

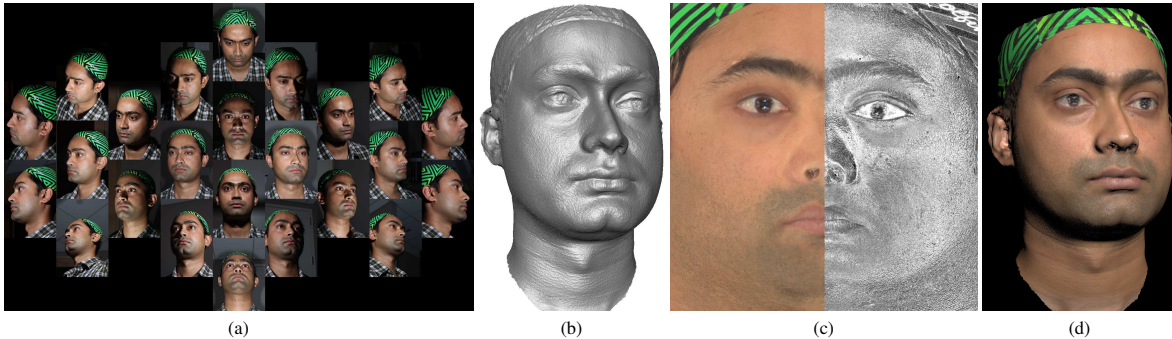
# Near-Instant Capture of High-Resolution Facial Geometry and Reflectance

## ICT Technical Report No. ICT-TR-01-2015

### May, 2015

Paul Graham\*    Graham Fyffe\*    Borom Tunwattanapong\*    Abhijeet Ghosh<sup>†</sup>    Paul Debevec\*

\*USC Institute for Creative Technologies    <sup>†</sup>Imperial College London



**Figure 1:** (a) Multi-view images shot under rapidly varying flash directions. (b) Refined geometry (c) Diffuse/specular maps (d) Rendering

## Abstract

We present a near-instant method for acquiring facial geometry and reflectance using commodity components consisting of a set of 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 multiview 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.

## 1 Introduction

Modeling realistic human characters is frequently done using 3D recordings of the shape and appearance of real people across a set of different facial expressions [Pighin et al. 1998; Alexander et al. 2010] to build blendshape facial models. Believable characters which cross the "Uncanny Valley" 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 which is near-instantaneous and relatively low-cost. While some facial capture techniques are instantaneous and inexpensive [Beeler et al. 2010; Bradley et al. 2010], these do not generally provide lighting-independent texture maps, specular reflectance information, or high-resolution surface normal detail for relighting. In contrast, techniques which use multiple photographs from spherical lighting setups [Weyrich et al. 2006; Ghosh et al. 2011] 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 near-instant facial capture technique which records high-quality facial geometry and reflectance using commodity hardware. We use a 24-camera DSLR photogrammetry setup similar to common commercial systems<sup>1</sup> and use six ring flash units to light the face. However, instead of the usual process 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 [Bixler et al. 1967]. 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. However, 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 separated diffuse and 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 multiview color-space analysis and novel photometric estimation of specular surface normals for geometry refinement.

## 2 Related work

We leverage a wide body of work in computer vision and graphics for facial geometry and reflectance measurement. While a compre-

<sup>1</sup>DSLR facial photogrammetry setups can be found at The Capture Lab (<http://www.capturelab.com/>), Autodesk (used in [Luo et al. 2013]), Ten24 (<http://www.ten24.info/>), and Infinite Realities (<http://ir-ltd.net/>)

hensive overview of the literature is beyond the scope of this paper, we discuss some of the most related previous work.

**Passive Multi-View Stereo** There is a rich history of work in the computer vision literature on passive multi-view stereo reconstruction of scenes including faces. Our facial base geometry reconstruction step takes advantage of the work of Furukawa and Ponce [Furukawa and Ponce 2009] 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 [Ramella-Roman 2008] 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. [Beeler et al. 2010] and Bradley et al. [Bradley et al. 2010] to reconstruct high quality facial geometry under diffuse illumination. Beeler et al. apply mesoscopic augmentation as in [Langer and Zucker 1994][Glencross et al. 2008] 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. [Valgaerts et al. 2012] 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. [Wu et al. 2011]. 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 [Rusinkiewicz et al. 2002; Zhang et al. 2004; Davis et al. 2005; Zhang and Huang 2006]. However, these techniques generally operate at lower resolution than is required to record high resolution facial detail and do not specifically address reflectance capture.

