High-Quality Capture of Eyes

Pascal Bérard[1,2] Derek Bradley[2] Maurizio Nitti[2] Thabo Beeler[2] Markus Gross[1,2]

1) ETH Zurich 2) Disney Research Zurich

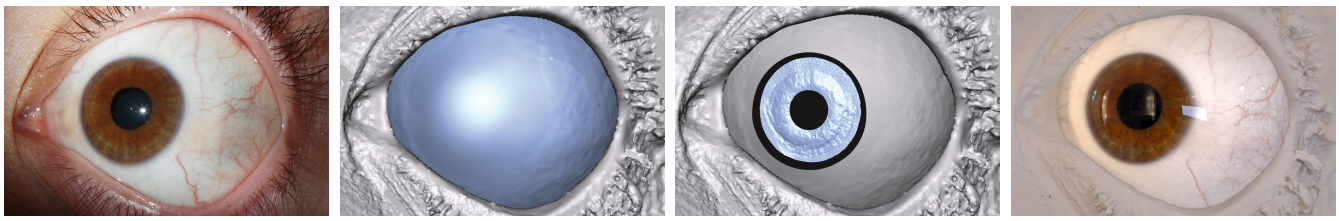

Figure 1: *We present a system to acquire the shape and texture of an eye at very high resolution. This figure shows one of the input images, the reconstructed eyeball and iris geometry, and a final render from a novel viewpoint under different illumination (left to right).*

Abstract

Even though the human eye is one of the central features of individual appearance, its shape has so far been mostly approximated in our community with gross simplifications. In this paper we demonstrate that there is a lot of individuality to every eye, a fact that common practices for 3D eye generation do not consider. To faithfully reproduce all the intricacies of the human eye we propose a novel capture system that is capable of accurately reconstructing all the visible parts of the eye: the white *sclera*, the transparent *cornea* and the non-rigidly deforming colored *iris*. These components exhibit very different appearance properties and thus we propose a hybrid reconstruction method that addresses them individually, resulting in a complete model of both spatio-temporal shape and texture at an unprecedented level of detail, enabling the creation of more believable digital humans. Finally, we believe that the findings of this paper will alter our community's current assumptions regarding human eyes, and our work has the potential to significantly impact the way that eyes will be modelled in the future.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and scanning;

Keywords: Eye capture, Eye modelling, Face reconstruction

Links: DL PDF WEB VIDEO

1 Introduction

Creating photo-realistic digital humans is a long-standing grand challenge in computer graphics. One of the cornerstones of producing digital doubles is capturing an actor's face. Over the past decade this area has been a topic of intense research, and many different approaches have been proposed [Ma et al. 2007; Alexander et al. 2010; Bradley et al. 2010; Beeler et al. 2010; Ghosh et al. 2011; Beeler et al. 2011; Graham et al. 2013; Garrido et al. 2013], most of which focus on reconstructing the skin surface and its appearance in increasing levels of detail. Only recently have researchers started to consider other facial features, such as facial hair [Beeler et al. 2012]. The eye, arguably the most important facial feature, has so far only received very little attention, especially its shape.

The visible portion of the eye is comprised of the white *sclera*, the transparent *cornea*, and the colored *iris*, as shown in Figure 2a. In computer graphics, the shape of the eye is generally approximated by two spheres, a big one for the sclera and a smaller one for the cornea [Lefohn et al. 2003; Ruhland et al. 2014]. The iris is often thought of as a planar disc, or as a cone to fake the refraction of the cornea. The constriction and dilation of the pupil is typically modelled as planar, radial motion and the out-of-plane deformation of the iris is generally neglected [Ruhland et al. 2014]. Figure 2b shows such a generic CG eye.

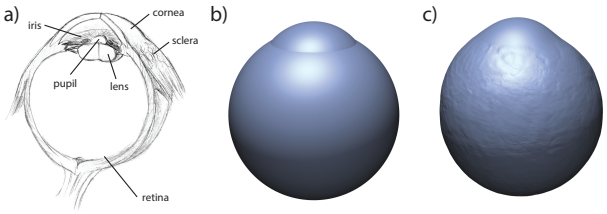

Figure 2: *a) Shows a schematic of the human eye denoting the components referred to in this paper. b) The shape of a generic CG eye represents only a low order approximation of an individual eye, while the proposed method reconstructs all its intricacies (c).*

In this paper, we argue that generic eye models typically used in computer graphics are insufficient for capturing the individual identity of a digital human. While a simple modeled or simulated eye may be sufficient for background characters, current industry practices spend significant effort to manually create eyes of hero characters. Our reconstruction method can greatly reduce the time spent and help increase the realism of the eye. As an example, Figure 2c presents an eye that is reconstructed by the method proposed in this paper. Our reconstruction specifically captures the overall shape and spatial surface variation of the sclera including a detailed vein texture, the complex shape, texture and deformation of the iris, and even properties of the transparent cornea including the exact curvature along with the refractive index at the boundary. This example demonstrates that the aforementioned assumptions only roughly approximate the true physiology of the eye, and thus cannot represent actor-specific details that can greatly increase the realism of a digital double. Furthermore, the eyeball exhibits strong asymmetry, contains microscopic surface details and imperfections such as

Pingueculas[1] - all of which are very person-specific. The microgeometry of the iris is as unique to every person as a fingerprint, and its position and deformation depends on the accommodation of the underlying lens. These are just a few examples of eye details that cannot be captured with traditional models. Through the results of this paper we will show several more examples, in particular when it comes to the dynamic deformation of the iris during *pupillary response*[2].

To overcome the limitations of generic eye models and accurately reproduce the intricacies of a human eye, we argue that eyes should be captured and reconstructed from images of real actors, analogous to the established practice of skin reconstruction through facial scanning. The eye, however, is more complex than skin, which is often assumed to be a diffuse Lambertian surface in most reconstruction methods. The human eye is a heterogeneous compound of opaque and transparent surfaces with a continuous transition between the two, and even surfaces that are visually distorted due to refraction. This complexity makes capturing an eye very challenging, requiring a novel algorithm that combines several complementary techniques for image-based reconstruction. In this work, we propose the first system capable of reconstructing the spatio-temporal shape of all visible parts of the eye; the sclera, the cornea, and the iris, representing a large step forward in realistic eye modeling. Our approach not only allows us to create more realistic digital humans for visual effects and computer games by scanning actors, but it also provides the ability to capture the accurate spatio-temporal shape of an eye in-vivo.

2 Related Work

According to a recent state-of-the-art report on eyes [Ruhland et al. 2014], there is not a single paper that is concerned with accurately reconstructing the shape of the visible components of the eye. Most research so far has focused solely on acquiring the iris, the most prominent part of an eye, typically only considering the color variation and neglecting its shape. An exception is the seminal work by François et al. [2009], which proposes to estimate the shape based on the color variation. Guided by the physiology of the iris, they develop a bright-is-deep model to hallucinate the microscopic details. While impressive and simple, the results are not physically correct and they have to manually remove spots from the iris, since these do not conform with their model. Lam et al. [2006] propose a biophysically-based light transport model to simulate the light scattering and absorption processes occurring within the iridal tissues for image synthesis applications. Sagar et al. [1994] model a complete eye including the surrounding face for use in a surgical simulator. However, the model is not based on captured data and only approximates the shape of a real eye.

In contrast, we will show that our proposed technique faithfully captures the intricacies specific to individual eyes with greater detail and fidelity (see Section 8).

Other authors have looked into the motion patterns of the iris, such as dilation or hippus[3] [Hachol et al. 2007]. Pamplona and colleagues study the deformation of the iris when the pupil dilates in 2D [Pamplona et al. 2009]. They manually annotate a sparse set of features on a sequence of images taken while the pupil dilates. The recovered tracks show that the individual structures present in the iris prevent it from dilating purely radially on linear trajectories. Our paper tracks the deformation of the iris densely since we do not require manual annotation and our measurements confirm these findings. More importantly, we capture the full three-dimensional deformation of the iris, which conveys the detailed shape changes during pupil dilation. In one of our proposed applications we complement our deformation model with the temporal model proposed by Pamplona et al. [2009].

In the medical community the situation is different. There, accurate eye measurements are fundamental, and thus several studies exist. These either analyze the eye ex-vivo [Eagle Jr 1988] or employ dedicated devices such as MRI to acquire the eye shape [Atchison et al. 2004] and slit lamps or keratography for the cornea [Vivino et al. 1993]. Optical coherence tomography (OCT) [Huang et al. 1991], in ophthalmology mostly employed to image the retina, can also be used to acquire the shape of cornea and iris at high accuracy. An overview of the current corneal assessment methods can be found in recent surveys [Rio-Cristobal and Martin 2014; Piñero 2013].

