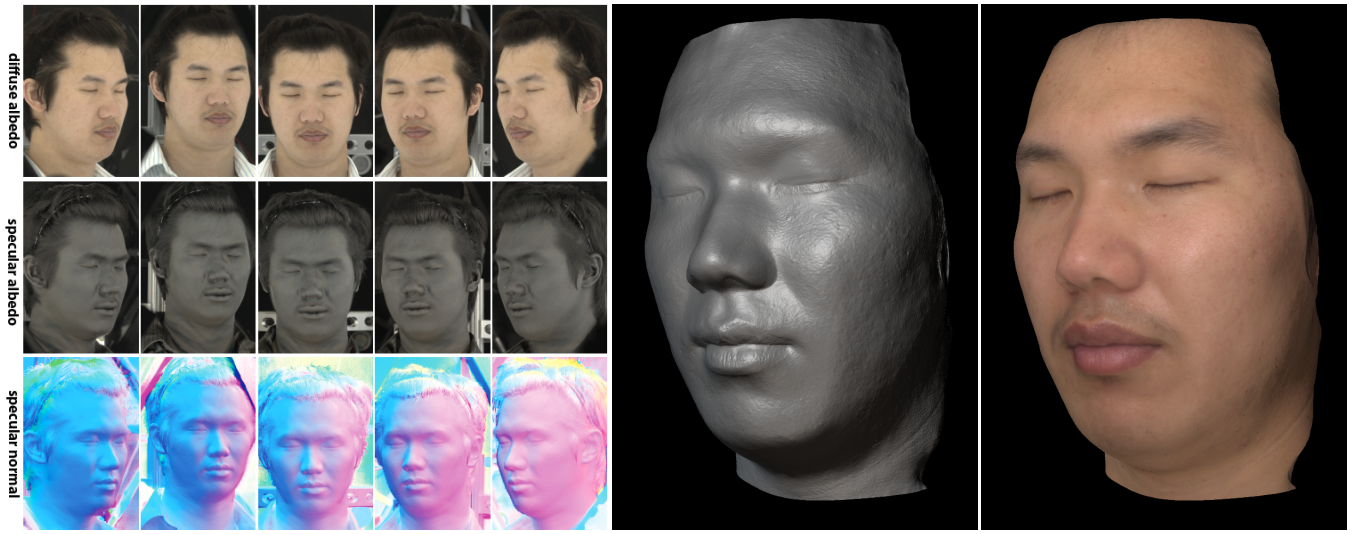
To appear in ACM TOG 30(6).

Multiview Face Capture using Polarized Spherical Gradient Illumination

Abhijeet Ghosh Graham Fyffe Borom Tunwattanapong Jay Busch Xueming Yu Paul Debevec
USC Institute for Creative Technologies



(a) Acquired data (b) Geometry (c) Rendering

Figure 1: *Multiview face capture using polarized spherical gradient illumination. (a) Acquired data from five viewpoints used for stereo reconstruction. (b) Reconstructed geometry. (c) Hybrid normal rendering* [Ma et al. 2007].

Abstract

We present a novel process for acquiring detailed facial geometry with high resolution diffuse and specular photometric information from multiple viewpoints using polarized spherical gradient illumination. Key to our method is a new pair of linearly polarized lighting patterns which enables *multiview* diffuse-specular separation under a given spherical illumination condition from just two photographs. The patterns – one following lines of latitude and one following lines of longitude – allow the use of fixed linear polarizers in front of the cameras, enabling more efficient acquisition of diffuse and specular albedo and normal maps from multiple viewpoints. In a second step, we employ these albedo and normal maps as input to a novel multi-resolution adaptive domain message passing stereo reconstruction algorithm to create high resolution facial geometry. To do this, we formulate the stereo reconstruction from multiple cameras in a commonly parameterized domain for multiview reconstruction. We show competitive results consisting of high-resolution facial geometry with relightable reflectance maps using five DSLR cameras. Our technique scales well for multiview acquisition without requiring specialized camera systems for sensing multiple polarization states.

Keywords: computational illumination, face capture, polarization, message passing.

1 Introduction

Digitally reproducing the shape and appearance of real-world subjects is a long-standing goal of computer graphics. In particular, the realistic reproduction of human faces has received increasing attention in recent years. Some of the best techniques use a combination of 3D scanning and photography under different lighting conditions to acquire models of a subject's shape and reflectance. When both of these characteristics are measured, the models can be used to faithfully render how the object would look from any viewpoint, reflecting the light of any environment. An ideal process would accurately model the subject's shape and reflectance with just a few photographs. However, in practice, significant compromises are typically made between the accuracy of the geometry and reflectance model and the amount of data which must be acquired. Ma et al. [2007] introduced polarized spherical gradient illumination for efficiently acquiring diffuse and specular photometric information and employed it in conjunction with structured light scanning to obtain high resolution scans of faces. In addition to the detail in the reconstructed 3D geometry, the photometric data acquired with this technique can be used for realistic rendering in either real-time or offline contexts. However, the technique has significant limitations. Chiefly, Ma et al.'s linear polarization pattern is effective only for the frontal camera viewpoint, forcing the subject to be moved to different positions to scan more than the front of the face. Also, Ma et al.'s lighting patterns require rapidly flipping a polarizer in front of the camera using custom hardware in order to observe both cross- and parallel-polarization states. Finally, Ma et al.'s reliance on structured light for base geometry acquisition adds scanning time and system complexity, while further restricting the process to single-viewpoint scanning.

In order to overcome the viewpoint restriction imposed by active illumination, recent work [Beeler et al. 2010; Bradley et al. 2010] has used advanced multiview stereo (MVS) to derive geometry from several high-resolution cameras under diffuse illumination. While the geometric detail derived by Bradley et al's dynamic system is not at the level of skin mesostructure, [Beeler et al. 2010] infers additional detail through a "dark-is-deep" interpretation of the diffuse shading, producing geometric detail correlating to skin pores and creases. These techniques are notable since just a single set of simultaneous photographs suffices as input, allowing even ephemeral poses to be recorded. However, the techniques are limited in that they record only a diffuse texture map to generate renderings rather than separated reflectance components, and the geometric detail inferrable from diffuse shading can vary significantly from the true surface detail which is more directly evidenced in specular reflections. Also, the single-shot nature of these techniques is not re-

quired for acquiring most facial expressions, as subjects can typically maintain the standard facial expressions used in building facial animation rigs for the handful of seconds required for multi-shot techniques [Alexander et al. 2010]. To make our multi-shot capture robust to subject motion, we leverage the joint optical flow technique of [Wilson et al. 2010].

Our work generalizes polarized spherical gradient illumination techniques to multiview acquisition and yields high quality facial scans including diffuse and specular reflectance albedo and normal maps. Specifically, our new pair of linearly polarized spherical illumination patterns enable camera placement anywhere near the equator of a subject while providing high quality diffuse-specular separation. Additionally, the technique only requires fixed static polarizers on the cameras enabling it to scale well for multiview acquisition. We then simultaneously leverage both the diffuse and specular photometric data in a novel multi-resolution adaptive-domain message passing stereo algorithm to reconstruct high resolution facial scans. We demonstrate the practicality of the proposed technique using data simultaneously acquired from five viewpoints.

