Real-time Geometry and Reflectance Capture for Digital Face Replacement
ICT Technical Report ICT-TR-04-2008

Andrew Jones, Jen-Yuan Chiang, Abhijeet Ghosh, Magnus Lang,
Matthias Hullin[†], Jay Busch, Paul Debevec

USC Institute for Creative Technologies

Max-Planck Institut für Informatik[†]

December 19, 2008

1 Introduction

In this work, we develop a real-time geometry capture approach to digital face replacement for a dynamic performance. Digital face replacement has major applications in visual effects for motion pictures as well as interactive applications such as video games and simulation and training environments. Our approach looks into extending the 3D face scanning technology developed at the ICT Graphics Lab [Ma 2008] to support seamless face replacement along with separated diffuse and specular albedo textures and surface normals for high quality post-production relighting of the captured performance (see Figure 1). Such an approach goes beyond the traditional scope of face replacement techniques that are either completely image based and hence view-dependent or typically capture a performance under a fixed lighting condition and hence cannot be relit for usage in other performances [Zhang and Yau 2006].

2 Method

Our approach builds upon the high resolution 3D face scanning technology developed at the ICT Graphics Lab [Jones et al. 2006, Ma et al. 2007, Ma 2008, Ma et al. 2008] for capturing dynamic facial performance. This includes capture of high resolution performance geometry and textures with high speed cameras (stereo pair of Phantom V10) and active illumination using a high speed MULE projector for structured light projection for stereo geometry reconstruction and spherical gradient illumination using

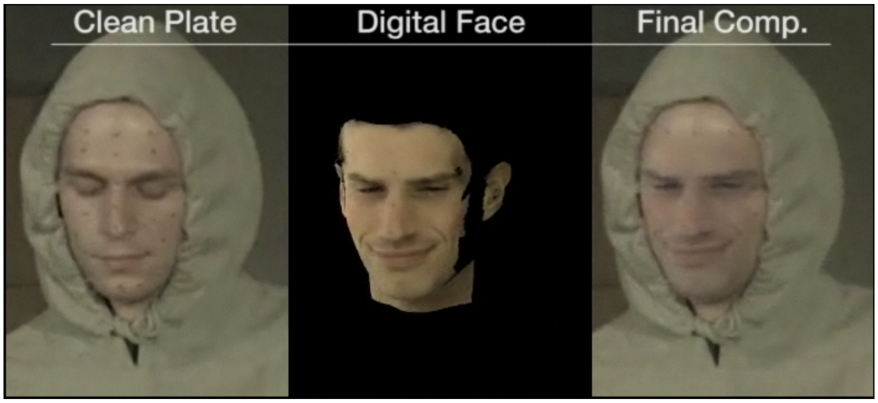

Figure 1: Digital face replacement based on real-time performance capture. Left: Recorded background plate. Center: Captured performance geometry rendered under novel lighting condition. Right: Captured performance geometry composited on the background plate with viewpoint tracking. Here, the digital face is rendered with illumination from a light probe corresponding to the background plate.

Light Stage 5 (see Figure 2). Furthermore, our approach involves tracking of the facial pose during performance in order to be able to be able to seamlessly composite the digital face on the captured background plate (e.g., a stunt performance) in a view-independent manner.

Another aspect of this work looks into techniques for obtaining separate specular and diffuse albedo textures and surface normals for high quality relighting of the captured facial performance to match the lighting of a given background plate. While the high resolution 3D face scanning technique for static expressions exploits polarization of light for this purpose [Ma et al. 2007], such a technique cannot be used while capturing a dynamic performance due to limitations in the hardware setup [Ma 2008]. Thus, we look into alternative separation techniques of diffuse and specular reflectance based on color-space analysis and computing separate diffuse and specular normals based on such separation as a post-process.

3 Tracking and Alignment

In order to align the dynamic geometry of the subject with the dynamic background plate, we perform 3D tracking based on a small set of painted marker points on both the target subject and the background plate actor. Using the tracking data of the subject, we stabilize the performance by inverting the rigid movement. The stabilized geometry is then transformed dynamically to the 3D face position in the background plate.

