Post-production Facial Performance Relighting using Reflectance Transfer

Pieter Peers* Naoki Tamura†‡ Wojciech Matusik‡ Paul Debevec*

University of Southern California*
Institute for Creative Technologies

The University of Tokyo† MERL‡

Figure 1: *From left to right: A reference (source) subject relit with the desired target illumination. A frame of the performance (target) actor captured under uniform illumination. The quotient image of the reference actor relit using the target illumination and the uniform performance illumination warped into the pose of the performance actor. The performance actor with the desired shading transferred from relit reference actor using the quotient image.*

Abstract

We propose a novel post-production facial performance relighting system for human actors. Our system uses just a dataset of view-dependent facial appearances with a neutral expression, captured for a static subject using a Light Stage apparatus. For the actual performance, however, a potentially different actor is captured under known, but static, illumination. During post-production, the reflectance field of the reference dataset actor is transferred onto the dynamic performance, enabling image-based relighting of the entire sequence. Our approach makes post-production relighting more practical and could easily be incorporated in a traditional production pipeline since it does not require additional hardware during principal photography. Additionally, we show that our system is suitable for real-time post-production illumination editing.

Keywords: image-based relighting, reflectance transfer, interactive lighting design.

1 Introduction

A central challenge in film production is for each shot to capture the best possible performance of the actor at the same time that it features the cinematographer's best choice of lighting. Coordinating the crafts of acting, directing, and cinematography to achieve the best results from each discipline is both difficult and expensive, as the frequent interruptions to set up lights and make camera adjustments limit the time available to obtain the best possible performances from the actors. Traditional filmmaking also requires committing to particular lighting choices at the time of filming, which is increasingly troublesome when so many other elements such as backgrounds and visual effects are created and refined during the post-production process.

Recent animated films such as *Monster House* have successfully used performance capture techniques [Williams 1990; Terzopoulos and Waters 1993] to drive animated 3D characters in a way that preserves much of the believability and emotional engagement of the actors' performances. In this way, performances are captured separately from determining how the characters should be illuminated and (virtually) photographed, which significantly reduced the time and expense of principal photography. While appropriate for animated films, these techniques do not leverage the real appearance and reflectance of the actors, and can miss fine-scale performance motion below the spacing of the motion capture markers.

Techniques for post-production performance relighting such as [Wenger et al. 2005; Einarsson et al. 2006] have been developed for recording live-action performances in a manner that allows subsequent relighting and photography after the performance has been recorded. To date, these techniques record the actor's performance under rapidly changing basis illumination conditions, which allows an image-based relighting process to be applied in post-production. While the results are realistic, existing techniques require complex, computer-controlled lighting systems and specialized, high-speed cameras. These limitations present challenges for current production workflows and prevent application of the techniques to performances shot under static lighting conditions or to a pre-existing footage of an actor's performance.

In this work we present a new technique for realistically relighting performances that uses only standard video of the actor's performance under static illumination. Changes in performance illumination are computed based on reflectance reference data of the actor (or of a different person similar in appearance) photographed in a neutral pose under many lighting conditions.

This technique is derived from three key observations with respect to lighting faces. First, relighting of an existing image of a subject without saturated or underexposed areas (e.g., ideally a video captured under a uniform illumination condition) amounts to deriving

ACM Reference Format
Peers, P., Tamura, N., Matusik, W., Debevec, P. 2007. Post-production Facial Performance Relighting using Reflectance Transfer. *ACM Trans. Graph. 26*, 3, Article 52 (July 2007), 10 pages. DOI = 10.1145/1239451.1239503 http://doi.acm.org/10.1145/1239451.1239503.

a quotient image [Riklin-Raviv and Shashua 1999], that when multiplied by the existing image, produces an image with the desired lighting. Second, the underlying geometry and skin properties (e.g., subsurface scattering, the specularity of the skin) do not differ significantly between subjects with a similar appearance. This makes it possible to seamlessly transfer reflectance (captured in quotient images) between two similar subjects. Third, we observe that we can relight a sequence of images by propagating the reflectance information through the consecutive frames using dense correspondences (computed via an optical flow algorithm) between frames.

To relight a performance, the reflectance reference information is derived from a static reflectance field of a single individual photographed under all possible lighting and view directions. The performance itself, is recorded under known (ideally uniform) illumination, potentially with a different, but similar looking person. Initially, we only transfer the reflectance, using quotient images, from the dataset to a limited number of key frames of the performance. Subsequently, the reflectance is propagated from the key frames to the remainder of the sequence using a bidirectional optical flow, such that temporal coherence is maintained.

Our approach is entirely image-based and does not require 3D geometry of the performance or the source reflectance field. The technique is straightforward and robust, and typically requires little or no user-assistance. Our results model all the major components of skin reflectance and shading including hard and soft shadows, specular highlights, and subsurface scattering. Our technique is especially well-suited to entertainment applications, where the highest priority is for the relit sequences to be perceptually convincing rather than being radiometrically accurate.

Our main contributions consist of:

- The first complete system for post-production performance relighting that **completely** decouples performance capture from lighting design.
- A method for all-frequency, view-dependent reflectance transfer from a static reflectance dataset of a source subject onto a dynamic performance of a different subject.
- A real-time relighting preview and design system.
- A user-assisted modification of an optical flow algorithm, and a new optical flow refinement algorithm.

2 Previous Work

Previous work can be divided into four categories: hardware-systems for post-production relighting, morphable face modeling and lighting, modeling faces using quotient images, and measuring skin reflectance.

Post-production Relighting Hardware-systems. The first systems for measuring reflectance of human faces [Debevec et al. 2000] collected images of a static face from a dense array of lighting directions. This allowed them to relight the subject using a linear combination of the original images. These systems have been extended to dynamic performances by acquiring reflectance fields for a large number of key poses and then blending between the reflectances [Hawkins et al. 2004]. Alternatively, dynamic reflectance fields have been captured by means of time-multiplexing using a high-speed camera and light sources [Wenger et al. 2005]. The main difference between these approaches and ours is that they are either not suited for capturing dynamic performances or require a sophisticated hardware setup during the performance shoot.

Morphable Face Modeling and Lighting. In morphable face modeling and lighting, a general face representation is built from a collection of examples. Pighin et al. [1999] represent a space of expressions for a given subject. Blanz and Vetter [1999] use a linear model to represent changes in 3D geometry and texture due to identity. This work has been extended to handle identity, expression, and visemes [Vlasic et al. 2005; Blanz et al. 2003], in order to resynthesize video footage. In the context of morphable models, only diffuse reflectance or simple analytic reflectance models have been used and these approaches offer limited realism of the relit images (e.g., no subsurface scattering).