Such devices however are not readily available and the data they produce is oftentimes less suited for graphics applications. We therefore chose to construct our own setup using commodity hardware and employ passive and active photogrammetry methods for the reconstruction. To reconstruct the shape of the sclera we employ the passive multi-view stereo reconstruction method from Beeler et al. [2010], but any other multi-view reconstruction method could be used as well. The cornea on the other hand is reconstructed using an active approach since it is transparent. Following the notation of Ihrke et al. [2008] the proposed method is a hybrid 'shape-from-specularity' and 'shape-from-distortion' technique. This method reconstructs the corneal shape using reflection constraints from a sparse set of LEDs combined with dense refraction constraints from the iris acquired from several viewpoints. Using only a sparse set of lights instead of a dense illumination pattern such as employed by Halstead et al. [1996] has the advantage that the underlying iris can contribute to the optimization in the form of refraction constraints, which allows us to also estimate the index of refraction and reconstruct the shape of the iris from the same data.

In summary, while there has been a substantial amount of research regarding the reconstruction of shape of various materials [Seitz et al. 2006; Ihrke et al. 2008; Hernández et al. 2008], none of these methods seem particularly suited to reconstruct the heterogeneous combination of materials present in the eye. As the individual components of the eye are all coupled, they require a unified reconstruction framework, which is what we propose in this paper.

3 Method Overview

The complexity of human eyes dictates a novel approach for capture and accurate reconstruction. We must pay particular attention to the appearance properties of the different components of the eye, and design different strategies for reconstructing each component. While it is possible to assume that the sclera is diffuse and Lambertian (such as often assumed for skin), the cornea is completely transparent, and the iris is viewed under unknown distortion due to refraction. Furthermore, there is a coupling of the eye components, for example the corneal shape should transition smoothly to the sclera, and the perceived iris position depends on both the corneal shape as well as the exact index of refraction (both of which *do* vary from person to person).

The above observations lead to a progressive algorithm for eye reconstruction. We start by recovering the sclera shape, followed by the cornea, and finally the iris. Each stage of the reconstruction requires a different approach, relying on constraints from the previous stages but tuned to the appearance properties at hand. The various reconstruction methods also require different (but complementary) capture data, which we acquire through a novel hardware setup of cameras, flashes and LED lights.

[1] A degeneration of the fibers of the sclera resulting in a small bump.

[2] Varying pupil size via relaxation/contraction of the iris dilator muscle.

[3] A rhythmic but irregular continuous change of pupil dilation.

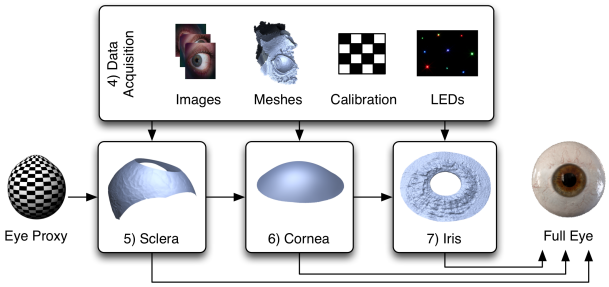

Figure 3: *This figure shows an overview of the system. First, several modalities of data are acquired (Section 4). From these plus a generic eye proxy, the system reconstructs the individual components of the eye, the sclera (Section 5), the cornea (Section 6), and the iris (Section 7) and combines them into a complete eye model.*

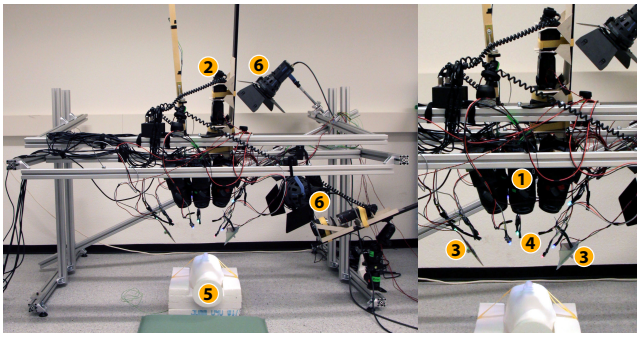

Figure 4: *Overview of the capture setup consisting of a camera array (1), a focused flash light (2), two high-power white LEDs (3) used to control the pupil dilation, and color LEDs (4) that produce highlights on the cornea. The subject is positioned in a headrest (5). The studio lamps (6) are used during camera calibration.*

To describe our method in detail, we organize the paper as illustrated in Figure 3. Section 4 explains the data acquisition phase including the capture hardware. Section 5 discusses our passive multi-view, multi-pose reconstruction method for obtaining the sclera. Given the approximate sclera shape, we design a photometric approach for computing the corneal shape given a set of known LED lights in the scene and multiple views of the refracted iris (Section 6). The iris itself is then reconstructed using a novel multi-view stereo approach that traces light paths through the corneal interface (Section 7). Irises are reconstructed for a sequence of different pupil dilations and we recover a deformable model for iris animation, parameterized by pupil radius. Our results demonstrate that each individual eye is unique in many ways, and that our reconstruction algorithm is able to capture the main characteristics required for rendering digital doubles (Section 8).

4 Data Acquisition

The first challenge in eye reconstruction is obtaining high-quality imagery of the eye. Human eyes are small, mostly occluded by the face, and have complex appearance properties. Additionally, it is difficult for a subject to keep their eye position fixed for extended periods of time. All of this makes capture challenging, and for these reasons we have designed a novel acquisition setup, and we image the eye with variation in gaze, focus and pupil dilation.

4.1 Capture Setup

Our capture setup consists of multiple cameras, a modified flash for primary illumination, and a variety of colored LEDs that will reflect off the cornea. To help the subject remain still during acquisition, we arrange the setup such that they can lie on the floor with their head in a headrest, situated under the camera array (Figure 4).

To get the best coverage in the space available, we place six cameras (Canon 650D) in a 2 by 3 configuration, with 100mm macro lenses focused on the iris. The lens is stepped down to $f11$ and the camera is set to ISO100. The exposure is set to 1 second since we capture in a dark room and the flash provides the primary illumination. The main flash light consist of three elements: a conventional flash (Canon 600EX-RT), a cardboard aperture mask and a lens. This assembly allows us to intensify and control the shape of the light so that reflections of the face and the eyelashes can be prevented as much as possible. We use 9 RGB LEDs and arrange them in a 3x3 pattern, ensuring that similar colors are not adjacent in order to maximize our ability to uniquely detect their reflections on the cornea. The pupil dilation is controlled with a high-power LED with adjustable brightness. We place this LED close to the eye that is *not* being captured. Since the pupil dilation of both eyes is linked we can control the dilation of the captured eye indirectly, avoiding an extra specular highlight on the captured eye. In order to measure the eye focusing at different depths, a focus pole with specifically marked distances is placed in front of the subject. Finally, additional studio lamps are used during camera calibration.

4.2 Calibration

Cameras are calibrated using a checkerboard of CALTag markers [Atcheson et al. 2010], which is acquired in approximately 15 positions throughout the capture volume. We calibrate the positions of the LEDs by imaging a mirrored sphere, which is also placed at several locations in the scene, close to where the eyeball is during acquisition. The highlights of the LEDs on the sphere are detected in each image by first applying a Difference-of-Gaussian filter followed by a non-maximum suppression operator, resulting in single pixels marking the positions of the highlights. The detected highlight positions from a specific LED in the different cameras form rays that should all intersect at the 3D position of that LED after reflection on the sphere with known radius (15mm). Thus, we can formulate a nonlinear optimization problem where the residuals are the distances between the reflected rays and the position estimates of the LEDs. We solve for the unknown LED and sphere positions with the Levenberg-Marquardt algorithm.

4.3 Image Acquisition

We wish to reconstruct as much of the visible eye as possible, so the subject is asked to open their eyes very wide. Even then, much of the sclera is occluded in any single view, so we acquire a series of images that contain a variety of eye poses, covering the possible gaze directions. Specifically we used 11 poses: *straight, left, left-up, up, right-up, right, right-down, down, left-down, far-left,* and *far-right*. The *straight* pose will be used as reference pose, as it neighbors all other poses except *far-left* and *far-right*.

We then acquire a second series of images, this time varying the pupil dilation. The intricate geometry of the iris deforms non-rigidly as the iris dilator muscle contracts and expands to open and close the pupil. The dilation is very person-specific, so we explicitly capture different amounts of dilation for each actor by gradually increasing the brightness of the high-power LED. In practice, we found that a series of 10 images was sufficient to capture the iris deformation parametrized by pupil dilation.

The acquisition of a complete data set takes approximately 5 minutes for positioning the hardware, 10 minutes for image acquisition,

and 5 minutes for calibration, during which time the subject lies comfortably on a cushion placed on the floor.

4.4 Initial Reconstruction

To initialize our eye capture method, we pre-compute partial reconstructions for each eye gaze using the facial scanning technique of Beeler et al. [2010]. Although this reconstruction method is designed for skin, the sclera region of the eye is similarly diffuse, and so partial sclera geometry is obtainable. These per-gaze reconstructions will be used in later stages of the pipeline. Additionally, the surrounding facial geometry that is visible will be used for providing context when rendering the eye in Section 8.

