Facial Expression Transfer from Video Via Deep Learning

Xiao S. Zeng*
stevennz@cs.ucla.edu
University of California, Los Angeles
Los Angeles, California, USA

Surya Dwarakanath*
suryadwar@g.ucla.edu
University of California, Los Angeles
Los Angeles, California, USA

Wuyue Lu
luwuyue@cs.ucla.edu
University of California, Los Angeles
Los Angeles, California, USA

Masaki Nakada
masaki@neuralx.ai
NeuralX, Inc
Los Angeles, California, USA

Demetri Terzopoulos
dt@cs.cula.edu
University of California, Los Angeles
Los Angeles, California, USA

ABSTRACT

The transfer of facial expressions from people to 3D face models is a classic computer graphics problem. In this paper, we present a novel, learning-based approach to transferring facial expressions and head movements from images and videos to a biomechanical model of the face-head-neck musculoskeletal complex. Specifically, leveraging the Facial Action Coding System (FACS) as an intermediate representation of the expression space, we train a deep neural network to take in FACS Action Units (AUs) and output suitable facial muscle and jaw activations for the biomechanical model. Through biomechanical simulation, the activations deform the face, thereby transferring the expression to the model. The success of our approach is demonstrated through experiments involving the transfer of a range of expressive facial images and videos onto our biomechanical face-head-neck complex.

CCS CONCEPTS

• **Computing methodologies** → **Animation**; *Computer vision*; *Machine learning*; **Physical simulation**.

KEYWORDS

Deep Learning, Physics-Based Facial Animation, Muscle-Driven Model, Facial Expression, Facial Action Coding System

ACM Reference Format:
Xiao S. Zeng, Surya Dwarakanath, Wuyue Lu, Masaki Nakada, and Demetri Terzopoulos. 2021. Facial Expression Transfer from Video Via Deep Learning. In *The ACM SIGGRAPH / Eurographics Symposium on Computer Animation (SCA '21), September 6–9, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3475946.3480959

1 INTRODUCTION

Biomechanical human musculoskeletal models aim to realistically capture the anatomy and physics underlying human motion generation. Much of the research in this subdomain of computer animation has focused on modeling the body, but only a few such models have focused on simulating the human face (e.g., [Lee et al. 1995]). These have succeeded in realistically emulating facial expression generation; however, they require significant effort in parameter setting and tuning to produce realistic results.

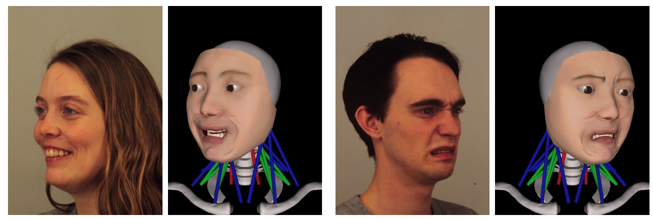

Figure 1: Our deep learning framework transfers facial expression and head pose from a real human image or video to a biomechanical face-head-neck model.

In this paper, we show how to endow a biomechanical, musculoskeletal model of the human face with the ability to perform facial expressions by machine learning from real-world reference images and videos (Fig. 1). To this end, we introduce a deep-neural-network-based method for learning the representation of human facial expressions through Ekman's Facial Action Coding System (FACS) [Cohn et al. 2007] in the context of the muscle actuators that drive the musculoskeletal face model. We furthermore augment the face animation system with a musculoskeletal cervicocephalic (neck-head) system to animate head movement during facial expression synthesis.

As a proof of concept, we demonstrate an automated processing pipeline (Fig. 2) for animating expressions and head poses using an improved version of the physics-based face-head-neck animation system developed by Lee and Terzopoulos [2006], but which can potentially be applied to any physics-based, muscle-actuated model.

2 MUSCULOSKELETAL MODEL

Our real-time musculoskeletal model is based on that of [Lee and Terzopoulos 2006], but both the underlying face-head-neck control system and the facial expression system are significantly improved.

The skeletal structure is an articulated multibody dynamics system, with bones and joints consistent with human anatomy. The skeletal model is driven by a Hill-type muscle actuator model. The biomechanical face component consists of a facial soft tissue model

*Both authors contributed equally to this research.

