

A Survey of Facial Modeling and Animation Techniques

Jun-yong Noh

Integrated Media Systems Center, University of Southern California

jnoh@cs.usc.edu <http://csuri.usc.edu/~noh>

Ulrich Neumann

Integrated Media Systems Center, University of Southern California

uneumann@graphics.usc.edu <http://www.usc.edu/dept/CGIT/un.html>

Realistic facial animation is achieved through geometric and image manipulations. Geometric deformations usually account for the shape and deformations unique to the physiology and expressions of a person. Image manipulations model the reflectance properties of the facial skin and hair to achieve small-scale detail that is difficult to model by geometric manipulation alone. Modeling and animation methods often exhibit elements of each realm. This paper summarizes the theoretical approaches used in published work and describes their strengths, weaknesses, and relative performance. Taxonomy groups the methods into classes that highlight their similarities and differences.

Categories and Subject Descriptors:

General Terms:

Additional Key Words and Phrases:

Introduction

Since the pioneering work of Frederic I. Parke [91] in 1972, many research efforts have attempted to generate realistic facial modeling and animation. The most ambitious attempts perform the modeling and rendering in real time. Because of the complexity of human facial anatomy, and our natural sensitivity to facial appearance, there is no real time system that captures subtle expressions and emotions realistically on an avatar. Although some recent work [43, 103] produces realistic results with relatively fast performance, the process for generating facial animation entails extensive human intervention or tedious tuning. The ultimate goal for research in facial modeling and animation is a system that 1) creates realistic animation, 2) operates in real time, 3) is automated as much as possible, and 4) adapts easily to individual faces.

Recent interest in facial modeling and animation is spurred by the increasing appearance of virtual characters in film and video, inexpensive desktop processing power, and the potential for a new 3D immersive communication metaphor for human-computer interaction. Much of the facial modeling and animation research is published in specific venues that are relatively unknown to the general graphics community. There are few surveys or detailed historical treatments of the subject [85]. This survey is intended as an accessible reference to the range of reported facial modeling and animation techniques.

Facial modeling and animation research falls into two major categories, those based on geometric manipulations and those based on image manipulations (Fig. 1). Each realm comprises several sub-categories. Geometric manipulations include key-framing and geometric interpolations [33, 86, 91], parameterizations [21, 88, 89, 90], finite element methods [6, 44, 102], muscle based modeling [70, 96, 101, 106, 107, 110, 122, 131], visual simulation using pseudo muscles [50, 71], spline models [79, 80, 125, 126, 127] and free-form deformations [24, 50]. Image manipulations include image morphing between

photographic images [10], texture manipulations [82], image blending [103], and vascular expressions [49]. At the preprocessing stage, a person-specific individual model may be constructed using anthropometry [25], scattered data interpolation [123], or by projecting target and source meshes onto spherical or cylindrical coordinates. Such individual models are often animated by feature tracking or performance driven animation [12, 35, 84, 93, 133].

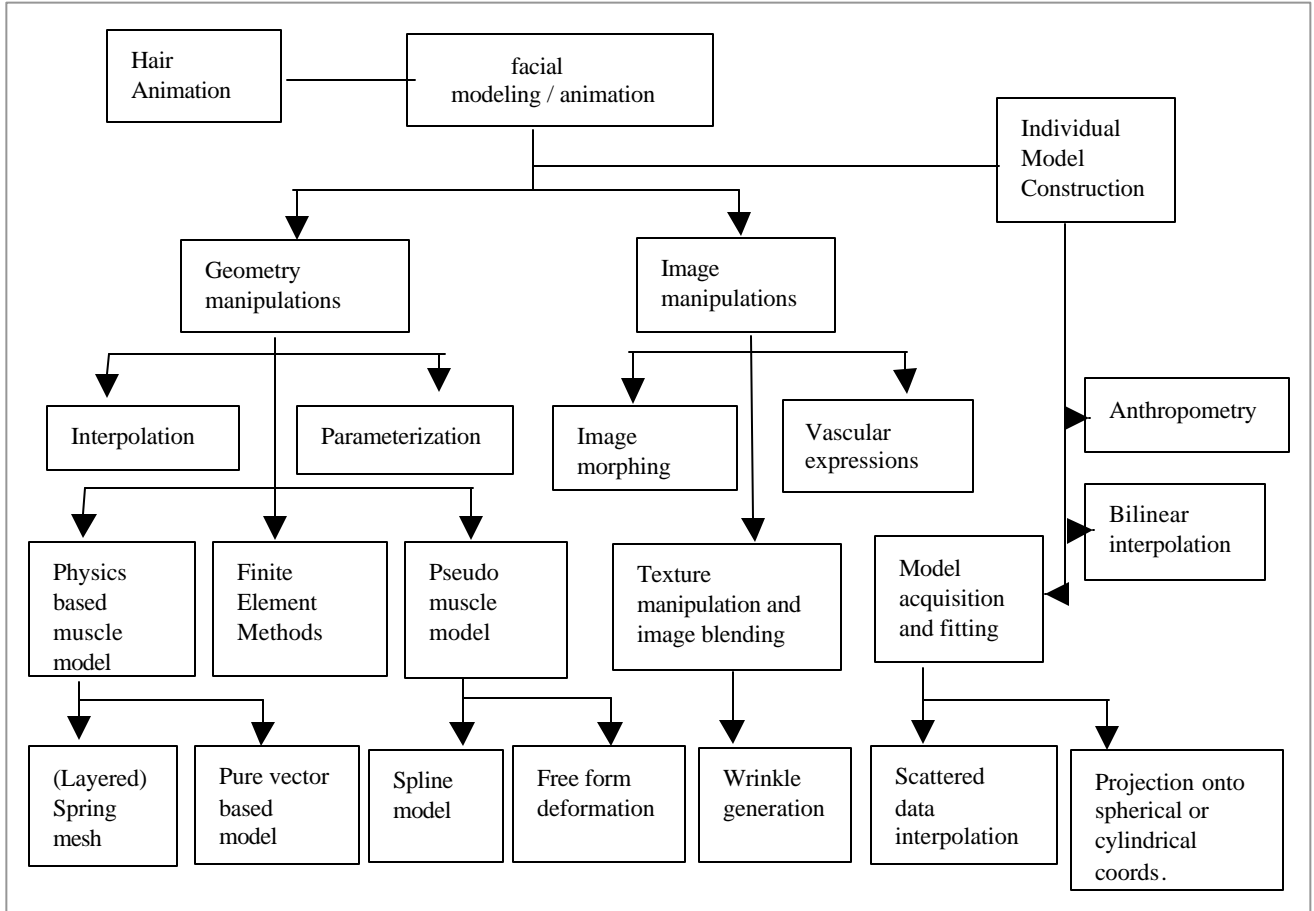


Fig. 1 - Classification of facial modeling and animation methods

This taxonomy in Figure 1 illustrates the diversity of approaches to facial animation. Exact classifications are complicated by the lack of exact boundaries between methods and the fact that recent approaches often integrate several methods to produce better results.

The survey proceeds as follows. Section 1 and 2 introduce the interpolation techniques and parameterizations followed by the animation methods using 2D and 3D morphing techniques in section 3. The Facial Action Coding System, a frequently used facial description tool, is summarized in section 4. Physics based modeling and simulated muscle modeling are discussed in sections 5 and 6, respectively. Techniques for increased realism, including wrinkle generation, vascular expression and texture manipulation, are surveyed in sections 7, 8, and 9. Individual modeling and model fitting are described in section 10, followed by animation from tracking data in section 11. Section 12 describes mouth animation research, followed by general conclusions and observations.

1. Interpolations

Interpolation techniques offer an intuitive approach to facial animation. Typically, an interpolation function specifies smooth motion between two key-frames at extreme positions, over a normalized time interval (Fig. 2).

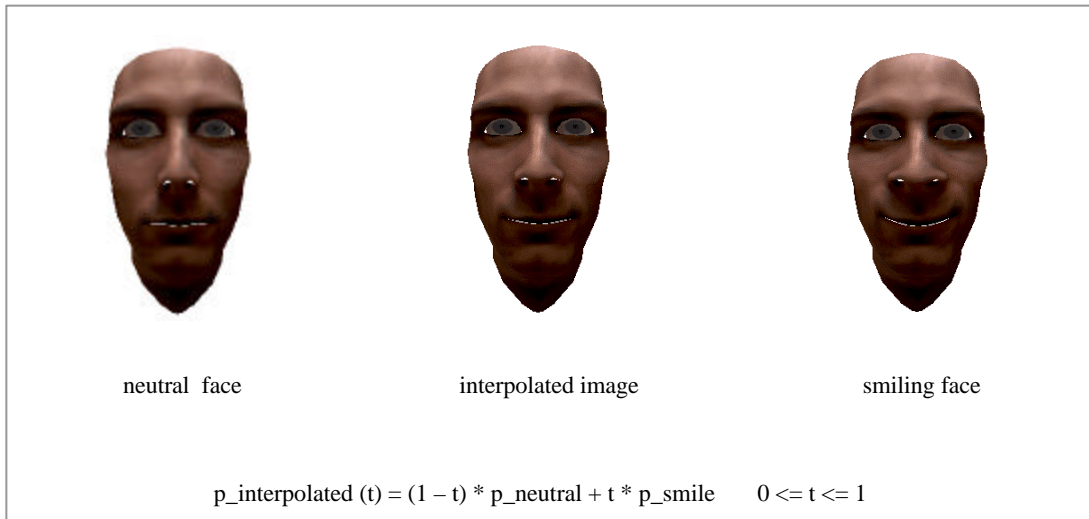


Fig. 2 - Linear interpolation is performed on muscle contraction values

Linear interpolation is commonly used [103] for simplicity, but a cosine interpolation function or other variations can provide acceleration and deceleration effects at the beginning and end of an animation [129]. When four key frames are involved, rather than two, bilinear interpolation generates a greater variety of facial expressions than linear interpolation [90]. Bilinear interpolation, when combined with simultaneous image morphing, creates a wide range of realistic facial expression changes [4].

Interpolated images are generated by varying the parameters of the interpolation functions. Geometric interpolation directly updates the 2D or 3D positions of the face mesh vertices, while parameter interpolation controls functions that indirectly move the vertices. For example, Sera et al. [115] perform a linear interpolation of the spring muscle force parameters, rather than the positions of the vertices, to achieve realistic mouth animation. Figure 2 shows two key frames and an interpolated image using linear interpolation of muscle contraction parameters.

Although interpolations are fast, and they easily generate primitive facial animations, their ability to create a wide range of realistic facial configurations is severely restricted. Combinations of independent face motions are difficult to produce. Interpolation is a good method to produce a small set of animations from a few key-frames.

2. Parameterizations

Parameterization techniques for facial animation [21, 88, 89, 90] overcome some of the limitations and restrictions of simple interpolations. Ideal parameterizations specify any possible face and expression by a combination of independent parameter values [85 pp. 188]. Unlike interpolation techniques, parameterizations allow explicit control of specific facial configurations. Combinations of parameters provide a large range of facial expressions with relatively low computational costs.

As Waters [128] indicates, there is no systematic way to arbitrate between two conflicting parameters to blend expressions that effect the same vertices, hence parameterization rarely produces natural human expressions or configurations when a conflict between parameters occurs. For this reason, parameterizations are designed to only affect specific facial regions, however this often introduces noticeable motion boundaries. Another limitation of parameterization is that the choice of the parameter

set depends on the facial mesh topology and, therefore, a complete generic parameterization is not possible. Furthermore, tedious manual tuning is required to set parameter values, and even after that, unrealistic motion or configurations may result. The limitations of parameterization led to the development of diverse techniques such as morphing between images, (pseudo) muscle based animation, and finite element methods.

3. 2D & 3D morphing ?