5 Sclera

Reconstructing the sclera is challenging because large parts are occluded by the eyelids and the eye socket at any given time. As indicated previously, the problem can be alleviated by acquiring the eye under multiple poses. In this section we explain our approach to register the different poses into a common frame and integrate the partial scans into a complete model of the eyeball. The individual steps are outlined in Figure 5.

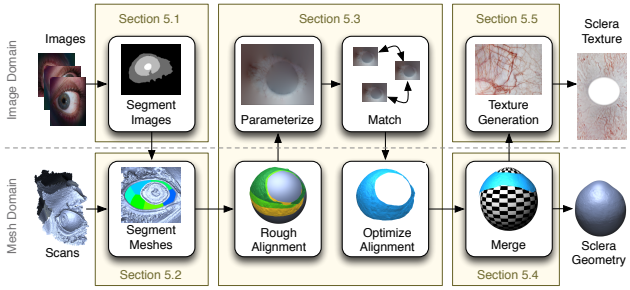

Figure 5: *The sclera reconstruction operates in both image and mesh domains. The input images and meshes are segmented (Sections 5.1 and 5.2). The partial scans from several eye poses are registered (Section 5.3) and combined into a single model of the sclera using a generic proxy (Section 5.4). A high-resolution texture of the sclera is acquired and extended via texture synthesis (Section 5.5).*

5.1 Image Segmentation

As indicated in Section 3, the individual components of the eye require dedicated treatment, and thus the first step is to segment the input images to identify skin, sclera, iris, and pupil regions. We acquire approximately 140 images for a single eye dataset, considering all the poses, pupil dilations and multiple cameras, which would make manual segmentation tedious. Therefore, a semi-supervised method is proposed to automate the process. All images are captured under similar conditions, and thus the appearance of the individual classes can be expected to remain similar. We therefore employ a nearest-neighbor classification. We manually segment one of the images into skin, sclera, iris and pupil regions (Figure 6a). These serve as examples, from which the algorithm labels the pixels of the other images automatically by assigning the label of the most similar example pixel. Similarity is computed in a lifted 21 dimensional feature space of 15 color and 6 Haralick texture features [Haralick 1979], and has proven to provide sufficiently accurate and robust results. This classification is fast since every pixel is treated independently. We obtain high quality classification by employing a post-processing step that uses the following topological rules:

- The iris is the largest connected component of iris pixels.
- There is only a single pupil and the pupil is inside the iris.
- The sclera part(s) are directly adjacent to the iris.

Figure 6b shows the final classification results for a subset of images, based on the manually annotated exemplar shown in (a).

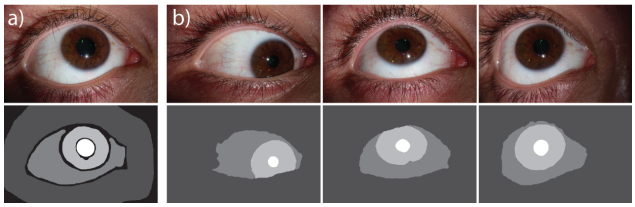

Figure 6: *Pupil, iris, sclera, and skin classification with manual labels (a) and examples of automatically labeled images (b).*

5.2 Mesh Segmentation

Given the image-based classification, we wish to extract the geometry of the sclera from the initial mesh reconstructions from Section 4.4. While the geometry is mostly accurate, the interface to the iris and skin may contain artifacts or exhibit over-smoothing, both of which are unwanted properties that we remove as follows.

While a single sphere only poorly approximates the shape of the eyeball globally (refer to Figure 12 in the results), locally the surface of the sclera may be approximated sufficiently well. We thus over-segment the sclera mesh into clusters of about 50mm^2 using k-means and fit a sphere with a 12.5mm radius (radius of the average eye) to each cluster. We then prune vertices that do not conform with the estimated spheres, either in that they are too far off surface or their normal deviates strongly from the normal of the sphere. We found empirically that a distance threshold of 0.3mm and normal threshold of 10 degrees provide good results in practice and we use these values for all examples in this paper. We iterate these steps of clustering, sphere fitting, and pruning until convergence, which is typically reached in less than 5 iterations. The result is a set of partial sclera meshes, one for each captured gaze direction.

5.3 Pose Registration

The poses are captured with different gaze directions *and* slightly different head positions, since it is difficult for the subject to remain perfectly still, even in the custom acquisition setup. To combine the partial sclera meshes into a single model, we must recover their rigid transformation with respect to the reference pose. ICP [Besl and McKay 1992] or other mesh-based alignment methods perform poorly due to the lack of mid-frequency geometric detail of the sclera. Feature-based methods like SIFT, FAST, etc. fail to extract reliable feature correspondences because the image consists mainly of edge-like structures instead of point-like or corner-like structures required by the aforementioned algorithms. Instead, we rely on optical flow [Brox et al. 2004] to compute dense pairwise correspondences.

Optical flow is an image based technique and typically only reliable on small displacements. We therefore align the poses first using the gaze direction and then parameterize the individual meshes jointly to a uv-plane. The correspondences provided by the flow are then employed to compute the rigid transformations of the individual meshes with respect to the reference pose. These steps are iterated, and convergence is typically reached in 4-5 iterations. In the following we will explain the individual steps.

Initial Alignment: The gaze direction is estimated for every pose using the segmented pupil. Since the head does not remain still during acquisition, the pose transformations are estimated by fitting a sphere to the reference mesh and aligning all other meshes so that their gaze directions match.

Joint Parameterization: The aligned meshes are parameterized to a common uv-space using spherical coordinates. Given the uv-parameterization, we compute textures for the individual poses by projecting them onto the image of the camera that is closest to the line of sight of the original pose. This naive texturing approach is sufficient for pose registration, and reduces view-dependent effects that could adversely impact the matching.

Correspondence Matching: We compute optical flow [Brox et al. 2004] of the individual sclera textures using the blue channel only, since it offers the highest contrast between the veins and the white of the sclera. The resulting flow field is sub-sampled to extract 3D correspondence constraints between any two neighboring sclera meshes. We only extract constraints which are both well localized and well matched. Matching quality is assessed using the normalized cross-correlation (NCC) within a $k \times k$ patch. Localization is directly related to the spatial frequency content present within this patch, quantified by the standard deviation (SD) of the intensity values. We set $k = 21$ pixels, NCC > 0, and SD < 0.015 in all our examples.

Optimization: The rigid transformations of all the poses are jointly optimized with a Levenberg-Marquardt optimizer so that the weighted squared distances between the correspondences are minimized. The weights reflect the local rigidity of the detected correspondences and are computed from the Euclidean residuals that remain when aligning a correspondence plus its 5 neighbors rigidly. The optimization is followed by a single ICP iteration to minimize the perpendicular distances between all the meshes.

5.4 Sclera Merging

After registering all partial scans of the sclera, they are combined into a single model of the eyeball. A generic eyeball proxy mesh, sculpted by an artist, is fit to the aligned meshes and the partial scans are merged into a single mesh, which is then combined with the proxy to complete the missing back of the eyeball.

Proxy Fitting: Due to the anatomy of the face, less of the sclera is recovered in the vertical direction and as a result the vertical shape is less constrained. We thus fit the proxy in a two step optimization. In the first step we optimize for uniform and in the second step for horizontal scaling only. In both steps we optimize for translation and rotation of the eyeball while keeping the rotation around the optical axis fixed.

Sclera Merging: The proxy geometry prescribes the topology of the eyeball. For every vertex of the proxy, a ray is cast along its normal and intersected with all sclera meshes. The weighted average position of all intersections along this ray is considered to be the target position for the vertex and the standard deviation of the intersections will serve as a confidence measure. The weights are a function of the distance of the intersection to the border of the mesh patch and provide continuity in the contributions.

Eyeball Merging: The previous step only deforms the proxy where scan data is available. To ensure a smooth eyeball, we propagate the deformation to the back of the eyeball using a Laplacian deformation framework [Sorkine et al. 2004]. The target vertex positions and confidences found in the previous step are included as weighted soft-constraints. The result is a single eyeball mesh that fits the captured sclera regions including the fine scale details and surface variation, and also smoothly completes the back of the eye.

5.5 Sclera Texturing

As a final step, we compute a color for each point on the reconstructed sclera surface by following traditional texture mapping approaches that project the 3D object onto multiple camera images. In our case, we must consider all images for all eye poses and use the computed sclera segmentation to identify occlusion. One approach is to naively choose the most front-facing viewpoint for each surface point, however this leads to visible seams when switching between views. Seams can be avoided by averaging over all views, but this then leads to texture blurring. An alternative is to solve the Poisson equation to combine patches from different views while enforcing the gradient between patches to be zero [Bradley et al. 2010], but this can lead to strong artifacts when neighboring pixels at the seam have high gradients - a situation that often occurs in our case due to the high contrast of a red blood vessel and white sclera. Our solution is to separate the high and low frequency content of the images. We then apply the Poisson patch combination approach only for the low frequency information, which is guaranteed to have low gradients. We use the naive best-view approach for the high frequencies, where seams are less noticeable because most seams come from shading differences and the shading on a smooth eye is low-frequency by nature. After texture mapping, the frequencies are recombined. Figure 7b shows the computed texture map for the eye in 7a.

