ABBA Voyage: High Volume Facial Likeness and Performance Pipeline

Jo Plaete
Industrial Light & Magic
UK
jo.plaete@gmail.com

Derek Bradley
DisneyResearch|Studios
CH
derek.bradley@disneyresearch.com

Paige Warner
Industrial Light & Magic
USA
paigewarner@ilm.com

Anthony Zwartouw
Industrial Light & Magic
CA
azwartouw@ilm.com

Figure 1: ABBA Today

ABSTRACT

For the ABBA: Voyage concert experience, Industrial Light & Magic (ILM) was tasked with digitally time traveling the iconic band's members Agnetha, Anni-Frida, Björn and Benny back to their prime time appearances. For the duration of this fully computer graphics generated concert four continuous photo-real digital human facial performances had to be synthesised driven by their original current day counterparts and stand-in young actors. This talk will dive into the extensive research and development that was undertaken to cater for high volume facial capture and processing, true-to-likeness face-retargeting and additional techniques for breaking through the uncanny valley.

ACM Reference Format:
Jo Plaete, Derek Bradley, Paige Warner, and Anthony Zwartouw. 2022. ABBA Voyage: High Volume Facial Likeness and Performance Pipeline. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '22 Talks), August 07-11, 2022.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532836.3536260

1 FACIAL RECONSTRUCTION

DisneyResearch|Studios' ANYMA [Wu et al. 2016] has been utilised by ILM for several projects to faithfully reconstruct facial performances. It starts with MEDUSA [Beeler et al. 2011] scanning the actors' neutral pose and FACS expressions which get compiled into an Anatomical Local Model. This enables ANYMA [Wu et al. 2016] to reconstruct 3D geometry from input facial performances captured by either a multi camera booth setup or head-mounted cameras (HMC). These feeds provided song performances for both current day ABBA band members and young stand-in actors respectively. Once reconstructed these per-frame 3D meshes get ingested into the ILM pipeline for cleanup and editing by the facial capture team. We had to streamline and link various pipeline components in order to facilitate automated reconstruction of over 90 minutes of long continuous performances for 8 input actors. A flexible blueprint configurable system was put in place that helped adjusting to the different input sources and actors and was extensible to include further facial pipeline steps later down the line. The facial dot-tracking step, required as an input to HMC ANYMA [Wu et al. 2016] solving and historically a manual tracking task, was replaced by an automated machine learning (ML) based solution. This meant that entire songs could be auto-ingested, processed and presented without manual intervention, a pipeline process we branded 'first-faces'. Subsequent facecap editing, processing and QC steps could also lean on this infrastructure to ensure performance continuity.

2 FACIAL RETARGETING

Once ingested, these faithful reconstructions of the source actors had to drive the facial performance of their young ABBATAR (digital ABBA Avatar) equivalents in a manner that would bring across the intent, emotion and energy but also, very importantly, retain their iconic likenesses. This process, known as retargeting, has been used for over a decade, and typically relies on either large blendshape rigs that are expensive to create, or direct deformation transfer algorithms that operate on individual geometric elements and are prone to artifacts and distortion of the target likeness by the input actor's differences in facial anatomy. After a stage of exploring various methods, with varying degrees of success, our colleagues at DisneyResearch|Studios in Zurich presented us with a new method for high-fidelity offline facial performance retargeting.

The two step method first transfers local expression details to the target, and is followed by a global face surface prediction that uses anatomical constraints in order to stay in the feasible shape space of the target character. As such, it turned out to be well suited for the complex task of human-to-human 3D facial performance retargeting retaining identity and likeness of the target, by sources that have different facial anatomies, either heavily aged current day band members or young stand-in actors respectively.

Alongside the neutral model poses, a set of 25 shape pairs covering a wide range of expressions fed into the retarget configuration as a base. This allowed our digital model shop team to carefully

SIGGRAPH '22 Talks, August 07-11, 2022, Vancouver, BC, Canada

ACM ISBN 978-1-4503-9371-3/22/08.
https://doi.org/10.1145/3532836.3536260

Figure 2: ABBA Today

craft a believable and on-model expression on the target ABBATAR for each of the input actor's equivalents. To expand on this concept within a shot context, a pipeline was installed that allowed adding additional extra shapes to the retarget config.

