

SIGGRAPH 2004 Course Notes
Facial Modeling and Animation

Jörg Haber
MPI Informatik
Saarbrücken, Germany
haberj@mpi-sb.mpg.de

Demetri Terzopoulos
New York University
New York, USA
dt@cs.nyu.edu

1 Abstract

In this course we present an overview of the concepts and current techniques in facial modeling and animation. We introduce this research area by its history and applications. As a necessary prerequisite for facial modeling, data acquisition is discussed in detail. We describe basic concepts of facial animation and present different approaches including parametric models, performance-, physics-, and learning-based methods. State-of-the-art techniques such as muscle-based facial animation, mass-spring networks for skin models, and morphable models are part of these approaches. We furthermore discuss texturing of head models and rendering of skin, addressing problems related to texture synthesis and bump mapping with graphics hardware. Typical applications for facial modeling and animation such as medical and forensic applications (craniofacial surgery simulation, facial reconstruction from skull data, virtual aging) and animation techniques for movie production (case study of *The Matrix* sequels) are presented and explained.

2 Syllabus

The course will be organized according to the following time schedule:

time	length	topic	presenter
08:30–08:35	5 min	outline of the tutorial	
08:35–09:05	30 min	history & applications	F. Parke
09:05–09:20	15 min	anatomy of the human head	J. Haber
09:20–10:00	40 min	data acquisition for facial modeling	L. Williams
10:00–10:15	15 min	overview: facial animation techniques	V. Blanz
10:30–11:10	40 min	parametric models	F. Parke
11:10–11:35	25 min	performance-based facial modeling/animation	L. Williams
11:35–12:15	40 min	physically based facial modeling/animation	D. Terzopoulos
13:45–14:30	45 min	learning-based approaches	V. Blanz
14:30–15:00	30 min	rendering techniques	J. Haber
15:00–15:30	30 min	forensic applications	J. Haber
15:45–16:45	60 min	movie production	G. Borshukov
16:45–17:15	30 min	medical applications and behavioral models	D. Terzopoulos
17:15–17:30	15 min	questions, discussion	all

3 Contents

The tutorial notes contain both the slides from the tutorial presentation and some selected publications, which serve as additional background information.

1. Slides: **Facial Animation: History & Applications**
2. Slides: **Anatomy of the Human Head**
3. Slides: **Overview: Facial Animation Techniques**
4. Slides: **Parameterized Face Models**
5. Slides: **Facial Performance Capture** (Data Acquisition + Performance-based Approaches)
6. Slides: **Physically based Facial Modeling and Animation**
7. Paper: Y. Lee, D. Terzopoulos, K. Waters: *Realistic Modeling for Facial Animations*, Proc. SIGGRAPH '95, 55–62, Aug. 1995.
8. Slides: **Learning-based Approaches**
9. Paper: V. Blanz, T. Vetter: *A Morphable Model for the Synthesis of 3D Faces*, Proc. SIGGRAPH '99, 187–194, Aug. 1999.
10. Slides: **Rendering Techniques for Facial Animation**
11. Paper: M. Tarini, H. Yamauchi, J. Haber, H.-P. Seidel: *Texturing Faces*, Proc. Graphics Interface 2002, 89–98, May 2002.
12. Slides: **Forensic Applications**
13. Paper: K. Kähler, J. Haber, H. Yamauchi, H.-P. Seidel: *Reanimating the Dead: Reconstruction of Expressive Faces from Skull Data*, ACM Trans. Graphics (Proc. SIGGRAPH 2003), 22(3), 554–561, July 2003.
14. Slides: **Image-based Facial Animation and Rendering for *The Matrix* Sequels**
15. Slides: **Medical Applications & Behavioral Models**



Facial Animation: History and Applications

*Fred Parke
Texas A&M University*

Applications of Facial Modeling and Animation



Entertainment animation and VFX
Interactive games
Human-computer interfaces
Telepresence
Perception research
Medical and educational

Facial Animation: History and Applications



A look back over the last 35 years

A Look Ahead – Future History

Convincing 'Realistic' Faces



- *The challenge has been the synthesis of artificial faces that look and act like your mother, brother, friend, or some well know celebrity*
- *A huge challenge because of familiarity*
- *The 'closer' you get the harder it is*

Facial Animation: Historical Perspective



Pre-history

Facial representation has been a major focus of art forms from ancient times up to the present

- *archeological artifacts*
- *sculpture*
- *drawing*
- *painting*
- *and traditional animation*

Facial Animation: Historical Perspective



1600's

First published investigations of facial expression

– *John Bulwer, London, 1648 and 1649*

1800's

'The mechanism of human facial expression'

– *G. Duchenne, Paris, 1862*

'Expression of the emotions in man and animals'

– *C. Darwin, London, 1872*

Broad Trends/Themes



- **Exponential increase in computer power**
– ~1000x every 15 years
- **Steady development of new and refinement of existing techniques, interspersed with flashes of insight**
- **Better and better tools**
- **Ever increasing expectations**
– speed, complexity, realism

Technique Categories



- **Sources of geometric data**
- **Modeling primitives**
- **Animation control**
- **Rendering**
- **Tools**

Sources of Geometric Data



- **Graph paper**
- **Direct surface measurement**
- **Photographic**
- **Laser scanners**
- **Structured light**
- **Interactive surface 'sculpting' systems**

Geometric Modeling



- **Vectors**
- **Polygonal surfaces**
- **Bi-cubic parametric surfaces**
– B-Splines, NURBS, ...
- **Subdivision surfaces**

Development of interactive modeling tools

Animation Control



- **Shape interpolation**
- **Direct parameterizations**
- **Muscle-based parameterizations**
- **Expression/Viseme level parameterizations**
- **Dynamic simulations**
- **Facial 'rigs' based on 'skeletons', deformers, blend shapes, ...**

Animation Control Handles



Scripted or interactive control of:

- **Interpolation coefficients**
- **Interpolation of parameter values**
– direct or muscle based parameters
- **Dynamic forces**
- **Facial rig 'handles'**

Key frame values, interactive curve editors

Rendering Techniques

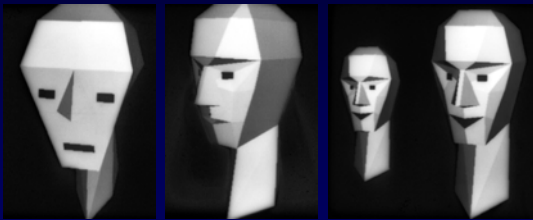
- *Vectors, flat shaded polygons*
- *Gouraud, Phong, Blinn shading*
- *Texture mapping*
- *Bump/displacement mapping*
- *Shader languages – Renderman, ...*
- *Global illumination techniques*
- *Video resolution → Theatrical resolution*

Facial Animation: Historical Perspective

Early 1970's

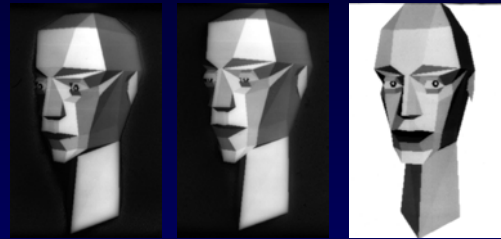
- Utah Graphics Class Project 1971
- Henri Gouraud's dissertation face 1971
- Chernoff's work 1971
- Interpolated Faces at Utah 1972 and 1973
- Gillenson at Ohio State 1973
- Parameterized Face Model at Utah 1974

Initial 3D Faces - 1971



F. Parke, University of Utah
Less than 100 polygons

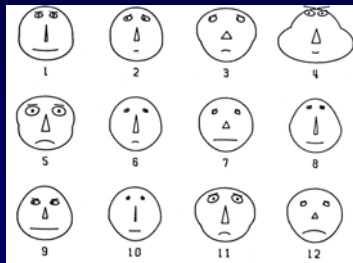
Initial Parametric Model - 1971



'Parameters' for eyes, eyelids, mouth
Used to create a 'flipbook' animation

Chernoff's work - 1971

Used faces to present n-dimensional data



Interpolated Faces - 1972

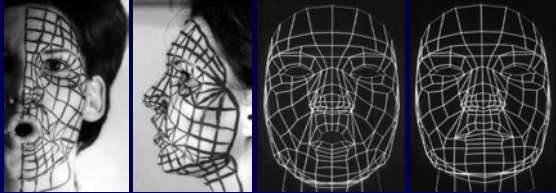
Facial Expression Interpolation



F. Parke – University of Utah

Interpolated Faces - 1972

Data Collection Technique



Interpolated Faces - 1972

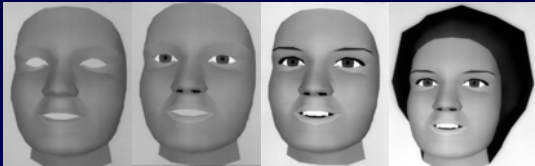
Interpolated Face Data Animation



Interpolated Faces - 1972

Face Components

Facial mask, eyes, eyebrows, teeth, hair



Interpolated Faces - 1972

Interpolated expression animation



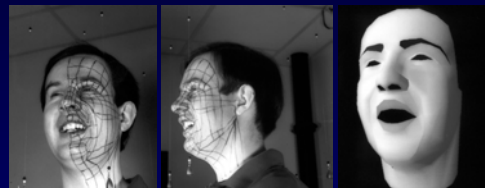
Interpolated Faces - 1973

Interpolation between individual faces



Interpolated Faces - 1973

Data Collection Technique



Interpolated Faces - 1973

Animation between individual faces



Parameterized Model - 1974

Expression and Conformation Control



F. Parke – University of Utah

Parameterized Model - 1974

Speech Synchronized Animation



Facial Animation: Historical Perspective

Late 1970's and Early 1980's

- **Facial Action Coding System (FACS)**
 - Ekman and Friesen - 1977
- **Interactive Parameterized Model - 1979**
 - Implemented on E&S CT-1 at Case Western
- **Parametric Model 'transported' to NYIT - 1980**
 - Later to U. Calgary and UCSC
 - Evolved into 'Baldi'
- **Muscle Based Expression Model - 1981**
 - Platt and Badler – University of Pennsylvania

1980's

Rise of the production studios

- Many started, a few survive

Bifurcation of development efforts

- Academic research
 - Goals – knowledge, understanding, new methods, grants, publications...
- Production studio development
 - Goals – get the job, get the job done – on time, make money, survive!

Facial Animation: Historical Perspective

Early to Mid 1980's

- 1981 – PC introduced, Wavefront software
- 1982 – SGI graphic workstations, Alias Research
- 'Caricature' Faces – 1982
 - S. Brennan - MIT
- 'Tony de Peltrie' – 1985
- Softimage - 1986

Facial Animation: Historical Perspective



Late 1980's

- **Automatic Speech Synchronization**
 - Lewis and Parke, NYIT 1987
 - Hill, et al, U. Calgary 1988
- **New Muscle Models**
 - K. Waters - 1987
 - Thalmann, et al – 1988
 - Waters and Terzopoulos - 1990

Facial Animation: Historical Perspective



Late 1980's

- **'Rendezvous in Montreal' – Thalmann 1987**
- **'Tin Toy' baby – Pixar 1988**
- **'The Abyss' water pseudopod face – 1989**
- **'Don't Touch Me' – Kleiser/Walczak - 1989**
- **Siggraph Facial Animation tutorials - 1989/90**
 - Simple parameterized model put in 'public domain'

Facial Animation: Historical Perspective



Early 1990's – increasing activity

- **Performance based Facial Animation**
- **SMILE multi-level animation system**
 - Kalra, et al, 1991
- **NSF Workshop on Facial Expression Understanding – 1992**
- **NSF Workshop on Facial Animation Standards – 1994**

Facial Animation: Historical Perspective



Mid 1990's

- **Real time speech synchronization**
 - Parke at IBM, Waters at DEC
- **Use in interfaces – agents/avatars**
- **Much activity in support of low bandwidth video conferencing**
- **'Babe', 'Toy Story', 'The End' - 1995**
- **First book on facial animation – 1996**
- **Speech Co-articulation – Pelachaud, et al, 1996**

Facial Animation: Historical Perspective



Late 1990's

- **Use in feature films**
 - Dragonheart - 1996
 - Geri's game – 1997 (subdivision surfaces)
 - A Bugs Life, ANTZ – 1998
 - Stuart Little – 1999
 - Star Wars Episode I – 1999
- **'Principle Component' Face Model**
 - Blanz and Vetter, 1999
- **'Voice Puppetry' – Brand 1999**
- **MPEG-4 Facial Model Coding**

Facial Animation: Historical Perspective



2000's

- **Commercially Successful!**
- **Synthetic characters in leading roles**
 - 2001 - Final Fantasy, Shrek, Jimmy Neutron, LOR
 - 2002 – LOR, Star Wars Episode II
 - 2003 – LOR (Gollum), The Hulk, The Matrix: Revolutions
- **Exponential Growth!**

Applications of Facial Modeling and Animation



Entertainment animation and VFX

Interactive games

Human-computer interfaces

Telepresence

Perception research

Medical and educational

Entertainment animation/VFX



- *Currently the major application and driving force*
- *Synthetic characters in leading and support roles*
- *Digital stand-ins*
- *Crowd simulation*

Interactive games



- *Another major application and driving force*
- *Quality expectations approaching those for entertainment animation*
- *Real-time performance required*
- *'Behavior' modeling important*

Human-computer interfaces



- *Requires interactive models*
- *Applications*
 - *Software agents*
 - *Social agents*
 - *Conversational interfaces*
 - *Kiosks*
 - *Stage shows, ...*

Agent Applications



- *Provides screen presence for agent software*
- *Provides an interaction 'focus'*
- *Conversational interfaces*
 - *Two way speech*
 - *Speech recognition*
 - *Synchronized speech animation response*

Kiosk Applications



- *Attracts attention*
 - *Initial 'patter'*
 - *Solicits user query interaction*
- *Provides response information*
 - *Guides query interaction*
 - *Spoken query feedback*

Stage Show Applications



- **As emcee or host**
 - Introduces show elements
 - Interacts with audience
- **As 'sidekick' for a real host**
 - Dialog with real host

Interactive Model Attributes



- **Expressive**
 - able to assume an appropriate range of expressions
- **Responsive and 'alive'**
 - synchronized speech and expression
- **'Intelligent Behavior'**
 - 'appropriate' behaviors
- **Visual realism vs. behavioral realism?**
 - these need to 'match'

Need to keep it 'Alive'



- **Believable eyes and eye motion**
 - Eyes are always moving, if just a little
 - Eye 'tracking'
 - Eye 'blinks'
- **Head motion**
 - Always moving, head 'follows' the eyes
- **Appropriate expressions**
- **'Good' synchronized speech**

Real Time Model



Screen shot - synchronized to real speech



Fred Parke ~ 1995

Telepresence



Low bandwidth 'video' conferencing

- Model based compression
- Model parameters extracted for transmission
- Only parameters sent over communication channel
- For reception, parameters drive model to recreate the facial images

Part of the MPEG-4 standard

Perception Research



- **Carefully controlled visual stimuli**
 - must to be 'correct'
- **Bi-Modal visual speech example**
 - Massaro & Cohen, UCSC
 - visual perception and aural perception work together
 - conflicts in visual and aural can induce misperceptions – McGurk effect
 - what you see can influence what you 'hear'

Medical and Educational



Medical

- Teaching anatomy
- Surgical simulation
- Model must be physically correct

Educational

- 'Tutor'
- Face must be interactive and engaging

Good Enough?



When will facial animation be good enough?

- Any face, any age, any expression, dramatic nuances, wide range of facial styles, 'easy' ...
- Visual and behavioral realism balanced

Appearance is getting very good, but not quite there yet – still hard to do well

Behavior modeling has a long way to go

Facial 'Turing' test

Future History



Looking ahead...

Just the Beginning!

- Animation only last 100 years
- Computer facial animation only last 35 years
- Most work in the last 10 years

Computation Capabilities

- 1,000 fold increase every 15 years!

Directions



- Much, much better models & tools
 - Subtle, more realistic detail and control
 - Behaviors, motivations
 - Idiosyncratic personality models
 - 'Director' level interfaces
-
- Something new – unexpected!

'Motivated' Facial Models



Action and expressions motivated by the character model, the situation, and the 'director' rather than manipulated by an animator

Fully Functional 'Actors'



Facial animation fully integrated

Anatomically 'correct'

Behavior driven

with personality, motivation

'Directable'

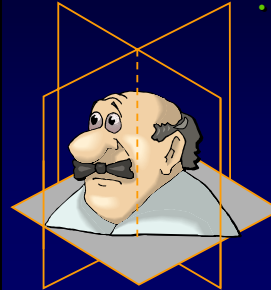
'Easy' to use



Anatomy of the Human Head

Jörg Haber
MPI Informatik

Terminology



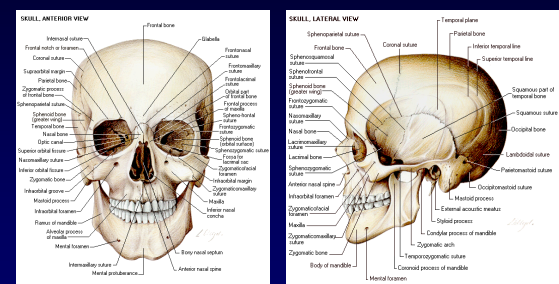
- positions of body parts are described relative to:
 - median (sagittal) plane:** vertical plane that divides the body into equal left and right halves; *medial / lateral* ⇔ closer to / further away from median plane
 - coronal plane:** vertical plane that divides the body into front and back halves; (*anterior / posterior*)
 - transverse (horizontal) plane:** any plane perpendicular to both median and coronal planes

The Human Head

Components of the human head:

- skull (lat. *cranium*)
- facial muscles (lat. *m. faciales et masticatores*)
- skin (lat. *integumentum commune*)
- eyes (lat. *oculi*)
- teeth (lat. *dentes*)
- tongue (lat. *lingua*)

Skull



Images: www.humanmuscles.8k.com

Skull

- cranium** (lat. *neurocranium*):
 - 7 bones; rigidly connected; lodges and protects brain and eyeballs; consists of *calvaria* and *cranial base*
- facial skeleton** (lat. *viscerocranium*):
 - 15 small bones that surround nasal and oral cavity mosaic-like; only the *mandible* (lat. *mandibula*) is movable
- bones of the skull are relocatable during birth, ossification completed at the age of 18 ⇒ proportions & shape of the skull change during growth

Facial Muscles

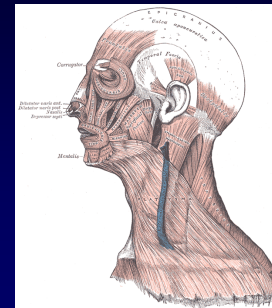
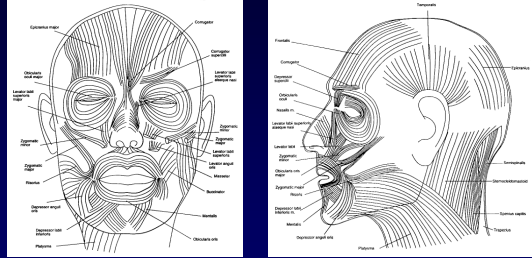


Image: Gray: "Anatomy of the Human Body" (1916)

Three groups:

- m. of facial expression:** two layers (superficial and deep)
- m. of mastication:** movement of the mandible
- epicranium:** tension / relaxation of facial skin

Facial Muscles



muscles connect a) two bones, b) bone and skin / muscle, or c) two different skin / muscle regions

Images: Parke/Waters: "Computer Facial Animation" (1996)

Types of Facial Muscles

- **sphincters**: contract radially towards a center point, e.g. *orbicularis oris*, *orbicularis oculi*
- **linear (parallel) muscles**: contract longitudinally towards their origin, e.g. *levator labii sup.*, *zygomaticus minor/major*
- **sheet muscles**: composed of several linear muscles side-by-side, e.g. *frontalis*

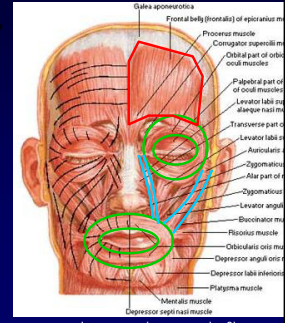


Image: www.humanmuscles.8k.com

Skin

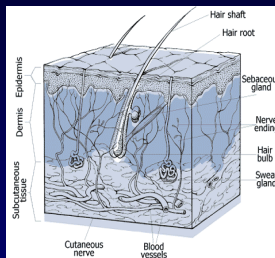
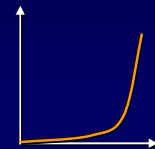


Image: www.humanmuscles.8k.com

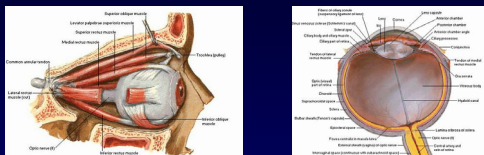
- **epidermis**: 0.02 mm thick, no vessels / glands, 5 layers of keratin
- **dermis**: 0.3-2.4 mm thick, 2 layers of soft connective tissue containing elastin fibers, blood and lymphatic vessels, and nerves
- **hypodermis (subcutis)**: adipose tissue built from collagen / fat cells, blood vessels, and nerves

Mechanical Properties of Skin

- skin composed of various layers with different elastic and viscous characteristics \Rightarrow skin exhibits significant **visco-elastic properties** (e.g. hysteresis, creep)
- skin has highly **non-linear stress-strain curve**:
 - low stress \Rightarrow low resistance against deformation (*collagen fibers unroll and stretch*)
 - high stress \Rightarrow sharp increase in resistance (*collagen fibers are completely stretched*)



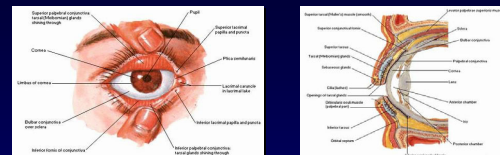
Eyes



- complex organ consisting of **eyeball** (lat. *bulbus oculi*) and **optic nerve**, embedded into the **skeletal**
- eyeball composed from **lens** and **vitreous body** (lat. *corpus vitreum*), enclosed by three concentric layers: **sclera / cornea**, **choroidea / iris**, and **retina**

Images: www.humanmuscles.8k.com

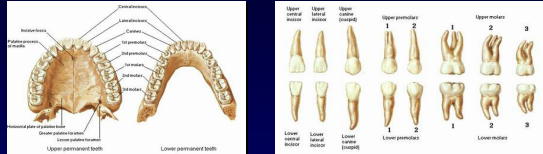
Eyes



- **eye muscles**: alignment of optical axis (external), focussing and adaptation to brightness (internal)
- **eyelids, connective tissue**: protect from contaminants
- **lachrymal**: secretion of tears to smooth the cornea, facilitate the motion of the eyeball, and wash away dust particles

Images: www.humanmuscles.8k.com

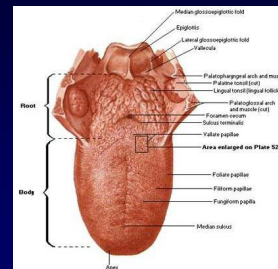
Teeth



- embedded into upper jaw (lat. *maxilla*) and lower jaw (lat. *mandibula*)
- 20 **milk teeth** are replaced gradually with 32 **permanent teeth** starting at the age of about six
- are used to chop up and squelch food, and for articulation

Images: www.humanmuscles.8k.com

Tongue



- consists of muscle tissue, nerves, blood vessels, and sensory cells (embedded in mucous membrane)
- can alter its shape and position in many ways
- most important sense organ for taste: sweet (tip), salty (front sides), bitter (back)
- support during chewing and swallowing
- use for articulation is learnt

Image: www.humanmuscles.8k.com

All that stuff...

Is it necessary to know all those details?

- it depends on the desired quality / realism of the head model:
 - *the more realism you want, the more precisely you have to simulate anatomy*
- at least: we need to know about the shape / structure / position of facial components and their interactions
- ... so don't be afraid to spend some money on medical textbooks or atlases



Overview Facial Animation Techniques

Volker Blanz
MPI Informatik

Facial Animation



Performance Driven

- Transfer performance of human actor to synthetic face model

Synthetic Motion

- From Text, Audio or defined by an Artist

Complete Script vs. Interactive Animation

Facial Animation: Two Levels



1. Dynamics of motion (temporal domain)

- Feature point coordinates $x_i(t)$
- Muscle contractions $c_i(t)$
- Action Units (AU, Ekman and Friesen 78) $a_i(t)$

2. Surface Deformation (spatial domain)

- Displacements of vertices of a high-resolution mesh
- Generate wrinkles
- May be solved statically at each moment t .

Dynamics of Motion: Performance Driven Animation



Performance of an Actor

- Tracking of marker points attached to skin
- Tracking of facial features

Feature Point i : $x_i(t)$

Performance-driven Animation

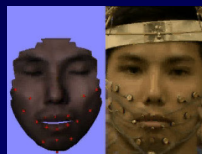


Acquisition of animation parameters

- specialized hardware (mechanical / electrical) transfers "deformation" of the human face to a synthetic face model



Virtual Actor system by SimGraphics (1994)



Movie: www.his.atr.co.jp/~kurate/movie/

Performance Driven Animation



Acquisition of animation parameters:

- video camera + software (\rightarrow computer vision)
- capture head movements, identify eyes and mouth, detect viewing direction and mouth configuration, control synthetic head model with these parameters



Movies: baback.media.mit.edu/~irfan/DFACE_demo/tracking.html

Dynamics of Motion: Voice Puppetry



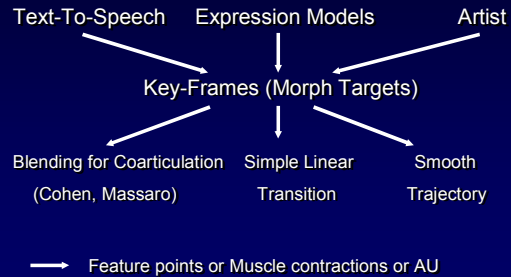
Brand, Siggraph99

Audio

- Hidden Markov Model
 - Trained from Video & Audio data

26 Feature Points $i: x_i(t)$

Dynamics of Motion: Key-Frame Animation



Key Frame Animation



Types of interpolation:

- convex combination (linear int., blending, morphing):
$$v = \alpha \cdot v_1 + (1 - \alpha) \cdot v_2 \quad (0 \leq \alpha \leq 1)$$

 v : scalar or vector (position, color,...)
- non-linear interpolation: e.g. trigonometric functions, splines, ...; useful for displaying dynamics (acceleration, slow-down)
- segmental interpolation: different interpolation values / types for independent regions (e.g. eyes, mouth);
⇒ decoupling of emotion and speech animation

Surface Deformations Main Approaches



1. Parametric Models

2. Physics-based Animation

3. Learning-Based Animation

- Image-Based
- 3D Models

Direct Parameterization



Idea:

- perform facial animation using a **set of control parameters** that manipulate (local) regions / features

What parameterization should be used?

- ideal universal parameterization:
 - small set of intuitive control parameters
 - any possible face with any possible expression can be specified

Parametric Models I



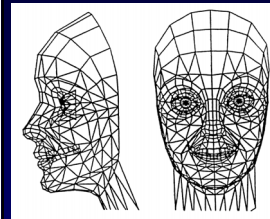
- F. I. Parke: "Parameterized Models for Facial Animation", IEEE CGA, 2(9):61-68, Nov. 1982
 - 10 control parameters for facial expressions
 - ~20 parameters for definition of facial conformation
- K. Waters: "A Muscle Model for Animating Three-Dimensional Facial Expression", SIGGRAPH '87, pp. 17-24, July 1987
 - deforms skin using "muscle vectors"

Parametric Models II



- N. Magnenat-Thalmann et al.: "Abstract Muscle Action Procedures for Human Face Animation", *The Visual Computer*, 3(5):290-297, March 1988
 - pseudo muscles based on empirical models
 - muscle actions are (complex) combinations of FACS action units
- J. E. Chadwick et al.: "Layered Construction for Deformable Animated Characters", SIGGRAPH '89, pp. 243-252, July 1989
 - freeform deformations (FFD), pseudo muscles

Parke's Parametric Face Model



- polygonal face mesh (~300 triangles + quads), symmetrical, edges aligned to facial feature lines
- two types of parameters:
 - 10 expression parameters
 - about 20 conformation parameters
- five different ways how parameters modify facial geometry

Parke: Expression Parameters



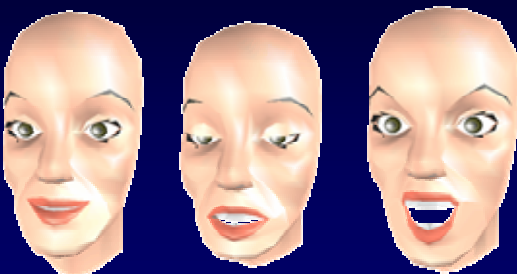
- eyes:
 - dilation of pupils, opening / closing of eyelids, position and shape of eyebrows, viewing direction
- mouth:
 - rotation of mandible, width and shape of the mouth, position of upper lip, position of mouth corners
- additional parameters (suggested):
 - head rotation, size of nostrils

Parke: Conformation Parameters

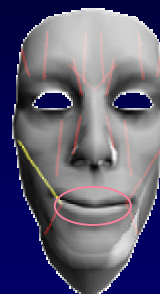


- aspect ratio of the face
- length and shape of the neck
- shape (= relative position of assigned vertices) of chin, forehead, cheeks, and cheekbones
- size of eyelids, eyeballs, iris; position of the eyes
- jaw width
- length of the nose; width of nose bridge and nostril
- relative size of chin, forehead, and mouth-nose-eyes-part w.r.t. remaining face parts
- color of skin, eyebrows, iris, and lips

Parke: Results



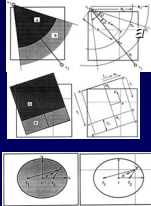
The Face Model by Waters



- polygonal face mesh:
 - 201 quads + 35 triangles
- 10 different muscles:
 - 9 linear muscles (symmetrical left/right)
 - 1 sphincter (orbicularis oris)
- additional parameters:
 - jaw rotation
 - viewing direction
 - opening of eyelids

Waters: Muscle Vectors

- muscles are represented by **muscle vectors**, which describe the effect of muscle contraction on the geometry of the skin surface
- muscle vectors are composed of:
 - a point of attachment and direction (for linear muscles)
 - a line of attachment and a direction (for sheet muscles)
 - a center point and two semi-axes defining an ellipse (for sphincters)



Images: Waters: "A Muscle Model for Animating Three-Dimensional Facial Expression" (1987)

Physics-based Models

Idea:

- represent and manipulate expressions based on **physical characteristics** of skin tissue and muscles

Real anatomy is too complex!

- no facial animation system has represented and simulated the complete, detailed anatomy of the human head yet.
- reduce complexity to obtain animatable model
- need to build appropriate models for muscles and skin tissue

Skin Tissue Mechanics

Viscoelastic response to stress / strain

- Elastic properties:
 - returns to rest shape when load is removed.
 - Non-linear relationship
 - Model: spring
- Viscous Properties
 - Energy is absorbed
 - Model: damper

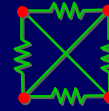
Mass-Spring Networks

- common technique for simulating dynamics of skin
- vertices = **mass points**, edges = **springs**
- Lagrangian equations of motion are integrated over time using numerical algorithms
- several variants with multiple layers of mass-spring networks (2D or 3D)

2D:



3D: tetrahedron



cube

Finite Element Method

- numerical technique for simulating deformation and flow processes (crash tests, weather forecast, ...); frequently used for surgery planning
- partitioning into 3D elements (tetrahedra, cubes, prisms,...)
- continuity conditions between elements are collected in global stiffness matrix M
⇒ time-consuming solution for high dimensional M

Learning-based Techniques

Observe facial deformations, Ignore underlying mechanisms

- Record keyframe shapes from
 - Images or Video (Multiple Views)
 - 3D Scans
- Keyframes reproduce natural appearance in a photorealistic way
 - Use *morphing* for smooth transitions between keyframes.





Parameterized Face Models

Fred Parke
Texas A&M University

What's the Goal?



All possible faces?
A specific face?
Realistic faces?
Caricature faces?
Fantasy faces?

Facial Attributes

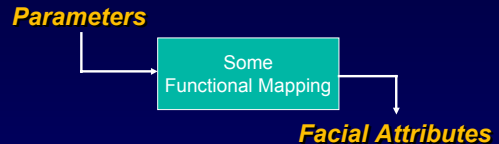


Facial conformation
Facial expression posture – shape
Head orientation, eye gaze
Skin texture, shading
Hair characteristics
Mouth/speech attributes
jaw rotation, lip and tongue shape, teeth,...