In summary, our principal contributions are:

- A new polarized spherical gradient illumination technique which enables multiview face scanning.
- A demonstration of multiview acquisition employing low-cost, static polarizers on both the cameras and light sources.
- A novel multi-resolution adaptive domain message passing stereo reconstruction algorithm which uses diffuse and specular albedo *and* normal maps for high quality facial geometry reconstruction.

2 Related Work

3D Facial Capture While there has been a wide body of work on 3D scanning of objects, we focus our discussion on scanning of human faces due to the specific challenges in obtaining high-quality geometry and reflectance information. There exist techniques for high resolution scanning of static facial expressions based on laser scanning a plaster cast, such as the scans performed by XYZRGB, Inc. However, such techniques are not well suited for scanning faces in non-neutral expressions and do not capture reflectance maps. Several real-time 3D scanning systems exist that are able to capture dynamic facial performances. These methods either rely on structured light [Rusinkiewicz et al. 2002; Zhang et al. 2004; Davis et al. 2005; Zhang and Huang 2006], unstructured painted face texture [Furukawa and Ponce 2009], or use photometric stereo [Wenger et al. 2005; Malzbender et al. 2006; Hernandez et al. 2007; Klaudiny et al. 2010]. However, these prior methods are limited: either they do not provide sufficient resolution to model facial details, they assume uniform albedo, or they are data-intensive. Bickel et al. [2007] take an alternate approach by first acquiring a detailed static scan of the face including reflectance data, augmenting it with traditional marker-based facial motion-capture data for large scale deformation, and integrate high resolution video data for medium scale expressive wrinkles. Recently, passive multiview face scanning systems have been proposed which exploit detail in the observed skin texture under diffuse illumination in order to reconstruct high resolution face scans [Beeler et al. 2010; Bradley et al. 2010]. While achieving impressive qualitative results for geometry reconstruction, these techniques rely on synthesis of mesoscopic detail from skin texture that may differ from true surface detail. Furthermore, these techniques do not capture specular reflectance maps which are useful for realistic rendering. At the other end of the spectrum, researchers have employed dense lighting and viewpoint measurements in order to capture detailed spatially varying facial reflectance [Debevec et al. 2000; Weyrich et al. 2006]. However, such techniques are data intensive and do not scale well for scanning of non-neutral facial expressions and dynamic facial performances.

Spherical Gradient Illumination Ma et al. [2007] introduced a technique for efficient high resolution face scanning of static expressions based on photometric surface normals computed from spherical gradient illumination patterns. They capture separate photometric albedo and normal maps for specular (surface) and diffuse (subsurface) reflection by employing polarization of incident lighting. Photometric normals – in particular the detailed specular normals – are used to add fine-scale detail to base geometry obtained from structured light as in [Nehab et al. 2005]. However, Ma et al.'s linear polarization pattern limits the acquisition to a single viewpoint providing limited coverage of the scanned subject. Subsequent work has extended the technique for capture of dynamic facial performance using high speed photography [Ma et al. 2008], as well as moderate acquisition rates using joint photometric alignment of complementary gradients [Wilson et al. 2010]. Recently, Fyffe et al. [2011] have applied the technique for acquiring facial performance from multiple viewpoints. However, the technique is limited to acquiring unpolarized data for viewpoint independence and employing heuristic post-processing for diffuse-specular separation. Similar to the approach of Fyffe et al., we employ a message passing based stereo reconstruction algorithm in this work. However, our proposed multi-resolution adaptive domain algorithm in conjunction with the acquired polarized data results in better reconstruction of surface detail. Ghosh et al. [2010] proposed view independent separation of diffuse and specular reflectance by measuring the Stokes parameters of circularly polarized spherical illumination. However, this technique requires four measurements per spherical lighting condition with a set of different linear and circular polarizers in front of the camera in order to compute the Stokes parameters and hence does not scale well for multiview acquistion of live subjects. In this work, we extend the polarized spherical gradient illumination technique for multivew face capture. In contrast to previous work, we propose a novel polarization technique that enables effective diffuse-specular separation for multiview acquisition in just *two* photographs for a given spherical lighting condition while *not* requiring special hardware in front of the camera to image multiple polarization states. Furthermore, our stereo reconstruction algorithm takes advantage of the acquired diffuse and specular albedo and normal maps from multiple viewpoints to reconstruct high resolution detailed facial scans.

3 Multiview Acquisition

In this section we describe our spherical polarization patterns and acquisition setup for multiview face capture.

3.1 Polarization Pattern

We propose a novel pair of *lat-long* polarized lighting patterns which allow multiview diffuse-specular separation under spherical illumination in just two photographs. The patterns are linearly polarized and locally orthogonal on the sphere, one following the horizontal lines of latitude (Figure 2, red) and one following vertical lines of longitude (Figure 2, blue), and each is symmetric about the up and down Y-axis. This symmetry allows measurement from any viewpoint around the equator of the sphere near the XZ plane. However, we emphasize that the usefulness these multiview patterns is restricted to viewpoints near the equator, making them less useful for capturing a subject from above or below. We also note that the diffuse-specular separation achieved by the lat-long patterns is slightly degraded compared to the optimal (but view-dependent) pattern of Ma et al. [2007]. Nonetheless, we show that in practice, the lat-long patterns very effectively record the most important range of viewpoints and surface orientations for

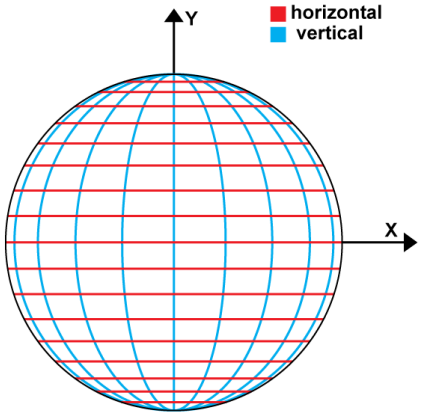


Figure 2: *Lines of latitude-longitude (lat-long) linear polarization patterns for multiview acquisition.*

multiview facial capture. In Figure 3, we compare a symmetric view-independent measurement obtained with circular polarization (column a) to that obtained with the proposed lat-long multiview polarization solution (column b). In order to better understand the comparison, we show simulated results for a perfectly specular sphere as well as real measurements of a plastic orange for both parallel-polarized (top) and cross-polarized (bottom) states. The simulations were generated according to Mueller calculus of polarized specular reflection [Ghosh et al. 2010] and employ a microfacet BRDF model that includes Fresnel reflectance. As can be seen, the cross-polarized state of circular polarization results in specular cancellation in the center of the sphere but strong specular reflections due to the opposite chirality of the reflected circularly polarized light beyond the Brewster angle. As such, Ghosh et al.'s technique requires *four* photographs (to measure complete Stokes parameters) for proper diffuse-specular separation under circularly polarized spherical lighting. In comparison, our lat-long linear polarization patterns, viewed with a fixed vertical linear polarizer on the camera (column b), result in high-quality diffuse-specular separation in just two photographs. Placing the cameras' polarizers vertically is important: since the lat-long patterns are symmetric only about the Y-axis, viewing them through a horizontal linear polarizer yields poor diffuse-specular separation (column c).

