Data-Driven Face Modeling and Animation

1. Research Team

Project Leader:	Prof. Ulrich Neumann, *IMSC and Computer Science*
Post Doc(s):	John P. Lewis
Graduate Students:	Zhigang Deng, Zhenyah Mo
Undergraduate Students:	Albin Cheenath

2. Statement of Project Goals

Avatars, computer animation, and other forms of human-centric computing all potentially require the generation of novel but realistic facial appearance and motion. Such face modeling and animation is currently the province of computer artists. While manually animated computer faces can be quite expressive, they usually fall short of complete realism in their appearance and motion (the recent movie Final Fantasy is one demonstration of the current state of the art). The manual effort required to model and animate faces also prevents their use in avatars and other contexts where the required behavior is not known in advance.

Data-Driven Face Modeling and Animation is a new project that explores the use of facial models made directly from captured data to address the goals of realism and automation. More specifically, a range of typical data (either face shapes or motions) is first captured and manually labeled. A high level model of this data is created. Once this model is available, its parameters can be explored to extrapolate or synthesize novel face shapes or motions. While this general data-driven approach is a recent theme across several research groups, IMSC already has some unique results, as shown below.

3. Project Role in Support of IMSC Strategic Plan

Data-Driven Face Modeling and Animation is part of the general IMSC effort towards expressive human interaction in virtual and augmented reality environments. It is one of a trio of technologies being developed specifically for placing humans in virtual spaces. The complementary Facial Expression Processing project is directed towards capturing human expression (including non-speech gestures) and adapting the results to different computer facial models. The third project in the trio, Hair Modeling and Animation, addresses a major gap in current computer graphics renditions of humans.

4. Discussion of Methodology Used

We have developed two initial demonstrations of a data-driven approach to face modeling. The first models the combined shape and appearance of a database of two dimensional faces. We demonstrate the capabilities of this model by extrapolating hidden regions of previously unprocessed faces (Figure 1). The second demonstration builds a mouth viseme and eye movement model from captured three-dimensional speech motion. This model is then used to synthesize facial motion on a 3D avatar in response to novel speech.

In the two-dimensional statistical face modeling demonstration, we adopt a principal component model of both the texture and shape of a number of faces. Specifically, a number of feature points are manually labeled on a large database of face photographs. A linear subspace (eigenspace) describing the shape variation across these feature points is obtained. The shape variation is then removed by morphing all faces to the mean shape and a second model describing the face texture variation is obtained.

This basic shape+texture linear subspace modeling approach was independently developed by a number of research groups and goes by various names including Active Appearance Models and Morphable Models. Our contributions include:

- *Group modeling and outlier rejection*. A comprehensive facial model should be able to represent faces of both sexes and of different ages and ethnicities. On the other hand, if abundant facial data for young asian females (for example) is available, adding additional data on (for example) older black males will not help much in improving the data model for the young asian female, and indeed it may degrade the model. In our work we replace the global principal component analysis with a representation that considers near neighbors to be part of the same "group". This enhances modeling and prediction even given limited data (see Figure 1).
- *Model extrapolation*. We use the statistics implicit in the data to extrapolate new face regions. In Figure 1, regions of novel faces not among the training pictures (left) are removed (center). The model then predicts the missing regions (right). The prediction is not always accurate, but if there are any statistical regularities in the data they will be revealed by our process. For example, if a particular "type" of face almost always has a wide mouth and thin lips, then the prediction will be of a wide mouth and thin lips. If the statistical trends are weak then the extrapolated result will appear blurry, as in the second row of Figure 1.

The human face is of paramount importance to other humans, and we can effortlessly make refined judgments about the attention and emotional qualities conveyed by a speaker. The eyes and mouth are the parts of the face that are most directly used to convey information, and are the areas that are most attended to by an observer. In our second exploration of a data-driven approach, a three-dimensional avatar is animated from novel audio, but using a data-driven approach to synthesizing the crucial mouth and eye movements. To obtain the mouth motion model, facial motion capture data from the mouth region is analyzed to obtain a movement subspace, then operating in this subspace a smooth viseme co-articulation path for any given utterance is obtained using a dynamic programming approach similar to that used in [1]. The resulting co-articulated viseme sequence is applied to a 3D avatar model using morphing by radial basis function interpolation.

Eye movements and blinks are not strongly correlated with speech sounds, so we adopted a statistical texture synthesis method to capture this lack of correlation while still producing characteristic movements. Specifically, we employ the popular non-parametric sampling method [2], operating on the one-dimensional eye-blink signal (see Figure 2) and the quaternion eye orientation signal (data obtained from motion capture and hand labeling). Eye vergence will be considered in future work.

