FaceSpace: A Facial Spatial-Domain Toolkit

Steve DiPaola

Simon Fraser University

10153 King George Hwy

Surrey, BC V3T 2W1 CANADA

+1 604 586-6162

steve@dipaola.org

Abstract

We will describe a visual development system for exploring face space, both in terms of facial types and animated expressions. Imagine an n-dimensional space describing every humanoid face, where each dimension represents a different facial characteristic. Within this continuous space, it would be possible to traverse a path from any face to any other face, morphing through faces along that path. It is also possible to combine elements of this space to create an expressive, emotive, talking 3D synthetic face of any given facial type.

This development toolkit called FaceSpace is based on a hierarchical parametric approach to facial animation and creation. We present our early results on exploring a face space and describe our preliminary investigation of the perceptual and cultural relationships between different facial types; as well as creating an additive language of hierarchical expressions, emotions and lip-sync sequences by using combined elements of a facial domain.

1. Introduction

The face is one of the most important surfaces we deal with. It is the showcase of the self, displaying our sex, age, and ethnic background as well as our health and mood. [1] It also is our main vehicle to transmit our feelings and explicit communications to others. We are all experts in interpreting, and communicating with the human face. We can recognize the face of someone we knew in the fifth grade, 30 years later. We can read volumes through one glance into our father's stare. Ironically, this universal expertise comes with little common understanding of what the mechanisms are that govern such talent. Most people can't even describe the features of a loved one's face accurately nor can they dissect how it is that they know when someone is lying to them; they just know.

With all of the books and research on user interface design, it is time to put together rigorous tools for understanding the communicative and expressive mechanisms of the human face – our greatest user interface. Furthermore, as 3D computer animation becomes ubiquitous in entertainment, communications and user interfaces, we need better synthetic facial tools for both authoring and understanding how to communicate with facial systems.

There are many authoring tools for facial animation, but they are based on general modeling and animation techniques (such as point morphing or boning systems) -- very few can be used in an intuitive way specific to facial creation and understanding. It is believed that a system that uses the conceptual idea of a correlated space of all faces, which can be browsed, explored and manipulated through various intuitive techniques, has many advantages. It can be used by expert and novice alike to explore both how we use and perceive human facial communication as well as to create and control synthetic faces for a myriad of uses.

2. The System

At the core of FaceSpace is a descriptive language for specifying both the features and state of a character's face. It is based on an approach used by the highly successful language PostScript™, which describes the layout of a page through an understanding of the standard elements of a document. In a similar way, FaceSpace uses knowledge of the human face, both anatomical and behavioral, to describe the type and expression in a compact representation. This description can be used both as an authoring scheme to describe the overall appearance of a character and as a real-time animation system to change the expression and move the lips.

In one feature of our FaceScript language, certain areas of the face have more resources than others because of their greater importance in face perception and expression, for instance, the mouth and eyes have many more controls, topology and computational resources associated with them in our system than say the less important forehead. The FaceScript method also allows for different low level techniques to specify or animate a facial sub-area, so we would not use the same general purpose method (say morphing or scaling) for the eyes that we would use for the mouth. The FaceScript approach lets the specifics of a given human facial feature and its inherent qualities dictate which type of low-level animation technique best suits that feature. So our system uses many different low-level techniques, each best suited for controlling and animating that part of the face. These techniques are then compartmentalized and represented by their higher-level parameter controls to maintain communication and consistency between them.

2.1. Parameter Space

The essence of FaceSpace is a set of numerical parameters, each of which controls some aspect of a character's face. Parameters are typically unitized vectors, each representing a sub-routine, which performs some low level complex transformations on the part of the face it controls. Because parameters are abstracted from their low level techniques, they have mathematically rigorous properties such as the ability to be combined, subtracted, added together, while still maintaining controllable and repeatable effects to their face model.

