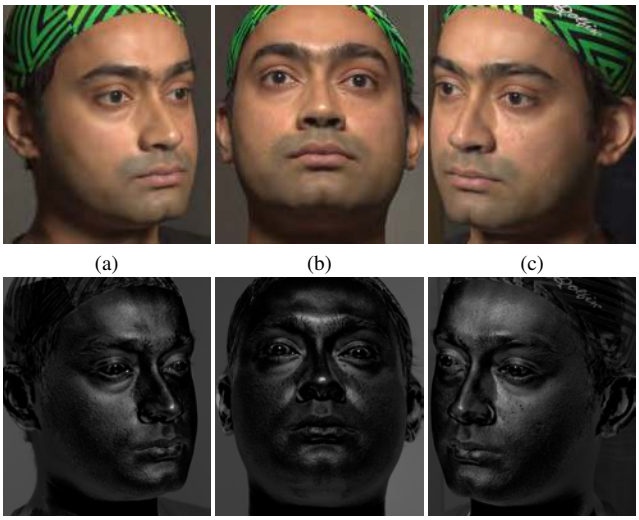


Figure 9: *Diffuse-Specular separation. (a-c) Three of the 24 original photographs. (d-f) Estimated specular components.*

4.4. Mesh Refinement

We refine the facial geometry mesh using the method of Nehab et al. [NRDR05]. We first resample the base mesh to produce a fine mesh using a regular 4096×4096 sampling in UV space. We then employ the *low-frequency rotation field* idea from [NRDR05] to remove any low-frequency disagreement between the photometric normals and the base mesh. One issue is that, due to occlusions, our photometric normal estimates may use a different set of views for different points on the surface and hence may contain seams. To alleviate the seam problem, we modify the method such that only points having the same set of visible views are blended together when computing the rotation field. We finally employ the *full model optimization* method of [NRDR05] to produce the final high resolution facial mesh (Fig. 8(c)). We then repeat the entire method once more, using the refined mesh as the base mesh for the second iteration. This reduces artifacts stemming from the coarse facets of the original base mesh.