Morphing effects a metamorphosis between two target images or models. A 2D image morph consists of a warp¹ between corresponding points in the target images and a simultaneous cross dissolve². Typically, the correspondences are manually selected to suit the needs of the application. Morphs between carefully acquired and corresponded images produce very realistic facial animations. Beier et al. [10] demonstrated 2D morphing between two images with manually specified corresponding features (line segments). The warp function is based upon a field of influence surrounding the corresponding features. Realism, with this approach, requires extensive manual interaction for color balancing, correspondence selection, and tuning of the warp and dissolve parameters. Variations in the target image viewpoints or features complicate the selection of correspondences. Realistic head motions are difficult to synthesize since target features become occluded or revealed during the animation.

To overcome the limitations of 2D morphs, Pighin et al. [104] combine 2D morphing with 3D transformations of a geometric model. Pighin et al. animate key facial expressions with 3D geometric interpolation, while image morphing is performed between corresponding texture maps. This approach achieves viewpoint independent realism, however, animations are still limited to interpolations between pre-defined key-expressions.

The 2D and 3D morphing methods can produce realistic facial expressions, but they share similar limitations with the interpolation approaches. Selecting corresponding points in target images is manually intensive, dependent on viewpoint, and not generalizable to different faces. Also, the animation viewpoint is constrained to approximately that of the target images.

4. Facial Action Coding System

AU	FACS Name	AU	FACS Name
1	Inner Brow Raiser	12	Lid Corner Puller
2	Outer Bow Raiser	14	Dimpler
4	Brow Lower	15	Lip Corner Depressor
5	Upper Lid Raiser	16	Lower Lip Depressor
6	Check Raiser	17	Chin Raiser
7	Lid Tightener	20	Lip Stretcher
9	Nose Wrinkler	23	Lip Tightener
10	Upper Lid Raiser	26	Jaw Drop

Table 1 - Sample single facial action units

Basic Expressions	Involved Action Units
Surprise	AU1, 2, 5, 15, 16, 20, 26
Fear	AU1, 2, 4, 5, 15, 20, 26
Disgust	AU2, 4, 9, 15, 17
Anger	AU2, 4, 7, 9, 10, 20, 26
Happiness	AU1, 6, 12, 14
Sadness	AU1, 4, 15, 23

Table 2 – Example sets of action units for basic expressions

The Facial Action Coding System (FACS) is a description of the movements of the facial muscles and jaw/tongue derived from an analysis of facial anatomy [32]. FACS includes 44 basic action units (AUs). Combinations of independent action units generate facial expressions. For example, combining the AU1

¹ Basic warping maps an image onto a regular shape such as a plane or a cylinder.

² In cross dissolving, one image is faded out while another is simultaneously faded in.

(Inner brow raiser), AU4 (Brow Raiser), AU15 (Lip Corner Depressor), and AU23 (Lip Tightener) creates a sad expression. A table of the sample action units and the basic expressions generated by the actions units are presented in Tables 1 and 2.

Animation methods using muscle models or simulated (pseudo) muscles overcome the correspondence and lighting difficulties of interpolation and morphing techniques. Physical muscle modeling mathematically describes the properties and the behavior of human skin, bone, and muscle systems. In contrast, pseudo muscle models mimic the dynamics of human tissue with heuristic geometric deformations. Approaches of either type often parallel the Facial Action Coding System and Action Units developed by Ekman and Friesen [32].

5. Physics Based Muscle Modeling

Physics-based muscle models fall into three categories: mass spring systems, vector representations, and layered spring meshes. Mass-spring methods propagate muscle forces in an elastic spring mesh that models skin deformation. The vector approach deforms a facial mesh using motion fields in delineated regions of influence. A layered spring mesh extends a mass spring structure into three connected mesh layers to model anatomical facial behavior more faithfully.

5.1. Spring Mesh Muscle

The work by Platt and Badler [106] is a forerunner of the research focused on muscle modeling and the structure of the human face. Forces applied to elastic meshes through muscle arcs generate realistic facial expressions. Platt's later work [105] presents a facial model with muscles represented as collections of functional blocks in defined regions of the facial structure. Platt's model consists of 38 regional muscle blocks interconnected by a spring network. Action units are created by applying muscle forces to deform the spring network.

5.2. Vector Muscle

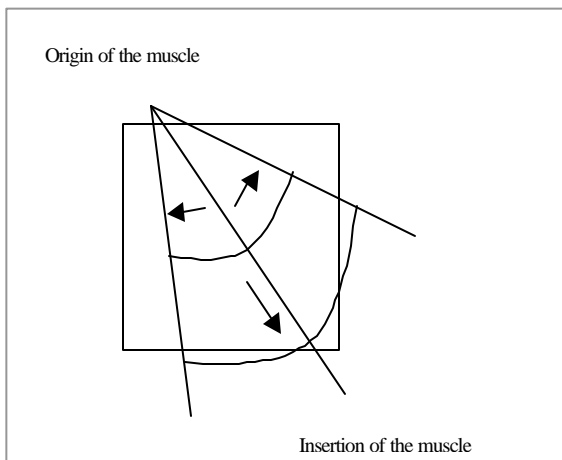


Fig. 3 - Zone of influence of Waters' linear muscle model. Deformation decreases in the directions of the arrows.

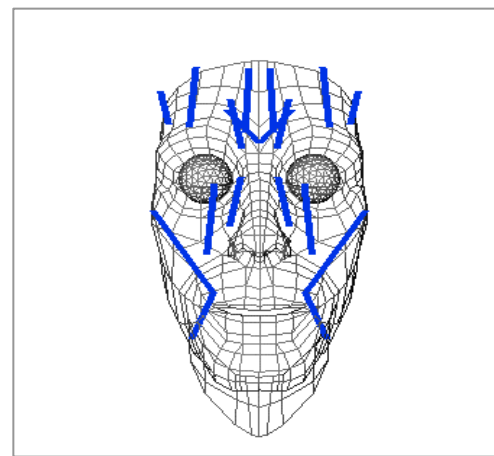


Fig. 4 - Waters' linear muscles

A very successful muscle model was proposed by Waters [131]. A delineated deformation field models the action of muscles upon skin. A muscle definition includes the vector field direction, an origin, and an insertion point (Fig. 3). The field extent is defined by cosine functions and fall off factors that produce a cone shape when visualized as a height field. Waters also models the mouth sphincter muscles as a simplified parametric ellipsoid. The sphincter muscle contracts around the center of the ellipsoid and is primarily responsible for the deformation of the mouth region. Waters animates human emotions such as

anger, fear, surprise, disgust, joy, and happiness using vector based linear and orbicularis oris muscles implementing the FACS. Figure 4 shows Waters’ muscles embedded in a facial mesh.

The positioning of vector muscles into anatomically correct positions can be a daunting task. No automatic way of placing muscles beneath a generic or person-specific mesh is reported. The process involves manual trial and error with no guarantee of efficient or optimal placement. Incorrect placement results in unnatural or undesirable animation of the mesh. Nevertheless, the vector muscle model is widely used because of its compact representation and independence of the facial mesh structure. An example of vector muscles is seen in Billy, the baby in the movie “Tin Toy”, who has 47 Waters’ muscles on his face.

5.3. Layered Spring Mesh Muscles

Terzopoulos and Waters [122] proposed a facial model that models detailed anatomical structure and dynamics of the human face. Their three-layers of deformable mesh correspond to skin, fatty tissue, and muscle tied to bone. Elastic spring elements connect each mesh node and each layer. Muscle forces propagate through the mesh systems to create animation. This model achieves great realism, however, simulating volumetric deformations with three-dimensional lattices requires extensive computation. A simplified mesh system reduces the computation time while still maintaining visual realism (Wu. et al [135]).

Lee et al. [61] presented models of physics-based synthetic skin and muscle layers based on earlier work [122]. The face model consists of three components: a biological tissue layer with nonlinear deformation properties, a muscle layer knit together under the skin, and an impenetrable skull structure beneath the muscle layer. The synthetic tissue is modeled as triangular prism elements that are divided into the epidermal surface, the fascia surface, and the skull surface (Fig. 5). Spring elements connecting the epidermal and fascia layers simulate skin elasticity. Spring elements that effect muscle forces connect the fascia and skull layers. The model achieves spectacular realism and fidelity, however tremendous computation is required, and extensive tuning is needed to model a specific face or characteristic.

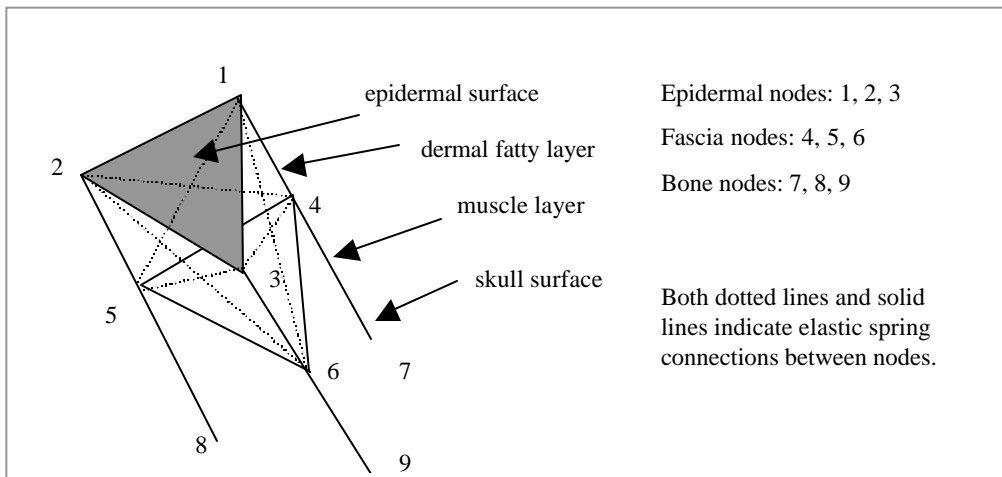


Fig. 5 - Triangular skin tissue prism element

6. Pseudo or Simulated Muscle

Physics-based muscle modeling produces realistic results by approximating human anatomy, but it is daunting to consider the exact modeling and parameters tuning needed to simulate a specific human’s facial structure. Simulated muscles offer an alternative approach by deforming the facial mesh in muscle-like fashion, but ignoring the complicated underlying anatomy. Deformation usually occurs only at the thin-

shell facial mesh. Muscle forces are simulated in the form of splines [79, 80, 125, 126, 127], wires [116], or free form deformations [24, 50].

6.1. Free form deformation

Free form deformation (FFD) deforms volumetric objects by manipulating control points arranged in a three-dimensional cubic lattice [114]. Conceptually, a flexible object is embedded in an imaginary, clear, and flexible control box containing a 3D grid of control points. As the control box is squashed, bent, or twisted into arbitrary shapes, the embedded object deforms accordingly (Fig. 6). The basis for the control points is a trivariate tensor product Bernstein polynomial. FFDs can deform many types of surface primitives, including polygons; quadric, parametric, and implicit surfaces; and solid models.

Extended free form deformation (EFFD) [24] allows the extension of the control point lattice into a cylindrical structure. A cylindrical lattice provides additional flexibility for shape deformation compared to regular cubic lattices. Rational free form deformation (RFFD) incorporates weight factors for each control point, adding another degree of freedom in specifying deformations. Hence, deformations are possible by changing the weight factors instead of changing the control point positions. When all weights are equal to one, then RFFD becomes a FFD. A main advantage of using FFD (EFFD, RFFD) to abstract deformation control from that of the actual surface description is that the transition of form is no longer dependent on the specifics of the surface itself [68 pp. 175].