SCA '21, September 06–09, 2021, Online

ACM ISBN 978-1-4503-8676-0/21/09.
https://doi.org/10.1145/3475946.3480959

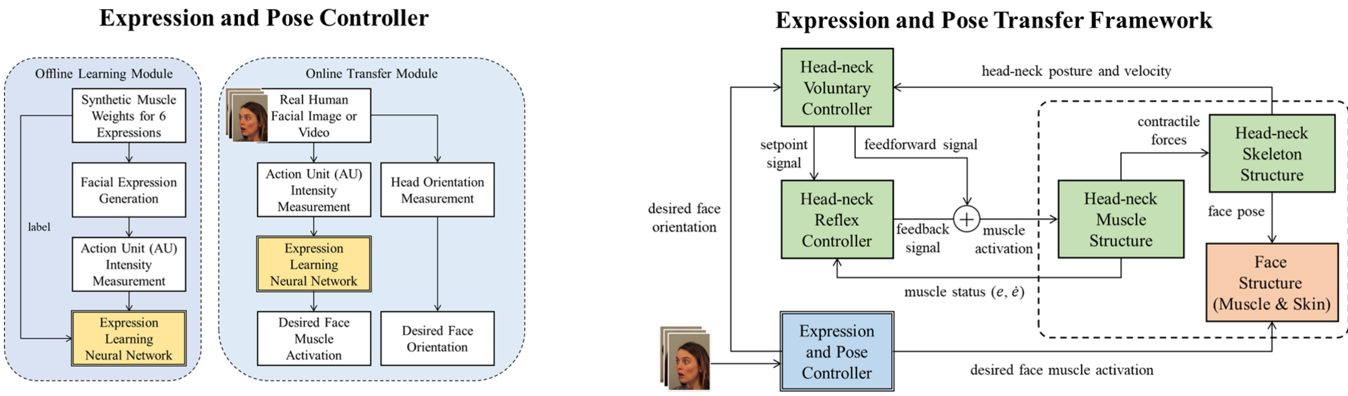

Figure 2: Expression and head pose controller (left) and the overall structure of our transfer framework (right). The expression learning neural network (yellow) is first trained offline. In a transfer task, an image or a video sequence of a real face is fed into the controller to output the desired facial muscle activations and head orientation information, which is then passed to the face model (orange) and the head-neck system (green) to produce expression and head pose.

comprising synthetic skin and muscle layers together with a skull beneath them, all of which are constructed based on the work by Lee et al. [1995]. The soft tissue model is a deformable model assembled from discrete uniaxial finite elements, which simulates dynamic facial deformation in an anatomically consistent yet simplified way. The contractions of the embedded 26 pairs of muscles apply forces to the facial tissue layers, which deform to produce meaningful facial expressions. We augment the expressive details such as wrinkles on the face model by applying multiple levels of subdivision to increase the number of facial nodes that can be influenced by muscle forces. We also adapt a high resolution texture image to our generic face mesh to provide a more natural look.

3 CONTROLLING THE MODEL

To control the face-head-neck system, our novel neural network-based expression and pose controller (Fig. 2) generates facial muscle activations that produce recognizable expressions. It simultaneously outputs head pose estimates to the head-neck complex, where voluntary and reflex neuromuscular control layers generate cervical muscle activation signals to achieve the desired head orientations.

3.1 Expression and Pose Learning

Our novel deep neuromuscular motor controller learns to map between FACS Action Units (AUs) extracted from human facial images and the activations of the muscle actuators that drive the biomechanical system. We leverage the aforementioned biomechanical face model and OpenFace 2.0 [Baltrusaitis et al. 2018] to synthesize a large quantity of training data pairs each consisting of (i) muscle and jaw activations and (ii) the associated normalized AUs and head orientations. These pairs are used to train our deep neural network to input AUs and output corresponding muscle and jaw activations offline. We then use a pipeline similar to the training module for the transfer of real facial expressions on the fly. We input an image of an expressive face into OpenFace to obtain AUs and head orientations. The AUs are then normalized and passed into the trained neural network which outputs predictions of the muscle and jaw activations, driving the biomechanical face to deform the muscles and transfer the expressions onto it. We transfer both image and video inputs. Each frame in a video is processed independently and a resulting video is created using the transferred frames.

4 RESULTS

We evaluate our expression and head pose transfer pipeline on different expressions and head orientations while using a variation of AUs and muscles in the biomechanical face-head-neck model. Fig. 1 shows example transfer results of a male subject and a female subject enacting joy and disgust expressions from the Karolinska Directed Emotional Faces dataset [Calvo and Lundqvist 2008].

Our expression transfer approach is uniquely advantageous because it is anatomically consistent and is based on the FACS, which is a widely adopted representation of facial expressions in the computer animation field. Further, our approach leverages the power of deep learning to establish a non-linear relation between the Action Units of the FACS and facial muscle activations.

ACKNOWLEDGMENTS

We thank CyberAgent, Inc., for providing the high resolution texture for the face mesh used in our biomechanical face model.

REFERENCES

Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 59–66.

Manuel G. Calvo and Daniel Lundqvist. 2008. Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods* 40, 1 (01 Feb 2008), 109–115.

Jeffrey F Cohn, Zara Ambadar, and Paul Ekman. 2007. Observer-based measurement of facial expression with the Facial Action Coding System. *The Handbook of Emotion Elicitation and Assessment* 1, 3 (2007), 203–221.

Sung-Hee Lee and Demetri Terzopoulos. 2006. Heads up! Biomechanical modeling and neuromuscular control of the neck. In *ACM SIGGRAPH 2006 Papers*. 1188–1198.

Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. 1995. Realistic modeling for facial animation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 55–62.