Parameters can be varied independently to modify specific features of the face (e.g. cheek-bone prominence, forehead height, jaw-width, etc.). This authoring paradigm is highly flexible, allowing a wide range of applications. The entire set of parameters can be exposed individually for full low level authoring control or a sub-set of these parameters with constraints can be presented to a novice user for customization and personalization. Higher-level constructs can be imposed on the basic parameter scheme by combining low-level parameters to create application-specific descriptive elements. For example a user could modify the character's appearance from "sophisticated" to "silly" with a single control that simultaneously modifies eye separation, forehead height, nose scale, etc.

Groups of high-level parameters can act on the face simultaneously, creating lip-sync speech with one channel while specifying an astonished look for the whole face on another independent channel. Because of their associative properties and their abstraction from the actual face topology, results always look natural, as noted in the example of Figure 3.

2.2. A Space of Facial Types

The encapsulation of a face's features into a defined set of parameters makes it possible to "search" for a desired character, rather than design one. By treating each possible face as a vector in a vast multi-dimensional space, it is possible to talk about similar faces as being "near-by" one another. By "navigating" through this space, a user can "find" a character that they recognize as appealing, as opposed to creating one from scratch.
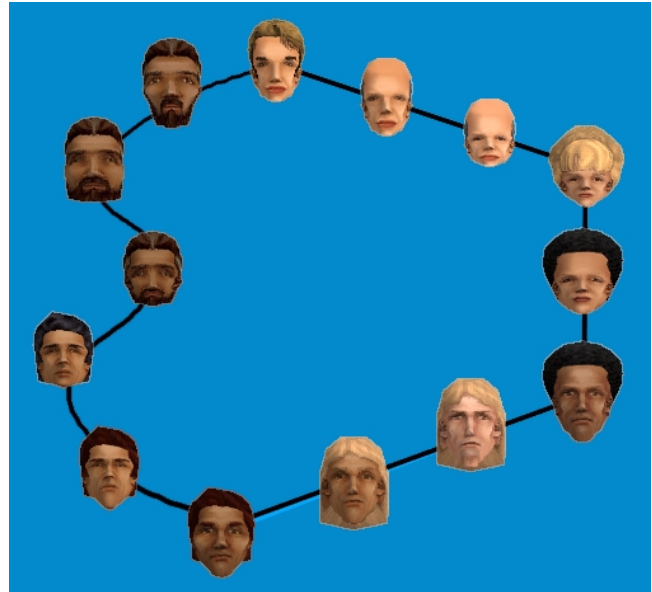
Figure 1. Fourteen distinct but related faces that were encountered through a session of navigating through "face space".

While a multi-dimensional space is a convenient way to generate a universe of faces, it has a slippery relationship to the ways we perceive facial similarity and sort faces into different categories. We currently measure facial similarity by values of low level facial parameters, such as chin width, eye radius, etc., based on our knowledge of creating facial animation systems for the graphics and avatar industries [2, 3], however perceived similarity operates on a different metric. The correlation between facial parameter values and perceived race, age, gender, or cultural stereotype is even less straightforward. It is this experimentation with classifications and their facial relationships, as well as how a viewer reacts to them, which have provided the motivation to create this development environment.

3. Preliminary Results

A prototype version of this system [4] was developed as a freely downloadable web tool, which allows users to create families of 3d heads for use in the popular simulation game "The Sims" by Maxis. FaceLift, as it was called, presents the user with a character and a set of its near-by neighbors in "face space". The user chooses one of the neighbors in some desired "direction" (i.e. more pirate-like or more grandmother-like, etc.) and then sees some of the faces surrounding their new choice. By narrowing the range of neighboring faces as the user converges on the character of choice, he or she eventually arrives at a desired character without ever directly modifying any of the low-level parameters.