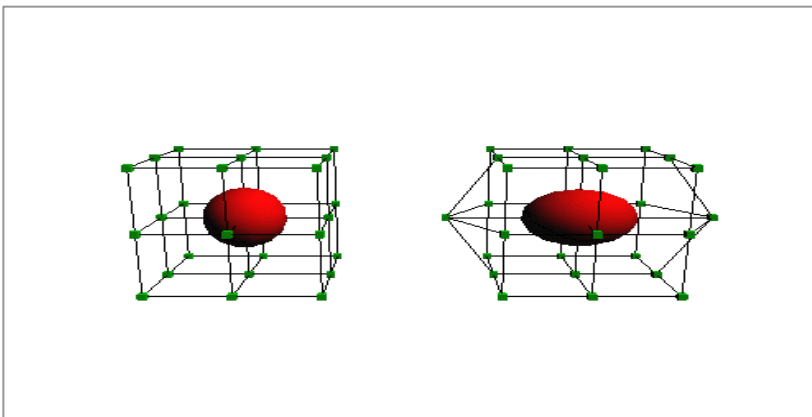


Fig. 6 - Free form deformation. Controlling box and embedded object are shown. When controlling box is deformed by manipulating control points, so is embedded object.

Kalra et al. [50] interactively simulates the visual effects of the muscles using Rational Free Form Deformation (RFFD) combined with region-based approach. To simulate the muscle action on the facial skin, surface regions corresponding to the anatomical description of the muscle actions are defined. A parallelepiped control volume is then defined on the region of interest. The skin deformations corresponding to stretching, squashing, expanding, and compressing inside the volume are simulated by interactively displacing the control points and by changing the weights associated with each control point. Linear interpolation is used to decide the deformation of the boundary points lying within the adjoining regions. Since the computation for the overall deformation is slow, larger regions are defined and stiffness factors associated with each control points are exploited to control the deformation. Displacing a control point is analogous to actuating a physically modeled muscle. Compared to Waters' physically based model [131], manipulating the positions or the weights of the control points is more intuitive and simpler than manipulating muscle vectors with delineated zone of influence. However, RFFD (FFD, EFFD) does not provide a precise simulation of the actual muscle and the skin behavior so that it fails to model furrows, bulges, and wrinkles in the skin. Furthermore, since RFFD (FFD, EFFD) is based upon surface deformation, volumetric changes occurring in the physical muscle is not accounted for.

In [50], facial animation is driven by a procedure called an Abstract Muscle Action (AMA) reported by Magnenat-Thalmann et al. [71]. These AMA procedures are similar to the action units of FACS and work on specific regions of the face. Each AMA procedure represents the behavior of a single or a group of related muscles. Facial expressions are formed by group of AMA procedures. When applied to form facial expression, the ordering of the action unit is important due to the dependency among the AMA procedures.

6.2. Spline Pseudo Muscles

Albeit polygonal models of the face are widely used, they often fail to adequately approximate the smoothness or flexibility of the human face. Fixed polygonal models do not deform smoothly in arbitrary regions, and planar vertices can not be twisted into curved surfaces without sub-division.

An ideal facial model has a surface representation that supports smooth and flexible deformations. Spline muscle models offer a solution. Splines are usually up to C2 continuous, hence a surface patch is guaranteed to be smooth, and they allow localized deformation on the surface. Furthermore, affine transformations are defined by the transformation of a small set of control points instead of all the vertices of the mesh, hence reducing the computational complexity.

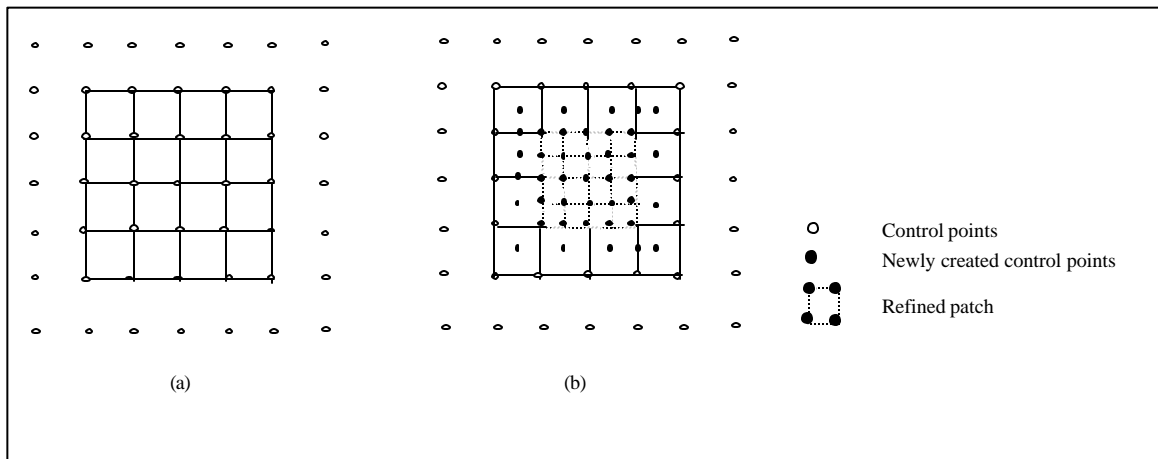


Fig. 7 - (a) shows a 16 patch surface with 49 control points and (b) shows the 4 patches in the middle refined to 16 patches. (Following the original description from [130])

Some spline-based animation can be found in [79, 80, 125]. Pixar used bicubic Catmull-Rom spline³ patches to model Billy, the baby in animation “Tin Toy”, and recently, used a variant of Catmull-Clark [19] subdivision surfaces to model Geri, a human character in short film Geri’s game. This technique is mainly adapted to model sharp creases on a surface or discontinuities between surfaces [27]. For a detailed description of Catmull-Rom splines and Catmull-Clark subdivision surfaces, refer to [20] and [19] respectively. Eisert and Girod [31] used triangular B-splines to overcome the drawback that conventional B-splines do not refine curved areas locally since they are defined on a rectangular topology.

A hierarchical spline model reduces the number of unnecessary control points. Wang et al. [127] showed a system that integrated hierarchical spline models with simulated muscles based on local surface deformations. Bicubic B-splines are used because they offer both smoothness and flexibility, which are hard to achieve with conventional polygonal models. The drawback of using naive B-splines for complex

³ A distinguishing property of Catmull-Rom splines is that the piecewise cubic polynomial segments pass through all the control points except the first and last when used for interpolation. Another is that the convex hull property is not observed in Catmull-Rom spline.

surfaces becomes clear, however, when a deformation is required to be finer than the patch resolution. To produce finer patch resolution, an entire row or column of the surface is subdivided. Thus, more detail (and control points) is added where none are needed. In contrast, hierarchical splines provide the local refinements of B-spline surfaces and new patches are only added within a specified region (Fig. 7). Hierarchical B-splines are an economical and compact way to represent a spline surface and achieve high rendering speed. Muscles coupled with hierarchical spline surfaces are capable of creating bulging skin surfaces and a variety of facial expressions.

Dubreuil et al. [29] used the animation model called DOGMA (Deformation of Geometrical Model Animated) [8] to define space deformation in terms of displacement constraints. The animation model of DOGMA [8, 9] is a four-dimensional deformation system, that is a subset of generalized n-dimensional model called DOGME [14]. A 4D deformation, where the fourth dimension is time, deforms both space and time, simultaneously. A limited set of spline muscles simulates the effects of muscle contractions without anatomic modeling.

7. Wrinkles

Wrinkles are important for realistic facial animation and modeling. They aid in recognizing facial expressions as well as a person's age. There are two types of wrinkles, temporary wrinkles that appear for a short time in expressions, and permanent wrinkles that form over time as permanent features of a face [135]. Wrinkles and creases are difficult to model with techniques such as simulated muscles or parameterization, since these methods are designed to produce smooth deformations. Physically based modeling with plasticity or viscosity, and texture techniques like bump mapping are more appropriate.

7.1. Wrinkles with Bump Mapping

Bump mapping produces perturbations of the surface normals that alter the shading of a surface. Arbitrary wrinkles can appear on a smooth geometric surface by defining wrinkle functions [13]. This technique easily generates wrinkles by varying wrinkle function parameters. Bump mapping technique is relatively computationally demanding as it requires about twice the computing effort needed for conventional color texture mapping. A bump mapped wrinkled surface is depicted in figure 8.

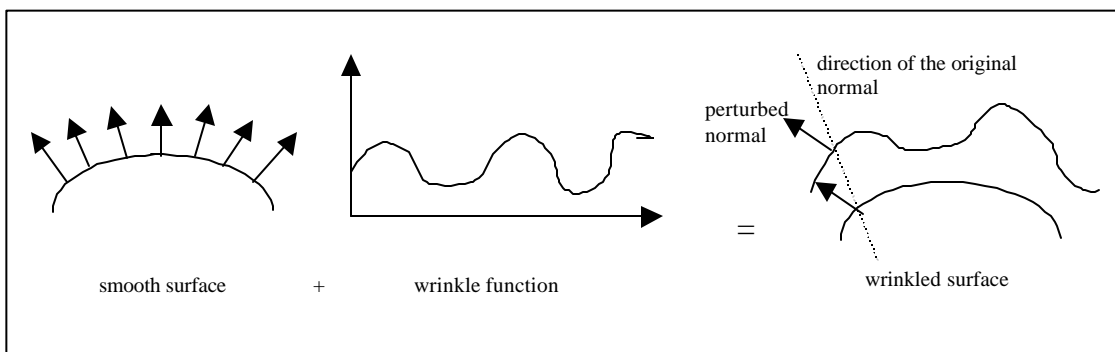


Fig. 8 - Generation of wrinkled surface using bump mapping technique.

Moubaraki et al. [77] presented a system using bump mapping to produce realistic synthetic wrinkles. The method synthesizes and animates wrinkles by morphing between wrinkled and unwrinkled textures. The texture construction starts with an intensity map (gray-level wrinkle texture) that is filtered using a simple averaging or Gaussian filter for noise-removal. Bump map gradients are extracted in orthogonal directions and used to perturb the normals of the unwrinkled texture. Correspondence is not a big issue because the wrinkled and unwrinkled images are essentially similar. Animations, as with any interpolation or morphing system, are limited to pre-defined target images. Until recently, bump mapping was also difficult to compute in real time.

7.2. Physically Based Wrinkles

Physically based wrinkle models using the plastic-visco-elastic properties of the facial skin and permanent skin aging effects are reported by Wu et al. [135]. Viscosity and plasticity are two of the canonical inelastic properties. Viscosity is responsible for time dependent deformation while plasticity is for non-invertible permanent deformation that occurs when an applied force goes beyond a threshold. Both viscosity and plasticity add to the simulations of inelasticity that moves the skin surface in smooth facial deformations. For generating immediate expressive wrinkles, the simulated skin surface deforms smoothly from muscle forces until the forces exceed the threshold; plasticity comes into play, reducing the restoring force caused by elasticity, and forming permanent wrinkles. The plasticity does not occur at all points simultaneously, rather it occurs at points that are most stressed by muscle contractions. By the repetition of this inelastic process over time, the permanent expressive wrinkles become increasingly salient on the facial model. The model is a simplified version of head anatomy, however, bones are ignored, and the muscle and fat layers are located according to the skin surface connections by simulated springs.

7.3. Other Wrinkle Approaches

Simpler inelastic models developed by Terzopoulos [123] compute only the visco-elastic property of the face. Spline segments model the bulges for the formation of wrinkles [125]. Mobaraki et al. [78] showed the animation of facial expressions using a time-varying homotopy⁴ based on the homotopy sweep technique [48]. In [78], emphasis was placed on the forehead and mouth motions accounting for the generation of wrinkles.

8. Vascular Expressions

Realistic face modeling and animation demand not only face deformation, but also skin color changes that depend on the emotional state of the person. Not much research is reported on this subject. The first notable computational model of vascular expression was reported by Kalra et al. [49] although simplistic approaches were conceived earlier [92].

Patel [92] added a skin tone effect to simulate the variation of the facial color by changing the color of all the polygons during strong emotion. Kalra et al. [49] developed a computational model of emotion that includes such visual characteristics as vascular effects and their pattern of change during the term of the emotions.

