Democratization of real-time facial tracking frameworks for Digital Humans

Carlos L. Vilchis*
Miguel Gonzalez Mendoza*
carlos.vilchis@tec.mx
mgonza@tec.mx
Tecnológico de Monterrey
N.L., Monterrey, México

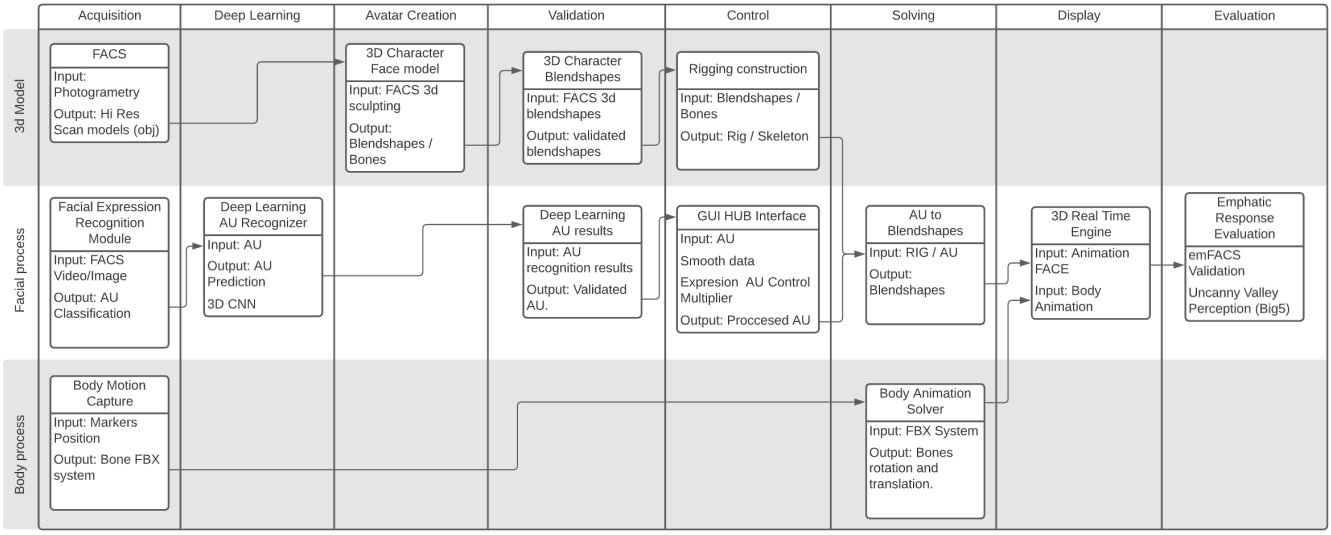

Figure 1: A visual representation of the proposed real-time facial tracking framework along eight steps.

ABSTRACT

The past two decades have brought about a rich scenario of techniques and methods that improve our perception of embodied avatars or digitals humans(the name now in vogue) linked to increased alikeness, realism, emphatic response, and interactivity, as well as a helpful workflow. Breakthrough milestones have been archived partly due to improved photogrammetry imagery, real-time graphics, and facial expression recognition. In particular, this last one has become popular today, thanks to the effectiveness of neural networks. In this work, we created a structure for a real-time tracking framework available to be replicated by peers to develop and operate digital humans. As well as analyze the effectiveness and efficiency in real-time environments.

CCS CONCEPTS

• **Computing methodologies** → *Motion capture.*

*Both authors contributed equally to this research.

MIG'21, November 10–12, 2021, Lausanne, Switzerland

KEYWORDS

Virtual Reality, Facial simulation, Interactivity, Avatars, FACS, Digital Humans, Affective computing

ACM Reference Format:
Carlos L. Vilchis and Miguel Gonzalez Mendoza. 2021. Democratization of real-time facial tracking frameworks for Digital Humans. Poster *MIG'21: Motion, Interaction and Games.*

1 INTRODUCTION

Gartner's recent publication of Hype Cycle of Emerging Technologies for 2021[Burke 2021] places Digital Humans with high expectations as an Innovation trigger. A wide array of applications in the industry make Digital Humans part of the elements needed to improve human interaction with computers, included e-commerce, agent services, government communication, customer services, etc. Just a few companies in the field have access to full integration of the process. Most of them using advanced photogrammetry scans from a real human being, analyzing their facial performance and extract the uniqueness of expressions to replicate in a 3d avatar. The final product, named Digital Human, is made with great detail, an advanced facial rig, motion capture and realistic shaders in a

real-time engine. All these processes mentioned can be summarized as a framework for digital humans.

Talking about the facial mocap to drive expressions in real-time into the digital human face is a separate topic. The tools needed to replicate it (hardware and software) are not easy to integrate, and few fully operational frameworks are available. There is a limited number of them that are public and open research.