3 PERFORMANCE REFINEMENT

For artists to get an editable handle on the retarget of the reconstructed performance, which at that point is still expressed as per-frame full face geometries and jaw channels, we rely on a process known as inversion. This analyses its input and converts it to a set of channels that drive the full blendshape rig, plus an additional residual delta layer that accounts for a small discrepancy between the input and the rig's result, typically describing higher frequency skin motion from the capture.

ILM's inversion toolset was revamped for this project with a focus on bringing it into the hands of the performance refinement team which was achieved by integrating it within Maya and by the addition of an easy-to-use yet flexible recipe system to tweak the output of the inversion and iterate quickly. Furthermore batch functionality was put in place to allow running inversion recipes on entire songs at once for efficiency but also continuity consistency of the performance.

4 FURTHER FACE REFINEMENT

To fine-tune and enhance the realism of the performances, further post facial reconstruction and retarget steps were introduced.

A first technique to foster facial connectivity and add dynamic motion to the face's skin was to run a tetrahedral simulation on it. The simulation would take in the input animation and underlying skull as drivers and collision objects and computed curvature to identify where muscle activation and skin folds were key to the expression read versus areas where the simulation could have a higher degree of freedom. It would also introduce volume preservation and perform de-intersection which worked particularly well on e.g. pressing lips.

It was also important to introduce a dynamic effect on the higher frequency details of the face such as fine wrinkles, which are expressed within the displacement layer of the model. Here a fairly straightforward but effective render-time dynamic displacement system was put in place that would source data generated by the blendshape rig and use it to reveal specifically crafted displacement details supportive of the expressions activated. Additionally, a subtle blood flow system was added modulating the face's colour textures due to strong compression or stretch.

Last but not least, sculpt artists had access to several of these extra delta layers so they could blend and tweak their effects by face regions and/or add manual corrective sculpting on top within a shot's context.

5 ML FACESWAP

Initial FaceSwap [Naruniec et al. 2020] models were trained on low resolution archival footage consisting of predominantly ABBA music videos and band member interviews. Once trained against CG playblasts or renders a tool named 'swapit' allowed animators and sculpt artists to run out a fast FaceSwap to get an example of photographic nature as to how the ABBATAR would have looked in their prime and how they would have articulated a certain performance. Despite outputs generated by the network being low resolution and soft, this was still hugely informative for both the facial expression sculpting process and animation development.

We were fortunate to be given access to the original scans of several ABBA music videos so they could be rescanned and remastered at higher resolution. Dedicated FaceSwap CG input renders were generated for song/shot specific refining of the FaceSwap models. The compositing team devised ways to fully or partially integrate this photographic likeness layer with its underlying CG whilst still sourcing detail and lighting information from the latter.

Despite the better models, resolution and sharpness remained a challenge for close-up FaceSwap to be successful. Furthermore the archival dataset was not all encompassing which meant certain face angles or lighting conditions would work better than others. This would also affect the performance generated by the network which would sometimes deviate from the input performance. This could result in broader performance problems such as lacking the input's energy or intent or very specific issues such as non-matching eye direction or lip-sync. To combat those issues, controllability advancements were made by ILM's R&D engineers that could nudge the network's output. This meant that FaceSwap artists could make subtle modifications to the networks output to address specific feedback.

6 CONCLUSION

The most prominent innovations were the introduction of a new human-to-human retarget method followed by the end-of-chain likeness enhancing FaceSwap, alongside of a high degree of pipeline automation.

REFERENCES

Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. 2011. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics* 30, Article 75 (2011), 10 pages. Issue 4.

J. Naruniec, L. Helminger, C. Schroers, and R.M. Weber. 2020. High-Resolution Neural Face Swapping for Visual Effects. *Computer Graphics Forum* 39, 4 (2020), 173–184.

Chenglei Wu, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. An Anatomically-constrained Local Deformation Model for Monocular Face Capture. *ACM Transactions on Graphics* 35, 4, Article 115 (2016), 12 pages.