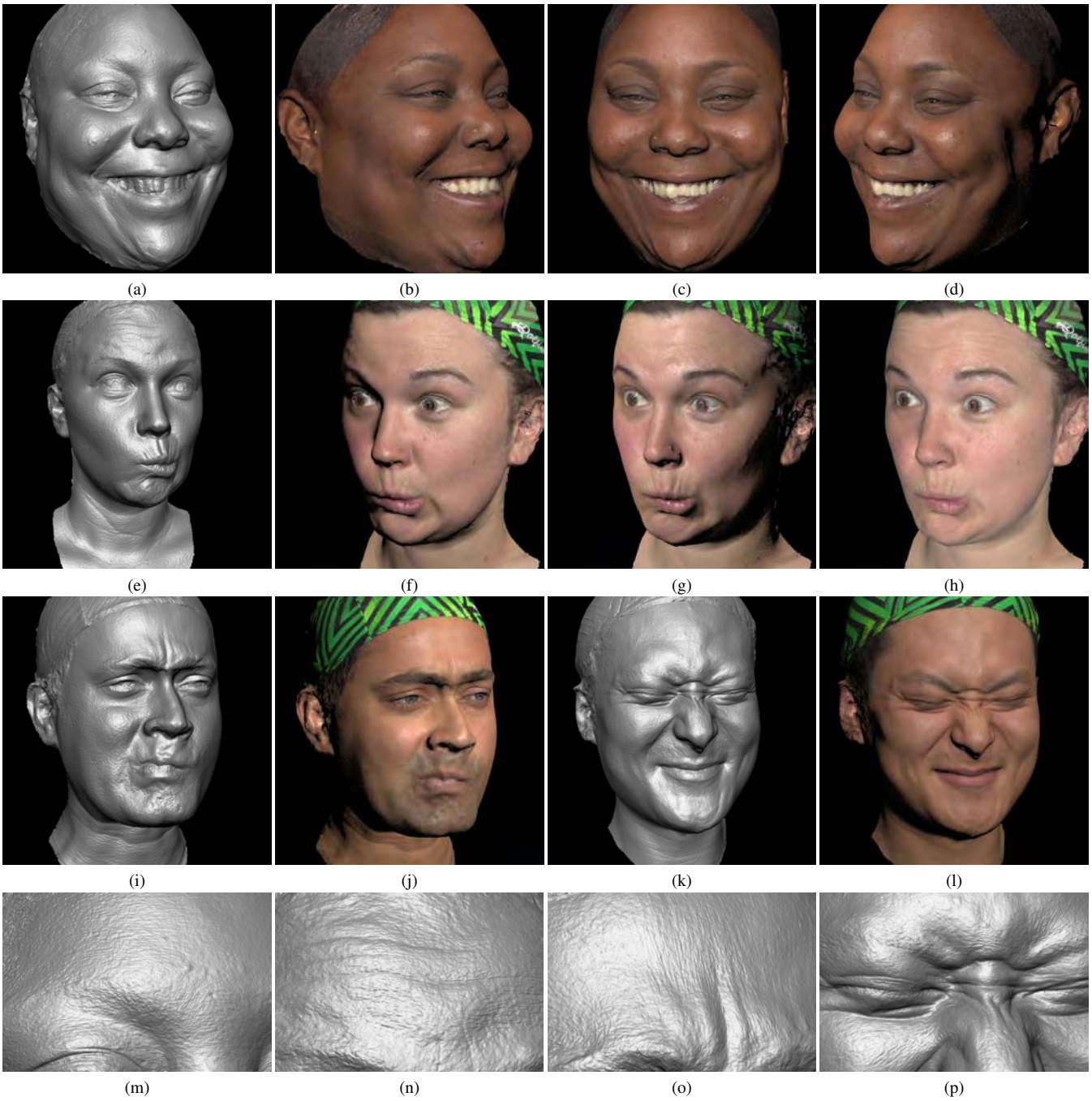


Figure 10: *(a-l) Renderings of recovered geometry (a,e,i,k) and reflectance maps for four subjects under novel viewpoint (b,c,d,j,l) and lighting (f,g,h,j,l). (m-p) Zooms of regions in (a,e,i,k) exhibiting fine skin detail unavailable from passive capture systems.*

5. Results

We employed our system to acquire a variety of subjects in differing facial expressions. Figs. 1, 7, and 10 show high-resolution geometry and renderings under novel viewpoint and lighting using our method, with complete results including all recovered reflectance maps shown in Fig. 7. Our acquisition system produces geometric quality competitive with more complex systems and reflectance maps not available from single-shot methods. Running on a dual quad-core 2.4 GHz Intel Xeon E5620 CPU with hyperthreading and an NVidia GTX Titan graphics card, the initial passive stereo geometry solve typically takes about 25 minutes, estimating the reflectance maps takes 15 minutes, and refining the geometry