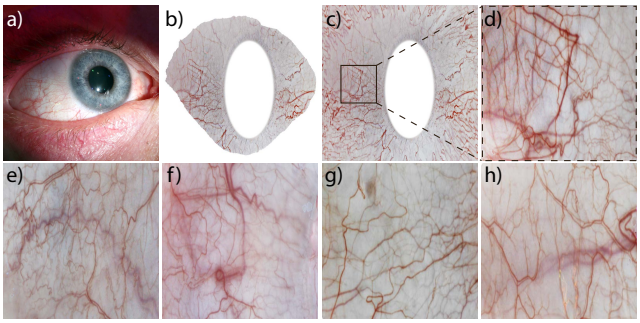

Figure 7: *The sclera is textured from multiple views of multiple different eye poses. The resulting texture map (b) for a given eye (a) contains all the visible parts of the sclera. We further complete the texture map through texture synthesis (c). Our textures can have very high resolution details (d-h).*

Our texturing approach will compute a color for each point that was seen by at least one camera, but the occluded points will remain colorless. Depending on the intended application of the eye reconstruction, it is possible that we may require texture at additional regions of the sclera, for example if an artist poses the eye into an extreme gaze direction that reveals part of the sclera that was never observed during capture. For this reason, we synthetically complete the sclera texture, using texture synthesis [Efros and Leung 1999]. In our setting, we wish to ensure consistency of blood vessels, which should naturally continue from the iris towards the back of the eye. We accomplish this by performing synthesis in Polar coordinates, where most veins traverse consistently in a vertical direction, and we seed the synthesis with a few vertical vein samples. Figure 8 demonstrates the rotated synthesis, which we perform only on the high frequencies in order to avoid synthesized shading artifacts. Corresponding low-frequency content is created by smooth extrapolation of the computed low-frequency texture.

Finally, we can also synthesize missing surface details in the back of the eye. We use the same texture synthesis approach, but in-

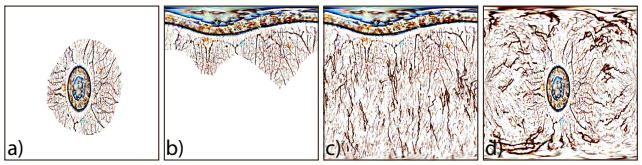

Figure 8: *Texture synthesis is performed on the high frequency information in order to complete the texture. A captured texture (a) is rotated to Polar coordinates (b) and synthesis is performed in a way that preserves vein orientation (c). The final texture is rotated back to Cartesian coordinates (d).*

stead we operate on a displacement map, which is computed as the difference between the original and a smoothed version of the reconstructed eyeball. The final result is a complete eyeball with continuous texture and displacement at all points. We show a complete texture and zoom region in Figure 7 (c-d), and highlight a few zoom regions of different eye textures in Figure 7 (e-h).

6 Cornea

Given the reconstructed sclera, we now describe our technique to reconstruct the transparent cornea. Although the cornea consists of several thin layers with different optical properties, we found it sufficient to model the cornea as a single surface with a single medium respectively index of refraction inside the eye. We use a surface optimization method that aims to satisfy constraints from features that are either reflected off or refracted through the cornea.

6.1 Theory

Reconstructing transparent surfaces requires different approaches than diffuse surface reconstruction since the surface is not directly visible. Transparent surfaces are generally not completely transmissive, but a fraction of light is reflected if the refractive indices of the media involved differ. Thus, a bright light placed in front of the cornea will cause a visible highlight that provides a cue about the surface. Unfortunately, the position of the highlight is view-dependent and cannot directly be used in a mulit-view setting.

On the other hand, for a single view there is an ambiguity between the depth along the viewing ray corresponding to a highlight and the normal of the surface. For every position along a viewing ray there exists a surface normal reflecting the ray to the origin of the light (Figure 9a, green). This creates a surface normal field defined by all possible viewing ray direction and depth combinations. A similar surface normal field is produced from refractions (Figure 9a, red).

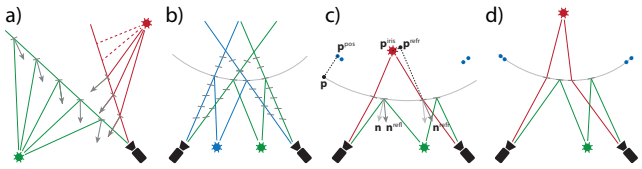

Figure 9: *The depth/normal ambiguity of a highlight (a) and the sparse normal field in a multi-view setting (b). Corneal constraints before (c) and after optimization (d).*

The reflection and refraction surface normal fields of different views only coincide at the position of the actual surface as illustrated in Figure 9b. We use this property to reconstruct the cornea.

Our system however produces only a sparse sampling of the normal fields as we employ only a few LEDs. We therefore need to add regularization to ensure a unique solution, which is provided through the chosen surface representation. We employ an open uniform B-spline surface with 100 control points. This surface has more representation power than the traditionally employed 4th order Zernike polynomials [Ares and Royo 2006; Smolek and Klyce 2003] yet can be controlled locally, which is beneficial for optimization. The control points are spaced regularly and initialized to the surface of the eyeball proxy introduced in Section 5.4. The position of the boundary control points are optimized such that the surface boundaries fit the proxy geometry. The boundary control points are kept fixed and are not part of the following surface optimization.

6.2 Constraint Initalization

The corneal surface is optimized using three different types of constraints: reflection, refraction and position constraints.

Reflection Constraints: The 9 calibrated LEDs placed in front of the cornea are imaged as highlights in the different views. From these highlights we extract reflection constraints, which prescribe the normal for any point along the viewing ray through the highlight. Since the cornea is convex, every LED-view pair contributes one constraint assuming the reflection of the LED is visible in the view. In addition, since we registered the poses in Section 5.3 we can combine constraints from all different poses. The highlights are detected and identified similarly as in the calibration phase (Section 4.2). While the highlights in the calibration images are acquired in complete darkness, now they appear superimposed on the iris in the input images, which can lead to false positive detections. Thus, we remove these unwanted detections by fitting a 2D Gaussian curve to the intensity profiles of all the highlight candidates to determine their width. Since all the LED highlights have a constant size we can remove false positives with a lower (3px) and upper (15px) threshold on the standard deviation of the Gaussian.

Refraction Constraints: Conceptually refraction constraints are very similar to reflection constraints. Instead of observing the reflected highlight of a known LED, we instead observe the refraction of a feature on the iris at unknown position. Furthermore, the angle of refraction depends on the refractive index. Both the position of the feature and the refractive index are included as unknowns into the optimization and solved for. A feature point on the iris contributes one refractive constraint per view. The corresponding image location in the different views is estimated via optical flow [Brox et al. 2004]. Features are filtered as described in Section 5.3 using NCC$>$0.6 and SD$<$0.02.

As for reflection constraints, we can combine refraction constraints from all poses. The distribution density of the features varies substantially, as we wont have any in the pupil for example. To account for this we weigh the constraints by the local density, approximated by the distance d to the 10th nearest constraint as $w^{refr} = NCC/d^2$ where NCC is the average normalized cross correlation score between corresponding image patches used as a measurement of the quality of the constraint.

Position Constraints: Position constraints are extracted from the merged sclera mesh (Section 5.4). Their purpose is to provide a continuous transition from the cornea to the sclera. We randomly sample position constraints on the sclera in the vicinity of the corneal boundary. To ensure a good distribution, we reject constraints that are closer than 1mm to each other.

6.3 Surface Reconstruction

With a given set of reflection, refraction and position constraints and an initial guess of the surface, the unknown parameters are optimized with a two stage approach. More specifically, we optimize

the control points of the B-Spline, the refractive index and the unknown positions of the feature points on the iris which are used for the refraction constraints. This amounts to a non-linear optimization which we solve using the Levenberg-Marquardt algorithm by minimizing the error

$$E^{tot} = \lambda^{pos} E^{pos} + \lambda^{refl} E^{refl} + \lambda^{refr} E^{refr}, \quad (1)$$

where $\lambda^{pos} = 0.1$, $\lambda^{refl} = 1$, and $\lambda^{refr} = 1$ are user-defined parameters. The error for the position constraints $\mathcal{P}$ is given as

$$E^{pos} = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} \|\mathbf{p}_i - \mathbf{p}_i^{pos}\|^2, \quad (2)$$

where $\mathbf{p}^{pos}$ denotes the position of the constraint and $\mathbf{p}$ the nearest point on the corneal surface. The error for the reflection constraints $\mathcal{Q}$ is given as

$$E^{refl} = \frac{1}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} \left\|\mathbf{n}_i - \mathbf{n}_i^{refl}\right\|^2, \quad (3)$$

where $\mathbf{n}$ is the current and $\mathbf{n}^{refl}$ the targeted surface normal. The error for the refraction constraints $\mathcal{R}$ is given as

$$E^{refr} = \frac{1}{|\mathcal{R}|} \sum_{i \in \mathcal{R}} w_i^{refr} \left\|\mathbf{p}_i^{iris} - \mathbf{p}_i^{refr}\right\|^2, \quad (4)$$

where $\mathbf{p}^{iris}$ is the point on the iris, $\mathbf{p}^{refr}$ the closest point on the refracted ray and w^{refr} its corresponding weight. Optimizing the distance to the closest point has proven to be more stable than optimizing the mismatch of the normals analogously to Equation 3.