Since both of our patterns are symmetric about the Y-axis, one can also consider the multiview diffuse-specular separation achievable when employing only *one* of these patterns. Using just one pattern, we can obtain approximately cross- and parallel-polarized states by flipping a linear polarizer in front of the camera as in [Ma et al. 2007]. If we employ just the longitudinal pattern, we obtain the parallel polarization state of Figure 3, (b) with good specular signal over most of the sphere. However, when we flip the polarizer on the camera to horizontal we obtain the cross polarized state of Figure 3, (c) with poor specular cancellation. Conversely, if we employ just the latitudinal pattern then we obtain good specular cancelation with a vertical polarizer on the camera (the cross-polarized state of Figure 3, (b)). However, flipping the polarizer on the camera to horizontal shows a loss of specular signal close to the Brewster angle as seen in the parallel-polarized state of Figure 3, (c). Instead, when we employ *both* the longitudinal and latitudinal patterns (with fixed vertical polarizers on the cameras), we obtain the best specular cancelation in the cross-polarized state and the strongest specular signal in the parallel-polarized state (Figure 3, (b)). The proposed lat-long polarization patterns have two implementation advantages as well. The first is that they require only a static (vertical) linear polarizer on each camera to observe both cross-polarized and parallel-polarized states. Secondly, the regular

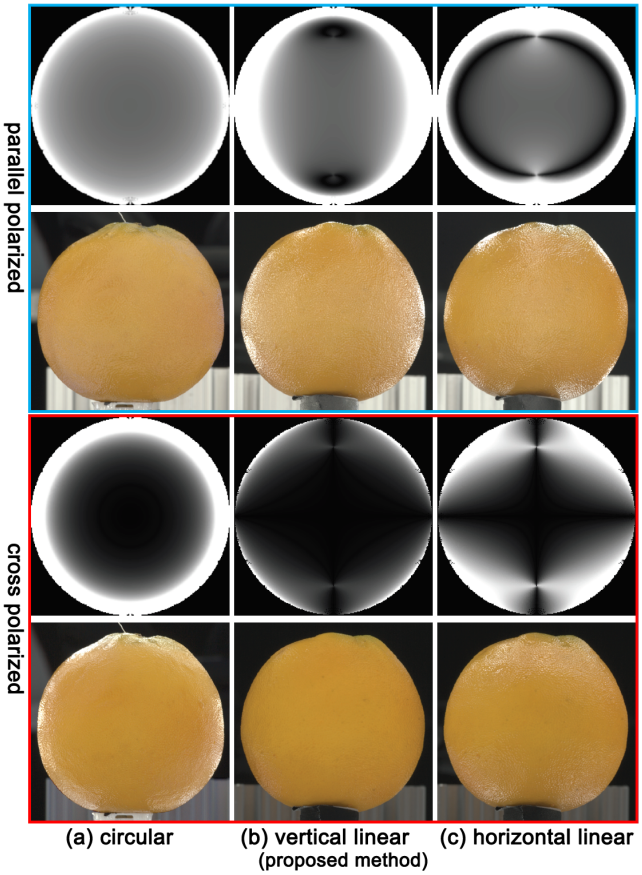


Figure 3: *Polarization-based multiview diffuse-specular separation. Rows one and three: Simulated polarized reflectance on a specular sphere. Rows two and four: Measured data of a plastic orange. Top two rows: Parallel polarization state with diffuse + specular. Bottom two rows: Cross polarization state with specular cancellation. (a) Circular polarization. (b) Proposed lines of lat-long linear polarization patterns with a vertical linear polarizer in front of the camera. (c) Lines of lat-long linear polarization patterns with a horizontal linear polarizer in front of the camera. Note that although the circular polarization separation is symmetric over the entire sphere (a), the proposed linear lines of lat-long provides a cleaner separation of reflected directions for a camera placed around the equator (b). However, the linear lines of lat-long are only symmetric about the Y-axis and hence rotating the linear polarizer in front of the camera to horizontal has a different result with poor diffuse-specular separation (c).*

grid structure of the polarization patterns makes it much simpler to mount polarizers on the lights without tuning polarizer orientations to cancel the reflections of a calibration object. In the next section, we describe a practical realization of these proposed polarization patterns for multiview face scanning.

3.2 Setup and Acquisition

Our setup for multiview face scanning consists of an LED sphere with 156 individually controllable lights that allow us to illuminate a subject with spherical illumination including the 1st-order gradient patterns of Ma et al. [2007] for obtaining surface normal estimates. We use five Canon 1D Mark III digital SLRs cameras operating in "burst" mode to rapidly acquire polarized gradient illumination data from multiple viewpoints as shown in Figure 4. We place fixed vertical linear polarizers on the camera lenses and illuminate the subject with spherical gradient illumination, switch-

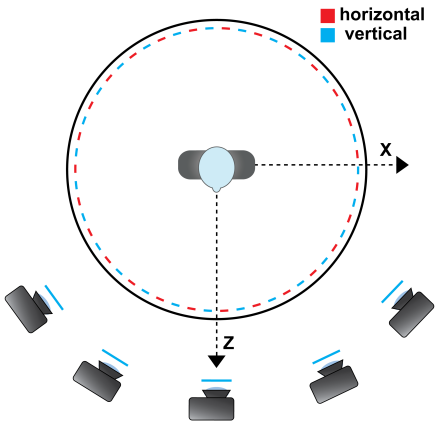


Figure 4: *Acquisition setup for multiview face scanning.*

ing between the latitudinal and longitudinal patterns using the LED sphere. To realize both patterns on one LED sphere, we partition the LED lights into two interleaved banks, one with vertical linear polarizers and one with horizontal. We take advantage of the low frequency nature of the spherical gradient illumination as the reflected light towards the camera integrates over the gradients covered by the diffuse and specular lobes of the surface BRDF. With this measurement setup, we rapidly capture a subject under the complementary spherical gradient illumination conditions of [Wilson et al. 2010] which are relatively robust to subject motion. Our cameras record the complemetary gradients in two polarization states in slightly over three seconds. From these measurements, we obtain diffuse and specular albedo and normal maps from multiple viewpoints (Figure 1, (a)). Figure 5, compares the quality of data acquired with the lat-long polarization patterns in our setup with those obtained with the view-dependent polarization pattern of Ma et al. [2007] and its alternative circular polarization approach. As can be seen, circular polarization suffers from specular pollution in the diffuse albedo and poor signal strength in the specular reflection around the sides of the face corresponding to Brewster angle (center-row). In contrast, the lat-long polarization patterns result in diffuse and specular albedo and normal maps comparable in quality to those obtained by the view-dependent linear polarization pattern of Ma et al., with the added advantage of multiview acquisition.

