Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models

Demetri Terzopoulos, *Member, IEEE,* and Keith Waters, *Member, IEEE*

Abstract— **We present a new approach to the analysis of dynamic facial images for the purposes of estimating and resynthesizing dynamic facial expressions. The approach exploits a sophisticated generative model of the human face originally developed for realistic facial animation. The face model, which may be simulated and rendered at interactive rates on a graphics workstation, incorporates a physics-based synthetic facial tissue and a set of anatomically motivated facial muscle actuators. We consider the estimation of dynamic facial muscle contractions from video sequences of expressive human faces. We develop an estimation technique that uses deformable contour models (snakes) to track the nonrigid motions of facial features in video images. The technique estimates muscle actuator controls with sufficient accuracy to permit the face model to resynthesize transient expressions.**

Index Terms—**Computer graphics, computer vision, deformable models, face modeling, facial image analysis, facial image synthesis, nonrigid motion analysis, physics-based modeling, snakes, tracking.**

I. Introduction

THE COMPLEXITY and expressiveness of the human face makes it a challenging subject for automated visual interpretation and recognition. Quick, robust facial image analysis is desirable for numerous applications. Among them is low-bandwidth teleconferencing, which may involve the real-time extraction of facial control parameters from live video at the transmission site and the reconstruction of a dynamic facsimile of the subject's face at a remote receiver. Teleconferencing and other applications require facial models that are not only computationally efficient but also realistic enough to accurately synthesize the various nuances of facial structure and motion. In this paper, we will show the following:

1) We present a 3-D dynamic model of the face that can be simulated in real time on graphics workstations. Our face model combines a physics-based model of facial tissue with an anatomically based facial muscle control process to synthesize realistic facial motions (Fig. 1). We enhance the apparent realism by employing geometric and photometric information acquired by scanning subjects with active sensors.

Manuscript received October 10, 1991; revised December 1, 1992. This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Information Technology Research Center of Ontario. Recommended for acceptance by T. Huang and P. Stucki.

D. Terzopoulos is with the Department of Computer Science, University of Toronto, Toronto, Canada M5S 1A4.

K. Waters is with Digital Equipment Corporation, Cambridge Research Laboratory, Cambridge, MA 02139.

IEEE Log Number 9209281.

Fig. 1. Images synthesized by the face model.

2) We develop a technique for analyzing video sequences of faces undergoing transient expressions. The goal is to estimate the dynamic muscle control parameters of the face model in order to reconstruct expressions. Our estimation technique employs interactive deformable contours (snakes) to track the nonrigid motions of extended facial features in video images.

Section II reviews prior research and motivates our approach. Section III presents the face model. The presentation includes a brief review of the histology and mechanical properties of facial tissue and the anatomical structure of facial muscles, a description of the synthetic tissue model and its real-time numerical simulation, a description of the muscle actuators embedded in our facial tissue model, and the facial action coding process that controls these muscles to produce recognizable expressions. Section IV presents techniques for enhancing the realism of the face model and personalizing it through the exploitation of geometric and photometric data acquired with active range sensors. Section V considers the analysis of video image sequences for the dynamic estimation of facial muscle parameters and demonstrates our approach

using an example. Section VI discusses our work and suggests some future research directions. Section VII concludes the paper.

II. Background and Motivation

The human face has attracted much attention in several disciplines, including psychology, computer vision, and computer graphics. Psychophysical investigations clearly indicate that faces are very special visual stimulii. Psychologists have studied various aspects of human face perception and recognition [5], [3]. They have also examined facial expression—the result of a confluence of voluntary muscle articulations that deform the neutral face into an expressive face. The facial pose space is immense. The face is capable of generating on the order of 55 000 distinguishable facial expressions with about 30 semantic distinctions. For example, Ekman and Friesen's facial action coding system (FACS) provides a quantification of facial expressions [8]. Studies have identified six primary expressions that communicate anger, disgust, fear, happiness, sadness, and surprise in all cultures [7]. The FACS quantifies facial expressions in terms of 44 action units (AU) involving one or more muscles and associated activation levels.

We employ a reduced version of Ekman and Friesen's FACS in a sophisticated computational model of the human face that we originally developed for the realistic animation of synthetic characters. Facial animation in computer graphics began with Parke's use of facial images as keyframes and his subsequent popularization of parameterized face models [22]. State-of-the-art parameterized models can produce impressive animation using parameters associated with facial muscle structures [33], [30]. Graphics researchers have devoted significant effort to parametric facial modeling but little effort to the inverse problem of extracting parameters from facial images. There is some relevant work on lip synchronization during continuous speech animation, but the parameter extraction techniques proposed remain predominantly manual [20], [30], [12]. Reflective markers have been placed on the face in order to extract parameters for performance-driven facial animation [35].

Automatic facial recognition had an early start in image understanding, but work on the problem has been sporadic over the years, evidently due to the difficulty of extracting meaningful information from facial images. Facial classification systems based on measurements derived from interactively selected fiducial points (eye and mouth corners, nose, top of head, etc.) go back to the mid 1960's [2], [15], [11]. Early attempts at recognition through automated facial feature identification include [25] and [13]. Part of the difficulty of facial image analysis is that the face is highly deformable, particularly around the forehead, eyes, and mouth, and these deformations convey a great deal of meaningful information. Techniques for dealing with the deformation of facial features include spring-loaded subtemplates [9], deformable contour models that are also known as snakes [14], and deformable templates [31], [26]. In our approach, we apply a variant of snakes. Snakes are dynamic deformable contour models that require some routine image processing and, in our application, a modest amount of user input during initialization. Snakes have also been applied to the related problem of determining the location of the head in images [32].

We argue that the anatomy and physics of the human face, especially the arrangement and actions of the primary facial muscles, provide a good basis for facial image analysis [29]. We use snakes to track the position of the head and the nonrigid motions of the eyebrows, nasal furrows, mouth, and jaw in the image plane. We are able to estimate dynamic facial muscle contractions directly from the snake state variables. These estimates make appropriate control parameters for resynthesizing facial expressions through our face model. The model resynthesizes facial images at real-time rates. Real-time synthesis is desirable for model-based analysis-synthesis coding of facial images (see e.g., [1] and [10]). Our approach is philosophically similar to that described in [1] and in [4] but differs in the details of the image analysis and resynthesis. In particular, we employ physical rather than geometric modeling methods.