Figure 1. Regions of novel faces not among the training data (left) are obscured (center). Our model then predicts the missing regions (right).

5. Short Description of Achievements in Previous Years

The Data-Driven Face Modeling and Animation effort is a new project, started within the last year. It uses experience we gained in the ongoing Facial Expression project, described in a separate project summary.

5a. Detail of Accomplishments During the Past Year

This is a new project undertaken during the last year. Section 4 describes the accomplishments to date.

6. Other Relevant Work Being Conducted and How this Project is Different

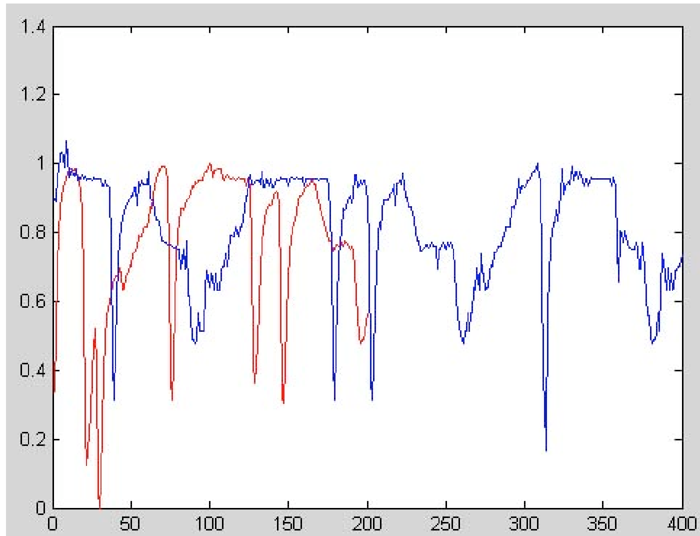

Figure 2. Motion capture data for eye blinks over time (red graph), and novel synthesized eye blink motion (blue graph).

"Data driven" modeling is a current trend in computer graphics research, and it has been applied to speech modeling [3] as well as other areas such as body animation and texture synthesis.

Our work is distinguished in several ways. The rejection of outliers in the two-dimensional statistical model refines the modeling capability offered by the basic appearance model approach, but without introducing additional complexity. We are the first group applying data-driven modeling to eye motion as well as speech.

7. Plan for the Next Year

It is most effective to evaluate research results in the context in which they will eventually be used. One of our main goals is to enable natural and expressive avatars. As such we will work towards a version of data-driven facial modeling that is suitable for avatars used in the Communication testbed.

Realism is a major challenge for computer face renditions – while stylized or cartoon-like faces are generally effective, attempts at realism often appear rigid or even "creepy". Since realism is not the focus of our research, we will explore data-driven approaches to stylized face renditions.

8. Expected Milestones and Deliverables

The current deliverables include several manually annotated facial databases, and the two prototype programs described above (roughly 10,000 lines of Java and C++ respectively). Our future milestones are indicated as follows:

- Stylized facial rendering from data
- Data-driven lip-synch: Improvements to quality and speed

9. Member Company Benefits

Sophisticated statistical models of face appearance such as the one we prototyped have a number of uses, including:

- Older or younger versions of a face can be approximately extrapolated, by first fitting the face using the model and then moving along an age axis in "face space" [4].

- The estimation of a missing face part may have application in reconstructive surgery: in cases where a face region is deformed at birth, there is no prior photograph to provide a guide for the surgery.
- Similarly, the estimation of missing face parts may be useful in security or other person identification contexts, as when a photograph of an individual is available but portions of the face are obscured.
- Existing face-finding algorithms often cannot localize parts of a found face other than the eyes and mouth center. By pairing such an algorithm with a statistical appearance model it will be possible to automatically locate many specific facial features such as the mouth corners, nostrils, points along the eyebrows, etc. This capability is in turn useful in diverse areas including previewing medical procedures, virtual cosmetic preview, and various computer gaming applications.

10. References

[1] CHUANG, E., AND BREGLER, C. 2002, Performance Driven Facial Animation using Blendshape Interpolation, *CS-TR-2002-02*, Department of Computer Science, Stanford University.

[2] EFROS, A.A., AND LEUNG, T.K. 1999, Texture Synthesis by Non-parametric Sampling, *ICCV'99*, 1033—1038.

[3] EZZAT, T., GEIGER, G., AND POGGIO, T. 2002, Trainable Videorealistic Speech Animation, *ACM Transaction on Graphics, 21,* 3, 388—398.

[4] B. Tiddeman, D. Burt, and D. Perret, Prototyping and Transforming Facial Textures for Perception Research, *IEEE Computer Graphics and Applications* 21, 5, 42-50, Sept./Oct 2001.