Achieving Photometric Consistency across Viewpoint The diffuse and specular albedo and specular normals acquired in this manner from multiple viewpoints are thereafter used as input to a stereo reconstruction algorithm (described in Section 4). The diffuse albedo map is view-independent and hence a suitable input to stereo matching. We compute the specular normal maps in world coordinates to make them suitable for stereo as well. Finally, the specular albedo maps exhibit view-dependent Fresnel gain toward grazing angles (Figure 6, (a)). We compensate for the Fresnel gain to make the specular albedo maps less view-dependent and more suitable for stereo. We do this using a data-driven procedure as follows: we first build a 1D curve of the observed Fresnel gain from a single viewpoint by averaging the observed intensities over the face as a function of the angle θ between the estimated specular normal N at a surface point and the camera view direction V (Figure 6, (c)). In a second step, we employ this 1D curve to scale the observed intensity at a surface point with a known surface orientation to that observed at $\theta = 0$ in order to obtain a view-independent specular albedo map (Figure 6, (b)). We apply the same 1D Frensel curve, built from data from a single viewpoint, to specular albedo maps captured from all camera viewpoints to obtain Fresnel-compensated albedo maps for stereo matching (Figure 1, (a)). Fresnel reflectance depends in principle on the index of refraction of the surface (skin).

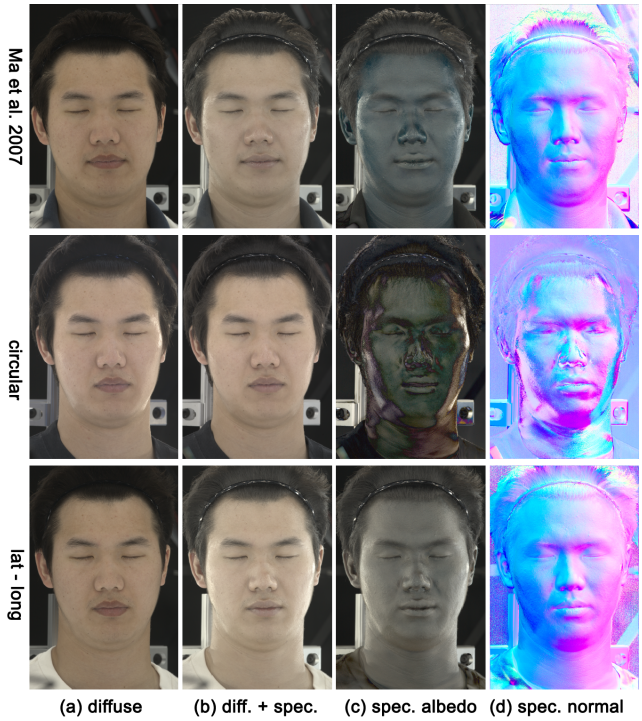
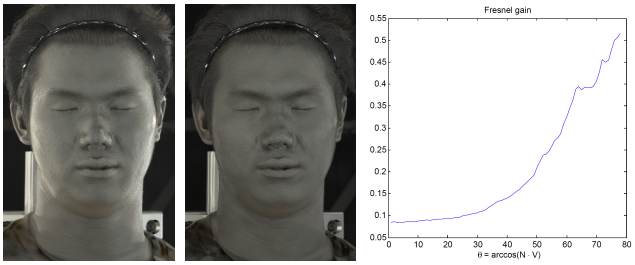


Figure 5: *Diffuse-specular separation comparison on a face. Top row: View-dependent separation using the linear polarization pattern of* [Ma et al. 2007]. *Center row: Multiview separation using circular polarization. Note the specular pollution in diffuse and poor specular signal-to-noise ratio close to the Brewster angle at the sides of the face. Bottom row: Multiview separation using the proposed linear lat-long polarization patterns. Note the clean separation and good specular signal-to-noise ratio over the entire face.*



(a) spec. albedo (b) Fresnel compensated (c) Measured Fresnel gain

Figure 6: *Data-driven Fresnel compensation from specular albedo. (a) Captured specular albedo map. (b) Specular albedo map after factoring out the measured Fresnel gain. (c) The measured view-dependent Fresnel gain (as a function of $N \cdot V$) in specular albedo map used for Fresnel compensation.*

Our approach of averaging the Fresnel curve across a face is motivated by the fact that recent measurements found very little spatial variation in the index of refraction across a face [Ghosh et al. 2010].

4 Geometry Reconstruction

Our multiview geometry reconstruction algorithm takes the acquired diffuse and specular albedo and normal maps from each of the cameras and derives a high resolution face mesh. We calibrate our cameras using the technique of Zhang [2000] in a common coordinate system for stereo reconstruction.

4.1 Stereo Reconstruction

Stereo reconstruction methods typically compute a depth map defined in the image spaces of the cameras used for acquisition. In multiview acquisition setups, this is usually followed by merging multiple depth maps into a single mesh. Further refinement may then be performed using the merged mesh as a base. We instead take an approach that requires no merging, and no separate refinement step. Similar to Fyffe et al. [2011], we represent facial geometry using a cylinder as a base surface plus a displacement map, where the displacement vectors point away from the cylinder axis. However, we compute a *single* mesh directly in the cylindrical parameterization domain, which eliminates the need for merging multiple depth maps. The cylindrical displacement map X is computed to minimize the following graphical cost function:

$$E(X) = \sum_{s \in \mathcal{V}} \phi_s(x_s) + \sum_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t), \tag{1}$$

where $\mathcal{V}$ is the set of all pixel sites in the displacement map, $\mathcal{E}$ is the set of edges connecting neighboring sites, x_s is the displacement (distance from cylinder axis) at site s, and ϕ_s, ψ_{st} are the data term and smoothing term, respectively (detailed in the paragraphs following). We employ the measured diffuse albedo map (3-channels), the (Fresnel corrected) specular albedo map (1-channel), and the photometric (specular) surface normal (3-channels) in the data term, while also employing the photometric normal in the smoothing term. These terms also make use of a visibility estimate.