Facial Animation Control



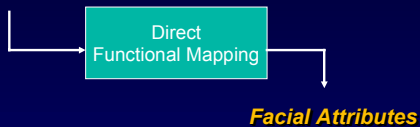
We can view all facial control systems as parameterizations



Direct Parameterizations



Parameters

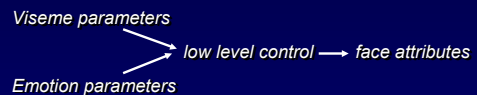


Where the functional mapping primarily consists of interpolations, affine transformations, translations, and generative procedures applied to subsets of the surface control points

Second Level Parameterizations



- Higher level parameters which allow specification and control of expressions, visemes, ...*
- Built on top of lower level parameterizations*
- Speech animation one example*



Universal Parameterization



Allows specification of any expression and facial attribute set, for any possible face

Don't exist yet

A lot of work on expression parameters

- FACS provides one basis

Not much work on conformation parameters

- Anthropometry, principle component analysis

Parameter Orthogonality



Expression parameters control expression for a given face

Conformation parameters select or specify a specific face from the universe of possible face

Should be orthogonal

- Manipulating expression should not effect conformation
- Manipulating conformation should not effect expression

Facial Expressions



Capable facial models allow wide range of expression

Including the universal expressions

- anger, fear, surprise, disgust, happiness, sadness

Capable facial animation are able to express and convey 'emotion'

Posture and expression display emotion

FACS



Facial Action Coding System

Developed by Ekman and Friesen to study and quantify facial expression across cultures

Consists of about 66 'facial actions'

While not intended, has been adopted by the facial animation community as an effective expression parameterization scheme

Animation Control Methods



- *Interpolation of expression poses*
- *Interpolation of control parameters to drive a parameterized model*
- *Emulation of muscle actions based on interpolated muscle parameters*

Shape Interpolation



Earliest (simplest) Animation Technique

Simple interpolation of entire face

- *earliest animation technique*

Interpolation of 'independent' facial regions

- *upper face, lower face - Kleiser 1989*

Interpolation in n-dimensional face spaces

Expression Interpolation

Various expression poses between two extremes

1 dimensional space, 3 parameters



2 dimensional pose space



6 parameters

Poses within a 2 dimensional interpolation space



Extrapolation in pose space



Interpolated Faces - 1973

Interpolation between individual faces



Parameterized Model - 1974

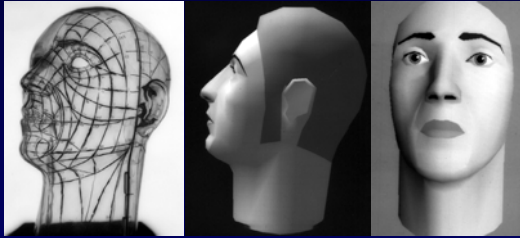
- F. Parke - University of Utah

Example Images



Expression and Conformation Control

Starting Point



Parameterized Model - 1974

About 50 parameters, ~10 most useful
Speech 'enabled'



Facial Features

Eyes, eyelashes

Lips, tongue, teeth and mouth interior

Skin, Hair

Nose, Ears

Most important features?

Expression Parameters

Eye region

- Eyebrow arch, separation
- Eyelid opening
- Eyeball size, eye gaze
- Pupil size, iris size

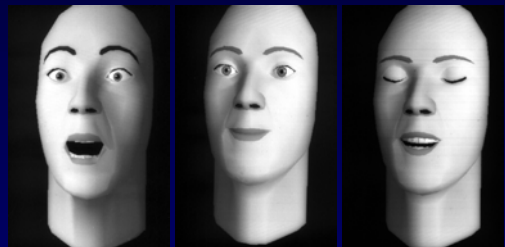
Mouth region

- Jaw rotation
- Mouth expression, width
- Upper lip position
- Control of mouth corners

Expression parameters



Expression parameters



Conformation Parameters



Shape

- Forehead, cheek, neck

Color – skin, eye, lips, teeth

Scaling

- Head scaling, eyelid scaling
- Widths of jaw, cheek, nose

Facial proportions

- Eye to forehead, chin to mouth, chin to eye

Offset

- Eyebrows, chin, end of nose, teeth

Conformation parameters



Conformation parameters



Implementation Techniques



Generative procedures

- Eyeballs, eyelids, eye gaze

Shape interpolation

- Forehead, cheeks, eyebrows, mouth expression

Transformations

- Aspect ratio and proportions of head and features such as nose, jaw, chin

Translation

- Chin, end of nose, eyebrows

Parameterized Model - 1974



Example Animation



Interactive parameterized model



~1990 on SGI with GL – F. Parke



Range of expression



Range of expression



Speech Animation



Support speech postures

- About 45 English phonemes
- 18 or so visually distinct speech postures

Synchronize postures to speech track

With coarticulation and expression overlays

Parameterized Model - 1974



Speech Synchronized Animation – F. Parke



Most Useful Speech Parameters



Lip Animation

- Jaw rotation
- Upper lip position
- Mouth width

Expression Animation

- Mouth expression, eye tracking
- Eyebrow arch, separation
- Eyelid opening, pupil size

Parameterized Model - 1974



Speech with a little more expression



Parameterized Model - 1982

Speech animation for a specific character

Expression and speech only – F. Parke



Coarticulation

Mouth posture influenced by phonemes prior to and after current phoneme

Mouth shape blends across phonemes

Due to dynamic motion limits

May span up to five phonemes

– see Pelachaud, et al - 1991

Eye Actions

Eye blinks

- keep eye wet
- synchronized with speech
- follow pause in speech
- listener blinks also synced to speaker

Eye gaze

- eye contact - allowed contact culturally dependent, degree of intimacy
- can communicate intention, ...

Pupil size

- reflects attitude, emotional state

Dialogue Mouth Action (Disney)

Action Leading Dialogue

- accent eyes lead sound by 2 to 5 frames - stronger accents have longer lead
- sync eye blinks should lead by 3 to 4 frames
- anticipate initial slow moves by 3 to 8 frames

Holds

- at end of phrase, retain mouth expression
- use "moving hold" on long mouth pose

Dialogue Mouth Action (Disney)

- The vowel sounds A, E, I, O, U always require some mouth opening
- The consonants B, M, P are all closed mouth
- T and G can also pucker like a U; Y and W can go into a very small O or U shape
- F and V lower lip under upper teeth
- E sounds generally show teeth
- 'White' teeth flash

Lip Sync (Madsen)

analyze speech track

- determine overall length, pauses, etc.

identify 'key frames'

- look for accented syllables, the b's, m's, and p's
- look for phonemes with distinctive shapes; oval o's and w's
- Consonants are the accents, need to be accurate
- locate frames where the lips meet

approximate the rest

Lip Movements (Madsen)

Realistic characters are the greatest challenge

- *invite comparison with real people*

For cartoon characters

- *simplicity is secret of success*
- *attempts at extreme accuracy appear forced and unnatural*

Head Tilt Angle (Blair)

Head angle, direction of 'look', and head motion relative to body all contribute to expression

Example - a hand puppet depends mostly on head tilt and body posture without any phonetic mouthing or facial action

Changes in head tilt or head turns convey different emotions

- *affirmative 'nod', negative sideways shake, ...*

Automated Synchronization

Text Driven

- *Synthesize speech audio and face images together*
- *Based on text-to-speech systems*

Speech Driven

- *Analysis of speech audio track for pauses, visemes*
 - *Simple energy tracking*
 - *Speech recognition acoustic preprocessor*
 - *LPC analysis – speech classification*
 - *Neural nets*

Automatic Lip Sync - 1987

Lewis and Parke



Emotional Overlays

Conversation always has emotional content

Facial expressions of emotion

- *'affect displays'*

Emotion includes visceral and muscular physiological responses

- *muscle tension*
- *variations in vocal tract*

Non-Emotional Overlays

Conversation Signals - illustrators - punctuate speech

- *eyebrows*

Punctuators - movements that occur at pauses

- *correspond to commas, periods, exclamation points*

Regulators - control speaker turn taking

- *speaker-turn-signals*
- *speaker state signals*
- *speaker within turn*
- *speaker continuation*

Muscle Based Parameterizations



Parameters control the face through functions which emulate or simulate muscle actions

*K. Waters – 1987
Thalmann, et al – 1988*

and many others since

Waters' Muscle Model - 1987



Models muscle induced displacement with geometric distortion functions which include first order elastic tissue properties

Three kinds of muscle functions

– Linear, sphincter, and sheet

The muscle functions are located and aligned independently of the skin geometry

They have defined regions of influence

Abstract Muscle Action Model



– Thalmann, et al, 1988

Empirical pseudomuscle action procedures

Each works on a specific region of the face

Each emulates a muscle or group of closely related muscles

Loosely patterned after FACS actions

Groups of 'actions' form 'expressions'

- such as 'emotions' and phonemes*

Principle Component Analysis



Use of principle component analysis to extract 'conformation parameters' from a data base of digitized real faces.

– Blanz and Vetter – 1999

The principle components become the parameters to specify a specific face

- Not an 'intuitive' parameter space*
- Requires an optimizing search to match a face*

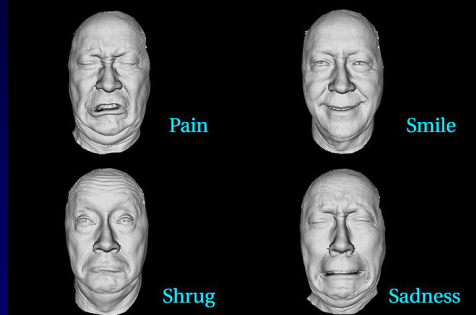


Facial Performance Capture

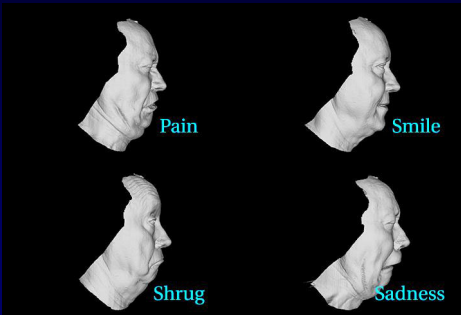
Lance Williams

Walt Disney Feature Animation

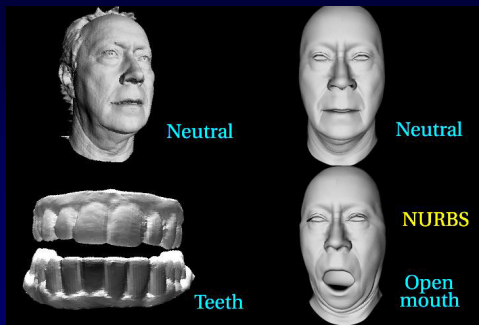
Laser-scanned facial expressions



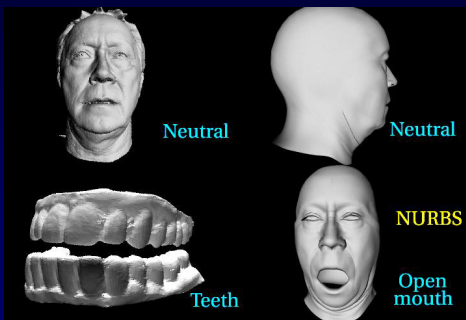
Laser-scanned facial expressions



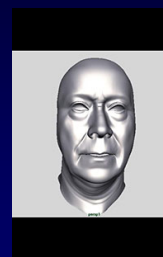
NURBS models sculpted from scan data



NURBS models sculpted from scan data



NURBS model sculpted from scan data



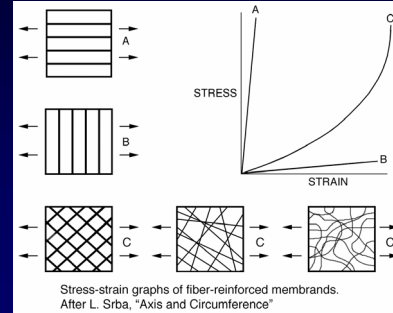
Model muscle blendshapes to fit scanned expressions



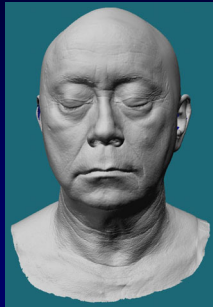
Expression: a linear superposition of shapes

- Approximately 60 blendshapes in facial model
 - Jaw and eyelid rotations are *piecewise linear*
 - Some “multitarget” blendshapes are used
- Generic muscles are posed to match expressions
- Differences are mapped by back propagation
- Process iterates through expressions repeatedly.

Linear regime for skin



Life mask scan data



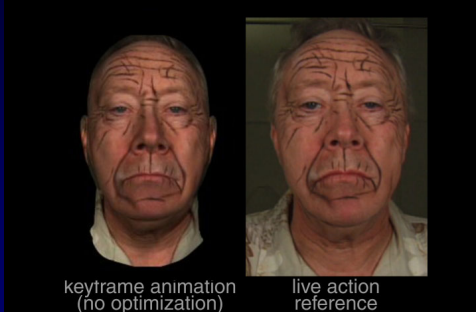
Mapping detail from lifemask scan



NURBS model conformed to scan data

- Approximately 3.5 million polygons in life mask scan
- NURBS model matched to rigid-body transformation
- Muscle blendshapes are posed to match expression
- NURBS CVs are sculpted to match model to scan
- Difference is extracted as a displacement map

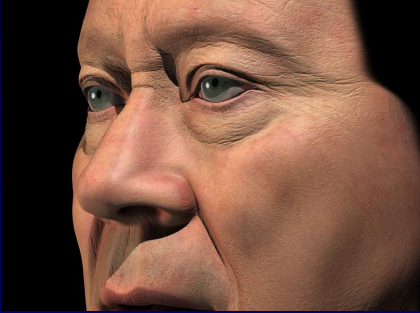
Model and image



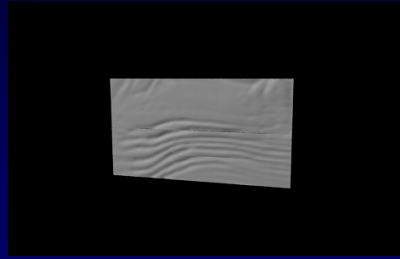
Fine hairs



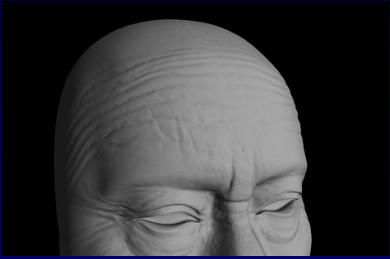
Pores and wrinkles



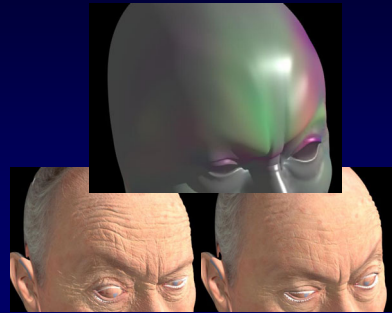
Modulated wrinkles



Modulated wrinkles



Modulated wrinkles



Modulated wrinkles

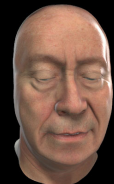


Modulated wrinkles

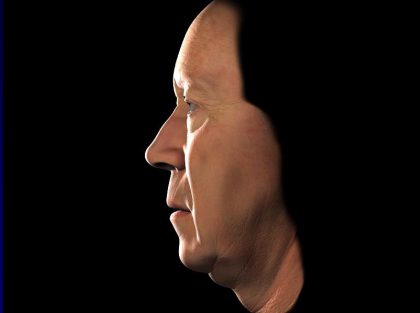


xinmin_test_4_17_wrinkles

xinmin_test_4_17_No_wrinkles



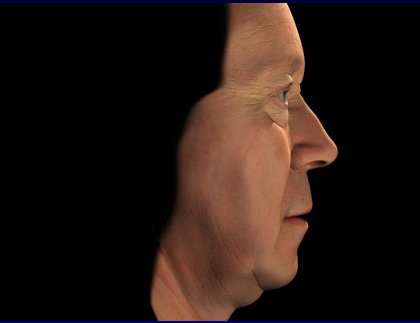
Marschner et. al.
BRDF skin model



Marschner et. al.
BRDF skin model



Marschner et. al.
BRDF skin model



Environment mapping



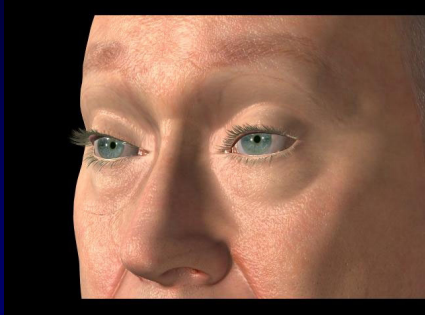
Environment mapping



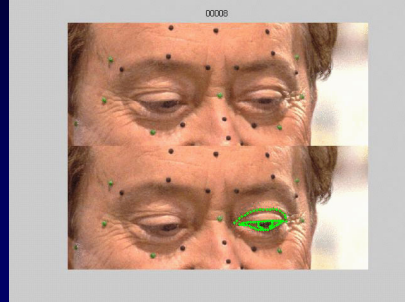
Environment mapping



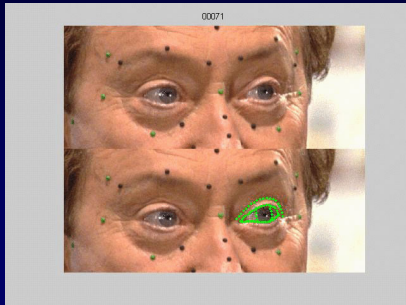
Bone constraints



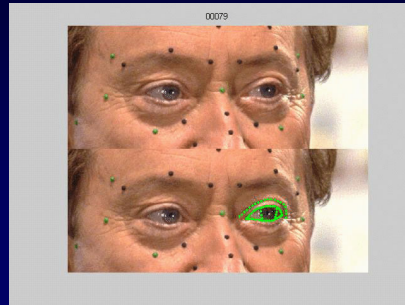
Eye tracking



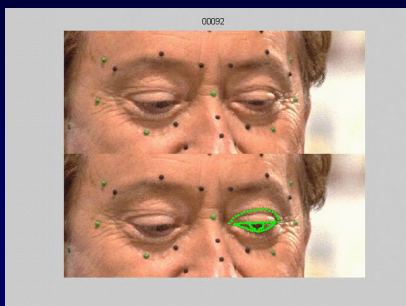
Eye tracking



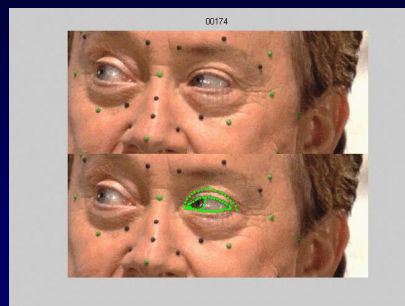
Eye tracking



Eye tracking



Eye tracking



Eye tracking



Image and model



Model and image



Tracking registration



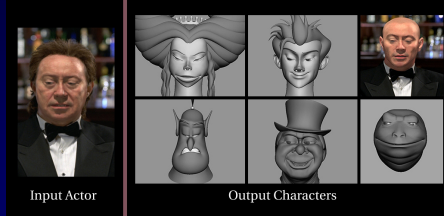
Together at last



Double take



Cross mapping performance



Next: automatic modeling, markerless tracking



Modeling



Hiroki Itokazu
"Hirokimation"

Tracking

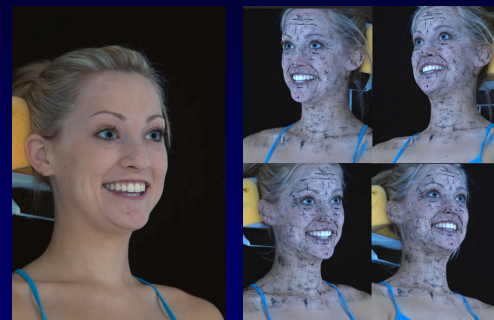


Xinmin Zhao
Numerical Optimization

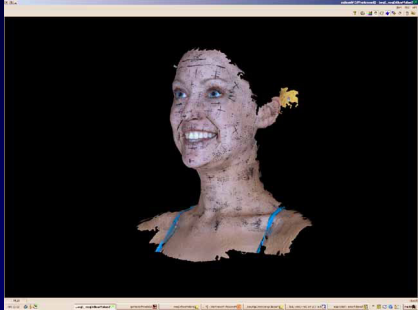
Motion and Emotion



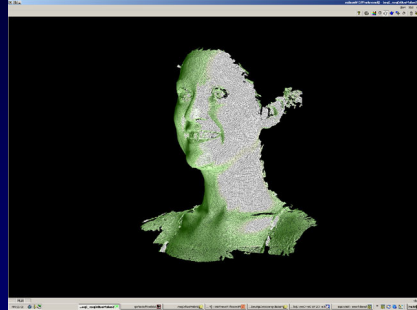
Flash capture: texture and model



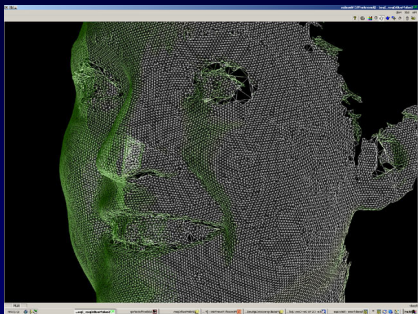
Captured model



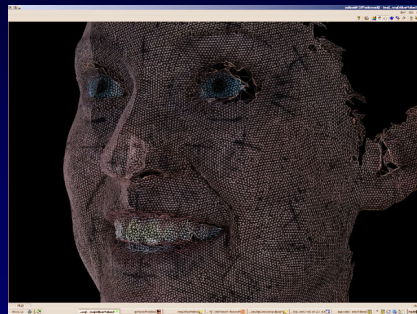
Captured model



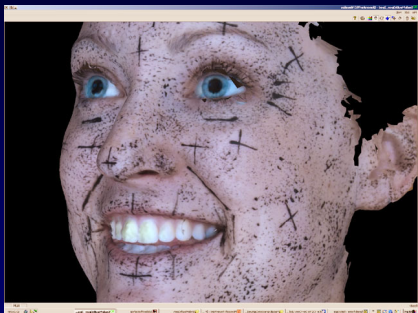
Captured model



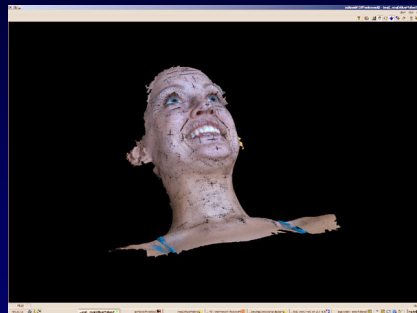
Captured model



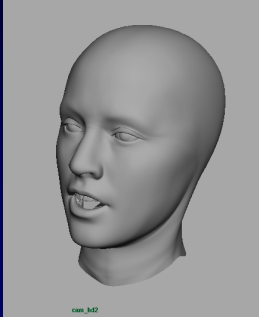
Captured model



Captured model



NURBS model



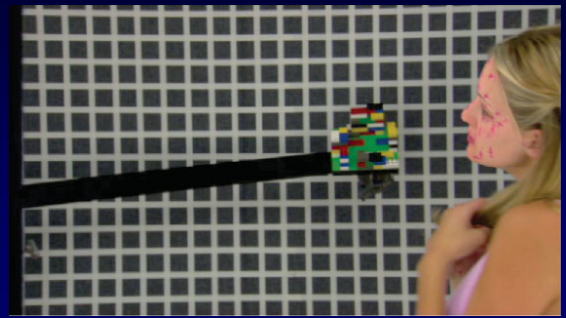
Tracking cameras



Matching images with model



Matching images with model



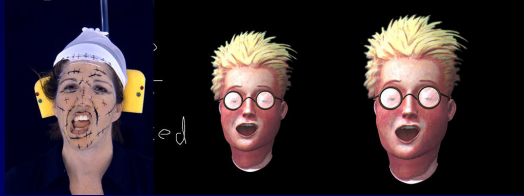
Voiceover camera setup



Performing with marks



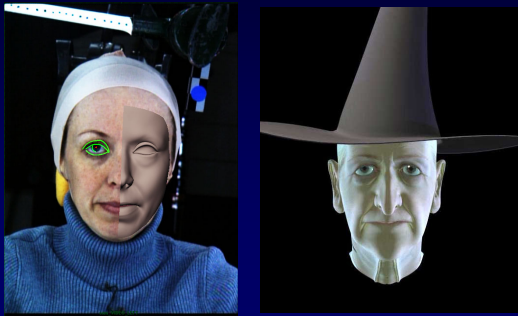
Driving facial animation



Automatic modeling, markerless tracking.



Driving facial animation

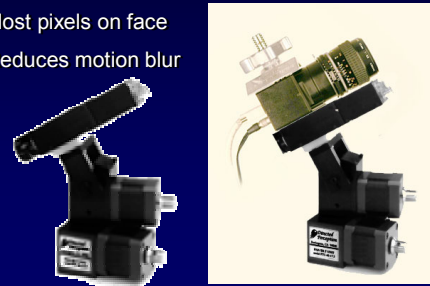


Tracking cameras



Servo camera on head marker

- Most pixels on face
- Reduces motion blur



Tracking cameras



Servo camera on head marker

Ross Lamm, Perceptivu Inc.

- Most pixels on face
- Reduces motion blur



- Azimuth / elevation:

Helmet cameras



2-camera helmetcam

- All pixels on face
- Eliminates motion blur
- Reduced room lights OR camera-mounted ring lights



Helmet cameras



2-camera helmetcam

- All pixels on face
- Eliminates motion blur
- Reduced room lights OR camera-mounted ring lights



Helmet cameras



2-camera helmetcam

- All pixels on face
- Eliminates motion blur
- Reduced room lights OR camera-mounted ring lights





Physics-Based Facial Modeling and Animation

Demetri Terzopoulos

New York University
University of Toronto

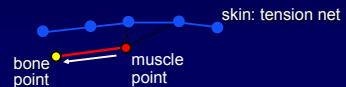
Platt & Badler, 1981

"Animating Facial Expression"



Pseudo-physical approach

- Muscle represented as group of fiber
- Contraction displaces muscle point
- Distribute "forces" → displace skin nodes



- Skin as an infinitesimally thin surface

A Physics-Based Face Model

[Terzopoulos & Waters 1990]



A Physics-Based Face Model

(Terzopoulos & Waters 1990)



Hierarchical structure

- Expression: Facial action coding system (FACS)
- Control: Coordinated facial actuator commands
- Muscles: Contractile muscle fibers exert forces
- Physics: Muscle forces deform 3D synthetic tissue
- Geometry: Expressive facial deformations
- Images: Rendering by graphics pipeline

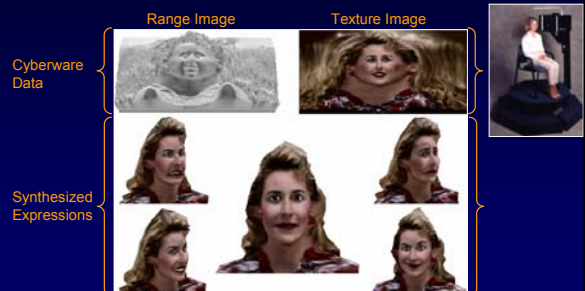
Physics-Based Facial Modeling

(Terzopoulos & Waters 1990)

(Lee & Terzopoulos 2002)



Artificial Humans Scanned Data → Synthetic Faces



Raw Input Dataset ("Heidi")

From CyberWare 3D Color Digitizer

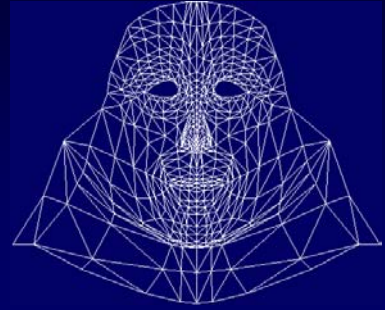


Range Image



RGB Texture Image

Generic Facial Mesh



Fitting the Generic Mesh

Feature-based image matching algorithm

localizes facial features in:

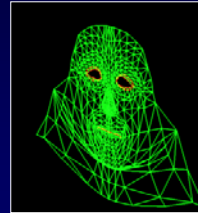
Processed range image

RGB texture image



Sampling Facial Shape

Fitted mesh nodes sample range data



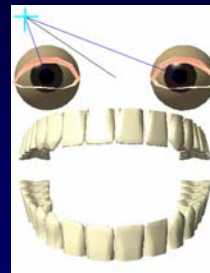
Textured 3D Geometric Model

Texture map coordinates

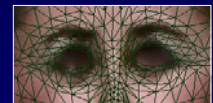
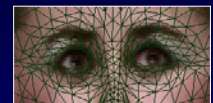
- Positions of fitted mesh nodes in RGB texture image



Auxiliary Geometric Models



Eyelid Texture Interpolation



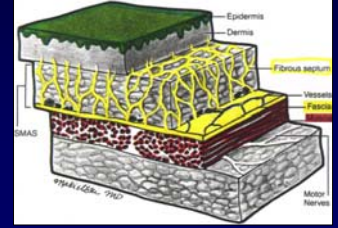
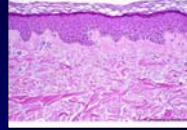
Complete Geometric Model

Neutral expression
is estimated



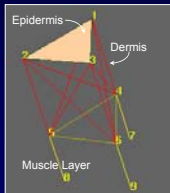
Facial Histology

A complex, multilayer structure

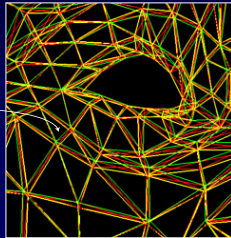


Biomechanical Skin Model

Deformable tissue element

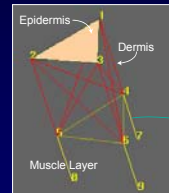


Single Element

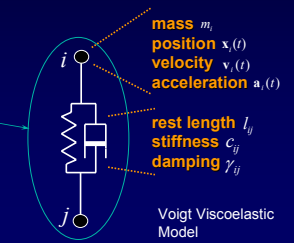


Biomechanical Skin Model

Viscoelastic uniaxial primitive



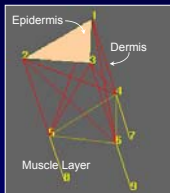
Single Element



Voigt Viscoelastic Model

Biomechanical Skin Model

Element dynamics

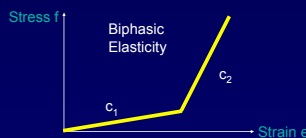


Single Element

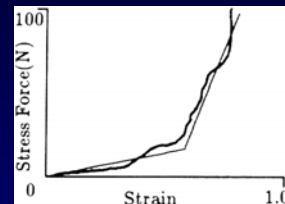
$$\mathbf{r}_{ij} = \mathbf{x}_j - \mathbf{x}_i \quad \text{Span}$$

$$e_{ij} = \frac{\|\mathbf{r}_{ij}\|}{l_{ij}} - 1 \quad \text{Deformation}$$

$$\mathbf{f}_{ij}^e = (c_{ij} e_{ij} + \gamma_{ij} \dot{e}_{ij}) \frac{\mathbf{r}_{ij}}{\|\mathbf{r}_{ij}\|} \quad \text{Viscoelastic Force}$$



Empirical Stress-Strain Curve



- Can represent $c(e)$ as a lookup table

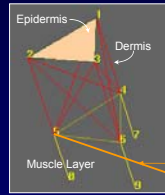
Langer's Lines

Non-isotropic stress-strain characteristics



Biomechanical Skin Model

Element dynamics



$$\mathbf{r}_{ij} = \mathbf{x}_j - \mathbf{x}_i \quad \text{Span}$$

$$\mathbf{e}_{ij} = \frac{\mathbf{r}_{ij}}{\|\mathbf{r}_{ij}\|} - \mathbf{l}_{ij} \quad \text{Deformation}$$

$$\mathbf{f}_{ij}^e = (c_{ij} \mathbf{e}_{ij} + \gamma_{ij} \dot{\mathbf{e}}_{ij}) \frac{\mathbf{r}_{ij}}{\|\mathbf{r}_{ij}\|} \quad \text{Viscoelastic Force}$$

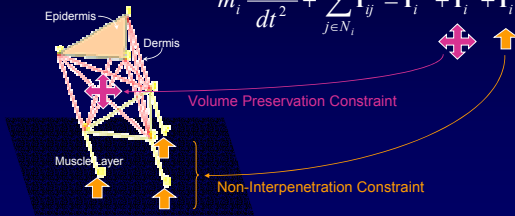
Differential Equations of Motion

$$m_i \frac{d^2 \mathbf{x}_i}{dt^2} + \sum_{j \in N_i} \mathbf{f}_{ij}^e = \mathbf{f}_i^m \quad \text{Muscle Forces}$$

Biomechanical Skin Model

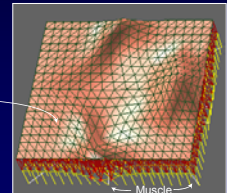
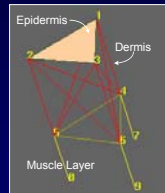
Element dynamics

$$m_i \frac{d^2 \mathbf{x}_i}{dt^2} + \sum_{j \in N_i} \mathbf{f}_{ij}^e = \mathbf{f}_i^m + \mathbf{f}_i^v + \mathbf{f}_i^c$$



Biomechanical Skin Model

Deformable tissue element and patch



Single Element

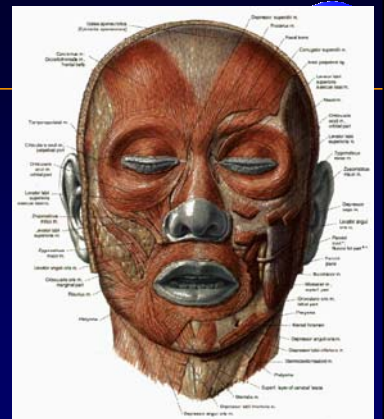
Skin Patch

Explicit Euler Time Integration Method

Efficient near stability limit for moderately deformable biomechanical skin model

$$\begin{cases} \mathbf{a}_i^t = \frac{1}{m_i} \left(\sum_{j \in N_i} \mathbf{f}_{ij}^e + \mathbf{f}_i^m + \mathbf{f}_i^v + \mathbf{f}_i^c \right) \\ \dot{\mathbf{x}}_i^{t+dt} = \dot{\mathbf{x}}_i^t + dt \mathbf{a}_i^t \\ \mathbf{x}_i^{t+dt} = \mathbf{x}_i^t + dt \dot{\mathbf{x}}_i^{t+dt} \end{cases}$$

Facial Musculature



Muscle-Actuated Expressions



Muscle-Actuated Expressions



Raw CyberScans



Functional Model of George



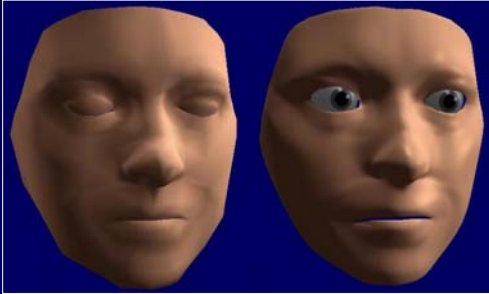
George in "Bureaucrat Too"



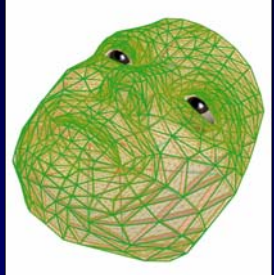
"Bureaucrat Too" (excerpt)



Facial Subdivision Surface



Facial Subdivision Surface



Interactive, Physics-Based Facial Animation



Runs at >30 fps on a dual 1.5 GHz AMD Athlon system



Real-Time Facial Simulation



Better Muscle Modeling

[Kahler, Haber, Seidel 2001]



Closer to the real thing

- Muscles insert into simulated skin tissue
- Forces applied by contraction
- Muscle geometry defines zone of influence
- Still: Muscle deformation itself is purely geometric
 - *Varying degree of realism*



Realistic Modeling for Facial Animation

Yuencheng Lee¹, Demetri Terzopoulos¹, and Keith Waters²
University of Toronto¹ and Digital Equipment Corporation²

Abstract

A major unsolved problem in computer graphics is the construction and animation of realistic human facial models. Traditionally, facial models have been built painstakingly by manual digitization and animated by ad hoc parametrically controlled facial mesh deformations or kinematic approximation of muscle actions. Fortunately, animators are now able to digitize facial geometries through the use of scanning range sensors and animate them through the dynamic simulation of facial tissues and muscles. However, these techniques require considerable user input to construct facial models of individuals suitable for animation. In this paper, we present a methodology for automating this challenging task. Starting with a structured facial mesh, we develop algorithms that automatically construct functional models of the heads of human subjects from laser-scanned range and reflectance data. These algorithms automatically insert contractile muscles at anatomically correct positions within a dynamic skin model and root them in an estimated skull structure with a hinged jaw. They also synthesize functional eyes, eyelids, teeth, and a neck and fit them to the final model. The constructed face may be animated via muscle actuations. In this way, we create the most authentic and functional facial models of individuals available to date and demonstrate their use in facial animation.

