



FaceLab: Scalable Facial Performance Capture for Visual Effects

Curtis Andrus
MPC-Film, Technicolor

Junghyun Ahn
Technicolor, InterDigital

Michele Alessi
Mr. X

Abdallah Dib
InterDigital

Philippe Gosselin
InterDigital

Cédric Thébault
InterDigital

Louis Chevallier
InterDigital

Marco Romeo
MPC-Film, Technicolor

ABSTRACT

In visual effects it is usually required to process multiple film plates in order to reproduce the performance of real actors using a digital 3D character. One of the most challenging aspects of this process is the complexity of facial animation.

Motivated by the work required to capture and digitally augment the performances for the 2019 film *Cats* [Cats 2019], we created FaceLab, a system that drives the animation of a 3D facial rig by optimizing its blendshape targets' weights and the head transformation matrix to match the reference plate. The system is designed to scale across different scenarios with varying complexity and produced tangible savings on the show.

CCS CONCEPTS

• **Applied computing** → **Media arts**; • **Computing methodologies** → **Motion capture**; **Machine learning**.

KEYWORDS

facial performance capture, optimization, visual effects

ACM Reference Format:

Curtis Andrus, Junghyun Ahn, Michele Alessi, Abdallah Dib, Philippe Gosselin, Cédric Thébault, Louis Chevallier, and Marco Romeo. 2020. FaceLab: Scalable Facial Performance Capture for Visual Effects. In *The Digital Production Symposium (DigiPro '20)*, July 18, 2020, Virtual Event, USA. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3403736.3403938>

1 INTRODUCTION

In visual effects it is common practice to recreate the performance of an actor on a digital character. Digital characters might be required to replace the real actor to perform dangerous stunts, to have a fantastic creature performing like the real actor or to augment the likeness of the original actor. The most challenging part of this work is to transfer the performance of the face.

At times it is possible to use facial capture rigs as a head mounted camera [FaceWare 2020], which points directly at the face of the actor, who is typically wearing a special make-up to help tracking the face, but this solution makes the actors uncomfortable, altering

their performance. Moreover the use of facial capture rigs makes the original performance of the actor on film unusable, requiring a complete digital replacement and the digital removal of the head-mounted rig from the plate. Other solutions require actors to perform inside a special booth that captures a high quality animated geometry cache of their performance [Anyma 2020], but this solution cannot be used to capture the live on-set performance of the actors. For many other scenarios where the aforementioned options are not viable, manual rotoscope animation of faces is the only solution, requiring a lot of resources to achieve high quality results.

In the next section we present FaceLab, a system we created to address this issue. In Section 3 and Section 4, we respectively describe the workflow we adopted in production and the results we achieved using the system during the production of *Cats* [Cats 2019]. Finally, in Section 5, we draw conclusions and discuss future work.

2 FACELAB

FaceLab is a system based on an optimization-based face reconstruction methodology similar to [Garrido et al. 2016]. The underlying modelling covers the face's shape (topology and expression), pose and reflectance, together with the light environment. The facial expression is represented using standard linear delta blendshapes interpolation. The blendshape targets' weights are optimized via an inverse rendering policy which gradually converges towards a face model compatible with the content of the scene. This approach has now become popular especially when extended with fast and robust direct deep learning based architectures [Tewari et al. 2017]. However, in production our main focus is on precision and high fidelity, which we found to be better achieved using iterative optimization. To improve robustness of the tracking, a robust rendering stage with fast auto-occlusion was integrated in the system.

In order to ease the process of any required manual post-processing, special attention was paid to the number of simultaneous active blendshapes. This resulted in a sparsity constraint in the formulation of the loss to be minimized. The system was built to take advantage of any available additional content like rotoscope animation or on-set data like lightprobes and witness cameras.

Following, we present a list of the most effective improvements we made to the original work of [Garrido et al. 2016], which made the system able to cope with real life shots:

- **Optimization policy** - In order to optimize the many parameters (ca. 100 per frames, plus 100 shared for each shot), we

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

DigiPro '20, July 18, 2020, Virtual Event, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8034-8/20/07.

<https://doi.org/10.1145/3403736.3403938>

adopted a gradient descent approach, along with a carefully designed policy. We first initialize the head pose and the blendshape targets' weights using solely landmarks. This fast optimization greatly speeds up the subsequent optimization steps. Then the other parameters - the light environment (a), the head pose (b), blendshape targets' weights (c) and the albedo (d) - are optimized for a selection of frames, starting with (a), then (a, b), (a, b, c) and finally (a, b, c, d). This policy ensures the robustness and the stability of the optimization. The parameters (except the albedo) are then fine tuned jointly for consecutive frames, while enforcing temporal smoothness.