In [49], emotion is defined as a function of two parameters in time, one tied to the intensities of the muscular expressions and the other with the color variations due to vascular expressions. The elementary muscular actions in this system are based on the Minimum Perceptible Actions (MPAs) [51] similar to FACS [32] (see section 4 for details about FACS). The notion of Minimum Perceptible Color Action (MPCA) analogous to MPA is also introduced to change the color attributes due to blood circulation in the different parts of the face. Modeling the color effects directly from blood flow is complicated. Texture maps and pixel valuation offer a simpler means of approximating vascular effects. Pixel valuation computes the parameter change for each pixel inside the Bezier planar patch mask that defines the affected region of an MPCA in the texture image. This pixel parameter modifies the color attributes of the texture image. With this technique, pallor and blushing of the face are demonstrated [49].

9. Texture Manipulation

Synthetic facial images derive color from either shading or texturing. Shading computes a color value for each pixel from the surface properties and a lighting model. Because of the subtlety of human skin coloring, simple shading models do not generally produce adequate realism. Textures enable complex variations of surface properties at each pixel, thereby creating the appearance of surface detail that is absent in the surface geometry. Consequently, textures are widely used to achieve facial image realism.

Using multiple photographs, Pighin et al. [103] developed a photorealistic textured 3D facial model. Both view-dependent and view-independent texture maps exploit weight-maps to blend multiple textures.

⁴ Homotopy is the notion that forms the basis of algebraic topology. Readers should refer to [42] for more information on Homotopy theory.

Weight maps are dependent on factors such as self-occlusion, smoothness, positional certainty, and view similarity.

A view-independent fusion of multiple textures often exhibits blurring from sampling and registration errors. In contrast, a view-dependent fusion dynamically adjusts the blending weights for the current view by rendering the model repeatedly, each time with different texture maps. The drawback of view-dependent textures is their higher memory and computing requirements. In addition, the resulting images are more sensitive to lighting variation in the original texture photographs.

Oka et al. [82] demonstrates a dynamic texture mapping system for the synthesis of realistic facial expressions and their animation. When the geometry of the 3D objects or the viewpoint change, new texture mapping occurs for the optimal display. This mapping takes place in real time (30 times per second) due to the simplicity and the efficiency of the proposed algorithm. In the proposed algorithm, a mapping function from the texture plane into the output screen is approximated by a locally linear function on each of the small regions that form the texture plane altogether. Its only constraint is that the mapping function needs to be smooth enough. Realistic facial expressions and their animations are synthesized by interpolation and extrapolation among multiple 3D facial surfaces and the dynamic texture mapping onto them depending on the viewpoint and geometry changes.

10. Fitting and Model Construction

An important problem in facial animation is to model a specific person, *i.e.*, modeling the 3D geometry of an individual face. A range scanner, digitizer probe, or stereo disparity can measure three-dimensional coordinates. However, the models obtained by those processes are often poorly suited for facial animation. Information about the facial structures is missing; measurement noise produces distracting artifacts; and model vertices are poorly distributed. Also, many measurement methods produce incomplete models, lacking hair, ears, eyes, etc.

An approach to person-specific modeling is to painstakingly prepare a prototype or generic animation mesh with all the necessary structure and animation information. This generic model is fitted or deformed to a measured geometric mesh of a specific person to create a personalized animation model. The geometric fit also facilitates the transfer of texture if it is captured with the measured mesh. If the generic model has fewer polygons than the measured mesh, decimation is implicit in the fitting process.

Person-specific modeling and fitting processes use various approaches such as scattered data interpolations [103, 124], anthropometry techniques [27, 59], and projections onto the cylindrical coordinates incorporated with a positive Laplacian field function [61]. Some methods attempt an automated fitting process, but most require significant manual intervention. Figure 9 depicts the general fitting process.

10.1. Bilinear interpolation

Parke [90] uses bilinear interpolation to create various facial shapes. His assumption is that a large variety of faces can be represented from variations of a single topology. He creates ten different faces by changing the conformation parameters of a generic face model. Parke's parametric model is restricted to the ranges that the conformation parameters can provide, and tuning the parameters for a specific face is difficult.

10.2. Scattered data interpolation

Radial basis functions⁵ are capable of closely approximate or interpolate smooth hyper-surfaces [109] such as human facial shapes. Some approaches morph a generic mesh into specific shapes with scattered data interpolation techniques based on radial basis functions. The advantages of this approach are as follows. First, the morph does not require equal numbers of nodes in the target meshes since missing points are interpolated [124]. Second, mathematical support ensures that a morphed mesh approaches the target mesh, if appropriate correspondences are selected [108, 109].

⁵ It is called radial basis because only the distance from a control point is considered and hence radially symmetric.

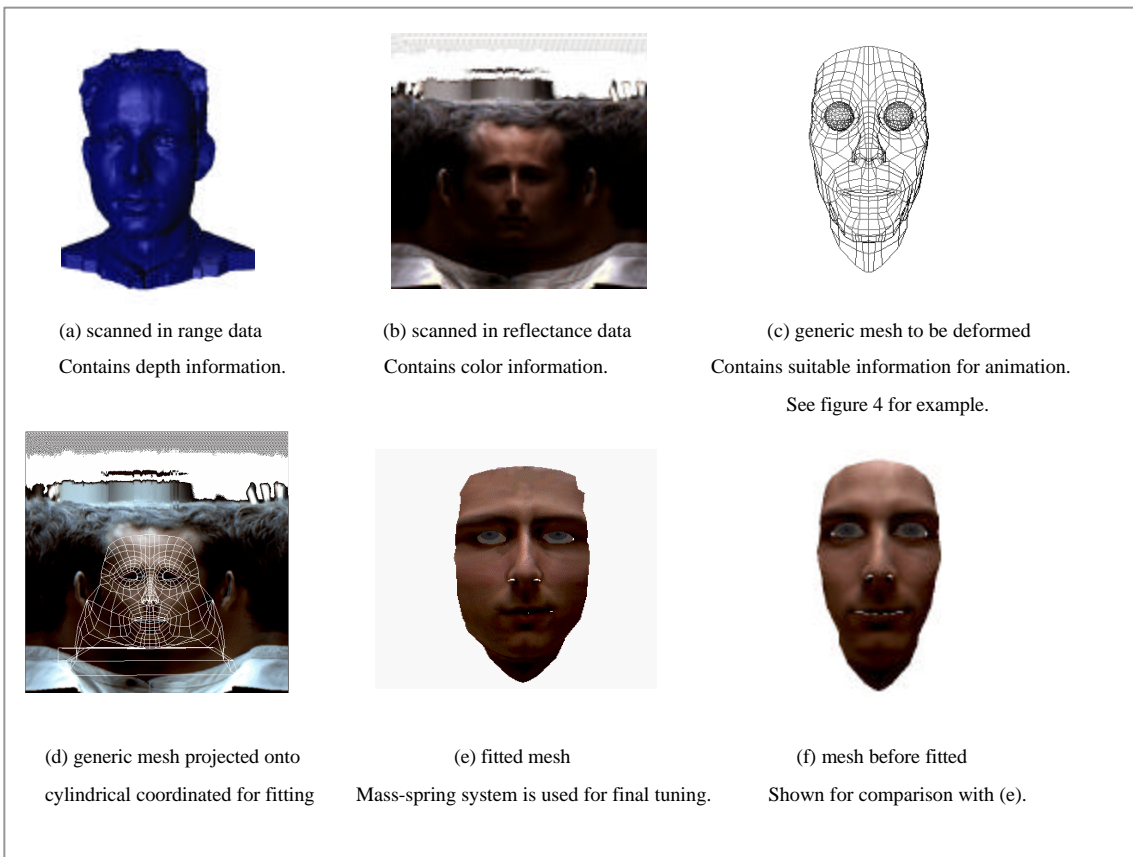


Fig. 9 - Example construction of a person specific model for animation from a generic model and laser scanned mesh.

Ulgen [124] uses 3D-volume morphing [38] to obtain a smooth transition from a generic facial model to a target model. First, biologically meaningful landmark points are selected (manually) around the eyes, nose, lips, and perimeters of both face models. Second, the landmark points define the coefficients of the Hardy multi-quadric radial basis function used to morph the volume. Finally, points in the generic mesh are interpolated using the coefficients computed from the landmark points. An example uses a generic face with 1251 polygons and a target face of 1157 polygons. Manually, 150 vertices are selected as landmark points, more than 50 around the nose. The success of the morphing depends strongly on the selection of the landmark points. Animation of the final fitted model is based on Platt's [105] muscle model facial animation system.

Pighin et al. employ a scattered data interpolation technique for a three-stage fitting process [104]. In the first stage, camera parameters (position, orientation, and focal length) are estimated. These are combined with manually selected correspondences to recover the 3D coordinates of feature points on the face. In the second stage, radial basis function coefficients are determined for the morph. In the third stage, additional correspondences facilitate fine-tuning. A sample generic mesh of under 400 polygons is morphed with 13 initial correspondence points, and 99 additional points for final tweaking.

10.3. Automatic correspondence points detection

Fitting intrinsically requires accurate correspondences between the source and target models. Incorrect or incomplete correspondences result in poor fitting. Manual correspondence selection is tedious at best, and increasingly error prone with large numbers of feature points. Several efforts at using the known properties of faces attempt to automate correspondence detection, and thereby automate the fitting process.

Yin et al [138] acquires two views of a person's face and identifies several fiducial points defined on a generic facial model. In the profile and front views, the head is segmented from the background with a threshold operation, and the profile of the head is extracted with an edge detector. The vertical positions of the fiducial points are determined by analyzing the profile curve with local maximum curvature tracking (LMCT), described in [139]. The vertical positions of fiducial points limit the correspondence search area and the computation cost. The interior fiducial points are located relative to the positions of strong edges in each search area. Finally, the extracted image fiducial points are matched with predefined fiducial points in the generic mesh.

In Yin's work [138], the generic model is modified separately by the 2D front and profile views. A 3D individual model is obtained by merging the fitting operations of each 2D view. Fiducial points in the generic model are moved to the positions of the corresponding points in the modified images, defining displacement vectors. Positions of non-fiducial points are interpolated from neighboring displacement vectors. To complete the model, the two view texture maps are blended based on the approximate orientation of localized facial regions. Animation of the complete individual model uses a Layered Force Spreading Method (LFSM). In LFSM, vertices are layered from a center to the periphery. The fiducial points constitute the center layer and the group of vertices connected to the center layer constitutes the second layer and so on. Spring forces propagate non-linearly from the center to the periphery layers, based on pre-assigned layer weights, to generate facial expression animation.

Lee et al. [61, 62] demonstrates the automatic construction of individual head models from laser-scanned range⁶ and reflectance⁷ data. To make facial features more evident for automatic detection, a modified Laplacian operator is first applied to the range map, producing a Laplacian field map. Mesh adaptation procedures on the Laplacian field map automatically identify feature points (as outlined below). The generic model with a priori labeled features is conformed to the 3D mesh geometry and texture according to a heuristic mesh adaptation procedure, summarized below.

- 1) Locate nose tip (highest range data point in the central area).
- 2) Locate chin tip (point below the nose with the greatest value of the positive Laplacian of range).
- 3) Locate mouth contour (point of the greatest positive Laplacian between the nose and the chin).
- 4) Locate chin contour (points whose latitudes lie in between the mouth and the chin).
- 5) Locate ears (points with a positive Laplacian larger than a threshold value around the longitudinal direction of the nose)
- 6) Locate eyes (points which have the greatest positive Laplacian around the estimated eyes region)
- 7) Activate spring forces to adapt facial regions of model and mesh (Located feature points are treated as fixed points.)
- 8) Adapt hair mesh (by extending the generic mesh geometrically over the rest of the range data to cover the hair)
- 9) Adapt body mesh (similar to above)
- 10) Store texture coordinates (by storing the adapted 2D nodal positions on the reflectance map)

⁶ data with depth information

⁷ data with color information

In the fitting process, contractile muscles are automatically inserted at anatomically plausible positions within a dynamic skin model and rooted to an estimated skull structure with a hinged jaw.