This system uses a genetic algorithm technique known as aesthetic selection [4] to search through face space, using approximately 25 parameters to specify the genotype. In this prototype, a menu of random faces is generated from random points in face space. By selecting one of the faces, the user is able to surf through the local region centered around the selected face and can increase or decrease the local region of similarly looking faces by the use of a mutation rate slider and other controls. In this way, novice users can "find" or breed a face or family of related faces that appeal to them for use in the game.

Figure 2. Using a genetic algorithm technique to browse through face space, this early prototype was commissioned by Maxis for their popular game "The Sims".

A large web based community is currently using this system to make thousands of faces [5]. In addition, these users have the ability to publish and trade faces on the web and even breed their creations with those of other users. Every head or family of heads becomes a local starting point or region in face space. What we have found is that a user will have a strong emotional response to certain faces they encounter (which appears to have cultural underpinnings) and that this emotional attachment continues as they surf that locally related face space.

4. Emotion and Expression Space

While our modified prototype system restricted itself to facial types, we have expanded our original system by also using a space of expression and lip-sync parameters for exploration of facial emotion and gesture space. These additional parameters represent more transient states (i.e. jaw-drop, eye-lid closure, eyebrow lift, etc.). By modifying these in real-time, a character's face can be animated with a high degree of articulation with a minimum level of specification. Just as in the earlier example, where characteristic parameters can be combined into higher-level constructs, the more transient parameters can be combined in real-time to express a change of emotional state.

Driving the mouth parameters from an audio source to perform real-time lip-synching is an example of animating groups of transient parameters in real-time. Because of the variety of parameters to modify, the quality of this voice-driven animation is limited only by the sophistication of the voice-stream analysis. For example, eye widening can be linked to increased emphasis or head tilting to rising inflection. Since the representation of these voice responses is extremely compact (i.e. small sets of parametric changes), these can be sent over-the-wire to create a character that responds to voice-over-IP data. The overall effect of this is a simulated "video-phone", where a proxy responds to the user's voice in a way chosen by the user. Combined with the earlier facial type system, it becomes possible for a user to pick (or browse to) any face then use that face to talk with someone over the Internet. Their communication partner would see (and hear) their chosen realistic 3D face, talking and expressing itself, in real-time, controlled via their own voice and high level commands.

By hierarchically combining characteristic parameters into higher-level time-based constructs we can create facial personality meta-parameters or "behaviors" which can be applied to any facial type. Behaviors, as we refer to them, are can be built up from simple key-framed animations (smiles, winks, nods, yawns, etc.) creating full personalities that define how the character will respond to a stream of voice data or other controls. Voice analysis tools with real-time feedback are provided to create dialog animations that include lip-synching and other facial animations relating to speech. The FaceSpace system allows a user to mix animations, voice-streams and behaviors at any time, in any combination and with any character, allowing maximum flexibility and re-use.

Figure 3. Two characters with the same "fear" expression applied to each. The "fear" behavior creates the impression of fear that is recognizable, but unique to each face. Note how the asymmetries inherent in the bottom character are preserved.

4.1. Behavior Meta-Parameters

The best way to describe these high-level personality meta-parameters (refer to as "behavior" controls within FaceSpace) is with an example. Say a game or interactive film developer wanted to use FaceSpace to create an interactive Spy character for an adventure game or film. The designer would create the face and head for the Spy in FaceSpace by adjusting low level parameters until the character had just the slim gaunt face with a shifty slant to the eyes. Alternatively he could browse FaceSpace and "find" the character he wanted. The designer could then save this set of parameters as a character library entry called "Spy".

