Simulation-aided face strain extraction for ML animation systems

Gergely Klár*
NVIDIA
Wellington
New Zealand
gklar@nvidia.com

Stephen J. Ward*
Apple Inc.
Cupertino, USA
s.ward@apple.com

Andrew Moffat
Wētā Digital
Wellington
New Zealand
amoffat@wetafx.co.nz

Eftychios Sifakis*
Univ. of Wisconsin-Madison
Madison, USA
sifakis@cs.wisc.edu

Ken Museth*
NVIDIA
Santa Clara, USA
kmuseth@nvidia.com

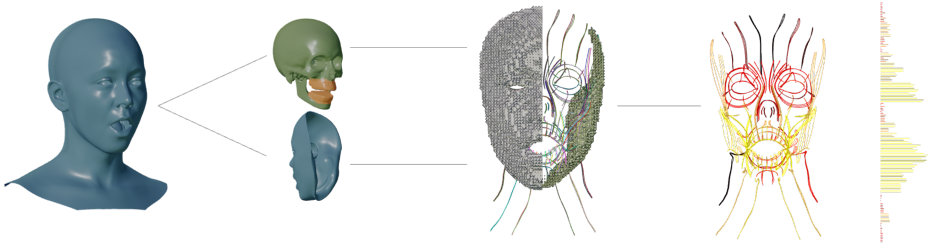

Figure 1: We leverage simulation to extrapolate dynamic performance capture into volumetric tissue simulation, from which deformation of muscle curves and strain measurements are extracted, to be used as features in ML-based animation tools.

ABSTRACT

We present a volumetric, simulation-based pipeline for the automatic creation of strain-based descriptors from facial performance capture provided as surface meshes. Strain descriptors encode facial poses via length elongation/contraction ratios of curves embedded in the flesh volume. Strains are anatomically motivated, correlate strongly to muscle action, and offer excellent coverage of the pose space. Our proposed framework extracts such descriptors from surface-only performance capture data, by extrapolating this deformation into the flesh volume in a physics-based fashion that respects collisions and filters non-physical capture defects. The result of our system feeds into Machine Learning facial animation tools, as employed in *Avatar: The Way of Water*.

CCS CONCEPTS

• **Computing methodologies** → **Physical simulation**.

KEYWORDS

Elasticity simulation, facial animation, data-driven animation.

ACM Reference Format:
Gergely Klár, Stephen J. Ward, Andrew Moffat, Eftychios Sifakis, and Ken Museth. 2023. Simulation-aided face strain extraction for ML animation systems. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '23 Talks), August 06–10, 2023, Los Angeles, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3587421.3595454

*Work done while at Wētā Digital.

SIGGRAPH '23 Talks, August 06–10, 2023, Los Angeles, CA, USA

ACM ISBN 979-8-4007-0143-6/23/08.
https://doi.org/10.1145/3587421.3595454

1 INTRODUCTION

The most common approach for facial animation is the use of blendshape rigs, where the deformed facial pose is realized as the weighted combination of blendshapes in a carefully crafted basis. Choi et al. [2022] present and demonstrate the benefits of an alternative system design that is based directly on facial capture data. They obtain a richer, anatomically based description of expressions by choosing muscle strain measurements as the fundamental representation of facial action.

Strain measurements are associated with one or more *muscle curves*. A muscle curve corresponds to a representative line of action for a section of a given muscle, chosen to capture its deformation; some muscles can be associated with several muscle curves to capture variability across their extent. *Strains* are defined via the ratio of the deformed and undeformed lengths of such curves. The benefits of this representation are that strain values are robust to variations of facial proportions and feature shapes, they are anatomically inspired, and muscle curves provide a good visual representation for local deformations to users. Strains are also an excellent representation for Machine Learning tools, as they capture well the *expression manifold* of plausible face animations. Muscle strains, however, are not natively a part of the output produced by traditional facial capture pipelines, and need to be computed from the deformed muscle curves. We produce these by embedding the muscle curves into a volumetric representation of the flesh tissue in the face, and proceed to deform this volume using a physics-based solver that tracks the performance-captured surface deformation.

Muscle activations are the traditional action descriptors in forward simulations, but would add unnecessary complexity for our use case. The results of an activation-based muscle simulation are sensitive to geometric and physical parameters of the model, are limited by the anatomical detail and sophistication of the model. These factors may lead to a significantly larger feature space without additional benefits. In contrast, strains convey the nuance of facial expressions without detailed knowledge of the actor's facial anatomy.

We employ simulation to extrapolate surface capture into the flesh volume, while respecting physical constraints (e.g. collisions) and filtering surface defects of the capture stage by targeting the capture with a volumetric elasticity simulation. In production, the output of our system was used as input to downstream Machine Learning animation tools that used strains as pose descriptors.

2 DATA PIPELINE

Our solver is part of the data preparation stage of the training pipeline for Machine Learning animation systems, and it operates on 3 classes of inputs: rig-specific, actor-specific, and frame-specific.

Rig-specific inputs are the number, position, and orientation of the cameras. These are optional inputs and are used for generating *confidence maps* for the capture data.

