DOI: 10.1111/cgf.14642
ACM SIGGRAPH / Eurographics Symposium on Computer Animation 2022
D. L. Michels and S. Pirk
(Guest Editors)
Volume 41 (2022), Number 8

Detailed Eye Region Capture and Animation

Glenn Kerbiriou[13], Quentin Avril[3], Fabien Danieau[3] and Maud Marchal[12]
Univ. Rennes, INSA, IRISA, CNRS, Inria[1] IUF[2] InterDigital[3]

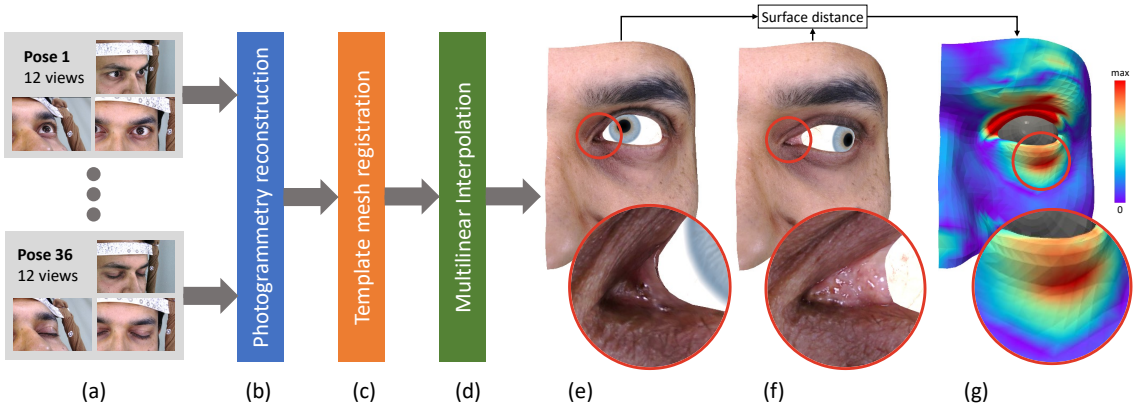

Figure 1: *Summary of our approach for capturing and animating detailed eye regions: (a) eye capture (Section 2), (b) mesh reconstruction, (c) mesh registration (Section 3) and (d) multi-linear interpolation (Section 4). Through this pipeline, we are able to produce detailed eye region animation by accurately reproducing specific details such as (e, f) eye inner corner tissues and g) bulging effect caused by eyeball orientation.*

Abstract

Even if the appearance and geometry of the human eye have been extensively studied during the last decade, the geometrical correlation between gaze direction, eyelids aperture and eyelids shape has not been empirically modeled. In this paper, we propose a data-driven approach for capturing and modeling the subtle features of the human eye region, such as the inner eye corner and the skin bulging effect due to globe orientation. Our approach consists of an original experimental setup to capture the eye region geometry variations combined with a 3D reconstruction method. Regarding the eye region capture, we scanned 55 participants doing 36 eyes poses. To animate a participant's eye region, we register the different poses to a vertex wise correspondence before blending them in a trilinear fashion. We show that our 3D animation results are visually pleasant and realistic while bringing novel eye features compared to state of the art models.

1. Introduction

Faithfully digitizing humans for applications such as video-games, Virtual Reality (VR) or visual effects has been a long-standing challenge in computer graphics. In particular, human face reconstruction has received a great amount of attention from the research community [DGO*19].

Among the human face features, eyes stand out by being the most salient [RWX17] and descriptive [KPKR17] region. In addition, eyes are extremely important to convey emotions, mental state or non-verbal communication [BCWH*01]. The need of faithful reconstruction of the eye region is therefore a key challenge of human face animation. In this work, we focus on the detailed capture, reconstruction and animation of the eye region using a data-driven approach. We refer to eye region as the soft tissues surrounding the eyeball, but not the eyeball itself. Our goal is to capture how these tissues behave when the gaze direction and the eye aperture vary. Realistic eyeball reconstruction and modeling has been addressed by Bérard and his colleagues in many of their works [BBN*14, BBGB16, BBGB19]. Bermano *et al.* [BBK*15] reconstructed the upper eyelid crease with great precision, but only for eyelid aperture variations, and the gaze-eyelid interaction was not modeled in depth. This correlation between gaze direction and eyelid motion has been learned by Neog *et al.* [NCRP16],

but their elastodynamic model mostly simulates lateral gaze direction changes and omits some fine details. In the same vein, Wood and his colleagues built a 3DMM of the eye [WBM*16a] for gaze direction estimation [WBZ*15] that they rendered and animated using generic *blendshapes i.e.* linear combination of per-vertex deformations. Then, they introduced a rotation-based animation and simulated the eyeball-eyelid contact using *shrinkwarping* [WBM*16b]. However, only upward-looking and downward-looking eyelid shapes were extracted from data. Real-time eyelid shape reconstruction has been achieved by Wen *et al.* [WXLY17] from a RGBD sensor. They detect semantic edges using a Deep Neural Network and fit a blendshapes model of eyelid in each frame. The trade-off of between real-time performance and reconstruction quality limits their system to the use of a linear model.

