Facial Animation Based on 3D Scans and Motion Capture

Martin Breidt, Christian Wallraven, Douglas W. Cunningham, Heinrich H. Buelthoff[*]
Max Planck Institute for Biological Cybernetics, Tuebingen, Germany

1 Introduction

One of the applications of realistic facial animation outside the film industry is psychophysical research in order to understand the perception of human facial motion. For this, an animation model close to physical reality is important. Through the combination of high-resolution 3D scans and 3D motion capture, we aim for such a model and provide a prototypical example in this sketch.

State-of-the art 3D scanning systems deliver very high spatial resolution but usually are too slow for real-time recording. Motion capture (mocap) systems on the other hand have fairly high temporal resolution for a small set of tracking points. The idea presented here is to combine these two in order to get high resolution data in both domains that is closely based upon real-world properties. While this is similar to previous work, for example [Choe et al. 2001] or [Pighin et al. 2002], the innovation of our approach lies in the combination of precision 3D geometry, high resolution motion tracking and photo-realistic textures.

2 Morph shapes

Facial geometry is captured using a structured light scanner from ABW GmbH. The scan takes about two seconds, so it is possible to capture facial expression without having to hold the pose for a long time. This produces about 3 million vertices for a single face scan.

The individual scans are cleaned and put into correspondence with each other in order to create a basis of morphable meshes. This is currently done by manually aligning a template control grid of 238 points to each scan. Now polygonal meshes of arbitrary resolution can be generated using subdivision schemes. The faces in Figure 2 each consist of 11,656 vertices. For this publication, a small set of facial expressions have been produced that way but it is planned to record a large range of facial poses.

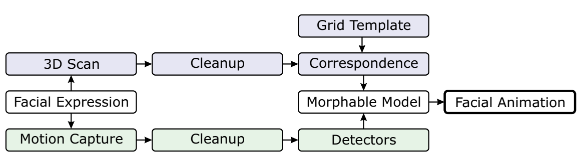

Figure 1: Our facial animation pipeline.

For better visual fidelity, we have fitted a seperate photography of the neutral face to the neutral morph shape. This texture map is deformed during morphing and partially suffers from stretching due to limited resolution and the frontal projection. In the finished system, high resolution pictures of the actual expression taken during the scanning process will be used to further increase the realism.

3 Motion Capture

Having generated a set of morphable expressions, we capture motion data with an optical Vicon system. For this, 72 markers are applied to the face of the person that was previously scanned. The person is then asked to perform the same expressions that were scanned. Since the mocap data is only used for timing and qualitative analysis, it is not necessary to have a precise geometric match between the movements while scanning and during mocap.

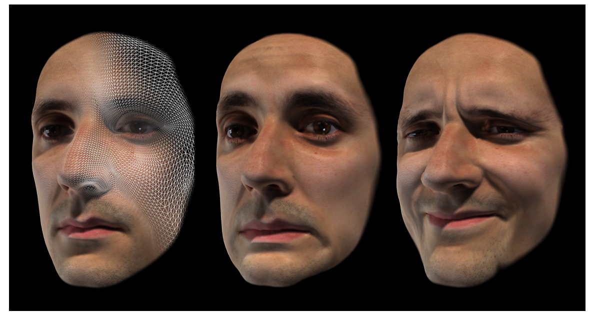

Figure 2: Rendered 3D morph shapes.

After clean-up of the mocap data, rigid head movement is temporarily removed in order to isolate the non-rigid facial movement. Currently, simple linear detectors for facial action elements use distances between markers for driving the morph channels. This produces morph animation based on the amplitude and timing of marker motion in the mocap data. Finally, rigid head movement is re-applied from the original mocap data.

Since mocap data is not directly used for facial animation but for deriving morph channels instead, it is easy to transfer the motion to other face models with the same morph basis while retaining the temporal quality of mocap. This will be used for real-time animation of an avatar in the IST research project COMIC[1].

4 Conclusion

As our video example[2] shows, it is beneficial to use best of both worlds: high definition scans for capturing the surface deformation and mocap data for amplitude, timing and coordination of the motion elements.

In order to make the described system practical for large-scale psychophysical research, we need to automate the process of correspondence calculation between facial expressions. A full set of individual muscle movements based on Ekman's FACS will be scanned and motion-captured. For the transcription of mocap data into FACS, both machine learning and optimization techniques are currently being investigated.

The authors would like to thank Barbara Knappmeyer, Winfried Ilg and Kai Wolf for their help in this project.

References

CHOE, B., LEE, H., AND KO, H.-S. 2001. Performance-driven muscle-based facial animation. *J. Visual. Comput. Animat. 12*, 67–79.

PIGHIN, F., SZELISKI, R., AND SALESIN, D. H. 2002. Modelling and animating realistic faces from images. *J. of Computer Vision 50*, 2, 143–169.

[*]e-mail: firstname.lastname@tuebingen.mpg.de

[1]`http://www.hcrc.ed.ac.uk/comic/`

[2]`http://www.kyb.tuebingen.mpg.de/~mbreidt/sig2003.html`