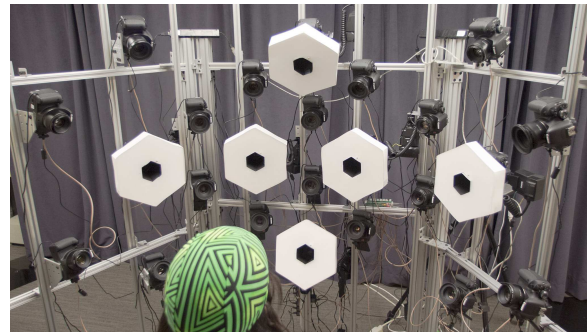
**Diffuse Photometric Stereo** Photometric stereo [Woodham 1978] has been applied to recover dynamic facial performances using simultaneous illumination from a set of red, green and blue lights [Hernandez et al. 2007; Klaudiny et al. 2010]. However, these techniques are either data intensive or do not recover reflectance information. An exception is Georghiades [Georghiades 2003], 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. [Zickler et al. 2008] 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. [Rushmeier et al. 1997] for creating bump maps, and Nehab et al. [Nehab et al. 2005] for embossing such surface orientations for improved 3D geometric models. Hertzmann and Seitz [Hertzmann and Seitz 2005] showed that with exemplar reflectance properties, photometric stereo can be applied accurately to materials with complex BRDF's, and Goldman et al. [Goldman et al. 2005] 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. [Zickler et al. 2002] exploits Helmholtz reciprocity to overcome this limitation for pairs of cameras and light sources. Significant work [Chen et al. 2006; Weyrich et al. 2006; Debevec et al. 2000] analyzes specular reflections to provide higher-resolution surface orientations for translucent surfaces. Ma et al. [Ma et al. 2007] and Ghosh et al. [Ghosh et al. 2011] 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 in separating diffuse and specular reflections [Nayar et al. 1997]. Mallick et al. [Mallick et al. 2005] use a linear transform from RGB color space to an SUV color space where S corresponds to the intensity of monochromatic specular reflectance and UV correspond to the orthogonal chroma of the diffuse reflectance. Our technique for diffuse-specular separation of the flash-lit facial data leverages this approach, but takes advantage of the multi-view dataset we have available.

### 3 Hardware Setup and Capture Process

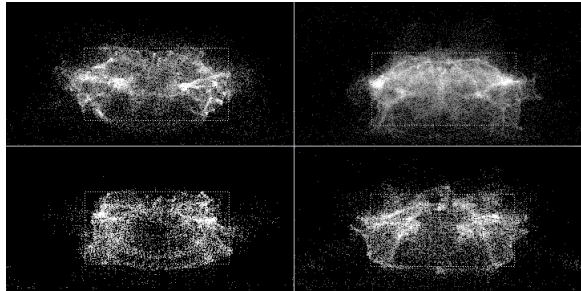
Our capture setup is designed to record accurate 3D geometry with both diffuse and specular reflectance information per pixel while minimizing cost and complexity and maximizing the speed of capture. In all, we use 24 entry-level DSLR cameras and a set of six ring flashes arranged on a gantry seen in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



**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.

**Camera and Flash Arrangement** The capture rig consists of 24 Canon EOS 600D entry-level consumer DSLR cameras, 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.

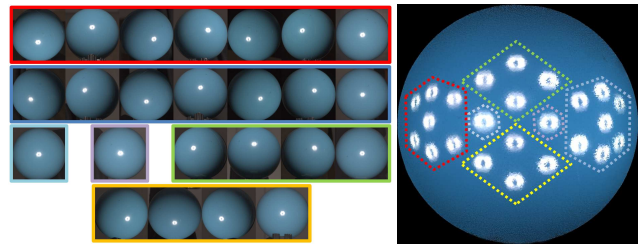
Since our processing algorithm determines fine-scale surface detail from specular reflections, we wish to observe a specular highlight from the majority of the surface orientations of the face. We tabulated the surface orientations for four scanned facial models and found, not surprisingly, that over 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