CR Categories: I.3.5 [Computer Graphics]: Physically based modeling; I.3.7 [Computer Graphics]: Animation.

Additional Keywords: Physics-based Facial Modeling, Facial Animation, RGB/Range Scanners, Feature-Based Facial Adaptation, Texture Mapping, Discrete Deformable Models.

1 Introduction

Two decades have passed since Parke's pioneering work in animating faces [13]. In the span of time, significant effort has been devoted to the development of computational models of the human face for applications in such diverse areas as entertainment, low bandwidth teleconferencing, surgical facial planning, and virtual reality. However, the task of accurately modeling the expressive human face by computer remains a major challenge.

Traditionally, computer facial animation follows three basic procedures: (1) design a 3D facial mesh, (2) digitize the 3D mesh, and (3) animate the 3D mesh in a controlled fashion to simulate facial actions.

In procedure (1), it is desirable to have a refined topological mesh that captures the facial geometry. Often this entails digitizing

¹Department of Computer Science, 10 King's College Road, Toronto, ON, Canada, M5S 1A4. {vlee | dt}@cs.toronto.edu

²Cambridge Research Lab., One Kendall Square, Cambridge, MA 02139. waters@crl.dec.com

as many nodes as possible. Care must be taken not to oversample the surface because there is a trade-off between the number of nodes and the computational cost of the model. Consequently, meshes developed to date capture the salient features of the face with as few nodes as possible (see [17, 14, 21, 9, 23] for several different mesh designs).

In procedure (2), a general 3D digitization technique uses photogrammetry of several images of the face taken from different angles. A common technique is to place markers on the face that can be seen from two or more cameras. An alternative technique is to manually digitize a plaster cast of the face using manual 3D digitization devices such as orthogonal magnetic fields sound captors [9], or one to two photographs [9, 7, 1]. More recently, automated laser range finders can digitize on the order of 10^5 3D points from a solid object such as a person's head and shoulders in just a few seconds [23].

In procedure (3), an animator must decide which mesh nodes to articulate and how much they should be displaced in order to produce a specific facial expression. Various approaches have been proposed for deforming a facial mesh to produce facial expressions; for example, parameterized models [14, 15], control-point models [12, 7], kinematic muscle models [21, 9], a texture-map-assembly model [25], a spline model [11], feature-tracking models [24, 16], a finite element model [6], and dynamic muscle models [17, 20, 8, 3].

1.1 Our Approach

The goal of our work is to automate the challenging task of creating realistic facial models of individuals suitable for animation. We develop an algorithm that begins with cylindrical range and reflectance data acquired by a Cyberware scanner and automatically constructs an efficient and fully functional model of the subject's head, as shown in Plate 1. The algorithm is applicable to various individuals (Plate 2 shows the raw scans of several individuals). It proceeds in two steps:

In step 1, the algorithm adapts a well-structured face mesh from [21] to the range and reflectance data acquired by scanning the subject, thereby capturing the shape of the subject's face. This approach has significant advantages because it avoids repeated manual modification of control parameters to compensate for geometric variations in the facial features from person to person. More specifically, it allows the automatic placement of facial muscles and enables the use of a single control process across different facial models.

The generic face mesh is adapted automatically through an image analysis technique that searches for salient local minima and maxima in the range image of the subject. The search is directed according to the known relative positions of the nose, eyes, chin, ears, and other facial features with respect to the generic mesh. Facial muscle emergence and attachment points are also known relative to the generic mesh and are adapted automatically as the mesh is conformed to the scanned data.

In step 2, the algorithm elaborates the geometric model constructed in step 1 into a functional, physics-based model of the subject's face which is capable of facial expression, as shown in the lower portion of Plate 1.

We follow the physics-based facial modeling approach proposed

by Terzopoulos and Waters [20]. Its basic features are that it animates facial expressions by contracting synthetic muscles embedded in an anatomically motivated model of skin composed of three spring-mass layers. The physical simulation propagates the muscle forces through the physics-based synthetic skin thereby deforming the skin to produce facial expressions. Among the advantages of the physics-based approach are that it greatly enhances the degree of realism over purely geometric facial modeling approaches, while reducing the amount of work that must be done by the animator. It can be computationally efficient. It is also amenable to improvement, with an increase in computational expense, through the use of more sophisticated biomechanical models and more accurate numerical simulation methods.

We propose a more accurate biomechanical model for facial animation compared to previous models. We develop a new biomechanical facial skin model which is simpler and better than the one proposed in [20]. Furthermore, we argue that the skull is an important biomechanical structure with regard to facial expression [22]. To date, the skin-skull interface has been underemphasized in facial animation despite its importance in the vicinity of the articulate jaw; therefore we improve upon previous facial models by developing an algorithm to estimate the skull structure from the acquired range data, and prevent the synthesized facial skin from penetrating the skull.

Finally, our algorithm includes an articulated neck and synthesizes subsidiary organs, including eyes, eyelids, and teeth, which cannot be adequately imaged or resolved in the scanned data, but which are nonetheless crucial for realistic facial animation.

2 Generic Face Mesh and Mesh Adaptation

The first step of our approach to constructing functional facial models of individuals is to scan a subject using a Cyberware Color Digitizer™. The scanner rotates 360 degrees around the subject, who sits motionless on a stool as a laser stripe is projected onto the head and shoulders. Once the scan is complete, the device has acquired two registered images of the subject: a range image (Figure 1) — a topographic map that records the distance from the sensor to points on the facial surface, and a reflectance (RGB) image (Figure 2) — which registers the color of the surface at those points. The images are in cylindrical coordinates, with longitude (0–360) degrees along the x axis and vertical height along the y axis. The resolution of the images is typically 512×256 pixels (cf. Plate 1)

The remainder of this section describes an algorithm which reduces the acquired geometric and photometric data to an efficient geometric model of the subject’s head. The algorithm is a two-part process which repairs defects in the acquired images and conforms a generic facial mesh to the processed images using a feature-based matching scheme. The resulting mesh captures the facial geometry as a polygonal surface that can be texture mapped with the full resolution reflectance image, thereby maintaining a realistic facsimile of the subject’s face.

2.1 Image Processing

One of the problems of range data digitization is illustrated in Figure 1(a). In the hair area, in the chin area, nostril area, and even in the pupils, laser beams tend to disperse and the sensor observes no range value for these corresponding 3D surface points. We must correct for missing range and texture information.

We use a *relaxation method* to interpolate the range data. In particular, we apply a membrane interpolation method described in [18]. The relaxation interpolates values for the missing points so as to bring them into successively closer agreement with surrounding points by repeatedly indexing nearest neighbor values. Intuitively, it stretches an elastic membrane over the gaps in the surface. The images interpolated through relaxation are shown in Figure 1(b) and

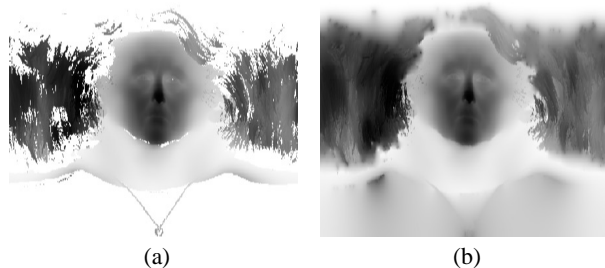


Figure 1: (a) Range data of “Grace” from a Cyberware scanner. (b) Recovered plain data.

illustrate improvements in the hair area and chin area. Relaxation works effectively when the range surface is smooth, and particularly in the case of human head range data, the smoothness requirement of the solutions is satisfied quite effectively.

Figure 2(a) shows two 512×256 reflectance (RGB) texture maps as monochrome images. Each reflectance value represents the surface color of the object in cylindrical coordinates with corresponding longitude (0–360 degrees) and latitude. Like range images, the acquired reflectance images are lacking color information at certain points. This situation is especially obvious in the hair area and the shoulder area (see Figure 2(a)). We employ the membrane relaxation approach to interpolate the texture image by repeated averaging of neighboring known colors. The original texture image in Figure 2(a) can be compared with the interpolated texture image in Figure 2(b).



Figure 2: (a) Texture data of “George” with void points displayed in white and (b) texture image interpolated using relaxation method.

The method is somewhat problematic in the hair area where range variations may be large and there is a relatively high percentage of missing surface points. A thin-plate relaxation algorithm [18] may be more effective in these regions because it would fill in the larger gaps with less “flattening” than a membrane [10].

Although the head structure in the cylindrical laser range data is distorted along the longitudinal direction, important features such as the slope changes of the nose, forehead, chin, and the contours of the mouth, eyes, and nose are still discernible. In order to locate the contours of those facial features for use in adaptation (see below), we use a modified Laplacian operator (applied to the discrete image through local pixel differencing) to detect edges from the range map shown in Figure 3(a) and produce the field function in Fig. 3(b). For details about the operator, see [8]. The field function highlights important features of interest. For example, the local maxima of the modified Laplacian reveals the boundaries of the lips, eyes, and chin.

2.2 Generic Face Mesh and Mesh Adaptation

The next step is to reduce the large arrays of data acquired by the scanner into a parsimonious geometric model of the face that can eventually be animated efficiently. Motivated by the adaptive meshing techniques [19] that were employed in [23], we significantly



Figure 3: (a) Original range map. (b) Modified Laplacian field function of (a).

improved the technique by adapting a generic face mesh to the data. Figure 4 shows the planar generic mesh which we obtain through a cylindrical projection of the 3D face mesh from [21]. One of the advantages of the generic mesh is that it has well-defined features which form the basis for accurate feature based adaptation to the scanned data and automatic scaling and positioning of facial muscles as the mesh is deformed to fit the images. Another advantage is that it automatically produces an efficient triangulation, with finer triangles over the highly curved and/or highly articulate regions of the face, such as the eyes and mouth, and larger triangles elsewhere.

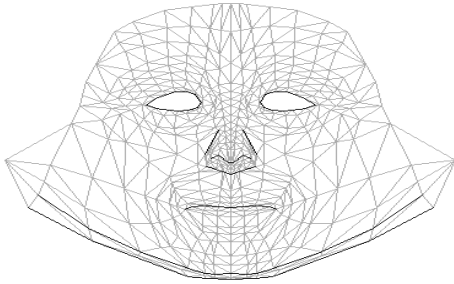


Figure 4: Facial portion of generic mesh in 2D cylindrical coordinates. Dark lines are features for adaptation.

We label all facial feature nodes in the generic face prior to the adaptation step. The feature nodes include eye contours, nose contours, mouth contours, and chin contours.

For any specific range image and its positive Laplacian field function (Figure 3), the generic mesh adaptation procedure performs the following steps to locate feature points in the range data (see [8] for details):

Mesh Adaptation Procedures

1. Locate nose tip
2. Locate chin tip
3. Locate mouth contour
4. Locate chin contour
5. Locate ears
6. Locate eyes
7. Activate spring forces
8. Adapt hair mesh
9. Adapt body mesh
10. Store texture coordinates

Once the mesh has been fitted by the above feature based matching technique (see Plate 3), the algorithm samples the range image at the location of the nodes of the face mesh to capture the facial geometry, as is illustrated in Figure 5.

The node positions also provide texture map coordinates that are used to map the full resolution color image onto the triangles (see Plate 3).

2.3 Estimation of Relaxed Face Model

Ideally, the subject’s face should be in a neutral, relaxed expression when he or she is being scanned. However, the scanned woman in

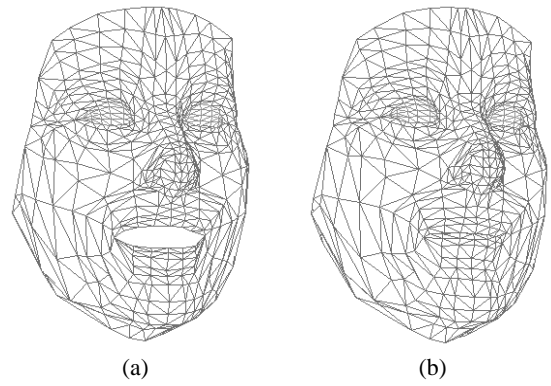


Figure 5: (a) Generic geometric model conformed to Cyberware scan of “Heidi”. (b) Same as (a). Note that “Heidi’s” mouth is now closed, subsequent to estimation of the relaxed face geometry.

the “Heidi” dataset is smiling and her mouth is open (see Plate 2). We have made our algorithm tolerant of these situations. To construct a functional model, it is important to first estimate the relaxed geometry. That is, we must infer what the “Heidi” subject would look like had her face been in a relaxed pose while she was being scanned. We therefore estimate the range values of the closed mouth contour from the range values of the open mouth contour by the following steps:

1. Perform adaptation procedures in Sec. 2.2 without step 3.
2. Store nodal longitude/latitude into adapted face model.
3. Perform lip adaptation in step 3 in sec. 2.2
4. Store nodal range values into adapted face model.

As a result, the final reconstructed face model in Figure 5(b) will have a relaxed mouth because the longitude and latitude recorded is the default shape of our closed mouth model (see Figure 4). Moreover, the shape of the final reconstructed face is still faithful to the head data because the range value at each facial nodal point is obtained correctly after the lip adaptation procedure has been performed. Relaxing the face shown in Figure 5(a) results in the image in Figure 5(b) (with eyelids inserted — see below).

3 The Dynamic Skin and Muscle Model

This section describes how our system proceeds with the construction of a fully functional model of the subject’s face from the facial mesh produced by the adaptation algorithm described in the previous section. To this end, we automatically create a dynamic model of facial tissue, estimate a skull surface, and insert the major muscles of facial expression into the model. The following sections describe each of these components. We also describe our high-performance parallel, numerical simulation of the dynamic facial tissue model.

3.1 Layered Synthetic Tissue Model

The skull is covered by deformable tissue which has five distinct layers [4]. Four layers—epidermis, dermis, sub-cutaneous connective tissue, and fascia—comprise the skin, and the fifth consists of the muscles of facial expression. Following [20], and in accordance with the structure of real skin [5], we have designed a new, synthetic tissue model (Figure 6(a)).

The tissue model is composed of triangular prism elements (see Figure 6(a)) which match the triangles in the adapted facial mesh. The epidermal surface is defined by nodes 1, 2, and 3, which are connected by epidermal springs. The epidermis nodes are also connected by dermal-fatty layer springs to nodes 4, 5, and 6, which define the fascia surface. Fascia nodes are interconnected by fascia

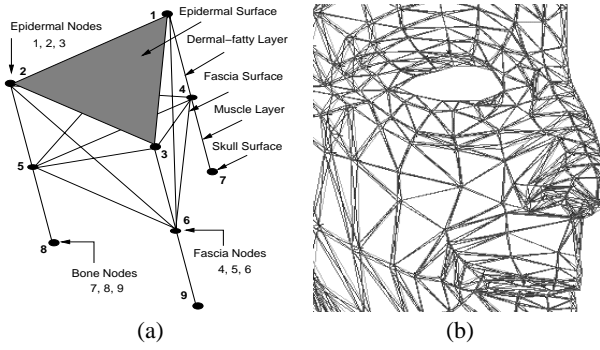


Figure 6: (a) Triangular skin tissue prism element. (b) Close-up view of right side of an individual with conformed elements.

springs. They are also connected by muscle layer springs to skull surface nodes 7, 8, 9.

Figure 9(b) shows 684 such skin elements assembled into an extended skin patch. Several synthetic muscles are embedded into the muscle layer of the skin patch and the figure shows the skin deformation due to muscle contraction. Muscles are fixed in an estimated bony subsurface at their point of emergence and are attached to fascia nodes as they run through several tissue elements. Figure 6(b) shows a close-up view of the right half of the facial tissue model adapted to an individual's face which consists of 432 elements.

3.2 Discrete Deformable Models (DDMs)

A discrete deformable model has a node-spring-node structure, which is a uniaxial finite element. The data structure for the node consists of the nodal mass m_i , position $\mathbf{x}_i(t) = [x_i(t), y_i(t), z_i(t)]'$, velocity $\mathbf{v}_i = d\mathbf{x}_i/dt$, acceleration $\mathbf{a}_i = d^2\mathbf{x}_i/dt^2$, and net nodal forces $\mathbf{f}_i^D(t)$. The data structure for the spring in this DDM consists of pointers to the head node i and the tail node j which the spring interconnects, the natural or rest length l_k of the spring, and the spring stiffness c_k .

3.3 Tissue Model Spring Forces

By assembling the discrete deformable model according to histological knowledge of skin (see Figure 6(a)), we are able to construct an anatomically consistent, albeit simplified, tissue model. Figure 6(b) shows a close-up view of the tissue model around its eye and nose parts of a face which is automatically assembled by following the above approach.

- The force spring j exerts on node i is

$$\mathbf{g}_j = c_j(l_j - l_j^r)\mathbf{s}_j$$

- each layer has its own stress-strain relationship c_j and the dermal-fatty layer uses biphasic springs (non-constant c_j) [20]
- l_j^r and $l_j = \|\mathbf{x}_j - \mathbf{x}_i\|$ are the rest and current lengths for spring j
- $\mathbf{s}_j = (\mathbf{x}_j - \mathbf{x}_i)/l_j$ is the spring direction vector for spring j

3.4 Linear Muscle Forces

The muscles of facial expression, or the muscular plate, spreads out below the facial tissue. The facial musculature is attached to the skin tissue by short elastic tendons at many places in the fascia, but is fixed to the facial skeleton only at a few points. Contractions of the facial muscles cause movement of the facial tissue. We model

28 of the primary facial muscles, including the zygomatic major and minor, frontalis, nasii, corrugator, mentalis, buccinator, and angullii depressor groups. Plate 4 illustrates the effects of automatic scaling and positioning of facial muscle vectors as the generic mesh adapts to different faces.

To better emulate the facial muscle attachments to the fascia layer in our model, a group of fascia nodes situated along the muscle path—i.e., within a predetermined distance from a central muscle vector, in accordance with the muscle width—experience forces from the contraction of the muscle. The face construction algorithm determines the nodes affected by each muscle in a precomputation step.

To apply muscle forces to the fascia nodes, we calculate a force for each node by multiplying the muscle vector with a force length scaling factor and a force width scaling factor (see Figure 7(a)). Function Θ_1 (Figure 8(a)) scales the muscle force according to the length ratio $\varepsilon_{j,i}$, while Θ_2 (Figure 8(b)) scales it according to the width $\omega_{j,i}$ at node i of muscle j :

$$\begin{aligned} \varepsilon_{j,i} &= ((\mathbf{m}_j^F - \mathbf{x}_i) \cdot \mathbf{m}_j) / (\|\mathbf{m}_j^A - \mathbf{m}_j^F\|) \\ \omega_{j,i} &= \|\mathbf{p}_i - (\mathbf{p}_i \cdot \mathbf{n}_j)\mathbf{n}_j\| \end{aligned}$$

- The force muscle j exerts on node i is

$$\mathbf{f}_i^j = \Theta_1(\varepsilon_{j,i})\Theta_2(\omega_{j,i})\mathbf{m}_j$$

- Θ_1 scales the force according to the distance ratio $\varepsilon_{j,i}$, where $\varepsilon_{j,i} = \rho_{j,i}/d_j$, with d_j the muscle j length.
- Θ_2 scales the force according to the width ratio $\omega_{j,i}/w_j$, with w_j the muscle j width.
- \mathbf{m}_j is the normalized muscle vector for muscle j

Note that the muscle force is scaled to zero at the root of the muscle fiber in the bone and reaches its full strength near the end of the muscle fiber. Figure 9(b) shows an example of the effect of muscle forces applied to a synthetic skin patch.

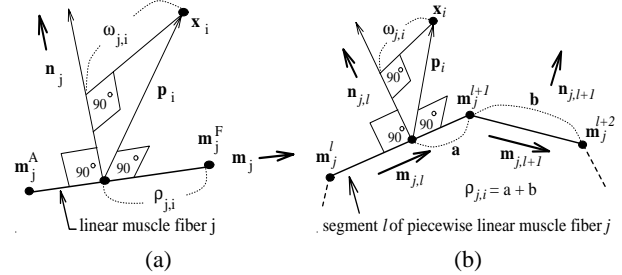


Figure 7: (a) Linear muscle fiber. (b) Piecewise linear muscle fiber.

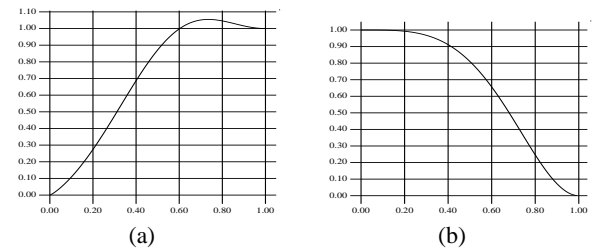


Figure 8: (a) Muscle force scaling function Θ_1 wrt $\varepsilon_{j,i}$, (b) Muscle force scaling function Θ_2 wrt $\omega_{j,i}/w_j$

3.5 Piecewise Linear Muscle Forces

In addition to using linear muscle fibers in section 3.4 to simulate sheet facial muscles like the frontalis and the zygomatics, we also model sphincter muscles, such as the orbicularis oris circling the mouth, by generalizing the linear muscle fibers to be piecewise

linear and allowing them to attach to fascia at each end of the segments. Figure 7(b) illustrates two segments of an N -segment piecewise linear muscle j showing three nodes \mathbf{m}_j^l , \mathbf{m}_j^{l+1} , and \mathbf{m}_j^{l+2} . The unit vectors $\mathbf{m}_{j,l}$, $\mathbf{m}_{j,l+1}$ and $\mathbf{n}_{j,l}$, $\mathbf{n}_{j,l+1}$ are parallel and normal to the segments, respectively. The figure indicates fascia node i at \mathbf{x}_i , as well as the distance $\rho_{j,i} = a + b$, the width $\omega_{j,i}$, and the perpendicular vector \mathbf{p}_i from fascia node i to the nearest segment of the muscle. The length ratio $\varepsilon_{j,i}$ for fascia node i in muscle fiber j is

$$\varepsilon_{j,i} = \frac{(\mathbf{m}_j^{l+1} - \mathbf{x}_i) \cdot \mathbf{m}_{j,l} + \sum_{k=l+1}^N \|\mathbf{m}_j^{k+1} - \mathbf{m}_j^k\|}{\sum_{k=1}^N \|\mathbf{m}_j^{k+1} - \mathbf{m}_j^k\|}$$

The width $\omega_{j,i}$ calculation is the same as for linear muscles. The remaining muscle force computations are the same as in section 3.4. Plate 4 shows all the linear muscles and the piecewise linear sphincter muscles around the mouth.

3.6 Volume Preservation Forces

In order to faithfully exhibit the incompressibility [2] of real human skin in our model, a volume constraint force based on the change of volume (see Figure 9(a)) and displacements of nodes is calculated and applied to nodes. In Figure 9(b) the expected effect of volume preservation is demonstrated. For example, near the origin of the muscle fiber, the epidermal skin is bulging out, and near the end of the muscle fiber, the epidermal skin is depressed.

- The volume preservation force element e exerts on nodes i in element e is

$$\mathbf{q}_i^e = k_1(V^e - \tilde{V}^e)\mathbf{n}_i^e + k_2(\mathbf{p}_i^e - \tilde{\mathbf{p}}_i^e)$$

- \tilde{V}^e and V^e are the rest and current volumes for e
- \mathbf{n}_i^e is the epidermal normal for epidermal node i
- $\tilde{\mathbf{p}}_i^e$ and \mathbf{p}_i^e are the rest and current nodal coordinates for node i with respect to the center of mass of e
- k_1, k_2 are force scaling constants

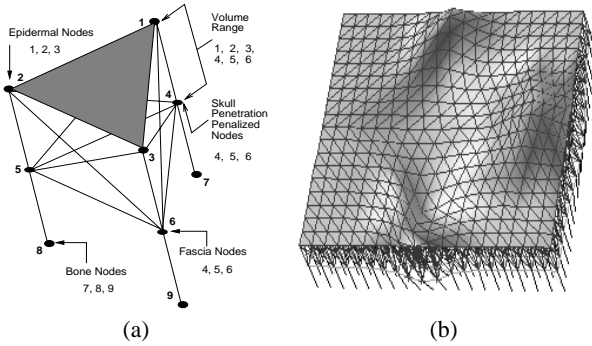


Figure 9: (a) Volume preservation and skull nonpenetration element. (b) Assembled layered tissue elements under multiple muscle forces.

3.7 Skull Penetration Constraint Forces

Because of the underlying impenetrable skull of a human head, the facial tissue during a facial expression will slide over the underlying bony structure. With this in mind, for each individual's face model reconstructed from the laser range data, we estimate the skull surface normals to be the surface normals in the range data image. The skull is then computed as an offset surface. To prevent nodes from penetrating the estimated skull (see Figure 9(a)), we apply a skull non-penetration constraint to cancel out the force component on the fascia node which points into the skull; therefore, the resulting force will make the nodes slide over the skull.

- The force to penalize fascia node i during motion is:

$$\mathbf{s}_i = \begin{cases} -(\mathbf{f}_i^n \cdot \mathbf{n}_i)\mathbf{n}_i & \text{when } \mathbf{f}_i^n \cdot \mathbf{n}_i < 0 \\ \mathbf{0} & \text{otherwise} \end{cases}$$

- \mathbf{f}_i^n is the net force on fascia node i
- \mathbf{n}_i is the nodal normal of node i

3.8 Equations of Motion for Tissue Model

Newton's law of motion governs the response of the tissue model to forces. This leads to a system of coupled second order ODEs that relate the node positions, velocities, and accelerations to the nodal forces. The equation for node i is

$$m_i \frac{d^2 \mathbf{x}_i}{dt^2} + \gamma_i \frac{d \mathbf{x}_i}{dt} + \tilde{\mathbf{g}}_i + \tilde{\mathbf{q}}_i + \tilde{\mathbf{s}}_i + \tilde{\mathbf{h}}_i = \tilde{\mathbf{f}}_i$$

- m_i is the nodal mass,
- γ_i is the damping coefficient,
- $\tilde{\mathbf{g}}_i$ is the total spring force at node i ,
- $\tilde{\mathbf{q}}_i$ is the total volume preservation force at node i ,
- $\tilde{\mathbf{s}}_i$ is the total skull penetration force at node i ,
- $\tilde{\mathbf{h}}_i$ is the total nodal restoration force at node i ,
- $\tilde{\mathbf{f}}_i$ is the total applied muscle force at node i ,

3.9 Numerical Simulation

The solution to the above system of ODEs is approximated by using the well-known, explicit Euler method. At each iteration, the nodal acceleration at time t is computed by dividing the net force by nodal mass. The nodal velocity is then calculated by integrating once, and another integration is done to compute the nodal positions at the next time step $t + \Delta t$, as follows:

$$\begin{aligned} \mathbf{a}_i^t &= \frac{1}{m_i}(\tilde{\mathbf{f}}_i^t - \gamma_i \mathbf{v}_i^t - \tilde{\mathbf{g}}_i^t - \tilde{\mathbf{q}}_i^t - \tilde{\mathbf{s}}_i^t - \tilde{\mathbf{h}}_i^t) \\ \mathbf{v}_i^{t+\Delta t} &= \mathbf{v}_i^t + \Delta t \mathbf{a}_i^t \\ \mathbf{x}_i^{t+\Delta t} &= \mathbf{x}_i^t + \Delta t \mathbf{v}_i^{t+\Delta t} \end{aligned}$$

3.10 Default Parameters

The default parameters for the physical/numerical simulation and the spring stiffness values of different layers are as follows:

Mass (m)	Time step (Δt)	Damping (γ)
0.5	0.01	30

	Epid	Derm-fat 1	Derm-fat 2	Fascia	Muscle
c	60	30	70	80	10

3.11 Parallel Processing for Facial Animation

The explicit Euler method allows us to easily carry out the numerical simulation of the dynamic skin/muscle model in parallel. This is because at each time step all the calculations are based on the results from the previous time step. Therefore, parallelization is achieved by evenly distributing calculations at each time step to all available processors. This parallel approach increases the animation speed to allow us to simulate facial expressions at interactive rates on our Silicon Graphics multiprocessor workstation.

4 Geometry Models for Other Head Components

To complete our physics-based face model, additional geometric models are combined along with the skin/muscle/skull models developed in the previous section. These include the eyes, eyelids, teeth, neck, hair, and bust (Figure 10). See Plate 5 for an example of a complete model.

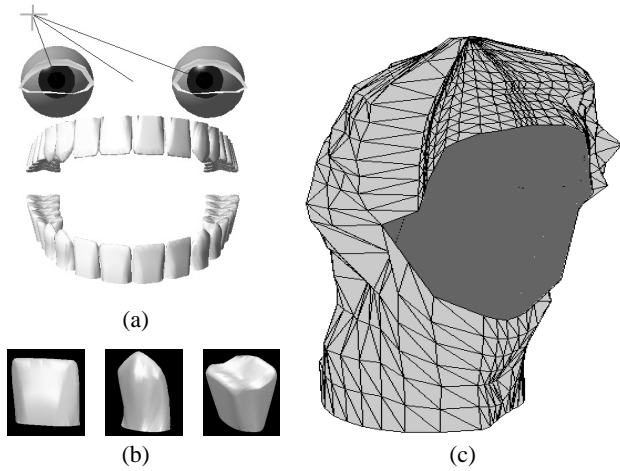


Figure 10: (a) Geometric models of eyes, eyelids, and teeth (b) Incisor, canine, and molar teeth. (c) hair and neck.

4.1 Eyes

Eyes are constructed from spheres with adjustable irises and adjustable pupils (Figure 10(a)). The eyes are automatically scaled to fit the facial model and are positioned into it. The eyes rotate kinematically in a coordinated fashion so that they will always converge on a specified fixation point in three-dimensional space that defines the field of view. Through a simple illumination computation, the eyes can automatically dilate and contract the pupil size in accordance with the amount of light entering the eye.

4.2 Eyelids

The eyelids are polygonal models which can blink kinematically during animation (see Figure 10(a)). Note that the eyelids are open in Figure 10(a).

If the subject is scanned with open eyes, the sensor will not observe the eyelid texture. An eyelid texture is synthesized by a relaxation based interpolation algorithm similar to the one described in section 2.1. The relaxation algorithm interpolates a suitable eyelid texture from the immediately surrounding texture map. Figure 11 shows the results of the eyelid texture interpolation.