10.4. Anthropometry

In individual model acquisition, laser scanning and stereo images are widely used because of their abilities to acquire detailed geometry and fine textures. However, as mentioned earlier, these methods also have several drawbacks. Scanned data or stereo images often miss regions due to occlusion. Spurious data and perimeter artifacts must be touched up by hand. Existing methods for automatically finding corresponding feature points are not robust, they still require manual adjustment if the features are not salient in the measured data.

The generation of individual models using anthropometry⁸ attempts to solve many of these problems for applications where facial variations are desirable, but absolute appearance is not important. Kuo et al. [59] proposes a method to synthesize a lateral face from only one 2D gray-level image of a frontal face with no depth information. Initially, a database is constructed, containing facial parameters measured according to anthropomorphic definitions. This database serves as a priori knowledge. Secondly, the lateral facial parameters are estimated from frontal facial parameters by using minimum mean square error (MMSE) estimation rules applied to the database. Specifically, the depth of one lateral facial parameter is determined by the linear combination of several frontal facial parameters. The 3D generic facial model is then adapted according to both the frontal plane coordinates extracted from the image and their estimated depths. Finally, the lateral face is synthesized from the feature data and texture-mapped.

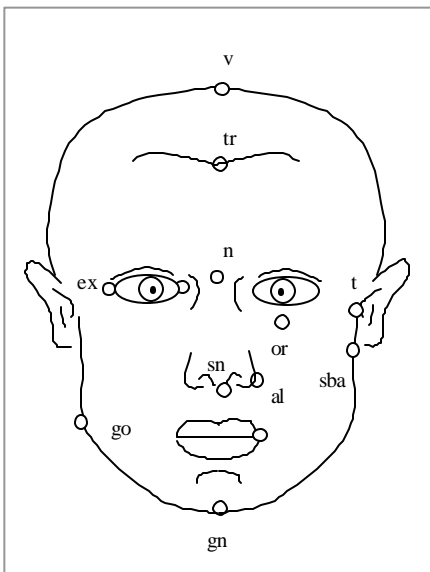


Fig. 10 - Some of the anthropometric landmarks on the face. The selected landmarks are widely used as measurements for describing the human face. (adapted from [39])

Whereas [59] uses anthropometry with one frontal image, Decarlo et al. [26] constructs various facial models purely based on anthropometry without assistance from images. This system constructs a new face model in two steps. The first step generates a random set of measurements that characterize the face. The form and values of these measurements are computed according to face anthropometry (see figure 10). The second step constructs the best surface that satisfies the geometric constraints using a variational constrained optimization technique [41, 119, 132]. In this technique, one imposes a variety of constraints

⁸ the science dedicated to the measurements of the human face

on the surface and then tries to create a smooth and fair surface while minimizing the deviation from a specified rest shape, subject to the constraints. In the case of [26], anthropometric measurements are the constraints, and the remainder of the face is determined by minimizing the deviation from the given surface objective function. Variational modeling enables the system to capture the shape similarities of faces, while allowing anthropometric differences. Although anthropometry has potential for rapidly generating plausible facial geometric variations, the approach does not model realistic variations in color, wrinkling, expressions, or hair.

10.5. Other Methods

Essa et al. [34] tackled the fitting problem using Modular Eigenspace methods⁹ [74, 99]. This method enables the automatic extraction of the positions of feature points such as the eyes, nose, and lips in the image. These features define the warping of a specific face image to match the generic face model. After warping, deformable nodes are extracted from the image for further refinement.

DiPaola's Facial Animation System (FAS) [28] is an extension of Parke's approach. New facial models are generated by digitizing live subjects or sculptures, or by manipulating existing models with free form deformations, stochastic noise deformations, or vertex editing.

Akimoto, et al. [3] uses front and profile images of a subject to automatically create 3D facial models. Additional fitting techniques are described in [58, 118, 137].

11. Animation using Tracking

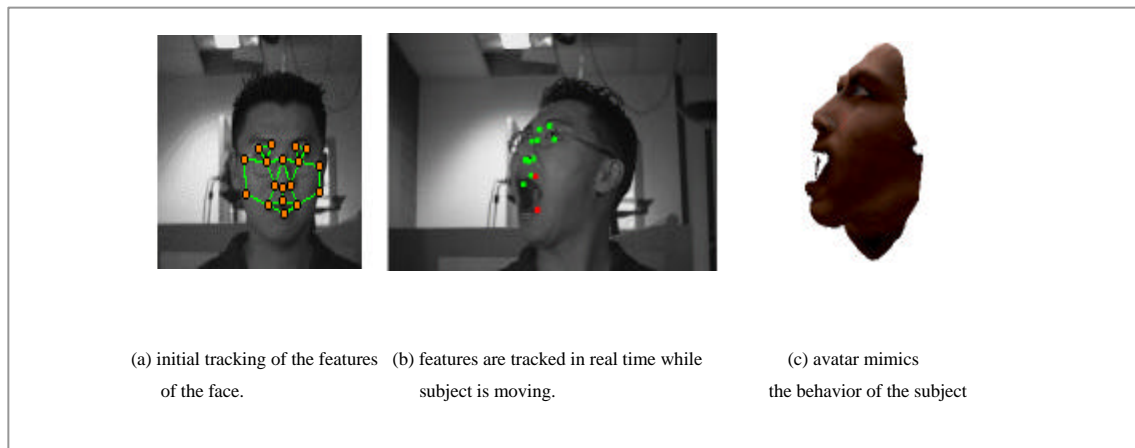


Fig. 11 - Real time tracking is performed without markups on the face using Eyematic Inc.'s face tracking system. Real time animation of the synthesized avatar is achieved based on the 11 tracked features.

The difficulties in achieving life-like character in facial animations led to the performance driven approach where tracked human actors control the animation. Real time video processing allows interactive animations where the actors observe the animations they create with their motions and expressions. Accurate tracking of feature points or edges is important to maintain a consistent and life-like quality of animation. Often the tracked 2D or 3D feature motions are filtered or transformed to generate the motion data needed for driving a specific animation system. Motion data can be used to directly generate facial animation [34] or to infer AUs of FACS in generating facial expressions. Figure 11 shows animation driven from a real time feature tracking system.

⁹ Modular Eigenspace Methods are primarily used for the recognition and detection of rigid, roughly convex objects, i.e. faces. It is modular in that it allows the incorporation of important facial features such as the eyes, nose, and mouth. The eigenspace method computes similarity from the image eigenvectors. For a detailed description of this approach, refer to [74, 99].

11. 1. Snakes and Markings

Snakes, or deformable minimum-energy curves, are widely used to track intentionally marked facial features [52]. The recognition of facial features with snakes is primarily based on color samples and edge detection. Many systems couple tracked snakes to underlying muscles mechanisms to drive facial animation [69, 120, 121, 122, 130]. Muscle contraction parameters are estimated from the tracked facial displacements in video sequences.

Tracking errors accumulate over long image sequences. Consequently, a snake may lose the contour it is attempting to track. In [84], tracking from frame to frame is done for the features that are relatively easy to track. A reliability test enables a re-initialization of a snake when error accumulations occur. Real time performance (10 frames/sec on a SGI Indy), is achieved for tracking a few (<10) features.

11.2. Optical Flow Tracking

Colored markers painted on the face or lips [18, 57, 66, 77, 81, 93, 115, 133] are extensively used to aid in tracking facial expressions or recognizing speech from video sequences. However, markings on the face are intrusive and impractical. Also, reliance on markings restricts the scope of acquired geometric information to the marked features. Optical flow¹⁰ [47] and spatio-temporal normalized correlation measurements¹¹ [25] perform natural feature tracking and therefore obviate the need for intentional markings on the face [34, 37].

Essa et al. [37] utilizes optical flow and physically based observations. The primary visual measurements of the system are sets of peak normalized correlation scores against a set of previously trained 2D templates. The normalized correlation matching [25] process allows the user to freely translate side-to-side and up-and-down, and minimizes the effects of illumination changes. The matching is also insensitive to small changes in scale or viewing distance (-15% ~ +15%) and small head rotations (-15 degrees ~ +15 degrees). For efficiency, the feature matching process is limited to a search of the neighborhood of the last observation. In the absence of a good match between the image and template expressions, interpolation based on a weighted combination of expressions is performed using the Radial Basis Function (RBF) method [108] with linear basis functions. The continuous time Kalman filter (CTKF) is incorporated to reduce noise. A 3D finite element mesh is adapted as a facial model, onto which muscles are attached based on the work of Pieper [101] and Waters [130]. In an offline process, the muscle parameters associated with each facial expression are first determined using Finite Element Methods [7] (see section 12.3. for the brief description of FEM).

Essa et al. discusses the advantage of performance driven animation using optical flow motion vectors over the systems based upon FACS [34]. The drawbacks of FACS are that 1) AUs are purely local patterns while actual facial motion is rarely completely localized and 2) FACS offers spatial motion descriptions but not temporal components. In terms of spatial motion, both video-based performance driven models and the FACS models show similar deformations in the primary activation regions. In the peripheral regions of the face, however, the performance driven model produces more deformations than FACS model. In the temporal domain, the performance driven animation utilizing motion vectors shows co-articulation effects, which are rarely observed in the FACS system. The limitation of the performance driven system is its confined animation range, limited to a set of facial motions.

¹⁰ an approximated vector representation of the displacements of group of pixels from one frame to the next frame

¹¹ Mean and variance (normalized correlation) of each pixel in the view (spatio) are computed for each frame in an image sequence (temporal).

Eisert and Girod [31] derive motion estimation and facial expression analysis from optical flow over the whole face. Since the errors of consecutive motion estimates tend to accumulate over multiple frames, a multiscale feedback loop is employed in the motion estimation process. First the motion parameters are approximated between consecutive low resolutions frames. The differences between a motion compensated frame and the current target frame is minimized. The procedure is repeated at higher resolutions, each time producing more accurate facial motion parameters. This iterative repetition at various image resolutions measures large displacement vectors (~30 pixels) between two successive video frames.

11.3. Other methods

Kato et al. [53] employ isodensity maps for the description and the synthesis of facial expressions. An isodensity map is constructed from the gray level histogram of the image based on the brightness of the region. The lightest gray level area is labeled the level-one isodensity line and the darkest is called the level-eight isodensity line. Together, these levels represent the 3D structure of the face. This method, akin to general shape-from-shading methods [46], is proposed as an alternative to feature tracking techniques.

Saji et al. [112] introduce the notion of Lighting Switch Photometry to extract 3D shapes from the moving face. The idea is to take the time sequence images illuminated in turn by separate light sources from the same viewpoint. The normal vector at each point on the surface is computed by measuring the intensity of radiance. The 3D shape of the face at a particular instant is then determined from these normal vectors. Even if the human face moves, the detailed facial shapes such as the wrinkles on a face are extracted by Lighting Switch Photometry.

Azarbayejani et al. [5] use an extended Kalman filter to recover the rigid motion parameters of a head. Saulnier et al. [113] report a template-based method for tracking and animation. Li et al. [64] use the Candid model for 3D-motion estimation for model based image coding. Masse et al. [73] use optical flow and principal direction analysis for automatic lip reading.

12. Mouth Animation

Among the regions of the face, the mouth is the most complicated in terms of its anatomical structure and its deformation behavior. Its complexity leads to considering the modeling and animation of the mouth independent from the remainder of the face. Many of the basic ideas and methods for modeling the mouth region are optimized variations of general facial animation methods. In this section, research specifically involved in the modeling and the animation of the mouth is categorized as muscle modeling with mass spring system, finite element methods, and parameterizations.