In the first step we optimize for the control point positions of the B-spline surface. They are optimized only along the opical axis of the eye and the boundary control points are kept fixed at all times. After convergence the surface is kept fixed and we optimize for the refraction constraint points on the iris ($\mathbf{p}^{iris}$) and the refractive index. We iterate by alternating the two steps until the overall improvement drops below $10e^{-10}$. The initial and optimized corneal surface plus constraints are visualized in Figure 10 for one dataset.

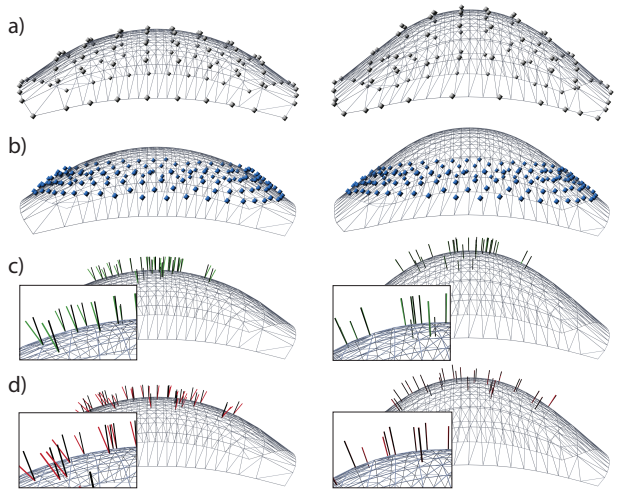

Figure 10: *Visualization of the B-spline control points (a), the position constraints (b), and a subset of the reflection (c) and refraction (d) constraints on the initial (left) and optimized (right) surfaces.*

6.4 Cornea-Eyeball Merging

We update the eyeball mesh with the optimized cornea by smoothly blending the corneal surface into the eyeball mesh. First, corneal samples are computed for each eyeball vertex by intersecting the cornea in the direction of the eyeball normals. Second, the iris masks are dilated, blurred, projected onto the cornea, and averaged to compute blending weights. Finally, the eyeball vertices are combined with the corneal samples by weighting them with the weights.

7 Iris

We now move to the final component of the eye, the iris. In contrast to the sclera, we cannot perform traditional multi-view reconstruction to obtain the iris geometry because the refractive cornea distorts the views of the iris. Additionally, the cornea transitions smoothly in opacity from fully transparent to fully opaque at the sclera, and this smooth transition can confuse multi-view correspondence matching. For these reasons, we create a specific iris reconstruction algorithm that is designed to handle these constraints. Since the iris is coupled with the pupil, our method begins by localizing the pupil in 3D. The iris geometry is then reconstructed and filtered, using the pupil as initialization. Finally, we combine iris reconstructions from captures with different pupil dilations, allowing us to parameterize and animate the deformation of the iris during pupillary response.

7.1 Pupil Reconstruction

The pupil has a very prominent position at the center of the eye, which makes it visually important and artifacts on its boundary would be clearly visible. Therefore, we require a reconstruction method for the pupil boundary that is robust with respect to perturbations like, for example, those caused by the flash highlight. This robust boundary is used to constrain the iris and also to guide the initial meshing of the iris.

Initialization: The pupil is initialized with the pupil mask boundaries that we detect in image space. Each boundary is triangulated from multiple views, taking into account refraction at the cornea, and we fit a circle to the triangulated points. The required image correspondences for the triangulation are obtained from the optical flow, which we already computed for the refraction constraints of the cornea optimization.

Refinement: As this initial estimate tends to be rather inaccurate due to inconsistencies between pupil masks, we refine the estimated 3D circle in an optimization that uses two data terms and two regularization terms. The data terms come from two additional cues about the pupil location, 1) an image term that incorporates the result of an image-based pupil detection algorithm, and 2) a mesh term that incorporates an approximate 3D surface reconstruction of the pupil region, triangulated from image correspondences found using optical flow. The two regularization terms control the overall shape and smoothness of the pupil. Based on these terms, we define an energy function for the pupil as

$$E = \lambda_I E_I + \lambda_M E_M + \lambda_C E_C + \lambda_S E_S, \quad (5)$$

which we minimize for a set of $n = 50$ pupil samples taken on the initial circle, with weights of $\lambda_I = 10$, $\lambda_M = 1000$, $\lambda_C = 10000$, and $\lambda_S = 1000$ for all data sets. In the following, we will describe each of the energy terms in more detail.

Image Term: We project the initial pupil circle into the cameras and blur the images radially along the produced ellipses. We then use a radial edge detector to locate the edge between the pupil and

the iris, and we apply radial non-maximum suppression (NMS) to the response. We define the image data term as

$$E_I = \frac{1}{n}\sum_{i=1}^{n}\left\|P(\mathbf{p}_i) - \mathbf{p}_i^{edge}\right\|^2, \quad (6)$$

where $P(\mathbf{p})$ is the projection of sample point $\mathbf{p}$ into the image plane through the cornea, and $\mathbf{p}^{edge}$ is the position of the closest point on the detected edge.

Mesh Term: We create an approximate 3D surface mesh in the vicinity of the pupil by triangulating rays from multiple views refracted at the corneal interface, again with the help of optical flow to provide correspondences. The mesh term for the pupil location then consists of the distances between the pupil samples and the generated mesh, given by

$$E_M = \frac{1}{\sum_{i=1}^{n} c_i}\sum_{i=1}^{n} c_i\left\|\mathbf{p}_i - \mathbf{p}_i^{mesh}\right\|^2, \quad (7)$$

where the distances are weighted with the triangulation confidences c of the mesh. The triangulation confidence is defined as a linear function of the triangulation residuals, which maps a residual of 0mm to a confidence of 1 and a residual of 0.05mm to a confidence of 0 and clamps all the values outside this range.

Regularization Terms: We allow the samples to deviate orthogonally from the perfect circle, but we penalize these deviations with

$$E_C = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{p}_i - \mathbf{p}_i^{circle}\right\|^2, \quad (8)$$

where $\mathbf{p}^{circle}$ is the corresponding point of $\mathbf{p}$ on the circle. To obtain a smooth pupil we also penalize strong changes in the deviations from one sample to the next, using the following smoothness term, where r is the radial and o the orthogonal component of the offset with respect to the circle.

$$E_S = \frac{1}{n}\sum_{i=1}^{n}\left[(2r_i - r_{i+1} - r_{i-1})^2 + (2o_i - o_{i+1} - o_{i-1})^2\right], \quad (9)$$

Finally, we minimize the sum of all these terms with the Levenberg-Marquardt algorithm to find the position, the radius, and the per-sample deviations from a circle of the pupil. During the optimization, we constrain the normal of the pupil circle to the normal of the plane fit to iris mesh samples taken 1 mm away from the initial pupil boundary estimate to be more robust. Figure 11 illustrates the resulting sample positions both in 3D and projected onto an image (in green), given the initial estimate (in red).

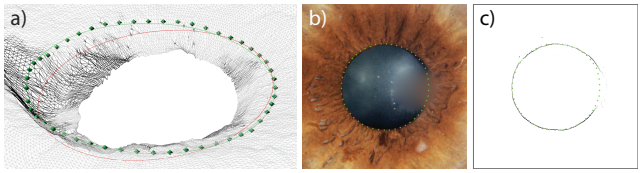

Figure 11: *Pupil reconstruction. Given an initial pupil boundary estimate (red) from the triangulated image-based pupil masks, we solve for the optimal pupil boundary (green). The resulting pupil samples are shown in 3D (a), projected onto one image (b), and overlaid onto the response of the pupil edge detector (c).*

7.2 Iris mesh generation

We use the reconstructed pupil boundary to initialize the iris mesh. Starting with a closed uniform B-Spline that we fit to the optimized pupil samples, we scale the spline radially in 0.025mm steps to create a sequence of larger and larger rings up to an iris radius of 7mm. These rings are sampled 600 times and a triangle mesh is created. This will serve as the topology for the iris.

In a second step, we reconstruct the correct position of each iris vertex. Each vertex is projected (through the cornea) into a reference camera, where flow-based correspondences to other views are computed. We triangulate the vertex position by minimizing the squared distances between the vertex and the refracted rays formed by the correspondences. This minimization is equivalent to minimizing the surface error defined in Section 6.3. In addition, the rays are weighted by the root mean square difference of the corresponding 7x7 pixel blocks in image space. In order to reduce high frequency noise, the entire mesh reconstruction process is repeated for a second reference camera to obtain a second mesh hypothesis which is combined with the first one through weighted averaging.