Data Term Our data term is a weighted average of normalized cross correlation costs (NCC) over all pairs of neighboring cameras i, j. We use $(1 - NCC)/2$ as the cost [Beeler et al. 2010] over a 3×3-sample window centered at the point p corresponding to the cylinder coordinate (s, x_s). We estimate a photometric surface normal as a weighted blend of the normals seen by each camera: $n_{ij} = (w_i n_i + w_j n_j)/|w_i n_i + w_j n_j|$, where n_i is the photometric normal seen at p in camera i, v_i is the view vector directed towards camera i, and $w_i = (n_i \cdot v_i)$ if p is visible to camera i (determined by the current visibility estimate) and 0 otherwise. We constrain the sample window in 3D to be perpendicular to n_{ij} (and as upright as possible), yielding samples that are roughly tangent to the surface. To avoid aliasing due to differences in foreshortening, we adjust the sample spacing on a per-camera-pair basis such that the projected samples are separated by roughly one pixel on both cameras in the pair. We sum the NCC cost over all data channels c in diffuse albedo, specular albedo, and specular normal. This provides increased localization compared to other works that use only surface color. The overall weight for the pair of cameras i, j is $w_{ij} = (w_i w_j (n_i \cdot n_j))^2$. The final data term is:

$$\phi_s(x_s) = \frac{\sum_{ij} w_{ij} \sum_c (1 - NCC_{ij;c}(p))/2}{\sum_{ij} w_{ij}}. \tag{2}$$

Smoothing Term First-order smoothing terms in stereo reconstruction favor piecewise-constant depth maps, since only constant-depth surfaces are admitted without penalty. Second-order smoothing terms allow for smoother geometry estimates since they admit any planar surface without penalty [Woodford et al. 2009], but are more difficult to optimize. Fyffe et al. [2011] propose a first-order term based on photometric surface normals, which eliminates the piecewise-constant artifact, but still suffers from cracks in the geometry wherever the photometric normals are biased away from the true geometric normals. Beeler et al. [2010] approximate second-order smoothing in an iterative framework, and compute anisotropic smoothing weights to avoid over-smoothing sharp features. We combine these two techniques in our framework: our smoothing term favors neighboring points in the plane defined by the photometric surface normal, weighted by anisotropic smoothing weights which we update between each iteration of message passing:

$$\psi_{st}(x_s, x_t) = w_{st} \frac{r^2}{x_s + x_t} \min_i ((n_{i;p_s} \cdot (p_s - p_t))^2 + (n_{i;p_t} \cdot (p_s - p_t))^2) \tag{3}$$

where r is the angular resolution of the cylinder displacement map, p_s is the point corresponding to the cylinder coordinate (s, x_s), $n_{i;p_s}$ is the photometric normal seen at p_s in camera i, $w_{st} = w_{h;s} + w_{h;t}$ if sites s and t are horizontal neighbors or $w_{v;s} + w_{v;t}$ if s and t are vertical neighbors, and $w_{h;s}, w_{v;s}$ are respectively the horizontal and vertical anisotropic smoothing weights at site s. The denominator $x_s + x_t$ makes the smoothing term invariant to the distance from the cylinder axis. For anisotropic smoothing weights, we employ the gradient of the diffuse albedo and the gradient of the photometric surface normal (which we obtain by finite differences), since these are available and often correlate to surface curvature. The horizontal weights are as follows, with vertical weights likewise:

$$w_{h;s} = W \exp(-\beta_\alpha (\alpha_{s+h} - \alpha_{s-h})^2 - \beta_n (n_{s+h} - n_{s-h})^2), \tag{4}$$

where $s + h$ is the next horizontal neighbor of site s, $s - h$ is the previous horizontal neighbor of site s, α_s and n_s are respectively the diffuse albedo and photometric surface normal at site s obtained as in the texture mapping step (detailed below), and W, β_α, β_n are user-tunable parameters.

4.2 Minimization of the Cost Function

Optimization of (1) is performed using a novel adaptive domain message passing framework, which extends the tree-reweighted sequential message passing algorithm (TRW-S) [Kolmogorov 2006] to continuous-valued unknowns. Fyffe et al. [2011] propose interleaved discrete and continuous TRW-S iterations to obtain a continuous-valued result. However, that method is limited in its adaptability to the objective function, primarily because the discrete iterations use a sliding window of samples with fixed sample spacing which may not conform well to the objective function, and cannot recover from errors inherited from a lower resolution in a multiresolution framework. We instead propose the following algorithm, which we call TRW-SAD (Algorithm 1), which suffers from neither of these drawbacks. A *domain vector* d_s of possible assignments is maintained for each variable s, initialized using stratified random sampling of the continuous range of possible values. A *message vector* m_{st} or m_{ts} is maintained on each edge (s,t) of the graph, where m_{st} is a message from node s to node t and m_{ts} is a message from node t to node s. All messages are initially set to zero. A *belief vector* b_s is computed for each node s of the graph in an iterative message passing procedure. The nodes are traversed sequentially according to an ordering $i(s)$, which is reversed after each iteration. Each message passing iteration follows Algorithm 1, where T is a temperature parameter (fixed at 10 in our tests), $\gamma_{st} = 1/\max(N_{st}, N_{ts})$ and $N_{st} = |\{(s,t) \in \mathcal{E} \mid i(s) > i(t)\}|$.

Note that by virtue of the sequential nature of the algorithm, m_{st} and m_{ts} may use the same memory for storage, and b_s need not be stored at all. During the final iteration, the solution may be extracted as $x_s = d_{s;j^\star}$, where $j^\star = \arg\min_j b_{s;j}$. Every time the belief of a variable is computed, we generate a set of domain proposals according to the updated belief and the incoming messages, and compute the beliefs for these proposals as well. We then conditionally replace domain values with domain proposals with an acceptance likelihood based on the beliefs. Importantly, we never replace the domain value with the least cost (lowest b_s), so that the retained least cost solution will be fused with the proposed samples in subsequent iterations. To enable these adaptive domain updates, we delay the min-marginalization of messages until after the message has been passed, instead of before it is passed (as in TRW-S).

Algorithm 1 The proposed TRW-SAD iteration.

for all nodes $s \in \mathcal{V}$ in the order of increasing $i(s)$ **do**
  // Compute belief:
  for $j \in 1 \ldots |d_s|$ **do**
    $b_{s;j} \leftarrow \phi_s(d_{s;j}) + \sum_{(t,s)\in\mathcal{E}} \min_k (m_{ts;k} + \psi_{st}(d_{s;j}, d_{t;k}))$
  Sort d_s, b_s by increasing b_s.
  // Generate domain proposals:
  for $j \in 1 \ldots |p_s|$ **do**
    $p_{s;j} \leftarrow$ new domain proposal.
    $\beta_{s;j} \leftarrow \phi_s(p_{s;j}) + \sum_{(t,s)\in\mathcal{E}} \min_k (m_{ts;k} + \psi_{st}(p_{s;j}, d_{t;k}))$
  Sort p_s, β_s by increasing β_s.
  // Conditionally accept domain proposals:
  for $j \in 2 \ldots |d_s|$ **do**
    for $k \in 1 \ldots |p_s|$ **do**
      if $\text{random}(0,1) \leq \exp(\frac{1}{T}(b_{s;j} - \beta_{s;k}))$ **then**
        // The proposal is accepted:
        $d_{s;j} \leftarrow p_{s;k}$
        $b_{s;j} \leftarrow \beta_{s;k}$
        // Mark the proposal as used:
        $\beta_{s;j} \leftarrow \infty$
  // Update messages:
  for $(s,t) \in \mathcal{E}$ with $i(s) < i(t)$ **do**
    for $j \in 1 \ldots |d_s|$ **do**
      $m_{st;j} \leftarrow \gamma_{st} b_{s;j} - \min_k (m_{ts;k} + \psi_{st}(d_{s;j}, d_{t;k}))$
// Reverse ordering:
$i(s) \leftarrow |\mathcal{V}| + 1 - i(s)$