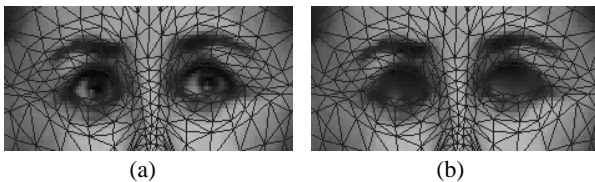


Figure 11: (a) Face texture image with adapted mesh before eyelid texture synthesis (b) after eyelid texture synthesis.

4.3 Teeth

We have constructed a full set of generic teeth based on dental images. Each tooth is a NURBS surfaces of degree 2. Three different teeth shapes, the incisor, canine, and molar, are modeled (Figure 10(b)). We use different orientations and scalings of these basic shapes to model the full set of upper and lower teeth shown in Figure 10(a). The dentures are automatically scaled to fit in length, curvature, etc., and are positioned behind the mouth of the facial model.

4.4 Hair, Neck, and Bust Geometry

The hair and bust are both rigid polygonal models (see Figure 10(c)). They are modeled from the range data directly, by extending the

facial mesh in a predetermined fashion to the boundaries of the range and reflectance data, and sampling the images as before.

The neck can be twisted, bent and rotated with three degrees of freedom. See Figure 12 for illustrations of the possible neck articulations.

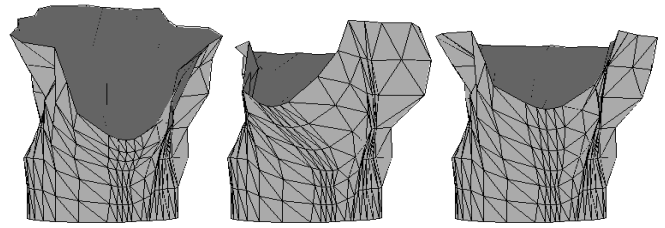


Figure 12: articulation of neck.

5 Animation Examples

Plate 1 illustrates several examples of animating the physics-based face model after conformation to the “Heidi” scanned data (see Plate 2).

- The *surprise* expression results from contraction of the outer frontalis, major frontalis, inner frontalis, zygomatics major, zygomatics minor, depressor labii, and mentalis, and rotation of the jaw.
- The *anger* expression results from contraction of the corrugator, lateral corrugator, levator labii, levator labii nasi, anguli depressor, depressor labii, and mentalis.
- The *quizzical look* results from an asymmetric contraction of the major frontalis, outer frontalis, corrugator, lateral corrugator, levator labii, and buccinator.
- The *sadness* expression results from a contraction of the inner frontalis, corrugator, lateral corrugator, anguli depressor, and depressor labii.

Plate 6 demonstrates the performance of our face model construction algorithm on two male individuals (“Giovanni” and “Mick”). Note that the algorithm is tolerant of some amount of facial hair.

Plate 7 shows a third individual “George.” Note the image at the lower left, which shows two additional expression effects—cheek puffing, and lip puckering—that combine to simulate the vigorous blowing of air through the lips. The cheek puffing was created by applying outwardly directed radial forces to “inflate” the deformable cheeks. The puckered lips were created by applying radial pursing forces and forward protruding forces to simulate the action of the orbicularis oris sphincter muscle which circles the mouth.

Finally, Plate 8 shows several frames from a two-minute animation “*Bureaucrat Too*” (a second-generation version of the 1990 “*Bureaucrat*” which was animated using the generic facial model in [20]). Here “George” tries to read landmark papers on facial modeling and deformable models in the SIGGRAPH ’87 proceedings, only to realize that he doesn’t yet have a brain!

6 Conclusion and Future Work

The human face consists of a biological tissue layer with nonlinear deformation properties, a muscle layer knit together under the skin, and an impenetrable skull structure beneath the muscle layer. We have presented a physics-based model of the face which takes all of these structures into account. Furthermore, we have demonstrated a new technique for automatically constructing face models of this sort and conforming them to individuals by exploiting high-resolution laser scanner data. The conformation process is carried out by a feature matching algorithm based on a reusable generic

mesh. The conformation process, efficiently captures facial geometry and photometry, positions and scales facial muscles, and also estimates the skull structure over which the new synthetic facial tissue model can slide. Our facial modeling approach achieves an unprecedented level of realism and fidelity to any specific individual. It also achieves a good compromise between the complete emulation of the complex biomechanical structures and functionality of the human face and real-time simulation performance on state-of-the-art computer graphics and animation hardware.

Although we formulate the synthetic facial skin as a layered tissue model, our work does not yet exploit knowledge of the variable thickness of the layers in different areas of the face. This issue will in all likelihood be addressed in the future by incorporating additional input data about the subject acquired using noninvasive medical scanners such as CT or MR.

Acknowledgments

The authors thank Lisa White and Jim Randall for developing the piecewise linear muscle model used to model the mouth. Range/RGB facial data were provided courtesy of Cyberware, Inc., Monterey, CA. The first two authors thank the Natural Science and Engineering Research Council of Canada for financial support. DT is a fellow of the Canadian Institute for Advanced Research.

References

- [1] T. Akimoto, Y. Suenaga, and R. Wallace. Automatic creation of 3D facial models. *IEEE Computer Graphics and Applications*, 13(5):16–22, September 1993.
- [2] James Doyle and James Philips. *Manual on Experimental Stress Analysis*. Society for Experimental Mechanics, fifth edition, 1989.
- [3] Irfan A. Essa. *Visual Interpretation of Facial Expressions using Dynamic Modeling*. PhD thesis, MIT, 1994.
- [4] Frick and Hans. *Human Anatomy*, volume 1. Thieme Medical Publishers, Stuttgart, 1991.
- [5] H. Gray. *Anatomy of the Human Body*. Lea & Febiger, Philadelphia, PA, 29th edition, 1985.
- [6] Brian Guenter. A system for simulating human facial expression. In *State of the Art in Computer Animation*, pages 191–202. Springer-Verlag, 1992.
- [7] T. Kurihara and K. Arai. A transformation method for modeling and animation of the human face from photographs. In *State of the Art in Computer Animation*, pages 45–57. Springer-Verlag, 1991.
- [8] Y.C. Lee, D. Terzopoulos, and K. Waters. Constructing physics-based facial models of individuals. In *Proceedings of Graphics Interface '93*, pages 1–8, Toronto, May 1993.
- [9] N. Magneneat-Thalman, H. Minh, M. Angelis, and D. Thalman. Design, transformation and animation of human faces. *Visual Computer*, 5:32–39, 1989.
- [10] D. Metaxas and E. Milios. Reconstruction of a color image from nonuniformly distributed sparse and noisy data. *Computer Vision, Graphics, and Image Processing*, 54(2):103–111, March 1992.
- [11] M. Nahas, H. Hutric, M. Rioux, and J. Domey. Facial image synthesis using skin texture recording. *Visual Computer*, 6(6):337–343, 1990.
- [12] M. Oka, K. Tsutsui, A. Ohba, Y. Kurauchi, and T. Tago. Real-time manipulation of texture-mapped surfaces. In *SIGGRAPH 21*, pages 181–188. ACM Computer Graphics, 1987.
- [13] F. Parke. Computer generated animation of faces. In *ACM National Conference*, pages 451–457. ACM, 1972.
- [14] F. Parke. Parameterized models for facial animation. *IEEE Computer Graphics and Applications*, 2(9):61–68, November 1982.
- [15] F. Parke. Parameterized models for facial animation revisited. In *SIGGRAPH Facial Animation Tutorial Notes*, pages 43–56. ACM SIGGRAPH, 1989.
- [16] Elizabeth C. Patterson, Peter C. Litwinowicz, and N. Greene. Facial animation by spatial mapping. In *State of the Art in Computer Animation*, pages 31–44. Springer-Verlag, 1991.
- [17] S. Platt and N. Badler. Animating facial expression. *Computer Graphics*, 15(3):245–252, August 1981.
- [18] D. Terzopoulos. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10(4):417–438, 1988.
- [19] D. Terzopoulos and M. Vasilescu. Sampling and reconstruction with adaptive meshes. In *Proceedings of Computer Vision and Pattern Recognition Conference*, pages 70–75. IEEE, June 1991.
- [20] D. Terzopoulos and K. Waters. Physically-based facial modeling, analysis, and animation. *Visualization and Computer Animation*, 1:73–80, 1990.
- [21] K. Waters. A muscle model for animating three-dimensional facial expression. *Computer Graphics*, 22(4):17–24, 1987.
- [22] K. Waters. A physical model of facial tissue and muscle articulation derived from computer tomography data. In *Visualization in Biomedical Computing*, pages 574–583. SPIE, Vol. 1808, 1992.
- [23] K. Waters and D. Terzopoulos. Modeling and animating faces using scanned data. *Visualization and Computer Animation*, 2:123–128, 1991.
- [24] L. Williams. Performance-driven facial animation. In *SIGGRAPH 24*, pages 235–242. ACM Computer Graphics, 1990.
- [25] J. Yau and N. Duffy. 3-D facial animation using image samples. In *New Trends in Computer Graphics*, pages 64–73. Springer-Verlag, 1988.



Plate 1: Objective. *Input*: Range map in 3D and texture map (top). *Output*: Functional face model for animation.



Plate 2: Raw 512×256 digitized data for Heidi (top left), George (top right), Giovanni (bottom left), Mick (bottom right).

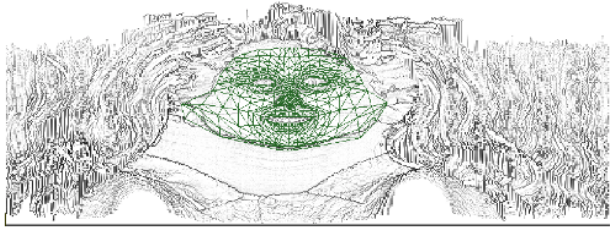


Plate 3: Adapted face mesh overlaying texture map and Laplacian filtered range map of Heidi.

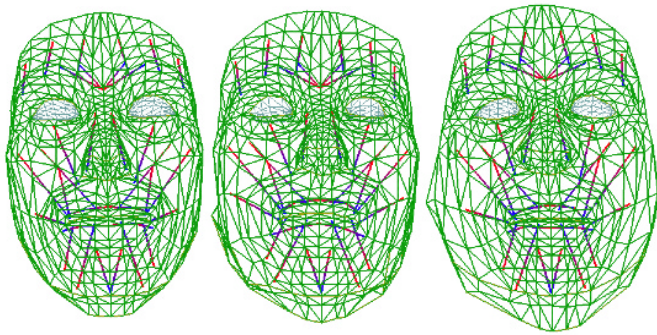


Plate 4: Muscle fiber vector embedded in generic face model and two adapted faces of Heidi and George.



Plate 5: Complete, functional head model of Heidi with physics-based face and geometric eyes, teeth, hair, neck, and shoulders (in Monument Valley).



Plate 6: Animation examples of Giovanni and Mick.



Plate 7: Animation example of George.



Plate 8: George in four scenes from "Bureaucrat Too".



Learning-Based Approaches

Volker Blanz
MPI Informatik

Learning-Based Approaches



- Measure movements of real faces
- Reproduce only the shape or appearance,
- Ignore underlying mechanisms
- Automatically learn by induction
- Transfer information to novel situations
 - No artist
 - No physical model and material properties
 - Potential for highly realistic results within the range of measured conditions

Learning-Based Approaches



...more and more popular due to

- Progress in Scanning Technology
 - high quality, high speed, low cost
- Large Data Bases possible
 - disk space, RAM, CPU, GPU
- Research in Machine Learning
 - we're only just beginning to use this potential

Overview



Image-Based Animation



3D Modeling and Animation



3D Animation applied to Images



Image-Based Animation



Learning from Video Data

- + Photo-realistic Results
- + Reproduce Dynamics of Motion
- Restricted in pose and illumination
- Need video footage of the person to be animated

2D Animation Paradigms



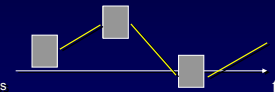
Video Rewrite:

Rearrange original frames



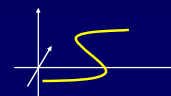
Linear Morphing:

Smooth transitions between keyframes



Smooth Trajectories

in Parameter Space:



Video Rewrite



Bregler et al., 1997, Graf et al., 2000

Re-arrange video-frames:



Search for the **Triphones** required in the novel utterance.

Speech Synthesis



Visemes = Basic mouth shapes in speech,
Visual analog of Phonemes.

Coarticulation = Influence of previous and subsequent viseme.

Triphones = Triplets of phonemes

Hello = /H - E - L/ + /E - L - OW/



Video Rewrite



Compensate for 3D head movements by 2D warping of

- Entire face (Bregler)
- Sub-regions (Graf)

- + Photo-realistic
- Possible appearances are limited
- Requires large corpus of video frames

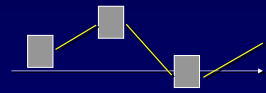
2D Morphing



"MikeTalk" by Ezzat&Poggio, 1998, 2000

Select keyframes = visemes from video

Morph in between:



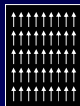
Movements in the Image Plane



Bad:
Cross-dissolve



Better: Warp
= shift pixels in image



$$I_{warp}(\mathbf{x} + \mathbf{v}(\mathbf{x})) = I(\mathbf{x}) \quad \mathbf{v}(\mathbf{x}) = \begin{pmatrix} v_x(x, y) \\ v_y(x, y) \end{pmatrix}$$

Vector Field of Pixel Displacements

2D Morphing



Warp Pixels and Interpolate Color Values:

$$\lambda \in [0,1]$$

Let corresponding points (e.g. corners of the mouth) be:

$$I_1(\mathbf{x}) \cong I_2(\mathbf{x} + \mathbf{v}(\mathbf{x}))$$

$$I_{morph}(\mathbf{x} + \lambda \mathbf{v}(\mathbf{x})) = (1 - \lambda) \cdot I_1(\mathbf{x}) + \lambda \cdot I_2(\mathbf{x} + \mathbf{v}(\mathbf{x}))$$

2D Morphing

"MikeTalk" by Ezzat&Poggio, 1998, 2000

Find corresponding points with an optical flow algorithm:



Morph between visemes:



2D Morphing

- + Need to store keyframes only.
- Morphing in 2D difficult due to occlusions (e.g. teeth)
 - *Pixels can only be displaced, but not appear or disappear.*
- Dynamics: Only linear transitions.

2D Vector Space for Animation

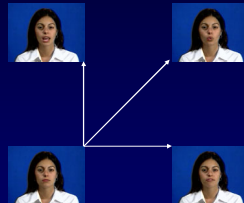
• *Vector Space of Images:*

Cosatto, Graf, 1998

• *Vector Space of*

Warp-fields and color values:

Ezzat, Geiger, Poggio, 2002

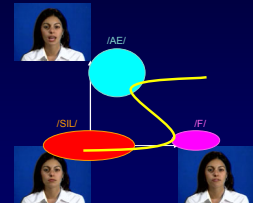


Trainable Speech Animation

Ezzat, Geiger, Poggio, 2002

Visemes:
modeled as
clusters in model space

Speech trajectory:
has to be close to targets,
but smooth.
Use regression methods.



Trainable Speech Animation

Ezzat, Geiger, Poggio, 2002

Corpus:
8 min. Video + Audio + Text

PCA on images

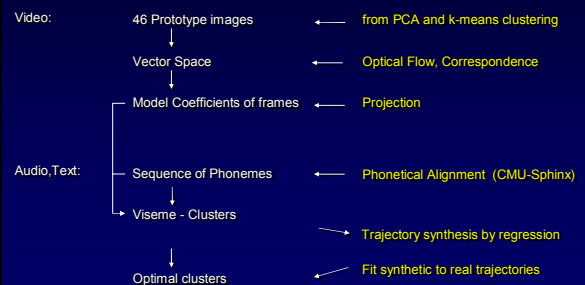
k-means clustering

46 Prototype images



Training

Ezzat, Geiger, Poggio, 2002



Animation

Ezzat, Geiger, Poggio, 2002

```

    graph TD
      A[Audio Input] --> B[Sequence of Phonemes]
      B --> C[Trajectory]
      C --> D[Mouth Images]
      D --> E[Compositing into random background video, including eye movements]
      B -.->|Phonetical Alignment (CMU-Sphinx)| B
      D -.->|Linear Combinations, Warping| D
  
```

Trainable Speech Animation

Ezzat, Geiger, Poggio, 2002

3D Animation

- Animation with 3D Rotation, Illumination, ...
- Occlusion of Teeth by Lips modeled correctly.
- Face can be integrated in Virtual Scene
- Unlike Motion-Capturing, generate new motion
 - e.g. *morph-targets as keyframes*

3D Animation

Exploit 3D Measurements from:

- Images from Multiple Viewpoints, using
 - *Facial features* (Pighin et al. 1998)
 - *Passive markers* (Reveret and Essa, 2001)
- Scans (Kalberer et al 2001, Blanz et al 2003)
 - *high-resolution scans capture details such as wrinkles.*

Example-Based Animation

Blanz et al., Eurographics 03

smile = -

Database of 3D scans of

- *Facial expressions*
- *Different persons' faces*

⇒ Converted to Face Vectors

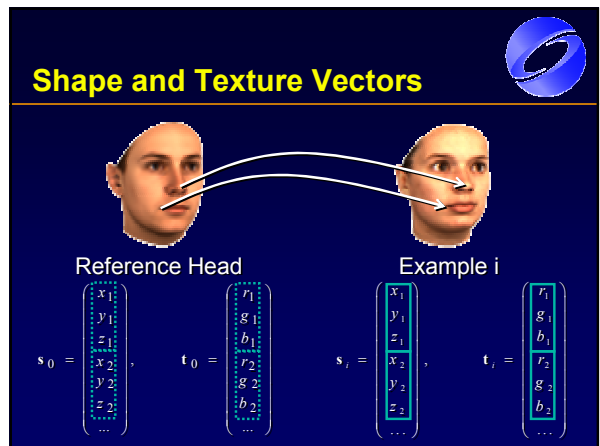
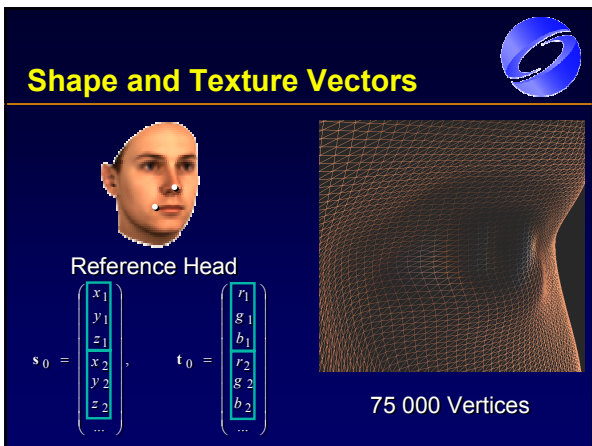
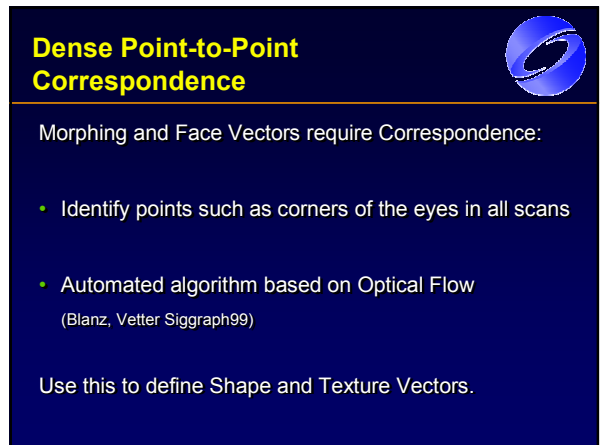
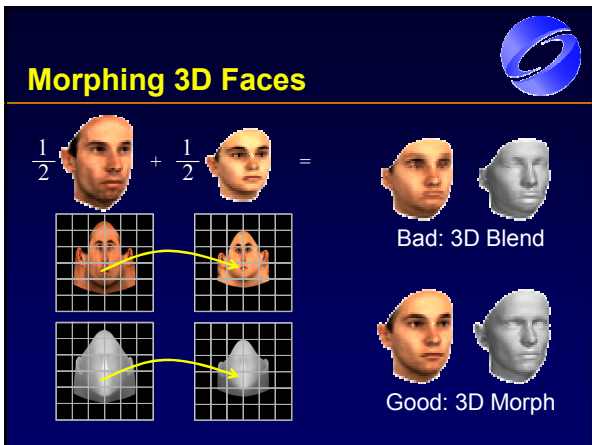
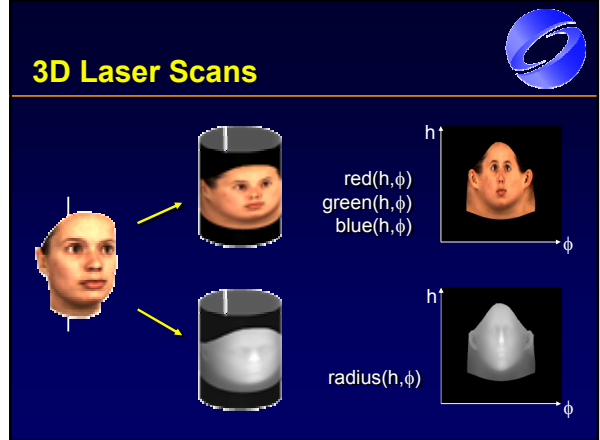
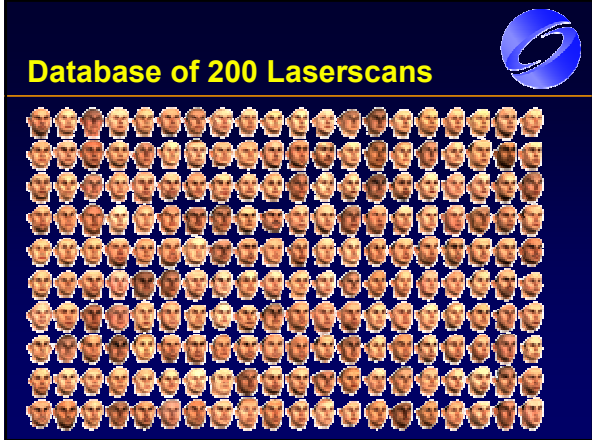
Vector Space of Shape and Texture

Blanz, Vetter Siggraph99

3D Morphable Face Model

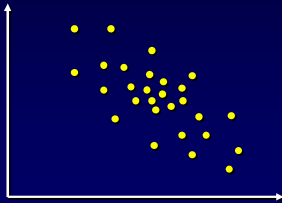
$$\alpha_1 \cdot \text{shape}_1 + \alpha_2 \cdot \text{shape}_2 + \alpha_3 \cdot \text{shape}_3 + \alpha_4 \cdot \text{shape}_4 + \dots$$

$$\beta_1 \cdot \text{texture}_1 + \beta_2 \cdot \text{texture}_2 + \beta_3 \cdot \text{texture}_3 + \beta_4 \cdot \text{texture}_4 + \dots$$



Statistics

Faces are Points in Face Space

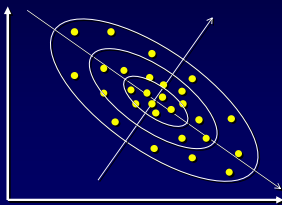


Principal Component Analysis PCA

- Estimate of Probability Density Function
- Order dimensions of face space according to the variance found in data
 - *Data compression*
 - *Coarse-to-fine strategies*

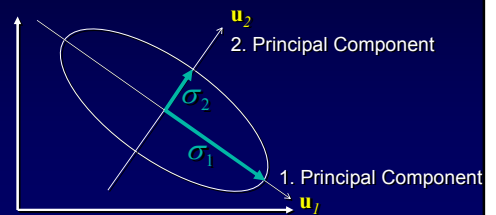
Principal Component Analysis PCA

Estimate Probability: Normal Distribution

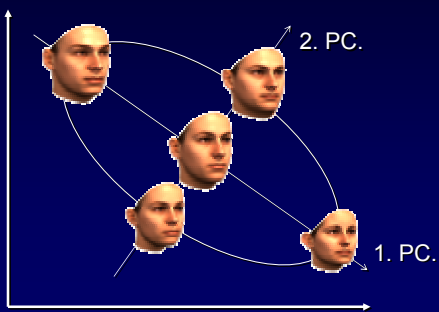


Principal Component Analysis PCA

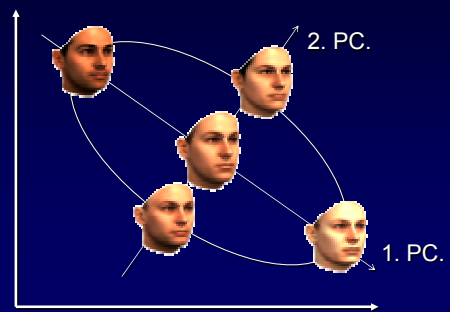
Estimate Probability: Normal Distribution



PCA of Shapes



PCA of Textures



Statistical Analysis

Find directions that are suited for describing and manipulating faces and facial expressions

- Unsupervised Methods (unlabeled data-points):
 - PCA
 - Directions are, in general, not meaningful for manipulations.
 - Independent Component Analysis
 - Statistically independent directions in face space
 - For Facial Animation: Kalberer et al. 2001:
- Supervised Learning:
 - From labeled examples

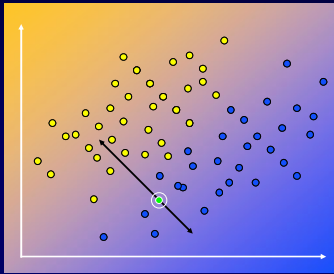
Facial Attributes

- Learn from labeled examples
- Fit a linear function to input data
 - Use *Linear Regression* or *Discriminant Analysis* or *Support Vector Machine*.
- Follow gradient to manipulate faces

Goal:

- Manipulate attribute, but leave individual characteristics unchanged: same person.

Learning from Labeled Examples



Facial Attributes

Gender



Weight



Original

Bianz, Vetter Siggraph99

Facial Attributes

Gender



Weight



Original

Bianz, Vetter Siggraph99

Facial Attributes

Hooked
Nose



Subjective
Attractiveness



Original

Bianz, Vetter Siggraph99

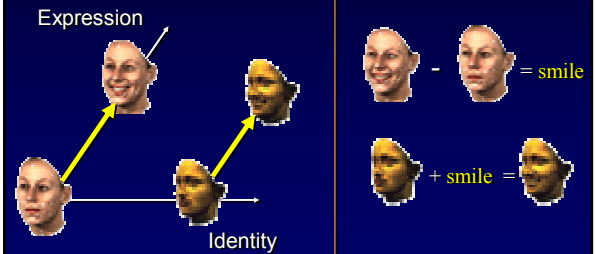
Example-Based Animation

Transfer 3D displacements of vertices to novel face.

- Requires correspondence of vertices (corners of the eyes, mouth...)
- Expressions differ across individuals. Still:
 - *Simple transfer of 3D vertex displacements causes no obvious artifacts.*
 - *More sophisticated methods may improve results.*

Identity and Expression

Blanz et al., Eurographics 03



Scans of Visemes



Strategy

- Mouth poses are learned from static scans.
- New reference scan with open mouth and teeth. Closing the mouth will occlude teeth in 3D.
- Upper jaw teeth remain fixed relative to the head.
- Lower jaw teeth move with tip of chin.
- Face and Lips: 3D Morphing.

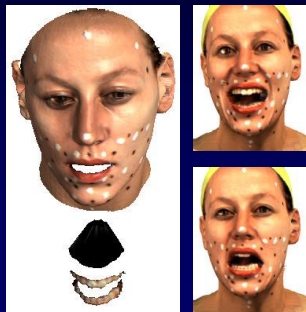
Teeth

Upper: fixed to head
Lower: move with chin

Same teeth for

- all expressions
- all persons

Inserted automatically



Occlusions



⇒ Occlusions make correspondence more difficult for optical flow than with neutral faces.

Use Bootstrapping:

- Start with set of similar expressions
- Extend vector space step-by-step

Mouth-Modeler based on PCA

Blanz et al., Eurographics 03

3D Animation applied to Images

Versatility of 3D Animation

- Works for any pose and illumination
- No video footage of animated face required

Photo-realism of 2D methods

- Animation in given scene context

Reanimation in Images and Video

Blanz et al., Eurographics 03

Animate

- unknown faces
- in given images or video
- at any pose and illumination

Approach

Blanz et al., Eurographics 03

1. Reconstruct 3D shape
2. Add 3D deformation
3. Draw 3D face into the image

head angle, position, illumination, ...

Facial Animation in Images

$I_{input} - I_{model} = \text{smile}$

$I_{input} \xrightarrow{3D} \text{3D model} + \text{smile} \rightarrow \text{Annotated 3D model} \rightarrow \text{Output Image}$

Fitting the Model to an Image

Blanz, Vetter Siggraph99

$$I_{input} \rightarrow I_{model} = R_{\rho} \left(\begin{matrix} \alpha_1 \cdot \text{Face}_1 + \alpha_2 \cdot \text{Face}_2 + \alpha_3 \cdot \text{Face}_3 + \alpha_4 \cdot \text{Face}_4 + \dots \\ \beta_1 \cdot \text{Face}_1 + \beta_2 \cdot \text{Face}_2 + \beta_3 \cdot \text{Face}_3 + \beta_4 \cdot \text{Face}_4 + \dots \end{matrix} \right)$$

$R = \text{Rendering}$ (Perspective Projection, Phong Illumination, Cast Shadows)

$\rho = \text{Pose, Illumination, ...}$ **Find optimal α, β, ρ**

Minimize Image Difference with Stochastic Newton Optimization.

Automated Parameter Estimation



- **Face Parameters** shape coefficients α_i
texture coefficients β_i
- **3D Geometry** head position
head orientation
focal length
- **Light and Color** Ambient: intensity, color
Parallel: intensity, color, **direction**
Color: contrast, gains, offsets

Error Function



Blanz, Vetter Siggraph99

- Image difference

$$E_{\text{Image}} = \sum_{x,y} (\mathbf{I}_{\text{model}}(x,y) - \mathbf{I}_{\text{input}}(x,y))^2$$

- Plausibility based on PCA

$$E_{\text{prior}} = -\log(p(\alpha_i, \beta_i, \dots))$$

- Minimize

$$E = E_{\text{image}} + E_{\text{prior}}$$

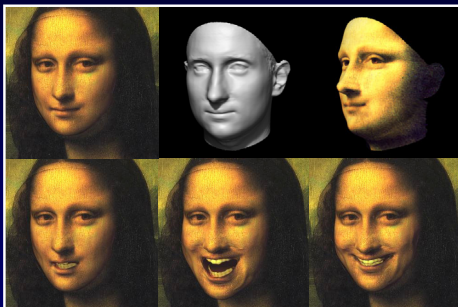


Blanz et al., Eurographics 03



Blanz et al., Eurographics 03

Mona Lisa



Blanz et al., Eurographics 03

Speech Animation



Audio + Text



Phoneme (t) CMU-SPHINX



Keyframe Animation soft accelerations

Reanimation of Video



Goal: Movie dubbing.

- *3D shape from 1 – 3 frames*
- *Track 3D motion*
- *Apply speech in 3D*
- *Draw into frames*

Reanimation of Video



Blanz et al., Eurographics 03

Conclusion



- Learning-based methods have a large potential for achieving photo-realistic results.
- Development of scanning technology is crucial for extensive datasets of high-quality scans.

A Morphable Model For The Synthesis Of 3D Faces

Volker Blanz

Thomas Vetter

Max-Planck-Institut für biologische Kybernetik,
Tübingen, Germany*

Abstract

In this paper, a new technique for modeling textured 3D faces is introduced. 3D faces can either be generated automatically from one or more photographs, or modeled directly through an intuitive user interface. Users are assisted in two key problems of computer aided face modeling. First, new face images or new 3D face models can be registered automatically by computing dense one-to-one correspondence to an internal face model. Second, the approach regulates the naturalness of modeled faces avoiding faces with an “unlikely” appearance.

Starting from an example set of 3D face models, we derive a morphable face model by transforming the shape and texture of the examples into a vector space representation. New faces and expressions can be modeled by forming linear combinations of the prototypes. Shape and texture constraints derived from the statistics of our example faces are used to guide manual modeling or automated matching algorithms.

We show 3D face reconstructions from single images and their applications for photo-realistic image manipulations. We also demonstrate face manipulations according to complex parameters such as gender, fullness of a face or its distinctiveness.

Keywords: facial modeling, registration, photogrammetry, morphing, facial animation, computer vision

1 Introduction

Computer aided modeling of human faces still requires a great deal of expertise and manual control to avoid unrealistic, non-face-like results. Most limitations of automated techniques for face synthesis, face animation or for general changes in the appearance of an individual face can be described either as the problem of finding corresponding feature locations in different faces or as the problem of separating realistic faces from faces that could never appear in the real world. The correspondence problem is crucial for all morphing techniques, both for the application of motion-capture data to pictures or 3D face models, and for most 3D face reconstruction techniques from images. A limited number of labeled feature points marked in one face, e.g., the tip of the nose, the eye corner and less prominent points on the cheek, must be located precisely in another face. The number of manually labeled feature points varies from

*MPI für biol. Kybernetik, Spemannstr. 38, 72076 Tübingen, Germany.
E-mail: {volker.blanz, thomas.vetter}@tuebingen.mpg.de

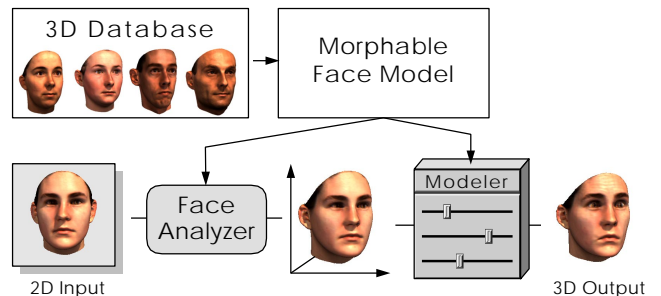


Figure 1: Derived from a dataset of prototypical 3D scans of faces, the morphable face model contributes to two main steps in face manipulation: (1) deriving a 3D face model from a novel image, and (2) modifying shape and texture in a natural way.

application to application, but usually ranges from 50 to 300.