Camera Calibration: Two stereo camera pairs are involved in the acquisition process:

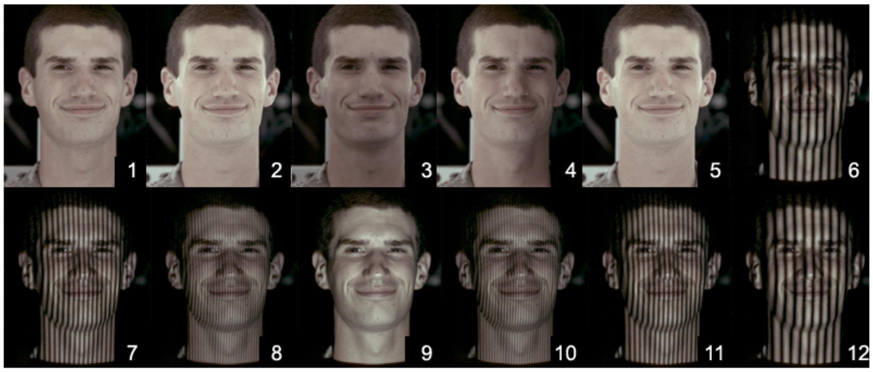

Figure 2: Sequence of images captured of the dynamic performance with our performance capture system at 24 Hz. The projected illumination conditions include structured light patterns (6 - 12) for 3D reconstruction of the geometry and gradient illumination patterns (1 - 5) for obtaining photometric surface normals and albedo textures for relighting.

one for the subjects performance and one for the background plate. In order to track both the performances, we need to calibrate each pair. For this purpose, we acquire a set of stereo images of a planar checkerboard pattern in various orientations and recover the intrinsic and extrinsic parameters of each camera using the technique of Zhang [Zhang 2000].

3D Tracking of Markers: In order to track the faces of the target subject and the background plate actor during the dynamic performances, we paint a sparse set of approximately 20 small marker points on the faces. For convinience of tracking, we ensure that at least one marker (the pivot point) is common to both performances and located in an area that deforms as little as possible. In our case, we place the pivot point on the ridge of the nose. After the performances have been recorded, we manually select a number of markers that can be reliably tracked automatically in both of the input videos. They are preferably chosen to be as rigidly connected as possible (e.g., on the forehead, nose ridge, or elsewhere depending on the nature of the facial performance). These marker points are automatically tracked through the input videos, yielding 3D marker coordinates via triangulation of the stereo data.

Estimating Rigid Transformation: For each frame in each sequence, we compute the translation T of the pivot point with respect to the first frame. We also determine the rotation R of the selected 3D marker set around the pivot point in a least squares sense. Although three valid marker trajectories would theoretically be sufficient to recover the rotation, using more markers yields more stable results. This process returns for each frame the translation T_s and the rotation R_s of the subject as well as the translation T_b and the rotation R_b of the background plate actor. The translations and rotations in each sequence are relative to the first frame in the respective sequences.

Alignment: The total transformation matrix that translates and rotates the subject geometry to match the background plate is then given as:

$$M_{total} = T_b R_b M_0 R_s^{-1} T_s^{-1}.$$

The constant transformation M_0 describes a one-time alignment of the stabilized replacement performance to the background plate. It can be obtained as a transformation matrix between a set of pairs of corresponding markers, or by manual alignment inside a modeling and animation software. If the actors both start in the same position and the world coordinate systems agree, then M_0 is identity.

We implemented the above transformation in multiple steps using the modeling and animation package Maya 8.0: the stabilization ($R_s^{-1} T_s^{-1}$) was baked into the subject performance, which was then positioned in Maya using a hierarchy of two locator objects corresponding to M_0 and $T_b R_b$, respectively. Since Maya does not by default support the import of transformation matrices, the rotation was converted to Euler angles.

4 Diffuse-Specular Separation

A limitation of the dynamic performance capture setup is that the observed images (see Figure 2) contain both the specular and diffuse reflection components. As mentioned in Section 2, the acquisition setup for capturing the dynamic performance cannot make use of polarization of incident light [Ma et al. 2007] to separate the two reflection components due to limitations in the existing hardware setup such as limited brightness levels and the rate at which a linear polarizer can be flipped in front of the camera [Ma 2008]. Hence, in this work we develop a technique to perfrom color-space separation of these two reflection components as a post-process to the acquired data.