The purely geometric nature of prior face models [33], [17], [30] limits their ability to synthesize realistic facial animation because it ignores the fact that the human face is an elaborate biomechanical system. Our face model takes a more fundamental, physics-based approach to synthesizing the many subtleties of facial tissue deformation in response to facial muscle actions (such as the skin wrinkles and furrows shown in Fig. 1). A wealth of biomedical literature on tissue mechanics [16] has provided motivation for finite element models of facial tissue that are suitable for surgical simulation [18], [6] (see also Pieper's deformable lattice model [23]). The next section describes our realistic face model that incorporates anatomically based muscle actuators with a physics-based synthetic tissue model.

III. A Realistic Face Model

We have developed a hierarchical model of the face that provides natural control parameters and is efficient enough to run at interactive rates. Conceptually, the model decomposes into six levels of abstraction. These representational levels encode specialized knowledge about the psychology of human facial expressions, the anatomy of facial muscle structures, the histology and biomechanics of facial tissues, and facial skeleton geometry and kinematics:

1) *Expression*: At the highest level of abstraction, the face model executes expression (or phoneme) commands. For instance, it can synthesize any of the six primary expressions within a specific time interval and with a specified degree of emphasis.
2) *Control*: A muscle control process (a subset of Ekman and Friesen's FACS) 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 fibers contract, they displace their points of attachment in the facial tissue or the jaw.

4) *Physics*: The face model incorporates a physical approximation to human facial tissue. The tissue model is a lattice of point masses connected by nonlinear elastic springs. Large-scale synthetic tissue deformations, which are subject to volume constraints, are simulated numerically by continuously propagating, through the tissue lattice, the stresses induced by activated muscle fibers.
5) *Geometry*: The geometric representation of the facial model is a nonuniform mesh of polyhedral elements whose sizes depend on the curvature of the neutral face. Muscle-induced synthetic tissue deformations distort the neutral geometry into an expressive geometry.
6) *Images*: After each simulation time step, standard visualization algorithms implemented in dedicated graphics hardware render the deformed facial geometry in accordance with viewpoint, light source, and skin reflectance information to produce the lowest level representatation in the modeling hierarchy: a continuous stream of facial images.

The hierarchical structure of the model hides from the user most of the complexities of the underlying representations, relegating the details of their computation to automatic procedures. At the higher levels of abstraction, our face model offers the user a semantically rich set of control parameters that reflect the natural constraints of real faces.

With the above top-down overview in mind, we will now present some of the details of our model in a bottom-up fashion. We explain the structure and functionality of the synthetic tissue model and then describe the facial muscle models and how they interact with the tissue. Finally, we explain the assembly of the face model from these components as well as a specified surface geometry.

A. Physics-Based Synthetic Tissue Model

Our synthetic facial tissue model is motivated by histology and tissue biomechanics. Human skin has a nonhomogeneous and nonisotropic layered structure consisting of the epidermis (a superficial layer of dead cells), which is about one tenth the thickness of the dermis that it protects [16], [18]. The dermis is primarily responsible for the mechanical properties of skin. 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 little resistance to stretch as the collagen fibers begin to uncoil in the direction of the strain, but under greater stress, the fully uncoiled collagen fibers resist stretch much more markedly. This yields an approximately biphasic stress-strain curve (Fig. 2). The incompressible ground substance retards the motion of the fibers and thereby gives rise to viscoelastic behavior. Finally, the elastin fibers act like elastic springs that return the collagen fibers to their coiled condition under zero load. A layer of subcutaneous fatty tissue that allows the skin to slide rather easily over fibrous fascia covering the underlying muscle layer is underneath the skin (see Section III–C).

The synthetic tissue is a deformable lattice, which is an assembly of point masses connected by springs, that is, a discrete deformable model [28]. Let node i, where $i = 1, \ldots, N$, represent a point mass m_i whose three-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 $\mathbf{a}_i = d^2\mathbf{x}_i/dt^2$.

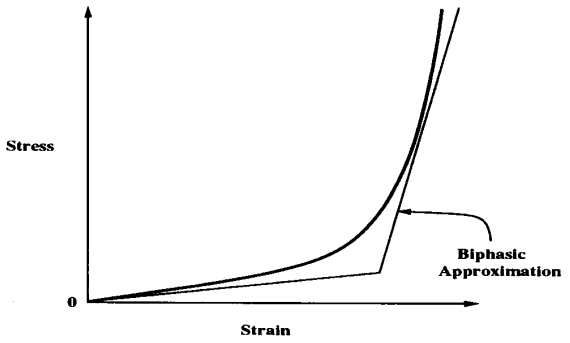


Fig. 2. Stress-strain curve of facial tissue and its biphasic approximation. The large-strain threshold e^c occurs at the intersection of the two lines.

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 $e_k = \|\mathbf{r}_k\| - l_k$. The (nonlinear) force that the spring exerts on node i is

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

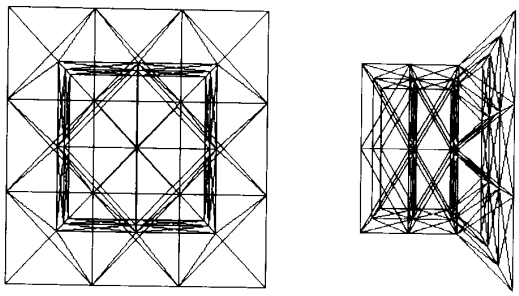
with

$$c_k = \begin{cases} \alpha_k & \text{when } e_k \le e_k^c, \\ \beta_k & \text{when } e_k > e_k^c \end{cases} \tag{2}$$

where the small-strain stiffness α_k is smaller than the large-strain stiffness β_k. Like real dermal tissue, this biphasic spring is readily extensible at low strains but exerts rapidly increasing restoring stresses after reaching a threshold e^c (Fig. 2).