As far as we know, no previous work addressed the animation of the eye region from high quality scans that correlates gaze direction and eye aperture. Moreover, existing methods are based on low to medium quality data [NCRP16, WXLY17] or on sparsely sampled poses [WBM*16b]. As a result, a number of dynamic details of the eye region are missed.

In this paper, we propose to address these limitations by introducing a data-driven approach for capturing and animating detailed eye regions. Our contributions are the following :

- A detailed protocol for capturing the eye region and the correlation between eyelids shape, gaze direction and eye aperture.
- An anatomically coherent multilinear animation model of the eye region.
- A preliminary analysis of eye inner corner tissues movements in a population of more than 50 individuals.

2. Detailed Eye Region Capture

The first step of our approach is to collect *in vivo* data in order to study how the possible movements of the eye affect the surrounding soft tissues. For that purpose, we designed an experimental setup in order to capture highly-detailed eye regions of 55 participants.

Experimental setup. Our setup consists of a rig (see Fig. 2) containing 12 Canon EOS 1300D cameras arranged concentrically around the participant's eye. The participant's head is fixed with a chin strap device similar to that of an ophthalmologist. Diffuse lighting is achieved with 14 round spot lights covered with diffuse paper mounted uniformly on the rig. Additionally, 8 physical targets are placed in a 3x3 grid pattern visible from participants' point of view. The central target of the grid corresponds to the center of the front camera lens.
To calibrate the colors of the sensors, a color mire device featuring April Tags [Ols11] is shot before each capture session. Before the capture, the participant's skin is cleaned with micellar water then matified using matifying wipes. If a participant had thick or bushy eyebrows, we used styling wax to flatten and comb them.

Experimental protocol. In our protocol, we chose to discretize the eye pose space in 36 poses (see Figure 3) that participants performed by looking at 9 physical targets. These targets allow to capture 9 different gaze angles together with 4 different eye apertures. At the end, our dataset contains eye close-up 3D scans of 55 participants performing 36 eye poses, totalling 1980 textured scans. The

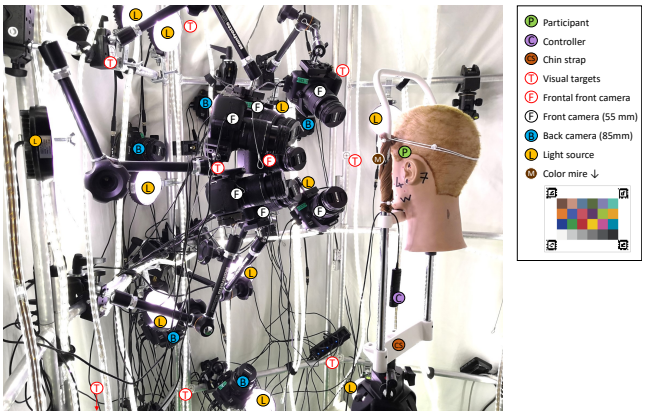

Figure 2: *Picture of our camera rig setup.*

captured population contains 40 males and 15 females (age ranging from 21 to 67 years old). No particular selection was made on the participants.

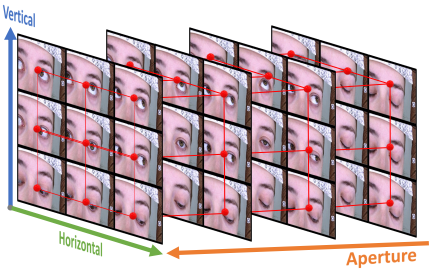

Figure 3: *The eye pose space is discretized in 36 poses, as the cartesian product of 3 levels for horizontal gaze angle ("right", "center" and "left"), 3 levels for vertical gaze angle ("down", "center" and "up") and 4 levels for eye aperture ("wide open", "open", "half-closed" and "closed"). The red spheres represent the data nodes used for the interpolation process of the animation (Section 4).*

3. 3D Reconstruction

The 3D reconstruction step consists in generating a 3D mesh from the data acquired with our experimental setup. For that purpose, we successively apply a photogrammetry method from the photographs, followed by a 3D registration to gather all the scans. Finally, we fit an ~~a~~ eyeball mesh to complete the generated 3D mesh.