Only a correct alignment of all these points allows acceptable intermediate morphs, a convincing mapping of motion data from the reference to a new model, or the adaptation of a 3D face model to 2D images for ‘video cloning’. Human knowledge and experience is necessary to compensate for the variations between individual faces and to guarantee a valid location assignment in the different faces. At present, automated matching techniques can be utilized only for very prominent feature points such as the corners of eyes and mouth.

A second type of problem in face modeling is the separation of natural faces from non faces. For this, human knowledge is even more critical. Many applications involve the design of completely new natural looking faces that can occur in the real world but which have no “real” counterpart. Others require the manipulation of an existing face according to changes in age, body weight or simply to emphasize the characteristics of the face. Such tasks usually require time-consuming manual work combined with the skills of an artist.

In this paper, we present a parametric face modeling technique that assists in both problems. First, arbitrary human faces can be created simultaneously controlling the likelihood of the generated faces. Second, the system is able to compute correspondence between new faces. Exploiting the statistics of a large dataset of 3D face scans (geometric and textural data, *CyberwareTM*) we built a morphable face model and recover domain knowledge about face variations by applying pattern classification methods. The morphable face model is a multidimensional 3D morphing function that is based on the linear combination of a large number of 3D face scans. Computing the average face and the main modes of variation in our dataset, a probability distribution is imposed on the morphing function to avoid unlikely faces. We also derive parametric descriptions of face attributes such as gender, distinctiveness, “hooked” noses or the weight of a person, by evaluating the distribution of exemplar faces for each attribute within our face space.

Having constructed a parametric face model that is able to generate almost any face, the correspondence problem turns into a mathematical optimization problem. New faces, images or 3D face scans, can be registered by minimizing the difference between the new face and its reconstruction by the face model function. We devel-

oped an algorithm that adjusts the model parameters automatically for an optimal reconstruction of the target, requiring only a minimum of manual initialization. The output of the matching procedure is a high quality 3D face model that is in full correspondence with our morphable face model. Consequently all face manipulations parameterized in our model function can be mapped to the target face. The prior knowledge about the shape and texture of faces in general that is captured in our model function is sufficient to make reasonable estimates of the full 3D shape and texture of a face even when only a single picture is available. When applying the method to several images of a person, the reconstructions reach almost the quality of laser scans.

1.1 Previous and related work

Modeling human faces has challenged researchers in computer graphics since its beginning. Since the pioneering work of Parke [25, 26], various techniques have been reported for modeling the geometry of faces [10, 11, 22, 34, 21] and for animating them [28, 14, 19, 32, 22, 38, 29]. A detailed overview can be found in the book of Parke and Waters [24].

The key part of our approach is a generalized model of human faces. Similar to the approach of DeCarlos et al. [10], we restrict the range of allowable faces according to constraints derived from prototypical human faces. However, instead of using a limited set of measurements and proportions between a set of facial landmarks, we directly use the densely sampled geometry of the exemplar faces obtained by laser scanning (*CyberwareTM*). The dense modeling of facial geometry (several thousand vertices per face) leads directly to a triangulation of the surface. Consequently, there is no need for variational surface interpolation techniques [10, 23, 33]. We also added a model of texture variations between faces. The morphable 3D face model is a consequent extension of the interpolation technique between face geometries, as introduced by Parke [26]. Computing correspondence between individual 3D face data automatically, we are able to increase the number of vertices used in the face representation from a few hundreds to tens of thousands. Moreover, we are able to use a higher number of faces, and thus to interpolate between hundreds of ‘basis’ faces rather than just a few. The goal of such an extended morphable face model is to represent any face as a linear combination of a limited basis set of face prototypes. Representing the face of an arbitrary person as a linear combination (morph) of ‘prototype’ faces was first formulated for image compression in telecommunications [8]. Image-based linear 2D face models that exploit large data sets of prototype faces were developed for face recognition and image coding [4, 18, 37].

Different approaches have been taken to automate the matching step necessary for building up morphable models. One class of techniques is based on optic flow algorithms [5, 4] and another on an active model matching strategy [12, 16]. Combinations of both techniques have been applied to the problem of image matching [36]. In this paper we extend this approach to the problem of matching 3D faces.

The correspondence problem between different three-dimensional face data has been addressed previously by Lee et al.[20]. Their shape-matching algorithm differs significantly from our approach in several respects. First, we compute the correspondence in high resolution, considering shape and texture data simultaneously. Second, instead of using a physical tissue model to constrain the range of allowed mesh deformations, we use the statistics of our example faces to keep deformations plausible. Third, we do not rely on routines that are specifically designed to detect the features exclusively found in faces, e.g., eyes, nose.

Our general matching strategy can be used not only to adapt the morphable model to a 3D face scan, but also to 2D images of faces. Unlike a previous approach [35], the morphable 3D face model is now directly matched to images, avoiding the detour of generat-

ing intermediate 2D morphable image models. As a consequence, head orientation, illumination conditions and other parameters can be free variables subject to optimization. It is sufficient to use rough estimates of their values as a starting point of the automated matching procedure.

Most techniques for ‘face cloning’, the reconstruction of a 3D face model from one or more images, still rely on manual assistance for matching a deformable 3D face model to the images [26, 1, 30]. The approach of Pighin et al. [28] demonstrates the high realism that can be achieved for the synthesis of faces and facial expressions from photographs where several images of a face are matched to a single 3D face model. Our automated matching procedure could be used to replace the manual initialization step, where several corresponding features have to be labeled in the presented images.

For the animation of faces, a variety of methods have been proposed. For a complete overview we again refer to the book of Parke and Waters [24]. The techniques can be roughly separated in those that rely on physical modeling of facial muscles [38, 17], and in those applying previously captured facial expressions to a face [25, 3]. These performance based animation techniques compute the correspondence between the different facial expressions of a person by tracking markers glued to the face from image to image. To obtain photo-realistic face animations, up to 182 markers are used [14]. Working directly on faces without markers, our automated approach extends this number to its limit. It matches the full number of vertices available in the face model to images. The resulting dense correspondence fields can even capture changes in wrinkles and map these from one face to another.

1.2 Organization of the paper

We start with a description of the database of 3D face scans from which our morphable model is built.

In Section 3, we introduce the concept of the morphable face model, assuming a set of 3D face scans that are in full correspondence. Exploiting the statistics of a dataset, we derive a parametric description of faces, as well as the range of plausible faces. Additionally, we define facial attributes, such as gender or fullness of faces, in the parameter space of the model.

In Section 4, we describe an algorithm for matching our flexible model to novel images or 3D scans of faces. Along with a 3D reconstruction, the algorithm can compute correspondence, based on the morphable model.

In Section 5, we introduce an iterative method for building a morphable model automatically from a raw data set of 3D face scans when no correspondences between the exemplar faces are available.

2 Database

Laser scans (*CyberwareTM*) of 200 heads of young adults (100 male and 100 female) were used. The laser scans provide head structure data in a cylindrical representation, with radii $r(h, \phi)$ of surface points sampled at 512 equally-spaced angles ϕ , and at 512 equally spaced vertical steps h . Additionally, the RGB-color values $R(h, \phi)$, $G(h, \phi)$, and $B(h, \phi)$, were recorded in the same spatial resolution and were stored in a texture map with 8 bit per channel.

All faces were without makeup, accessories, and facial hair. The subjects were scanned wearing bathing caps, that were removed digitally. Additional automatic pre-processing of the scans, which for most heads required no human interaction, consisted of a vertical cut behind the ears, a horizontal cut to remove the shoulders, and a normalization routine that brought each face to a standard orientation and position in space. The resultant faces were represented by approximately 70,000 vertices and the same number of color values.

3 Morphable 3D Face Model

The morphable model is based on a data set of 3D faces. Morphing between faces requires full correspondence between all of the faces. In this section, we will assume that all exemplar faces are in full correspondence. The algorithm for computing correspondence will be described in Section 5.

We represent the geometry of a face with a shape-vector $S = (X_1, Y_1, Z_1, X_2, \dots, Y_n, Z_n)^T \in \mathbb{R}^{3n}$, that contains the X, Y, Z -coordinates of its n vertices. For simplicity, we assume that the number of valid texture values in the texture map is equal to the number of vertices. We therefore represent the texture of a face by a texture-vector $T = (R_1, G_1, B_1, R_2, \dots, G_n, B_n)^T \in \mathbb{R}^{3n}$, that contains the R, G, B color values of the n corresponding vertices. A morphable face model was then constructed using a data set of m exemplar faces, each represented by its shape-vector S_i and texture-vector T_i . Since we assume all faces in full correspondence (see Section 5), new shapes S_{model} and new textures T_{model} can be expressed in barycentric coordinates as a linear combination of the shapes and textures of the m exemplar faces:

$$S_{model} = \sum_{i=1}^m a_i S_i, \quad T_{model} = \sum_{i=1}^m b_i T_i, \quad \sum_{i=1}^m a_i = \sum_{i=1}^m b_i = 1.$$

We define the morphable model as the set of faces $(S_{model}(\vec{a}), T_{model}(\vec{b}))$, parameterized by the coefficients $\vec{a} = (a_1, a_2, \dots, a_m)^T$ and $\vec{b} = (b_1, b_2, \dots, b_m)^T$.¹ Arbitrary new faces can be generated by varying the parameters \vec{a} and \vec{b} that control shape and texture.

For a useful face synthesis system, it is important to be able to quantify the results in terms of their plausibility of being faces. We therefore estimated the probability distribution for the coefficients a_i and b_i from our example set of faces. This distribution enables us to control the likelihood of the coefficients a_i and b_i and consequently regulates the likelihood of the appearance of the generated faces.

We fit a multivariate normal distribution to our data set of 200 faces, based on the averages of shape \bar{S} and texture \bar{T} and the covariance matrices C_S and C_T computed over the shape and texture differences $\Delta S_i = S_i - \bar{S}$ and $\Delta T_i = T_i - \bar{T}$.

A common technique for data compression known as Principal Component Analysis (PCA) [15, 31] performs a basis transformation to an orthogonal coordinate system formed by the eigenvectors s_i and t_i of the covariance matrices (in descending order according to their eigenvalues)²:

$$S_{model} = \bar{S} + \sum_{i=1}^{m-1} \alpha_i s_i, \quad T_{model} = \bar{T} + \sum_{i=1}^{m-1} \beta_i t_i, \quad (1)$$

$\vec{\alpha}, \vec{\beta} \in \mathbb{R}^{m-1}$. The probability for coefficients $\vec{\alpha}$ is given by

$$p(\vec{\alpha}) \sim \exp\left[-\frac{1}{2} \sum_{i=1}^{m-1} (\alpha_i / \sigma_i)^2\right], \quad (2)$$

with σ_i^2 being the eigenvalues of the shape covariance matrix C_S . The probability $p(\vec{\beta})$ is computed similarly.

Segmented morphable model: The morphable model described in equation (1), has $m - 1$ degrees of freedom for texture and $m - 1$ for shape. The expressiveness of the model can

¹Standard morphing between two faces ($m = 2$) is obtained if the parameters a_1, b_1 are varied between 0 and 1, setting $a_2 = 1 - a_1$ and $b_2 = 1 - b_1$.

²Due to the subtracted average vectors \bar{S} and \bar{T} , the dimensions of $Span\{\Delta S_i\}$ and $Span\{\Delta T_i\}$ are at most $m - 1$.

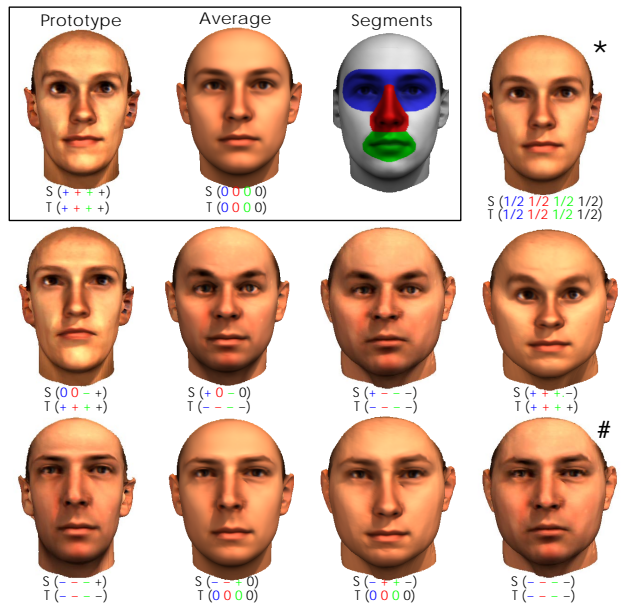


Figure 2: A single prototype adds a large variety of new faces to the morphable model. The deviation of a prototype from the average is added (+) or subtracted (-) from the average. A standard morph (*) is located halfway between average and the prototype. Subtracting the differences from the average yields an 'anti'-face (#). Adding and subtracting deviations independently for shape (S) and texture (T) on each of four segments produces a number of distinct faces.

be increased by dividing faces into independent subregions that are morphed independently, for example into eyes, nose, mouth and a surrounding region (see Figure 2). Since all faces are assumed to be in correspondence, it is sufficient to define these regions on a reference face. This segmentation is equivalent to subdividing the vector space of faces into independent subspaces. A complete 3D face is generated by computing linear combinations for each segment separately and blending them at the borders according to an algorithm proposed for images by [7].

3.1 Facial attributes

Shape and texture coefficients α_i and β_i in our morphable face model do not correspond to the facial attributes used in human language. While some facial attributes can easily be related to biophysical measurements [13, 10], such as the width of the mouth, others such as facial femininity or being more or less bony can hardly be described by numbers. In this section, we describe a method for mapping facial attributes, defined by a hand-labeled set of example faces, to the parameter space of our morphable model. At each position in face space (that is for any possible face), we define shape and texture vectors that, when added to or subtracted from a face, will manipulate a specific attribute while keeping all other attributes as constant as possible.

In a performance based technique [25], facial expressions can be transferred by recording two scans of the same individual with different expressions, and adding the differences $\Delta S = S_{expression} - S_{neutral}$, $\Delta T = T_{expression} - T_{neutral}$, to a different individual in a neutral expression.

Unlike facial expressions, attributes that are invariant for each individual are more difficult to isolate. The following method allows us to model facial attributes such as gender, fullness of faces, darkness of eyebrows, double chins, and hooked versus concave noses (Figure 3). Based on a set of faces (S_i, T_i) with manually assigned labels μ_i describing the markedness of the attribute, we compute

weighted sums

$$\Delta S = \sum_{i=1}^m \mu_i (S_i - \bar{S}), \quad \Delta T = \sum_{i=1}^m \mu_i (T_i - \bar{T}). \quad (3)$$

Multiples of $(\Delta S, \Delta T)$ can now be added to or subtracted from any individual face. For binary attributes, such as gender, we assign constant values μ_A for all m_A faces in class A , and $\mu_B \neq \mu_A$ for all m_B faces in B . Affecting only the scaling of ΔS and ΔT , the choice of μ_A, μ_B is arbitrary.

To justify this method, let $\mu(S, T)$ be the overall function describing the markedness of the attribute in a face (S, T) . Since $\mu(S, T)$ is not available per se for all (S, T) , the regression problem of estimating $\mu(S, T)$ from a sample set of labeled faces has to be solved. Our technique assumes that $\mu(S, T)$ is a linear function. Consequently, in order to achieve a change $\Delta\mu$ of the attribute, there is only a single optimal direction $(\Delta S, \Delta T)$ for the whole space of faces. It can be shown that Equation (3) defines the direction with minimal variance-normalized length $\|\Delta S\|_M^2 = \langle \Delta S, C_S^{-1} \Delta S \rangle, \|\Delta T\|_M^2 = \langle \Delta T, C_T^{-1} \Delta T \rangle$.

A different kind of facial attribute is its “distinctiveness”, which is commonly manipulated in caricatures. The automated production of caricatures has been possible for many years [6]. This technique can easily be extended from 2D images to our morphable face model. Individual faces are caricatured by increasing their distance from the average face. In our representation, shape and texture coefficients α_i, β_i are simply multiplied by a constant factor.

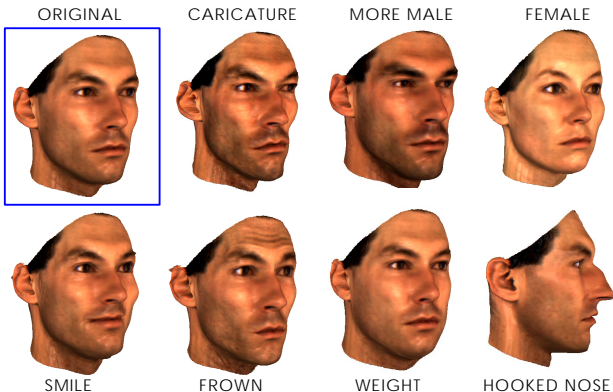


Figure 3: Variation of facial attributes of a single face. The appearance of an original face can be changed by adding or subtracting shape and texture vectors specific to the attribute.

4 Matching a morphable model to images

A crucial element of our framework is an algorithm for automatically matching the morphable face model to one or more images. Providing an estimate of the face’s 3D structure (Figure 4), it closes the gap between the specific manipulations described in Section 3.1, and the type of data available in typical applications.

Coefficients of the 3D model are optimized along with a set of rendering parameters such that they produce an image as close as possible to the input image. In an analysis-by-synthesis loop, the algorithm creates a texture mapped 3D face from the current model parameters, renders an image, and updates the parameters according to the residual difference. It starts with the average head and with rendering parameters roughly estimated by the user.

Model Parameters: Facial shape and texture are defined by coefficients α_j and $\beta_j, j = 1, \dots, m - 1$ (Equation 1). Rendering parameters $\vec{\rho}$ contain camera position (azimuth and elevation), object scale, image plane rotation and translation, intensity $i_{r,amb}, i_{g,amb}, i_{b,amb}$ of ambient light, and intensity

2D Input

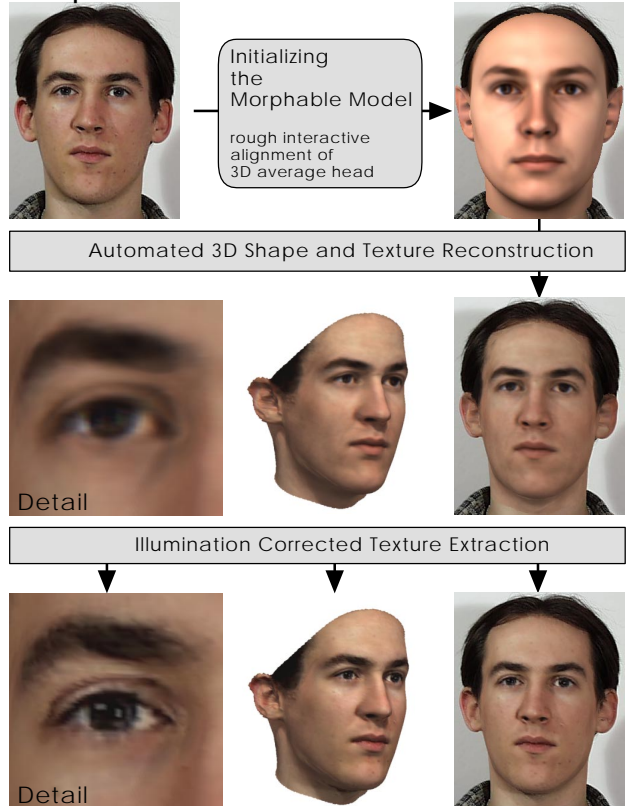


Figure 4: Processing steps for reconstructing 3D shape and texture of a new face from a single image. After a rough manual alignment of the average 3D head (top row), the automated matching procedure fits the 3D morphable model to the image (center row). In the right column, the model is rendered on top of the input image. Details in texture can be improved by illumination-corrected texture extraction from the input (bottom row).

of directed light. In order to handle photographs taken under a wide variety of conditions, $\vec{\rho}$ also includes color contrast as well as offset and gain in the red, green, and blue channel. Other parameters, such as camera distance, light direction, and surface shininess, remain fixed to the values estimated by the user.

From parameters $(\vec{\alpha}, \vec{\beta}, \vec{\rho})$, colored images

$$\mathbf{I}_{model}(x, y) = (I_{r,mod}(x, y), I_{g,mod}(x, y), I_{b,mod}(x, y))^T \quad (4)$$

are rendered using perspective projection and the Phong illumination model. The reconstructed image is supposed to be closest to the input image in terms of Euclidean distance

$$E_I = \sum_{x,y} \|\mathbf{I}_{input}(x, y) - \mathbf{I}_{model}(x, y)\|^2.$$

Matching a 3D surface to a given image is an ill-posed problem. Along with the desired solution, many non-face-like surfaces lead to the same image. It is therefore essential to impose constraints on the set of solutions. In our morphable model, shape and texture vectors are restricted to the vector space spanned by the database.

Within the vector space of faces, solutions can be further restricted by a tradeoff between matching quality and prior probabilities, using $P(\vec{\alpha}), P(\vec{\beta})$ from Section 3 and an ad-hoc estimate of $P(\vec{\rho})$. In terms of Bayes decision theory, the problem is to find the set of parameters $(\vec{\alpha}, \vec{\beta}, \vec{\rho})$ with maximum posterior probability, given an image \mathbf{I}_{input} . While $\vec{\alpha}, \vec{\beta}$, and rendering parameters $\vec{\rho}$ completely determine the predicted image \mathbf{I}_{model} , the observed image \mathbf{I}_{input} may vary due to noise. For Gaussian noise

with a standard deviation σ_N , the likelihood to observe I_{input} is $p(\mathbf{I}_{input}|\bar{\alpha}, \bar{\beta}, \bar{\rho}) \sim \exp[-\frac{1}{2\sigma_N^2} \cdot E_I]$. Maximum posterior probability is then achieved by minimizing the cost function

$$E = \frac{1}{\sigma_N^2} E_I + \sum_{j=1}^{m-1} \frac{\alpha_j^2}{\sigma_{S,j}^2} + \sum_{j=1}^{m-1} \frac{\beta_j^2}{\sigma_{T,j}^2} + \sum_j \frac{(\rho_j - \bar{\rho}_j)^2}{\sigma_{\rho,j}^2} \quad (5)$$

The optimization algorithm described below uses an estimate of E based on a random selection of surface points. Predicted color values \mathbf{I}_{model} are easiest to evaluate in the centers of triangles. In the center of triangle k , texture $(\bar{R}_k, \bar{G}_k, \bar{B}_k)^T$ and 3D location $(\bar{X}_k, \bar{Y}_k, \bar{Z}_k)^T$ are averages of the values at the corners. Perspective projection maps these points to image locations $(\bar{p}_{x,k}, \bar{p}_{y,k})^T$. Surface normals \mathbf{n}_k of each triangle k are determined by the 3D locations of the corners. According to Phong illumination, the color components $I_{r,model}$, $I_{g,model}$ and $I_{b,model}$ take the form

$$I_{r,model,k} = (i_{r,amb} + i_{r,dir} \cdot (\mathbf{n}_k \mathbf{l}) \bar{R}_k + i_{r,dir} s \cdot (\mathbf{r}_k \mathbf{v}_k)^\nu) \quad (6)$$

where \mathbf{l} is the direction of illumination, \mathbf{v}_k the normalized difference of camera position and the position of the triangle's center, and $\mathbf{r}_k = 2(\mathbf{n}_k \mathbf{l}) \mathbf{n} - \mathbf{l}$ the direction of the reflected ray. s denotes surface shininess, and ν controls the angular distribution of the specular reflection. Equation (6) reduces to $I_{r,model,k} = i_{r,amb} \bar{R}_k$ if a shadow is cast on the center of the triangle, which is tested in a method described below.

For high resolution 3D meshes, variations in \mathbf{I}_{model} across each triangle $k \in \{1, \dots, n_t\}$ are small, so E_I may be approximated by

$$E_I \approx \sum_{k=1}^{n_t} a_k \cdot \|\mathbf{I}_{input}(\bar{p}_{x,k}, \bar{p}_{y,k}) - \mathbf{I}_{model,k}\|^2,$$

where a_k is the image area covered by triangle k . If the triangle is occluded, $a_k = 0$.

In gradient descent, contributions from different triangles of the mesh would be redundant. In each iteration, we therefore select a random subset $\mathcal{K} \subset \{1, \dots, n_t\}$ of 40 triangles k and replace E_I by

$$E_{\mathcal{K}} = \sum_{k \in \mathcal{K}} \|\mathbf{I}_{input}(\bar{p}_{x,k}, \bar{p}_{y,k}) - \mathbf{I}_{model,k}\|^2. \quad (7)$$

The probability of selecting k is $p(k \in \mathcal{K}) \sim a_k$. This method of stochastic gradient descent [16] is not only more efficient computationally, but also helps to avoid local minima by adding noise to the gradient estimate.

Before the first iteration, and once every 1000 steps, the algorithm computes the full 3D shape of the current model, and 2D positions $(p_x, p_y)^T$ of all vertices. It then determines a_k , and detects hidden surfaces and cast shadows in a two-pass z-buffer technique. We assume that occlusions and cast shadows are constant during each subset of iterations.

Parameters are updated depending on analytical derivatives of the cost function E , using $\alpha_j \mapsto \alpha_j - \lambda_j \cdot \frac{\partial E}{\partial \alpha_j}$, and similarly for β_j and ρ_j , with suitable factors λ_j .

Derivatives of texture and shape (Equation 1) yield derivatives of 2D locations $(\bar{p}_{x,k}, \bar{p}_{y,k})^T$, surface normals \mathbf{n}_k , vectors \mathbf{v}_k and \mathbf{r}_k , and $\mathbf{I}_{model,k}$ (Equation 6) using chain rule. From Equation (7), partial derivatives $\frac{\partial E_{\mathcal{K}}}{\partial \alpha_j}$, $\frac{\partial E_{\mathcal{K}}}{\partial \beta_j}$, and $\frac{\partial E_{\mathcal{K}}}{\partial \rho_j}$ can be obtained.

Coarse-to-Fine: In order to avoid local minima, the algorithm follows a coarse-to-fine strategy in several respects:

- The first set of iterations is performed on a down-sampled version of the input image with a low resolution morphable model.
- We start by optimizing only the first coefficients α_j and β_j controlling the first principal components, along with all parameters

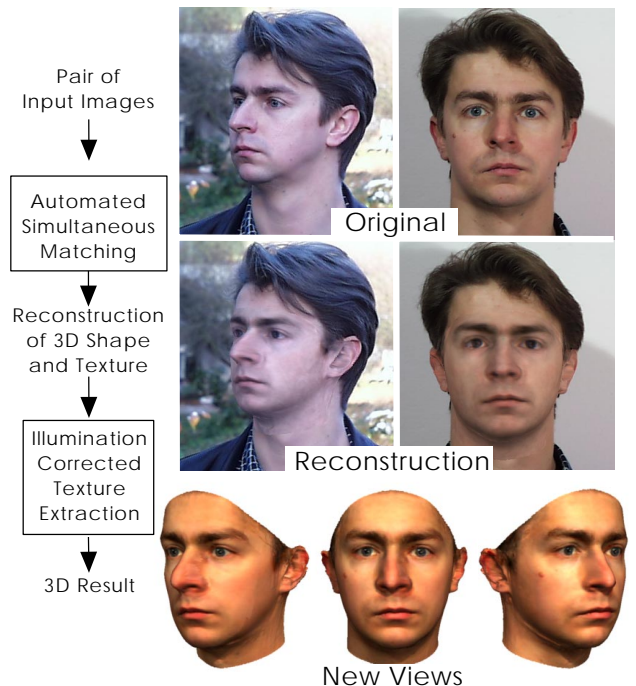


Figure 5: Simultaneous reconstruction of 3D shape and texture of a new face from two images taken under different conditions. In the center row, the 3D face is rendered on top of the input images.

ρ_j . In subsequent iterations, more and more principal components are added.

c) Starting with a relatively large σ_N , which puts a strong weight on prior probability in equation (5) and ties the optimum towards the prior expectation value, we later reduce σ_N to obtain maximum matching quality.

d) In the last iterations, the face model is broken down into segments (Section 3). With parameters ρ_j fixed, coefficients α_j and β_j are optimized independently for each segment. This increased number of degrees of freedom significantly improves facial details.

Multiple Images: It is straightforward to extend this technique to the case where several images of a person are available (Figure 5). While shape and texture are still described by a common set of α_j and β_j , there is now a separate set of ρ_j for each input image. E_I is replaced by a sum of image distances for each pair of input and model images, and all parameters are optimized simultaneously.

Illumination-Corrected Texture Extraction: Specific features of individual faces that are not captured by the morphable model, such as blemishes, are extracted from the image in a subsequent texture adaptation process. Extracting texture from images is a technique widely used in constructing 3D models from images (e.g. [28]). However, in order to be able to change pose and illumination, it is important to separate pure albedo at any given point from the influence of shading and cast shadows in the image. In our approach, this can be achieved because our matching procedure provides an estimate of 3D shape, pose, and illumination conditions. Subsequent to matching, we compare the prediction $\mathbf{I}_{mod,i}$ for each vertex i with $\mathbf{I}_{input}(p_{x,i}, p_{y,i})$, and compute the change in texture (R_i, G_i, B_i) that accounts for the difference. In areas occluded in the image, we rely on the prediction made by the model. Data from multiple images can be blended using methods similar to [28].

4.1 Matching a morphable model to 3D scans

The method described above can also be applied to register new 3D faces. Analogous to images, where perspective projection

$P : \mathcal{R}^3 \rightarrow \mathcal{R}^2$ and an illumination model define a colored image $\mathbf{I}(x, y) = (R(x, y), G(x, y), B(x, y))^T$, laser scans provide a two-dimensional cylindrical parameterization of the surface by means of a mapping $C : \mathcal{R}^3 \rightarrow \mathcal{R}^2$, $(x, y, z) \mapsto (h, \phi)$. Hence, a scan can be represented as

$$\mathbf{I}(h, \phi) = (R(h, \phi), G(h, \phi), B(h, \phi), r(h, \phi))^T. \quad (8)$$

In a face (S, T) , defined by shape and texture coefficients α_j and β_j (Equation 1), vertex i with texture values (R_i, G_i, B_i) and cylindrical coordinates (r_i, h_i, ϕ_i) is mapped to $\mathbf{I}_{model}(h_i, \phi_i) = (R_i, G_i, B_i, r_i)^T$. The matching algorithm from the previous section now determines α_j and β_j minimizing

$$E = \sum_{h, \phi} \|\mathbf{I}_{input}(h, \phi) - \mathbf{I}_{model}(h, \phi)\|^2.$$

5 Building a morphable model

In this section, we describe how to build the morphable model from a set of unregistered 3D prototypes, and to add a new face to the existing morphable model, increasing its dimensionality.

The key problem is to compute a dense point-to-point correspondence between the vertices of the faces. Since the method described in Section 4.1 finds the best match of a given face only within the range of the morphable model, it cannot add new dimensions to the vector space of faces. To determine residual deviations between a novel face and the best match within the model, as well as to set unregistered prototypes in correspondence, we use an optic flow algorithm that computes correspondence between two faces without the need of a morphable model [35]. The following section summarizes this technique.

5.1 3D Correspondence using Optic Flow

Initially designed to find corresponding points in grey-level images $I(x, y)$, a gradient-based optic flow algorithm [2] is modified to establish correspondence between a pair of 3D scans $\mathbf{I}(h, \phi)$ (Equation 8), taking into account color and radius values simultaneously [35]. The algorithm computes a flow field $(\delta h(h, \phi), \delta \phi(h, \phi))$ that minimizes differences of $\|\mathbf{I}_1(h, \phi) - \mathbf{I}_2(h + \delta h, \phi + \delta \phi)\|$ in a norm that weights variations in texture and shape equally. Surface properties from differential geometry, such as mean curvature, may be used as additional components in $\mathbf{I}(h, \phi)$.

On facial regions with little structure in texture and shape, such as forehead and cheeks, the results of the optic flow algorithm are sometimes spurious. We therefore perform a smooth interpolation based on simulated relaxation of a system of flow vectors that are coupled with their neighbors. The quadratic coupling potential is equal for all flow vectors. On high-contrast areas, components of flow vectors orthogonal to edges are bound to the result of the previous optic flow computation. The system is otherwise free to take on a smooth minimum-energy arrangement. Unlike simple filtering routines, our technique fully retains matching quality wherever the flow field is reliable. Optic flow and smooth interpolation are computed on several consecutive levels of resolution.