We assemble the tissue model by arranging biphasic springs into structurally stable tetrahedral and hexahedral elements. Diagonal springs strut each face of the hexahedral elements so that they will resist shearing. Fig. 3 illustrates a small patch of the facial tissue model consisting of three layers of elements representing the cutaneous tissue, subcutaneous tissue, and muscle layer (the layers are not shown to scale). The biphasic springs (line segments) in each layer have different stiffnesses in accordance with the inhomogeneity of real facial tissue. The top-most surface represents the epidermis (which is a rather stiff layer of keratin and collagen), and we set the spring stiffnesses 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 bottom-most surface of the second layer represent the fascia to which the muscle fibers in the third layer are attached. Nodes on the bottom surface of the third layer are 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 that minimizes the deviation of the volume V_j of a deformed element E_j from its natural volume V_j^0 at rest. The volumes of elements are readily computable using vector algebra. The tissue incompressibility constraint is



(a) (b)

Fig. 3. Trilayer facial tissue model: (a) Top view; (b) side view showing (right to left) epidermal surface, dermal layer (pentahedral elements), and subcutaneous and muscle layers (hexahedral elements).

given by $Q = \sum_j (V_j - V_j^0)^2$. Differentiation of the constraint yields a net volume restoration force $\mathbf{q}_i$ for each node i: $\mathbf{q}_i = dQ/d\mathbf{x}_i$. Note that the derivative at a given node involves nonzero terms only over elements that share the node.

B. Numerical Simulation of Facial Tissue

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

$$\mathbf{g}_i(t) = \sum_{j \in \mathcal{N}_i} \mathbf{s}_k. \tag{3}$$

The discrete Lagrange equations of motion for the dynamic node/spring system is the system of 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 γ_i is the coefficient of velocity-proportional damping dissipating kinetic energy in the lattice, $\mathbf{g}_i$ is the net spring force (3), $\mathbf{q}_i$ is the net volume restoration force, and $\mathbf{f}_i$ is the net external force acting on node i. It is possible for facial muscle fibers to displace specific attachment nodes by applying driving forces $\mathbf{f}_i$ to them. In our current face model, however, the $\mathbf{f}_i$ are not used. Instead, inextensible muscle fibers displace attachment nodes by directly modifying their positions $\mathbf{x}_i$, as the following section will explain.

To simulate the dynamics of the deformable lattice, initial positions $\mathbf{x}_i^0$ (as determined by the face assembly procedure; see below) and velocities $\mathbf{v}_i^0 = \mathbf{0}$ are provided for each node i, and the equations of motion are numerically integrated forward though time. Each time step requires the evaluation of forces, accelerations, velocities, and positions for all of the nodes.

The explicit Euler method is a simple and quick time-integration method, but it has a limited range of stability [24]. Unfortunately, the greater computational complexity per time step of inherently more stable numerical methods can compromise interactive performance. A satisfactory solution to the stability/complexity tradeoff is provided by a second-order Runge-Kutta method, which requires two evaluations of the nodal forces per time step.

We choose m_i and γ_i such that the facial tissue exhibits a slightly overdamped behavior. The overdamped dynamics, the high flexibility of the biphasic springs in the small-strain region, and the use of muscle fiber displacements rather than driving forces all contribute to enhance the stability of the numerical simulation.

C. Facial Muscle Control Process

Muscles are bundles of muscle fibers working in unison. 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 attaches to and 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 attaches to and 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 toward a virtual center; an example is the orbicularis oris, which circles the mouth and can pout the lips.

In the human face, more than 200 voluntary muscles can exert traction on the facial tissue to create expressions. When the muscles contract, they pull the facial soft tissue to which they *attach* toward the place where they *emerge* from the underlying bony framework of the skull. Waters [33] and others have achieved a broad range of facial expressiveness by incorporating about 20 muscle actuators into their geometric face models.

In our physics-based face model, muscle actuators run through the third layer of the synthetic tissue (Fig. 3). Muscles fibers emerge from some nodes fixed in "bone" at the bottom of the third layer and attach to mobile nodes on the upper surface of the layer (fascia).

Let $\mathbf{m}_i^e$ denote the point where muscle i emerges from the "bone," and $\mathbf{m}_i^a$ 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 in the fascia layer from $\mathbf{x}_j$ to $\mathbf{x}_j'$ due to muscle contraction is a weighted sum of m muscle activities acting on node j:

$$\mathbf{x}_j' = \mathbf{x}_j + \sum_{i=1}^{m} c_i b_{ij} \mathbf{m}_i \tag{5}$$

where $0 \le c_i \le 1$ is a contraction factor, and b_{ij} is a muscle blend function that specifies a radial zone of influence for the muscle fiber. Defining $\mathbf{r}_{ij} = \mathbf{m}_i^a - \mathbf{x}_j$

$$b_{ij} = \begin{cases} \cos\left(\frac{\|\mathbf{r}_{ij}\|}{r_i}\frac{\pi}{2}\right): & \text{for } \| \mathbf{r}_{ij} \| \le r_i \\ 0: & \text{otherwise} \end{cases} \tag{6}$$

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

Once all muscle interactions have been computed, the positions $\mathbf{x}_j$ of nodes that are subject to muscle actions are displaced to their new positions $\mathbf{x}_j'$. As a result, the nodes in the fatty, dermal, and epidermal layers that are not directly influenced by muscle contractions are in an unstable state, and unbalanced forces propagate through the lattice to establish a new equilibrium position.

The face model incorporates a subset of the FACS representation [8], which was implemented as part of Water's prior

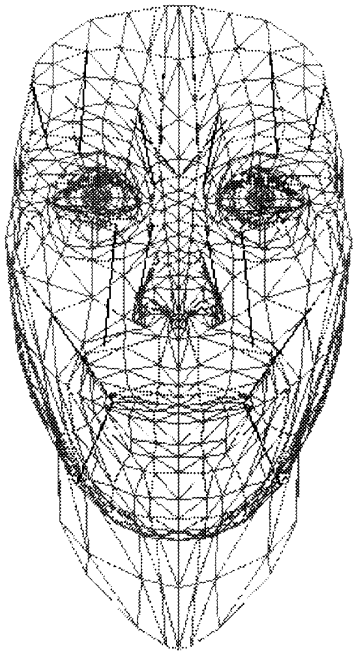


Fig. 4. Epidermal mesh and 16 muscle vectors (dark lines).

geometric model [33]. The FACS AU's are grouped into those that affect the upper and lower faces, 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. Through the FACS abstraction, it is possible to suppress the low-level details of coordinated muscle actuation and provide an interface to the model in terms of high-level expression commands.

