Published in the *Journal of Visualization and Computer Animation*, **1**(2):73–80, 1990.

Physically-Based Facial Modeling, Analysis, and Animation

Demetri Terzopoulos*†‡ and Keith Waters†

†Schlumberger Laboratory for Computer Science, P.O. Box 200015, Austin, TX 78720

‡Department of Computer Science, University of Toronto, Toronto, Ontario, M5S 1A4

Abstract

We develop a new 3D hierarchical model of the human face. The model incorporates a physically-based approximation to facial tissue and a set of anatomically-motivated facial muscle actuators. Despite its sophistication, the model is efficient enough to produce facial animation at interactive rates on a high-end graphics workstation. A second contribution of this paper is a technique for estimating muscle contractions from video sequences of human faces performing expressive articulations. These estimates may be input as dynamic control parameters to the face model in order to produce realistic animation. Using an example, we demonstrate that our technique yields sufficiently accurate muscle contraction estimates for the model to reproduce expressions from dynamic images of faces.

Keywords: Facial Animation, Physically Based Modeling, Deformable Models, Image Analysis

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism (Animation); I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling (Curve, surface, solid, and object representations); I.6.3 [Simulation and Modeling]: Applications; I.2.10 [Artificial Intelligence] Vision and Scene Understanding.

1 Introduction

The expressive power of the face makes it an attractive but elusive target for computer graphics modelers and animators. One of the hardest challenges has been to develop computational models of the face capable of synthesizing the various nuances of facial motion quickly and convincingly. There is a prevalent need for such models, not just for the animation of synthetic human characters, but also for a variety of other applications ranging from low bandwidth teleconferencing, to plastic surgery, to criminal investigation. This paper advances the state-of-the-art of realistic facial modeling and investigates two converse problems—facial image synthesis and facial image analysis.

[0]* Fellow, Canadian Institute for Advanced Research.

1. We develop a 3D dynamic face model which can be simulated and rendered on a high-end graphics workstation to synthesize real-time animated facial images. An important innovation of our model is the combination of an anatomically-based facial muscle process with a physically-based model of human facial tissue. This combination significantly enhances the realism of the animation compared to what is achievable with reasonable effort using earlier geometric models of the face.

2. From videos of human faces, we derive time-varying control parameters suitable for animating the face model. The anatomically-based control set corresponds to the contractions of the major muscles underlying human facial expression. Estimating dynamic muscle contractions from facial image sequences poses a problem inverse to that of generating synthetic facial images. We develop a physically-based solution which employs deformable contour models to track facial features as they move nonrigidly in the image plane.

It is difficult to devise a model of the face which is at once convenient for animators to use, physically realistic, and efficient enough to run at interactive rates. We therefore tackle the complexities of the facial modeling and animation task at different levels of abstraction. These levels include representations motivated by what is known about the psychology of human facial expressions, the anatomy of facial muscle structures, the histology of facial tissues, the mechanics of deformable materials, and the geometry and kinematics of facial skeletons. Our model spans six levels of abstraction; from most to least abstract, they are:

1. *Expression.* It is often convenient for animators to think in terms of facial expressions. The face model will execute commands to synthesize any of the six canonical expressions (see Section 4) within a given time interval and with a specified degree of emphasis (The expression level could also include phonetics).

2. *Control.* A muscle control process translates expression (or phoneme) instructions into a coordinated activation of actuator groups in the facial model.

3. *Muscles.* As in real faces, muscles comprise the basic actuation mechanism of the model. Each muscle model consists of a bundle of muscle fibers. When they contract, the fibers displace their attachment points in the facial tissue (or in the jaw).

4. *Physics.* The face model incorporates a physical approximation to human facial tissue. The synthetic tissue is a layered deformable lattice of point masses connected by elastic springs. A numerical simulation computes large-scale tissue deformation by continuously propagating through the lattice the local stresses induced by activated muscle fibers.

5. *Geometry.* The geometric representation of the facial model is a non-uniform mesh of polyhedral elements whose sizes depend on the curvature of the neutral face. Muscle-induced tissue deformations distort the neutral geometry into an expressive geometry.

6. *Images.* After each time step of the numerical simulation procedure, standard rendering techniques, accelerated by graphics hardware, project the deformed facial geometry in accordance with viewpoint, light source, and skin reflectance information to synthesize a continuous stream of facial images.

The multilevel structure of the model hides from the animator as many of the complexities of the underlying representations as possible. Internal procedures handle the details in the lower level structures automatically. At the higher levels of abstraction, our face model offers the animator a natural and semantically rich set of control parameters that reflect the constraints of real faces.

1.1 Overview

Section 2 presents a perspective on our work before the background of prior investigations into facial modeling and animation. After a brief review of the structure and mechanical properties of facial tissue, Section 3 describes our tissue model and the real-time numerical simulation of its mechanical behavior. Section 4 reviews the anatomical structure of facial muscles, and it describes the muscle actuators of the model along with the process which controls muscle groups to generate recognizable expressions. Section 5 describes the automatic construction and interactive animation of the physically-based face model. Section 6 explains how we analyze video image sequences to estimate dynamic facial muscle contractions, and it demonstrates this technique using an example. Section 7 concludes the paper.

2 Background

The modeling and animation of faces has attracted much interest in computer graphics [16]. The first attempts at facial animation involved keyframing, in which two or more complete facial poses are captured and inbetween contortions are computed by interpolation [13]. The immense pose space of the human face makes this approach extremely cumbersome for full 3D facial animation. This prompted Parke and others to develop parameterized models for facial animation [14, 19, 15, 1].

Using parameterized models, animators can create facial expressions by specifying appropriate sets of parameter value sequences; for instance, by interpolating the parameter values rather than directly keyframing face shapes. The parameters of these models shape facial features, such as the mouth, by specifying lip opening height, width, and protrusion. But unless the animator specifies the shape parameters with care, the model will produce incorrect shapes, unrealistic motions, and other spurious effects.

The limitations of *ad hoc* parameterized models prompted a movement towards models whose parameters are based on the anatomy of the human face [18, 19, 26, 25, 12]. Such models can produce more natural facial motions using a relatively small number of parameters based on facial muscle structures. When anatomically-based models incorporate facial action coding schemes [4] as control procedures, it becomes relatively straightforward to synthesize a wide range of recognizable expressions.