Constructing a morphable face model from a set of unregistered 3D scans requires the computation of the flow fields between each face and an arbitrary reference face. Given a definition of shape and texture vectors S_{ref} and T_{ref} for the reference face, S and T for each face in the database can be obtained by means of the point-to-point correspondence provided by $(\delta h(h, \phi), \delta \phi(h, \phi))$.

5.2 Bootstrapping the model

Because the optic flow algorithm does not incorporate any constraints on the set of solutions, it fails on some of the more unusual

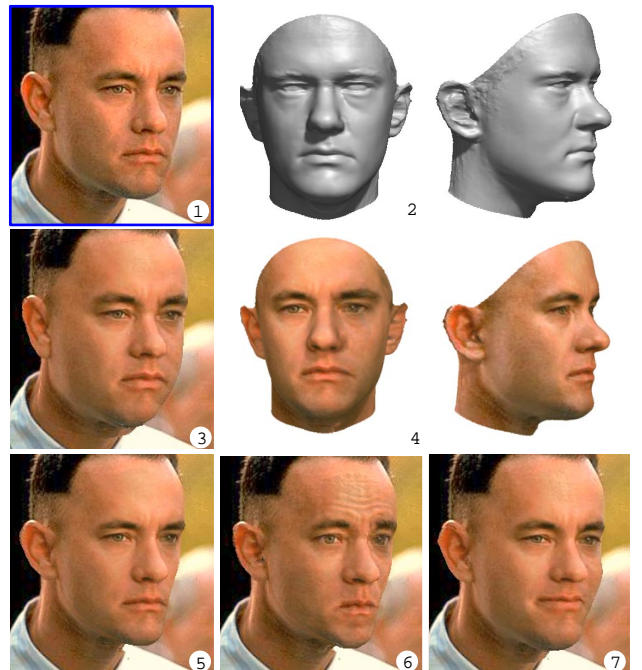


Figure 6: Matching a morphable model to a single image (1) of a face results in a 3D shape (2) and a texture map estimate. The texture estimate can be improved by additional texture extraction (4). The 3D model is rendered back into the image after changing facial attributes, such as gaining (3) and losing weight (5), frowning (6), or being forced to smile (7).

faces in the database. Therefore, we modified a bootstrapping algorithm to iteratively improve correspondence, a method that has been used previously to build linear image models [36].

The basic recursive step: Suppose that an existing morphable model is not powerful enough to match a new face and thereby find correspondence with it. The idea is first to find rough correspondences to the novel face using the (inadequate) morphable model and then to improve these correspondences by using an optic flow algorithm.

Starting from an arbitrary face as the temporary reference, preliminary correspondence between all other faces and this reference is computed using the optic flow algorithm. On the basis of these correspondences, shape and texture vectors S and T can be computed. Their average serves as a new reference face. The first morphable model is then formed by the most significant components as provided by a standard PCA decomposition. The current morphable model is now matched to each of the 3D faces according to the method described in Section 4.1. Then, the optic flow algorithm computes correspondence between the 3D face and the approximation provided by the morphable model. Combined with the correspondence implied by the matched model, this defines a new correspondence between the reference face and the example.

Iterating this procedure with increasing expressive power of the model (by increasing the number of principal components) leads to reliable correspondences between the reference face and the examples, and finally to a complete morphable face model.

6 Results

We built a morphable face model by automatically establishing correspondence between all of our 200 exemplar faces. Our interactive

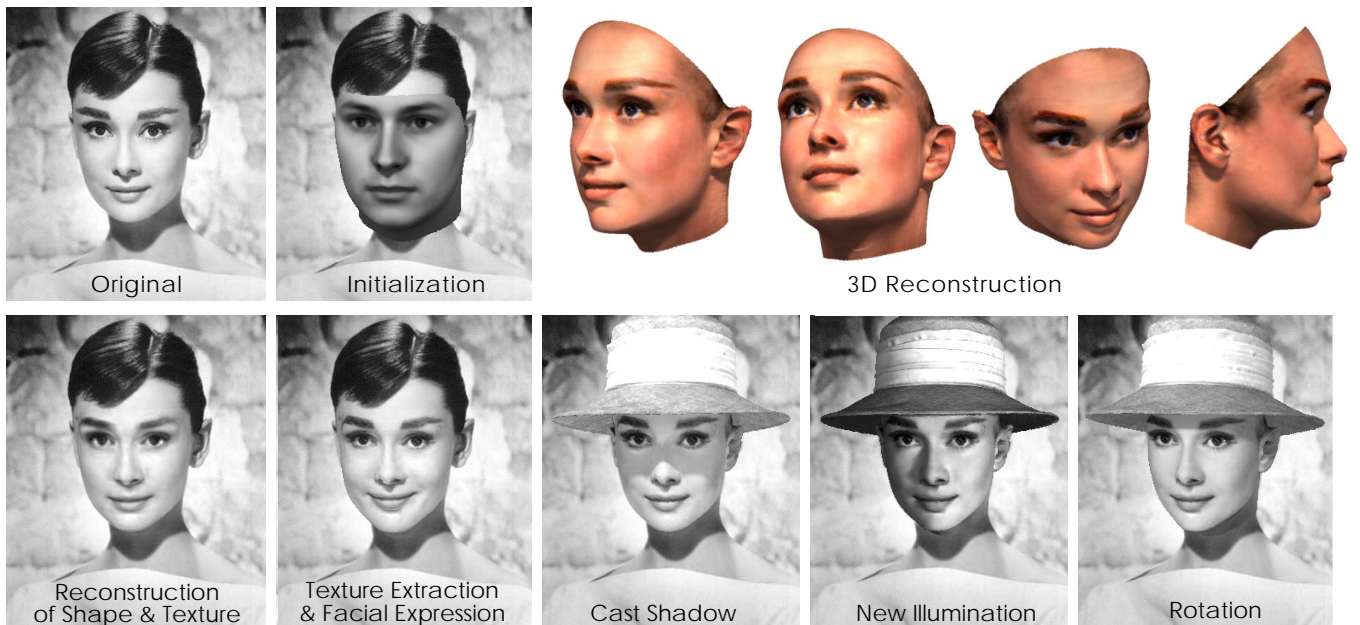


Figure 7: After manual initialization, the algorithm automatically matches a colored morphable model (color contrast set to zero) to the image. Rendering the inner part of the 3D face on top of the image, new shadows, facial expressions and poses can be generated.

face modeling system enables human users to create new characters and to modify facial attributes by varying the model coefficients. Within the constraints imposed by prior probability, there is a large variability of possible faces, and all linear combinations of the exemplar faces look natural.

We tested the expressive power of our morphable model by automatically reconstructing 3D faces from photographs of arbitrary Caucasian faces of middle age that were not in the database. The images were either taken by us using a digital camera (Figures 4, 5), or taken under arbitrary unknown conditions (Figures 6, 7).

In all examples, we matched a morphable model built from the first 100 shape and the first 100 texture principal components that were derived from the whole dataset of 200 faces. Each component was additionally segmented in 4 parts (see Figure 2). The whole matching procedure was performed in 10^5 iterations. On an SGI R10000 processor, computation time was 50 minutes.

Reconstructing the true 3D shape and texture of a face from a single image is an ill-posed problem. However, to human observers who also know only the input image, the results obtained with our method look correct. When compared with a real image of the rotated face, differences usually become only visible for large rotations of more than 60° .

There is a wide variety of applications for 3D face reconstruction from 2D images. As demonstrated in Figures 6 and 7, the results can be used for automatic post-processing of a face within the original picture or movie sequence.

Knowing the 3D shape of a face in an image provides a segmentation of the image into face area and background. The face can be combined with other 3D graphic objects, such as glasses or hats, and then be rendered in front of the background, computing cast shadows or new illumination conditions (Fig. 7). Furthermore, we can change the appearance of the face by adding or subtracting specific attributes. If previously unseen backgrounds become visible, we fill the holes with neighboring background pixels (Fig. 6).

We also applied the method to paintings such as Leonardo's Mona Lisa (Figure 8). Due to unusual (maybe unrealistic) lighting, illumination-corrected texture extraction is difficult here. We therefore apply a different method for transferring all details of the

painting to novel views. For new illumination, we render two images of the reconstructed 3D face with different illumination, and multiply relative changes in pixel values (Figure 8, bottom left) by the original values in the painting (bottom center). For a new pose (bottom right), differences in shading are transferred in a similar way, and the painting is then warped according to the 2D projections of 3D vertex displacements of the reconstructed shape.

7 Future work

Issues of implementation: We plan to speed up our matching algorithm by implementing a simplified Newton-method for minimizing the cost function (Equation 5). Instead of the time consuming computation of derivatives for each iteration step, a global mapping of the matching error into parameter space can be used [9].

Data reduction applied to shape and texture data will reduce redundancy of our representation, saving additional computation time.

Extending the database: While the current database is sufficient to model Caucasian faces of middle age, we would like to extend it to children, to elderly people as well as to other races.

We also plan to incorporate additional 3D face examples representing the time course of facial expressions and visemes, the face variations during speech.

The laser scanning technology we used, unfortunately, does not allow us to collect dynamical 3D face data, as each scanning cycle takes at least 10 seconds. Consequently, our current example set of facial expressions is restricted to those that can be kept static by the scanned subjects. However, the development of fast optical 3D digitizers [27] will allow us to apply our method to streams of 3D data during speech and facial expressions.

Extending the face model: Our current morphable model is restricted to the face area, because a sufficient 3D model of hair cannot be obtained with our laser scanner. For animation, the missing part of the head can be automatically replaced by a standard hair style or a hat, or by hair that is modeled using interactive manual segmentation and adaptation to a 3D model [30, 28]. Automated reconstruction of hair styles from images is one of the future challenges.

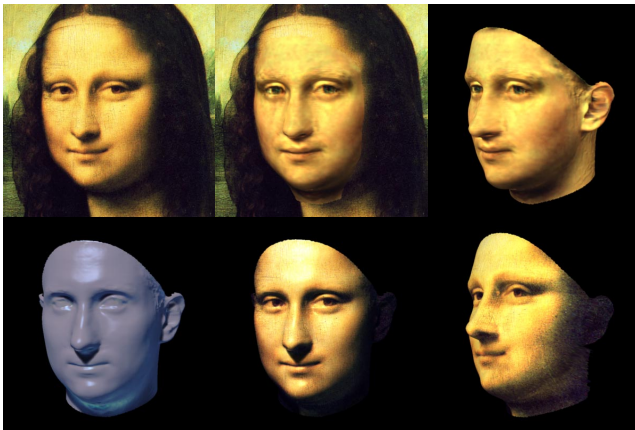


Figure 8: Reconstructed 3D face of Mona Lisa (top center and right). For modifying the illumination, relative changes in color (bottom left) are computed on the 3D face, and then multiplied by the color values in the painting (bottom center). Additional warping generates new orientations (bottom right, see text), while details of the painting, such as brush strokes or cracks, are retained.

8 Acknowledgment

We thank Michael Langer, Alice O'Toole, Tomaso Poggio, Heinrich Bülthoff and Wolfgang Straßer for reading the manuscript and for many insightful and constructive comments. In particular, we thank Marney Smyth and Alice O'Toole for their perseverance in helping us to obtain the following. **Photo Credits:** Original image in Fig. 6: Courtesy of Paramount/VIACOM. Original image in Fig. 7: MPTV/interTOPICS.

References

- [1] T. Akimoto, Y. Suenaga, and R.S. Wallace. Automatic creation of 3D facial models. *IEEE Computer Graphics and Applications*, 13(3):16–22, 1993.
- [2] J.R. Bergen and R. Hingorani. Hierarchical motion-based frame rate conversion. Technical report, David Sarnoff Research Center Princeton NJ 08540, 1990.
- [3] P. Bergeron and P. Lachapelle. Controlling facial expressions and body movements. In *Advanced Computer Animation, SIGGRAPH '85 Tutorials*, volume 2, pages 61–79, New York, 1985. ACM.
- [4] D. Beymer and T. Poggio. Image representation for visual learning. *Science*, 272:1905–1909, 1996.
- [5] D. Beymer, A. Shashua, and T. Poggio. Example-based image analysis and synthesis. A.I. Memo No. 1431, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1993.
- [6] S. E. Brennan. The caricature generator. *Leonardo*, 18:170–178, 1985.
- [7] P.J. Burt and E.H. Adelson. Merging images through pattern decomposition. In *Applications of Digital Image Processing VIII*, number 575, pages 173–181. SPIE The International Society for Optical Engineering, 1985.
- [8] C.S. Choi, T. Okazaki, H. Harashima, and T. Takebe. A system of analyzing and synthesizing facial images. In *Proc. IEEE Int. Symposium of Circuit and Systems (ISCAS91)*, pages 2665–2668, 1991.
- [9] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. In Burkhardt and Neumann, editors, *Computer Vision – ECCV'98 Vol. II*, Freiburg, Germany, 1998. Springer, Lecture Notes in Computer Science 1407.
- [10] D. DeCarlos, D. Metaxas, and M. Stone. An anthropometric face model using variational techniques. In *Computer Graphics Proceedings SIGGRAPH'98*, pages 67–74, 1998.
- [11] S. DiPaola. Extending the range of facial types. *Journal of Visualization and Computer Animation*, 2(4):129–131, 1991.
- [12] G.J. Edwards, A. Lanitis, C.J. Taylor, and T.F. Cootes. Modelling the variability in face images. In *Proc. of the 2nd Int. Conf. on Automatic Face and Gesture Recognition*, IEEE Comp. Soc. Press, Los Alamitos, CA, 1996.
- [13] L.G. Farkas. *Anthropometry of the Head and Face*. RavenPress, New York, 1994.
- [14] B. Guenter, C. Grimm, D. Wolf, H. Malvar, and F. Pighin. Making faces. In *Computer Graphics Proceedings SIGGRAPH '98*, pages 55–66, 1998.
- [15] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.
- [16] M. Jones and T. Poggio. Multidimensional morphable models: A framework for representing and matching object classes. In *Proceedings of the Sixth International Conference on Computer Vision*, Bombay, India, 1998.
- [17] R. M. Koch, M. H. Gross, and A. A. Bosshard. Emotion editing using finite elements. In *Proceedings of the Eurographics '98, COMPUTER GRAPHICS Forum, Vol. 17, No. 3*, pages C295–C302, Lisbon, Portugal, 1998.
- [18] A. Lanitis, C.J. Taylor, and T.F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.
- [19] Y.C. Lee, D. Terzopoulos, and Keith Waters. Constructing physics-based facial models of individuals. *Visual Computer*, Proceedings of Graphics Interface '93:1–8, 1993.
- [20] Y.C. Lee, D. Terzopoulos, and Keith Waters. Realistic modeling for facial animation. In *SIGGRAPH '95 Conference Proceedings*, pages 55–62, Los Angeles, 1995. ACM.
- [21] J. P. Lewis. Algorithms for solid noise synthesis. In *SIGGRAPH '89 Conference Proceedings*, pages 263–270. ACM, 1989.
- [22] N. Magnenat-Thalmann, H. Minh, M. Angelis, and D. Thalmann. Design, transformation and animation of human faces. *Visual Computer*, 5:32–39, 1989.
- [23] L. Moccozet and N. Magnenat-Thalmann. Dirichlet free-form deformation and their application to hand simulation. In *Computer Animation '97*, 1997.
- [24] F. I. Parke and K. Waters. *Computer Facial Animation*. AKPeters, Wellesley, Massachusetts, 1996.
- [25] F.I. Parke. Computer generated animation of faces. In *ACM National Conference*. ACM, November 1972.
- [26] F.I. Parke. *A Parametric Model of Human Faces*. PhD thesis, University of Utah, Salt Lake City, 1974.
- [27] M. Petrow, A. Talapov, T. Robertson, A. Lebedev, A. Zhilyaev, and L. Polonskiy. Optical 3D digitizer: Bringing life to virtual world. *IEEE Computer Graphics and Applications*, 18(3):28–37, 1998.
- [28] F. Pighin, J. Hecker, D. Lischinski, Szeliski R, and D. Salesin. Synthesizing realistic facial expressions from photographs. In *Computer Graphics Proceedings SIGGRAPH'98*, pages 75–84, 1998.
- [29] S. Platt and N. Badler. Animating facial expression. *Computer Graphics*, 15(3):245–252, 1981.
- [30] G. Sannier and N. Magnenat-Thalmann. A user-friendly texture-fitting methodology for virtual humans. In *Computer Graphics International '97*, 1997.
- [31] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4:519–554, 1987.
- [32] D. Terzopoulos and Keith Waters. Physically-based facial modeling, analysis, and animation. *Visualization and Computer Animation*, 1:73–80, 1990.
- [33] Demetri Terzopoulos and Hong Qin. Dynamic NURBS with geometric constraints to interactive sculpting. *ACM Transactions on Graphics*, 13(2):103–136, April 1994.
- [34] J. T. Todd, S. M. Leonard, R. E. Shaw, and J. B. Pittenger. The perception of human growth. *Scientific American*, 1242:106–114, 1980.
- [35] T. Vetter and V. Blanz. Estimating coloured 3d face models from single images: An example based approach. In Burkhardt and Neumann, editors, *Computer Vision – ECCV'98 Vol. II*, Freiburg, Germany, 1998. Springer, Lecture Notes in Computer Science 1407.
- [36] T. Vetter, M. J. Jones, and T. Poggio. A bootstrapping algorithm for learning linear models of object classes. In *IEEE Conference on Computer Vision and Pattern Recognition – CVPR '97*, Puerto Rico, USA, 1997. IEEE Computer Society Press.
- [37] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–742, 1997.
- [38] Keith Waters. A muscle model for animating three-dimensional facial expression. *Computer Graphics*, 22(4):17–24, 1987.



Rendering Techniques for Facial Animation

Jörg Haber
MPI Informatik

Rendering Faces



- skin rendering:
 - textures for skin and facial components
 - bump mapping for skin dimples and wrinkles
- hair modeling and rendering:
 - course #9 : “Photorealistic Hair Modeling, Animation, and Rendering “

Textures are...



- a cheap means of conveying realism
- a tool for LoD management
- available both on graphics hardware and in modeling / rendering software
- useful for many rendering “tricks”

How to create textures from input images?

Cylindrical Textures



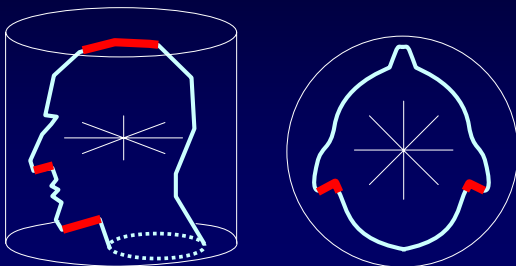
Common approach:

- created from input photographs:
 - L. Williams: “Performance-Driven Facial Animation”, SIGGRAPH '90, 235-242, Aug. 1990
 - F. Pighin et al.: “Synthesizing Realistic Facial Expressions from Photographs”, SIGGRAPH '98, 75-84, July 1998
- acquired during range scanning process (→ Cyberware scanners)

Cylindrical Textures



A head is similar to a cylinder ...is it?



Cylindrical Textures



Problems:

- limited texture resolution (Cyberware)
- need accurate geometry for registration (from photos)
- visual artifacts:
 - on top of the head
 - behind the ears
 - under the chin
- limited animation (eyes, teeth)

Textures from Photographs



Given:

- 3D mesh
- uncalibrated images (digitized photographs)

Assumptions:

- mesh represents real object (head) sufficiently precise
- images cover all areas of real object

Solution:

- register images using Tsai algorithm
- create texture patches

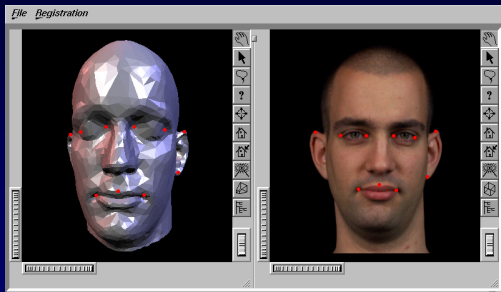
Tsai Algorithm



R. Y. Tsai: "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-Shelf TV Cameras and Lenses", IEEE J. of Robotics and Automation, RA-3(4), Aug. 1987

- compute **intrinsic camera parameters** (effective focal length, radial distortion, optical center) *once* from images of calibration pattern for different points of view using non-linear optimization
- compute **extrinsic camera parameters** (rotation & translation) *for each input image* using corresponding points (3D geometry \leftrightarrow 2D image) and linear optimization

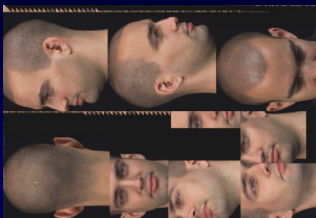
Corresponding Points



Texture Binding



Texture Combination



Important aspects:

- optimal packing of individual segments
- smooth transition between segments (*blending*)

Texture Atlases

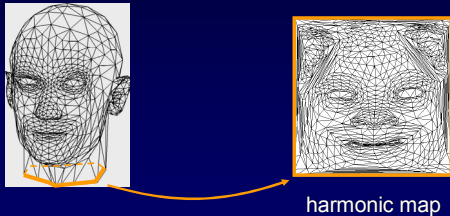


Problems:

- not suitable for mip-mapping
- waste of texture space:
 - optimal packing of patches is difficult
 - patches contain redundant information

A Different Parameterization

A head is topologically similar to a disk:



Harmonic Maps

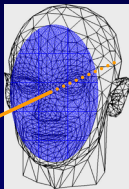
Characteristics:

- results in single texture patch suitable for mip-mapping
- 3D object must be topologically equivalent to a disk
- need to control distortion, e.g.:
 - P. V. Sander et al.: "Texture Mapping Progressive Meshes", SIGGRAPH '01, 409-416, Aug. 2001
- may introduce additional weights

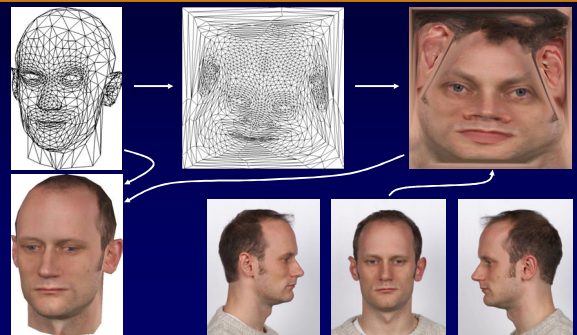
Weighted Parameterization

Facial region is most important:

- assign amount of texture space through weights
- triangles on the face become larger in the texture, backfacing triangles become smaller
- weights are computed automatically using dot product of triangle normal and viewing direction V of head model

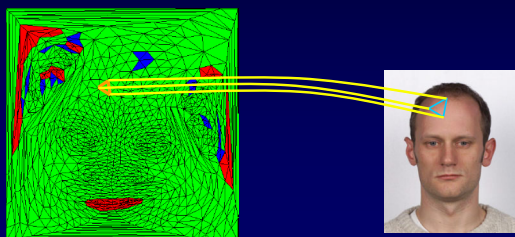


Process Overview



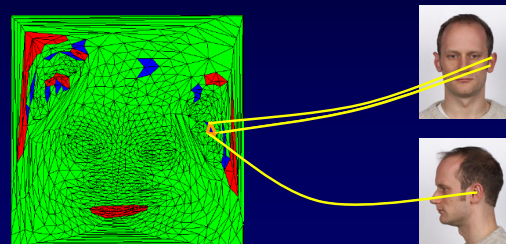
Texture Resampling: Resampling

common image for all vertices: resample triangle



Texture Resampling: Interpolation

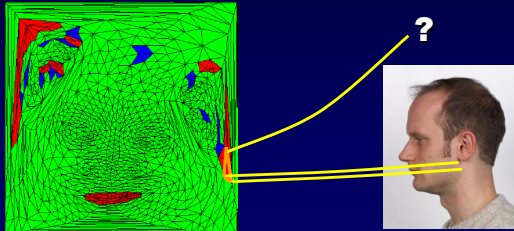
all vertices bound, no common image: interpolate



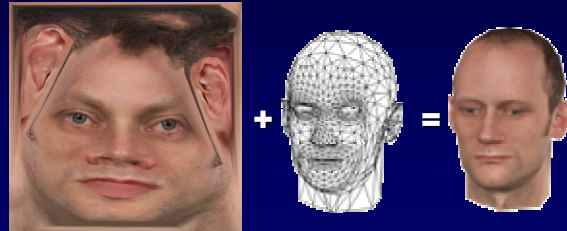
Texture Resampling: Filling holes



unbound vertices: apply iterative interpolation scheme



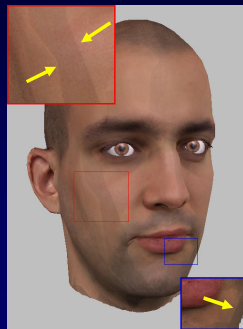
Result



Uncontrolled Illumination



- different skin color \Rightarrow discontinuities in the resampled texture



Removing Discontinuities



- P. J. Burt, E. H. Adelson: "A Multiresolution Spline with Application to Image Mosaics", ACM TOG, 2(4):217-236, Oct. 1983

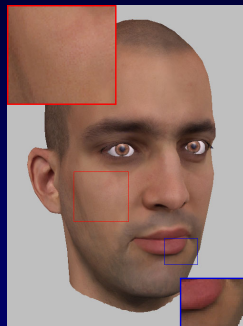


Removing Discontinuities



Multiresolution spline:

- removes discontinuities
- keeps fine detail



"De-lighting" Textures



- S. Marschner, B. Guenter, S. Raghupathy: "Modeling and Rendering for Realistic Facial Animation", Proc. EG Rendering Workshop 2000, 231-242, June 2000
- P. Debevec et al.: "Acquiring the Reflectance Field of a Human Head", SIGGRAPH 2000, 145-156, July 2000
- extract diffuse reflectivity (albedo map) from photographs taken under controlled illumination conditions (relative position of object, camera, and light sources)
- diffuse reflectivity is computed per texel from viewing direction, direction of incident light, surface normal and radiance (= color from photograph)

Facial Components



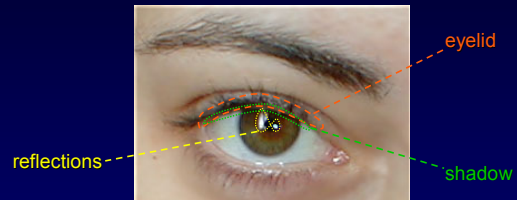
Observations:

- individual facial components (eyes, teeth) are crucial for realistic modeling
- difficult to acquire data for modeling these components

Solution:

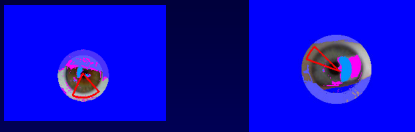
- Tarini et al.: "Texturing Faces", Proc. Graphics Interface 2002, 89-98, May 2002
- use generic models with individual textures
- create individual textures from plain photographs

Eyeball Textures: Problem



Many pixels must be discarded!

Eyeball Textures: Discarding Pixels



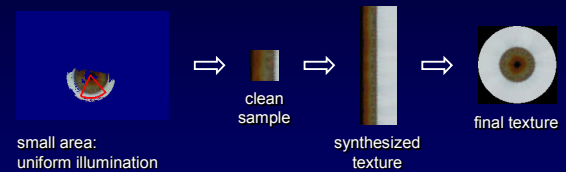
- remove pixels with a color similar to skin
- remove pixels with a color dissimilar to the pixels at the same radial distance from the center

Just a small clean part is needed as a seed...

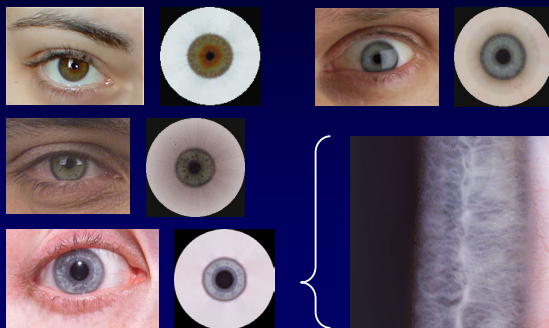
Eyeball Textures: Texture Synthesis



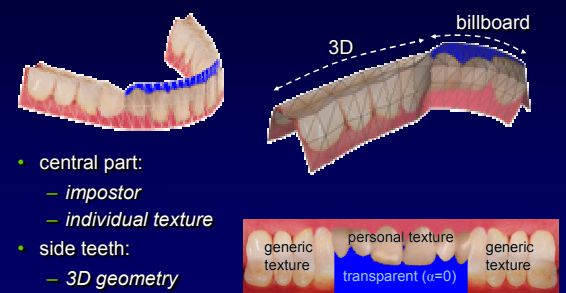
Texture synthesis in polar coordinates:



Eyeball Textures: Results



Generic Teeth Model



Teeth: Results

Skin Rendering

- speed vs. quality trade-off (i.e. real-time applications vs. offline computations)
- different techniques for modeling/rendering skin:
 - simple geometry + texture
 - simple geometry + bump mapping + texture
 - simple geometry + displacement mapping + texture
 - complex geometry + texture

Bump Mapping

- simulate complex geometry using coarse geometry and “faked” per-pixel surface normals

Rendering Wrinkles

- encode surface normals into RGB texture
- use modern graphics hardware for real-time rendering

Rendering Skin

- T. Ishii et al.: “A Generation Model for Human Skin Texture”, Proc. CGI '93, 39-150, 1993
- presents a method for generating skin structure bump maps and an appropriate illumination model for rendering skin
- surface normals are computed from recursively generated, hierarchical micro-geometry during preprocessing
- illumination model simulates multi-layered skin structure taking into account subsurface scattering

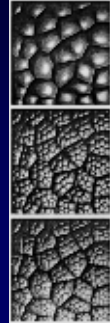
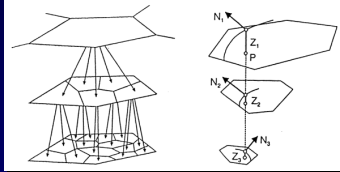
“Pattern Generation”

- skin cells are represented by Voronoi cells

Images: Ishii et al.: “A Generation Model for Human Skin Texture”

“Hierarchical Skin Structure”

- recursive Voronoi subdivision of skin cells (3 levels)

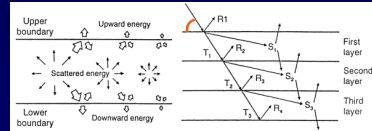
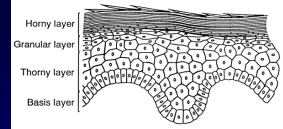


hierarchy level

Images: Ishii et al.: “A Generation Model for Human Skin Texture”

“Multiple Light Reflections”

- multi-layered skin structure results in complex light transport mechanisms
- model: parallel layers; reflection & transmission & scattering at each layer boundary

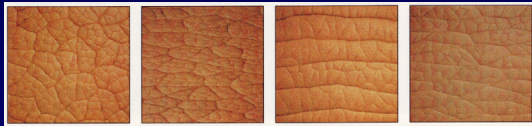


precompute lighting w.r.t. angle of incidence at skin surface

Images: Parke/Waters: “Computer Facial Animation” (1996)

Ishii et al.: Results

- generic model for rendering skin
- orientation of skin cells can be aligned to wrinkles
- anisotropic scaling of skin cells (\rightarrow wrist)
- skin structure can be rendered in real-time using graphics hardware bump mapping; illumination model not (yet) suitable for real-time rendering



Images: Ishii et al.: “A Generation Model for Human Skin Texture”

Texturing Faces

Marco Tarini^{1,2}

Hitoshi Yamauchi¹

Jörg Haber¹

Hans-Peter Seidel¹

¹ Max-Planck-Institut für Informatik, Saarbrücken, Germany

² Visual Computing Group, IEI, CNR Pisa, Italy

mtarini@di.unipi.it, {hitoshi, haberj, hpseidel}@mpi-sb.mpg.de

Abstract

We present a number of techniques to facilitate the generation of textures for facial modeling. In particular, we address the generation of facial skin textures from uncalibrated input photographs as well as the creation of individual textures for facial components such as eyes or teeth. Apart from an initial feature point selection for the skin texturing, all our methods work fully automatically without any user interaction. The resulting textures show a high quality and are suitable for both photo-realistic and real-time facial animation.

Key words: texture mapping, texture synthesis, mesh parameterization, facial modeling, real-time rendering

1 Introduction

Over the past decades, facial modeling and animation has achieved a degree of realism close to photo-realism. Although the trained viewer is still able to detect minor flaws in both animation and rendering of recent full-feature movies such as *Final Fantasy*, the overall quality and especially the modeling and texturing are quite impressive. However, several man-years went into the modeling of each individual character from that movie. Trying to model a real person becomes even more tricky: the artistic licence to create geometry and textures that “look good” is replaced by the demand to create models that “look real”.

A common approach towards creating models of real persons for facial animation uses range scanners such as, for instance, Cyberware scanners to acquire both the head geometry and texture. Unfortunately, the texture resolution of such range scanning devices is often low compared to the resolution of digital cameras. In addition, the textures are typically created using a cylindrical projection. Such cylindrical textures have the drawback to introduce visual artifacts, for instance on top of the head, behind the ears, or under the chin. Finally, there is no automatic mechanism provided to generate textures for individual facial components such as eyes and teeth.