We employ a dichromatic color model [Shafer 1985] based separation technique. A dichromatic color model assumption is valid for faces since skin is a dielectric medium. Under this assumption, the sensor response to the skin reflectance is given as

$$I = (D f_d(\theta) + S f_s(\theta)) \hat{n} \cdot \hat{l}, \qquad (1)$$

where D is the diffuse color of skin, S the color of the light source, f_d (apporximately) Lambertian reflectance, and f_s specular reflectance. The camera measurement I is an RGB color vector $I = [I_R, I_G, I_B]^T$, with similar definitions for S and D.

Similar to [Mallick et al. 2005], we define a data-dependent RGB to SUV rotation matrix R_{SUV} such that one of the three axes (say red) aligns with the direction of the source color S, while the other two axes align with components of the diffuse color. Applying such a transformation to Equation 1 leads to

$$I_{SUV} = (D' f_d(\theta) + S' f_s(\theta)) \hat{n} \cdot \hat{l}, \qquad (2)$$

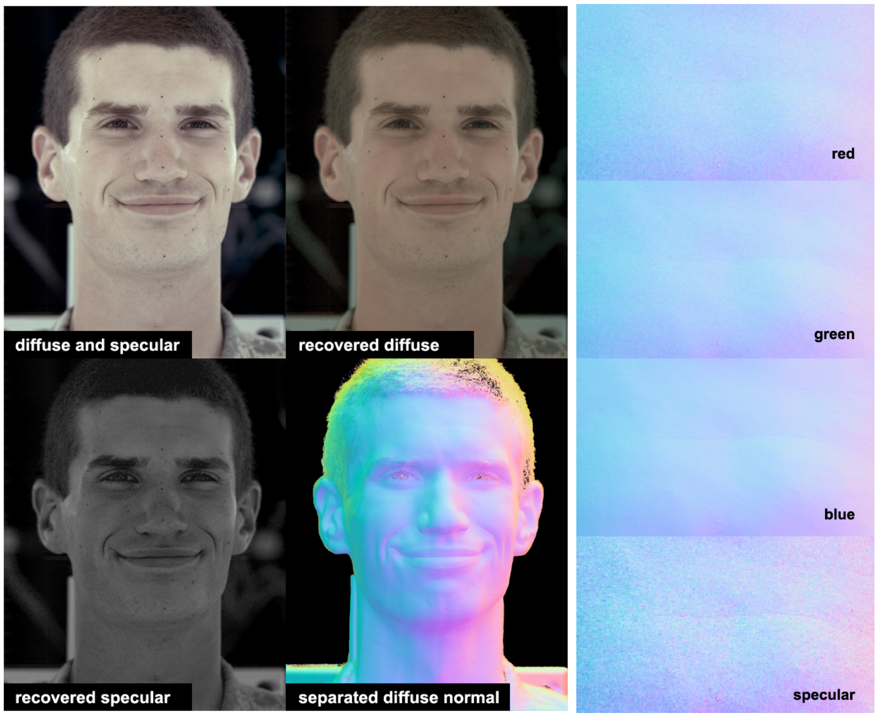

Figure 3: Color space separation of diffue and specular components of the acquired images. Clockwise from top left: captured albedo under uniform spherical illumination, separated diffuse albedo, separated diffuse normal, separated specular albedo. Right colum: separated (red, green and blue) diffuse and specular normals.

where $D' = R_{SUV}D$ and $S' = R_{SUV}S = [1, 0, 0]^T$.

The above transformation has two important properties. First, it separates the diffuse and specular reflection components using a local procedure and hence can handle textured surfaces. The second important property is that the transformation preserves shading information and is hence suitable for photometric stereo. Hence, we employ this transormation for computing separate diffuse and specular normals and albedo textures from the acquired data.

One limitation of the dichromatic model based separation technique is that it performs poorly if the diffuse and specular color are similar. In our case, this is true for light colored skin observed under spherical gradient illumination. Hence, instead of directly transforming the acquired data into SUV color space, we first perfrom the following operation on the color channels:

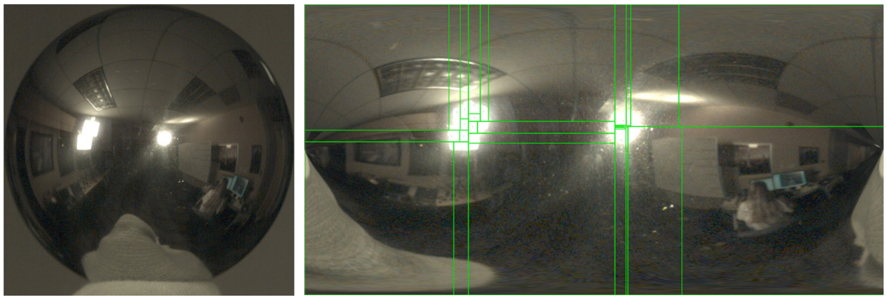

Figure 4: Left: The light probe that was also captured on-site along with the background plate. Right: 64 sampled lights according to the Median Cut algorithm [Debevec 2005].