A serious limitation of earlier face models is that they are purely geometric. Geometric modeling limits realism because it ignores the fact that the human face is an elaborate biomechanical system. In particular, earlier models approximate the skin as an infinitesimally thin surface with no underlying structure, and they generate deformations by geometrically distorting the surface [26, 9, 12]. The geometric approach makes it difficult to reproduce many of the subtleties of facial tissue deformation, such as wrinkles and furrows.

We contend that facial modeling based on anatomy and physics is in many ways superior to conventional geometric facial modeling. The present paper develops a realistic face model

which incorporates an anatomically-based arrangement of muscle models with a physically-based tissue model. In developing our tissue model, we have taken into account some of the biomedical literature on tissue mechanics [8]. These studies have led to finite element models of facial tissue for application to surgical simulation [10, 3]. As in the work of Pieper [17] (see also [27]), however, the present paper describes a tissue model constructed from deformable lattices, a class of discrete deformable models [23].

Significant effort has been devoted to facial modeling by the graphics community. By contrast, the inverse problem of extracting parameters suitable for animation from images of real faces has received little attention. Some relevant work has focussed on the subproblem of lip synchronization during continuous speech animation, but existing parameter extraction techniques remain essentially manual [2, 11, 12, 6]. As Hill *et al.* observe in reference to Water's muscle-based model: "It remains very difficult to extract the parameters from real faces..." The present paper contributes a physically-based technique for estimating a useful subset of facial muscle parameters from video sequences of faces in articulatory motion.

3 A Physically-Based Facial Tissue Model

This section proposes a physically-based model of facial tissue which transcends the approximation of skin as an infinitesimally thin geometric surface and provides a more realistic approach to modeling faces. We briefly review the structure and mechanical properties of real facial tissue, propose a deformable lattice model that approximates these properties, and describe the numerical simulations of the equations of motion governing the lattice model.

3.1 Histology and Mechanics of Facial Tissue

Human skin has a layered structure. The epidermis, a superficial layer of dead cells, is about one tenth the thickness of the dermal layer, which it protects. The layered, network-like structure of skin makes it nonhomogeneous and nonisotropic (it has lower stiffness along Langer's lines than across them) [8, 10].

The mechanical properties of skin are due mostly to the dermis. Dermal tissue is composed of collagen (72%) and elastin (4%) fibers forming a densely convoluted network in a gelatinous ground substance (20%). Under low stress, dermal tissue offers low resistance to stretch as the collagen fibers begin to uncoil in the direction of the strain, but under higher stress the fully uncoiled collagen fibers become much more resistant to stretch. This yields an approximately biphasic stress-strain curve (Fig. 1). The incompressible ground substance retards the motion of the fibers and thereby gives rise to time-dependent, viscoelastic behavior: stress relaxation at constant strain, strain creep at constant stress, and hysteresis under cyclic loading (see [24] for a discussion of graphics models inspired by the behavior of viscoelastic materials). Finally, the elastin fibers act like elastic springs which return the collagen fibers to their coiled condition under zero load.

The skin rests on a layer of subcutaneous fatty tissue that allows it to slide rather easily over fibrous fascia that cover the underlying muscle layer (see Section 4).

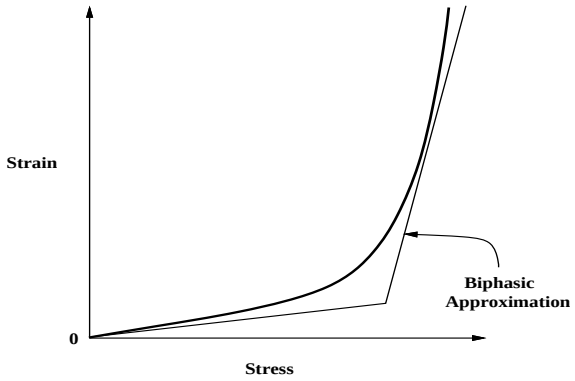


Figure 1: Stress-strain curve of facial tissue and its biphasic approximation.

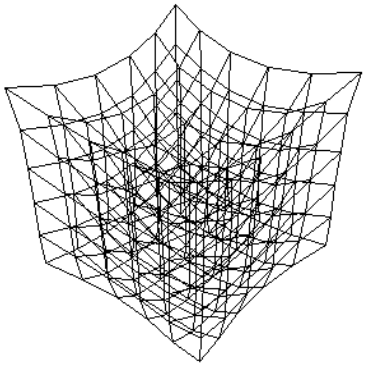


Figure 2: Deformable hexahedral lattice with corners constrained, at equilibrium under influence of gravity.

3.2 Deformable Lattices

A deformable lattice is a type of discrete deformable model constructed from point masses connected by springs [5]. Fig. 2 shows an example of a 3D deformable lattice constructed from hexahedral elements.

Let node i, where $i = 1, \ldots, N$, be a point mass m_i whose 3-space position is $\mathbf{x}_i(t) = [x(t), y(t), z(t)]'$. The velocity of the node is $\mathbf{v}_i = d\mathbf{x}_i/dt$ and its acceleration is $\mathbf{a}_i = d^2\mathbf{x}_i/dt^2$.

Let spring k have natural length l_k and stiffness c_k. Suppose the spring connects node i to node j, where $\mathbf{r}_k = \mathbf{x}_j - \mathbf{x}_i$ is the vector separation of the nodes. The actual length of the spring is $\|\mathbf{r}_k\|$. The deformation of the spring is $e_k = \|\mathbf{r}_k\| - l_k$. Then, we define the force the spring exerts on node i as

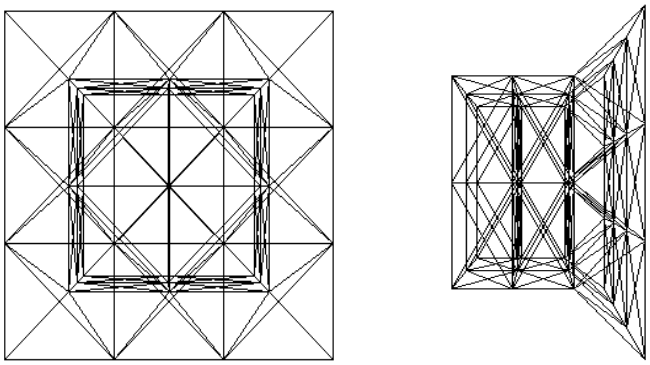
$$\mathbf{s}_k = \frac{c_k e_k}{\|\mathbf{r}_k\|}\mathbf{r}_k. \tag{1}$$

The spring force is a nonlinear function of node positions because $\|\mathbf{r}_k\|$ involves roots of sums of squares.