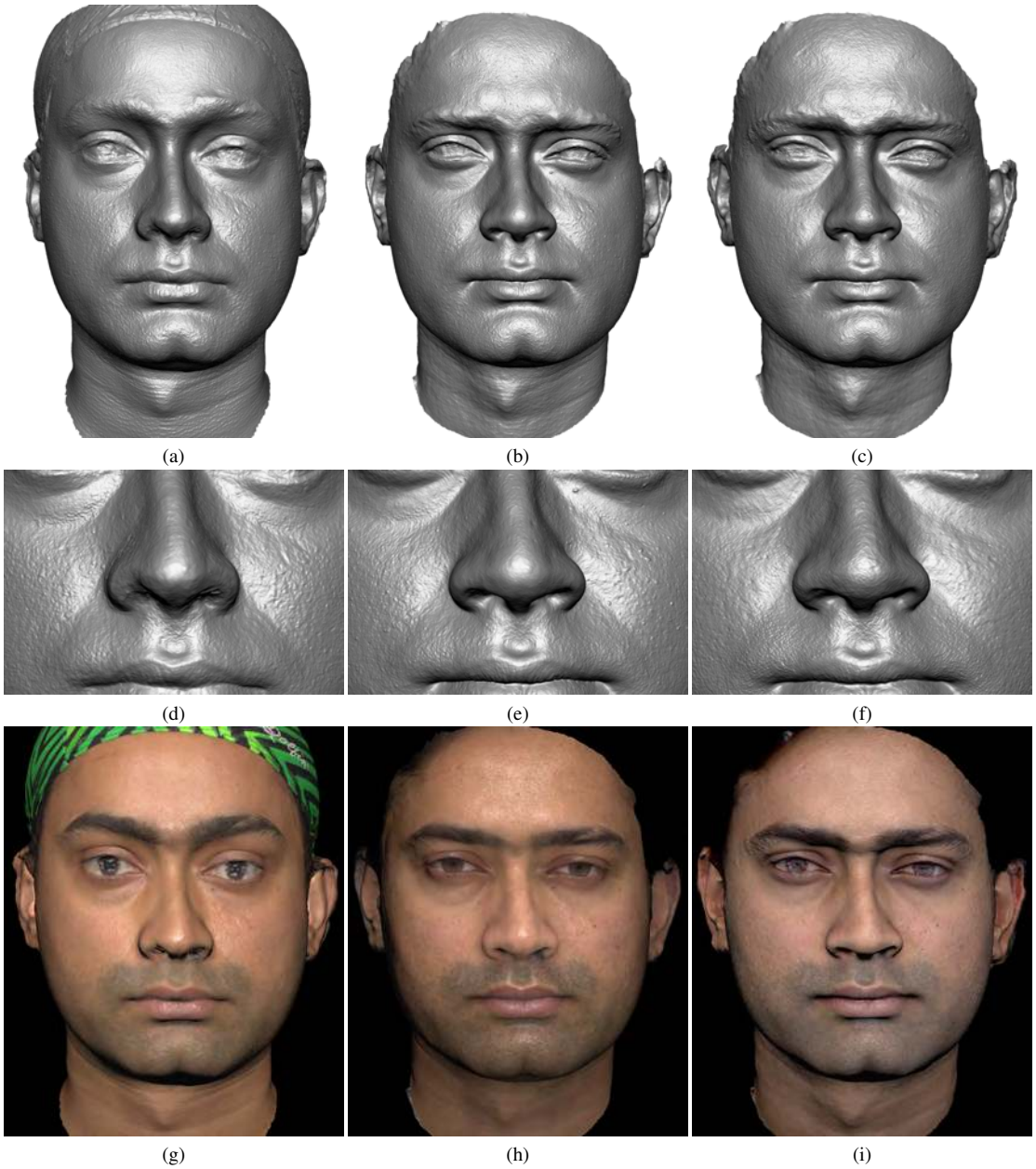


Figure 11: *Reconstructed geometry using: (a) Our method. (b) Polarized spherical gradient illumination [GFT*11]. (c) Passive stereo reconstruction using the flat-lit photographs from (b) and "dark is deep" detail emboss. (d-f) Zooms of (a-c). Note that (d) and (e) are often in agreement in regard to fine skin surface details, while (f) often disagrees. (g-i) Renderings using the results from (a-c) with the same view and illumination. All reflectance parameters in (g) are estimated automatically. The specular exponent in (h) is tuned manually as it is not measured. In (i), all reflectance parameters are chosen manually, using the flat-lit photographs as diffuse albedo.*

takes an additional 20 minutes. Fig. 11(a-c) show geometry reconstruction comparing our method (a) to the method of [GFT*11] (b), and also to passive stereo reconstruction (c) using the "dark is deep" heuristic to emboss surface detail [LZ94, GWM*08, BBB*10]. It is worth noting that the polarized spherical gradient illumination result employs 7 high end DSLR cameras (Canon EOS 1D X). In comparison, though our acquisition method employs more cameras (24), we employ relatively inexpensive entry level DSLR cameras (Canon EOS 600D) resulting in a lower overall cost. The fine scale surface mesostructure is faithfully reconstructed using our method (obtained from specular reflections) without requiring a complex LED sphere setup as in [GFT*11]. Note that the surface details in (c) are predominantly concave, whereas (a) and (b) exhibit a mix of convex and concave features, and are largely in agreement. For