In this paper, we present an approach to generate high-resolution textures for both facial skin and facial compo-

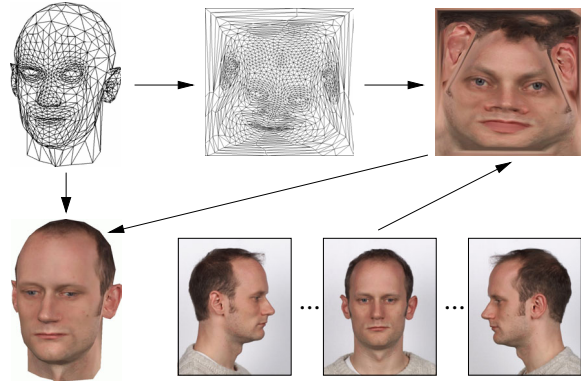


Figure 1: Overview of our skin texture generation process: the 3D face mesh is parameterized over a 2D domain and the texture is resampled from several input photographs.

nents from several uncalibrated photographs. The generation of these textures is automated to a large extent, and the resulting textures do not exhibit any patch structures, i.e. they can be used for mip-mapping. Our approach combines several standard techniques from texture mapping and texture synthesis. In addition, we introduce the following contributions:

- a view-dependent parameterization of the 2D texture domain to enhance the visual quality of textures with a fixed resolution;
- a texture resampling method that includes color interpolation for non-textured regions and visual boundary removal using multiresolution splines with a fully automatic mask generation;
- a radial texture synthesis approach with automatic center finding, which robustly produces individual eyeball textures from a single input photograph;
- a technique that uses a single natural teeth photograph to generate a teeth texture, which is applied to an appropriate 3D model to resemble the appearance of the subject’s mouth.

All of these techniques are fully automated to minimize the construction time for creating textures for facial modeling. However, we do not address the topic of facial modeling itself in this paper. We apply the textures generated by the techniques presented in this paper in our facial animation system [12], which has been designed to produce physically based facial animations that perform in real-time on common PC hardware. Thus the focus of our texture generation methods is primarily on the applicability of the textures for OpenGL rendering and a simple but efficient acquisition step, which does not require sophisticated camera setups and calibration steps.

2 Previous and Related Work

Research on either texturing or facial animation has provided a large number of techniques and insights over the years, see the surveys and textbooks in [13, 6] and [25] for an overview. Texturing in the context of facial animation is, however, an often neglected issue. Many sophisticated facial animation approaches, e.g. [32, 18, 19], simply use the textures generated by Cyberware scanners. In [35], Williams presents an approach to generate and register a cylindrical texture map from a peripheral photograph. This approach is meanwhile superseded by the ability of Cyberware scanners to acquire geometry and texture in one step. The method presented in [1] generates an individual head geometry and texture by linear combination of head geometries and textures from a large database that has been acquired using a Cyberware scanner in a costly preprocessing step. Marschner *et al.* describe a technique that uses several input photographs taken under controlled illumination with known camera and light source locations to generate an albedo texture map of the human face along with the parameters of a BRDF [23]. Several other approaches such as [26, 11, 16, 17] are image-based and use a small number of input photographs (or video streams) for the reconstruction of both geometry and texture. Although these approaches could potentially yield a higher texture quality compared to the Cyberware textures, they typically suffer from a less accurate geometry reconstruction, limited animation, and reduced texture quality by using cylindrical texture mapping.

Creating textures from multiple, unregistered photographs has been addressed in the literature by several authors [28, 3, 24]. First, they perform a camera calibration for each input photograph based on corresponding feature points. Next, a texture patch is created for each triangle of the input mesh. The approaches differ in the way these texture patches are created, blended, and combined into a common texture. However, the resulting textures always exhibit some patch structure,

which makes it impossible to generate mip-maps from these textures. Creating textures that can be mip-mapped requires to construct a parameterization of the mesh over a two-dimensional domain. To this end, generic techniques based on spring meshes have been presented in [10, 15, 7]. Special parameterizations that minimize distortion during texture mapping for different kinds of surfaces have been investigated by several authors, see for instance [27, 29, 22, 21].

Texture synthesis [9, 33] has become an active area of research in the last few years. Recent publications focus on texture synthesis on surfaces [34, 31, 36] or on texture transfer [8, 14]. All of the methods presented so far use a Euclidean coordinate system for the synthesis of textures. In contrast, we use a polar coordinate system to synthesize textures that exhibit some kind of radial similarity.

3 Texturing Facial Skin

To generate a skin texture for a head model, we first take about three to five photographs of the person's head from different, uncalibrated camera positions. All photographs are taken with a high-resolution digital camera (3040×2008 pixels). The camera positions should be chosen in such a way that the resulting images roughly cover the whole head. During the acquisition, no special illumination is necessary. However, the quality of the final texture will benefit from a uniform, diffuse illumination. In addition, we acquire the geometry of the head using a structured-light range scanner. As a result, we obtain a triangle mesh that consists of up to a few hundred thousand triangles. After the texture registration step, this triangle mesh is reduced to about 1.5k triangles for real-time rendering using a standard mesh simplification technique. Each photograph is registered with the high-resolution triangle mesh using the camera calibration technique developed by Tsai [30]. Since the intrinsic parameters of our camera/lens have been determined with sub-pixel accuracy in a preprocessing step, we need to identify about 12–15 corresponding feature points on the mesh and in the image to robustly compute the extrinsic camera parameters for each image. This manual selection of feature points is the only step during our texture generation process that requires user interaction.

Next, we automatically construct a parameterization of the 3D input mesh over the unit square $[0, 1]^2$. This step is described in detail in the following Section 3.1. Finally, every triangle of the 2D texture mesh is resampled from the input photographs. A multiresolution spline method is employed to remove visual boundaries that might arise from uncontrolled illumination conditions during the photo session. Details about this resam-

pling and blending step are given in Section 3.2. Figure 1 shows an overview of our texture generation process.

3.1 Mesh Parameterization

We want to parameterize the 3D input mesh over the 2D domain $[0, 1]^2$ in order to obtain a single texture map for the whole mesh. To obtain a mip-mappable texture, the texture should not contain individual patches (*texture atlas*) but rather consist of a single patch. Clearly, this goal cannot be achieved for arbitrary meshes. In our case, the face mesh is topologically equivalent to a part of a plane, since it has a boundary around the neck and does not contain any handles. Thus we can “flatten” the face mesh to a part of a plane that is bounded by its boundary curve around the neck. We represent the original face mesh by a spring mesh and use the L^2 stretch norm presented in [29] to minimize texture stretch. In our simulations, this L^2 norm performs better than the L^∞ norm that is recommended by the authors of [29].

By applying the texture stretch norm, texture stretch is minimized over the whole mesh. In the following step, we introduce some controlled texture stretch again. Since the size of textures that can be handled by graphics hardware is typically limited, we would like to use as much texture space as possible for the “important” regions of a head model while minimizing the texture space allocated to “unimportant” regions. Obviously, the face is more important for the viewer than the ears or even the back of the head. To accomplish some biased texture stretch, we have introduced an additional weighting function ω into the L^2 stretch norm presented in [29]:

$$L^2(M) := \sqrt{\frac{\sum_{T_i \in M} (L^2(T_i))^2 \omega(T_i) A'(T_i)}{\sum_{T_i \in M} \omega(T_i) A'(T_i)}}$$

with

$$\omega(T_i) := \frac{1}{\langle N(T_i), V \rangle + k},$$

where $M = \{T_i\}$ denotes the triangle mesh, $A'(T_i)$ is the surface area of triangle T_i in 3D, $N(T_i)$ is the triangle normal of T_i , V is the direction into which the head model looks, and $k > 1$ is a weighting parameter. The weighting function ω thus favors the triangles on the face by diminishing their error while penalizing the triangles on the back of the head by amplifying their error. As a consequence, triangles on the face become larger in the texture mesh while backfacing triangles become smaller. Useful values for k are from within $[1.01, 2]$.

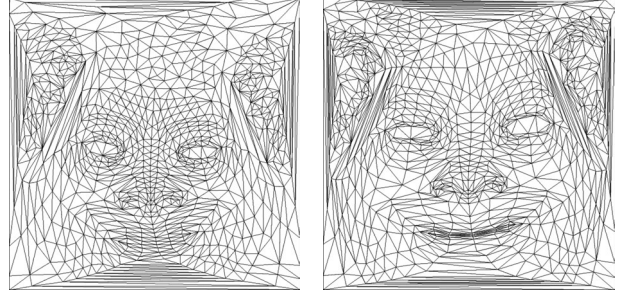


Figure 2: Comparison between a view-independent texture mesh parameterization according to [29] (left) and our view-dependent parameterization (right).

Figure 2 shows a view-independent texture mesh parameterization obtained with the original L^2 stretch norm as well as a view-dependent parameterization with our modified stretch norm for $k = 1.2$.

The difference between our *view-dependent texture mesh parameterization* and the *view-dependent texture mapping* proposed in [5, 26] is the following: the latter performs an adaptive blending of several photographs for each novel view, whereas we create a static texture that has its texture space adaptively allocated to regions of different visual importance.

3.2 Texture Resampling

After having created the 2D texture mesh from the 3D face mesh, we resample the texture mesh from the input photographs that have been registered with the face mesh. First, we perform a vertex-to-image binding for all vertices of the 3D face mesh. This step is carried out as suggested in [28]: Each mesh vertex v is assigned a set of *valid photographs*, which is defined as that subset of the input photographs such that v is visible in each photograph and v is a non-silhouette vertex. A vertex v is visible in a photograph, if the projection of v on the image plane is contained in the photograph **and** the normal vector of v is directed towards the viewpoint **and** there are no other intersections of the face mesh with the line that connects v and the viewpoint. A vertex v is called a silhouette vertex, if at least one of the triangles in the fan around v is oriented opposite to the viewpoint. For further details see [28]. In contrast to the approach in [28], we do not require that all vertices of the face mesh are actually bound to at least one photograph, i.e. the set of valid photographs for a vertex may be empty.

Let $\Delta = \{v_1, v_2, v_3\}$ denote a triangle of the face mesh and $\tilde{\Delta} = \{\tilde{v}_1, \tilde{v}_2, \tilde{v}_3\}$ be the corresponding triangle in the texture mesh. For each triangle Δ , exactly one of the following situations might occur (see also Figure 3):

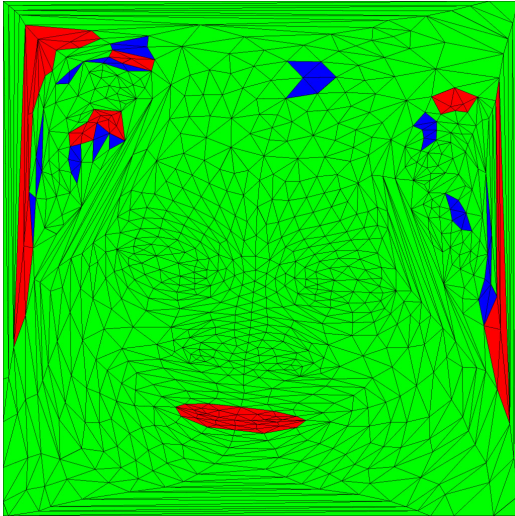


Figure 3: Color-coded triangles of the texture mesh: each green triangle has at least one common photograph to which all of its vertices are bound; the vertices of blue triangles don't have a common photograph, but they are all bound; red triangles have at least one unbound vertex.

1. There exists at least one common photograph in the sets of valid photographs of the three vertices v_1, v_2, v_3 of Δ (green triangles).
2. All of the vertices of Δ are bound to at least one photograph, but no common photograph can be found for all three vertices (blue triangles).
3. At least one vertex of Δ is not bound to any photograph (red triangles).

In the first case, we rasterize $\tilde{\Delta}$ in texture space. For each texel T , we determine its barycentric coordinates ρ, σ, τ w.r.t. $\tilde{\Delta}$ and compute the corresponding normal N by interpolating the vertex normals of Δ : $N = \rho N(v_1) + \sigma N(v_2) + \tau N(v_3)$. For each common photograph i in the sets of valid photographs of all vertices of Δ , we compute the dot product between N and the viewing direction V_i for the pixel P_i that corresponds to T . Finally, we color T with the color obtained by the weighted sum of pixel colors $\sum_i \langle N, V_i \rangle \cdot \text{Color}(P_i) / \sum_i \langle N, V_i \rangle$.

In the second case, we color each vertex \tilde{v}_j of $\tilde{\Delta}$ individually by summing up the weighted pixel colors of the corresponding pixels in all valid photographs i of \tilde{v}_j similarly as in the first case: $\text{Color}(\tilde{v}_j) := \sum_i \langle N(v_j), V_i \rangle \cdot \text{Color}(P_i) / \sum_i \langle N(v_j), V_i \rangle$. The texels of the rasterization of $\tilde{\Delta}$ are then colored by barycentric interpolation of the colors of the vertices $\tilde{v}_1, \tilde{v}_2, \tilde{v}_3$. Alternatively, we tried to use as much information as possible from the

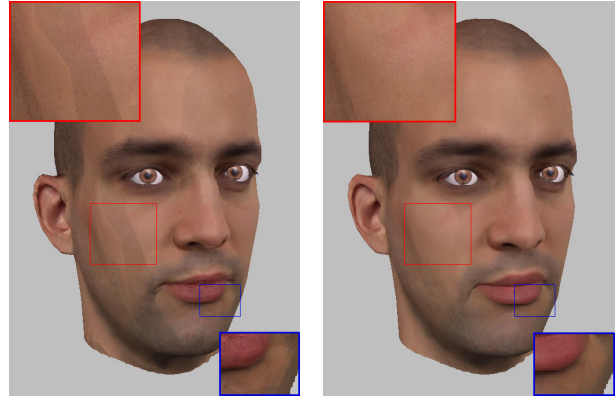


Figure 4: Boundaries in the skin texture (left) are removed using multiresolution spline techniques (right).

input photographs if, for instance, the vertices v_1, v_2 of Δ share a photograph and the vertices v_2, v_3 share another photograph. However, we found that this second case does not occur very often (cf. Figure 3) and that the difference between plain color interpolation and a more sophisticated approach is almost invisible.

Since we do not require that each vertex of the face mesh is bound to at least one photograph, there might exist some vertices that cannot be colored by any of the previously described schemes. We address this problem in a two-stage process: First, we iteratively assign an interpolated color to each unbound vertex. Next, we perform the color interpolation scheme from the second case for the remaining triangles of $\tilde{\Delta}$ that have not yet been colored. The first step iteratively loops over all unbound and uncolored vertices of the face mesh. For each unbound vertex v , we check if at least $p = 80\%$ of the vertices in the one-ring around v are colored (either by being bound to a photograph or by having an interpolated color). If this is true, we assign to v the average color of all the colored vertices around v , otherwise we continue with the next unbound vertex. We repeat this procedure until there are no further vertex updates. Next, we start the same procedure again, but this time we only require $p = 60\%$ of the vertices in the one-ring around v to be colored. As soon as there are no more updates, we repeat this step twice again with $p = 40\%$ and $p = 20\%$. Finally, we update each unbound vertex that has at least one colored neighbor. Upon termination of this last step, all vertices of the face mesh are either bound or colored and the remaining triangles of $\tilde{\Delta}$ can be colored.

If the input photographs have been taken under uncontrolled illumination, the skin color might differ noticeably between the images. In this case, boundaries might appear in the resampled texture. We then apply a multires-



Figure 5: Multiresolution spline masks: three different regions in the texture mesh resampled from different input photographs (top) and their corresponding masks shown in red (bottom).

olution spline method as proposed in [2, 17] to remove visual boundaries. Figure 4 shows a comparison between a textured head model with and without multiresolution spline method applied. To smoothly combine texture regions that have been resampled from different input photographs, we automatically compute a mask for each region by removing the outmost ring of triangles around the region, see Figure 5. Such a shrinking is necessary to ensure that there is still some valid color information on the outside of the mask boundary, because these adjacent pixels might contribute to the color of the boundary pixels during the construction of Gaussian and Laplacian pyramids. In addition to the masks for each input photograph, we create one more mask that is defined as the complement of the sum of all the other masks. This mask is used together with the resampled texture to provide some color information in those regions that are not covered by any input photograph (e.g. the inner part of the lips). As described above, these regions have been filled by color interpolation in the resampled texture. By blending all of the masked input photographs and the masked resampled texture with a multiresolution spline, we obtain a final texture with no visual boundaries and crispy detail.

4 Texturing Facial Components

Both human eyes and teeth are important for realistic facial animation while, at the same time, it is difficult to acquire data from a human being to precisely model these facial components. Thus we use generic models of these components as shown in Figure 8. The design of our generic models has been chosen such that they look convincingly realistic when inserted into a face mesh while still being rendered efficiently using OpenGL hardware.

On the other hand, both eyes and teeth (especially the more visible middle ones) are crucial features to visually differentiate one individual from another. Hence, it would be very desirable to use individual models for each person. Luckily, texturing can do the trick alone: indeed it is sufficient to apply a personal texture to a generic model to get the desired effect. Moreover, it is possible to automatically and quickly generate these textures each from a single input photograph of the subject’s eye and teeth, respectively. Details about this process will be given in the next two subsections.

4.1 Texturing Eyes

In order to realistically animate our head model, we must be able to perform rotations of the eyeball and dilation of the pupil. While the latter can be achieved by transforming the texture coordinates, we need an eye texture that covers the whole frontal hemisphere of the eyeball for the rotations.

Our goal to generate such an eyeball texture from a single input photograph is complicated by several factors such as the presence of occluding eyelids, shadows of eyelashes, highlights, etc. Still, all these factors are local and can be detected and removed. A new texture can then be synthesized from an input image consisting of the surviving pixels. In our current approach, we focus our effort on the iris, since it is obviously the most characteristic part of the eye.

Both the detection and the synthesis phase rely on the simplicity of the eye structure, i.e. an almost perfect point symmetry about the center, assuming our photograph represents an eye looking at the camera. To take advantage of this symmetry, we must first know precisely where the center of the eye is located. Since this would encumber the user, the center finding is done automatically by refining a rough estimation to sub-pixel precision using the following heuristic: we progressively enlarge an initially point-sized circle while checking the pixels on the circle at every iteration. If these pixels are too bright, they are assumed to be outside the iris and we thus move the center of the circle away from them. When most of the circle is composed by too bright pixels, we assume its center is the eye center and its radius is the iris radius. This approach runs robustly as long as the initial estimation is inside the pupil or the iris.

At this point, removal of occluded, shadowed, and highlighted pixels is done by:

- removing pixels with a color too similar to the skin;
- removing pixels with a color too dissimilar to the pixels at the same radial distance from the center.

For the second case, we compute the average color and standard deviation of the pixels at the same radial dis-

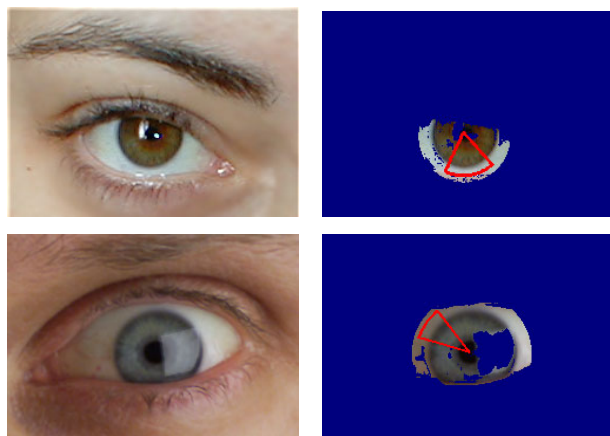


Figure 6: Two input photographs (left) and the resulting reference patches outlined by red sectors (right). Occluded, shadowed, highlighted, and skin-colored pixels (shown in blue) have been removed automatically.

tance and remove those pixels that are at least α times the standard deviation away from the average. The parameter α should be chosen within [2, 3]. We typically use a rather small value of $\alpha = 2.3$, as it empirically proved to remove the problematic (occluded, shadowed, highlighted, etc.) pixels in most cases. In addition, we remove pixels too close to the skin to better take into account small shadows cast by eyelids. Actually, the decision of which pixel to remove does not need excessively fine tuning: due to the regularity of the eye, we can be pretty conservative and remove many pixels, since the reconstruction phase requires only a small zone of pixels in order to synthesize more. Figure 6 shows the remaining set of pixels for two different input photographs.

For the reconstruction phase it is natural to resort to some texture synthesis from samples approach like e.g. [33]. In our case, we need to work in polar coordinates, because the eyeball texture behaves like a texture as defined in [33] only along the *angle* axis. This means that subregions of the eyeball texture are perceived to be similar if their *radius* coordinates are the same, cf. Figure 10. To take this into account, when choosing a candidate pixel p in the input image for filling a pixel p' in the output texture, we constrain the radius coordinate of p to be within a small threshold of the radius coordinate of p' .

A robust approach for texture synthesis is to use only a small patch of the original input image as the reference image and synthesize the texture from scratch. Although larger reference images theoretically result in more faithful textures, we obtained very good results with small reference patches covering a sector of about 30 degrees around the pupil. Small reference patches have the advan-

tage of being more uniform and thus bypassing problems related to uneven lighting in the original photograph. In our approach, we simply use the largest sector of valid pixels of at most 60 degrees as the reference patch. In the rare cases where the largest sector is too small, e.g. spanning less than 20 degrees, the entire set of valid pixels with a valid neighborhood is used as the reference image.

Since the detail frequencies of human irises are roughly the same, it is sufficient to use a texture synthesis scheme with a fixed neighborhood size rather than a multiresolution approach. In our case, the size of the neighborhood mask depends only on the resolution of the input image. For instance, for an image of an iris with a diameter of approximately 80 pixels, we use a 3×6 pixel mask (radius \times angle). For other iris diameters, the pixel mask is set proportionally. Depending on the value of the radius coordinate, a neighborhood with a fixed size in polar coordinates covers areas of different sizes in the input image. Our simulations showed, however, that no correction is needed, since the human iris usually exhibits higher frequency detail towards the center. Thus an iris resampled in polar coordinates shows quite uniform frequency distribution. Figure 9 shows several input photographs together with the resulting eye textures for various individuals.

To speed-up the reconstruction step, we use a one-dimensional texture synthesis approach along the angle axis alone, modeling the texture as a Markov chain rather than a Markov random field. Each symbol of the chain is an entire row of texels at a given angle coordinate. We output each new row accordingly to the previous rows. This approach gives similar results (even if it requires slightly larger reference textures) and is much faster, not even requiring any vector quantization for finding the best neighborhood row. If, however, the size of the reference patch is very small, we apply a two-dimensional texture synthesis approach as described earlier in this section.

4.2 Texturing Teeth

Geometry and color of teeth are difficult to capture and, at the same time, crucial to reflect personal appearance. We address this problem by distinguishing between

- the six middle teeth (incisors and canines) and
- the rest of the teeth (4–5 on each side).

The middle teeth are much more visible than the other teeth. This means that they account for most of the visual appearance of an individual person, but also that it is much easier to reconstruct them from a photograph. In addition, the middle teeth have an almost two-dimensional structure: they are shaped to have the function of a blade. Their small width allows us to model

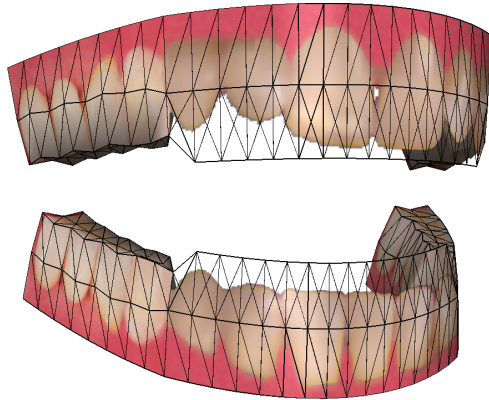


Figure 7: Teeth arch model using the texture shown in Figure 11. The wireframe shows the geometry of the teeth model, which consists of 384 triangles.

them using a billboard (impostor). Being a 2D data structure, the billboard can be easily extracted directly from a normal photograph of the subject exposing the teeth in a similar way as shown in Figure 11 (left). Using local transparency, it is straightforward to make the texture embed the teeth shape and size including gaps between teeth. This approach allows us to use the same (billboarded) 3D model for every face model and just change the texture from person to person.

The rest of the teeth, while being more voluminous and less accessible and visible, do not allow this useful shortcut. But, for the same reason, it is also less important to model them faithfully and individually for each single person. Thus it seems reasonable to use a standard 3D model and a standard texture (up to recoloring, see below) for this part of the teeth arch.

Following these considerations, we have built a generic 3D model for the teeth, which is non-uniformly scaled according to the individual skull and jaw geometry to fit into every head model. For each individual head model, we only need to vary the texture (including the billboard), which is created fully automatically. The generic teeth model is constructed such that the transition between the billboard (in the middle) and the 3D structure (left and right) is smooth, see Figure 7. The billboard, which is bent for better realism, could cause undesired artifacts when seen from above. To avoid this, only the upper part of the lower teeth and the lower part of the upper ones is actually modeled as a billboard. The remaining parts of the upper and lower middle teeth smoothly gain some width as they go up and down, respectively.

To automatically create a texture for the teeth, we start from a normal photograph of the subject showing his/her teeth. Several stages of the whole process of generating

a teeth texture are shown in Figure 11. We color-code dark parts that represent voids with a blue color, which is replaced by a transparent alpha value during rendering. Similarly, we identify and remove gums, lips, and skin, recoloring it with some standard gums color. To make this color-coding more robust, we identify the different regions using threshold values, which are obtained by finding the biggest jumps in the histograms of the color distances to the target color (red for gums and black for voids). In addition, we expand teeth into those parts of the gums that have been covered by the lips in the input photograph. We use some simple heuristics to include the missing part of the tooth roots, cf. Figure 11.

During rendering, our teeth model is shaded using a Phong shading model, which means that we have to desaturate our teeth texture. In order to do so for uncontrolled illumination, we equalize the color of the teeth, supposing they have approximately the same albedo. First, we define a target color by computing the average color of all teeth pixels and setting its brightness (but not the hue) to a predefined value. Next, we subdivide the texture in six vertical stripes and compute the average color of each stripe. We then add to the pixels in each column the difference between the target color and the stripe average, taking care of enforcing continuity in this correction by using a piecewise linear function. Similarly, we use the target color to correct the color of the “generic” part of the texture, which is applied to the side teeth. Finally, we composite the middle teeth texture into our generic texture using a curved boundary that follows the silhouettes of the canines.

5 Results

We have created facial textures for several individuals who have also been range-scanned to acquire their head geometry. Rendering of our head model is performed in real-time using OpenGL hardware (about 100 fps on a 1.7 GHz PC with a GeForce3 graphics board). A physics-based simulation is used to control the facial animation. Several images of our head models are distributed over this paper, see for instance Figures 1, 4, 8, and especially Figure 12. For each skin texture, the only interactive step is the initial identification of corresponding feature points. This step takes about five minutes per input photograph, which sums up to about 15–25 minutes spent interactively for three to five photographs. Computing an optimized parameterization of the face mesh (approx. 1600 triangles) takes about 80 minutes on a fast PC (1.7 GHz Pentium 4). Resampling a 2048×2048 texture from five input photographs takes about one minute, additional multiresolution spline blending (if necessary)

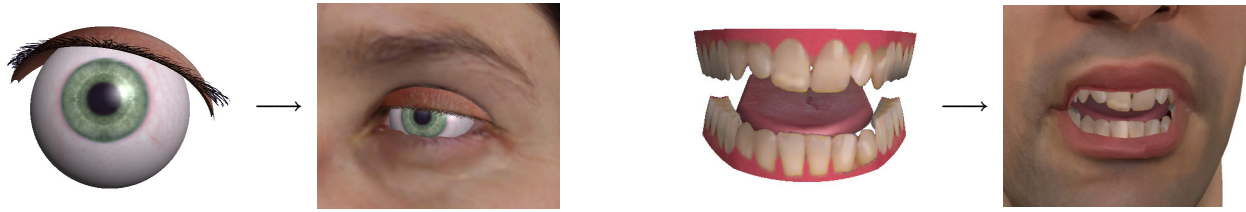


Figure 8: Generic models of eyes, teeth, and tongue are fitted into individual face meshes.

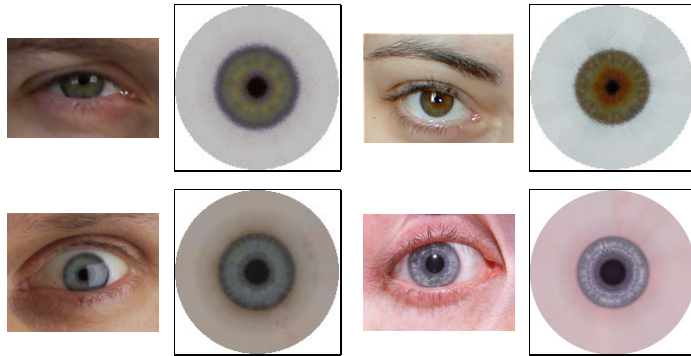


Figure 9: Input photographs and resulting eye textures: the input images have been taken under various illumination conditions with different resolutions. The size of the resulting textures changes from 128×128 (top left) to 1024×1024 (bottom right).

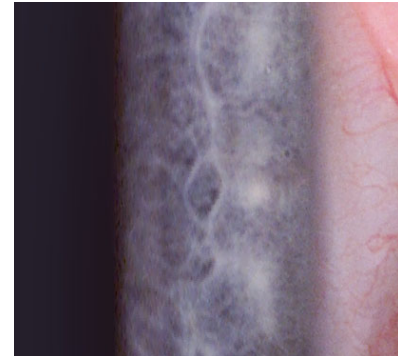


Figure 10: A detail of the texture from Figure 9 (bottom right) shown in polar coordinates. The abscissa represents the radius axis and the ordinate represents the angle axis.

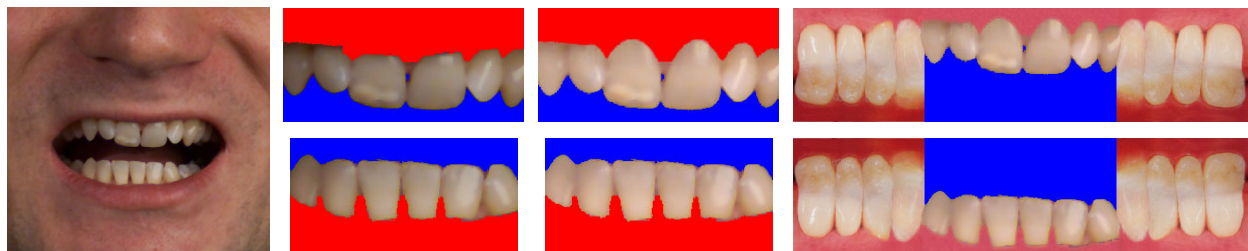


Figure 11: Teeth texture generation. Left to right: starting from an input photograph, we extract the upper and lower middle teeth, fill in missing parts and adjust the color, and composite the new image with a generic teeth texture. The blue pixels in the final texture (right) will be rendered transparently.

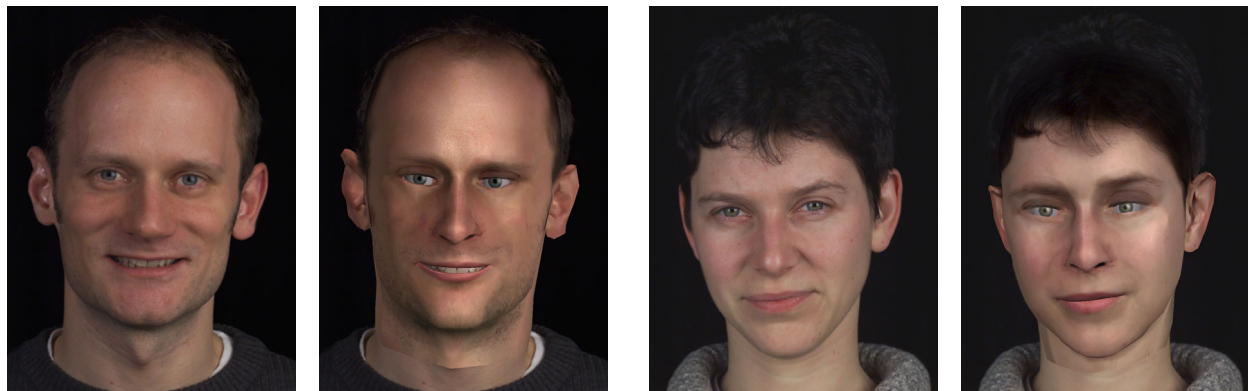


Figure 12: Side-by-side comparison of photographs (left) and head models (right) for plain OpenGL rendering.

takes about ten minutes. Currently, our algorithms are optimized with respect to robustness but not to speed.

Generating the teeth and eye textures takes only a few seconds even for large textures using the 1D Markov chain method for the texture synthesis. If a full Markov field is used, construction time may go up to several minutes, depending on the size of the texture being created.

6 Conclusion and Future Work

We have introduced a number of techniques that help to minimize the time and effort that goes into the creation of textures for facial modeling. With the exception of the initial feature point selection for the skin texturing, our methods are fully automated and do not require any user interaction.

For the generation of skin textures from uncalibrated input photographs, we propose a view-dependent parameterization of the texture domain and a texture resampling method including color interpolation for non-textured regions and multiresolution splining for the removal of visual boundaries. Using our methods, both eye and teeth textures can be created fully automatically from single input photographs, adding greatly to a realistic appearance of individual subjects during facial animation.