7.3 Mesh cleanup

The reconstructed iris mesh can be noisy and distorted at the boundaries due to the translucent sclera affecting the optical flow. We perform four operations to filter the iris mesh.

Spike Filtering: Spikes are detected by computing a 3-ring neighborhood around each vertex. If the distance between the vertex and the mean of the neighboring vertices exceeds a threshold (set to 0.05mm), then the vertices inside the ring are smoothed by solving a Laplacian system, keeping the rest of the vertices fixed.

Boundary Deformation: Two criteria are used to label distorted boundary vertices: a threshold on the triangulation residuals (set to 0.05mm) and an angle threshold between the smoothed vertex normal and the normal of the pupil set to 30 degrees. We dilate the labeled region and smooth those vertices in the normal direction.

Mesh Relaxation: The mesh is relaxed locally to improve the triangulation by removing skinny or overlapping triangles.

Pupil Constraint: The vertices at the pupil boundary are constrained to the detected pupil shape. The constraint is enforced with a local Laplacian system, where the pupil vertices as well as all mesh vertices farther than 1mm from the pupil are constrained. The vertices in-between are deformed but the local shape is preserved.

Finally, the two independently triangulated and cleaned mesh hypotheses are averaged to create the iris mesh.

7.4 Mesh Propagation

We now combine iris reconstructions from captures with different pupil dilations. Each mesh is reconstructed independently, with different topology and vertex counts. We wish to compute a new set of iris meshes that are in vertex-correspondence, allowing us to compute a per vertex deformation model.

We begin by computing per camera optical flow [Brox et al. 2004] between neighboring poses. Since the vertices are propagated from one pose to the next, drift might accumulate. To minimize the total amount of drift we select a reference pose in the middle of the dilation sequence and compute the optical flow in both dilation directions from there. To find the vertex correspondences we project each vertex from the source mesh into all the target pose cameras taking into account the refraction at the cornea. With the resulting image positions and the optical flows we compute a set of rays that

we refract at the cornea and intersect with the iris of the target pose. The target pose vertex is computed as the median of all the intersections. To ensure a clean pupil we enforce the pupil constraint and relax the mesh in the same way as described in section 7.3.

7.5 Temporal Smoothing and Interpolation

In order to animate the pupil dilation, we will use the captured pupil poses as keyframes and interpolate linearly in-between. In practice we found that the dilation of the pupil cannot be accurately controlled, and so the pupil diameter tends to decrease in irregular steps. This can lead to multiple poses with very similar diameters and geometry, but with different high frequency reconstruction noise, which leads to artifacts when interpolating. In order to smoothly integrate meshes from similar pupil radii, we compute two linear regression models for all poses within a distance of 1mm pupil radius. The first regression model expresses the vertex position and the second model the Laplacian vector as a function of the pupil radius. We solve for the smoothed mesh by evaluating both models and solving the resulting Laplacian system with equal weights given to the Laplacians and the positions.

7.6 Iris Texturing

Iris textures can be computed from a single view, but these textures will contain undesired artifacts like highlights, washed out regions close to the boundary, dust on the cornea, etc. These artifacts can be attenuated by combining the textures from multiple views of the same iris dilation. We compute a contribution map for each view which is set to 1 if the pixel is the most saturated from all the candidates and to 0 otherwise. These maps are then blurred with a small Gaussian kernel of 3 pixels. Based on these contribution maps, the textures from the different views are blended into a single texture. Picking the most saturated pixels will reduce artefacts caused by illumination pollution from the flash light and by superposition of the white sclera at the semi-transparent sclera-cornea transition alike. Then, we combine the textures from several iris dilations using the median to attenuate shading changes caused by the deforming iris.

8 Results

In this section we highlight the results of our eye capture technique by illustrating the reconstructions of a variety of human eyes, each with its own intricacies and details.

We begin by analyzing the common assumption that eyes can be modelled as two spheres, a large one for the eyeball and a smaller one for the cornea. In our work we show that this assumption is inaccurate, which we can illustrate by overlaying a cross-section of a captured eye on top of the simple model (Figure 12, left). Furthermore, it is often assumed that an eye is symmetric about the view vector and that the left and right eye can be modelled similarly. By capturing both the left and right eye of an actor, we demonstrate that each eye is in fact unique and shows strong asymmetry individually, but when combined the expected left/right symmetry is clearly visible. We believe these results have the potential to change how eyes are traditionally modelled in computer graphics.

Our eye capture method is robust, which we highlight by reconstructing 9 different eyes from 6 different actors. The full set of reconstructions, shown in Figure 13, contains a variety of different iris colors, surface details, textures, and overall eye shapes. Each eye has unique details, but we observed that the differences between people are more significant than the differences between the two eyes of the same person, an expected phenomenon that helps to validate our reconstruction results. For example, the two brown eyes in the center (5th and 6th from left) are larger than the rest. These represent the eyes of an actor with severe *myopia*

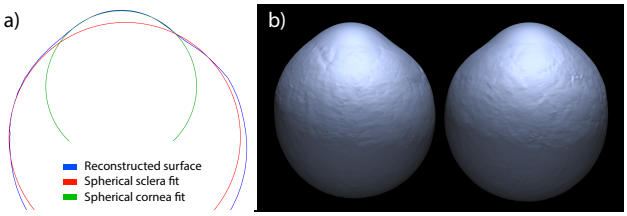

Figure 12: *The traditional assumption that an eye can be modelled as two spheres (red and green) is inaccurate, as indicated by a top-view cross-section of our reconstruction in blue (left). Eyes also exhibit strong asymmetry, which we show by reconstructing both the left and right eyes of the same actor (right).*

(or short-sightedness), which is often correlated with larger-than-normal eyes [Atchison et al. 2004].

Every human eye is unique and contains minor intricacies that add to the identity of the person. Our capture approach aims to reconstruct all the visible intricacies. In particular, our sclera reconstruction is able to acquire high-resolution surface variation including small details and Pingueculas, as shown in Figure 14.

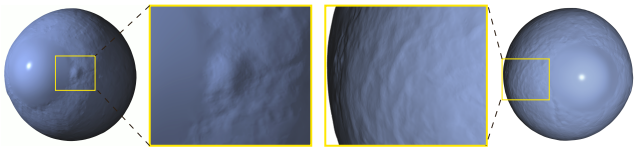

Figure 14: *Our sclera reconstruction technique is able to acquire fine scale details including Pingueculas (left) and surface variation (right) that is unique to each eye.*

Even more unique is the iris. Figure 15 illustrates one pose of the reconstructed irises for our 9 actors, visualized on their own with blue shading for comparing the geometry. The individuality of iris shape from eye to eye is clearly visible, again highlighting the importance of capturing real eyes using the proposed technique. Figure 16 shows a close-up view of a captured iris with both surface details and texture, rendered with refraction through the cornea.

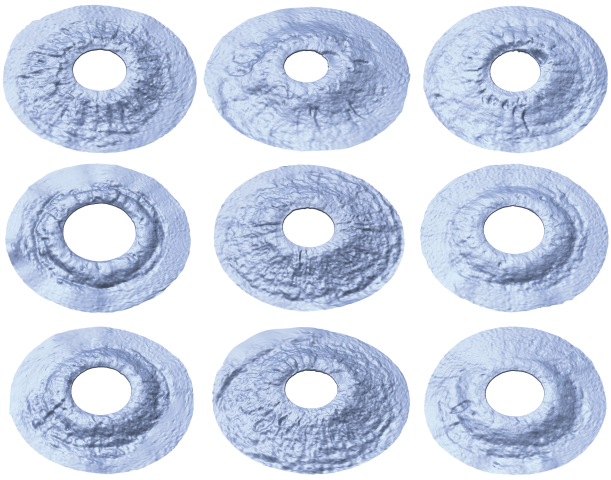

Figure 15: *We highlight the uniqueness of each individual iris by visualizing the 9 captured irises with blue shading.*

One of the most interesting features of human eyes is the time-varying deformation of the iris during pupillary response. Our

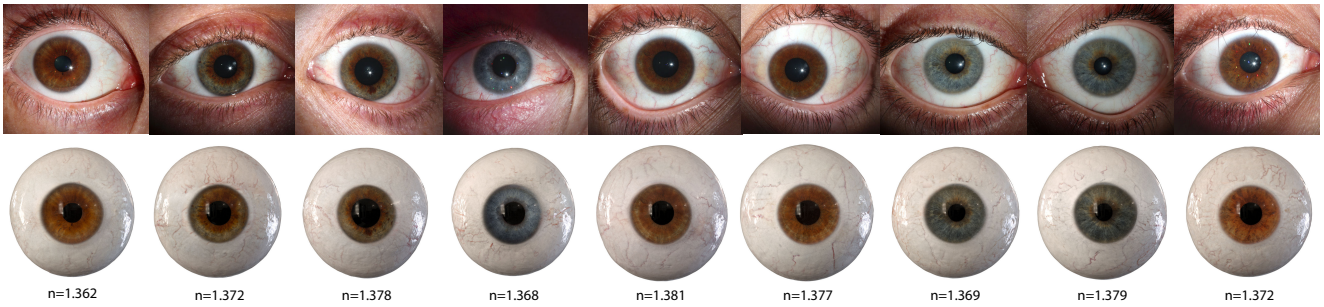

Figure 13: *We highlight the robustness of our technique by capturing a wide variety of eyes. This dataset consists of different iris colors, individual sclera textures, and unique geometry for each eye. In particular, we can see that the two brown eyes in the center are larger than the others - further highlighting the importance of capture over generic modelling. The measured index of refraction is listed under each eye.*