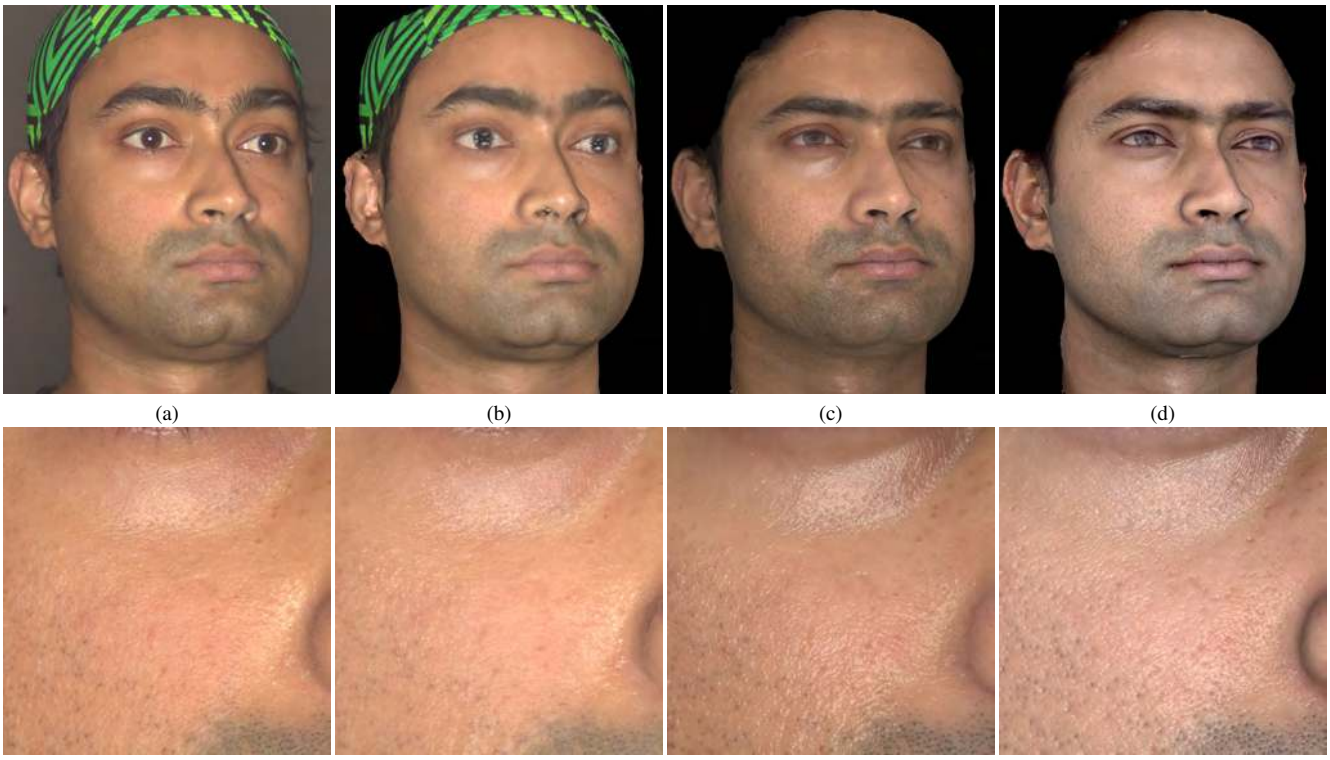


Figure 12: *(a) Ground-truth photograph with frontal illumination. (b-d) Synthetic renderings with the same view and illumiation using (b) our proposed method, (c) polarized spherical gradient illumination [GFT*11], and (d) passive stereo reconstruction using the flat-lit photographs from (c) and a "dark is deep" detail emboss. Some of the reflectance maps in (c) and (d) are not estimated automatically and are tuned manually (see text for details). (e-h) Zooms of cheek region in (a-d). Note that the fine-scale details in the specular highlights on the skin in (f) and (g) are generally in agreement with ground truth, while the highlight details in (h) differ significantly.*