The total force on node i due to springs which connect it to other nodes $j \in \mathcal{N}_i$ in the deformable lattice is

$$\mathbf{g}_i(t) = \sum_{j \in \mathcal{N}_i} \mathbf{s}_k. \tag{2}$$

To create a more accurate model, we define a biphasic spring which, like real dermal tissue,



(a) (b)

Figure 3: Trilayer facial tissue model. (a) Top view. (b) Side view.

is readily extensible at low strains, but exerts rapidly increasing restoring stresses after reaching a strain e^c. The biphasic spring exerts a force (1) with stiffness

$$c_k = \begin{cases} \alpha_k & \text{when } e_k \leq e_k^c; \\ \beta_k & \text{when } e_k > e_k^c; \end{cases} \tag{3}$$

where the small-strain stiffness α_k is smaller than than the large-strain stiffness β_k (Fig. 1).

3.3 A Trilayer Model of Facial Tissue

The deformable lattice in Fig. 2 will collapse if the supports at the corners are removed because although each hexahedral element resists extension and compression forces, it does not resist twist and shear forces. However, if the faces of the elements are cross-strutted with springs, the lattice will resist twisting and shearing, yielding a structurally stable model. An advantage of tetrahedral and pentahedral elements over hexahedra is that they achieve structural stability with fewer springs. We use combinations of structurally stable elements to construct deformable lattice models of facial tissues.

Fig. 3 illustrates the geometry of a facial tissue model consisting of three layers of elements representing the cutaneous tissue, subcutaneous tissue, and muscle layer. The springs (line segments) in each layer have different stiffness parameters in accordance with the nonhomogeneity of real facial tissue. The topmost surface represents the epidermis (a rather stiff layer of keratin and collagen) and we set the spring stiffnesses so as to make it moderately resistant to deformation. The biphasic springs underneath the epidermis represent the dermis. The springs in the second layer are highly deformable, reflecting the nature of subcutaneous fatty tissue. Nodes on the bottommost surface of the second layer represent the fascia to which the muscle fibers in the third layer are attached. The bottom surface of the third layer is fixed in bone.

To account for the incompressibility of the cutaneous ground substance and the subcutaneous fatty tissues, we include a constraint into each element which minimizes the deviation of the volume of the deformed element from its natural volume at rest. The volumes are computed using simple vector algebra. Differentiation of the constraints for all elements sharing node i yields a net volume restoration force $\mathbf{q}_i$ for that node (see, e.g., [20], Section 8.7 for details about imposing volume constraints).

3.4 Numerical Simulation of Facial Tissue

The discrete Lagrange equations of motion for the dynamic node/spring system is the system of coupled, second order ordinary differential equations

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} + \gamma_i \frac{d\mathbf{x}_i}{dt} + \mathbf{g}_i + \mathbf{q}_i = \mathbf{f}_i; \qquad i = 1, \ldots, N, \tag{4}$$

where $\mathbf{g}_i$ is the net spring force from (2), $\mathbf{q}_i$ is the net volume restoration force, and $\mathbf{f}_i$ is the net driving force acting on node i. The quantity γ_i is a velocity-dependent damping coefficient which dissipates kinetic energy in the lattice through friction. We set the γ_i such that the facial tissue exhibits a slightly overdamped behavior.

To simulate the dynamics of the deformable lattice, we provide initial positions $\mathbf{x}_i^0$ and velocities $\mathbf{v}_i^0$ for each node i for $i = 1, \ldots, N$, and numerically integrate the equations of motion forward though time. At each time step, $\Delta t, 2\Delta t, \ldots, t, t + \Delta t, \ldots$, we must evaluate the forces, accelerations, velocities, and positions for each node.

A simple and quick time-integration procedure is the explicit Euler method [21]:

$$\begin{aligned} \mathbf{a}_i^t &= \frac{1}{m_i}(\mathbf{f}_i^t - \gamma_i \mathbf{v}_i^t - \mathbf{g}_i^t - \mathbf{q}_i^t); \\ \mathbf{v}_i^{t+\Delta t} &= \mathbf{v}_i^t + \Delta t\, \mathbf{a}_i^t; \\ \mathbf{x}_i^{t+\Delta t} &= \mathbf{x}_i^t + \Delta t\, \mathbf{v}_i^{t+\Delta t}. \end{aligned} \tag{5}$$

The stability of this procedure is inherently limited, but its convergence is facilitated by the overdamped dynamics of the tissue model and the high flexibility of the biphasic springs in the small-strain region (see [23] for a discussion of the effect that deformable model flexibility has on the numerical procedure). However, modest increases in the size of the time step or too large a value for the large-strain stiffness β_k will cause divergence.

It is possible to maintain convergence by using a more stable numerical procedure such as the Adams-Bashforth-Moulton method, but its computational complexity per time step erodes interactive performance. As a compromise solution to the stability/complexity tradeoff, we chose a second-order Runge-Kutta method which requires two evaluations of the forces per time step [21].

We point out that since the springs of deformable lattice are perfectly elastic, they do not model the viscoelasticity of real skin. There are several ways to include viscous behavior into the elastic spring model (see [24]). Note, however, that the explicit time integration procedure introduces an "artificial viscosity" which is analogous to placing a dashpot in parallel with each spring, so our simulated facial tissue model exhibits a somewhat viscoelastic response.

4 A Facial Muscle Control Process

4.1 Facial Muscle Anatomy

268 voluntary muscles can exert traction on the facial tissue to create expressions. When the muscles contract, they usually pull the facial soft tissue to which they *attach* towards the place where they *emerge* from the underlying skeletal framework.