Modeling Faces using Quotient (Ratio) Images. Marschner and Greenberg [1997] compute ratios of synthesized images under two different lighting conditions and modify a photograph taken under the first illumination to generate the photo under the second illumination. However, Riklin-Raviv and Shashua [1999] were the first to introduce the notion of a quotient image for image-based relighting of faces. In particular, they model human faces as Lambertian surfaces and use the fact that the image space of Lambertian faces can be modeled using a low-dimensional representation. They assume a fixed (frontal) viewpoint and no dense correspondence. Stoschek [2000] extends this to arbitrary viewpoints and lighting using image morphing. Liu et al. [2001] capture the illumination change due to one person's expression and map it to a different person using an expression ratio image. Wen et al. [2003] relight faces using a reference sphere (radiance environment map) to model the skin reflectance. They represent the radiance environment map using spherical harmonics. Thus, they are able to rotate the lighting or change the coefficients of the basis, but in principle only low-frequency shading can be achieved. These methods, in contrast to the presented method, are limited to still images. Furthermore, we use a data-driven method based on detailed, high resolution reflectance fields instead of assuming a simple analytical reflectance model. Additionally, we introduce a spatial filtering technique to remove artifacts from the obtained quotient images in order to achieve movie-quality relit results.

Measuring Skin Reflectance. Marschner et al. [1999] use image-based measurements of skin reflectance to estimate homogenous reflectance parameters and skin albedo. Sim and Kanade [2002] use a collection of images of faces under different illumination conditions to build a statistical model of surface normals. Next, they show how to recover surface normals from a single image of a face under an unknown light source. They model the deviation from Lambertian reflectance and are able to relight the face from arbitrary viewpoints. Georghiades [2003] uses a Torrance-Sparrow reflectance model to estimated both shape and reflectance based on a small number of images from a single viewpoint to relight faces. Fuchs et al. [2005] estimate parameters of analytic spatially varying *BRDF* models in order to relight faces and transfer reflectance to other faces. Weyrich et al. [2006] measure, analyze, and transfer parameters of spatially varying analytic *BSSRDF* models for a large collection of people. All these methods acquire skin reflectance models of static subjects with varying radiometric accuracy. In contrast, in this paper we focus on obtaining visually pleasing results for video sequences.

3 Acquisition

The acquisition consists of two steps: capturing a reflectance dataset of a source subject and recording the target actor's performance. The first step can be treated as a pre-processing step. The second step, recording the actor's performance, is a simple process that can be performed at any time.

Face Reflectance Dataset. Our reflectance acquisition system is similar to the Light Stage systems [Debevec et al. 2000; Wenger et al. 2005], and consists of 16 synchronized cameras and 150 light sources mounted on a geodesic sphere. During reflectance acquisition, each light is turned on sequentially, while the cameras capture images. The lighting sequence is repeated for two different exposure settings and the corresponding images are assembled into a high dynamic range (HDR) image. A single reflectance dataset contains 2,400 HDR images. Capturing a complete reflectance field takes about 20 seconds during which the subject must remain static.

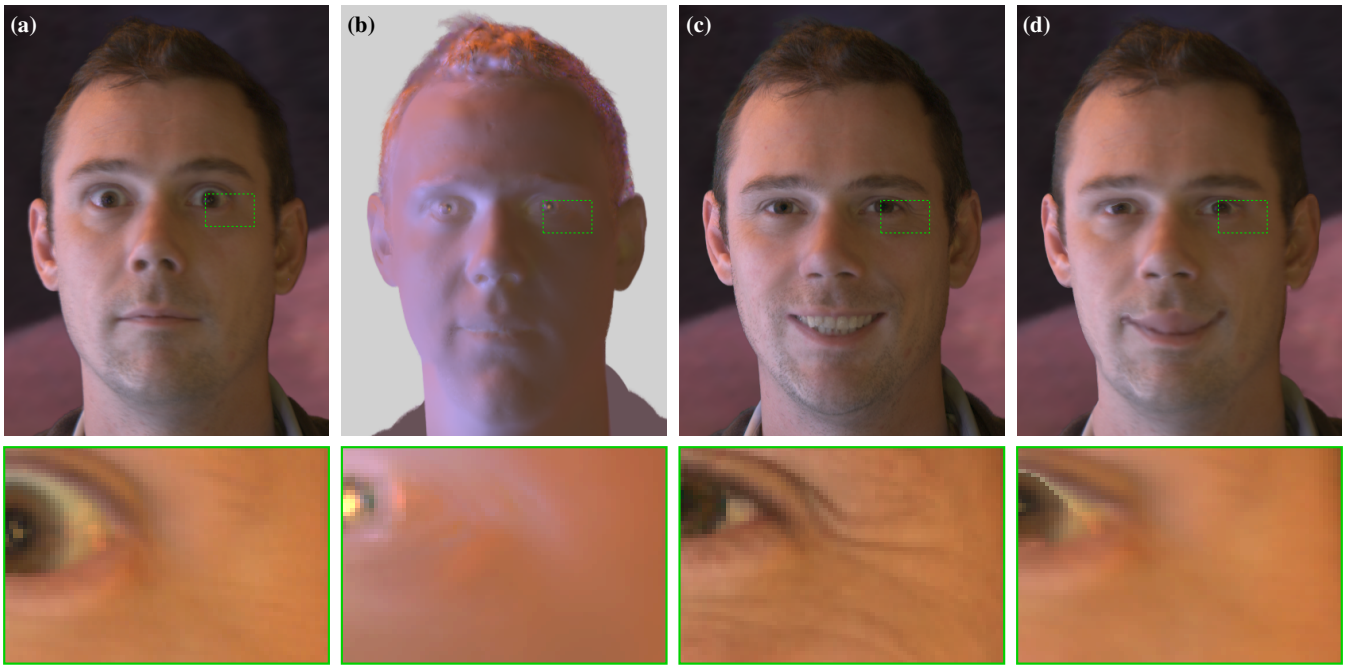

Figure 2: ***Same-subject relighting.*** *(a) The reference dataset pose, relit using an environment map. (b) The quotient image of the reference dataset pose lit by the desired target illumination, and the uniform performance illumination. warped into the target pose. (c) Transferring reflectance using the quotient image maintains fine-scale geometrical details in the resulting relit image. (d) Directly warping image (a) will miss these fine-scale geometrical details. For each image a (gamma-enhanced) detail of the subject's left eye is shown to better illustrate the fine-scale geometrical expression details in (c), absent in the relit reference dataset pose, quotient image, and directly warped relit image.*

Performance Capture. Capturing the performance of the actor is relatively straightforward. We use a Basler A-101fc high-resolution camera with a 1300×1030 pixel resolution running at 15 frames per second. Ideally, the performance is captured at 24 frames per second or higher with a digital motion picture camera. The presented algorithms should work equally well or better at such frame rates. For lighting, we use relatively uniform frontal illumination which ensures that no pixels are oversaturated or underexposed. Furthermore, this minimizes the relative contribution of specular reflections from the skin. Our process requires a known lighting environment. This can either be measured using a light probe [Debevec 1998] or from the reflection in the actor's eyes [Nishino and Nayar 2004]. During the acquisition we capture the actor in front of a green screen, allowing us to compute an alpha-matte for the performance.