example, moles that are clearly visible in (b) under the subject's left tear duct and on the left labial-nasal fold are also visible in (a) (though more faintly), but are entirely absent in (c), and facial hair stubble appears as small dents in the neck in (c) while both (a) and (b) exhibit a convex bump located at each follicle. (Please see the electronic copy for high-resolution images.) Fig. 11(g-i) show renderings of the results in (a-c), using the hybrid normal technique. The specular component in (h, i) is a dual-lobe phong model with manually tuned exponents, while (g) is our automatically estimated single-lobe model. The diffuse-specular separation in the passive result (i) is not measured, so specular albedo is set to a constant value of 1 (with Fresnel) and the diffuse albedo is simply the values from the flat-lit photographs. Further, the soft diffuse surface normal used in hybrid normal rendering is not measured photometrically in (i), so it is approximated as a blurred version of the geometry surface normal. Despite the manual effort to produce the reflectance maps in (i), the "dark is deep" surface detail produces a uniform, dimpled appearance that lacks the structural variation visible in human skin. Fig. 12 shows a ground-truth photograph and synthetic renderings from an additional viewpoint. The automatically estimated parameters from our proposed method produce similar specular highlights as ground truth, while the uniform lobe parameters used in the polarized spherical gradient result do not capture the subtle variations in the fine-scale highlight details. The "dark is deep" surface detail in the passive stereo result produces fine-scale highlight details that differ significantly from the ground truth, despite some manual effort tuning the reflectance values. These results support the conclusion in [GFT*11] that "dark is deep" detail embossing hallucinates geometric detail that may have little correlation to the actual fine-scale geometry of the subject. For surfaces with little albedo variation such as the undulating surfaces in [LZ94] or the validation mask in [BBB*10], the heuristic is of course valid. For surfaces with more varied albedo, textureless renderings of "dark is deep" geometry are visually pleasing, perhaps because it is aesthetically similar to carving relief into a marble statue to mimic albedo variation. However the skin of real subjects has natural variations in albedo that violate the "dark is deep" heuristic, and also a degree of translucency that further confuses the heuristic when structures are visible just underneath the skin such as closely shaven facial hair. Thus for photorealistic renderings of live human subjects, it is necessary to capture the geometry of the actual outer surface of the skin for accurate specular reflections, in addition to the other reflectance properties. Our proposed method captures these important properties, while maintaining a near-instant capture time competitive with passive stereo. Fig. 13 shows synthetic renderings using the hybrid nor-