12.1. Mass Spring Muscle Systems

In mouth modeling and speech animation, mass-spring systems often model the phonetic structure of speech animation. Kelso et al. [56] qualitatively analyzes a real person's face in reiterant speech production and models it with a simple spring mass system. Browman et al. [17] showed the control of vocal-tract simulation with two mass spring systems. One spring controlled the lip aperture and the other the protrusion.

Exploiting the simplicity and generality of mass-spring systems, Waters et al. [128] develop a two-dimensional mouth muscle model and animation method. Since mouth animation is generated from relatively few muscle actions, motion realism is largely independent of the number of surface model elements. Texture mapping provides additional realism for simple geometric mouth models.

Attempts to automatically synchronize computer generated faces with synthetic speech were made by Waters et al. [129] for ASCII text input. Two different mouth animation approaches are analyzed. First, each viseme¹² mouth node is defined with positions in the topology of the mouth. Intermediate node positions between consecutive visemes are interpolated using a cosine function rather than a linear function to produce acceleration and deceleration effects at the start and end of each viseme animation. However, during fluent speech, mouth shape rarely converges to discrete viseme targets due to the continuity of

¹² a group of phonemes with similar mouth shapes when pronounced.

speech and the physical properties of the mouth. To emulate fluent speech, the calculation of co-articulated¹³ visemes is needed. The second animation method exploits Newtonian physics, Hookean elastic force, and velocity dependent damping coefficients to construct the dynamic equations of nodal displacements. The dynamic system adapts itself as the rate of speech increases, thus reducing lip displacement as it tries to accommodate each new position. This behavior is characteristics of real lip motion. A real time (15 frames / sec) animation rate was achieved using a 2D wire frame of 200 polygons representing only the frontal view on a DEC Alpha AXP 3000/500 workstation (150MHz).

12.2. Layered Spring Mesh Muscles

Sera et al. [115] add a mouth shape control mechanism to the facial skin modeled as a three-layer spring mesh with appropriately chosen elasticity coefficients (following the approach of [61]). Muscle contraction values for each phoneme are determined by the comparison of corresponding points on photos and the model (see figure 11 for muscle placements around the mouth). During speech animation, intermediate mouth shapes are defined by a linear interpolation of the muscle spring force parameters. High computation cost keeps this system from working in real time.

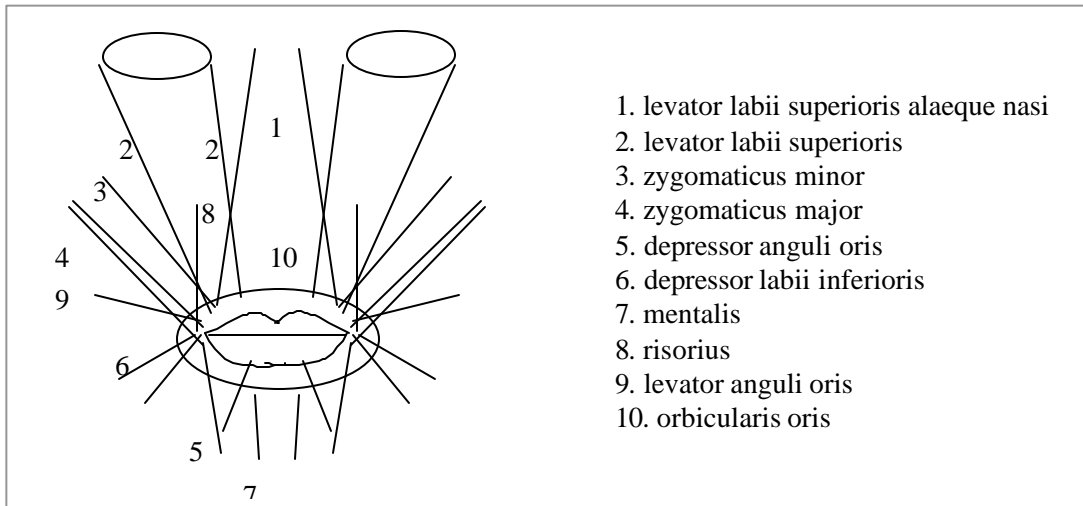


Fig. 11 - Muscle placements around the mouth. Although 5, 6, and 7 are attached to the mouth radially in reality, they are modeled linearly here.

12.3. Finite Element Method

The finite element method (FEM) is a numerical approach to approximating the physics of an arbitrary complex object [7]. It implicitly defines interpolation functions between nodes for the physical properties of the material, typically a stress-strain relationship. An object is decomposed into area or volume elements, each endowed with physical parameters. The dynamic element relationships are computed by integrating the piecewise components over the entire object.

Basu et al. [6] built a Finite Element Method (FEM) 3D model of the lips. The model parameters are determined from a training set of measured lip motions to minimize the strain felt throughout the linear elastic FEM structure. (It is worth noting that the tracking system uses normalized or chromatic color information ($r' = r / (r + g + b)$) to make it robust against variations in lighting conditions.) The goal was to extend similar ideas from 2D [15, 67] to 3D structure. The 2D models in [15, 67] suffered from

¹³ Rapid sequences of speech require that the posture for one phoneme anticipate the posture for the next phonemes. Conversely, the posture for the current phoneme is modified by the previous phonemes. This overlap between phonetic segments is referred to as co-articulation [55].

complications caused by changes in projected lip shapes from rigid rotations. By modeling the true three-dimensional structure of the lips, complex and nonlinear variations in 2D projections become simple linear parameter changes [6]. The difficult control problem associated with muscle-based approaches [36, 128] are minimized by the training stage, as are the accuracy problems that result from using only key-frames for mouth animation [128].

12.4. Parameterization

Parametric techniques for mouth animation usually require a significant number of input parameters for realistic control. Mouth animation from only two parameters is demonstrated by Moubaraki et al. [76]. The width and height of the mouth opening are the parameter pair that determine the opening angle at the corners of the mouth as well as the protrusion coefficients, derived from a radial basis function. The lip shape is obtained from a piecewise spline interpolation. For each of a set of scanned facial expression, the opening angle at the lip corner and the z-components of protrusion are measured and associated with the measured height and width of the mouth opening. This set of associations is the training set for a radial basis neural network. At run time, detected feature points from a video sequences are input to the trained network that computes the lip shape and protrusion for animation. Teeth are modeled using two texture-mapped portions of a cylinder.

12.5. Tongue modeling

In most facial animation, the tongue and its movement is omitted or oversimplified. When modeled, it is often represented as a simple parallelepiped [21, 63, 72, 87]. Although only a small portion of the tongue is visible during normal speech, the tongue shape is important for realistic synthesized mouth animation.

Stone [117] proposes a 3D model of the tongue defined as five segments in the coronal plane and five segments in the sagittal plane¹⁴. This model may deform into twisted, asymmetric, and groove shapes. This relatively accurate tongue model is carefully simplified by Pelachaud et al. [95] for speech animation.

Pelachaud et al. [95] model the tongue as a blobby object [136]. This approach assumes a pseudo skeleton comprised of geometric primitives (9 triangles) that serve as a charge distribution mechanism, creating a spatial potential field. Modifying the skeleton modifies the equi-potential surface that represents the tongue shape. The palate is modeled as a semi-sphere and the upper teeth are simulated by a planar strip. Collision detection is also proposed using implicit functions. The shape of the tongue changes to maintain volume preservation. Equi-potential surfaces are expensive to render directly, but an automatic method adaptively computes a triangular mesh during animation. The adaptive method produces triangles of sizes that are inversely proportional to the local curvature of the equi-potential surface. In addition, isotropically curved surface areas are represented by equilateral triangles and anisotropically curved surface areas produce acute triangles [83]. Lip animation as well as tongue animation was performed based on FACS [32] taking co-articulation¹⁵ into account.

12.6. Other Methods

Lip modeling by algebraic functions [45] adjusts the coefficients of a set of continuous functions to best fit the contours of 22 reference lip shapes. To predict all the algebraic equations of the various lip shape contours, five parameters are measured from real video sequences. The model even computes the contact forces during the lip interaction by virtue of a volumetric model created from an implicit surface. High-resolution realistic lip animation is successfully produced with this method.

With an image-based approach, Duchnowski et al. [30] fed raw pixel intensities into a neural net to classify lip shapes for lip reading. Adjoudani et al. [2] associated a small set of observed mouth shape parameters with a polygonal lip mesh. Petajan [100] exploited several image features in order to parameterize the lip

¹⁴ In anatomy, the coronal plane divides the body into front and back halves while the sagittal plane cuts through the center of the body dividing it into right and left halves.

¹⁵ Rapid sequences of speech require that the posture for one phoneme anticipate the posture for the next phonemes. Conversely, the posture for the current phoneme is modified by the previous phonemes. This overlap between phonetic segments is referred to as co-articulation [55].

shape. Other methods for synthesized speech and modeling of lip shapes are found in [1, 11, 16, 22, 23, 65, 75, 94, 97, 111].

Conclusion

In this paper, we describe and survey the issues associated with facial modeling and animation. We organize a wide range of approaches into categories that reflect the similarities between methods. Two major themes in facial modeling and animation are geometry manipulations and image manipulations. Balanced and coupled in various ways, variations of these themes often achieve realistic facial animations.

Generation of facial modeling and animation can be summarized as follows. First, an individual specific model is obtained using a laser scanner or stereo images and fitted into the prearranged prototype mesh by scattered data interpolation technique or by some others as discussed in section 8. Second, the constructed individual facial model is deformed to produce facial expression based on (simulated) muscles mechanism, Finite Element Method, or 2D & 3D morphing technique etc. Wrinkles and vascular effects are also considered for added realism. Third, the complete facial animation is performed by Facial Action Coding System or by tracking human actor in the video footage.

Although not mentioned in this paper, there have been approaches in facial animation and expression synthesis in the context of neural networks [54], using genetic algorithms [98] and many more. The goal of the research related to the synthesis of the face, achieving realism in real time in automated way, has not been reached yet. However, the success in each realm is recently reported.

References

- [1] C. Abry, L. J. Boe, Laws for lips, *Speech Communication*, 1986, vol. 5, pp. 97-104
- [2] A. Adjoudani, C. Benoit. On the Integration of Auditory and Visual Parameters in an HMM-based ASR. In *NATO Advanced Study Institute Speech reading by Man and Machine*, 1995
- [3] T. Akimoto, Y. Suenaga, R. Wallace, Automatic creation of 3D facial models. *IEEE computer Graphics and Application*, 1993, vol. 13(5), pp. 16-22
- [4] K. Arai, T. Kurihara, K. Anjyo, Bilinear Interpolation for Facial Expression and Metamorphosis in Real-Time Animation, *The Visual Computer*, 1996 vol. 12 pp. 105–116
- [5] A. Azarbayejani, T. Starner, B. Horowitz, A. Pentland, 'Visually Controlled Graphics, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, June 1993, vol. 15, No 6, pp. 602-605
- [6] S. Basu, N. Oliver, A. Pentland, 3D Modeling and Tracking of Human Lip Motions, *ICCV*, 1998 pp. 337-343
- [7] Klaus-Jurgen Bathe. *Finite Element Procedures in Engineering Analysis*. Prentice-Hall, 1982
- [8] D. Bechmann, N. Dubreuil, Order-controlled Free form Animation, *The Journal of Visualization and computer Animation*, 1995, vol. 6, pp. 11 - 32
- [9] D. Bechmann, N. Dubreuil, Animation Through Space and Time Based on a Space Deformation Model, *The Journal of Visualization and Computer Animation*, 1993, vol. 4, pp. 165–184
- [10] T. Beier, S. Neely, Feature-based image metamorphosis, *Computer Graphics (Siggraph proceedings 1992)*, vol. 26, pp. 35-42
- [11] C. Benoit, C. Abry, L.J. Boe, The effect of context on labiality in French, *Eurospeech*, 1991, *Proceedings*, vol. 1, pp. 153–156
- [12] P. Bergeron, P. Lachapelle, Controlling facial expressions and body movements in the computer-generated animated short "tony de peltrie". In *Siggraph, Advanced Computer Animation Seminar Notes*, July 1985
- [13] J. F. Blinn, Simulation of wrinkled surfaces, *Siggraph*, 1978 pp. 286–292
- [14] P. Borrel, D. Bechmann, Deformation of N-dimensional objects, *Symposium on Solid Modeling Foundations and CAD/CAM Application*, ACM press, Texas, 1991