4 Relighting using Reflectance Transfer

Our post-production relighting process consists of a number of important steps. We first discuss how to transfer reflectance from a reference reflectance field of a subject to a still image of a possibly different subject captured under known illumination. Next, we describe how to propagate the reflectance through the entire sequence without introducing temporal artifacts.

4.1 Reflectance Transfer for a Still Image

The goal of reflectance transfer is to compute a relit image under some user-defined environmental lighting L of a frame t of a performance by an actor B. To achieve this, information from a static reflectance field $R_A(\cdot)$ of an actor A is used as input, whose performance is captured under some known uniform illumination U. We denote the *unknown* reflectance field of the performance actor B and frame t as: $R_{B,t}(\cdot)$. The acquired frames are thus $R_{B,t}(U)$, and the desired relit frames are $R_{B,t}(L)$.

Same-subject Reflectance Transfer. We first consider the case in which both the source and performance actor are the same ($A = B$). It is highly unlikely that the performance actor assumes a pose/expression in any frame t similar to the pose/expression in the reference reflectance field. Therefore, we compute an image-based function that warps the dataset pose to the uniformly lit frame t: $f_t(R_A(U)) \approx R_{A,t}(U)$. The computation of this warp function is detailed in Subsection 5.2. We assume that the spatial relations between arbitrarily lit frames and a relit dataset image can be warped with sufficient accuracy using the same warp function:

$$f_t(R_A(L)) \approx R_{A,t}(L). \tag{1}$$

A straightforward solution to transferring the reflectance is to directly warp a relit image $R_A(L)$. Because the large scale geometry is very similar, the effects of the incident illumination L will be globally correct on $f_t(R_A(L))$. This implies that on a global scale the occlusions and normals are *nearly* the same. As a result, large shadows, such as those cast by the nose, and the overall appearance of highlights will be similar. Locally, however, the effects of small geometrical details, such as wrinkles caused by the different expressions/poses, will introduce localized variances in shading and occlusion. An example of this effect is shown in Figure 2 where a relit image (Figure 2.a) is warped into the target pose (Figure 2.d). Apart from obvious errors around the mouth (due to the appearance of teeth), important details conveying the expression are missing, such as (lack of) wrinkle detail around the subject's left eye.

In order to transfer the global reflectance information, while maintaining the local details, we take an alternate approach. We observe that occlusions are one of the major contributing factors to the differences due to small scale geometry, and that the shading on corresponding surface points is similar. This observation is supported by the fact that the reflectance of faces has a significant diffuse component due to subsurface scattering. Finding inspiration in *ambient occlusion*, we approximate the reflectance field $R_{A,t}(\cdot)$ as the product of an ambient occlusion-like term $M_{A,t}$ (independent of the incident illumination) and the shading information $S_{A,t}(\cdot)$ (dependent on the incident illumination). Note that similar to ambient occlusion, this approximation is not radiometrically correct, but provides a convincing approximation. We can now write:

$$\frac{R_{A,t}(L)}{R_{A,t}(U)} \approx \frac{S_{A,t}(L)M_{A,t}}{S_{A,t}(U)M_{A,t}} = \frac{S_{A,t}(L)}{S_{A,t}(U)}. \tag{2}$$

This ratio is independent of occlusions and only encodes the shading information. An identical approximation can be made for $\frac{f_t(R_A(L))}{f_t(R_A(U))}$. Since we assumed that $S_{A,t}(\cdot) \approx f_t(S_A(\cdot))$, we can combine both equations to:

$$\frac{R_{A,t}(L)}{R_{A,t}(U)} \approx \frac{f_t(R_A(L))}{f_t(R_A U)}. \quad (3)$$

To express the desired relit image $R_{A,t}(L)$ in terms of the acquired $R_{A,t}(U)$, we can define the quotient image:

$$Q_t(L,U) = \frac{f_t(R_A(L))}{f_t(R_A(U))}. \quad (4)$$

Using this quotient image, a relit frame can now be approximated by: $R_{A,t}(L) \approx Q_t(L,U)R_{A,t}(U)$. Our relit image contains the globally correct reflectance transferred from $R_A(L)$, together with the local occlusion information from $R_{A,t}(U)$. Approximation error is proportional to how closely similar the expressions and poses are in the two images. If expression and pose are identical, no approximation error will be made. An example of a quotient image and reflectance transfer are shown in Figures 2.b and 2.c.

Cross-subject Reflectance Transfer. Transferring reflectance between two different individuals A and B produces more pronounced variances in geometry and reflectance then we observe in the same-subject case. Again, we compute a warp function which maps subject A to the pose and expression of subject B in the t^{th} frame: $f_t(R_A(L)) \approx R_{B,t}(L)$. This warp function not only matches pose and expression, but also aligns the large scale geometrical features. The computation of this warp function is detailed in Subsection 5.2.

In a similar manner to the same-subject case, we can derive a quotient image $Q_t(L,U)$. However, the assumption that $S_{B,t}(\cdot) \approx f_t(S_A(\cdot))$ is now less accurate. The quality of the transfer, will be greatly dependent on feature similarity between the source and target actor in terms of geometry (i.e., normal information), and reflectance. However, good results can still be obtained when both subjects are similar in appearance, as evidenced by the examples shown in Figures 1 and 3.

Summary: Relighting using Quotient Images. To compute a relit frame of a captured performance, we use the following algorithm. First, the best matching viewpoint is selected from the reflectance dataset. This can be easily done by comparing the head orientation [Jones and Viola 2003] of the target frame and each of the viewpoints in the dataset. Next, given this (fixed viewpoint) reflectance field, two relit images are generated under (1) the desired target illumination and (2) the illumination in which the performance is captured. A quotient image is computed by dividing each pixel in the target illuminated image by the corresponding pixel in the performance illuminated image. Next, a warp function is computed between the target frame of the sequence, and the identically relit reference dataset image. This warp function is subsequently used to warp the quotient image to the same pose and expression as the target frame of the performance. Alternatively, it is also valid to compute the quotient image on the warped relit source images. Finally, a relit frame is generated by multiplying the warped quotient image and the target frame. This process is illustrated in Figure 4.