Mocap technology related to digital human facial real-time tracking has become hard to find recently. The players changed drastically due to the purchase of Dynamixyz by Take-Two Interactive[Take-Two News Release 2021], making it no more available. In the same price-solution range, Faceware looks effective, but Facegood is a new player to keep an eye on. The priced product/service option based in Cubic Motion has the most effective results yet, but the purchase of this company by Epic games in 2020 lets us see an unclear scenario here. The only affordable (but not open source) option is Apple ARKit based Live Link Face for Unreal Engine. Last case, you have no more options than work developing tools based on computer vision and artificial intelligence to archive professional results.

A custom facial tracking deep learning tool needs to address it in small methods and techniques. This paper proposes a validated framework easy to replicate how to create and operate a facial tracking set of tools into Digital Humans. Specifically, we chose a validated custom Facial Actor Code System (FACS) set of expressions to be recognized by a Deep Neural network. Next, the streaming of coded labels in a custom tool to control/normalize the expression. Finally, the data is retargeted live into Unreal Engine facial rig to be validated with a set of tests to know if the effectiveness of the digital human helps to archive the Uncanny valley tests. Figure 1 is a representation of this framework.

2 METHODOLOGY

This section briefly describes the proposed framework's design.

2.1 Facial Codification

The FACS[Ekman 1997] facial codification has been chosen due to its advantage in representing muscular/skin expressions and the holistic representation of emotions with the emFACS model. Finally, the possibility to be certified by a coder expert in FACS helps to settle a ground of truth about validating facial scans, blendshapes, and final real-time rendering.

2.2 Data Capture and Deep Learning.

To archive a highly customizable tool, the framework describes the use of an Artificial Intelligence tool. The advantage of free and open-to-use pre-trained models based on Convolutional Neural Networks (CNN) makes it easy to replicate. A complete description of expressions needed to train and merge into a Lightweight Neural Network will improve time and speed to archive realistic results, a key feature to break the latency working with a real-time conversational agent inside an experimental enviroment like this[Seymour et al. 2017].

2.3 Asset creation and rig control.

The next step is to create the digital human asset and the photogrammetry scans of the FACS, cleaning, and correct retopology to keep lightweight blendshapes. Extra features to improve quality, like hair cards, eyebrows, facial hair, realistic skin shaders, etc., each of those needs optimized for a real-time engine like Unreal Engine and using state of the art rigging techniques like eyelids deformation, flushed cheeks due to the blood, etc.

2.4 Real-time solving and display.

As other researchers did[Aneja et al. 2019], the framework includes developing a software tool based on Python. This tool listens via TCP to the Actor Units(AU) delivered by the Neural Network on a priority basis into Unreal Engine. Final rendering is archived thanks to the set of tools inside the engine like LiveLink, for body/fingers motion capture.

2.5 Evaluation.

Recent authors[Zibrek et al. 2019] define a new set of tools to validate the emphatic response of Digital Humans. The emFACS model can be rated using traditional survey methods and double-checked with FACS certified coder. Last, the Uncanny valley perception using the live experiment "Wizzard of Oz" on a Likert scale.

3 IMPLEMENTATION DETAILS AND RESULTS

Neural Networks can always be improved with new techniques and models to improve accuracy and time. Better results can be archived by doing new iterations of the model. In this work, we tested each step individually; this allows us to understand the full set of steps and areas to improve. Our early results let us get a recognition rate of emFACS of 80%, higher than the results presented by HapFACS[Amini et al. 2015] of 72% in similar conditions.

4 CONCLUSION AND DISCUSSION

In this work, we propose an open framework to do real-time facial tracking into digital humans using a Neural Network. In the future, we will work on improving the model to increase the emphatic response of the proposed framework

REFERENCES

Reza Amini, Christine Lisetti, and Guido Ruiz. 2015. HapFACS 3.0: FACS-based facial expression generator for 3D speaking virtual characters. *IEEE Transactions on Affective Computing* 6, 4 (2015), 348–360.

Deepali Aneja, Daniel McDuff, and Shital Shah. 2019. A High-Fidelity Open Embodied Avatar with Lip Syncing and Expression Capabilities. In *2019 International Conference on Multimodal Interaction* (Suzhou, China) *(ICMI '19)*. Association for Computing Machinery, New York, NY, USA, 69–73. https://doi.org/10.1145/3340555.3353744

Brian Burke. 2021. *Hype Cycle for Emerging Technologies, 2021.* Technical Report. Stamford, CT 06902 USA.

Rosenberg Ekman. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS).* Oxford University Press, USA.

Michael Seymour, Kai Riemer, and Judy Kay. 2017. Interactive Realistic Digital Avatars-Revisiting the Uncanny Valley. (2017).

Take-Two News Release 2021. Take-Two Interactive Software Acquires Dynamixyz. https://www.dynamixyz.com/.

Katja Zibrek, Sean Martin, and Rachel McDonnell. 2019. Is Photorealism Important for Perception of Expressive Virtual Humans in Virtual Reality? *ACM Trans. Appl. Percept.* 16, 3, Article 14 (Sept. 2019), 19 pages. https://doi.org/10.1145/3349609