D. Assembling and Simulating the Model

The automatic face model assembly procedure starts with a nonuniform triangular facial mesh whose nodes and springs represent the epidermis. First, it projects normal vectors from the center of gravity of each triangle into the face to establish subcutaneous nodes and forms terahedral dermal elements by connecting them to epidermal nodes using dermal springs. Second, it forms hexahedral subcutaneous elements by attaching short weak springs from the subcutaneous nodes downwards to muscle layer nodes. Third, it adds the muscle layer of hexahedral elements, whose lower nodes are constrained, anchoring them in "bone." Finally, it inserts the muscle fibers through the muscle layer from their emergence in "bone" to their attachments at muscle layer nodes.

Fig. 4 shows the epidermal triangles and 14 muscle vectors (the dermal and subcutaneous layers are suppressed for clarity) after the automatic assembly starting from the facial mesh employed in [33]. The synthetic tissue includes about 960 elements with approximately 6500 springs in total. The physics-based face model can be simulated and rendered at interactive rates on a single CPU of a Silicon Graphics Iris 4D-340VGX workstation. Fig. 1 shows several frames from a videotape recorded in real-time as the user interacted with the model through a menu-driven, mid-level interface enabling the contraction of individual muscles.

IV. Personal Face Models from Scanned Data

It is possible to enhance the realism of the face model dramatically through texture mapping, which is a widely adopted technique in model-based facial image coding. We describe initial work in this direction in [34] as well as recent work in [19].

More specifically, our polygonal face model is useful for capturing the 3-D geometry of faces from scanned data. For example, Fig. 5 shows a 360° head-to-shoulder scan of a woman (Heidi, which was acquired by Cyberware, Inc.) using a Cyberware Color 3-D Digitizer. The data set consists of a radial range map (Fig. 5(a)) and a registered RGB photometric map (Fig. 5(b)). The range and RGB maps are high-resolution 512×256 arrays in cylindrical coordinates, where the x axis is the latitudinal angle around the head, and the y axis is vertical distance. Fig. 5(c) shows the epidermal mesh of Fig. 4 radially projected into the 2-D cylindrical domain and overlayed on the RGB map. The triangle edges in the mesh are stretchy springs, and the mesh has been conformed semi-interactively to the woman's face using both the range and RGB maps [34], [19]. The nodes of the conformed mesh serve as sample points in the range map. Their cylindrical coordinates and the sampled range values are employed to compute 3-D Euclidean space coordinates for the polygon vertices. In addition, the nodal coordinates serve as polygon vertex texture map coordinates into the RGB map. Fig. 5(d) shows the 3-D facial mesh with the texture-mapped photometric data.

The visual quality of the face model is comparable to a 3-D display of the original high resolution data, despite the significantly coarser mesh geometry. We can visualize the texture-mapped model from arbitrary viewpoints at interactive rates on the SGI workstation that implements texture mapping in hardware.

Once we have reduced the scanned data to the 3-D epidermal mesh of Fig. 5(d), we can assemble a physics-based face model of Heidi using the assembly procedure described in the previous section. Fig. 5(e) and (f) demonstrates that we can animate the resulting face model by activating muscles.

V. Analysis of Dynamic Facial Images

In this section, we consider the inverse problem to facial image synthesis, i.e., the analysis of images of expressive faces. Our specific goal is to infer dynamic muscle contraction parameters that may be used to drive the physics-based model. This problem is challenging because it requires the reliable estimation of quantitative information about extended facial features that are moving nonrigidly. We develop a method that enables us to capture dynamic facial expressions directly from video sequences.

Through straightforward image processing, we convert digitized image frames into 2-D potential functions whose ravines (extended local minima) correspond to salient facial features such as the eyebrows, mouth, and chin. We employ a variant