The Spy character has to deliver various pieces of dialog, some of which must be delivered nervously and others slyly. The designer could create a sly mood (with solid eye-contact and smooth, shifty gestures) and a nervous mood (with darting glances, rapid blinking and jerky motions), also saving each mood as expression library entrees. A voice talent would record the speeches and the

FaceSpace system would analyzes them for lip-syncing and inflection. The spy character can then be made to play back any of the speeches with either the sly or nervous mood, as determined by the programmed logic of the interactive film or game. The Spy's face will lip-sync to the words and respond to the inflections, using whatever mood is specified at run-time. If a new Spy character is introduced into the story line, with different speeches and voice talent, the same sly and nervous moods could be used to accent the new character's performances. Individual facial expressions (smiles, frowns, ticks, etc.) can be created, stored as libraries and overlaid on top of the speeches at run-time, under program control. For example, a player may perform some action in the middle of a Spy's speech, causing him to be displeased and frown or to be surprised and look startled. In short, the player's interaction with the Spy characters can be varied, subtle and life-like to whatever extent the designer desires.

In this way we have begun to build up a hierarchical library of behaviors, expressions and character types – which all can be combined and changed in any number of ways. These then become a large continuous domain of facial expression space. Just as we have described exploring or browsing a space of facial types, we now can begin to explore a space of facial expressions and emotions. Because this emotion space has a time-based component, and because "atoms" of emotion are less straight forward in where they begin and end (i.e., when a sly expression animates into one of embarrassment, it is not obvious where one began and the other finished), further research into the controlling structure of this emotion space is needed.

We believe that using the notion of correlated space or domain to explore facial types and their emotion language can be useful in a myriad of applications in communication, entertainment and the sciences. We therefore have begun to combine our systems into a configurable development kit, which can be quickly adapted for specific uses and investigations.

5. Future Research

By expanding this system into a full development tool, we hope to better understand the cultural relationships for both specific faces as well as their relationship to other faces in face space. Specifically, we are interested in the following areas:

- The system allows us to traverse morphing face paths through face space, with the ability to affect the trajectory of path (straight, curved) and the key points (faces) to target along the way. Music can be described as paths through tonal space with significant cultural biases described by certain paths or key points (i.e. notes, intervals, scales)

determining whether certain motifs sound appealing. Can we similarly write a composition in face space? Imagine a representation of two intertwining paths through face space that use their relationships, both harmonic and discordant, to create a composition.

- Can we create a syntax or language using face symbology that can be recognized rapidly, and span national and cultural boundaries?

- We are interested in using FaceSpace as a social science tool for understanding how different cultures perceive faces. The development kit could be used in a controlled study where participants of different countries are asked to find specific face types such as a heroic face. Data would be generated from this study that could correlate the perceived focal points of facial stereotypes from different groups.

- We have described how our system can be used as a new "more natural" type of email or instant messenger system, where a user can pick any head of his choosing, and speak through it over the Internet, communicating in a more expressive way than text based systems. We are exploring this further.

- We have also already described how FaceSpace can be used to make more realistic and more expressive 3D characters, both more efficiently and with greater emotional control. We are currently exploring this goal with the interactive entertainment industry.

- Can we create a social interface system using FaceSpace that is very human, expressive and intuitive to the user, while still maintaining full programmatic control and repeatability for the developer?

6. Conclusion

It is well known that we humans have specific neuro-programming for recognizing and interpreting faces. It is hoped that we can use the FaceSpace development system to better understand the conscious and intuitive meaning of faces and the universal language they appear to represent (facial meaning, inter-relationships and expressions) for use in the arts, cultural theory and communication.

We hope that by creating a development kit that our team or other developers can use and customize, that it is possible to both gain a greater understanding of how we communicate and beguile using our faces as well as create an authoring environment that will let us take advantage of this new understanding via an intuitive synthetic facial system.

7. References

[1] D. McNeil, *The Face*, Little Brown and Co., 1998
[2] S. DiPaola, "Extending the Range of Facial Types", *Journal of Visualization and Computer Animation*, 1991
[3] Parke & Waters, *Computer Facial Animation*, A.K. Peters, 1996
[4] K. Sims, "Artificial Evolution for Computer Graphics", *ACM SIGGRAPH*, 1991
[5] Facelift download - http://thesims.ea.com/us/, Electronic Arts , The Sims website