Muscles are bundles of fibers working in unison. Shorter fibers are more powerful but have a smaller range of movement than longer fibers. The shape of the fiber bundle determines the muscle

type and its functionality. There are three main types of facial muscles, linear, sphincter, and sheet. Linear muscle, such as the zygomaticus major which raises the corner of the mouth, consists of a bundle of fibers that share a common emergence point in bone. Sheet muscle, such as the occipito frontalis which raises the eyebrow, is a broad, flat sheet of muscle fiber strands without a localized emergence point. Sphincter muscle consists of fibers that loop around facial orifices and can draw towards a virtual center; an example is the orbicularis oris which circles the mouth and can purse the lips into a pout.

4.2 Expressions and the Facial Action Coding System

A facial expression is the result of a confluence of muscle contractions which together deform the neutral face into an expressive face. Facial expressions are a primary form of human visual communication. Ekman has cataloged on the order of 55,000 distinguishable facial expressions, with about 30 semantic distinctions, and he has identified six primary expressions that communicate emotions—anger, disgust, fear, happiness, sadness, and surprise.

Ekman and Friesen have proposed the *Facial Action Coding System* (FACS), a quantified abstraction of the actions of facial muscles, as a means of recording facial expressions independent of cultural or personal interpretation [4]. The FACS represents facial expressions in terms of 66 *action units* (AU), which involve one or more muscles and associated activation levels. The AUs are grouped into those that affect the upper face and the lower face, and they include vertical actions, horizontal actions, oblique actions, orbital actions, and miscellaneous actions such as nostril shape, jaw drop, and head and eye position.

4.3 Muscle Actuators in the Model

The model incorporates the FACS representation implemented as part of Water's earlier geometric model [26]. Through the FACS abstraction, it is possible to suppress the low-level details of coordinated muscle actuation, and provide a more convenient interface to the model in terms of the higher-level language of expressions.

Although it is difficult to simulate the actions of all the muscles in the human face, Waters [26] and others have achieved a broad range of facial expression using on the order of 20 muscle actuators. These actuators run through the bottom layer of the trilayer tissue model. Muscles fibers emerge from nodes fixed in bone at the bottom of the layer and attach to mobile nodes on the upper surface of the layer.

Let $\mathbf{m}_i^e$ denote the point where muscle i emerges from the bone, while $\mathbf{m}_i^a$ is its point of attachment in the tissue. These two points specify a muscle vector

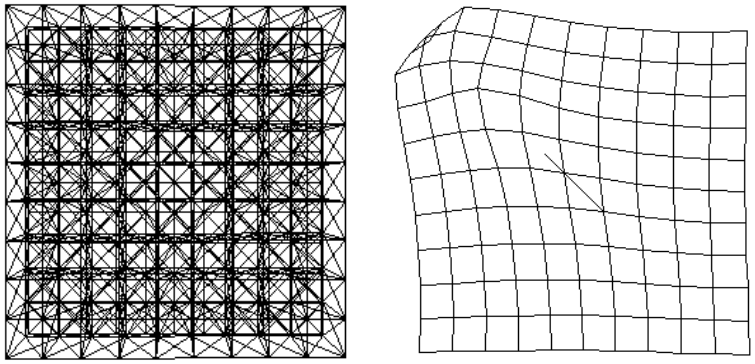
$$\mathbf{m}_i = \mathbf{m}_i^e - \mathbf{m}_i^a.$$

The displacement of node j from $\mathbf{x}_j$ to $\mathbf{x}_j'$ due to muscle contraction is a weighted sum of m muscle activities acting on node j:

$$\mathbf{x}_j' = \mathbf{x}_j + \sum_{i=1}^{m} c_i b_{ij} \mathbf{m}_i, \tag{6}$$

where $0 \leq c_i \leq 1$ is a contraction factor and b_{ij} is a muscle blend function. Defining $\mathbf{r}_{ij} = \mathbf{m}_i^a - \mathbf{x}_j$,

$$b_{ij} = \begin{cases} \cos\left(\frac{\|\mathbf{r}_{ij}\|}{a_i}\frac{\pi}{2}\right); & \text{for } \| \mathbf{r}_{ij} \| \leq a_i \\ 0; & \text{otherwise} \end{cases}, \tag{7}$$



(a) (b)

Figure 4: Tissue lattice with embedded muscle. (a) Initial geometry with muscle relaxed. (b) Equilibrium position with muscle contracted and lower layers suppressed for clarity. (1,802 springs)

where a_i is the radius of influence of the cosine blend profile.

Once all the muscle interactions have been computed, the lattice nodes $\mathbf{x}_j$ are displaced to their new positions $\mathbf{x}'_j$ as in (6). As a result, those nodes not influenced by the muscle contraction are in an unstable state and unbalanced forces propagate through the lattice to establish a new equilibrium position. Fig. 4 illustrates the stable states of the lattice before and after the contraction of a muscle.

5 Construction and Animation of the Face Model

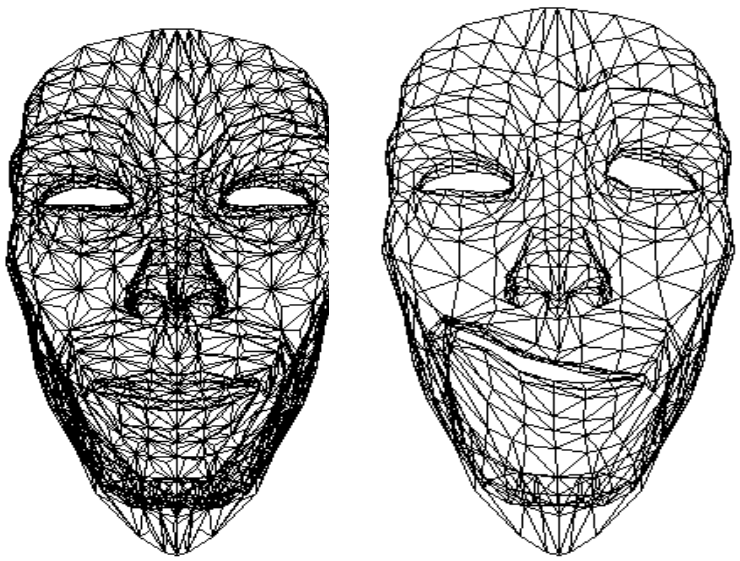
5.1 Assembling the Face Model

An automatic procedure assembles the facial model starting from a nonuniformly triangulated mesh of a face. The nodes and springs of the initial mesh represent the epidermis. Normal vectors from the center of gravity of each triangle are projected below the surface of the face to establish nodes at the subcutaneous tissue interface. Tetrahedral units are then constructed by attaching springs from the epidermis triangle nodes to these new nodes. The new springs represent the dermal layer. Short springs are then attached from the second layer to another set of nodes to form the subcutaneous layer. A final set of springs are attached from these nodes and anchored at the other end to the bone, thus establishing the muscle layer. The muscle fibers are then automatically inserted through the muscle layer from their emergence in bone to their nodes of attachment.