- **Keypoints** - Image keypoints are a viable alternative to dense optical flow to align facial features (and makeup) from one frame to another. An additional - computationally efficient - loss term is derived from the keypoints' locations. This term improves the quality of the facial performance capture, as well as the temporal smoothness.
- **Personalized rigs** - Similarly to [Garrido et al. 2016], our system works on unknown actors by inferring their neutral face geometry (i.e. *identity*). In addition, our system can also handle personalized facial rigs, which consist of a scanned identity and actors' blendshape targets based on the Facial Action Coding System (FACS).
- **Visibility check** - Accurate and fast vertex visibility from the camera was added to the system using ray casting. This provided enhanced robustness for profile pose as compared to standard back-face culling test.
- **Landmarks** - Like [Garrido et al. 2016], we also resort to facial landmark prediction ([Bulat and Tzimiropoulos 2018]) to obtain an initial estimation of the pose of the head and the blendshape targets' weights. Predicted landmarks tend to be noisy, and for some of them the correspondences between the landmarks and the vertices vary with the pose. To address this, we used 3D coordinates prediction and added a model to learn this relationship for each personalized rig. This comes as a replacement of the standard ICP (Iterated Closest Point) algorithm for better stability.
- **Interactivity** - Under difficult conditions like hard lateral lighting with cast shadows, high specular reflections and/or blurred faces, the result of the performance reconstruction requires corrections. An interactive mechanism was implemented which enabled artists to guide the optimization with direct image annotations. This is done by adding vertex-pixel associations which are translated into additional constraints to be satisfied by the optimization engine.
- **Pytorch** - We used the Pytorch framework [Paszke et al. 2017] to implement the full pipeline. This framework made the formulation of the modelling, the rendering and the many loss terms easy, thanks to its automatic differentiation capability, it proved to be flexible enough to implement our complex optimization policy. Finally, it supports GPU and CPU backends transparently.

3 CAPTURE WORKFLOW

We integrated FaceLab in our production workflow by defining three separated processes:

- (1) **Preparation** - Typically we receive reference data for the actor (scans and pictures), which enables us to create a three dimensional representation of the actor's face. Once the geometry for the digital actor's face is ready, we create a facial rig using FACS and compute the correspondence between the digital actor's scanned identity and the FaceLab generic face meshes.
- (2) **Setup** - Using FaceLab we identify the actors on the shot plate and relate the identity of the actor to the 3D information previously prepared.
- (3) **Solve** - We distribute the computation of the performance of different actors on different shots across the resources available on our computation farm. The FaceLab command-line interface (*CLI*) allowed to batch process multiple plates during the night. The results from FaceLab are then assetized as animation curves that we can reassign in Maya to the blendshape targets' weights.

This integration of FaceLab within our production pipeline allowed our artists and supervisors to take advantage of the system by using existing production rigs and compute an animated facial rig immediately usable in Maya. Finally, any existing matchmove or rotoscoping preliminary results could also be fed to the system to increase the quality of the output and reduce shots' turnaround.

4 RESULTS

We used FaceLab to create the visual effects for the 2019 film *Cats* [Cats 2019]. The show required us to augment the look of the real actors by adding digital fur and other cat-like features (e.g. ears) to their performances. Because of the uniqueness of the planned shots, actors had to be filmed without wearing any special head-mounted camera rig and without using any witness camera on set. The actors were wearing special facial makeup to improve the tracking of their facial features, which was helpful for close-ups, however this makeup was not perceivable for faces of actors in the distance.



Figure 1: From left to right: the original plate, the result achieved with FaceLab and the final render using the performance captured. ©2019 Universal Pictures. All rights reserved.

Thanks to FaceLab, we could automatically compute the facial animation on the farm in 4 hours for each actor, considering an

average shot length of six seconds. Supervisors found that the results qualitatively compared well to the results typically achieved over three days of work, independently from the distance of the actors. With an average of three actors per processed shot and a 15% of the total 1765 shots using FaceLab, a substantial amount of time was saved, accelerating the global workflow, allowing timely delivery of complex shots and allowing artists to focus on perfecting other creative tasks.

Overall the system performed very well, producing results comparable to those of artists. FaceLab was used on different shots that ranged from close-ups to mid-long shots and was also used to compute the performance of faces appearing de-focused on the plate, demonstrating an all-around fit for purpose. The system showed some limitations under challenging light conditions as hard shadows and strong highlights.

5 CONCLUSIONS AND FUTURE WORK

The system proved to rapidly produce high quality performance capture under challenging scenarios, requiring manual intervention only to setup the solver for the required shot.

To improve FaceLab further and efficiently process shots with difficult lighting conditions, we plan to integrate differentiable ray tracing and Cook-Torrance BRDF approaches [Cook and Torrance 1982; Dib et al. 2019] to better model specularities and self-shadows in FaceLab.

We continuously work on improving the user experience with the tool and we are researching deep learning techniques aimed to increase the performance of the optimization process.

ACKNOWLEDGMENTS

We thank our creative teams for their feedback and the Max Planck Institute for sharing their expertise with us.

REFERENCES

- Anyma 2020. *Disney's Anyma*. Retrieved June 8, 2020 from <https://studios.disneyresearch.com/anyma/>
- Adrian Bulat and Georgios Tzimiropoulos. 2018. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 109–117.
- Cats 2019. *Universal Pictures, Cats*. Retrieved June 10, 2020 from <https://www.universalpictures.com/movies/cats>
- Robert L Cook and Kenneth E. Torrance. 1982. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)* 1, 1 (1982), 7–24.
- Abdallah Dib, Gaurav Bharaj, Junghyun Ahn, Cedric Thebault, Philippe-Henri Gosselin, and Louis Chevallier. 2019. Face Reflectance and Geometry Modeling via Differentiable Ray Tracing. *ACM SIGGRAPH European Conference on Visual Media Production (CVMP)* (2019).
- FaceWare 2020. *FaceWare Technologies*. Retrieved June 8, 2020 from <https://facewaretech.com>
- Pablo Garrido, Michael Zollhöfer, Dan Casas, Levi Valgaerts, Kiran Varanasi, Patrick Pérez, and Christian Theobalt. 2016. Reconstruction of personalized 3D face rigs from monocular video. *ACM Transactions on Graphics (TOG)* 35, 3 (2016), 1–15.
- Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. (2017).
- Ayush Tewari, Michael Zollhofer, Hyeonwoo Kim, Pablo Garrido, Florian Bernard, Patrick Pérez, and Christian Theobalt. 2017. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 1274–1283.