Photogrammetry process. The surfaces are extracted from the photographs of the twelve cameras using Agisoft Metashape photogrammetry software †. The "multi-frame chunk" functionality is used to calibrate and optimize jointly the cameras for the thirty-six eye poses. The depth maps are downscaled eight times and the diffuse texture map is computed in 4096 by 4096 pixels.

3D Registration. To achieve dense correspondence across all

† Agisoft Metashape, http://www.agisoft.com

the scans, we employ non rigid registration [SP04] with a template mesh. The template mesh used is a subset of the ICT-FaceKit topology [LBZ*20] focusing on the left eye containing 1407 vertices and 2664 triangles. We manually place a set of 32 3D landmarks on each scan following the pattern depicted in Figure 4. The landmarks are placed on anatomical key points, and are decomposed in 7 eye region features.

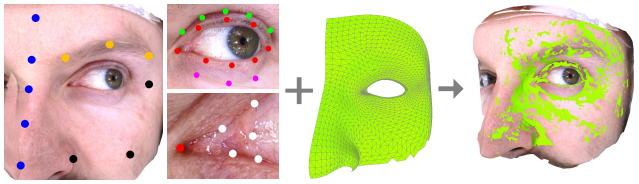

Figure 4: *Left, our semantic landmark pattern. 5 landmarks for the vertical face symmetry axis (blue), 3 for the eyebrow (orange), 3 for the bottom exterior region (black), 8 for the eyelid contour (red), 5 for the eyelid crease (green), 3 for the bottom eyelid bulge (pink), 5 for the inner tissues (white). Middle, the template mesh used. Right, superimposition of the registered template and the scan.*

To ensure a denser correspondence mapping between the scans of a same subject, we first register the neutral pose (eyes open, look forward), then we find dense correspondence in the texture maps in order to add more landmarks and reduce skin sliding. Once the geometry registration is done, we generate diffuse color map, height map and normal map by comparing the fitted template mesh to the scan.

Eyeball fitting. Finally, we fit a generic eyeball mesh on each scan for visual purposes and to fill the gaps between the eyelid borders and the eyeball (see Section 4). We extract the surface of the visible eyeball by thresholding the distance to the registered template. Then, we fit a sphere at the least squares sense on this surface. The pupil location is determined as the center of the pit of the cornea. Finally, we perform Iterative Closest Point (ICP) to align our generic eyeball mesh to the scan part and we fix the eyeball radius such that its scale is constant along the poses.

4. Animation

Animation System. To show the shape and appearance representation capacity of our capture and registration method, we computed an animation based on our 3D data (Figure 5). We linearly interpolate the meshes vertices, textures and eyeball transforms between gaze angles and eye aperture levels. This leads to a trilinear interpolation inside the 2 voxels wide, 2 voxels long and 3 voxels deep volume represented as a red wireframe in Figure 3. The desired pose is obtained from an aperture value $a \in [0,3]$, and horizontal and vertical angle levels $x, y \in [0,2]$. The eyelid borders are then *shrinkwarped* to the eyeball with increasing distance as done by Wood *et al.* [WBM*16b].

Preliminary Data Analysis. Our animation features unprecedented level of details compared to state-of-the-art methods. First, as the pose space is densely sampled (not only along the semantic axes), we strongly reduce the risk of creating anatomically invalid poses. The equivalent 1-D linear interpolation model of our animation would be performed with three blendshapes for aperture ("open" to "wide open" and to "half-closed" and "half-closed" to "closed") and four blendshapes for gaze direction ("forward" to "up", "down", "right" and "left"). Unless creating a specific rig with strong constraints, this model can easily generate undesired artifacts or even anatomically invalid poses by combining blendshapes acting on the same vertices. Figure 7 provides (a) an example of such an invalid pose, fixed with (b) our multilinear interpolation.