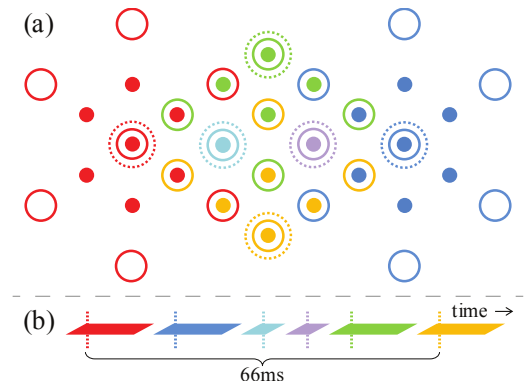
**Figure 3:** Surface normal distributions for four faces, covering ears, forehead, and the front of the neck. The extents of the dotted rectangles are  $\pm 90^\circ$  horizontally by  $\pm 45^\circ$  vertically, each containing more than 90% of the normals.

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. 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 [Bixler et al. 1967].

**Implementation Details** The one custom component in our system is a USB-programmable 80MHz Microchip PIC32 microcontroller 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 cal-



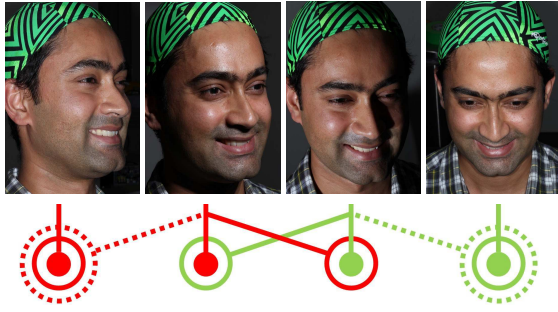
**Figure 4:** (Left) 24 images shot with the apparatus of a shiny blue plastic ball. (Right) All 24 images added together after being reprojected onto the ball's spherical shape as seen from the front, showing 24 evenly-spaced specular reflections from the six flash lighting conditions. The colored lines indicate which images correspond to each flash.



**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).

ibration object is used to focus the cameras and to geometrically calibrate the camera's intrinsic, extrinsic, and distortion parameters, with reprojection errors of below a pixel. 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 be 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.

**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 [Zickler et al. 2002], 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) respectively, we did not use them since each would either require adding additional cameras and/or lights to the system for reflectance acquisition, or not achieve reflectance separation/estimation when employing flood lighting for acquisition.



**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.*

## 4 Deriving Geometry and Reflectance

Our technique to process the photographs into an accurate 3D model plus maps of diffuse and specular reflectance proceeds as follows: We first leverage passive stereo reconstruction to build an approximate geometric mesh of the face from the photographs. We then separate the diffuse and specular components of reflectance from the acquired photographs using a novel multiview color-space analysis and further employ photometric analysis to estimate diffuse and specular photometric normals and albedo. We also employ an inverse rendering approach to estimate a per-pixel specular exponent. All reflectance 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.

### 4.1 Constructing the Base Mesh

We begin by building a base mesh using AGISoft's PhotoScan<sup>2</sup> software. Similar base mesh results can be obtained with PMVS2 [Furukawa and Ponce 2010] or Autodesk's 123D Catch<sup>3</sup> softwares. The base mesh is reconstructed with passive multiview stereo using the camera calibration from our calibration object 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; we call this the *base mesh*. We manually trim away extraneous surfaces (Fig. 7(a)), and create a minimally-distorted  $4,096 \times 4,096$  pixel  $(u, v)$  texture map space using the commercial software UnFold 3D<sup>4</sup>.

### 4.2 Diffuse-Specular Separation

We perform a per-surface-point diffuse-specular separation of the acquired photographs using a multiview color-subspace analysis. 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) for rendering with

high resolution surface details and skin reflectance. We omit views from the diffuse-specular photometric estimation where the point is occluded or in shadow, using the base mesh to compute depth maps for visibility testing.

We separate the images of the face into diffuse and specular reflectance components to facilitate recovery of the reflectance properties of the face. 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 could know the diffuse color  $\vec{d}$ , i.e., the RGB color of the subsurface scattered light at a given pixel, then it is trivial to decompose the pixel's RGB color into its diffuse and specular components. Mallick et al. [Mallick et al. 2006] proposed a color-space separation approach that uses neighboring pixel areas to infer the diffuse color which works well for relatively homogeneous dielectric materials. However, due to different amounts of melanin and hemoglobin in the face, the diffuse color varies across the face. We can however, like [Debevec et al. 2000; Weyrich et al. 2006], leverage multiple images of a surface point in our dataset for diffuse-specular separation.

