Emotion Challenge: Building a New Photoreal Facial Performance Pipeline for Games

Alex Smith, Activision Publishing, Inc.
Sven Pohle, Activision Publishing, Inc.
Wan-Chun Ma, Activision Publishing, Inc.
Chongyang Ma, Activision Publishing, Inc.
Xian-Chun Wu, Activision Publishing, Inc.
Yanbing Chen, Activision Publishing, Inc.
Etienne Danvoye, Activision Publishing, Inc.
Jorge Jimenez, Activision Publishing, Inc.
Sanjit Patel, Activision Publishing, Inc.
Mike Sanders, Activision Publishing, Inc.
Cyrus A. Wilson, Activision Publishing, Inc.

Figure 1: Left to right: Lightstage likeness acquisition; HMC performance capture; in-engine animation.

CCS CONCEPTS

• **Computing methodologies → Animation**; *Shape modeling*; *Motion capture*; *Reflectance modeling*;

KEYWORDS

facial scanning, performance driven animation, games

ACM Reference format:
Alex Smith, Sven Pohle, Wan-Chun Ma, Chongyang Ma, Xian-Chun Wu, Yanbing Chen, Etienne Danvoye, Jorge Jimenez, Sanjit Patel, Mike Sanders, and Cyrus A. Wilson. 2017. Emotion Challenge: Building a New Photoreal Facial Performance Pipeline for Games. In *Proceedings of DigiPro '17, Los Angeles, CA, USA, July 29, 2017,* 2 pages.
https://doi.org/10.1145/3105692.3105695

1 INTRODUCTION

In recent years the expected standard for facial animation and character performance in AAA video games has dramatically increased. The use of photogrammetric capture techniques for actor-likeness acquisition, coupled with video-based facial capture and solving methods, has improved quality across the industry. However, due to variability across project pipelines, increased per-project scope for performance capture, and a reliance on external vendors, it is often challenging to maintain visual consistency from project to project, and even from character to character within a single project. Given these factors, we identified the need for a unified, robust, and scalable pipeline for actor likeness acquisition, character art, performance capture, and character animation.

1.1 Novel Challenges

The desired quality bar has previously been the exclusive domain of proprietary performance-driven facial pipelines at large motion picture visual effects facilities. Achieving such a quality level in-game carries with it certain unique challenges.

Working within rig limitations. In most cases, facial animation rigs for runtime are limited to linear blend skinning (bones) and blendshapes. Crucially, the number of blendshapes supported is significantly less than in advanced facial animation rigs for motion pictures. The same limited set of blendshapes for a given character must support the entire range of performances that the character will deliver over the course of a game, and potentially over the course of subsequent expansions to the game.

Efficiently handling a large volume of performances. Performance-driven facial animation pipelines typically rely on clean-up and polish work by animators. For small animation teams, this presents a quality/quantity tradeoff. Our solution aims to produce a result which does not require clean-up or polish animation to faithfully reproduce the actor's original performance, while still supporting subsequent animation changes called for by creative direction.

Achieving consistency with pre-rendered cinematics. Several games incorporate pre-rendered cinematics produced using dedicated animation/VFX pipelines, as well as dedicated assets which are separate from the assets used in-engine. We wish for our solution to support both pre-rendered and in-engine facial animation, and asset re-use.

Supporting a wide range of art/animation pipelines and game engines. One of our biggest challenges is not necessarily intrinsic to games, but is significant within Activision Blizzard. For our solution

DigiPro '17, July 29, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5102-7/17/07.
https://doi.org/10.1145/3105692.3105695

to be a viable option across the company, it must support varying pipelines and engines developed by the different studios and game teams. Therefore our solution must be somewhat pipeline-agnostic.

1.2 The Emotion Challenge Project

With a number of related-yet-disparate research projects already under way, including the introduction of a Lightstage, we created an internal testbed project (known internally as the 'Emotion Challenge') to analyze and document each piece of technology and assess it for production viability. We would then devise and implement new techniques to replace or augment existing methods, and consolidate the entire character pipeline.

2 PIPELINE OVERVIEW

Likeness Acquisition. To faithfully capture an actor's likeness for recreation as a digital character in-game, we employ an expanded version of the ICT Lightstage system. The Activision lightstage leverages 16 DSLRs for high-resolution geometry and reflectance reconstruction using polarized complementary gradient illumination. We augmented this with 35 machine vision cameras capturing at 70 fps the movement of the actor's face between neutral and each extreme pose. This enables correspondence of the animation mesh topology, not only with each extreme pose, but also with all the intermediate poses. Game character art teams then regionalize mesh deltas to derive blendshapes for their specific face rigs. For more on Activision's extensions, as well as recent research into alternate methods of generating blendshapes, see [Ma et al. 2016].