We have constructed a physically-based face model starting with the geometry of Water's earlier model [26]. Fig. 5 shows the skin topology after lattice construction. A total of 960 polygons are used in the tissue model, which results in approximately 6,500 spring units. The figure illustrates the tissue distortion after the influence of the zygomaticus major muscle (AU12), which raises the corner of the lip, and the inner frontalis muscle (AU1) which raises the inner portion of the brow.

5.2 A Real-Time Animation Example

The physically-based facial model is efficient enough to simulate and render at screen refresh rates greater than 8Hz on a single cpu of a Silicon Graphics Iris 4D-240GTX workstation. Fig. 6 illus-



(a) (b)

Figure 5: 3D face model. (a) Undeformed geometry of the skin layer. (b) Deformed geometry under the influence of AU1 and AU12 (only epidermis is displayed for clarity).

trates a variety of stills from an interactive session. The parameters of this facial model simulation using equations (4) and (6) were $\Delta t = 0.02$, $m_i = 0.02$, $\gamma_i = 0.5$, and $c_k = 0.49$. The user interacts with the model through a menu driven interface which allows the specification of jaw rotation and individual muscle contractions. The simulation runs continuously and rendering occurs every 5 time steps, which turns out to be suffient time for the tissue model to reach dynamic equilibrium in response to muscle activation. The epidermis of the face is Gouraud shaded and the eyes and teeth are positioned and rendered as in [26].

Fig. 6(a) and Fig. 6(b) show the undeformed facial tissue model, while Fig. 6(c) shows the jaw rotated. Figs. 6(d) and 6(e) demonstrate asymmetric contractions on the zygomaticus major, frontalis and the labii superioris muscles. Fig. 6(f) illustrates an expression of happiness. Figs. 6(g) and 6(h) show the contraction of the lateral corrugator and the anguli depressors, and Fig. 6(i) illustrates an expression of anger.

Note the wrinkles in the skin that appear around the frowning mouth, the nasolabial furrows, and the bulging of the inner portion of the brow at the root of the nose. It is nontrivial to produce such effects with a purely geometric model, but in our physically-based model they emerge automatically through the simulation of the facial tissue, primarily due to the volume preservation constraints.

6 Analysis of Dynamic Facial Images

In this section we consider the analysis of images of real faces in motion; i.e., the inverse problem to facial image synthesis. Our goal is to extract directly from video images a set of dynamic muscle parameters that may be used to animate our physically-based model. This is a difficult problem because it requires the reliable estimation of quantitative information about extended facial features, such as the eyebrows and mouth, which typically undergo nonrigid motion in the image plane.