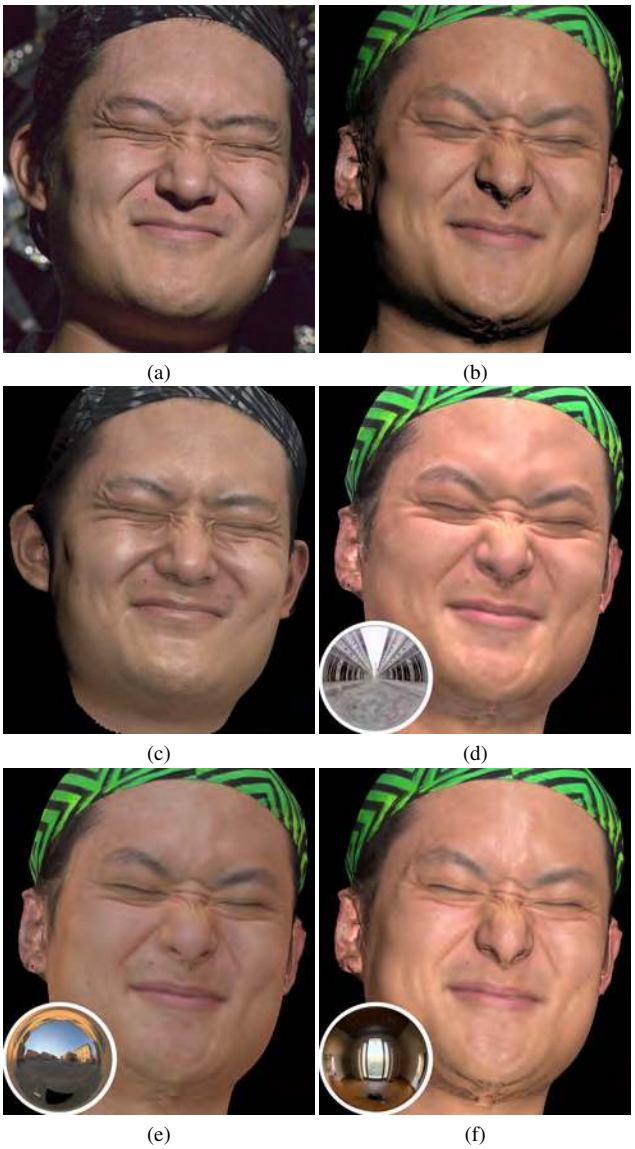


Figure 13: *(a) Ground truth photograph of a subject under point light illumination (not used in the reconstruction). (b) Rendering with similar lighting to (a) of automatically reconstructed geometry and reflectance maps using our method, for a similar expression captured in a different session. (c) Rendering with similar lighting to (a) of reconstructed geometry and reflectance maps using polarized spherical gradient illumination [GFT*11] and a hand-tuned specular exponent. (d-f) Additional renderings of our result using high dynamic range image based lighting (shown inset).*

mal technique [MHP*07] to produce a skin-like appearance under point lighting and under high dynamic range image based lighting, comparing our method to ground truth and to the polarized spherical gradient method of [GFT*11]. Both methods produce lifelike synthetic faces with similar attributes as ground truth, though the method of [GFT*11] requires hand-tuning a specular exponent.

6. Discussion and Future Work

We reiterate that our framework allows exploration of reflectance capture techniques combined with facial geometry capture, with comparable cost and complexity to passive systems. Active systems such as [GFT*11] require far more expensive sports photography cameras to capture multiple images in rapid succession, and complex programmable polarized light sources. We simply place and fire our entry-level DSLR cameras and flashes in a specific order. The results from our system suggest several avenues for future work. The surfaces of the eyes do not reconstruct well, due in part to the disparity between the diffuse reflection of the iris and the specular reflection of the cornea. Detecting eyes and modeling them specifically as in [BBN*14] would be of interest. Adding more cameras with views of high-curvature features such as nostrils and ears could improve results without increasing capture time. If more flashes were added to extend the system further around the head, capture time would increase but views from these angles might not see the eyes blinking anyway, so that a complete model of the head could be captured. Modeling facial hair as in [BBN*12] would expand the utility of the system. Since the number of lighting conditions is small, the technique could in principal be applied to dynamic facial performances, using optical flow for temporal alignment with video cameras synchronized to alternating light sources. We are also clearly not exploiting all of the reflectance cues present within our data. The high resolution surface detail allows much of the spatially-varying skin BRDF to be exhibited directly from the geometry; however, using reflectance sharing [ZREB06], it may be possible to derive improved diffuse and specular BRDFs of the skin. Also, the shadow transitions seen in the data could be analyzed to estimate subsurface scattering parameters for regions of the face. Although some of our subjects wore modest amounts of makeup without negatively impacting the reconstructions, the technique may require refinement to faithfully record and render the subtle reflectance effects that makeup is engineered to create.

7. Conclusion

We have presented a near-instant capture technique for recording the geometry and reflectance of a face from a set of still photographs lit by flash illumination. The technique leverages photoconsistency, photometric stereo, and specular reflections simultaneously to solve for facial shape and reflectance that explain the input photographs. It is the first near-instant capture technique able to produce such data at high resolution and at substantially lower cost than more complex reflectance measurement setups.