Actor-specific inputs need to be authored for each actor, and capture their morphology. These inputs are the bones (skull and mandible), the facial slab (watertight volumetric geometry of the simulated fleshy parts of the frontal face), the neutral face-mask (the face-capture geometry in a neutral pose), passive elements (muscle geometry and muscle curves), and attachment maps. All of these are defined in the neutral pose of the actor.

Frame-specific inputs are the outputs of the dynamic scan and the jaw alignment steps of the acquisition pipeline. The dynamic scans provide the face-mask, while the jaw alignment defines the transformation of the mandible. These frames of data are grouped into motion clips, corresponding to FACS actions, emotions, and phonemes. However, this organization serves the convenience of the users, and is not a requirement by the solver.

3 SOLVER SETUP AND SIMULATION

The solver we developed to simulate the facial-slab's deformation is a quasi-static Newton solver for elasticity. The solver computes the equilibrium state of the elastic deformations of a tetrahedral mesh encasing the face-slab. The deformation is driven by the difference between the current and the neutral face-mask. Since we focus on the composite effect of all the constituents of the facial anatomy, we model the facial slab as homogeneous material. Our solver supports the Shape Targeting [Klár et al. 2020], the Stable Neo-Hookean, and the Fixed Co-rotated material models.

The simulation mesh is a BCC lattice-based tetrahedral mesh that we generate around the facial-slab. We use adaptive coarsening of the inside of the tetmesh to reduce the number of simulated elements. Our solver supports user-guided refinement to increase the resolution around the eyes and mouth. However, in practice the refinement does not have a noticeable effect on the strain measurements. Furthermore, our simulation tetmesh is non-manifold, allowing faces to have more than two neighbors. This helps greatly in capturing the sharp corners of the mouth and eyes.

The muscle curves and muscle geometries are passively embedded into the simulation tetmesh using the vertices' barycentric coordinates, and whose positions are updated at equilibrium. The resulting deformed curves are the primary output of the solver. The rest of the pipeline uses these to compute muscle strains. The muscle geometries are only used for visualization, and do not contribute to the deformation.

Collision handling requires special considerations because of the quasi-static nature of the problem. Vertices of the lips can have large displacements with trajectories that have many collision-free but physically impossible positions, such as vertices stuck on the inside of the teeth. To alleviate this, we disable collision detection for the first roughly 80% of Newton iterations, allowing the vertices to settle into the proximity of their final positions, at which point we enable collision detection for the remaining iterations.

Even with this modification, however, snagged vertices may still occur, for example with a tight *lip corner puller* with an open jaw. For this reason, we replaced the accurate thin teeth models in favor of "bulked up" proxies.

Parallel processing is trivially possible with quasi-static solvers as the results are independent of inertial effects. At the same time, sequential simulation of consecutive frames has the benefit of each solved frame providing a great first estimate for the following one, significantly improving simulation performance.

Attachment maps define the regions where the flesh is attached to the bone. The attachments are modeled by pinning the nearby vertices of the facial-slab to the skull or mandible. In other areas, the facial-slab is free to slide over the bones.

Confidence maps modulate the attachment strength between the face-mask and the simulation mesh, preventing overfitting to regions of poor reconstruction, e.g. under the chin of an open mouth.

4 REFINEMENT AND AUGMENTATION

Filtering captured expressions through simulation has the added benefit of refining and augmenting the results of the dynamic scans:

- Through the confidence maps, regions occluded from the cameras will follow the bulk motion of the tissues.
- Collisions of the lips with each other or the teeth will be resolved, resulting in a mesh free of interpenetration.
- The insides of the lips and cheeks, that are absent from the dynamic scans, are recreated through the simulation.

5 CONCLUSION

By tailoring a quasi-static elasticity solver we created a tool that computes muscle strains from frontal facial shells acquired by 4D scanning. These strains are the ground truth descriptions of facial expressions used in training the Wētā FX facial animation system.

ACKNOWLEDGMENTS

We would like to express our gratitude to Millie Maier, Stephen Cullingford, and the Wētā FX leadership for their support; Joe Letteri and Marco Revelant for their invaluable feedback; Nikolay Ilinov for his excellent work on the BCC generator tool; BK Choi, HK Eom, Benjamin Mouscadet, and the rest of the Facial R&D team.

REFERENCES

B. Choi, H. Eom, B. Mouscadet, S. Cullingford, K. Ma, S. Gassel, S. Kim, A. Moffat, M. Maier, M. Revelant, J. Letteri, and K. Singh. 2022. Animatomy: An Animator-Centric, Anatomically Inspired System for 3D Facial Modeling, Animation and Transfer. In *SIGGRAPH Asia 2022 Conference Papers (SA '22)*. Article 16, 9 pages.

G. Klár, A. Moffat, K. Museth, and E. Sifakis. 2020. Shape Targeting: A Versatile Active Elasticity Constitutive Model. In *ACM SIGGRAPH 2020 Talks*. Association for Computing Machinery, New York, NY, USA, Article 59, 2 pages.