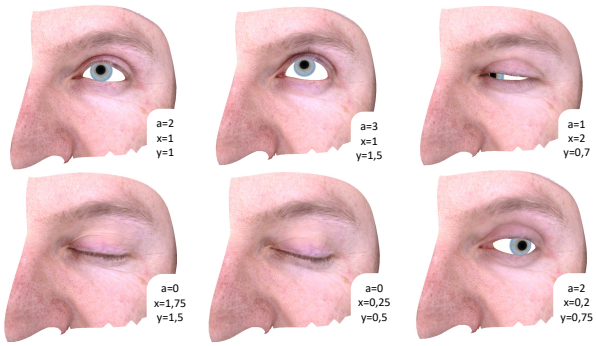

Figure 5: *Illustration of multilinear pose interpolations for a participant.*

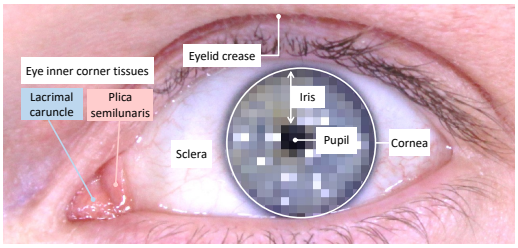

Figure 6: *Anatomy features of the eye that are relevant for this work. We can notice the clear distinction between the lacrimal caruncle and the plica semilunaris.*

Second, our animation accounts for bulges appearing relatively far from the visible sclera (Figure 6) with upward gaze directions (Figure 8, (a, b)). Eyelid deformation with closed eye and a varying gaze direction is also well modeled (c, d). Third, the eye inner corner tissues size is strongly correlated with an outward-looking eye. In Figure 9, we show that our method (c) is able to correctly animate this area compared to (a, b) previous works. We observe that the *plica semilunaris* (Figure 6) is mainly responsible for the expansion. We measured the area of the inner tissues compared to the area of the globe plus the inner tissues for all the participants. It appears that the inner tissues occupy in average $13.4 \pm 3.8\%$ of this reference area with an outward gaze, against $3.5 \pm 1.5\%$ with an inward gaze direction. Moreover, the expansion is measured as an increase of $9.8 \pm 3.4\%$ in average. This tells us that the tissues expansion is not only significant, but also person specific (great variance), which enforces the need of personalized eye region capture, modeling and animation for first leading characters.

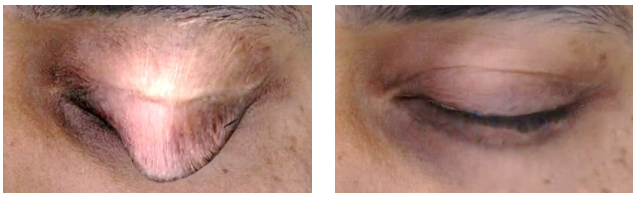

Figure 7: *Interpolated pose with both downward gaze and closed eye : (a) blendshapes-driven equivalent of our method, (b) our trilinear interpolation.*

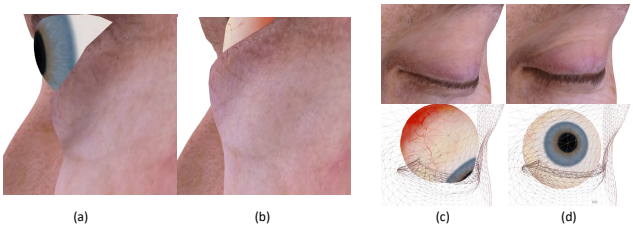

Figure 8: *(a, b) Skin bulge appearing with an upward gaze. (c, d) Eyelid deformation with closed eye and a varying gaze direction.*

5. Conclusion

In this paper, we proposed an original data-driven approach for capturing and modeling the subtle features of the human eye region. We first designed a dedicated photogrammetry rig for capturing with precision a combination of multiple eye poses of more than 50 participants. We then proposed a 3D reconstruction methodology to generate 3D meshes of the eye region. The dense sampling of the eye pose space allows to correlate gaze direction and eye aperture, resulting in accurate animation with only a linear interpolation scheme. Unlike previous techniques, our animation also features person-specific details such as the expansion of the *plica semilunaris* and skin bulges due to eyeball orientation. Future work will tackle statistical modeling of the eye region using a linear method or a neural network. We trust that our pose-oriented data will lead to good results in terms of reconstruction, pose transfer and fitting to images [WBM*16a]. We plan to investigate other interpolation techniques including physical constraints. A strong limitation of our approach is the manual landmark placement that will also be addressed later. Thanks to the level of detail captured through our setup, our approach paves the way for personalized human eye region modeling and animation.