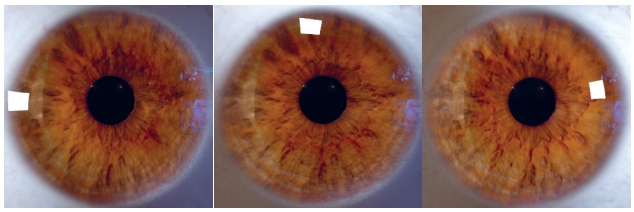

Figure 16: *A close-up of detailed iris geometry and texture captured with our method, rendered in high-quality with refraction through the cornea and three different light positions.*

method is able to recover this deformation, which we illustrate for one actor in Figure 17. As the pupil changes size, our reconstruction shows that the iris dilator muscle creates significant out-of-plane deformation, which largely contributes to the realistic appearance of the eye. To further illustrate how unique this effect is for each iris, we provide side-view renders for two additional irises and three pupil radii in Figure 18.

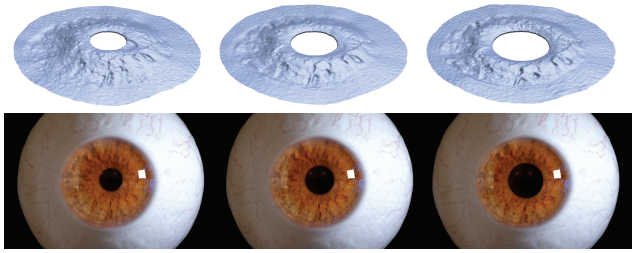

Figure 17: *We measure the iris under various amounts of pupil dilation. As can be seen, the iris dilator muscle creates significant out-of-plane deformation as the pupil becomes larger (left to right).*

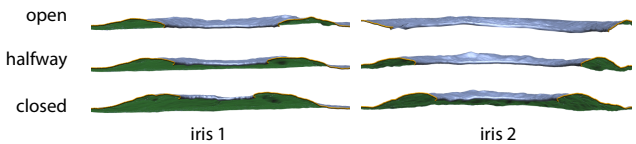

Figure 18: *We highlight the uniqueness of each eye's iris deformation during pupil dilation by showing the deformations from a side view for two different eyes and three different pupil sizes.*

The ability to reconstruct a per-vertex deformation model for the iris during pupil dilation allows us to animate the captured eyes. We show two different applications for iris animation in Figure 19. The first is a motion capture scenario. Analogous to the way facial animation rigs are often built from high-quality scan data and then later animated from low-resolution mo-cap markers, our captured irises can be animated from a single low-quality video stream. As a demonstration, we detect the pupil size of an actor in each frame of such a video and compute the corresponding iris shape for a captured actor (Figure 19, top). A second application is to automatically make a digital character respond to lighting changes in the 3D environment. Using predicted pupillary response curves introduced in the seminal work of Pamplona et al. [2009], we can animate the captured iris geometry to show a character dynamically responding to a light source turning on and off (Figure 19, bottom). As these applications target iris animation, the results are best viewed in the accompanying supplemental video.

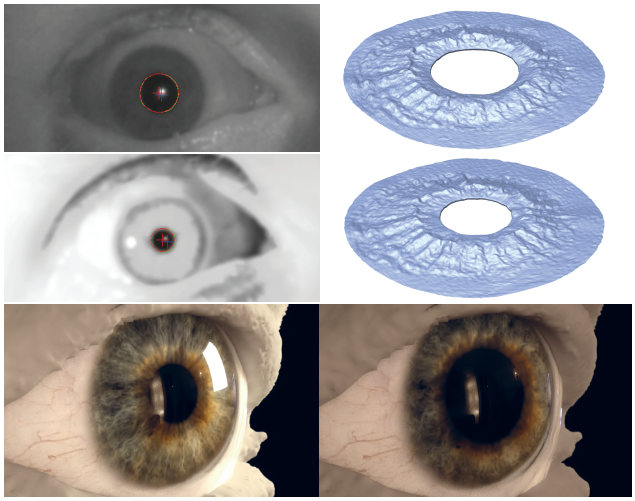

Figure 19: *We can apply the measured iris deformation in a pupil dilation animation. Here we show two applications: one where an actor's pupil is tracked in a single low-quality infra-red video and the corresponding radius is applied to our model (top). A second application is to automatically make a digital double respond to lighting changes in the 3D environment (bottom).*

We compare our results qualitatively with the seminal work of François et al. [François et al. 2009] in Figure 20. While the strength of their approach is its simplicity, our method arguably excels in quality. Since we aim to accurately reconstruct all the intricacies of the eye, we more faithfully capture the uniqueness and realism of eyes. In particular, our reconstructions show the asymmetric shape of the sclera and fine scale surface variation. Our iris geometry is reconstructed rather than heuristically synthesized, and we even recover small defects like the aforementioned Pingueculas and the non-circular transition between sclera and iris in Figure 20.

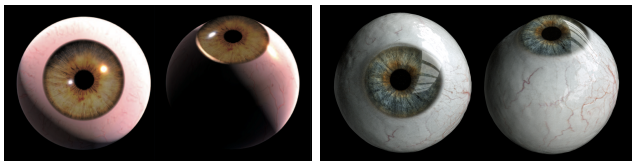

Figure 20: *We show a comparison with François et al. [2009] on the left. They employ a generic eyeball model combined with a heuristic to synthesize the iris morphology. Note how our results shown on the right faithfully capture the intricacies of this particular eye, such as its asymmetric shape, the small surface variation, and the non-circular iris-sclera transition.*

In order to provide context for visualizing the captured eyes we combine them with the partially reconstructed face scans of the actors. We use a simple combination process that automatically fits the face geometry around the back of the eyeball using a Laplacian deformation scheme. While the approach is rudimentary, the result is sufficient to simulate an eye socket for holding the reconstructed eye. Several results for different actors are shown in Figure 21, rendered from multiple viewpoints. We note that more sophisticated methods for capturing the face region around the eyeball would be ideal topics for future research.

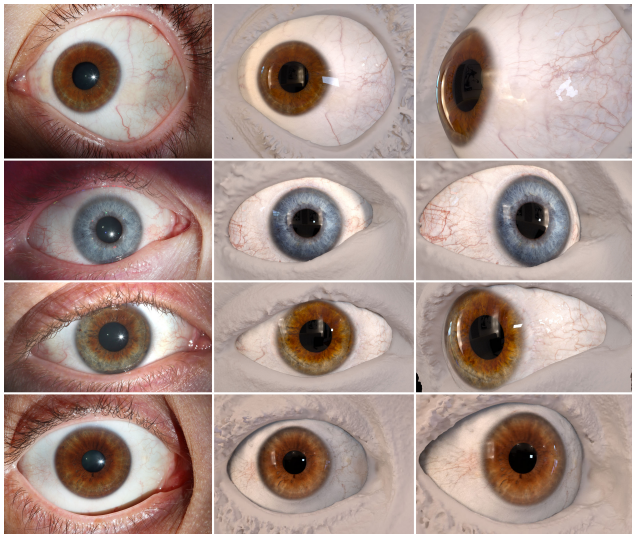

Figure 21: *We further demonstrate our results by combining the captured eyes with partial face scans, and rendering from various viewpoints with different environment lighting. This figure shows how the reconstruction results could be used in the visual effects industry for creating digital doubles.*

Finally, we wish to highlight the potential impact that capturing real eyes can have in creating artistic digital doubles - a task that is often performed for visual effects in films. To this end, we combine both of the captured eyes of an actor together with a face scan to create a compelling rendition of an artistically designed digital human character, as shown in Figure 22. Such a result would traditionally take significant artistic skill and man-hours to generate, in particular if the digital character should closely resemble a real actor. Our result was created with very little effort, thanks to our new method for capturing real human eyes.

All our textured results are rendered in a single pass using Autodesk Maya with Octane Render. We use built-in diffuse materials with subsurface scattering for the sclera and the iris, and reflective/refractive materials for the cornea plus a water layer created by extruding the sclera by 0.1 mm. The total processing time to reconstruct a complete eye on a standard Windows PC with a 3.2 Ghz 6-core CPU is approximately 4 hours (2 hours for initial reconstruction, 20 minutes for the sclera, 5-10 minutes for the cornea, 1 hour for the iris, and 40 minutes for unoptimized texture synthesis). The main bottleneck is the computation of optical flow.