- [15] C. Bregler, Stephen, M. Omohundro, Nonlinear Image Interpolation using Manifold Learning, In NIPS 7, 1995
- [16] N. Brooke and Q. Summerfield, Analysis, Synthesis, and Perception of visible articulatory movements. *Journal of Phonetics*, 1983, vol. 11, pp. 63-76
- [17] C. Browman, L. Goldstein, Dynamic modeling of phonetic structure. In V. Fromkin, editor, *Phonetic Linguistics*, 1985, pp. 35-53, Academic Press, New York
- [18] E. M. Caldognetto, K. Vaggies, N. A. Borghese, G. Ferrigno, Automatic Analysis of Lips and Jaw Kinematics in VCV Sequences, *Proceedings of Eurospeech Conference*, 1989, vol. 2, pp. 453-456
- [19] E. Catmull, J. Clark, Recursively generated b-spline surfaces on arbitrary topological meshes, *Computer Aided Design*, 1978, vol. 10(6), pp. 350-355
- [20] E. Catmull, *Subdivision Algorithm for the Display of Curved Surfaces*, Ph.D. Thesis, University of Utah, 1974
- [21] M. Cohen, D. Massara, Modeling co-articulation in synthetic visual speech. In N. Magnenat-Thalman, and D. Thalman editors, *Model and Technique in Computer Animation*, 1993, pp. 139-156, Springer-Verlag, Tokyo
- [22] M. Cohen and D. Massaro, Synthesis of visible speech. *Behavior Research Methods, Instruments & Computers*, 1990, vol. 22(2), pp. 260-263
- [23] T. Coianiz, L. Torresani, B. Caprile, 2D Deformable Models for Visual Speech Analysis. In *NATO Advanced Study Institute: Speech reading by Man and Machine*, 1995
- [24] S. Coquillart, Extended Free-Form Deformation: A Sculpturing Tool for 3D Geometric Modeling, *Computer Graphics*, 1990, vol. 24, pp. 187 - 193
- [25] T. Darrell, A. Pentland, Space-time gestures. In *Computer Vision and Pattern Recognition*, 1993
- [26] D. DeCarlo, D. Metasas and M. Stone, An Anthropometric Face Model using Variational Technique, 1998, *Siggraph proceedings*
- [27] T. Derose, M. Kass, T. Truong, Subdivision Surfaces in Character Animation, *Siggraph proceedings*, 1998, pp. 85-94
- [28] S. Dipaola, Extending the range of facial types, *The Journals of Visualization and Computer Animation*, 1991, vol 2(4), pp. 129-131
- [29] N. Dubreuil, D. Bechmann, Facial Animation, *Computer Animation*, 1996, *IEEE proceedings*, pp. 98-109
- [30] P. Duchnowski, U. Meier, and A. Waibel, See Me, Hear Me: Integrating Automatic Speech Recognition and Lip-Reading. In *Int'l Conf. On Spoken Language processing*, 1994
- [31] P. Eisert and B. Girod, Analyzing Facial Expressions for Virtual Conferencing, *IEEE, Computer Graphics and Applications*, 1998, vol. 18, no. 5, pp. 70-78
- [32] P. Ekman, W. V. Friesen, *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, CA, 1978
- [33] A. Enmett, Digital portfolio: Tony de peltrie. *Computer Graphics World*, 1985, vol. 8(10), pp. 72-77
- [34] I. A. Essa, S. Basu, T. Darrell, A. Pentland, Modeling, Tracking and Interactive Animation of Faces and Heads using Input from Video, *Proceedings of Computer Animation June 1996 Conference*, Geneva, Switzerland, *IEEE Computer Society Press*
- [35] I. A. Essa, A. Pentland, Facial Expression Recognition using a dynamic model and motion energy, *Proc. of Int. Conf. on Computer Vision*, pp. 360 - 367, CA, 1995
- [36] I. A. Essa, Analysis, Interpretation, and Synthesis of Facial Expressions. Ph.D. thesis, MIT Department of Media Arts and Sciences, 1995

- [37] I. A. Essa, T. Darrell, A. Pentland, Tracking Facial Motion, Proceedings of the IEEE Workshop on Non-rigid and Articulate Motion, Austin, Texas, November, 1994
- [38] L. Farkas, Anthropometry of the Head and Face, Raven Press, 1994
- [39] S. Fang, R. Raghavan, J. T. Richtsmeier, Volume Morphing Methods for Landmark Based 3D Image Deformation, SPIE Int. Symp. on Medical Imaging, CA, 1996
- [40] D. R. Forsey, R. H. Bartels, Hierarchical B-spline Refinement, Computer Graphics (Siggraph 1998), vol. 22(4), pp. 205 – 212
- [41] S. Gortler, M. Cohen, Hierarchical and variational geometric modeling with wavelets, Symposium on Interactive 3D Graphics, 1995, pp. 35-42
- [42] B. Gray, Homotopy Theory, An Introduction to Algebraic Topology, Academic Press, 1975, ISBN 0-12-296050-5
- [43] B. Guenter, C. Grimm, D. Wood, H. Malvar, F. Pighin, Making Faces, Siggraph proceedings, 1998, pp. 55-66
- [44] B. Guenter, A system for simulating human facial expression. In State of the Art in Computer Animation, 1992, pp. 191–202
- [45] T. Guiard-Marigny, N. Tsingos, A. Adjoudani, C. Benoit, M. P. Gascuel, 3D Models of the Lips for Realistic Speech Animation, IEEE proceedings of Computer Animation, 1996, pp. 80–89
- [46] B.K.P. Horn, and M.J. Brooks, (Eds.), Shape from Shading, Cambridge: MIT Press 1989. ISBN 0-262-08159-8.
- [47] B. K. P. Horn, B. G. Schunck, Determining optical flow. Artificial Intelligence, 1981, vol. 17, pp. 185-203
- [48] S. Kajiwara, H. Tanaka, Y. Kitamura, J. Ohya, F. Kishino, Time-Varying Homotopy and the Animation of Facial Expression for 3D Virtual Space teleconferencing, SPIE, 1993, vol. 2094/37
- [49] P. Kalra, N. Magnenat-Thalmann, Modeling of Vascular Expressions in Facial Animation, Computer Animation, 1994, pp. 50 -58
- [50] P. Kalra, A. Mangili, N. M. Thalmann, D. Thalmann, Simulation of Facial Muscle Actions Based on Rational Free From Deformations, Eurographics 1992, vol. 11(3), pp. 59–69
- [51] P. Kalra, A. Mangili, N. Magnenat-Thalmann, D. Thalmann (1991), SMILE: A multi-layered Facial Animation System, Proc. IFIP WG 5. 10, Tokyo, Japan (Ed Kunii TL) pp. 189–198
- [52] M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models. International Journal of Computer Vision, 1987, vol. 1(4), pp. 321–331
- [53] M. Kato, I. So, Y. Hishinuma, O. Nakamura, T. Minami, Description and Synthesis of Facial Expressions based on Isodensity Maps, In L. Tosiyasu (Ed.), Visual Computing, Springer-Verlag, Tokyo, 1992, pp. 39-56
- [54] F. Kawakami, M. Ohkura, H. Yamada, H. Harashima, S. Morishima, 3-D Emotion Space for Interactive Communication, Third International Computer Science Conference proceedings, ICSC, Image Analysis Applications and Computer Graphics, 1995
- [55] R. D. Kent and F. D. Minifie, Coarticulation in recent speech production models, Journal of Phonetics, 1977, vol. 5 pp. 115 - 135
- [56] J. Keslo, E. Vatikiotis-Bateson, E. Saltzman, B. Kay, A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. J. Acoust. Soc. Am, 1985, vol. 1(77), pp. 266–288
- [57] F. Kishino, Virtual Space Teleconferencing System - Real Time Detection and Reproduction of Human Images, Proc. Imagina 94, 109-118
- [58] K. Komatsu, Surface model of face for animation, Trans. IPSJ, 30. 1989

- [59] C. J. Kuo, R. S. Huang, T. G. Lin, Synthesizing Lateral Face from Frontal Facial Image Using Anthropometric Estimation, proceedings of International Conference on Image Processing, 1997, Vol. 1, pp. 133-136
- [60] T. Kurihara and K. Arai, A transformation method for modeling and animation of the human face from photographs. In State of the Art in Computer Animation, 1991, pp. 45-57
- [61] Y. C. Lee, D. Terzopoulos, K. Waters. Realistic face modeling for animation. Siggraph proceedings, 1995, pp. 55-62
- [62] Y. C. Lee, D. Terzopoulos, K. Waters, Constructing physics-based facial models of individuals. In Proceedings of Graphics Interface, 1993, pp. 1 - 8
- [63] J. P. Lewis, F. I. Parke. Automated lipsynch and speech synthesis for character animation. In Proceedings Human Factors in Computing Systems and Graphics Interface 1987, pp. 143-147
- [64] H. Li, P. Roivainen, R. Forchheimer, 3-D Motion Estimation in Model Based Facial Image Coding, IEEE Transaction on Pattern Analysis and Machine Intelligence, June 1993, vol. 15, No 6, pp. 545-555
- [65] B. Lindblom, J. Sunberg, Acoustical consequences of lip, tongue, jaw, and larynx movement. The Journal of the Acoustical Society of America, 1971, vol. 50(4), pp. 1166-1179
- [66] P. Litwinowicz, and L. Williams, Animating images with drawings. ACM Siggraph Conference Proceedings, 1994, pp. 409-412, Annual Conference Series
- [67] J. Luetin, N. Thacker, S. Beet, Visual Speech Recognition Using Active Shape Models and Hidden Markov Models. In ICASSP 96, pp. 817-820. IEEE Signal Processing Society, 1996
- [68] N. Magnenat-Thalmann, D. Thalmann Editors, Interactive Computer Animation, Prentice Hall, 1996, ISBN 0-13-518309-X
- [69] N. Magnenat-Thalmann, A. Cazedevs, D. Thalmann, Modeling Facial Communication Between an Animator and a Synthetic Actor in Real Time, Proc. Modeling in Computer Graphics, Genova, Italy, June 1993, (Eds. B. Falcidieno and L. Kunii), pp. 387-396.
- [70] N. Magnenat-Thalmann, H. Minh, M. Angelis, D. Thalmann, Design, transformation and animation of human faces. Visual Computer, 1988, vol. 5 pp. 32-39
- [71] N. Magnenat-Thalmann, N. E. Primeau, D. Thalmann, Abstract muscle actions procedures for human face animation. Visual Computer, 1988, vol. 3(5), pp. 290-297
- [72] N. Magnenat-Thalmann, D. Thalmann. The direction of synthetic actors in the film rendez-vous a montreal. IEEE Computer Graphics and Applications, 1987, pp. 9-19
- [73] K. Masse, A. Pentland, Automatic Lip reading by Computer, Trans. Inst. Elec., Info. And Comm. Eng. 1990. Vol. J73-D-II, No.6. pp.796-803
- [74] B. Moghaddam, A. Pentland, Face Recognition using View-Based and Modular Eigenspaces, In Automatic Systems for the Identification and Inspection of Humans, SPIE, 1994
- [75] S. Morishima, K. Aizawa, H. Harashima, A real-time facial action image synthesis system driven by speech and text. SPIE Visual Communications and Image Processing, 1360:1151-1157, 1990
- [76] L. Moubaraki, J. Ohya, Realistic 3D Mouth Animation Using a Minimal Number of Parameters, IEEE International Workshop on Robot and Human Communication, 1996 pp. 201-206
- [77] L. Moubaraki, J. Ohya, F. Kishino, Realistic 3D Facial Animation in Virtual Space Teleconferencing, 4th IEEE International workshop on Robot and Human Communication, 1995, pp. 253-258
- [78] L. Moubaraki, H. Tanaka, Y. Kitamura, J. Ohya, F. Kishino, Homotopy-Based 3D Animation of Facial Expression Technical Report of IEICE, IE 94-37, 1994