4.2 Reflectance Propagation for an Image Sequence

The previous subsection discussed how to transfer reflectance from a reflectance reference dataset to a still image, i.e., a single frame in the performance sequence. Repeating this procedure for every frame in the sequence, however, can result in temporal discontinuities in the transferred reflectance. Visually this translates into a distracting jittering and waving of shading and shadows. Therefore, we compute quotient images for a select number of key frames, and use optical flow to propagate the reflectance to non-key frames through the entire sequence.

Figure 3: ***Cross-subject reflectance transfer.*** *The reflectance from the top subject is transferred to the bottom subject. The left column shows the subjects relit using an environment map. The right column shows the subjects relit by a spot light.*

Propagation. To propagate the reflectance, we use optical flow to warp the quotient image at a key frame to the target frame. Currently, we manually select key frames at 10 to 20 frame intervals. A forward and backward optical flow is computed for in-between frames to the bounding key frames. The optical flow computation is detailed in Subsection 5.1.

To compute a relit in-between frame, the quotient images of the next and previous key frames are computed, and subsequently warped using the forward and backward optical flow to the target frame pose. Note that the bounding key frame quotient images do not necessary need to be computed with respect to the same viewpoint in the reflectance dataset. As a result, the relative coordinate system of the head can change with respect to the incident illumination. Therefore, we rotate the incident illumination in the opposite direction of the rotation of the head to compensate for the implicit change of viewpoint in the warp functions. Since the quotient image pair will be very similar but not identical, especially if two different viewpoints are used, we blend these two quotient images, applying blending weights linearly depending on the distance to the key frames. Using the blended quotient images, a relit frame is created. This process is illustrated in Figure 5.

Cross-bilateral Quotient Filtering. Although the warp function and optical flow give the most optimal warp function feasible, it is

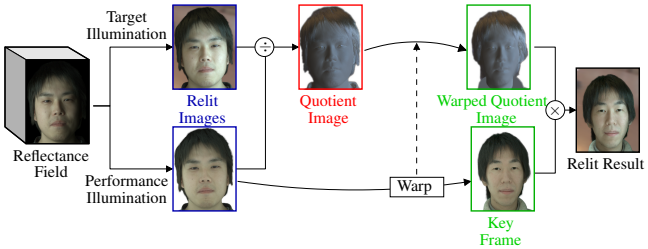

Figure 4: *An iconic representation of the reflectance transfer algorithm.*

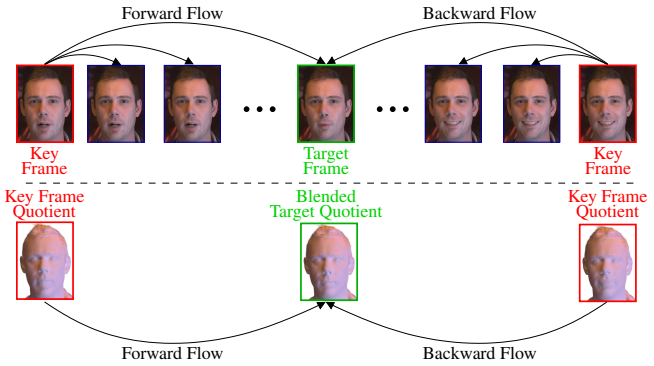

Figure 5: *An iconic illustration of the reflectance propagation process.*

still possible that some areas are not matched very well. For example, errors in the optical flow due to drifts, disappearance and appearance of teeth and blinking eyes, and feature differences between the source and target subject will result in an incorrect reflectance transfer in localized areas. These localized areas must be treated specially.

We propose to filter the the quotient images spatially and in proportion to the error on the warped uniformly relit reflectance dataset image and the target frame of the sequence. The motivation is that pixels with a similar albedo have a similar reflectance function, and thus can be used to correct pixels that have an incorrect quotient-value. If the error is large, then the quotient image should be smoothed out. However, if the error is small, then the quotient image is maintained in full detail. Inspired by the cross-bilateral filter [Eisemann and Durand 2004] and the joint bilateral filter [Petschnigg et al. 2004], we introduce a cross-bilateral quotient filter that filters the quotient image based on (1) the quality of the warp functions, and on (2) the similarity of pixel albedo in the uniformly lit dataset image and the performance frame.

Denote the current frame index by t, and the key frame index by t_0. Furthermore, denote the $\mathbf{p}^{th}$ pixel of the quotient image of frame t by $Q_{t,\mathbf{p}}$. $I_{t,\mathbf{p}}$ is the intensity of $\mathbf{p}^{th}$ pixel of the t^{th} frame, and $f_{t_0\to t}(\cdot)$ is the optical flow warping function from key frame t_0 to a frame t. Then, the filtered propagated quotient image of a frame t is:

$$Q_{t,\mathbf{p}} = \frac{1}{n(\mathbf{p})} \sum_{q\in\Omega} \left(G_s(\mathbf{p},\mathbf{q}) G_i(\mathbf{p},\mathbf{q},t_0,t) G_d(\mathbf{p},\mathbf{q},t_0,t) f_{t_0\to t}(Q_{t_0,\mathbf{q}}) \right).$$

where:

$$\begin{aligned} G_s(\mathbf{p},\mathbf{q}) &= G(|\mathbf{p}-\mathbf{q}|), \\ G_i(\mathbf{p},\mathbf{q},t_0,t) &= G(|I_{t,\mathbf{p}} - f_{t_0\to t}(I_{t_0,\mathbf{q}})|), \\ G_d(\mathbf{p},\mathbf{q},t_0,t) &= G(|I_{t,\mathbf{q}} - f_{t_0\to t}(I_{t_0,\mathbf{q}})|), \end{aligned}$$

G is a Gaussian weighting function, $n(\mathbf{p})$ a normalization constant, and Ω a neighborhood around $\mathbf{p}$. The G_s term is a spatial filter, ensuring that high frequency details are filtered out. The G_i term ensures that pixels with similar albedo in the target frame and in the warped key frame are weighted more. Finally, the G_d term gives more weight to pixels for which the optical flow is accurate. We use the following deviations: $\sigma_d = \sigma_i = 0.01$, and $\sigma_s = 40.0\,|I_{t,\mathbf{p}} - f_{t_0\to t}(I_{t_0,\mathbf{p}})|$.