Performance Capture. For facial performance capture, we acquire stereo 60 fps footage using a Head-Mounted Camera (HMC) system with a vertical stereo camera pair. An on-actor facial acquisition system allows facial performances and body mocap to be acquired simultaneously. We apply makeup dots to the actor's face, then track the dots using a semi-automated tool, producing a sparse input to our solver. For dense inputs we compute a dense stereo reconstruction between views each frame, as well as a dense optical flow from each frame to the next in each view. This is all fed into an optimization which fits the input data using the blendshape basis constructed for that actor based on the above Likeness Acquisition.

In-engine Animation. The per-frame blendshape weights can be exported directly to engine to drive facial blendshapes at runtime. Or, depending on the project pipeline, they can be used as inputs to a facial animation rig for additional animator polish prior to export.

3 VALIDATION

While the primary research focus of the 'Emotion Challenge' project was technological development, the end goal was a creative target: to capture and convey a genuine emotional performance using a pipeline viable for game development. In addition, we had a goal of demonstrating good results without animator polish. Removing the variable of animator skill—and effort spend—gave us a consistent quality baseline which can then be considered universally-applicable, independent of a given project's animation resources.

Prior experience had indicated that subtle emotions—joy, sadness, yearning, stoicism—were more difficult to capture and reproduce convincingly than overtly visual emotions such as anger or surprise, so we set ourselves the challenge of targeting those in our content.

The **supporting video** shows an example facial animation result without animator polish. (Note that eye animation was achieved using a third-party tool; and tongue animation was hand-keyed.)

4 TRANSITION INTO PRODUCTION

The Likeness Acquisition pipeline has proven scalable in production, thus far having been applied in the creation of approximately 150 in-game characters.

The 'Emotion Challenge' project demonstrated that the Performance Capture process was viable under ideal conditions. The non-ideal conditions encountered in production presented certain additional challenges for the Performance Capture process.

Input data quality. In production, the facial performance capture process is simultaneous with body mocap and on-actor audio recording; HMC footage quality can be impacted in several ways:

- Motion blur as the cameras move relative to the head during fast head movements.
- Lighting changes as actors tilt their heads up or down or move about the mocap volume.
- Specular highlights due to sweat during intense activity.
- General lighting and focus issues throughout shoot days with large casts, where the number of HMCs in use can make it difficult for a small crew to monitor each HMC carefully.

These data issues motivated the introduction of dense optical flow (from each frame to the next in each view) as a source of dense detail, substituting for, or complementing, dense stereo depending on the conditions. In practice, optical flow and/or dense stereo are leveraged when they help provide additional fidelity beyond marker tracks alone, but are excluded when they are detrimental due to quality issues in the input footage.

Solver shape basis. As discussed in Sec. 1.1, a central goal of this facial performance pipeline is to support various engines and art and animation pipelines across the company. The approaches to derive animation rig blendshapes from the Likeness Acquisition process can differ between game development teams, introducing a variable which significantly affects the robustness of the Performance Capture solver in practice.

Therefore we have introduced an intermediate solve step, using a PCA shapes basis derived for the actor. This enables an initial robust, high-fidelity solve. (Use of the PCA basis can also increase robustness to quality issues in input HMC footage.) The PCA solve is then retargeted to the animation rig shapes, such that subsequent animation work is still possible, if desired due to creative direction; and no modifications are required to the specific pipeline/engine.

Use of the PCA intermediate, and subsequent retarget, are optional; the example performance in the **supporting video** was solved to the animation rig blendshapes directly. The overall philosophy for accommodating challenges encountered in production has been to structure the Performance Capture technology in terms of components which can be used as needed for each scenario.

REFERENCES

Wan-Chun Ma, Mathieu Lamarre, Etienne Danvoye, Chongyang Ma, Manny Ko, Javier von der Pahlen, and Cyrus A. Wilson. 2016. Semantically-aware Blendshape Rigs from Facial Performance Measurements. In *SIGGRAPH ASIA 2016 Technical Briefs (SA '16)*. ACM, New York, NY, USA, Article 3, 4 pages. https://doi.org/10.1145/3005358.3005378