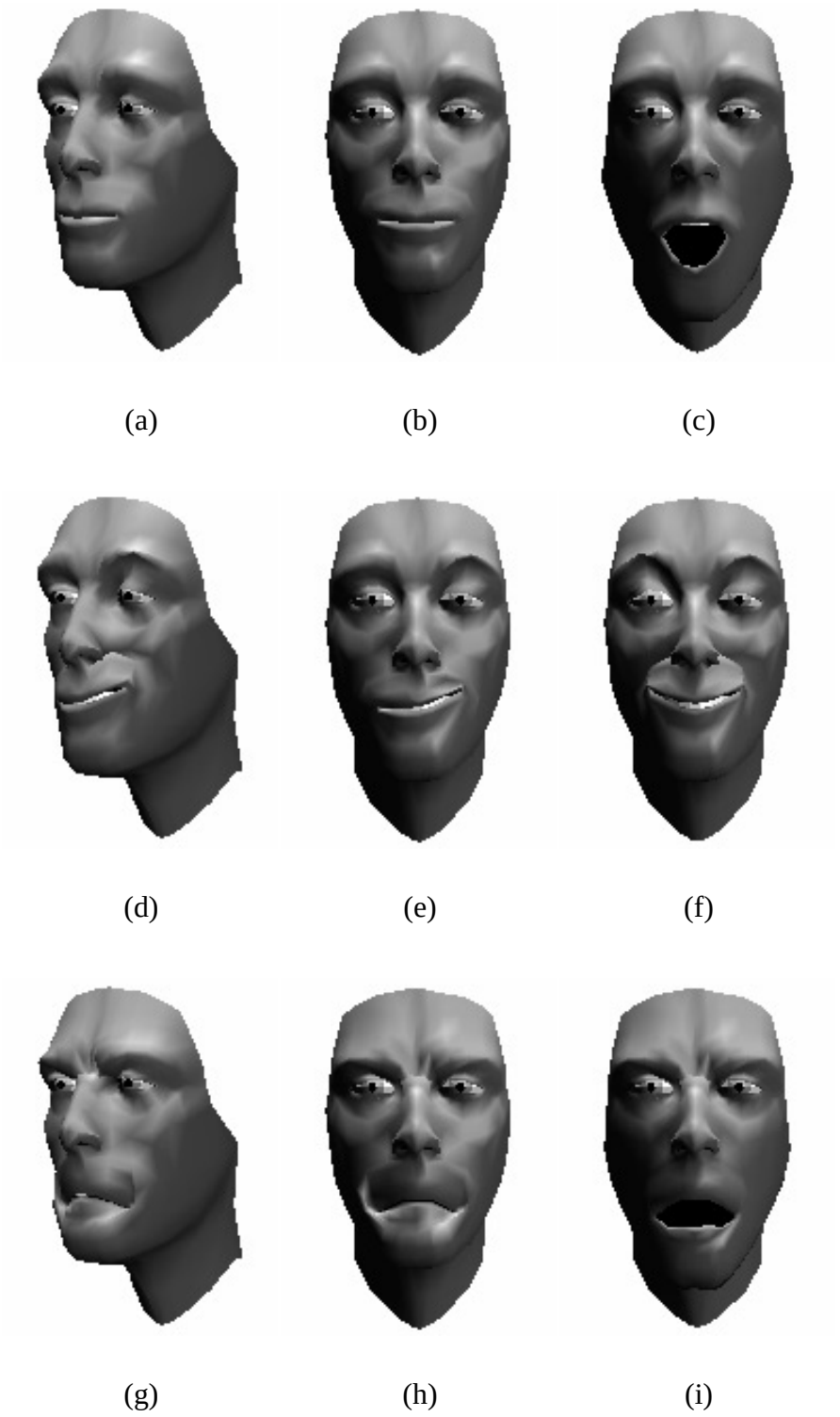


Figure 6: Rendered frames from interactive physically-based facial animation. (a) Relaxed face; oblique view. (b) Relaxed face; frontal view. (c) Jaw rotated. (d) Asymmetric contraction of the zygomaticus major, frontalis, and labii superioris; oblique view. (e) Frontal view of (d). (f) Symmetric contraction yielding an expression of happiness. (g) Symmetric contraction of the anguli depressors, lateral corrugators, and labii superioris; oblique view. (h) frontal view of (g). (i) Symmetric contraction of the anguli depressors, lateral corrugators, labii superioris, and jaw rotation yielding an expression of anger.

6.1 Approach

We develop a method for capturing the essential characteristics of dynamic facial expressions by making use of some recently developed physically-based vision techniques. Briefly, our approach is as follows: Through straightforward image processing, we convert digitized image frames into 2D potential functions. The ravines (extended local minima) of these potentials correspond to salient facial features such as the eyebrows, mouth, and chin. We employ a discrete variant of deformable contours (a.k.a. *snakes*) first proposed in [7, 22]. The deformable contours lock onto the ravines, thereby tracking the facial features from frame to frame. Our method estimates dynamic muscle parameters for the physically-based face model by automatically interpreting the state variables of the deformable contours in successive image frames.

6.2 Deformable Contours

A deformable contour can be thought of as an energy minimizing spline in the x-y image plane. The present application calls for deformable contours that have some rigidity, can stretch arbitrarily, but will resist shrinking beyond a prespecified amount. We define a discrete deformable contour as a set of nodes indexed by $i = 1, \ldots, n$. We associate with these nodes time varying positions $\mathbf{x}_i(t) = [x_i(t), y_i(t)]'$, along with "tension" forces $\boldsymbol{\alpha}_i(t)$, "rigidity" forces $\boldsymbol{\beta}_i(t)$, and external forces $\mathbf{f}_i(t)$. Note that all the forces have two components and act in the image plane.

We connect the nodes in series using nonlinear springs. Following the formulation of (1), let l_i be the given rest length of the spring connecting node i to node $i+1$ and let $\mathbf{r}_i(t) = \mathbf{x}_{i+1} - \mathbf{x}_i$ be the separation of the nodes. We want the spring to resist compression only when its actual length $\|\mathbf{r}_i\|$ is less than l_i. Hence, given the deformation $e_i(t) = \|\mathbf{r}_i\| - l_i$ we define the tension force

$$\boldsymbol{\alpha}_i = \frac{a_i e_i}{\|\mathbf{r}_i\|}\mathbf{r}_i - \frac{a_{i-1} e_{i-1}}{\|\mathbf{r}_{i-1}\|}\mathbf{r}_{i-1}, \tag{8}$$

where the tension variable $a_i(t)$ is

$$a_i = \begin{cases} a & \text{if } e_i < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

The rigidity of the continuous contour in [7, 22] stems from second order variational splines. The associated rigidity force may be expressed in terms of the fourth-order parametric derivative of the position function of the contour. Introducing rigidity variables b_i, a discrete version of the rigidity force is given by

$$\boldsymbol{\beta}_i = b_{i+1}(\mathbf{x}_{i+2} - 2\mathbf{x}_{i+1} + \mathbf{x}_i) - 2b_i(\mathbf{x}_{i+1} - 2\mathbf{x}_i + \mathbf{x}_{i-1}) + b_{i-1}(\mathbf{x}_i - 2\mathbf{x}_{i-1} + \mathbf{x}_{i-2}). \tag{10}$$

To create an interactive deformable contour, we simulate the first-order dynamic system

$$\gamma\frac{d\mathbf{x}_i}{dt} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_i = \mathbf{f}_i; \qquad i = 1, \ldots, n, \tag{11}$$

where γ is a velocity-dependent damping coefficient, and $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ control the local tension and rigidity of the contour. Tension and rigidity are locally adjustable though the a_i and b_i variables. In particular, we want to be able to break a long deformable contour to create several shorter contours on an image. Setting $a_i = b_i = 0$ permits a position discontinuity to occur at between nodes i and $i+1$ (note that setting only $b_i = 0$ permits a tangent discontinuity to occur between these nodes).

The deformable contour is responsive to a force field, derived from the image, which influences its shape and motion (see below). It is convenient to express the force field though a time-varying potential function $P(x, y, t)$. A user may also interact with the deformable contour by applying forces $\mathbf{f}_i^u(t)$ using a mouse (see [7] for details about user forces). Combining the two types of forces, we have

$$\mathbf{f}_i = p\nabla P(\mathbf{x}_i) + \mathbf{f}_i^u, \tag{12}$$

where p is the strength of the image forces and $\nabla = [\partial/\partial x, \partial/\partial y]'$ is the gradient operator in the image plane.