Proposal Generation Strategy The domain proposals generated within our algorithm are intended to sample the domain in the neighborhood of the least cost solution. To generate a set of proposals, we start with a set of *suggested* domain values S_s containing the two domain values $d_{s;1}$ and $d_{s;2}$ corresponding to the two least-cost belief values $b_{s;1}$ and $b_{s;2}$. We then add to S_s two *suggestions* from each neighboring node t: the two domain values $\arg\min_{x_s} \psi_{st}(x_s, d_{t;k_1})$ and $\arg\min_{x_s} \psi_{st}(x_s, d_{t;k_2})$ corresponding to the two least-cost message values $m_{ts;k_1}$ and $m_{ts;k_2}$, with $k_1 = \arg\min_k m_{ts;k}$ and $k_2 = \arg\min_{k\neq k_1} m_{ts;k}$. This allows our method to recover from poor sampling by encouraging samples that are consistent with neighboring nodes. Finally we add to S_s the minimum and maximum possible displacement values. This results in set S_s with up to 12 domain values for our four-connected stereo matching cost function. We then sort the values in S_s to produce a discretely sampled function $F_s(u)$ which maps values from $\{1, 2, \ldots |S_s|\}$ to domain values. We then draw continuous-valued samples $u_j \sim \text{uniform}(1, |S_s|)$ and finally evaluate $p_{s;j} = F_s(u_j)$ with linear interpolation. We find that domain proposals generated in this fashion produce acceptable results, while being faster than sampling/importance resampling on a fine grid. We also find it beneficial to not generate any domain proposals at all for the first two iterations of the algorithm, to allow the beliefs to converge somewhat before the domains begin to adapt.

Multi-Resolution Optimization To improve efficiency, we adopt a multi-resolution strategy in this work. We initialize the method by down-sampling the resolution of the input and cylindrical domain by a factor of 16 in each dimension. We then perform TRW-SAD with 16 domain samples, for 16 iterations. Then we continue with the next higher resolution (by a factor of 2) until we reach the original resolution. We initialize the TRW-SAD domain of each higher resolution using the final domain of the previous resolution, up-sampled by a factor of 2. Each higher resolution uses half as many TRW-SAD iterations and half as many domain samples as the previous resolution (but no fewer than 4), since up-sampling the previous domain provides a warm start. The domain samples are pruned by truncating the vector d_s, which is always ordered by increasing b_s. Additionally, after each resolution is processed, we use the current geometry estimate to update the visibility estimate, and apply a low-frequency correction to the photometric normals to reduce bias, related to the correction used in [Ma et al. 2007]. The entire procedure is outlined in Algorithm 2. The lack of an initial visibility estimate creates artifacts that are not easily removed in later iterations. To combat this issue, we first run our algorithm on only the two smallest resolutions to obtain a coarse visibility estimate. We then re-start the algorithm all over again, but retain the coarse visibility estimate as the initial visibility estimate.

Algorithm 2 Multi-resolution geometry estimation.

for pass p from coarsest to finest resolution **do**
  // Update resolution:
  Scale all data to resolution required for pass p.
  Up-sample (or initialize) cylindrical domain.
  Prune TRW-SAD domains to $\max(2^{4-p}, 4)$ domain samples.
  // Message passing:
  Execute $\max(2^{4-p}, 4)$ steps of TRW-SAD using (1).
  // Update result:
  Compute vertex world coordinates from displacement map.
  Update visibility estimate from geometry estimate.
  Compute geometry normals from geometry estimate.
  Correct bias in photometric normals using geometry normals.

Final Refinement The TRW-SAD algorithm becomes costly with larger resolutions, especially with terms employing normalized cross correlation. We observe that the geometry estimate obtained in the second-to-last resolution pass is close to the final result, and so we may employ a simplified cost function during the final resolution pass without significant effect on the result. The data term is simplified to a quadratic cost centered around the previous geometry estimate, with a fixed magnitude determined by a user-tunable parameter. The smoothing term is simplified by setting the surface normal to a fixed value obtained as in the texture mapping step. The simplified cost function executes an order of magnitude faster than the original cost function. The entire processing time from photographs to final face mesh with textures is one hour running on a single core of a 2008 Intel Xeon processor.

Texture Mapping We sample textures for the specular albedo, specular normal, diffuse albedo, and red, green, and blue diffuse normals by projecting the geometry vertices back into the camera views, and blending the pixel values in the corresponding photographs. The result is a set of maps in the cylindrical parameterization domain, aligned one-to-one with the geometry. We weight the contribution of each camera with the same camera weighting factor used in the data term. To avoid seams caused by weight discontinuities, we feather the weights in the cylindrical domain before computing the final blend.

5 Results

We now present some results of multiview face capture with the proposed technique. As discussed in Section 3.2, we capture a subject from five viewpoints near the equator of the LED sphere using DSLR cameras operating in burst mode. We photograph the subject under the complementary spherical gradient illumination conditions of [Wilson et al. 2010] in both cross- and parallel-polarization states and perform automatic photometric alignment of the acquired data to compensate for any subject motion during acquisition. Following the alignment step, we compute diffuse and specular albedo and normal maps from multiple viewpoints which

(a) Ma et al. (b) Our method

Figure 7: *Geometry reconstruction comparison with the view-dependent technique of [Ma et al. 2007]. (a) Structured light scanning + specular detail embossing according to the technique of Ma et al. (b) Proposed reconstruction based on the separated diffuse and specular albedo and normal maps.*

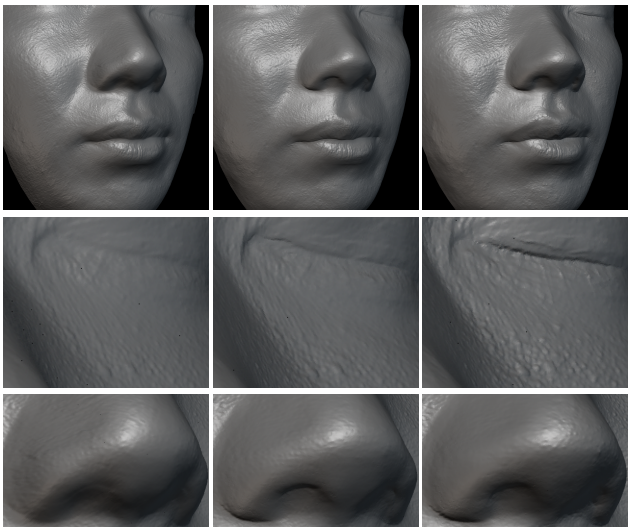
are then used as input for the message passing stereo reconstruction algorithm. Figure 1 presents the acquired data of a subject (a), as well as the result of the detailed geometry reconstruction (b) and rendering with the acquired hybrid normals as described in [Ma et al. 2007] (c). We present additional results of faces scanned in relatively extreme expressions in Figure 12. The accompanying video contains more renderings with varying viewpoint and lighting for further evaluation of the results.