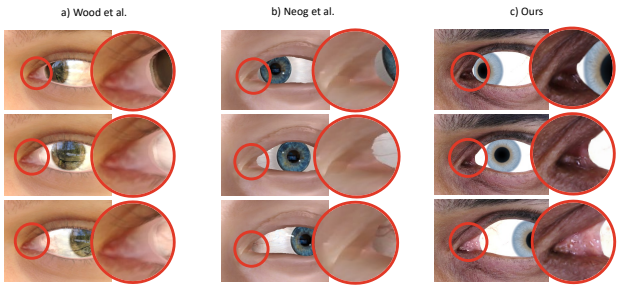

Figure 9: *An illustration of our eye inner corner tissues expansion animation compared to two other eye and eye region simulation methods. From left to right : [WBM*16b], [NCRP16], Ours.*

6. Acknowledgments

This project is supported by CIFRE grant No 2020/1213 and received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 952147.

References

[BBGB16] BÉRARD P., BRADLEY D., GROSS M., BEELER T.: Lightweight eye capture using a parametric model. *ACM Trans. Graph. 35*, 4 (2016). 1

[BBGB19] BÉRARD P., BRADLEY D., GROSS M., BEELER T.: Practical person-specific eye rigging. *Comp. Graph. Forum 38* (2019). 1

[BBK*15] BERMANO A., BEELER T., KOZLOV Y., BRADLEY D., BICKEL B., GROSS M.: Detailed spatio-temporal reconstruction of eyelids. *ACM Trans. Graph. 34*, 4 (2015). 1

[BBN*14] BÉRARD P., BRADLEY D., NITTI M., BEELER T., GROSS M.: High-quality capture of eyes. *ACM Trans. Graph. 33*, 6 (2014). 1

[BCWH*01] BARON-COHEN S., WHEELWRIGHT S. J., HILL J., RASTE Y., PLUMB I.: The "reading the mind in the eyes" test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry, and allied disciplines 42 2* (2001), 241–51. 1

[DGO*19] DANIEAU F., GUBINS I., OLIVIER N., DUMAS O., DENIS B., LOPEZ T., MOLLET N., FRAGER B., AVRIL Q.: Automatic generation and stylization of 3d facial rigs. *IEEE-VR* (2019), 784–792. 1

[KPKR17] KARCZMAREK P., PEDRYCZ W., KIERSZTYN A., RUTKA P.: A study in facial features saliency in face recognition: an analytic hierarchy process approach. *Soft Computing 21* (12 2017). 1

[LBZ*20] LI R., BLADIN K., ZHAO Y., CHINARA C., INGRAHAM O., XIANG P., REN X., PRASAD P., KISHORE B., XING J., LI H.: Learning formation of physically-based face attributes. In *IEEE/CVF CVPR* (2020), pp. 3407–3416. 3

[NCRP16] NEOG D. R., CARDOSO J. A. L., RANJAN A., PAI D. K.: Interactive gaze driven animation of the eye region. In *Proc. of the Int. Conf. on Web3D Technology* (2016), p. 51–59. 1, 2, 4

[Ols11] OLSON E.: AprilTag: A robust and flexible visual fiducial system. In *Proc.of the IEEE International Conference on Robotics and Automation* (2011), pp. 3400–3407. 2

[RWX17] REN Y., WANG Z., XU M.: Learning-based saliency detection of face images. *IEEE Access 5* (2017), 6502–6514. 1

[SP04] SUMNER R. W., POPOVIĆ J.: Deformation transfer for triangle meshes. *ACM Trans. Graph. 23*, 3 (2004), 399–405. 3

[WBM*16a] WOOD E., BALTRUŠAITIS T., MORENCY L.-P., ROBINSON P., BULLING A.: A 3d morphable eye region model for gaze estimation. In *Proc. of ECCV* (2016), pp. 297–313. 2, 4

[WBM*16b] WOOD E., BALTRUŠAITIS T., MORENCY L.-P., ROBINSON P., BULLING A.: Learning an appearance-based gaze estimator from one million synthesised images. In *Proc. of ACM Symp. on Eye Tracking Research and Applications* (2016), p. 131–138. 2, 3, 4

[WBZ*15] WOOD E., BALTRUSAITIS T., ZHANG X., SUGANO Y., ROBINSON P., BULLING A.: Rendering of eyes for eye-shape registration and gaze estimation. In *Proc. of IEEE Int. Conf. on Computer Vision* (12 2015), pp. 3756–3764. 2

[WXLY17] WEN Q., XU F., LU M., YONG J.-H.: Real-time 3d eyelids tracking from semantic edges. *ACM Trans. Graph. 36*, 6 (2017). 2