The quotient images corresponding to the key frames are filtered in a similar manner with respect to the reflectance dataset, but instead of using the optical flow $f_{t_0\to t}(I_{t_0,\mathbf{q}})$, the warp of the uniformly relit dataset image is used.

An example of a reflectance transfer from a filtered versus an unfiltered quotient image is shown in Figure 6. Not only does the filter solve warping problems around the ears, mouth and nose, it also removes small scale geometrical effects that are not originally present on the target subject. The latter is illustrated by the smoothing of the wrinkles on the forehead. Furthermore, the cross-bilateral filtering allows us to correct for errors in the optical flow, lowering the accuracy requirements for the correspondence alignment.

5 Correspondence Estimation

In our system, correspondences need to be estimated in two separate cases. This is not always trivial because faces do not have much texture. First, the optical flow of the performance needs to be estimated with respect to the key frames in order to propagate the reflectance along the performance. Second, we need to compute the key frame warpings that provide a consistent mapping between the source and performance actor for each key frame.

For both correspondence estimations we use an extension of the powerful optical flow algorithm of Brox et al. [2004]. A key feature of this optical flow algorithm is that the error-functional is not linearized, but is rather iteratively solved using a numerical scheme based on two nested fixed point iterations. Additionally, a coarse-to-fine strategy is followed to allow for large scale displacements. [Brox et al. 2004] use an atypical ratio of 0.9 between two levels in the coarse-to-fine refinement rather then the *standard* 0.5 ratio.

5.1 Optical Flow Computation

The algorithm of Brox et al. [2004] is well suited to compute optical flows between individual frames. Occasionally, however, a difficult frame is encountered for which the algorithm is not completely successful. Furthermore, when concatenating individual flows, undesirable drifts can occur due to sub-pixel errors. To counter these problems we present two new extentions. First, we introduce a way for the user to steer the algorithm in difficult areas by providing additional *hints*. Second, we present an additional refinement step that is able to correct partially inaccurate flows.

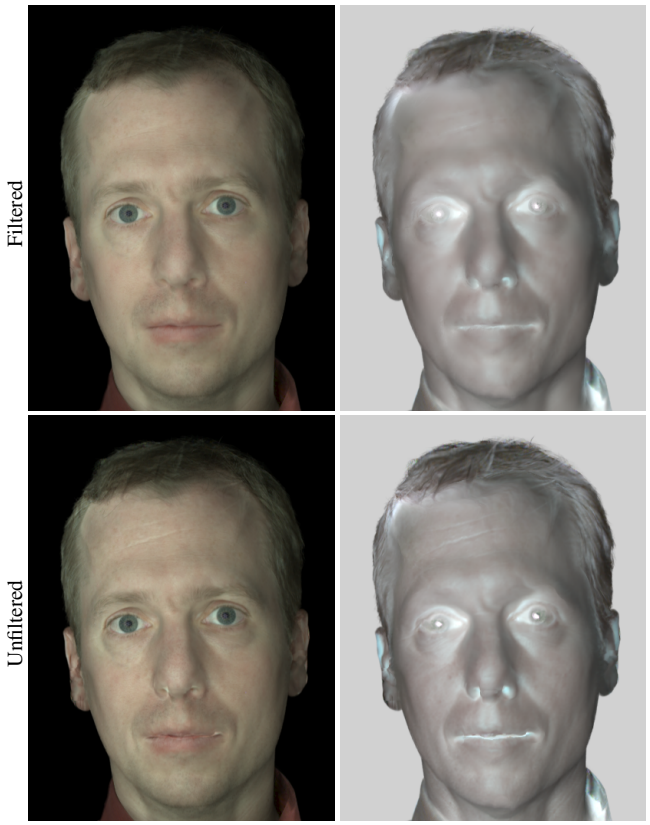

Figure 6: ***Reflectance transfer using filtered and unfiltered quotient images.*** *The filtering not only removes warping artifacts, but also removes small-scale geometrical effects not present on the target subject (e.g., the wrinkles on the forehead).*

User-assistance. To steer the optical algorithm, *hints* in the form of point correspondences are provided. The correspondence points can be selected manually, or generated automatically using, for example, the well-known *KLT* technique [Lucas and Kanade 1981]. To take these hints into account, we add an additional error term to the energy functional:

$$E_{Hints} = \int \left(\sum_i \phi(|\mathbf{x} - \mathbf{x}_{\mathbf{h}_i}|^2) \Psi(|\mathbf{w} - \mathbf{h}_i|^2) \right) d\mathbf{x}, \quad (5)$$

where $\mathbf{x}$ are image coordinates, $\mathbf{w} = (u, v)$ the flow vectors, $\Psi(\cdot)$ is the robust norm used by [Brox et al. 2004], $\mathbf{h}_i$ is the i^{th} hint (also a flow vector), $\mathbf{x}_{\mathbf{h}_i}$ is the image coordinate for which hint $\mathbf{h}_i$ is given, and $\phi(\cdot)$ is a weighting function, monotonically decreasing as a function of the distance between two image coordinates. A key observation is that the flow field is locally relatively smooth around these points, and thus, the neighboring pixels have a similar flow vector. This error function will guide the optical flow field around the user provided correspondence points, in the direction of the hints.

Refinement of Existing Flows. In a number of cases a flow is obtained by means other than the optical flow algorithm. For instance, when computing the flow from a key frame to a frame more than one time-step apart, it is necessary to concatenate the individual flows between all consecutive frames, such that the resulting optical flow is temporally coherent. However, due to small errors in the individual flows, undesirable drifts can occur in the concatenated optical flow.

We propose a coarse-to-fine refinement technique based on the optical flow algorithm discussed previously. The key idea is to retain the accurate portions of the optical flow, and refine just those flow vectors that are inaccurate. At each coarse-to-fine level, we warp a downsampled source frame l, using a downsampled optical flow, to the target frame k. If the error on both images for a specific pixel is small, then we force the algorithm to output a similar vector. To achieve this, we introduce an additional error term to the energy functional:

$$E_{Flow} = \int \rho(|I_{k,\mathbf{x}} - I_{l,\mathbf{x}+\mathbf{w}'}|^2) \Psi(|\mathbf{w} - \mathbf{w}'|^2) d\mathbf{x}, \quad (6)$$

where $\rho(\cdot)$ is a weighting function decreasing monotonically from 1 (no error) towards 0 at some maximum error threshold, and $\mathbf{w}'$ is the user-provided flow vector for pixel $\mathbf{x}$.

This error term resembles the previously-introduced hint error term E_{Hints} in the sense that every vector in the user-provided optical flow can be seen as a *hint*, but now relative weighting dependents on its accuracy, instead of spatial distance as in Equation 5.