Acknowledgements

We wish to thank our scan subjects, and the following for their support and assistance: Xueming Yu, Jay Busch, Kathleen Haase, Bill Swartout, Cheryl Birch, Randall Hill and Randolph Hall. This work was sponsored by the University of Southern California Office of the Provost and the U.S. Army Research, Development, and Engineering Command (RDECOM), and partially supported by a Royal Society Wolfson Research Merit Award. The content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

[ARL*10] ALEXANDER O., ROGERS M., LAMBETH W., CHIANG J.-Y., MA W.-C., WANG C.-C., DEBEVEC P.: The Digital Emily Project: Achieving a photoreal digital actor. *IEEE Computer Graphics and Applications 30* (July 2010), 20–31. 1

[BBB*10] BEELER T., BICKEL B., BEARDSLEY P., SUMNER B., GROSS M.: High-quality single-shot capture of facial geometry. *ACM Trans. Graph. 29* (July 2010), 40:1–40:9. 1, 2, 8, 9

[BBL67] BIXLER E. O., BARTLETT N. R., LANSING R. W.: Latency of the blink reflex and stimulus intensity. *Perception & Psychophysics 2*, 11 (1967), 559–560. 2, 4

[BBN*12] BEELER T., BICKEL B., NORIS G., BEARDSLEY P., MARSCHNER S., SUMNER R. W., GROSS M.: Coupled 3d reconstruction of sparse facial hair and skin. *ACM Trans. Graph. 31*, 4 (July 2012), 117:1–117:10. 10

[BBN*14] BÉRARD P., BRADLEY D., NITTI M., BEELER T., GROSS M.: High-quality capture of eyes. *ACM Trans. Graph. 33*, 6 (Nov. 2014). 10

[BHPS10] BRADLEY D., HEIDRICH W., POPA T., SHEFFER A.: High resolution passive facial performance capture. *ACM Trans. Graph. 29* (July 2010), 41:1–41:10. 1, 2

[CGS06] CHEN T., GOESELE M., SEIDEL H. P.: Mesostructure from specularities. In *CVPR* (2006), pp. 1825–1832. 2

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2000), SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., pp. 145–156. 2, 5

[DNRR05] DAVIS J., NEHAB D., RAMAMOORTHI R., RUSINKIEWICZ S.: Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Trans. Pattern Anal. Mach. Intell. 27*, 2 (2005), 296–302. 2

[FP09] FURUKAWA Y., PONCE J.: Dense 3D motion capture for human faces. In *Proc. of CVPR 09* (2009). 2

[FP10] FURUKAWA Y., PONCE J.: Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell. 32*, 8 (Aug. 2010), 1362–1376. 4

[GCHS05] GOLDMAN D. B., CURLESS B., HERTZMANN A., SEITZ S. M.: Shape and spatially-varying brdfs from photometric stereo. In *ICCV* (2005), pp. 341–348. 2

[Geo03] GEORGHIADES A.: Recovering 3-D shape and reflectance from a small number of photographs. In *Rendering Techniques* (2003), pp. 230–240. 2

[GFT*11] GHOSH A., FYFFE G., TUNWATTANAPONG B., BUSCH J., YU X., DEBEVEC P.: Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 30*, 6 (2011). 1, 3, 8, 9, 10

[GWM*08] GLENCROSS M., WARD G. J., MELENDEZ F., JAY C., LIU J., HUBBOLD R.: A perceptually validated model for surface depth hallucination. *ACM Trans. Graph. 27*, 3 (Aug. 2008), 59:1–59:8. 2, 8

[HS05] HERTZMANN A., SEITZ S. M.: Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *PAMI 27*, 8 (2005), 1254–1264. 2

[HVB*07] HERNANDEZ C., VOGIATZIS G., BROSTOW G. J., STENGER B., CIPOLLA R.: Non-rigid photometric stereo with colored lights. In *Proc. IEEE International Conference on Computer Vision* (2007), pp. 1–8. 2

[KHE10] KLAUDINY M., HILTON A., EDGE J.: High-detail 3D capture of facial performance. In *International Symposium 3D Data Processing, Visualization and Transmission (3DPVT)* (2010). 2

[LLR13] LUO L., LI H., RUSINKIEWICZ S.: Structure-aware hair capture. *ACM Trans. Graph. 32*, 4 (July 2013). 2

[LZ94] LANGER M. S., ZUCKER S. W.: Shape-from-shading on a cloudy day. *J. Opt. Soc. Am. A 11*, 2 (Feb 1994), 467–478. 2, 8, 9

[MHP*07] MA W.-C., HAWKINS T., PEERS P., CHABERT C.-F., WEISS M., DEBEVEC P.: Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Rendering Techniques* (2007), pp. 183–194. 2, 5, 10