Figure 7 presents a qualitative comparison of our technique for facial geometry reconstruction with the approach of Ma et al. [2007] that employs structured light scanning for base geometry reconstruction followed by embossing of specular detail according to [Nehab et al. 2005]. Here, both techniques employ a single stereo pair of cameras. As can be seen, our approach achieves very comparable high quality reconstruction without requiring structured light scanning or restricting the acquisition to a single viewpoint.

Figure 8 presents a qualitative comparison of our technique for face capture with the recent approach of Fyffe et al. [2011] which employs heuristic diffuse-specular separation of albedo and normal maps for input to a message passing stereo reconstruction algorithm. Here, we simulated unpolarized input data for comparison with Fyffe et al. (a) by adding together the parallel and cross polarized images obtained with our setup. Compared to Fyffe et al., our proposed TRW-SAD algorithm results in an improved geometry reconstruction with the same heuristic based diffuse-specular separation of albedo and normals, particularly around discontinuities such as eyelids, nostrils and lips (b). When employed in conjunction with polarization-based diffuse-specular separation, our technique results in an even more accurate reconstruction of the facial geometry with greater surface detail (c).

Figure 9 presents a comparison of the mesoscopic detail synthesized from skin texture under uniform diffuse illumination [Beeler et al. 2010] with that obtained from specular normal maps. While considerable qualitative surface detail can be derived from the texture alone, not all of this detail matches up to more directly observable surface detail obtained from specular reflection. In particular, some convexities on the surface and dark hair can be misinterpreted as concavities on the surface due to skin pores and wrinkles while some fine wrinkle detail that is captured in the specular reflection can be completely missing in the skin texture due to subsurface scattering.

Finally, we present an application of our proposed polarization technique in Figure 10 to a recently proposed passive illumination



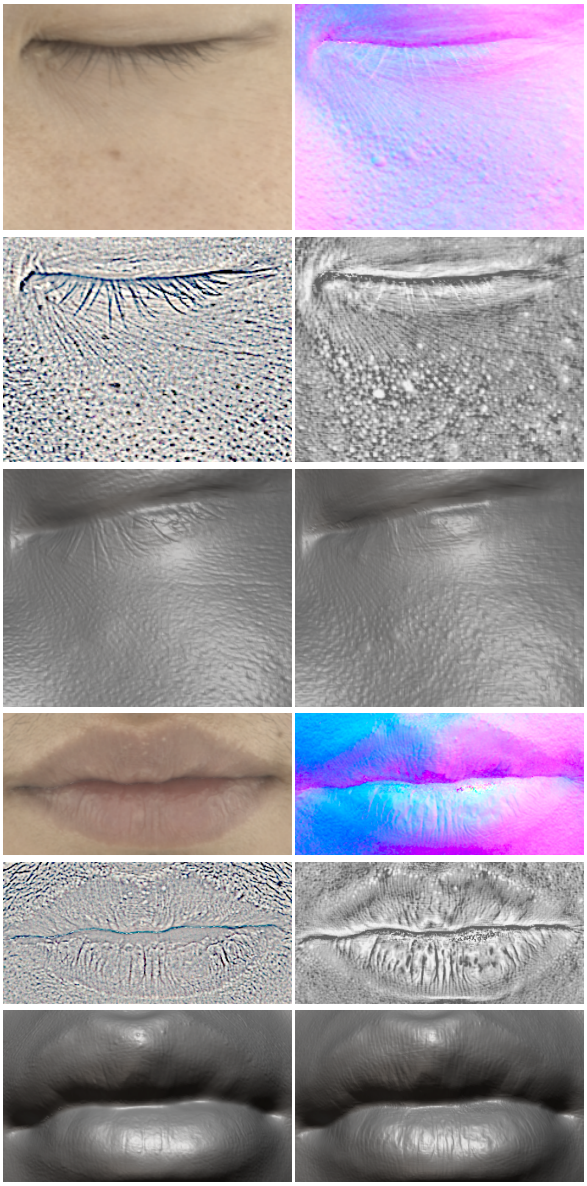
(a) Fyffe et al. (b) TRW-SAD (heuristic separation) (c) Our method (polarized separation)

Figure 8: *Geometry reconstruction comparison with the multiview technique of [Fyffe et al. 2011]. a) Reconstruction based on the technique of Fyffe et al. employing heuristic diffuse-specular separation of albedo and normals. (b) Proposed reconstruction algorithm with heuristic diffuse-specular separation. (c) Proposed reconstruction algorithm with the proposed polarization based diffuse-specular separation.*

setup for multiview face scanning [Bradley et al. 2010]. Here, we photograph a plastic macquette illuminated by nine flat light panels to create a uniform illumination condition similar to the approach of Bradley et al. When employing unpolarized lighting (a), there is significant specular reflection in the photographs that can adversely affect stereo correspondance. Bradley et al. reported having to apply specular canceling makeup to the subjects' faces in order to aid the multiview stereo which is not an ideal solution as it interferes with natural facial appearance. Instead, employing sheets of linear polarizer on the light panels oriented along the lines of latitude while mounting vertical linear polarizers on the cameras eliminates most of the undesirable specular reflections in the photographs (b).

5.1 Discussion of Limitations

Lat-long polarization Our lat-long patterns cannot achieve perfect specular cancelation in the cross-polarized state compared to the view-dependent pattern of Ma et al. [2007]. But the performance is remarkably good both visually and in simulation, canceling 99.88% of the specular reflection over the surface of the sphere, with the worst performance of only 99.63% near grazing angles. Like Ma et al., the lat-long patterns also produce attenuated specular reflections for upward- and downward-pointing surface orientations due to the Brewster angle. However, these areas are typically outside the region of interest for face scanning and hence this is not a problem in practice. The two lat-long polarization patterns were realized by partitioning the LED sphere into every-other-light banks of polarized lights, reducing the angular resolution of the illumination by half. While this still resulted in sufficient resolution for facial reflection, the lower resolution could cause worse artifacts for more specular materials. A future LED sphere could be envisioned to produce either polarization state at each light position. The lat-long patterns also result in progressive degradation in the quality of diffuse-specular separation as the camera viewpoint is moved above or below the equator (Figure 11). However, this degradation is very gradual and we have found the method to work well up to 20



