Facial Type, Expression, and Viseme Generation (sap_0202)

James Skorupski Jerry Yee Josh McCoy James Davis
University of California, Santa Cruz
{jskorups, yee, mccoyjo, davis}@cs.ucsc.edu

1 Introduction

The process of generating facial models and various poses of these models is a necessary part of most present-day movies, and usually required for any interactive game which features humans as a primary character. The generation of this face data can be approached in ways varying from pure computation to pure data acquisition. Computational models are flexible but can lack realism and intuitive or simple controls, while data-driven models produce realistic faces, but necessitate the often slow and cumbersome capture of new scan data for every desired set of face attributes. Our method is a hybrid approach, which combines a relatively small set of real world facial data with a computational algorithm that learns the underlying variations in this geometric information automatically. Given a sparse data set that spans variation in viseme, face type, and expression, we are able to generate new faces that exhibit combinations of these attributes, and were never part of the original data set. We rely on user-assisted categorization of our sparse data set to associate each piece of face data with a small set of attribute contributions, and then use this categorization data as a guide for binding abstract variation to concrete parameters. This process takes the complex, subtle, and often subjective qualities associated with visemes, expressions, and face types, and correlates them to known geometric features, in order to facilitate the creation of entirely new face poses.

2 Method

Our algorithm is separated into three primary stages: tagging, mesh analysis, and generation. During the tagging step, a user is presented with each model from the original facial data set, and he or she judges the contributions of various attributes for that specific model. Each attribute contribution is a continuous value from zero to one, represented by a slider in our user interface. The three categories of attributes, or axes of variation, are face **types** (1), **expressions** (2), and **visemes** (3). Face types can be arbitrary qualities that vary from high level qualities such as femininity or weight, to specific features such as chin definition. Expression attributes are the contribution of a basic set of emotions, including joy, sadness, anger, fear, disgust and surprise. Our viseme attributes are derived from the known set of visemes and their associated phonemes, studied in previous research.

After tagging, we then analyze each of our three sets of faces, and determine how these user-generated attributes correlate directly to the principle components of geometric variation. First, for each set of face types, expressions, and visemes, we perform per-vertex principle component analysis to generate a common set of eigenvectors and multiple sets of unique eigenvalues for all analyzed faces in a specific data set. We next compare these eigenvalues to the user-submitted tagging information, and learn a relationship between the contributions of each eigenvector, and the strength of the user-provided attributes. We perform this regression across all attributes and eigenvalues by generating a relationship matrix **M** between the concrete eigenvalues in matrix **W** and subjective attributes in matrix **A**, where $\mathbf{M} = \mathbf{W}\mathbf{A}^{-1}$. These matrix arrangement and correlation methods are similar to the methods used in previous work with body shape deformation [Allen 2004]. This matrix calculation is performed three times, once for each separate axis of facial variation, resulting in three relationship matrices.

Finally, we generate new faces by acquiring from a user a set of desired values for each predetermined face attribute, **a**, and transforming them, with a relationship matrix **M,** into a set of generated eigenvalues, **w**: $\mathbf{w}_{generated} = \mathbf{M}\mathbf{a}_{desired}$. This calculation is performed three times with the three sets of requested attribute values relating to type, expression, and viseme, to generate three new sets of eigenvalues. Each of these new eigenvalue vectors is multiplied by their corresponding set of pre-calculated eigenvectors and summed together to generate new vertex offsets. A unique new face is then generated by taking the average face type, and perturbing it with those type, expression, and viseme vertex offsets.

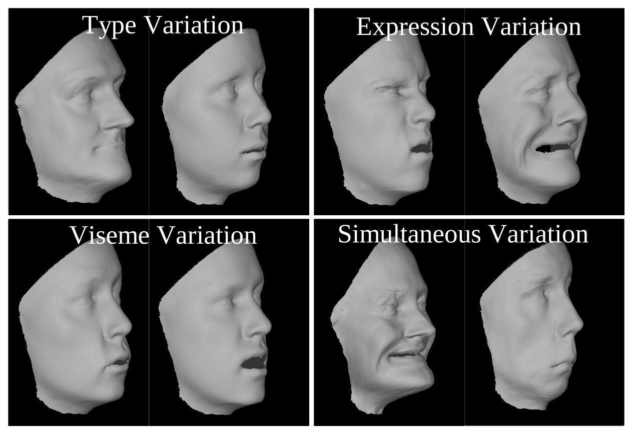

Figure 1. Generated faces demonstrating three axes of variation.

3 Results

Our three-axis data set consisted of 75 face types models with neutral expressions and visemes from the work of Blanz et al [1999], and 384 viseme and expression models of a single face type from the work of Li Zhang et al [2004]. Figure 1 shows the results of varying the type, expression, viseme axes separately, and in the final frame, the combination of geometric variation. Our method is ideal for generating realistic faces, because it leverages the use of a sparse and manageable set of real-world data, and at the same time is able to capture and intuitively represent the complex interactions between face types, expressions, and visemes throughout an entire human face.

References

ALLEN, B., CURLESS, B., and POPOVIĆ, Z. 2004. Exploring the space of human body shapes: data-driven synthesis under anthropometric control. In *Proc. Digital Human Modeling for Design and Engineering Conference*. SAE International.

BLANZ, V. and T. A. VETTER. 1999. Morphable Model for the Synthesis of 3D Faces. In *Proc. of ACM SIGGRAPH 1999*. 187-194.

ZHANG, L., ET AL. 2004. Spacetime faces: high resolution capture for modeling and animation. In *Proc. of ACM SIGGRAPH 2004*. 548-558.