Assume we are examining a point on a surface that projects into the different views in our dataset to pixel values  $\vec{p}_i = [p_r^i \ p_g^i \ p_b^i]^T$ ,  $i \in (1 \dots k)$ . Following [Zickler et al. 2008], we rotate the RGB colors into the so-called *su $v$*  color space via a simple matrix transform such that the *s* component aligns with  $\vec{s}$ , yielding  $[p_s^i \ p_u^i \ p_v^i]^T$ . Then the *chroma intensities*  $p_{uv}^i = \sqrt{p_u^i + p_v^i}$  may be 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. [Ma et al. 2007]. However, the chroma normal is unsuitable for constructing detailed surface geometry. We therefore desire a normal map constructed from *specular* reflection 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  $\vec{p}_i$ ), which is empirically designed to suppress specular highlights. With this ratio and the chroma surface normal, it is trivial to establish the diffuse albedo and to remove the diffuse component from all pixel values, leaving only specular highlights. Fig. 8 shows separation results for some example views. We then employ *Blinn-Phong* photometric stereo to extract detailed specular surface normals from the specular highlight intensities (detailed in Sec. 4.3).

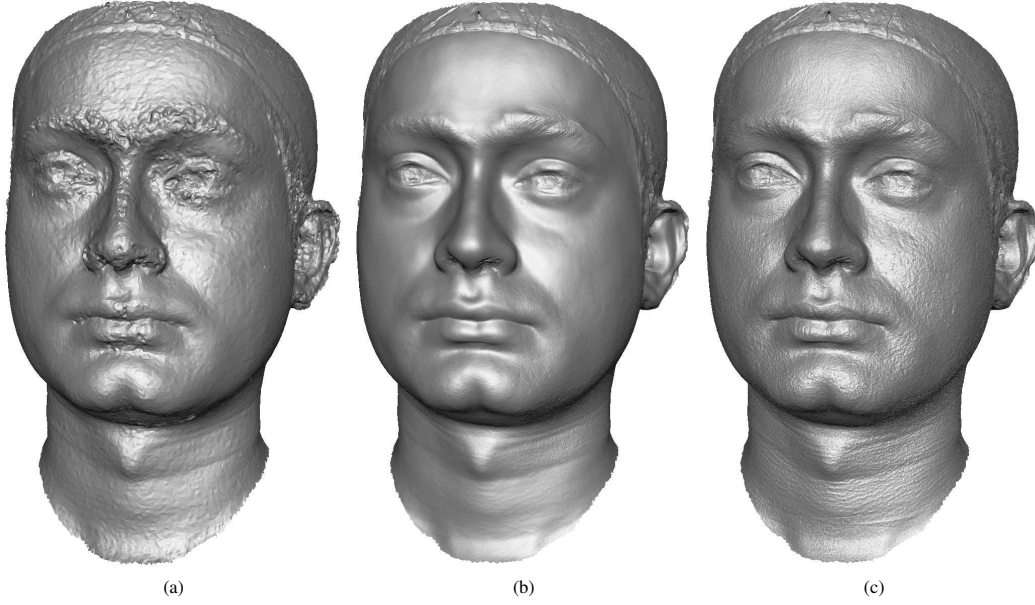
### 4.3 Diffuse and Specular Photometric Stereo

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  $\rho$  by leveraging certain assumptions about the reflectance properties of the surface using. This process is known as *photometric stereo* [Woodham 1978]. The photometric stereo equations are presented with a distant light assumption, and light intensity  $\pi$ . 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

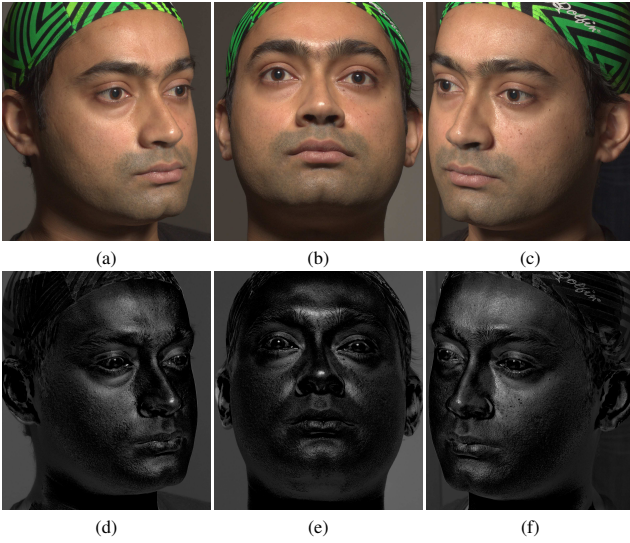
<sup>2</sup><http://www.agisoft.ru/products/photoscan/>