5.2 Key Frame Warp Computation

The visual differences between the (relit) source image and performance key frame can be significant (e.g., different skin albedo, different facial geometry, different pose, etc...) in such a way that the optical flow assumptions are not completely valid anymore. Therefore, we first compute an initial homography from a number of easily detectable points (e.g., corners of the eyes, mouth, etc). This homography is used to warp the source image, and a local histogram adjustment is applied. Next, the optical flow algorithm is executed to further refine the warp.

6 Results

Comparison to Ground Truth. To evaluate our method, we make a comparison to ground truth relit images. Although our goal is not to create radiometrically correct relit images, it is interesting to note how well our technique performs. We consider two cases: the effects of different expressions and poses, and the effects of cross-subject transfers.

First we study the influence of the expression/pose differences between the source and target. To do so, we capture two reflectance fields of a single person, each with a different expression. One of the reflectance fields is used to create a quotient image, while the other is used to create a relit reference image, and a uniformly lit target frame. Figure 7 offers such a comparison. The obtained results are visually very close to the ground truth images. The visual differences are more noticeable in the single light source case. However, this case is quite difficult, and without a reference image with which to compare, these discrepancies are not easily noticed. The transfers from a neutral expression provide slightly better results than do transfers from an extreme expression. Finally, the specular highlights are globally located at the correct locations, although some detail is missing. This is clearly visible on the forehead.

Using a similar approach, we study the effects of a cross-subject transfer. Such a comparison is shown in Figure 8. The results of the cross-subject reflectance transfer are similar to the ground truth images. However, the differences are more noticeable than in the same-subject case. Some of the medium scale features such as the forehead ridge are transferred from the source subject. However, the resulting relit images are still quite convincing.

Additional Results. Each column of Figure 10 shows a five frame sequence of different subjects under various lighting conditions. The first column shows a static frame illuminated by a single light source moving from the top to the bottom. These frames are generated using a same-subject transfer. Note the correct shading and shadowing of the eye sockets. In the second column a cross-subject transfer is illustrated, relit by an environment map. We used the same subjects as in Figure 3. The relit results look convincing, and without reference images it is difficult to detected artifacts. The third column also shows a cross-subject transfer. In this case we used the same subjects as in Figure 8. Finally, in the last column, we show a same-subject transfer example in which a significant rotation of the head occurs. The quotient images of the prior and next key frame for this five frame sequence are computed from different viewpoints in the reflectance dataset.

To this point, our example target frames were captured under uniform illumination. Figure 9 provides an example in which the target frame is captured under (low frequency) non-uniform illumination (i.e., the Uffizi light probe). The obtained results illustrate that even in this case our algorithm performs very well.

Real-time Relighting. An important application of our facial performance relighting system is cinematic lighting design. In such an application it is important for the director or lighting artist to change the lighting interactively during playback of the performance. Except for the correspondence estimation, all parts of our system can be executed in real-time on a GPU. Fortunately, the correspondence estimation does not change during relighting, and can be computed beforehand as a preprocessing step.

The average length of a performance shot in movies is about 4 seconds. Current graphics cards provide sufficient onboard video memory to load a typical performance sequence, its corresponding optical flows, a single reflectance field, and the key frame warping functions. Computing a relit frame can now be accomplished by warping the quotient image to the current frame, applying the cross-bilateral quotient filter, and multiplying each pixel with this filtered warped quotient image. Computing a quotient image is only necessary when the illumination or key frame changes, and is computationally the most expensive step. Our real-time system is able to playback a sequence with changing incident illumination at a resolution of 515×650 at $25 - 50$ fps ($500 - 800$ fps without changing illumination) on an *n*VIDIA 8800 GTS with 640MB video memory. If only one viewpoint of the reflectance field needs to be loaded, a sequence of approximately 7 seconds at 30 fps can be fit into the

Figure 7: ***Reflectance transfers between different expressions of a single subject.*** *The top row shows reference images computed directly from acquired reflectance fields (under an environment map, and under a single light source). The bottom row shows the respective transfers of columns 1 ↔ 2, and columns 3 ↔ 4.*

video memory. Not only is it possible to rotate the illumination in real-time, but also to *paint* new illumination. This would for example allow a lighting artist to add light sources and immediately see their effect on the sequence.

We refer the reader to the accompanying video, demonstrating our real-time system on a number of different sequences relit with same-subject and cross-subject reflectance transfers.

Limitations and Discussion. The presented system is able to relight facial performances captured using a lightweight acquisition process. No special hardware is required when filming the performance, and only a reflectance field of a static pose of a subject similar in appearance is required.

Quotient images, as noted in the previous work section have been used in many applications to relight faces or to transfer expressions. The presented technique differs in a number of aspects. Quotient images were introduced by Riklin-Raviv and Shashua [1999] for still images of faces recorded from a fixed viewpoint. Stoscheck [2000] extended this to re-render the face for a variable viewpoint. However, both are limited to a still image, and assume an underlying Lambertian reflectance model. The presented technique differs from these methods in that we do not make the assumption that the underlying reflectance model is Lambertian. Instead of a sparse training set in terms of lighting directions and viewpoints, we use a densely captured reflectance field. Finally, Liu et al. [2001] use ratio images to transfer expressions between different faces. However, they consider only fixed lighting. Unlike previous methods, ours is not limited to a still image, but is able to relight whole video sequences.

Although the goal of the system of Wenger et al. [2005] is similar, the accuracy, hardware of the setup, and application are significantly different. Our system trades off accuracy for an easier, less hardware driven acquisition process. Our results are less radiometrically accurate compared to those of Wenger et al., but yield a reasonably plausible relit performance sequence. Finally, our system is currently limited to facial performances, while systems similar to Wenger et al. [2005] can potentially handle full-body subjects [Einarsson et al. 2006]. On the other hand, our system is less data intensive and illumination can be manipulated in real-time during post-production. The system of Hawkins et al. [2004] is similar to ours. The main difference is that they capture a large collection of expressions, while we use just a single static expression, significantly simplifying acquisition of the reflectance dataset. Furthermore, we do not require that the subject for which the reflectance dataset is recorded be the same as the performance subject.

The presented method is a purely image-based technique. An advantage is that the geometry need not be known, so that issues associated with real-time geometry capture are avoided. However, this also introduces some limitations. Effects due to fine detail geometrical changes, such as wrinkles, are not completely taken into account. Conversely, since we transfer only reflectance, and not geometrical details, the artifacts due to these fine detail geometrical changes are less noticeable. Large scale geometrical deformations, however, can still result in significant changes in light transport (e.g., differently shaped shadows, stronger inter-reflections). About 3% of the frames, and about 25% of the key frames required manually specifying hints to help the optical flow and warping algorithm. This took about 15 minutes per sequence, and is insignificant compared to the computation time of optical flow.