$$\begin{aligned} R' &= R - \alpha \cdot B \\ G' &= G - \beta \cdot B \\ B' &= \gamma \cdot B \end{aligned} \tag{3}$$

The rationale for the above operation is that the specular reflectance in skin is typically of the order of the blue component of the diffuse albedo and much less in magnitude compared to the red and the green components. Hence, the above operation removes some of the specular reflection in the observed image, thereby enhancing the color difference between the diffuse and specular components, making it suitable for an SUV color transformation thereafter. In our implementation, we emperically found $\alpha = 0.4, \beta = 0.7, \gamma = 0.5$ to work well for this diffuse color enhancement operation.

The entire separation process is then as follows: We first perform the color enhancement operation (Equation 3) and transform the image I into SUV color space to obtain I_{SUV}. Then we set the U and the V components of the transformed image to zero and transform I_{SUV} back into RGB to obtain just the specular component I_S. Finally, we subtract I_S from I to obtain I_D. We perform these operations on all of the gradient illumination patterns to obtain the diffuse and specular albedo maps as well as separate diffuse and specular normals (see Figure 3).

5 Rendering

The captured facial performance is finally rendered with incident illumination from a light probe [Debevec 1998] (see Figure 4, left) that is captured on-site along with the background plate in order to match the appearance of the digital face for compositing with the background plate. Here, we sample the captured light probe into a set of 64 point lights using the Median Cut algorithm [Debevec 2005] (see Figure 4, right) and employ the hybrid normal map technique [Ma et al. 2007] for efficiently relighting the captured performance with a local shading model using the separated diffuse and specular normals and albedo textures. The separated albedo texture is used as the per-pixel specular roughness while we emperically choose a specular lobe that is appropriate for a face. The relit face is then seamlessly composited on the background plate for final replacement (see Figure 1). Note that the performance in the background plate was captured with a hood around the face in order to simplify the compositing. However, additional effort would make the technique extend for performances without such hoods.

6 Conclusions

In this work, we present a digital face replacement technique designed for a dynamic performance. Our method is based on high-speed capture of performance geometry and photometric normals and albedo textures. We also propose color-space separation of the diffuse and specular reflections components of the acquired data for high-quality relighting of the performance geometry. Our technique enables very realistic face replacements for performances with viewpoint freedom that is charectersitic of geometry based methods as well as relightability for compositing in arbitrary background plates.

References

[Debevec 1998] DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of ACM SIGGRAPH 98*, Computer Graphics Proceedings, Annual Conference Series.

[Debevec 2005] DEBEVEC, P., 2005. Median cut algorithm for light probe sampling. SIGGRAPH poster.

[Jones et al. 2006] JONES, A., GARDNER, A., BOLAS, M., MCDOWALL, I., AND DEBEVEC, P. 2006. Performance geometry capture for spatially varying relighting. In *European Conference on Visual Media Production (CVMP).*

[Ma et al. 2007] 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 et al. 2008] 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 Transactions on Graphics (Proc. SIGGRAPH Asia).*

[Ma 2008] Ma, W.-C. 2008. *A Framework for Capture and Synthesis of High Resolution Facial Geometry and Performance*. PhD thesis, National Taiwan University.

[Mallick et al. 2005] Mallick, S. P., Zickler, T. E., Kriegman, D. J., and Belhumeur, P. N. 2005. Beyond lambert: Reconstructing specular surfaces using color. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*.

[Shafer 1985] Shafer, S. 1985. Using color to separate reflection components. *COLOR research and Applications 10*, 4, 210–218.

[Zhang and Yau 2006] Zhang, S., and Yau, S.-T. 2006. High-resolution, real-time 3D absolute coordinate measurement based on a phase-shifting method. In *Optics Express*, vol. 14, 2644–2649.

[Zhang 2000] Zhang, Z. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22*, 11, 1330–1334.