<sup>3</sup><http://www.123dapp.com/catch>

<sup>4</sup><http://www.polygonal-design.fr/>



**Figure 7:** (a) The base mesh from mutiview stereo. (b) Refined mesh using the diffuse (chroma) normals (c) Refined mesh using the specular reflectance analysis, which exhibits skin mesostructure details.



**Figure 8:** Diffuse-Specular separation. (a-c) Three of the 24 original photographs. (d-f) Estimated specular components.

stereo equations for exposition. In the Lambertian case, the lighting equation is  $L\vec{\beta} = P$ , where  $L = [\vec{l}_1 \vec{l}_2 \dots \vec{l}_k]^\top$ ,  $\vec{\beta} = \rho\vec{n}$ , and  $P = [p_1 p_2 \dots p_k]^\top$ . Any  $i$  with  $p_i = 0$  are omitted, as the lighting equation does not hold. The solution via pseudoinverse is:

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

In the Blinn-Phong case, the lighting equation is expressed in terms of halfway vectors  $\vec{h}_i$  instead of lighting directions, and is more complicated. The dot product has an exponent  $\alpha$ , and an associated normalization factor to conserve energy, leading to the following:

$$H\vec{\gamma} = P, \text{ where } H = [\vec{h}'_1 \vec{h}'_2 \dots \vec{h}'_k]^\top,$$

$$\vec{h}'_i = p_i^{1-1/\alpha} (f_i \frac{\alpha+8}{8})^{1/\alpha} \frac{\vec{v}_i + \vec{l}_i}{\|\vec{v}_i + \vec{l}_i\|}, \text{ and } \vec{\gamma} = \rho^{1/\alpha} \vec{n}, \quad (2)$$

with  $v_i$  the direction towards the viewer,  $\alpha$  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  $\alpha$  and  $f_i$  are known, the solution via pseudoinverse has the same form:

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

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:

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

where  $P_s$  indicates the use of the specular signal values in (2),  $P_{uv}$  represents the chroma signal values,  $\rho_{uv}$  is the chroma albedo (already computed), and  $\lambda$  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  $\alpha$  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  $t = 1 \dots 8$ ) 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.

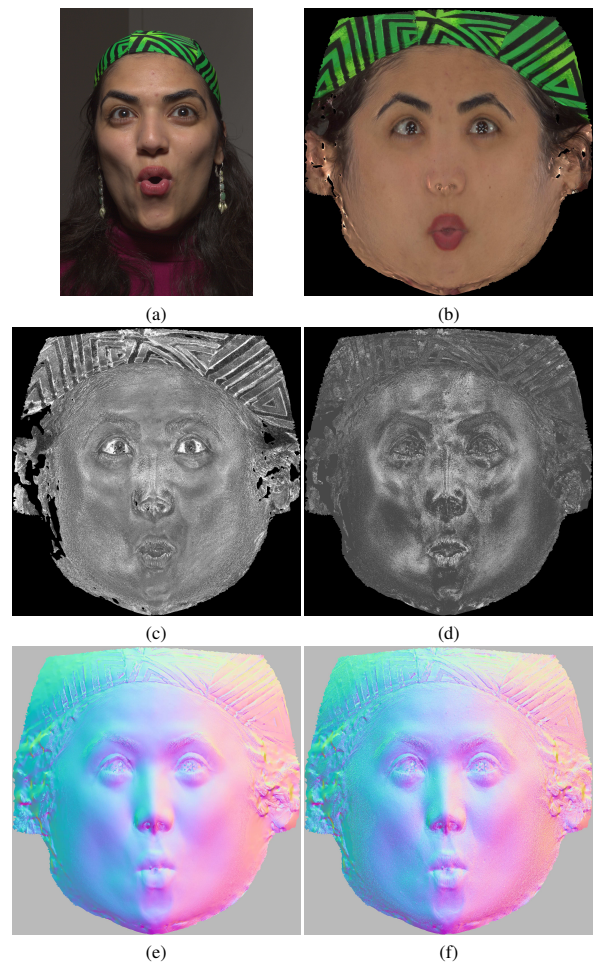
#### 4.4 Mesh Refinement

We refine the facial geometry mesh using the method of Nehab et al. [Nehab et al. 2005]. We first resample the base mesh to produce a fine mesh using a regular  $4096 \times 4096$  sampling in UV space. We then employ the *low-frequency rotation field* idea from [Nehab et al. 2005] 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 [Nehab et al. 2005] to produce the final high resolution facial mesh (Fig. 7(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.