To simulate the deformable contour we integrate the system of ordinary differential equations (11) forward through time using a semi-implicit Euler procedure [21]. We obtain the procedure by replacing the time derivative in (11) with the forward finite difference approximation $d\mathbf{x}_i/dt \approx (\mathbf{x}_i^{t+\Delta t} - \mathbf{x}_i^t)/\Delta t$. Observing that the rigidity forces $\boldsymbol{\beta}_i$ are linear in the $\mathbf{x}_i$ variables, and moving the nonlinear tension force $\boldsymbol{\alpha}_i$ and external forces $\mathbf{f}_i$ to the right hand side yields the pentadiagonal system of algebraic equations

$$\frac{\gamma}{\Delta t}\mathbf{x}_i^{t+\Delta t} + \boldsymbol{\beta}_i^{t+\Delta t} = \frac{\gamma}{\Delta t}\mathbf{x}_i^t - \boldsymbol{\alpha}_i^t + \mathbf{f}_i^t, \tag{13}$$

which gives the node positions at the next time instant $\mathbf{x}_i^{t+\Delta t}$ in terms of the positions at the current time $\mathbf{x}_i^t$. This pentadiagonal system is solved efficiently using a direct LDU factorization method. Since the system has a constant matrix of coefficients, we factorize it only once at the beginning of the deformable contour simulation and then efficiently resolve with different right hand sides at each time step (see [22] for details).

6.3 Facial Image Processing

To apply deformable contours to facial image analysis, we first transform the image intensity function $I(x, y, t)$ at time t into a planar force field using simple image processing techniques. In the present application, we are concerned with extended image features such as the eyebrow and lip boundaries. Usually there is a prominent transition or edge in image intensity along these boundaries.

We want intensity edges to attract the deformable contours. To do this, we create a 2D potential function $P(x, y, t)$ whose ravines correspond with intensity edges by simply computing the magnitude of the gradient of the image intensity

$$P(x, y, t) = -\left\|\nabla G_\sigma * I(x, y, t)\right\|,$$

where $G_\sigma *$ denotes convolution with a 2D Gaussian smoothing filter of width σ. The smoothing filter broadens the ravines so that they attract the contours from a distance.

6.4 Facial Feature Tracking

In a few simulation time steps the deformable contours slide downhill in $P(x, y, t)$ (for fixed t) and come to equilibrium at the bottoms of the nearest ravines. As they do so, the contours deform freely and they eventually conform to the shapes of the intensity edges, thereby tracing the facial features of interest.

As soon as the contours have settled into ravines associated with a particular image frame, we replace it with the next frame in the video sequence. Starting from their previous equilibrium positions, the contours slide downhill into the perturbed ravines, so long as the motion of the facial features of interest is small enough to retain the contours on the slopes of the perturbed ravines along most of their lengths. If part of a contour escapes the attractive zone of a ravine, the rest of the contour can pull it back into place.

We repeat the procedure on subsequent frames. The evolving ravines capture and convey the contours, and the contours conform freely to them. Thus the contours track the nonrigid motions of extended image features. As the deformable contours evolve in successive frames, their dynamic state variables $\mathbf{x}_i^t$ provide quantitative information about the nonrigid shapes and motions of the facial features. Using an example, we will explain our procedure for automatically estimating from these variables a set of facial muscle contractions through time.

6.5 A Facial Image Analysis Example

We have applied our approach to facial image analysis to a sample image sequence. One of the authors (DT) subjected himself to a humiliating makeup job in order to enhance the contrast of his facial features (lips, eyebrows, and nasolabial furrows). He then performed several facial expressions in frontal view before a video camera. A surprise expression was digitized as a sequence of 256x256x8-bit images and submitted to analysis using deformable contours.

Fig. 7(b) shows the (negative) potential function computed from the frame in Fig. 7(a) which occurs near the start of the surprise sequence. To compute the potential, we apply a discrete smoothing filter $G(i,j)$ which consists of two applications of 4-neighbor local averaging of the pixel intensities, followed by the application of the discrete approximation to the gradient operator: $\nabla v(i,j) = [(v(i+1,j) - v(i,j)), (v(i,j+1) - v(i,j))]'$. The outputs of the filters on the discrete array (i,j) are interpolated bilinearly between pixels to obtain the continuous potential $P(x,y,t)$.

Using the mouse, the user draws initial deformable contours along the hairline, the left and right eyebrows, the left and right nasolabial furrows, the tip of the nose, the upper and lower lips, and the chin boss. The initialization procedure places the deformable contour nodes roughly 1 pixel apart and sets the rest lengths l_i in (9) to the initial node separations. The parameter values of the deformable contour simulation are $\gamma/\Delta t = 0.5$, $a_i = 1.0$ and $b_i = 0.5$ (except at the jump discontinuities between the contours, where $a_i = b_i = 0.0$), and $p = 0.001$. The figures show the (red) deformable contours at equilibrium locked onto the facial features.

From the position information stored in the state variable arrays of the deformable contours, an automatic procedure estimates:

1. Contractions of the left and right zygomaticus major from the positions of the endpoints of the upper lip contour;

2. Contraction of the left and right levator labii superioris alaeque nasi from the positions of the uppermost points of the associated nasolabial furrow contours;

3. Contractions of the left and right inner, major, and outer occipitofrontalis, respectively, from the positions of the innermost, center, and outermost points of the associated eyebrow contours;

4. Jaw rotation from the average position of the chin boss contour;

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 7: Analysis of facial images using deformable contours. (a) Image frame near start of surprise sequence with (red) deformable contours tracking facial features. (b) Potential function (magnitude of intensity gradient of smoothed image (a)) with deformable contours. (c) Rendered facial model with muscle contractions as estimated from deformable contours in (a). (d)–(f) Same as above for frame near middle of surprise sequence. (g)–(i) Same as above for a later frame.

5. Head reference frame from the average position of the hairline contour.

The positions of all facial feature points are computed relative to the head reference frame which, assuming a relatively stable hairline, will move with the head in the image. The first frame in the sequence shows DT's face in a relaxed state. From this frame, the analysis procedure estimates the rest lengths of his facial muscles in the head reference frame and calibrates them against the rest lengths of the muscles in the face model. Fig. 7(c) shows the rendered equilibrium position of the face model with the calibrated muscle rest lengths as inputs.