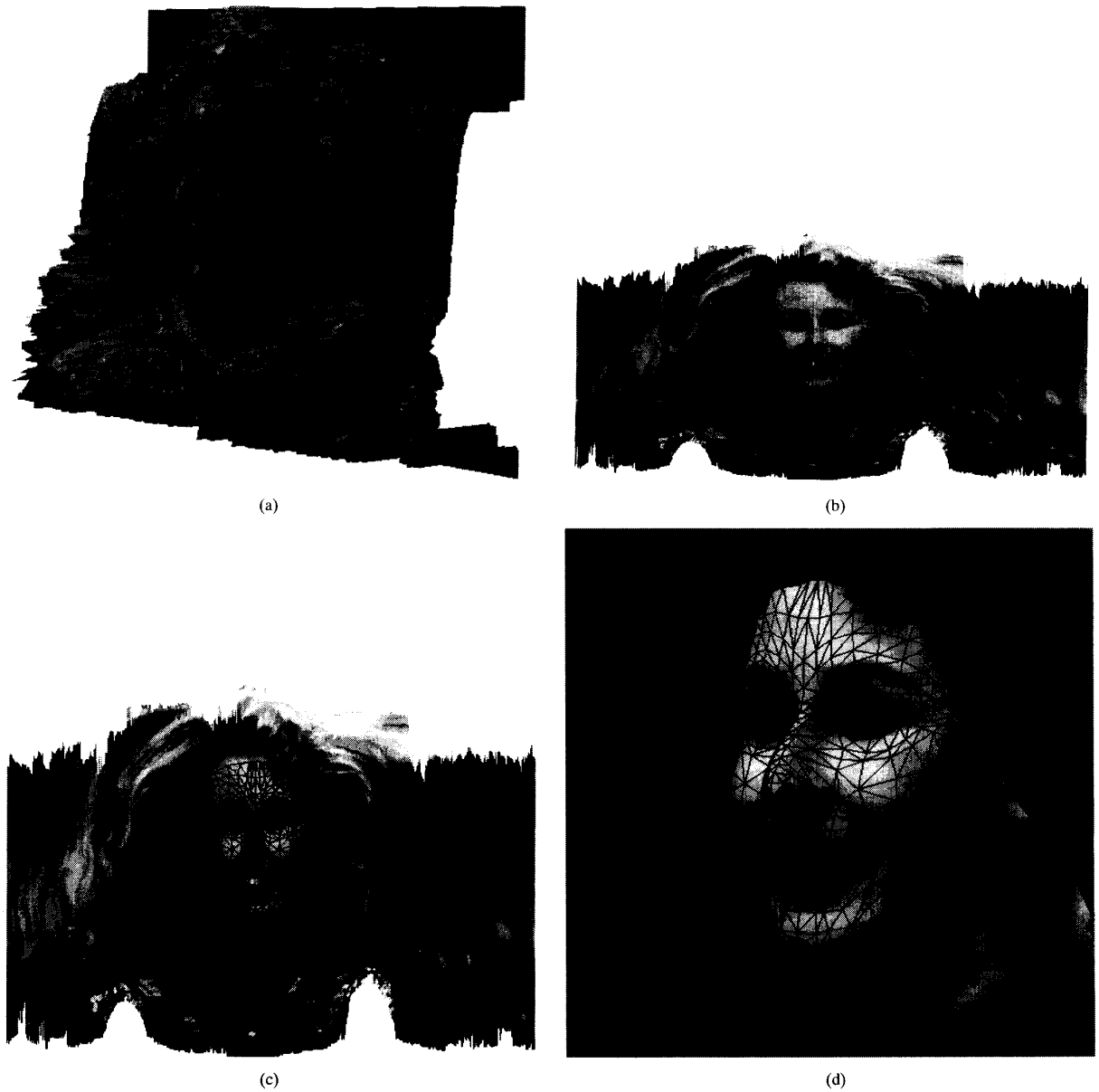


Fig. 5. Facial modeling using scanned data: (a) Radial range map; (b) RGB photometric map; (c) RGB map with conformed epidermal mesh overlayed; (d) 3-D mesh and texture mapped triangles.

of deformable contour models, or snakes, introduced in [14] and [27]. The deformable contours lock onto the ravines, tracking them from frame to frame. The deformable model's state variables provide quantitative information about the nonrigid shapes and motions of the evolving facial features. The automatic interpretation of this information leads to dynamic muscle parameters that allow the face model to reconstruct motions.

A. Discrete Deformable Contour Models

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 viscoelasticity and rigidity. We define a discrete deformable contour as a set of n 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 $\mathbf{f}_i(t)$ that act in the image plane.[1]

We connect the nodes in series using nonlinear springs. Following the formulation of (1), let l_i be the given reference

[1] Note that although the vectors $\mathbf{x}_i$ and $\mathbf{f}_i$ for the deformable contour are analogous to those found in the dynamics equations (4) of the synthetic facial tissue model, they are different, two-component vectors in the ensuing discussion.

(e) (f)

Fig. 5. Facial modeling using scanned data: (e), (f) Animate face model.

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

$$\boldsymbol{\alpha}_i = \frac{a_i e_i}{\|\mathbf{r}_i\|}\mathbf{r}_i \tag{7}$$

where the a_i's are tension variables. A viscoelastic contour may be obtained by letting

$$\frac{dl_i}{dt} = \nu_i e_i \tag{8}$$

where ν_i is a coefficient of viscoelasticity. Introducing rigidity variables b_i, the rigidity force is

$$\begin{aligned}\boldsymbol{\beta}_i = {} & b_{i+1}(\mathbf{x}_{i+2} - 2\mathbf{x}_{i+1} + \mathbf{x}_i) - 2b_i(\mathbf{x}_{i+1} - 2\mathbf{x}_i + \mathbf{x}_{i-1}) \\ & + b_{i-1}(\mathbf{x}_i - 2\mathbf{x}_{i-1} + \mathbf{x}_{i-2}).\end{aligned} \tag{9}$$

The behavior of an interactive deformable contour is governed by the first-order dynamic system

$$\gamma\frac{d\mathbf{x}_i}{dt} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_i = \mathbf{f}_i: \qquad i = 1, \ldots, n \tag{10}$$

where γ is a velocity-dependent damping coefficient. Tension and rigidity are locally adjustable though the a_i and b_i variables. In particular, by setting $a_i = b_i = 0$, we are able to break a deformable contour to create several shorter contours on an image.

To simulate the deformable contour, we integrate the system of ordinary differential equations (10) forward through time using a semi-implicit Euler method [24]. Applying the forward finite difference approximation $d\mathbf{x}_i/dt \approx (\mathbf{x}_i^{t+\Delta t} - \mathbf{x}_i^t)/\Delta t$ to (10), evaluating the linear terms in the $\mathbf{x}_i$ (i.e., $\boldsymbol{\beta}_i$) at time $t + \Delta t$ and the nonlinear terms at time t yields the pentadiagonal system of algebraic equations

$$\frac{\gamma}{\Delta t}\mathbf{x}_i^{t+\Delta t} + \boldsymbol{\beta}_i^{t+\Delta t} = \frac{\gamma}{\Delta t}\mathbf{x}_i^t - \boldsymbol{\alpha}_i^t + \mathbf{f}_i^t \tag{11}$$

for the new node positions $\mathbf{x}_i^{t+\Delta t}$ in terms of the current positions $\mathbf{x}_i^t$. Since the system has a constant coefficient matrix, we factorize it only once at the beginning of the deformable contour simulation using a direct LDU factorization method and then efficiently resolve with different right-hand sides at each time step (see [27] for details).

B. External Forces and Image Processing

The deformable contour is responsive to an image force field that influences its shape and motion. It is convenient to express the force field as the gradient of 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 [14] 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 $\nabla = [\partial/\partial x, \partial/\partial y]'$.

In the present application, we are concerned with the localization of extended image features such as the eyebrow and lip boundaries. Usually, these features correspond to high-contrast regions in the image intensity function $I(x, y, t)$. To make these regions deformable contour attractors, we use

$$P(x, y, t) = -\|\nabla G_\sigma * I(x, y, t)\| \tag{13}$$

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

C. Tracking Nonrigid Facial Features

In a few simulation time steps, the deformable contours slide downhill in $P(x, y, t_k)$ (image frame k), conforming to the shapes of its ravines as they come to equilibrium at their bottoms. Once they equilibrate, the contours accurately trace the facial features of interest. As soon as the contours have equilibrated in $P(x, y, t_k)$, we replace it with $P(x, y, t_{k+1})$ associated with the next video frame. Continuing from their previous equilibrium positions, the contours slide downhill to again equilibrate in the perturbed ravines, thus tracking their nonrigid motions. We repeat the process on successive frames.

This simple tracking scheme works if 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. Should part of a contour escape the attractive zone of a ravine, however, the rest of the contour will usually pull it back into place.