- [79] M. Nahas, H. Hutric, M. Rioux, and J. Domey, Facial image synthesis using skin texture recording. *Visual Computer*, 1990, vol. 6(6) pp. 337 – 343
- [80] M. Nahas, H. Huitric, and M. Saintourens, Animation of a B-spline figure, *The Visual Computer*, 1988, vol. 3(5), pp. 272-276
- [81] J. Ohya, Y. Kitamura, H. Takemura, H. Ishi, F. Kishino, N. Terashima, Virtual Space Teleconferencing: Real-Time Reproduction of 3D Human Images", *Journal of Visual Communications and Image Representation*, 1995, vol. 6, No.1, March, pp. 1-25
- [82] M. Oka, K. Tsutsui, A. Ohba, Y. Jurauchi, T. Tago, Real-time manipulation of texture-mapped surfaces. In *Siggraph 21*, 1987, pp. 181–188. *ACM Computer Graphics*
- [83] C. W. A. M. van Overveld and B. Wyvill, Potentials, polygons and penguins: An adaptive algorithm for triangulating and equi-potential surface, 1993
- [84] I. S. Pandzic, P. Kalra, N. Magnenat-Thalmann, Real time Facial Interaction, *Displays* (Butterworth-Heinemann), Vol. 15, No. 3, 1994
- [85] F. I. Parke, K. Waters, *Computer Facial Animation*, 1996, ISBN 1-56881-014-8
- [86] F. I. Parke, Techniques of facial animation, In N. Magnenat-Thalmann and D. Thalmann, editors, *New Trends in Animation and Visualization*, 1991, Chapter 16, pp. 229 – 241, John Wiley and Sons
- [87] F. I. Parke, Control parameterization for facial animation. In N. Magnenat-Thalmann, D. Thalmann, editors, *Computer Animation 1991*, pp. 3–14, Springer-Verlag
- [88] F. I. Parke, Parameterized models for facial animation revisited. In *ACM Siggraph Facial Animation Tutorial Notes*, 1989, pp. 53–56
- [89] F. I. Parke, Parameterized models for facial animation. *IEEE Computer Graphics and Applications*, 1982, vol. 2(9) pp. 61 – 68
- [90] F. I. Parke, A Parametric Model for Human Faces, Ph.D. Thesis, University of Utah, Salt Lake City, Utah, 1974, UTEC-CSc-75-047
- [91] F. I. Parke, Computer Generated Animation of Faces. *Proc. ACM annual conf.*, 1972
- [92] M. Patel (1992), *FACES*, Technical Report 92-55 (Ph.D. Thesis), University of Bath, 1992
- [93] E. C. Patterson, P. C. Litwinowicz, N. Greene, Facial Animation by Spatial Mapping, *Proc. Computer Animation 1991*, N. Magnenat-Thalmann, D. Thalmann (Eds.), Springer-Verlag, pp. 31-44
- [94] A. Pearce, G. Wyvill, D. Hill, Speech and expression: A Computer solution to face animation. *Proceedings of Graphics Interface 1986, Vision Interface 1986*, pp. 136-140
- [95] C. Pelachaud, C.W.A.M. van Overveld, C. Seah, Modeling and Animating the Human Tongue during Speech Production, *IEEE, Proceedings of Computer Animation*, 1994, pp. 40-49
- [96] C. Pelachaud, Communication and Co-articulation in Facial Animation, Ph.D. thesis, Department of Computer Science and Information Science, School of Engineering and Applied Science, U. of Penn. PA, October 1991
- [97] C. Pelachaud, N. Badler, M. Steedman, Linguistic issues in facial animation. In N. Magnenat-Thalmann, D. Thalmann, editors, *Proceedings of Computer Animation 1991*, pp. 15–29, Tokyo, Springer-Verlag
- [98] A. Peng and M. H. Hayes, Iterative Human Facial Expression Modeling, *Third International Computer Science Conference proceedings, ICSC, Image Analysis Applications and Computer Graphics*, 1995
- [99] A. Pentland, B. Moghaddam, T. Starner, View-Based and Modular Eigenspaces for Face Recognition. In *Computer Vision and Pattern Recognition Conference*, 1994, pp. 84–91. *IEEE Computer Society*, 1994

- [100] E. D. Petajan, Automatic Lipreading to Enhance Speech Recognition. In Proc. IEEE Communications Society Global Telecom. Conf., 1984
- [101] S. Pieper, J. Rosen, and D. Zeltzer, Interactive Graphics for plastic surgery: A task level analysis and implementation. Computer Graphics, Special Issue: ACM Siggraph, 1992 Symposium on Interactive 3D Graphics, pp. 127–134
- [102] S. D. Pieper, More than skin deep: Physical modeling of facial tissue. Master's thesis, MIT, 1989
- [103] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, D. H. Salesin, Synthesizing Realistic Facial Expressions from Photographs, Siggraph proceedings, 1998, pp. 75-84
- [104] F. Pighin, J. Auslander, D. Lischinski, D. H. Salesin, R. Szeliski, Realistic Facial Animation Using Image-Based 3D Morphing, 1997, Technical report UW-CSE-97-01-03
- [105] S. M. Platt, A Structural Model of the Human Face, Ph.D. Thesis, University of Pennsylvania, 1985
- [106] S. Platt, N. Badler, Animating facial expression. Computer Graphics, 1981, vol. 15(3) pp. 245-252
- [107] S. M. Platt, A system for computer simulation of the human face, Master's thesis, The Moore School, Pennsylvania, 1980
- [108] T. Poggio, F. Giros., A theory of networks for approximation and learning. Technical Report A.I. Memo No. 1140, Artificial Intelligence Lab, MIT, Cambridge, MA, July 1989
- [109] M. J. D. Powell, Radial basis functions for multivariate interpolation: a review. In J.C. Mason and M.G. Cox, editors, Algorithms for Approximation, Clarendon Press, Oxford, 1987
- [110] W. T. Reeves, Simple and Complex facial animation: Case Studies. In State of the Art in Facial Animation: Siggraph 1990 Course Notes #26, pp. 88-106. 17th International Conference on Computer Graphics and Interactive Technique, Dallas
- [111] M. Saintourens, M-H. Tramus, H. Huitric, and M. Nahas. Creation of a synthetic face speaking in real time with a synthetic voice. In Proceedings of the ETRW on Speech Synthesis, pp. 249 – 252, Grenoble, France, 1990. ESCA
- [112] H. Saji, H. Hioki, Y. Shinagawa, K. Yoshida, T. Junii , Extraction of 3D Shapes from the Moving Human Face using Lighting Switch Photometry, in N. Magnenat-Thalmann, D. Thanlmann (Ed.), Creating and Animating the Virtual World, Springer-Verlag Tokyo 1992, pp. 69–86
- [113] A. Saulnier, M. L. Viaud, D. Geldreich, Real-time facial analysis and synthesis chain. In International Workshop on Automatic Face and Gesture Recognition, 1995, pp. 86–91, Zurich, Switzerland, Editor, M. Bichsel
- [114] T. W. Sederberg, S. R. Parry, Free-Form deformation of solid geometry models, Computer Graphics (Siggraph 1996), vol. 20(4), pp. 151 - 160
- [115] H. Sera, S. Morishma, D. Terzopoulos, Physics-based Muscle Model for Mouth Shape Control, IEEE International Workshop on Robot and Human Communication, 1996, pp. 207–212
- [116] K. Singh, E. Fiume, Wires: A Geometric Deformation Technique, Siggraph proceedings, 1998, pp. 405 – 414
- [117] M. Stone, Toward a model of three-dimensional tongue movement. Journal of Phonetics, 1991, vol. 19, pp. 309-320
- [118] L. Strub, et al., Automatic facial conformation for model-based videophone coding, IEEE ICIP, 1995
- [119] D. Terzopoulos, H. Qin, Dynamic nurbs with geometric constraints for interactive sculpting, ACM Transactions on Graphics, 1994, vol. 13(2), pp. 103-136
- [120] D. Terzopoulos, R. Szeliski, Tracking with Kalman snakes, In A. Blake and A. Yuille, editors, Active Vision, 1993, pp. 3-20. MIT Press

- [121] D. Terzopoulos, K. Waters, Techniques for Realistic Facial Modeling and Animation, Proc. Computer Animation 1991, Geneva, Switzerland, Springer-Verlag, Tokyo, pp. 59–74
- [122] D. Terzopoulos and K. Waters, Physically-based facial modeling, analysis, and animation. J. of Visualization and Computer Animation, March, 1990, vol. 1(4), pp. 73-80
- [123] D. Terzopoulos, K. Fleisher (1988), Modeling Inelastic Deformation: Visco-elasticity, Plasticity, Fracture, Computer Graphics, Proc. Siggraph 1988, Vol. 22, No. 4, pp. 269–278
- [124] F. Ulgen, A step Toward universal facial animation via volume morphing, 6th IEEE International Workshop on Robot and Human communication, 1997, pp. 358-363
- [125] M. L. Viad and H. Yahia, Facial animation with wrinkles. In D. Forsey and G. Hegron, editors, Proceedings of the Third Eurographics Workshop on Animation and Simulation, 1992
- [126] C. T. Waite, The facial action control editor, FACE: A parametric facial expression editor for computer generated animation. Master's thesis, MIT, 1989, spline model with muscles
- [127] C. L. Y. Wang, D. R. Forsey, Langwidere: A New Facial Animation System, proceedings of Computer Animation, 1994, pp. 59-68
- [128] K. Waters, J. Frisbie, A Coordinated Muscle Model for Speech Animation, Graphics Interface, 1995 pp. 163 – 170
- [129] K. Waters, T. M. Levergood, Decface: An Automatic Lip-Synchronization Algorithm for Synthetic Faces, 1993, DEC. Cambridge Research Laboratory Technical Report Series
- [130] K. Waters, S. Terzopoulos, Modeling and Animating Faces using Scanned Data, Journal of Visualization and Computer Animation, 1991, Vol. 2, No. 4, pp. 123–128
- [131] K. Waters. A muscle model for animating three-dimensional facial expression. In Maureen C. Stone, editor, Computer Graphics (Siggraph proceedings, 1987) vol. 21 pp. 17-24
- [132] W. Welch, A. Witkin, Variational surface modeling. Siggraph proceedings, 1992 pp. 157-166
- [133] L. Williams, Toward Automatic Motion Control, ACM Siggraph, 1990, vol. 24 (4), pp. 235–242
- [134] G. Wolberg, Digital Image Warpings, IEEE Computer Society Press, Los Alamitos, CA, 1991
- [135] Y. Wu, N. Magnenat-Thalmann, D. Thalmann, A Plastic-Visco-Elastic Model for Wrinkles in Facial Animation and Skin Aging, Proc. 2nd Pacific Conference on Computer Graphics and Applications, Pacific Graphics, 1994
- [136] G. Wyvill, C. McPheeters, B. Wyvill, Data structure for Soft Objects. The Visual Computer, 1986, vol. 2(4), pp. 227-234
- [137] G. Xu et al., Three-dimensional Face Modeling for virtual space teleconferencing systems, Trans. IEICE, E73, 1990
- [138] L. Yin, A. Basu, MPEG4 Face Modeling Using Fiducial Points, proceedings of International Conference on Image Processing, 1997, vol. 1, pp. 109-112
- [139] L. Yin, A fast feature detection algorithm for human face contour based on local maximum curvature tracking, Technique Report, ICG, Department of Computing Science, City U of HK, 1995