The estimated muscle lengths from successive frames are scaled by the calibration factors and input to the physically-based model in sequence, the model quickly attains dynamic equilibrium on each frame, and the state variables are rendered in real-time to produce an animated sequence. Fig. 7(d)–(f) illustrates the deformable contours and the rendered face model for a frame near the middle of the surprise sequence, and Fig. 7(g)–(i) show a later frame. The results demonstrate that our technique can robustly estimate muscle contractions from images of a real face and use them to animate the expression convincingly in a 3D face model with a different facial geometry.

7 Conclusion

This paper has presented a new 3D model of the human face which incorporates a physically-based deformable model of facial tissue and a set of anatomically-motivated facial muscle actuators. The tissue model, a lattice of nonlinear springs connecting point masses, is simulated by numerically integrating a coupled system of Lagrangian equations of motion. Coordinated facial muscle contractions induce local stresses which propagate through the facial tissue lattice deforming it to produce meaningful expressions. Realistic effects such as wrinkling around an articulating mouth result from the action of volume preservation constraints in the tissue lattice. Despite its sophistication, the face model runs efficiently and can produce animation at interactive rates on a high-end graphics workstation.

A second contribution of the paper is a physically-based technique for estimating dynamic muscle contractions for the face model from video sequences of articulating human faces. We demonstrated that the estimated contractions are robust enough to allow the model to reconstruct a transient surprise expression convincingly. Our work to date suggests that it may be possible to run our hierarchical face model backwards from the image level all the way up to the expression level, thereby solving a longstanding vision problem—the analysis and recognition of human expressions from dynamic images. This possibility remains open for future investigation.

Acknowledgements

We thank Tony Crossley for his assistance with image digitization and video equipment, and Reid Smith for supporting our research in facial modeling at SLCS.

References

[1] P. Bergeron, P. Lachapelle, "Techniques for animating characters," "SIGGRAPH'85 Tutorial: Advanced Computer Graphics Animation."

[2] N.M. Brooke and Q. Summerfield, "Analysis, synthesis, and perception of visible articulatory movements," *Journal of Phonetics*, 11, 1983, 63–76.

[3] X. Deng, "A Finite Element Analysis of Surgery of the Human Facial Tissues," PhD Dissertation, University of Columbia, 1988.

[4] P. Ekman and W.V. Friesen, *Manual for the facial action coding system*, Consulting Psychologists Press, Palo Alto, 1977.

[5] D. Greenspan, *Discrete Models*, Addision Wesley, Reading, MA, 1973.

[6] D. Hill, A. Pearce, and B. Wyvill, "Animating speech: A automated approach using speech synthesised by rules," *The Visual Computer*, 3, 1988, 277–287.

[7] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, **1**, 4, 1987, 321–331.

[8] R.M. Kenedi, T. Gibson, J.H. Evans, and J.C. Barbenel, "Tissue mechanics," Phys. Med. Biol., 20(5), 1975, 699–717.

[9] K. Komatsu, "Human skin model capable of natural shape variation," *The Visual Computer*, 3, 1988, 265–271.

[10] W. Larrabee, "A finite element model of skin deformation. I Biomechanics of skin and soft tissue: A review," *Laryngoscope* 96, 1986, 399–419. "II An experimental model of skin deformation," *Laryngoscope* 96, 1986, 406–412. "III The finite element model," *Laryngoscope* 96, 1986, 413–419.

[11] J.P. Lewis and F.I. Parke, "Automated lipsynch and speech synthesis for character animation," Proc. Human Factors in Computing Systems and Graphics Interface'87, Toronto, 1987, 143–147.

[12] N. Magnenat-Thalmann, E. Primeau, and D. Thalmann, "Abstract muscle action procedures for face animation," *The Visual Computer*, 3, 1988, 290–297.

[13] F.I. Parke, "Computer generated animation of faces," MS Thesis, Technical Report UTEC-CSc-72-120, Department of Computer Science, University of Utah, Salt Lake City, Utah, 1972.

[14] F.I. Parke, "A parametric model for human faces," PhD Thesis, Technical Report UTEC-CSc-75-047, Department of Computer Science, University of Utah, Salt Lake City, Utah, 1974.

[15] F.I. Parke, "Parameterized models for facial animation," *IEEE Computer Graphics and Applications*, 2(9), November, 1982, 61–68.

[16] F.I. Parke, "SIGGRAPH'89 Course Notes Vol 22: State of the art in facial animation," July, 1989.

[17] S. Pieper. "More than Skin Deep: Physical Modeling of Facial Tissue," MSc thesis, Massachusetts Institute of Technology, January 1989.

[18] S.M. Platt, "A system for computer simulation of the human face," MSc thesis, The Moore School, University of Pennsylvania, August, 1980.

[19] S.M. Platt and N.I. Badler, "Animating facial expressions," *Computer Graphics*, 15(3), August, 1981, 245–252.

[20] J. Platt, "Constraint methods for neural networks and computer graphics," PhD Thesis, Dept. of Computer Science, California Institute of Technology, Pasadena, CA (Caltech-CS-TR-89-07).

[21] W. Press, B. Flanney, S. Teukolsky and W. Verttering, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986.

[22] D. Terzopoulos, "On matching deformable models to images," *Topical Meeting on Machine Vision*, Technical Digest Series, Vol. 12 (Optical Society of America, Washington, DC) 1987, 160–163. Also Technical Report No. 60, Schlumberger Palo Alto Research, Palo Alto, CA, November, 1986.

[23] D. Terzopoulos and K. Fleischer. Deformable models. *The Visual Computer*, 4(6):306–331, December, 1988.

[24] D. Terzopoulos and K. Fleischer. Modeling inelastic deformation: Viscoelasticity, plasticity, fracture. *Computer Graphics*, 22(4):269–278, August, 1988.

[25] C. Waite, "The facial action control editor, FACE: A parametric facial expression editor for computer generated animation," MSc thesis, Massachusetts Institute of Technology, February 1989.

[26] K. Waters. A Muscle Model for Animating Three-Dimensional Facial Expression. *Computer Graphics*, 22(4):17–24, July, 1987.

[27] K. Waters and D. Terzopoulos "A physical model of facial tissue and muscle articulation," Proc. First Conf. on Visualization in Biomedical Computing (VBC'90), Atlanta, GA, May, 1990, 77–82.