(a) uniform illumination (b) specular detail

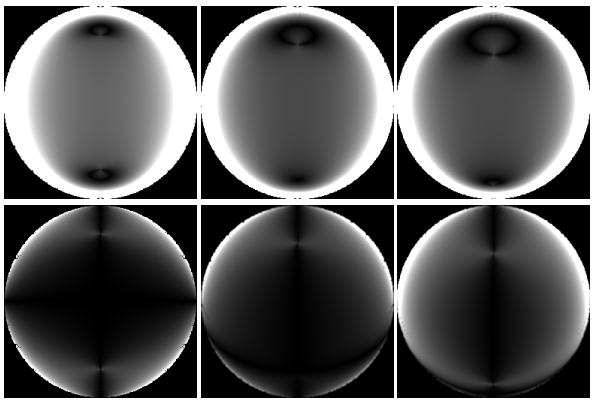
Figure 9: *Mesoscopic detail comparison for two different areas of a face. Top-three rows: upper cheek. Bottom-three rows: lips. (a) Skin texture under uniform illumination. (b) Specular normal map from polarized spherical gradients. Mesoscopic detail from skin texture (rows two and five) incorrectly classifies some convexities and dark hair as concavities, while failing to record some fine wrinkles captured in the specular normals when embossed on base geometry (rows three and six).*

degrees away from the equator; at 15 degrees, the average amount of uncancelled specular resolution is still just 1%.

Geometry reconstruction Our geometry reconstruction has some limitations that we would like to address in future work. The cylindrical parameterization domain, though convenient, is unable to represent concavities in any direction other than towards the cylinder axis, and care must be taken in choosing the cylinder axis to avoid problems under the chin. Using a template mesh may lessen this issue. Our method is also strongly dependent on photometric surface normals, and suffers wherever the surface is largely occluded from illumination, such as inside the nostrils and mouth, and also under the chin.

(a): unpolarized (b): lines of lat (c): $(a) - (b)$

Figure 10: *Lines-of-lat polarization with the setup of Bradley et al.[2010]. Top-row: frontal viewpoint. Bottom-row: side viewpoint. (a): Unpolarized illumination. (b): Light panels polarized in the lines of latitude pattern. (c): $(a) - (b)$ depicting the specular reflection that is cancelled by the proposed polarization pattern.*



(a) 0 degrees (b) 15 degrees (c) 30 degrees

Figure 11: *Simulated lat-long polarization with change in viewpoint. Top-row: Parallel-polarization. Bottom-row: Cross-polarization. (a) Camera viewing at the equator. (b) Camera viewing from 15 degree above the equator. (c) Camera viewing from 30 degree above the equator.*

6 Conclusion

We generalized the polarized spherical gradient illumination technique to multiview acquisition, demonstrating high quality facial capture and rendering with acquired diffuse and specular reflectance information. Our proposed *lat-long* polarization patterns enable high quality multiview diffuse-specular separation of spherical illumination from just two photographs, and it scales well to many cameras as it only requires low-cost fixed linear polarizers in front of cameras and light sources. We demonstrate the application of the polarization pattern to an alternate passive illumination face capture setup. The proposed polarization technique should have

applications in general multiview acquisition of real world objects and material reflectance.

We also presented a novel multiresolution adaptive domain message passing stereo reconstruction algorithm which derives detailed facial geometry from both the diffuse and specular reflectance of the face. Here, we eliminated the need for merging multiple depth maps by formulating the multiview stereo reconstruction in a common parameterization domain. In future work, it would be of interest to apply our technique to dynamic facial performances, and also to investigate other parameterization domains for stereo reconstruction of more complex shapes such as human bodies.

7 Acknowledgments

We would like to thank Jina Lee and Joel Jurik for sitting as subjects and Alex Ma, Santa Datta, Kathleen Haase, Bill Swartout, Randy Hill, and Randolph Hall for their support and assistance with this work. We also thank our anonymous reviewers for their helpful suggestions and comments. This work was sponsored in part by NSF grant IIS-1016703, the University of Southern California Office of the Provost and the U.S. Army Research, Development, and 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.

BICKEL, B., BOTSCH, M., ANGST, R., MATUSIK, W., OTADUY, M., PFISTER, H., AND GROSS, M. 2007. Multi-scale capture of facial geometry and motion. *ACM Transactions on Graphics 26*, 3, 33: 1–10.

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.

DAVIS, J., NEHAB, D., RAMAMOORTHI, R., AND RUSINKIEWICZ, S. 2005. Spacetime stereo: A unifying framework for depth from triangulation. *PAMI 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 ACM SIGGRAPH 2000*, 145–156.

FURUKAWA, Y., AND PONCE, J. 2009. Dense 3D motion capture for human faces. In *Proc. of CVPR 09*.

FYFFE, G., HAWKINS, T., WATTS, C., MA, W.-C., AND DEBEVEC, P. 2011. Comprehensive facial performance capture. *Computer Graphics Forum (Proc. EUROGRAPHICS) 30*, 2.

GHOSH, A., CHEN, T., PEERS, P., WILSON, C. A., AND DEBEVEC, P. 2010. Circularly polarized spherical illumination reflectometry. *ACM Trans. Graph. 29* (December), 162:1–162:12.

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.

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)*.

KOLMOGOROV, V. 2006. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Mach. Intell. 28* (October), 1568–1583.

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.

MA, W.-C., JONES, A., CHIANG, J.-Y., HAWKINS, T., FREDERIKSEN, S., PEERS, P., VUKOVIC, M., OUHYOUNG, M., AND DEBEVEC, P. 2008. Facial performance synthesis using deformation-driven polynomial displacement maps. *ACM TOG (Proc. SIGGRAPH Asia)*.

MALZBENDER, T., WILBURN, B., GELB, D., AND AMBRISCO, B. 2006. Surface enhancement using real-time photometric stereo and reflectance transformation. In *Rendering Techniques*, 245–250.

NEHAB, D., RUSINKIEWICZ, S., DAVIS, J., AND RAMAMOORTHI, R. 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM TOG 24*, 3, 536–543.

RUSINKIEWICZ, S., HALL-HOLT, O., AND LEVOY, M. 2002. Real-time 3D model acquisition. *ACM TOG 21*, 3, 438–446.

WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM TOG 24*, 3, 756–764.

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.

WILSON, C. A., GHOSH, A., PEERS, P., CHIANG, J.-Y., BUSCH, J., AND DEBEVEC, P. 2010. Temporal upsampling of performance geometry using photometric alignment. *ACM Trans. Graph. 29* (April), 17:1–17:11.

WOODFORD, O. J., TORR, P. H. S., REID, I. D., AND FITZGIBBON, A. W. 2009. Global stereo reconstruction under second order smoothness priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence 31*, 12, 2115–2128.

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.

ZHANG, Z. 2000. A flexible new technique for camera calibration. *PAMI 22*, 11, 1330–1334.

(a) geometry (b) rendering (c) geometry (d) rendering

Figure 12: *Reconstructed geometry (a),(c) and hybrid normal renderings (b),(d) of subjects in various non-neutral expressions.*