## 5 Results

We employed our system to acquire a variety of subjects in differing facial expressions. In addition to Figure 1, Figure 9 shows the high-resolution geometry and several renderings under novel viewpoint and lighting conditions using our method. The recovered reflectance maps for one of the faces are shown in Fig. 10. Our acquisition system produces geometric quality which is competitive with more complex systems and reflectance maps not available from single-shot methods.

Fig. 11(a-c) show geometry reconstruction comparing our method (a) to the polarized spherical gradient method of [Ghosh et al. 2011] (b), and also to passive stereo reconstruction (c) using the “dark is deep” heuristic to emboss surface detail [Langer and Zucker 1994][Beeler et al. 2010]. (Please see the electronic copy for high-resolution images.) The fine scale surface mesostructure is faithfully reconstructed using our method (obtained from specular reflections) without requiring a complex setup of an LED sphere as in [Ghosh et al. 2011]. 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 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) instead exhibit a convex bump located at each follicle. Fig. 11(d-f) show renderings of the results in (a-c), using the hybrid normal technique [Ma et al. 2007] to produce a skin-like appearance. The specular component in (e, f) is a dual-lobe phong model with a manually tuned exponent, while (d) is the automatically estimated single-lobe model. The diffuse-specular separation in the passive result (f) is not measured, so specular albedo is heuristically approximated as half of the blue channel of the flat-lit photography. Further, the soft diffuse surface normal used in hybrid normal rendering is not measured photometrically in (f), so it is approximated as a blurred version of the geometry surface normal. Despite requiring manual or heuristic effort to produce the reflectance maps in (f), 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



**Figure 10:** Recovered reflectance maps: (a) sample input photograph (b) diffuse albedo (c) specular albedo (d) specular exponent  $\times 0.02$  (e) diffuse normal (f) specular normal.

as ground truth, while the uniform lobe parameters used in the polarized spherical gradient result do not capture some of 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 heuristically generated reflectance values. These results support the conclusion in [Ghosh et al. 2011] that “dark is deep” detail embossing produces hallucinated geometric detail that may have little correlation to the actual geometry of the subject. For surfaces with little albedo variation such as the undulating surfaces in [Langer and Zucker 1994] or the validation mask in [Beeler et al. 2010], the “dark is deep” heuristic is of course valid. However the skin of real subjects has natural variations in albedo that violate the “dark is deep” heuristic, and also a degree of translucency such that the heuristic is further confused by structures visible just underneath the skin such as closely shaven facial hair. We suggest that “dark is deep” results are generally well received by viewers evaluating textureless renderings of the geometry, perhaps because it is aesthetically similar to carving relief into a marble statue to mimic albedo variation. However 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 main-



**Figure 9:** Renderings of recovered geometry and reflectance maps for four subjects under novel viewpoint and lighting. The maps used to create the lower-right subject can be seen in Fig. 10.

taining a near-instant capture time competitive with the effectively instant capture time of passive stereo.

## 6 Discussion and Future Work

The results we achieve with our system suggest several avenues for future work. Currently, the surfaces of the eyes do not reconstruct well, due in part to the geometric disparity between the diffuse reflection of the iris and the specular reflection of the cornea. Detecting eyes from the sharp specularities and modeling them specifically as in [Bérard et al. 2014] would be of interest. Modeling facial hair as in [Beeler et al. 2012] 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 to bring the streams into temporal alignment with video cameras synchronized to alternating light sources. We are also clearly not yet 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 [Zickler et al. 2006], it could be possible to derive improved diffuse and specular BRDFs of the skin. Also, the shadow transitions seen in the data could be analyzed to solve for subsurface scattering parameters for certain areas of the face. Although some of our subjects wore modest

amounts of makeup without negatively impacting the reconstructions, we believe the technique would require considerable 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 which best matches the input photographs. As such, 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