[MZBK06] MALLICK S. P., ZICKLER T., BELHUMEUR P. N., KRIEGMAN D. J.: Specularity removal in images and videos: A pde approach. In *ECCV* (2006). 4

[MZKB05] MALLICK S. P., ZICKLER T. E., KRIEGMAN D. J., BELHUMEUR P. N.: Beyond lambert: Reconstructing specular surfaces using color. In *CVPR* (2005). 3

[NFB97] NAYAR S., FANG X., BOULT T.: Separation of reflection components using color and polarization. *IJCV 21*, 3 (1997), 163–186. 3

[NRDR05] NEHAB D., RUSINKIEWICZ S., DAVIS J., RAMAMOORTHI R.: Efficiently combining positions and normals for precise 3D geometry. *ACM TOG 24*, 3 (2005), 536–543. 2, 6

[PHL*98] PIGHIN F., HECKER J., LISCHINSKI D., SZELISKI R., SALESIN D. H.: Synthesizing realistic facial expressions from photographs. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1998), SIGGRAPH '98, ACM, pp. 75–84. 1

[RHHL02] RUSINKIEWICZ S., HALL-HOLT O., LEVOY M.: Real-time 3D model acquisition. *ACM TOG 21*, 3 (2002), 438–446. 2

[RR08] RAMELLA-ROMAN J. C.: Out of plane polarimetric imaging of skin: Surface and subsurface effect. In *Optical Waveguide Sensing and Imaging*, Bock W. J., Gannot I., Tanev S., (Eds.), NATO Science for Peace and Security Series B: Physics and Biophysics. Springer Netherlands, 2008, pp. 259–269. "10.1007/978-1-4020-6952-9_12". 2

[RTG97] RUSHMEIER H., TAUBIN G., GUÉZIEC A.: Applying shape from lighting variation to bump map capture. In *Rendering Techniques* (1997), pp. 35–44. 2

[VWB*12] VALGAERTS L., WU C., BRUHN A., SEIDEL H.-P., THEOBALT C.: Lightweight binocular facial performance capture under uncontrolled lighting. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2012) 31*, 6 (November 2012), 187:1–187:11. 2

[WMP*06] WEYRICH T., MATUSIK W., PFISTER H., BICKEL B., DONNER C., TU C., MCANDLESS J., LEE J., NGAN A., JENSEN H. W., GROSS M.: Analysis of human faces using a measurement-based skin reflectance model. *ACM TOG 25*, 3 (2006), 1013–1024. 1, 2, 5, 6

[Woo78] WOODHAM R. J.: Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Proc. SPIE's 22nd Annual Technical Symposium* (1978), vol. 155. 2, 6

[WVL*11] WU C., VARANASI K., LIU Y., SEIDEL H.-P., THEOBALT C.: Shading-based dynamic shape refinement from multi-view video under general illumination. In *Proceedings of the 2011 International Conference on Computer Vision* (2011), ICCV '11, pp. 1108–1115. 2

[ZBK02] ZICKLER T. E., BELHUMEUR P. N., KRIEGMAN D. J.: Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *Int. J. Comput. Vision 49*, 2-3 (2002), 215–227. 2, 4

[ZH06] ZHANG S., HUANG P.: High-resolution, real-time three-dimensional shape measurement. *Optical Engineering 45*, 12 (2006). 2

[ZMKB08] ZICKLER T., MALLICK S. P., KRIEGMAN D. J., BELHUMEUR P. N.: Color subspaces as photometric invariants. *Int. J. Comput. Vision 79*, 1 (Aug. 2008), 13–30. 2, 5

[ZREB06] ZICKLER T., RAMAMOORTHI R., ENRIQUE S., BELHUMEUR P. N.: Reflectance sharing: Predicting appearance from a sparse set of images of a known shape. *PAMI 28*, 8 (2006), 1287–1302. 10

[ZSCS04] ZHANG L., SNAVELY N., CURLESS B., SEITZ S. M.: Spacetime faces: high resolution capture for modeling and animation. *ACM TOG 23*, 3 (2004), 548–558. 2