Figure 22: *We combine both captured eyes of an actor together with a face scan to further demonstrate how our results can be used to create artistic digital doubles.*

9 Conclusion

We present a unified framework to capture shape and texture of the most prominent components of the eye at unprecedented level of detail, including the *sclera*, the *cornea*, and the *iris*. We demonstrate that the generic eye models typically used in our community are not sufficient to represent all the intricacies of an eye, which are very person-specific, and we believe that the findings of this paper have the potential to alter our community's current assumptions regarding human eyes. In addition, we present the first method for reconstructing detailed iris deformation during pupil dilation, and demonstrate two applications of how data-driven iris animations combined with our high-quality eye reconstructions can be used to add realism to digital doubles in visual effects.

Our method focuses purely on the static eye combined with the dynamic deformation of the iris. The dynamic aspects of an eyeball with its complex motion patterns would be an exciting future extension. It would also be interesting to model the variation present in the individual eyes, similar as has been done for faces [Blanz and Vetter 1999], to provide the user with a descriptive and parametric eye model that can be used to synthesize novel eyes.

Limitations. Our system is designed to capture the geometry of the eye with different pupil dilations, however we currently do not capture dynamic behavior such as saccades, hippus, or tremor. Also, since the iris is a volumetric object with partially translucent tissue it is difficult to reconstruct accurately. For this reason we approximate the iris as a surface, however we believe that the optical flow correspondences used to reconstruct that surface are sufficiently accurate to represent the iris with adequate details suitable for rendering, already a step forward from traditional practices that approximate the iris as a plane or cone. In this work we do not capture reflectance properties of the eyes, such as BRDF or BSSRDF parameters. All these topics could be addressed in future work.

Acknowledgements

We wish to thank Prof. Dr. Dr. Jens Funk and his team and Dr. med. Peter Maloca for the helpful discussions. Thanks to all our eye models and Lewis Siegel for his industry perspective.

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 photorealistic digital actor. *IEEE CG&A 30*, 4, 20–31.

ARES, M., AND ROYO, S. 2006. Comparison of cubic b-spline and zernike-fitting techniques in complex wavefront reconstruction. *Applied optics 45*, 27, 6954–6964.

ATCHESON, B., HEIDE, F., AND HEIDRICH, W. 2010. CALTag: High precision fiducial markers for camera calibration. In *International Workshop on Vision, Modeling and Visualization*.

ATCHISON, D. A., JONES, C. E., SCHMID, K. L., PRITCHARD, N., POPE, J. M., STRUGNELL, W. E., AND RILEY, R. A. 2004. Eye shape in emmetropia and myopia. *Investigative Ophthalmology & Visual Science 45*, 10, 3380–3386.

BEELER, T., BICKEL, B., SUMNER, R., BEARDSLEY, P., AND GROSS, M. 2010. High-quality single-shot capture of facial geometry. *ACM Trans. Graphics (Proc. SIGGRAPH) 29*, 40:1–40:9.

BEELER, T., HAHN, F., BRADLEY, D., BICKEL, B., BEARDSLEY, P., GOTSMAN, C., SUMNER, R. W., AND GROSS, M. 2011. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graphics (Proc. SIGGRAPH) 30*, 4, 75.

BEELER, T., BICKEL, B., NORIS, G., BEARDSLEY, P., MARSCHNER, S., SUMNER, R. W., AND GROSS, M. 2012. Coupled 3D reconstruction of sparse facial hair and skin. *ACM Trans. Graphics (Proc. SIGGRAPH) 31*, 4, 117:1–117:10.

BESL, P. J., AND MCKAY, N. D. 1992. A method for registration of 3-d shapes. *IEEE Trans. on PAMI 14*, 2, 239–256.

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3d faces. In *Proc. of the 26th annual conference on Computer graphics and interactive techniques*, 187–194.

BRADLEY, D., HEIDRICH, W., POPA, T., AND SHEFFER, A. 2010. High resolution passive facial performance capture. *ACM Trans. Graphics (Proc. SIGGRAPH) 29*, 4, 41.

BROX, T., BRUHN, A., PAPENBERG, N., AND WEICKERT, J. 2004. High accuracy optical flow estimation based on a theory for warping. In *ECCV*. Springer, 25–36.

EAGLE JR, R. 1988. Iris pigmentation and pigmented lesions: an ultrastructural study. *Trans. of the American Ophthalmological Society 86*, 581.

EFROS, A. A., AND LEUNG, T. K. 1999. Texture synthesis by non-parametric sampling. In *IEEE ICCV*, 1033–1038.

FRANÇOIS, G., GAUTRON, P., BRETON, G., AND BOUATOUCH, K. 2009. Image-based modeling of the human eye. *IEEE TVCG 15*, 5, 815–827.

GARRIDO, P., VALGAERTS, L., WU, C., AND THEOBALT, C. 2013. Reconstructing detailed dynamic face geometry from monocular video. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 32*, 6, 158.

GHOSH, A., FYFFE, G., TUNWATTANAPONG, B., BUSCH, J., YU, X., AND DEBEVEC, P. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graphics (Proc. SIGGRAPH Asia) 30*, 6, 129.

GRAHAM, P., TUNWATTANAPONG, B., BUSCH, J., YU, X., JONES, A., DEBEVEC, P., AND GHOSH, A. 2013. Measurement-based synthesis of facial microgeometry. In *Computer Graphics Forum*, vol. 32, 335–344.

HACHOL, A., SZCZEPANOWSKA-NOWAK, W., KASPRZAK, H., ZAWOJSKA, I., DUDZINSKI, A., KINASZ, R., AND WYGLEDOWSKA-PROMIENSKA, D. 2007. Measurement of pupil reactivity using fast pupillometry. *Physiological measurement 28*, 1, 61.

HALSTEAD, M. A., BARSKY, B. A., KLEIN, S. A., AND MANDELL, R. B. 1996. Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals. In *Proceedings of Computer graphics and interactive techniques*, 335–342.

HARALICK, R. M. 1979. Statistical and structural approaches to texture. *Proc. IEEE 67*, 5, 786–804.

HERNÁNDEZ, C., VOGIATZIS, G., AND CIPOLLA, R. 2008. Multiview photometric stereo. *IEEE PAMI 30*, 3, 548–554.

HUANG, D., SWANSON, E. A., LIN, C. P., SCHUMAN, J. S., STINSON, W. G., CHANG, W., HEE, M. R., FLOTTE, T., GREGORY, K., PULIAFITO, C. A., ET AL. 1991. Optical coherence tomography. *Science 254*, 5035, 1178–1181.

IHRKE, I., KUTULAKOS, K. N., LENSCH, H. P., MAGNOR, M., AND HEIDRICH, W. 2008. State of the art in transparent and specular object reconstruction. In *Eurographics 2008 - State of the Art Reports*.

LAM, M. W. Y., AND BARANOSKI, G. V. G. 2006. A predictive light transport model for the human iris. In *Computer Graphics Forum*, vol. 25, 359–368.

LEFOHN, A., BUDGE, B., SHIRLEY, P., CARUSO, R., AND REINHARD, E. 2003. An ocularist's approach to human iris synthesis. *IEEE CG&A 23*, 6, 70–75.

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 *Proc. Rendering Techniques*, 183–194.

PAMPLONA, V. F., OLIVEIRA, M. M., AND BARANOSKI, G. V. 2009. Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Trans. Graphics (TOG) 28*, 4, 106.

PIÑERO, D. P. 2013. Technologies for anatomical and geometric characterization of the corneal structure and anterior segment: a review. In *Seminars in Ophthalmology*, no. 0, 1–10.

RIO-CRISTOBAL, A., AND MARTIN, R. 2014. Corneal assessment technologies: Current status. *Survey of Ophthalmology*.

RUHLAND, K., ANDRIST, S., BADLER, J., PETERS, C., BADLER, N., GLEICHER, M., MUTLU, B., AND MCDONNELL, R. 2014. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Eurographics State of the Art Reports*, 69–91.

SAGAR, M. A., BULLIVANT, D., MALLINSON, G. D., AND HUNTER, P. J. 1994. A virtual environment and model of the eye for surgical simulation. In *Proceedings of Computer Graphics and Interactive Techniques*, 205–212.

SEITZ, S. M., CURLESS, B., DIEBEL, J., SCHARSTEIN, D., AND SZELISKI, R. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *IEEE CVPR*, vol. 1, 519–528.

SMOLEK, M. K., AND KLYCE, S. D. 2003. Zernike polynomial fitting fails to represent all visually significant corneal aberrations. *Investigative ophthalmology & visual science 44*, 11, 4676–4681.

SORKINE, O., COHEN-OR, D., LIPMAN, Y., ALEXA, M., RÖSSL, C., AND SEIDEL, H.-P. 2004. Laplacian surface editing. In *Proc. SGP*, 175–184.

VIVINO, M., CHINTALAGIRI, S., TRUS, B., AND DATILES, M. 1993. Development of a scheimpflug slit lamp camera system for quantitative densitometric analysis. *Eye 7*, 6, 791–798.