The authors 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



**Figure 11:** Reconstructed geometry using: (a) Our method. (b) Polarized spherical gradient illumination [Ghosh et al. 2011]. (c) Passive stereo reconstruction using the flat-lit photographs from (b) and a “dark is deep” detail emboss. Note that (a) and (b) are often in agreement in regard to fine skin surface details, while (c) often disagrees. (d-f) Renderings using the results from (a-c) with the same view and illumination. All reflectance parameters in (d) are estimated automatically. The specular exponent in (e) is tuned manually as it is not measured. In (f), all reflectance maps are manually generated, using the flat-lit photography as a starting point. Note the uniform dimple-like appearance of (f).

Engineering Command (RDECOM). 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

- ALEXANDER, O., ROGERS, M., LAMBETH, W., CHIANG, J.-Y., MA, W.-C., WANG, C.-C., AND DEBEVEC, P. 2010. The Digital Emily Project: Achieving a photoreal digital actor. *IEEE Computer Graphics and Applications* 30 (July), 20–31.
- BEELER, T., BICKEL, B., BEARDSLEY, P., SUMNER, B., AND GROSS, M. 2010. High-quality single-shot capture of facial geometry. *ACM Trans. Graph.* 29 (July), 40:1–40:9.
- BEELER, T., BICKEL, B., NORIS, G., BEARDSLEY, P., MARSCHNER, S., SUMNER, R. W., AND GROSS, M. 2012. Coupled 3d reconstruction of sparse facial hair and skin. *ACM Trans. Graph.* 31, 4 (July), 117:1–117:10.
- BÉRARD, P., BRADLEY, D., NITTI, M., BEELER, T., AND GROSS, M. 2014. High-quality capture of eyes. *ACM Trans. Graph.* 33, 6 (Nov.).
- BIXLER, E. O., BARTLETT, N. R., AND LANSING, R. W. 1967. Latency of the blink reflex and stimulus intensity. *Perception & Psychophysics* 2, 11, 559–560.
- BRADLEY, D., HEIDRICH, W., POPA, T., AND SHEFFER, A. 2010. High resolution passive facial performance capture. *ACM Trans. Graph.* 29 (July), 41:1–41:10.
- CHEN, T., GOESELE, M., AND SEIDEL, H. P. 2006. Mesostructure from specularities. In *CVPR*, 1825–1832.
- DAVIS, J., NEHAB, D., RAMAMOORTHY, R., AND RUSINKIEWICZ, S. 2005. Spacetime stereo: A unifying framework for depth from triangulation. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 2, 296–302.



- DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, SIGGRAPH '00, 145–156.
- FURUKAWA, Y., AND PONCE, J. 2009. Dense 3D motion capture for human faces. In *Proc. of CVPR 09*.
- FURUKAWA, Y., AND PONCE, J. 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 8 (Aug.), 1362–1376.
- GEORGHIADES, A. 2003. Recovering 3-D shape and reflectance from a small number of photographs. In *Rendering Techniques*, 230–240.
- GHOSH, A., FYFFE, G., TUNWATTANAPONG, B., BUSCH, J., YU, X., AND DEBEVEC, P. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 30, 6.
- GLENCROSS, M., WARD, G. J., MELENDEZ, F., JAY, C., LIU, J., AND HUBBOLD, R. 2008. A perceptually validated model for surface depth hallucination. *ACM Trans. Graph.* 27, 3 (Aug.), 59:1–59:8.
- GOLDMAN, D. B., CURLESS, B., HERTZMANN, A., AND SEITZ, S. M. 2005. Shape and spatially-varying brdfs from photometric stereo. In *ICCV*, 341–348.
- HERNANDEZ, C., VOGIATZIS, G., BROSTOW, G. J., STENGER, B., AND CIPOLLA, R. 2007. Non-rigid photometric stereo with colored lights. In *Proc. IEEE International Conference on Computer Vision*, 1–8.
- HERTZMANN, A., AND SEITZ, S. M. 2005. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *PAMI* 27, 8, 1254–1264.
- KLAUDINY, M., HILTON, A., AND EDGE, J. 2010. High-detail 3D capture of facial performance. In *International Symposium 3D Data Processing, Visualization and Transmission (3DPVT)*.
- LANGER, M. S., AND ZUCKER, S. W. 1994. Shape-from-shading on a cloudy day. *J. Opt. Soc. Am. A* 11, 2 (Feb), 467–478.
- LUO, L., LI, H., AND RUSINKIEWICZ, S. 2013. Structure-aware hair capture. *ACM Trans. Graph.* 32, 4 (July).
- MA, W.-C., HAWKINS, T., PEERS, P., CHABERT, C.-F., WEISS, M., AND DEBEVEC, P. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Rendering Techniques*, 183–194.
- MALLICK, S. P., ZICKLER, T. E., KRIEGMAN, D. J., AND BELHUMEUR, P. N. 2005. Beyond lambert: Reconstructing specular surfaces using color. In *CVPR*.
- MALLICK, S. P., ZICKLER, T., BELHUMEUR, P. N., AND KRIEGMAN, D. J. 2006. Specularity removal in images and videos: A pde approach. In *ECCV*.
- NAYAR, S., FANG, X., AND BOULT, T. 1997. Separation of reflection components using color and polarization. *IJCV* 21, 3, 163–186.
- NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHY, R. 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM TOG* 24, 3, 536–543.
- PIGHIN, F., HECKER, J., LISCHINSKI, D., SZELISKI, R., AND SALESIN, D. H. 1998. Synthesizing realistic facial expressions from photographs. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, NY, USA, SIGGRAPH '98, 75–84.
- RAMELLA-ROMAN, J. C. 2008. Out of plane polarimetric imaging of skin: Surface and subsurface effect. In *Optical Waveguide Sensing and Imaging*, W. J. Bock, I. Gannot, and S. Tanev, Eds., NATO Science for Peace and Security Series B: Physics and Biophysics. Springer Netherlands, 259–269. "10.1007/978-1-4020-6952-9\_12".
- RUSHMEIER, H., TAUBIN, G., AND GUÉZIEC, A. 1997. Applying shape from lighting variation to bump map capture. In *Rendering Techniques*, 35–44.
- RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3D model acquisition. *ACM TOG* 21, 3, 438–446.
- VALGAERTS, L., WU, C., BRUHN, A., SEIDEL, H.-P., AND THEOBALT, C. 2012. Lightweight binocular facial performance capture under uncontrolled lighting. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2012)* 31, 6 (November), 187:1–187:11.
- WEYRICH, T., MATUSIK, W., PFISTER, H., BICKEL, B., DONNER, C., TU, C., MCANDLESS, J., LEE, J., NGAN, A., JENSEN, H. W., AND GROSS, M. 2006. Analysis of human faces using a measurement-based skin reflectance model. *ACM TOG* 25, 3, 1013–1024.
- WOODHAM, R. J. 1978. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Proc. SPIE's 22nd Annual Technical Symposium*, vol. 155.
- WU, C., VARANASI, K., LIU, Y., SEIDEL, H.-P., AND THEOBALT, C. 2011. Shading-based dynamic shape refinement from multi-view video under general illumination. In *Proceedings of the 2011 International Conference on Computer Vision, ICCV '11*, 1108–1115.
- ZHANG, S., AND HUANG, P. 2006. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering* 45, 12.
- ZHANG, L., SNAVELY, N., CURLESS, B., AND SEITZ, S. M. 2004. Spacetime faces: high resolution capture for modeling and animation. *ACM TOG* 23, 3, 548–558.
- ZICKLER, T. E., BELHUMEUR, P. N., AND KRIEGMAN, D. J. 2002. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *Int. J. Comput. Vision* 49, 2-3, 215–227.
- ZICKLER, T., RAMAMOORTHY, R., ENRIQUE, S., AND BELHUMEUR, P. N. 2006. Reflectance sharing: Predicting appearance from a sparse set of images of a known shape. *PAMI* 28, 8, 1287–1302.
- ZICKLER, T., MALLICK, S. P., KRIEGMAN, D. J., AND BELHUMEUR, P. N. 2008. Color subspaces as photometric invariants. *Int. J. Comput. Vision* 79, 1 (Aug.), 13–30.



**Figure 12:** (a) Ground-truth photograph with frontal illumination. (b-d) synthetic renderings with the same view and illumination using (b) our proposed method, (c) polarized spherical gradient illumination [Ghosh et al. 2011], 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). Note that the fine-scale details in the specular highlights on the skin in (b) and (c) are in general agreement with ground truth, while the highlight details in (d) differ significantly.