D. Estimating Facial Muscle Contractions

As the deformable contours evolve from frame to frame, their dynamic state variables $\mathbf{x}_i^t$ and their time derivatives provide explicit information about the nonrigid shapes and motions of the facial features. The information is reduced to a head reference frame and 11 dynamic fiducial points.

In our muscle contraction estimation process, we have employed, to date, nine deformable contour sections. These localize and track 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. Using the deformable contour state variables, an automatic procedure first calibrates the input image to the face model and then computes the following:

1) A head reference frame from the average position of the hairline contour
2) contractions of the left and right inner, major, and outer occipitofrontalis from the positions of the inner-most, center, and outer-most points of the associated eyebrow contours, respectively
3) contractions of the left and right zygomaticus major and depressor labii inferioris from the positions of the endpoints of the upper lip contour
4) contraction of the left and right levator labii superioris alaeque nasi from the positions of the upper-most points of the associated nasolabial furrow contours
5) jaw rotation from the average position of the chin boss contour.

Fig. 6 illustrates the positions of the nine deformable contours in equilibrium at two different frames of an image sequence that will be described in the next section. The dots indicate the 11 fiducial points mentioned above, which are computed from the snakes. The positions of the points are computed relative to the head reference frame, whose origin is marked by the crosshair in the figure. Assuming a frontal view and relatively stable hairline, the head reference frame will track the head motion in the image.

The muscle contraction estimation scheme makes the simplifying assumption of orthographic projection. It estimates

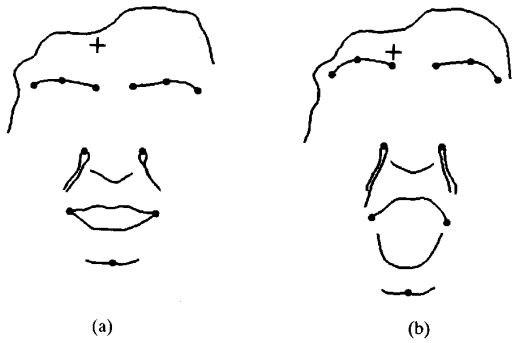


Fig. 6. Snakes and fiducial points used for muscle contraction estimation: (a) Neutral face; (b) surprise expression.

$c_i(t)$ in the x-y image plane using (5) for each muscle independently while ignoring z coordinates. Starting with an image of the neutral face, the calibration procedure establishes the origin of the head reference frame and positions the muscle vector emergence points appropriately with respect to it. The snake fiducial points in Fig. 6(a) serve to approximate the positions of the muscle attachments. Once the natural lengths of the primary facial muscles have been established, the muscle contractions that are responsible for a dynamic expression may be estimated immediately as the snake fiducial points move from frame to frame. The jaw rotation is also estimated as the chin snake fiducial point descends from its neutral position in the head reference frame.

The next section illustrates this dynamic estimation procedure with an example.

E. An Experiment in Facial Video Analysis

We have applied our facial image analysis technique to a video sequence of one of the authors (DT) performing facial expressions in frontal view before a CCD camera.[2] A surprise expression was digitized as a sequence of 256×256×8-b images and analyzed using deformable contours. Fig. 7 illustrates the facial image analysis and the results of the muscle contraction estimation on three image frames. Fig. 7(b) shows the (negative) potential functions computed from the frames in Fig. 7(a). To compute the potential, we apply a discrete smoothing filter $G(i, j)$ consisting of two four-neighbor local intensity averaging steps followed by the discrete gradient operator $\nabla v(i, j) = [(v(i + 1, j) - v(i, j)), (v(i, j + 1) - v(i, j))]'$. We bilinearly interpolate the result between pixels (i, j) to obtain the continuous potential function $P(x, y, t_k)$.

We initialize the deformable contours on the first frame of the sequence using the mouse. The initialization procedure places their nodes roughly 1 pixel apart and sets the rest of the lengths l_i in (8) 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 $a_i = b_i = 0.0$), $\nu_i = 0.2$, and

[2] Using the available video camera and lighting in our lab environment, it was necessary to enhance DT's lips, eyebrows, and nasolabial furrows by subjecting him to a humiliating makeup job. Under more favorable imaging situations, makeup may not be necessary, depending on the individual.

(a) (b) (c)

Fig. 7. Dynamic facial image analysis and expression resynthesis. Sample video frames with superimposed deformable contours trackingfacial features; (a) intensity images with black snakes, (b) imagepotentials with white snakes. (c) Facial model resynthesizes surpriseexpression from estimated muscle contractions.

$p = 0.001$. Fig. 7(a) and (b) shows the deformable contours at equilibrium locked onto the facial features.

From the first frame in the video sequence, which captures DT's face in a relaxed state, the analysis procedure first calibrates the face model to the frontal view of the subject and then estimates dynamic muscle contractions $c_i(t)$, as we explained in the previous section. Fig. 8 shows a plot of the estimated contractions versus the frame number. They are input to the physics-based model as a time sequence. The model quickly attains dynamic equilibrium on each frame input, and the state variables are rendered in real time on the SGI workstation to synthesize an animated sequence of facial images, three of which are shown in Fig. 7(c).

VI. Discussion

Our experiment has demonstrated the estimation of muscle contractions from video of a subject's face and their use in resynthesizing facial expression. Clearly, our technique is tolerant of the significant discrepancy between the 3-D geometry of the subject's face and the face model. It is difficult to assess quantitative accuracy, however, because ground-truth data are not readily available. Even though some contraction estimation errors may be quantitatively significant, we have noted that the expression resynthesis remains qualitatively robust.

The simple feed-forward analysis/synthesis scheme that we describe in this paper has some limitations. At present, the snakes require manual initialization, but we are confident that some heuristic facial feature detection procedure, like the one described in [13], can be modified to initialize them adequately. Although it is very efficient, the 2-D nature of the muscle-estimation scheme can cause problems in general, e.g., when the head turns significantly. Three-dimensional muscle contraction estimation assuming full, and not necessarily frontal, perspective projection is desirable. We can probably accomplish this in the future by exploiting multiple views of the face. It is evident that we can also improve the fidelity of the image-based facial expression analysis and graphical resynthesis loop with a more complete modeling of facial musculature. A limitation of the present facial model is the lack of an adequate model of the orbicularis oris, which is the highly articulate sphincter muscle that defines the lips. Once we incorporate a more sophisticated lip actuator, it will make sense to exploit more of the nodal variables available in the lip tracking snakes, rather than the two fiducial points that we currently employ. Another deficiency, at present, is the lack of eyelid and eye position estimation that results in discernible differences between the input expression and resynthesized expression. The eyelid is particularly difficult to track because of its speed. A more feasible solution would be to simply detect whether the eyelids are open or shut and input this information to the model.