7 Conclusions and Future Work

We have presented a novel system for image-based post-production performance relighting. Our system allows decoupling of the performance capture and lighting design, which can greatly simplify

Figure 8: ***Reflectance transfer between different individuals.*** *The top row shows a reference image of each subject. The bottom row shows the reflectance transfers from columns* $1 \leftrightarrow 2$*, and columns* $3 \leftrightarrow 4$*.*

the film-making process. The main advantage of our approach is that it does not modify the existing production process and it only relies on a single reflectance dataset of an actor. The acquisition of the reference datasets need not to occur at any particular time, either before or after the performance acquisition. This allows one to potentially capture a database of face reflectance datasets, and select the best match from this database rather than acquiring a reflectance field from each particular actor. If the reflectance of the performing actor is similar to the actor in the reflectance dataset, the relit video more nearly approaches ground truth. However, if the actor differs from the reference dataset subject, our system still produces realistic results that are qualitatively close to ground truth.

There are several avenues for future work. First, we currently only handle distant illumination; the next step would be to reproduce the effects of general incident light fields. Second, we would like to extend this method to handle complete bodies instead of just faces. Finally, we would like to give the designer a way to modify the surface detail that is transferred (e.g., virtual make-up).

Acknowledgements. We would like to thank Tomas Pereira and Saskia Mordijck for their valuable input, Patrick Pletscher, Barton Nicholls, Clifton Forlines, Derek Schwenke, Yuki Hashimoto, Jay Thronton, and Michael Jones for sitting as subjects, Janet McAndless for proofreading and data acquisition, Brian Miller for video editing, Malte Weiss and Peter Sand for their input on the optical flow algorithm, and Silicon Studio Corporation, Bill Swartout, Randolph Hall, and Max Nikias for their support and assistance with this work. This work was sponsored by the University of Southern California Office of the Provost and the U.S. Army Research, Development, and Engineering Command (RDECOM). 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

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3D faces. In *Proceedings of SIGGRAPH 99*, Computer Graphics Proceedings, Annual Conference Series, 187–194.

BLANZ, V., BASSO, C., POGGIO, T., AND VETTER, T. 2003. Reanimating faces in images and video. *Computer Graphics Forum 22*, 3 (Sept.), 641–650.

BROX, T., BRUHN, A., PAPENBERG, N., AND WEICKERT, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *Proceedings of the 8th European Conference on Computer Vision*, 25–36.

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*, Computer Graphics Proceedings, Annual Conference Series, 145–156.

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 SIGGRAPH 98*, Computer Graphics Proceedings, Annual Conference Series, 189–198.

EINARSSON, P., CHABERT, C.-F., JONES, A., MA, W.-C., LAMOND, B., HAWKINS, T., BOLAS, M., SYLWAN, S., AND DEBEVEC, P. 2006. Relighting human locomotion with flowed reflectance fields. In *Rendering Techniques 2006: 17th Eurographics Workshop on Rendering*, 183–194.

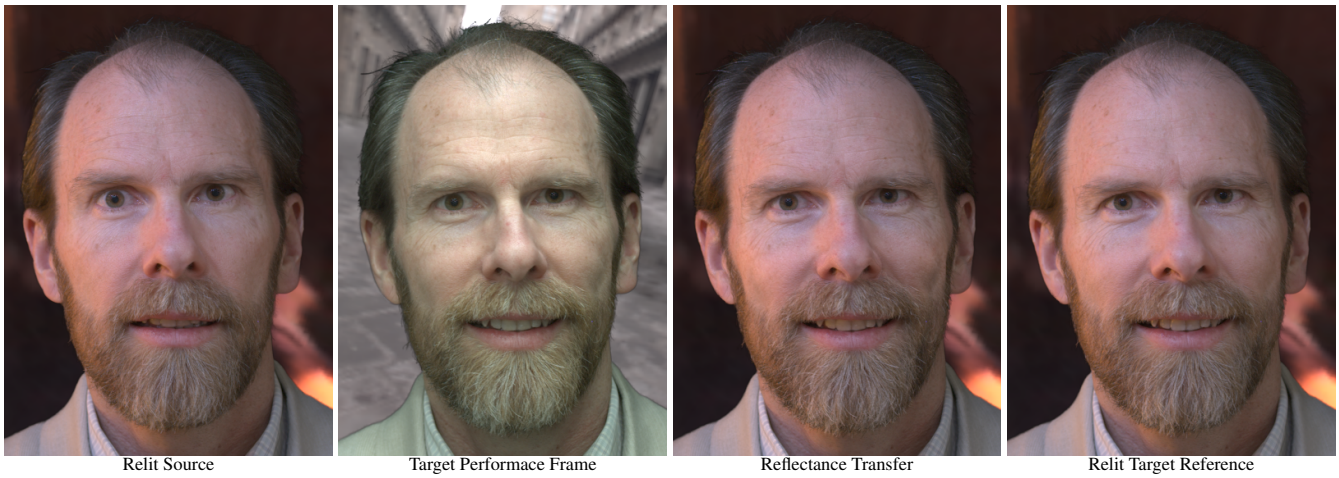

Figure 9: ***Reflectance transfer from a non-uniformly lit performance.*** *The performance frame is acquired under non-uniform illumination corresponding to the Uffizi light probe. Reflectance transfer is obtained by computing the quotient image with respect to this non-uniform illumination. A reference image is provided for comparison.*

EISEMANN, E., AND DURAND, F. 2004. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics 23*, 3, 673–678.

FUCHS, M., BLANZ, V., LENSCH, H., AND SEIDEL, H.-P. 2005. Reflectance from images: a model-based approach for human faces. *IEEE Transactions on Visualization and Computer Graphics 11*, 3 (May-June), 296–305.

GEORGHIADES, A. S. 2003. Recovering 3-D shape and reflectance from a small number of photographs. In *Eurographics Symposium on Rendering: 14th Eurographics Workshop on Rendering*, 230–240.

HAWKINS, T., WENGER, A., TCHOU, C., GARDNER, A., GÖRANSSON, F., AND DEBEVEC, P. 2004. Animatable facial reflectance fields. In *Rendering Techniques 2004: 15th Eurographics Workshop on Rendering*, 309–320.

JONES, M., AND VIOLA, P. 2003. Fast multi-view face detection. Tech. rep., MERL. TR2003-096.