Our work opens up many avenues for further research. For example, it seems possible to further automate the modeling approach that we have developed for working with scanned data and possibly extend it to the reconstruction of faces from grey-level images. Another interesting research direction would investigate the possibility of running the hierarchical model backward from the image level all the way up to the expression level, thereby addressing the problems of measuring, classifying, and recognizing dynamic human expressions from video sequences of the face. The FACS representation promises to be useful in addressing the expression recognition problem, as Mase [21] has argued recently.

Our demonstration that it is possible to analyze a particular face captured on video and reconstruct, with reasonable degree of accuracy, the expression in the different facial geometry of the model affirms the notion that muscle actions are the salient features of expression that are common across individual faces.

VII. Conclusion

A solid foundation for facial image analysis is the anatomy of the face, especially the arrangements and actions of the primary facial muscles. This paper has presented a new approach to facial image analysis using a realistic facial model. We have described a hierarchical model of the human face that incorporates a physics-based synthetic 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 graphics workstation. We use snakes to track the position of the head and the nonrigid motions of the eyebrows, nasal furrows, mouth, and jaw in the image plane. Reducing the snake measurements to fiducial points within a head reference frame, we are able

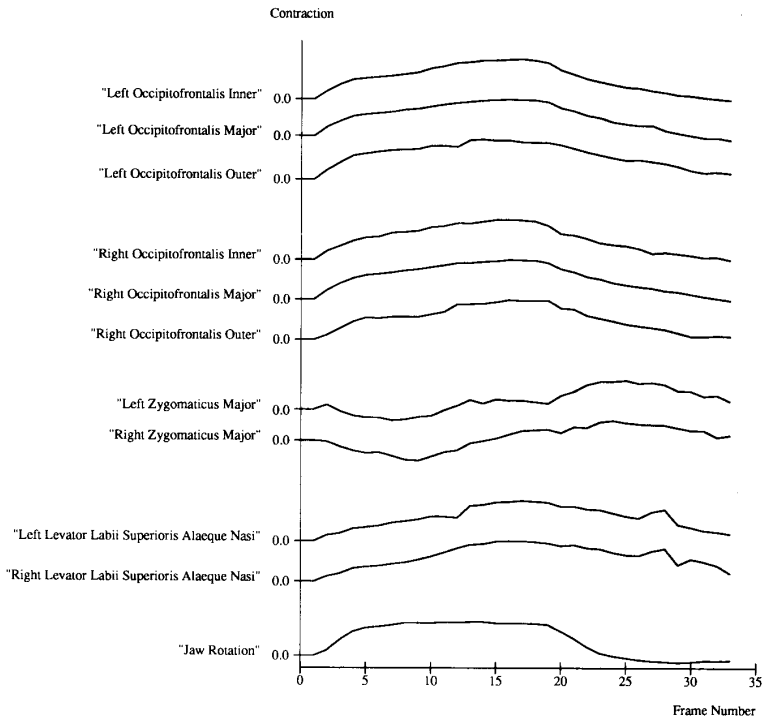


Fig. 8. Estimated muscle contractions plotted as time series.

to estimate the dynamic contractions of the primary facial muscles. These estimates make appropriate control parameters for resynthesizing facial expressions through our face model.

ACKNOWLEDGMENT

We thank Y. Lee for his contributions to our facial modeling research. He wrote the algorithms that generated the images for Fig. 5 and constructed the physics-based face models of us shown in the biography photos. Scanned data were provided courtesy of Cyberware, Inc., Monterey, CA. We also thank R. Smith and I. Chakravarty for their support at Schlumberger LCS and T. Crossley for his assistance with image digitization.

REFERENCES

[1] K. Aizawa, H. Harashima, and T. Saito, "Model-based analysis synthesis image coding (MBASIC) system for a person's face," *Signal Processing: Image Commun.*, vol. 1, no. 2, pp. 139–152, 1989.

[2] W. W. Bledsoe, "Man-machine facial recognition," Panoramic Res. Inc., Palo Alto, CA, Rep. PRI:22, Aug. 1966.

[3] V. Bruce, *Recognizing Faces*. Hillsdale: Lawrence Erlbaum, 1988.

[4] C. S. Choi, H. Harashima, and T. Takebe, "Analysis and synthesis of facial expressions in knowledge-based coding of facial image sequences," in *Proc. Int. Conf. Acoustics Speech Signal Processing* (Toronto), 1991, pp. 2737–2740.

[5] G. M. Davies, H. D. Ellis, and G. M. Shepherd, *Perceiving and Remembering Faces*. New York: Academic, 1981.

[6] X. Deng, "A finite element analysis of surgery of the human facial tissues," Ph.D. thesis, Graduate Sch. Arts Sci., Columbia Univ., New York, NY, 1988.

[7] P. Ekman, *Unmasking the Human Face*. New York: Prentice-Hall, 1971.

[8] P. Ekman and W. V. Friesen, *Manual for the Facial Action Coding System*. Palo Alto: Consulting Psychologists, 1977.

[9] M. A. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures," *IEEE Trans. Comput.*, vol. C-22, pp. 67–92 1973.

[10] R. Forchheimer and T. Kronander, "Image coding—From waveforms to animation," *IEEE Trans. Acoustics Speech Signal Processing*, vol. ASSP-37, no. 12, pp. 2008–2023, 1989.

[11] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, "Man-machine interaction in human face identification," *Bell System Tech. J.*, vol. 51, no. 2, pp. 399–427, 1972.

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

[13] T. Kanade, "Picture processing system by computer complex and recognition of human faces," Ph.D. Thesis, Dept. of Inform. Sci., Kyoto Univ., Nov. 1973.

[14] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vision*, vol. 1, no. 4, pp. 321–331, 1987.

[15] Y. Kaya and K. Kobayashi, "A basic study on human face recognition," in *Frontiers of Pattern Recognition* (S. Watanabe, Ed.). 1972, p. 265.

[16] R. M. Kenedi, T. Gibson, J. H. Evans, and J. C. Barbenel, "Tissue mechanics," *Phys. Med. Biol.*, vol. 20, no. 5, pp. 699–717, 1975.

[17] K. Komatsu, "Human skin model capable of natural shape variation," *Visual Comput.*, vol. 3, pp. 265–271, 1988.

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

[19] Y. Lee, D. Terzopoulos, and K. Waters, "Constructing physics based facial models of individuals," in *Proc. Graphics Interface '93* (Toronto), to be published in May 1993.

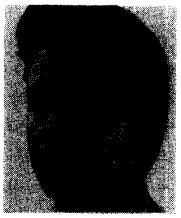
[20] J. P. Lewis and F. I. Parke, "Automated lipsynch and speech synthesis for character animation," in *Proc. Human Factors Comput. Syst. Graphics Interface '87* (Toronto), 1987, pp. 143–147.

[21] K. Mase, "Recognition of facial expression from optical flow," *IEICE Trans.*, vol. E74, no. 10, pp. 3474–3483, Oct. 1991.

[22] F. I. Parke, "Parameterized models for facial animation," *IEEE Comput. Graphics Applications*, vol. 2, no. 9, pp. 61–68, Nov. 1982.

[23] S. Pieper, "More than skin deep: Physical modeling of facial tissue," M.Sc. thesis, Dept. of Media Arts Sci., Mass. Inst. of Technol.,

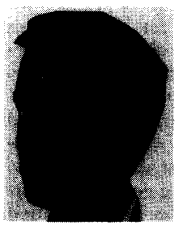
Cambridge MA, Jan. 1989.
[24] W. Press, B. Flanney, S. Teukolsky, and W. Verttering, *Numerical Recipes: The Art of Scientific Computing,*. Cambridge, UK: Cambridge University Press, 1986.
[25] T. Sakai, M. Nagao, and S. Jujibayashi, "Line extraction and pattern detection in a photograph," *Patt. Recogn.*, vol. 1, p. 233, 1969.
[26] M. A. Shakleton and W. J. Welsh, "Classification of facial features for recognition," in *Proc. Comput. Vision Patt. Recogn. Conf. (CVPR'91)* (Lahaina, HI), 1991, pp. 573–578.
[27] D. Terzopoulos, "On matching deformable models to images," *in Proc. Topical Meeting Machine Vision*, Tech. Digest Series, vol. 12, Opt. Soc. Amer., Washington, DC, 1987, pp. 160–163; also Tech. Rep. 60, Schlumberger Palo Alto Res., Palo Alto, CA, Nov. 1986.
[28] D. Terzopoulos and K. Fleischer. "Deformable models," *Visual Comput.*, vol. 4, no. 6, pp. 306–331, 1988.
[29] D. Terzopoulos and K. Waters. "Analysis of dynamic facial images using physical and anatomical models," *in Proc. Third Int. Conf. Comput. Vision (ICCV'90)* (Osaka), 1990, pp. 727–732.
[30] N. Magnenat-Thalmann, E. Primeau, and D. Thalmann, "Abstract muscle action procedures for face animation," *Visual Comput.*, vol. 3, pp. 290–297, 1988.
[31] A. L. Yuille, D. S. Cohen, and P. W. Hallinan, "Feature extraction from faces using deformable templates," in *Proc. Comput. Vision Patt. Recogn. Conf.* (San Diego, CA), June 1989, pp. 104–109.
[32] J. Waite and W. Welsh, "Head boundary location using snakes," *Brit. Telecom Tech. J.*, vol. 8, no. 3, pp. 127–136, 1990.
[33] K. Waters, "A muscle model for animating three-dimensional facial expression," *Comput. Graphics*, vol. 22, no. 4, pp. 17–24, 1987.
[34] K. Waters and D. Terzopoulos, "Modeling and animating faces using scanned data," *J. Visualization Comput. Animation*, vol. 2, no. 4, pp. 123–128, 1991.
[35] L. Williams, "Performance-driven facial animation," *Comput. Graphics*, vol. 24, no. 4, pp. 235–242.



Demetri Terzopoulos (S'78–M'85) was born in Krestena, Greece, in 1956. He received the B.Eng. degree with distinction in honours electrical engineering and the M.Eng. degree in electrical engineering from McGill University, Montreal, Canada, in 1978 and 1980, respectively, and the Ph.D. degree in artificial intelligence from the Massachusetts Institute of Technology, Cambridge, MA, in 1984.

Since September 1989, he has been Associate Professor of Computer Science and Electrical and Computer Engineering at the University of Toronto and Fellow of the Canadian Institute for Advanced Research. From 1985–1992, he was affiliated with Schlumberger, Inc., serving as Program Leader at the Laboratory for Computer Science, Austin, TX, and at the former Palo Alto Research Laboratory. During 1984–1985, he was a Research Scientist at the MIT Artificial Intelligence Laboratory, Cambridge, MA. Previously, he held summer positions at the National Research Council of Canada, Ottawa, Canada, (1977–1978) and Bell-Northern Research, Montreal, Canada (1980). He has published research papers in computational vision, computer graphics, biomedical image analysis, digital speech and image processing, and parallel numerical algorithms and has been invited to speak internationally on these topics.

Dr. Terzopoulos has received several scholarships and awards, including an AAAI-87 Conference Best Paper Award from the American Association for Artificial Intelligence. He serves on the editorial boards of *CVGIP: Graphical Models and Image Processing* and the *Journal of Visualization and Computer Animation*. He is a member of the New York Academy of Sciences and Sigma Xi.



Keith Waters (M'92) was born in Kent, England, in 1962. He received the B.A. honours degree in graphic design in 1984 and the Ph.D. degree in computer graphics in 1988 from Middlesex Polytechnic, London.

Since August 1991, he has been a member of the research staff at Digital Equipment Corporation's Cambridge Research Lab, where he is involved in scientific visualization. From 1988 to 1991, he was a member of the technical staff at the Schlumberger Laboratory for Computer Science, Austin, TX, where he worked on 3-D visualization of seismic and borehole data. His current research interests include medical facial applications, physics-based modeling, computer-based facial synthesis, and volume visualization. He has published papers in computer graphics, computer vision, and biomedical visualization.

Dr. Waters serves on the editorial board for the *Journal of Visualization and Computer Animation* and is a member of the ACM Siggraph.