LIU, Z., SHAN, Y., AND ZHANG, Z. 2001. Expressive expression mapping with ratio images. In *Proceedings of ACM SIGGRAPH 2001*, Computer Graphics Proceedings, Annual Conference Series, 271–276.

LUCAS, B., AND KANADE, T. 1981. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence (IJCAI81)*, 674–679.

MARSCHNER, S. R., AND GREENBERG, D. P. 1997. Inverse lighting for photography. In *Proceedings of the Fifth Color Imaging Conference, Society for Imaging Science and Technology*, 262–265.

MARSCHNER, S., WESTIN, S., LAFORTUNE, E., TORRANCE, K., AND GREENBERG, D. 1999. Image-based BRDF measurement including human skin. In *Rendering Techniques '99: 10th Eurographics Workshop on Rendering*, 139–152.

NISHINO, K., AND NAYAR, S. K. 2004. Eyes for relighting. *ACM Transactions on Graphics 23*, 3 (Aug.), 704–711.

PETSCHNIGG, G., SZELISKI, R., AGRAWALA, M., COHEN, M., HOPPE, H., AND TOYAMA, K. 2004. Digital photography with flash and no-flash image pairs. *ACM Transactions Graphics 23*, 3, 664–672.

PIGHIN, F. H., SZELISKI, R., AND SALESIN, D. 1999. Resynthesizing facial animation through 3D model-based tracking. In *International Conference on Computer Vision*, 143–150.

RIKLIN-RAVIV, T., AND SHASHUA, A. 1999. The quotient image:class based recognition and synthesis under varying illumination conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 02*, 262–265.

SIM, T., AND KANADE, T. 2002. Illuminating the face. Tech. rep., CMU. CMU-RI-TR-01-31.

STOSCHEK, A. 2000. Image-based re-rendering of faces for continuous pose and illumination directions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 582–587.

TERZOPOULOS, D., AND WATERS, K. 1993. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 15*, 6 (June), 569–579.

VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIĆ, J. 2005. Face transfer with multilinear models. *ACM Transactions on Graphics 24*, 3 (Aug.), 426–433.

WEN, Z., LIU, Z., AND HUANG, T. 2003. Face relighting with radiance environment maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 157–165.

WENGER, A., GARDNER, A., TCHOU, C., UNGER, J., HAWKINS, T., AND DEBEVEC, P. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics 24*, 3 (Aug.), 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 Transactions on Graphics 25*, 3 (July), 1013–1024.

WILLIAMS, L. 1990. Performance-driven facial animation. In *Computer Graphics (Proceedings of SIGGRAPH 90)*, vol. 24, 235–242.

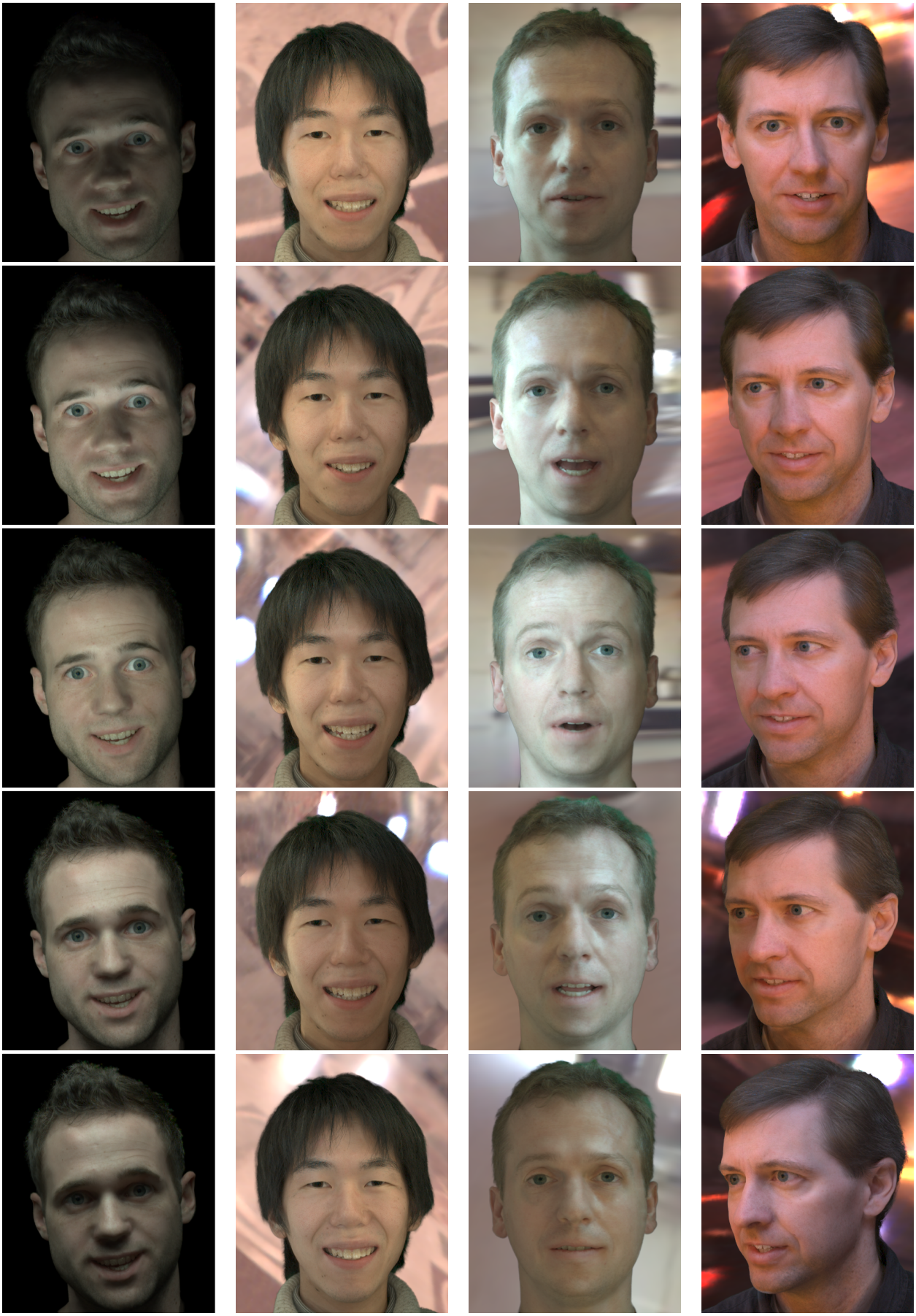

Figure 10: ***Additional Results.*** *Each column shows a five frame sequence from a performance. Columns one and four are computed using a same-subject reflectance transfer. Columns two and three are computed using a cross-subject reflectance transfer.*