One of the main goals of ongoing research is to get rid of the interactive camera calibration step for skin texturing. Given that the resulting texture should contain fine detail, this is a tough problem, indeed. Automatic approaches such as [20] fail simply due to the fact that the silhouette of a human head looks more or less identical when viewed from within a cone of viewing directions from the front or the back. Furthermore, it would be desirable to account for lighting artifacts in the input photographs. Although a uniform, diffuse illumination during the photo session helps a lot, there are still contributions from diffuse and specular lighting in the photographs. Approaches to overcome these problems have been suggested [4, 23], but they require sophisticated camera setups and calibration steps. Finally, it would be very helpful to speed-up the computation time of the current bottleneck, namely the mesh parameterization, using a hierarchical coarse-to-fine approach.

Acknowledgments

The authors would like to thank their models Letizia, Claudia, and Kolja for all the smiles during the photo sessions. Many thanks also to our colleagues, who gave helpful comments during the development of our techniques, and to the anonymous reviewers for their suggestions.

References

- [1] V. Blanz and T. Vetter. A Morphable Model for the Synthesis of 3D Faces. In *Computer Graphics (SIGGRAPH '99 Conf. Proc.)*, pages 187–194, August 1999.
- [2] P. J. Burt and E. H. Adelson. A Multiresolution Spline with Application to Image Mosaics. *ACM Transactions on Graphics*, 2(4):217–236, October 1983.
- [3] P. Cignoni, C. Montani, C. Rocchini, R. Scopigno, and M. Tarini. Preserving Attribute Values on Simplified Meshes by Resampling Detail Textures. *The Visual Computer*, 15(10):519–539, 1999.
- [4] P. E. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the Reflectance Field of a Human Face. In *Computer Graphics (SIGGRAPH '00 Conf. Proc.)*, pages 145–156, July 2000.
- [5] P. E. Debevec, C. J. Taylor, and J. Malik. Modeling and Rendering Architecture from Photographs: A Hybrid Geometry- and Image-based Approach. In *Computer Graphics (SIGGRAPH '96 Conf. Proc.)*, pages 11–20, August 1996.
- [6] D. S. Ebert, F. K. Musgrave, D. Peachey, K. Perlin, and S. Worley. *Texturing & Modeling: A Procedural Approach*. Academic Press, London, 2 edition, 1998.
- [7] M. Eck, T. DeRose, T. Duchamp, H. Hoppe, M. Lounsbery, and W. Stuetzle. Multiresolution Analysis of Arbitrary Meshes. In *Computer Graphics (SIGGRAPH '95 Conf. Proc.)*, pages 173–182, August 1995.
- [8] A. A. Efros and W. T. Freeman. Image Quilting for Texture Synthesis and Transfer. In *Computer Graphics (SIGGRAPH '01 Conf. Proc.)*, pages 341–346, August 2001.
- [9] A. A. Efros and T. K. Leung. Texture Synthesis by Non-parametric Sampling. In *IEEE Int'l Conf. Computer Vision*, volume 2, pages 1033–1038, September 1999.
- [10] M. S. Floater. Parametrization and Smooth Approximation of Surface Triangulations. *Computer Aided Geometric Design*, 14(3):231–250, 1997.
- [11] B. Guenter, C. Grimm, D. Wood, H. Malvar, and F. Pighin. Making Faces. In *Computer Graphics (SIGGRAPH '98 Conf. Proc.)*, pages 55–66, July 1998.
- [12] J. Haber, K. Kähler, I. Albrecht, H. Yamauchi, and H.-P. Seidel. Face to Face: From Real Humans to Realistic Facial Animation. In *Proc. Israel-Korea Binational Conf. on Geometrical Modeling and Computer Graphics*, pages 73–82, October 2001.
- [13] P. S. Heckbert. Survey of Texture Mapping. *IEEE Computer Graphics and Applications*, 6(11):56–67, November 1986.
- [14] A. Hertzmann, Ch. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image Analogies. In *Computer Graphics (SIGGRAPH '01 Conf. Proc.)*, pages 327–340, August 2001.
- [15] K. Hormann and G. Greiner. MIPS: An Efficient Global Parametrization Method. In *Curve and Surface Design: Saint-Malo 1999*, pages 153–162. Vanderbilt University Press, 2000.

- [16] W.-S. Lee, J. Gu, and N. Magnenat-Thalmann. Generating Animatable 3D Virtual Humans from Photographs. In *Computer Graphics Forum (Proc. EG 2000)*, volume 19, pages C1–C10, August 2000.
- [17] W.-S. Lee and N. Magnenat-Thalmann. Fast Head Modeling for Animation. *Image and Vision Computing*, 18(4):355–364, March 2000.
- [18] Y. Lee, D. Terzopoulos, and K. Waters. Constructing Physics-based Facial Models of Individuals. In *Proc. Graphics Interface '93*, pages 1–8, May 1993.
- [19] Y. Lee, D. Terzopoulos, and K. Waters. Realistic Modeling for Facial Animations. In *Computer Graphics (SIGGRAPH '95 Conf. Proc.)*, pages 55–62, August 1995.
- [20] H. P. A. Lensch, W. Heidrich, and H.-P. Seidel. Automated Texture Registration and Stitching for Real World Models. In *Proc. Pacific Graphics 2000*, pages 317–326, October 2000.
- [21] B. Lévy. Constrained Texture Mapping for Polygonal Meshes. In *Computer Graphics (SIGGRAPH '01 Conf. Proc.)*, pages 417–424, August 2001.
- [22] J. Maillot, H. Yahia, and A. Verroust. Interactive Texture Mapping. In *Computer Graphics (SIGGRAPH '93 Conf. Proc.)*, pages 27–34, August 1993.
- [23] S. R. Marschner, B. Guenter, and S. Raghupathy. Modeling and Rendering for Realistic Facial Animation. In *Rendering Techniques 2000 (Proc. 11th EG Workshop on Rendering)*, pages 231–242, 2000.
- [24] P. J. Neugebauer and K. Klein. Texturing 3D Models of Real World Objects from Multiple Unregistered Photographic Views. In *Computer Graphics Forum (Proc. EG '99)*, volume 18, pages C245–C256, September 1999.
- [25] F. I. Parke and K. Waters, editors. *Computer Facial Animation*. A K Peters, Wellesley, MA, 1996.
- [26] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin. Synthesizing Realistic Facial Expressions from Photographs. In *Computer Graphics (SIGGRAPH '98 Conf. Proc.)*, pages 75–84, July 1998.
- [27] D. Piponi and G. D. Borshukov. Seamless Texture Mapping of Subdivision Surfaces by Model Pelting and Texture Blending. In *Computer Graphics (SIGGRAPH '00 Conf. Proc.)*, pages 471–478, July 2000.
- [28] C. Rocchini, P. Cignoni, C. Montani, and R. Scopigno. Multiple Textures Stitching and Blending on 3D Objects. In *Rendering Techniques '99 (Proc. 10th EG Workshop on Rendering)*, pages 119–130, 1999.
- [29] P. V. Sander, J. Snyder, S. J. Gortler, and H. Hoppe. Texture Mapping Progressive Meshes. In *Computer Graphics (SIGGRAPH '01 Conf. Proc.)*, pages 409–416, August 2001.
- [30] R. Y. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 364–374, June 1986.
- [31] G. Turk. Texture Synthesis on Surfaces. In *Computer Graphics (SIGGRAPH '01 Conf. Proc.)*, pages 347–354, August 2001.
- [32] K. Waters and D. Terzopoulos. Modeling and Animating Faces Using Scanned Data. *J. Visualization and Computer Animation*, 2(4):123–128, October–December 1991.
- [33] L.-Y. Wei and M. Levoy. Fast Texture Synthesis Using Tree-Structured Vector Quantization. In *Computer Graphics (SIGGRAPH '00 Conf. Proc.)*, pages 479–488, July 2000.
- [34] L.-Y. Wei and M. Levoy. Texture Synthesis over Arbitrary Manifold Surfaces. In *Computer Graphics (SIGGRAPH '01 Conf. Proc.)*, pages 355–360, August 2001.
- [35] L. Williams. Performance-Driven Facial Animation. In *Computer Graphics (SIGGRAPH '90 Conf. Proc.)*, volume 24, pages 235–242, August 1990.
- [36] L. Ying, A. Hertzmann, H. Biermann, and D. Zorin. Texture and Shape Synthesis on Surfaces. In *Rendering Techniques 2001 (Proc. 12th EG Workshop on Rendering)*, pages 301–312, 2001.



Forensic Applications

Jörg Haber
MPI Informatik

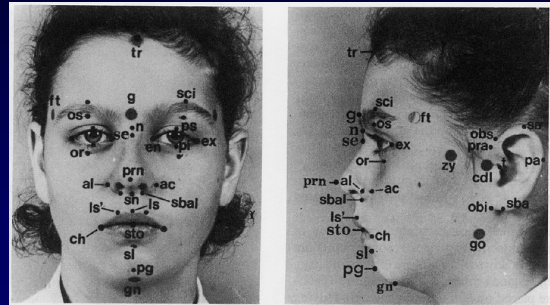
Typical Applications

- police work:
 - growth simulation / aging of missing children
 - facial reconstruction from skeletal remains
- important for tracing and identification
- based on anthropometric data

Anthropometric Data

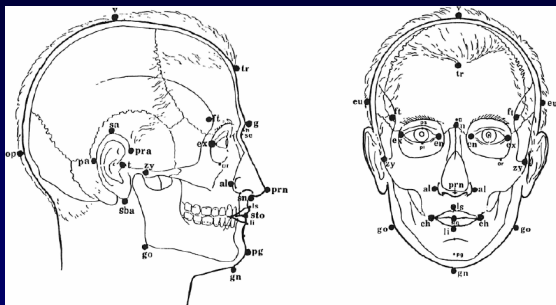
- data collected over decades
- facial measurements: **landmarks**
- populations vary by:
 - ethnicity (caucasian, asian, ...)
 - age (1-25 yrs. for growth measurements)
 - gender
- measurements consist of:
 - distances: axis-aligned, euclidean, arc-length
 - angles
 - proportions

Landmarks



Images: Farkas: "Anthropometry of the Head and Face", 1994

Landmarks

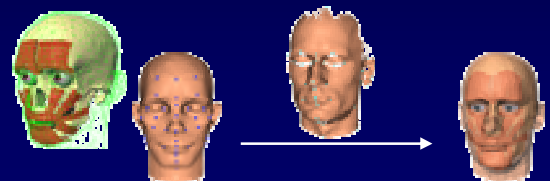


Images: Farkas: "Anthropometry of the Head and Face", 1994

Deformable Head Model

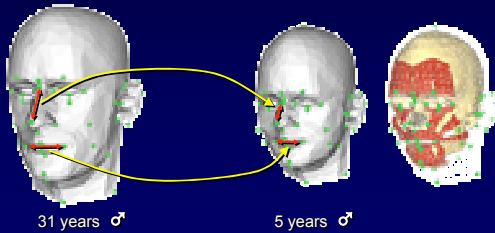
Idea: use landmarks for head deformation

- structured, animatable reference head model
- tagged with landmarks
- thin-plate spline interpolation for deformation



Growth Simulation

- analyze landmark positions of given age
- compute new landmarks for target age
- deform reference head model to fit new landmarks



Growth Simulation

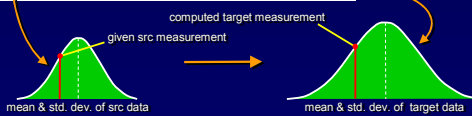
Derive new measurements for age change

- given input head model, age, sex, ethnicity
- examine landmarks on input model:
 - find deviation from statistical data
- look up statistics for target age using same deviation
- compute new landmarks:
 - best fit for target age measurements

Growth Simulation

Age (years)	Male			Female		
	N	Mean	SD	N	Mean	SD
0-5 months	8	70.0	4.5	5	68.0	4.5
6-12 months	20	70.5	4.8	8	72.7	4.4
1	18	82.8	4.8	20	77.2	4.8
2	31	87.5	3.5	31	83.8	3.9
3	30	88.5	3.5	26	85.0	3.4
4	30	90.4	4.3	30	87.6	3.5
5	30	90.7	3.0	20	86.5	4.1
6	50	98.5	5.0	50	90.7	4.4
7	50	99.5	5.0	50	90.9	3.7
8	51	101.8	4.9	51	98.1	5.4
9	51	102.7	5.3	50	101.3	5.3
10	50	105.2	4.5	49	103.9	5.0
11	50	107.1	6.0	51	104.7	5.0
12	52	109.1	5.4	53	108.2	4.6
13	50	111.6	5.7	49	109.1	5.0
14	49	114	6.5	51	110.7	5.3
15	50	116.1	5.7	51	111.0	5.1
16	50	120.9	4.6	51	113.0	4.6
17	49	120.9	7.1	51	112.0	4.7
18	49	121.8	6.6	45	113.6	4.6
19-25	109	124.7	5.7	200	111.4	4.8

Table: Farkas: "Anthropometry of the Head and Face", 1994

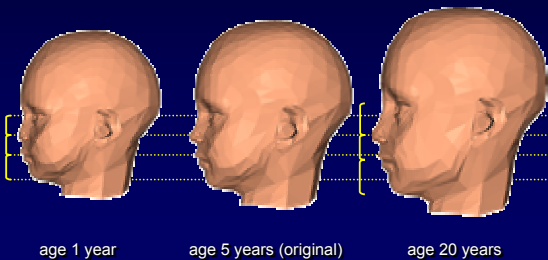


Growth Simulation Movie



Growth Simulation Examples

- growth simulation \neq scaling



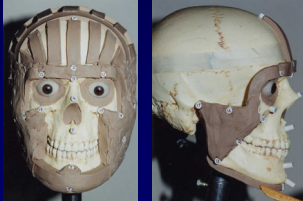
Limitations

It's only statistics!

- landmarks are sparsely distributed
 - lots of source characteristics are maintained
- positioning in normal distribution valid?
 - does a child with a big nose have a big nose as an adult?
- accuracy depends on physical measurements taken decades ago
 - could be improved using 3D scanning
 - build up a big database of measurements?

Reconstruction of Faces

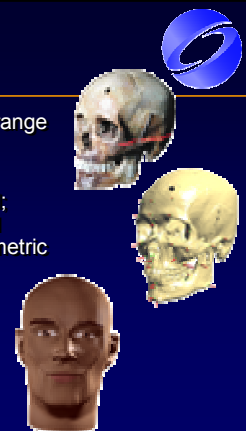
- traditional clay sculpting approach:
 - place tissue depth markers on the skull; length of pegs corresponds to anthropometric data
 - face is modeled using clay (→ artistic licence)



Images: Taylor: "Forensic Art and Illustration", 2001

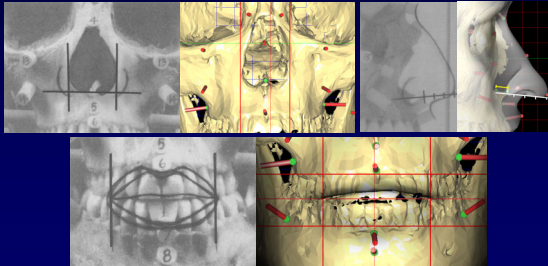
CG Approach

1. acquisition of skull data (3D range scan, computer tomography)
2. interactive placement of landmarks on the virtual skull; tissue depth values assigned automatically from anthropometric data tables
3. automatic fitting of the reference head model to the prescribed skin surface positions ⇒ instantly animatable head model



Additional Reconstruction Hints

- forensic art: many "rules of the thumb" to locate certain features of the face based on the skull shape



Results



Discussion

Again: it's only statistics!

- method mirrors the manual tissue depth method ⇒ same prediction power
- results show plausible reproduction of facial shape and proportions + some well-matched details!
- advantages: very fast (a few hours instead of weeks), does not damage original skull
- need additional editing tools for hair, beards, wrinkles
- most promising: gather lots of data through simulation and evaluation; update tissue thickness tables with these data

Reanimating the Dead: Reconstruction of Expressive Faces from Skull Data

Kolja Kähler* Jörg Haber† Hans-Peter Seidel‡

MPI Informatik, Saarbrücken, Germany

Abstract

Facial reconstruction for postmortem identification of humans from their skeletal remains is a challenging and fascinating part of forensic art. The former look of a face can be approximated by predicting and modeling the layers of tissue on the skull. This work is as of today carried out solely by physical sculpting with clay, where experienced artists invest up to hundreds of hours to craft a reconstructed face model. Remarkably, one of the most popular tissue reconstruction methods bears many resemblances with surface fitting techniques used in computer graphics, thus suggesting the possibility of a transfer of the manual approach to the computer. In this paper, we present a facial reconstruction approach that fits an anatomy-based virtual head model, incorporating skin and muscles, to a scanned skull using statistical data on skull / tissue relationships. The approach has many advantages over the traditional process: a reconstruction can be completed in about an hour from acquired skull data; also, variations such as a slender or a more obese build of the modeled individual are easily created. Last not least, by matching not only skin geometry but also virtual muscle layers, an animatable head model is generated that can be used to form facial expressions beyond the neutral face typically used in physical reconstructions.

CR Categories: I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Physically based modeling I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation G.3 [Probability and Statistics]—Multivariate statistics G.1.2 [Numerical Analysis]: Approximation—Approximation of surfaces and contours

Keywords: facial modeling, forensic art, face reconstruction

1 Introduction

1.1 Background

For well over a hundred years, forensic art and science has been assisting law enforcement. One of the major areas of concern in this area is facial reconstruction for postmortem identification of humans from their physical remains. Manual reconstruction and identification techniques build on the tight shape relationships between the human skull and skin: for instance, the presumed identity

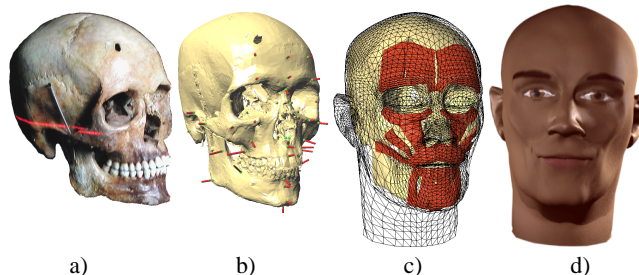


Figure 1: Reconstruction of a face from the skull: a) scanning the skull; b) skull mesh tagged with landmarks; c) skin mesh with muscles fitted to the skull; d) textured skin mesh, smiling expression.

of a murder victim can be confirmed by superimposing a facial photograph with a properly aligned and sized image of the skull. If no photograph is available, the look of the face can be reconstructed to a certain degree by modeling the missing tissue layers directly onto the skull or a plaster cast made from it.

The first documented case using three-dimensional facial reconstruction from the skull dates back to 1935 [Taylor 2001]. A key experiment was later performed by KROGMAN [1946]: given the body of a deceased person, he took a picture of the cadaver head before extracting the skull. The skull was provided to a sculptor along with information about sex, origin, and age of the late owner, plus data on the average tissue thicknesses at several positions in the face. From this material, a reconstruction sculpture was created that could be compared to the original head. Since that time, three-dimensional facial reconstruction from the skull has been much refined, but the method has essentially remained the same. Researchers have examined the skull / skin relationships for different ethnic groups [Lebedinskaya et al. 1993] and analyzed the correspondences of skull morphology and facial features [Fedosytukin and Nainys 1993]. Others found correlations between muscle activity and skull shape [Moore and Lavelle 1974; Weijs and Hillen 1986]. In her comprehensive textbook, TAYLOR [2001] describes the craft in great detail.

Much of the fascination of the topic is due to the combined efforts of science and art, resulting in often astonishingly lifelike reconstructions, given the little available input (see Fig. 2). Many parameters of the outward appearance of an individual cannot be readily derived from the skull, though. The process is thus highly dependent on rules of thumb, the experience of the artist, and some guesswork. It is, for instance, next to impossible to reconstruct the shape of the ears based on scientific reasoning, although empirically there seems to be a relation of ear height to the length of the nose.

1.2 The Manual Reconstruction Process

The traditional work process for facial reconstruction begins with preparation of the skull. Since the skull is often evidence in a criminal case, great care needs to be taken in handling it: some parts are extremely thin and fragile, especially in the nose and the orbits. For identification, the teeth often provide a lot of useful informa-

*e-mail: kkaehler@acm.org

†e-mail: haberj@acm.org

‡e-mail: hpseidel@mpi-sb.mpg.de



Figure 2: Comparison of sculpted reconstructions with photographs. Left: male subject; right: female subject. (Images: Copyright ©[Helmer et al. 1993], reprinted by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.)

tion, so a dental analysis is usually performed at this stage. For the reconstruction of the lower face, the mandible needs to be properly aligned and secured to the skull. In cooperation with an anthropologist, and possibly given more information from the remains of the victim, an estimation of age, ancestry, sex, and stature can now be obtained.

The actual face reconstruction proceeds with one of two available approaches: the *anatomical method* and the *tissue depth method*. The anatomical method attempts reconstruction by sculpting muscles, glands, and cartilage, fleshing out the skull layer by layer. This technique is more often used in the reconstruction of fossil faces, where no statistical population data exists [Zollikofer et al. 1998]. As TAYLOR states, this technique is very time consuming, occupying “many hundreds of hours”. It also requires a great deal of detailed anatomical knowledge. Therefore, the alternative tissue depth method has become the more popular reconstruction technique in law enforcement. Here, standard sets of statistical tissue thickness measurements at specific points on the face are used. Each measurement describes the total distance from skin surface to the skull, including fat and muscle layers. The method is thus more rapid than the anatomical method and does not require as much anatomical knowledge. Such measurements have been collected for males and females of several racial groups, using needles, X-rays, or ultrasound techniques. The tissue depth data most often used by police artists today was collected primarily by RHINE *et al.* [Rhine and Campbell 1980; Rhine and Moore 1984]. The data is sorted into “slender”, “normal”, and “obese” groups, as well as by sex and race.

Given the set of measurements, tissue depth markers are now placed on the skull or a cast made from it, reflecting the tissue thickness at the sample points. These markers are oriented orthogonally to the skull surface, corresponding to the direction of the tissue thickness measurements. Using the markers and other features on the skull for guidance, the face is modeled on top of the skull using clay. A snapshot of the beginning stages of a reconstruction using the tissue depth method is shown in Fig. 3.

1.3 Our approach

Looking at the facial reconstruction process as described above from a computer graphics perspective, it essentially boils down to a surface interpolation problem. We thus implement the manual “dowel placement” method as an interactive procedure, obtaining position and distance constraints that define the relation between skin and skull at selected sample positions. The sculpting of the skin surface is mapped to a volume deformation applied to a head model template, satisfying these constraints. The deformation approach has the additional advantage of being applicable to additional structures attached to the template: in our system, we map a muscle structure to the fitted head model (see Fig. 1), enabling

animation on the reconstructed head in a physics-based facial animation framework.

The remainder of this paper is organized as follows: after reviewing related work in Section 2, we discuss acquisition of skull data and interactive landmark placement for setting up surface constraints in Section 3. Section 4 describes the structure of our generic head model and how it is fitted to the skull. Animation and texture generation for the resulting head model are touched upon in Section 5. We present examples in Section 6 and draw conclusions from our results in Section 7.

2 Previous and Related Work

2.1 Computer-Aided Face Reconstruction

Perhaps due to the lack of rigid taxonomies and hard rules, the use of computers and computer graphics in this forensic application is still very limited. The procedures described above cannot be cast easily into a computer program that produces good results in an automated manner—the experience and judgment of the practitioner remain a vital part of the system.

In law enforcement practice, computer-aided techniques restrict to relatively simple image and video manipulation: face photographs are used for skull superimposition [Grüner 1993; Miyasaka et al. 1995], while image warping and retouching enable a basic simulation of aging [Taylor 2001, p. 253]. This situation is unfortunate, since the traditional three-dimensional face reconstruction process is extremely time-consuming and expensive. It is hardly feasible to produce a variety of different plausible reconstructions from one skull, simply due to the effort that has to be put into the creation of each model. Also, repeated physical handling of the original skull increases the risk of damage.

One prototypical computer-based face reconstruction system, allowing fitting of a generic hierarchical B-spline head model to a skull mesh, is described by ARCHER in her Master’s thesis [1997]. The user places dowels on a skull model with prescribed tissue thickness values, resulting in targets for a B-spline surface fitting process. The interpolation process is tricky and requires careful preparation of the template head model.

In the approach presented by MICHAEL and CHEN [1996], a source head model H_s that includes a skull S_s is deformed using a volume distortion function V such that the deformed source skull approximately matches the target skull S_t : $V(S_s) \approx S_t$. It is assumed that the deformed source head model $V(H_s)$ bears a good resemblance to the (unknown) target head model. The volume distortion function V is set up as a field warp using forty pairs of disc fields, which are manually placed around the skull. No details are given about the placement of these control fields.

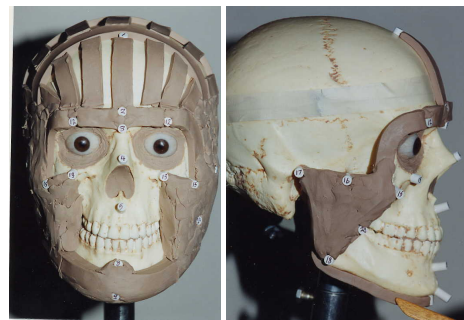


Figure 3: Modeling the face with clay on top of the skull using the tissue depth method. (Images [Taylor 2001], reprinted by permission.)

A deformation technique similar to the one used in our approach is employed by VANEZIS *et al.* [2000]. A facial template chosen from a database of scanned faces is deformed to match the position of target face landmarks, which have been derived from adding statistical tissue thickness values to the corresponding skull landmarks. The resulting reconstructed heads are not always complete (for instance, the top of the head is usually missing). The authors suggest to export an image of the reconstructed head and to apply a final image-processing step to add eyes, facial and head hair.

The above methods require a lot of manual assistance in setting up the interpolation function [Archer 1997; Michael and Chen 1996], or rely on a database of head templates [Vanezis *et al.* 2000]. In contrast, we develop reconstructions from one head template with relatively few markers, and use additional mechanisms to improve reconstruction results (see Section 4.3). Our approach always generates complete head models. Instead of using higher-order surfaces or point samples, the surface of our deformable head template is an arbitrary triangle mesh, simplifying later artistic modifications of the result using standard modeling tools. To the best of our knowledge, integration of expressive facial animation is not discussed by any other computer-aided facial reconstruction approach.

Other than explicit treatment of facial reconstruction, the creation of virtual head models based on human anatomy is well researched and documented in the computer graphics literature. Major developments in this area are discussed in the following section.

2.2 Human Head Modeling

A variety of techniques exists to create a face model from images or scan data. In the method presented by LEE *et al.* [1995], animatable head models are constructed semi-automatically from range scans. A generic face mesh with embedded muscle vectors is adapted to range scans of human heads. This process relies on the planar parameterization of the range scans as delivered, for instance, by the Cyberware digitizers. PIGHIN *et al.* [1998] interactively mark corresponding facial features in several photographs of an individual to deform a generic head model using radial basis functions. Animation is possible by capturing facial expressions in the process and blending between them. CARR *et al.* [2001] use radial basis functions to generate consistent meshes from incomplete scan data. Employing a large database of several hundred scanned faces, BLANZ *et al.* [1999] are able to create a geometric head model from only a single photograph. This model has the same resolution as the range scans in the database and cannot be readily animated. In the context of medical imaging, SZELISKI *et al.* [1996] minimize the distance between two surfaces obtained from volume scans of human heads by applying local free-form deformations [Sederberg and Parry 1986] and global polynomial deformations. The method does not require specification of corresponding features on the geometries.

Several facial animation systems use an approximation of the layered anatomical structure. WATERS [1987] represents skin and muscles as separate entities, where muscle vectors and radial functions derived from linear and sphincter muscles specify deformations on a skin mesh. In contrast to this purely geometric technique, physics-based approaches attempt to model the influence of muscle contraction onto the skin surface by approximating the biomechanical properties of skin. Typically, mass-spring or finite element networks are used for numerical simulation [Platt and Badler 1981; Lee *et al.* 1995; Koch *et al.* 1998]. From an initial triangle mesh, TERZOPOULOS and WATERS [1990] automatically construct a layered model of the human face. The model structure consists of three layers representing the muscle layer, dermis, and epidermis. The skull is approximated as an offset surface from the skin. Free-form deformations are employed by CHADWICK *et al.* [1989] to shape the skin in a multi-layer model, which contains bones, muscles, fat

tissue, and skin. SCHEEPERS *et al.* [1997] as well as WILHELMS and VAN GELDER [1997] introduce anatomy-based muscle models for animating humans and animals, focusing on the skeletal musculature. Skin tissue is represented only by an implicit surface with zero thickness [Wilhelms and Van Gelder 1997].

We build our system on the deformable, anatomy-based head model described by KÄHLER *et al.* [2002]. There, a generic face mesh with underlying muscle and bone layers is deformed to match scanned skin geometry. This process is adopted here to match the muscle and skin layers to given skull data instead.

3 Preparation of the Skull

Our approach uses three-dimensional skull data acquired, for instance, from volume scans and extraction of the bone layers, or by range scanning a physical skull. The test data used for the examples in Section 6 was acquired using both types of scans. To speed up processing, a triangle mesh of the skull model comprised of 50-250k polygons is produced by mesh decimation techniques [Garland and Heckbert 1997]. In general, the original data should be simplified as little as possible since minute details on the skull can give important clues for the reconstruction. The mesh resolution is chosen for adequate responsiveness of our interactive skull editor application. In practice, it is helpful to have the original data set (or the physical skull) ready as a reference during editing.

In the editor, the skull model is equipped with landmarks, as shown in Fig. 4. Points on the skull surface are simply picked to create a landmark, which can then be moved around on the surface for fine positioning. Each landmark is associated with a vector in surface normal direction, corresponding to the typical direction of thickness measurements. As can be seen on the right image in Fig. 4, some skull / skin correspondences are in fact non-orthogonal to the skull surface in the area of the lips. This is corrected for at a later step of the fitting process, as described in Section 4.3. The landmark vector is scaled to the local tissue thickness, which is looked up automatically by the landmark's assigned name in a table based on RHINE's data (see Section 1.2). The specific set of landmarks used in our system is listed in Appendix A.

4 Fitting the Deformable Head Model

4.1 Head Model Structure

When the skull is tagged with landmarks, it serves as the target for deformation of the generic head model shown in Fig. 5. Since the head model is used in a physics-based animation system, it does

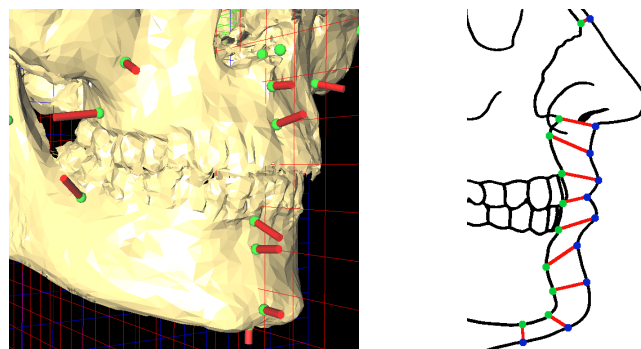


Figure 4: Skull landmark specification in the mouth area. Left: snapshot from our landmark editor; right: correspondences between skull and skin markers (Image after [y'Edynak and İşcan 1993])

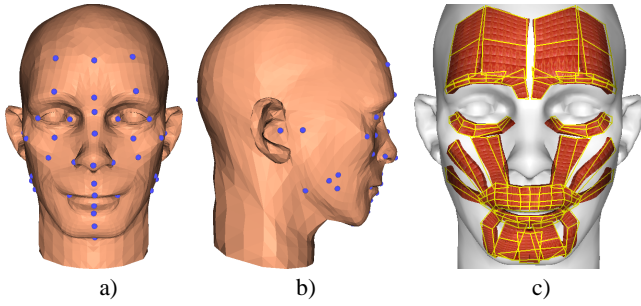


Figure 5: The deformable head model: a) head geometry with landmarks (blue dots), front view; b) side view; c) underlying muscles (red) created from layout grids (yellow).

not only consist of the visible outer geometry. The encapsulated structure includes:

the skin surface represented as a triangle mesh. The mesh resolution should be high enough to ensure good fitting results. Our template head mesh consists of 8164 triangles.

virtual muscles to control the animation. Each muscle is specified by a grid laid out on the skin, the actual muscle shape being computed automatically to fit underneath the skin surface. Each muscle consists of an array of fibers, which can contract in a linear or circular fashion. Our model includes 24 facial muscles responsible for facial expressions. Fig. 5(c) shows the muscle layout on the head template.

a mass-spring system connecting skin, muscles, and skull, built after the head model is fitted to the skull. For animation, muscles pull at spring nodes attached to their surface, in turn causing deformation of the spring mesh in the skin surface layer.

landmarks defined on the skin surface, as shown in Fig. 5(a) and (b). The majority of these landmarks corresponds to the landmarks interactively specified on the skull. These landmark pairs control the basic fitting of the head structure as described in Section 4.2. A few additional landmarks are only defined on the skin and are used for the final adjustments of the reconstructed shapes discussed in Section 4.3.

The head model is similar to the one in [Kähler et al. 2002], where detailed descriptions of the muscle model and animation approach can also be found.

4.2 Landmark-Based RBF Deformation

Given the deformable head model with n predefined skin landmark positions $\mathbf{p}_i \in \mathbb{R}^3$ and the corresponding landmarks $\mathbf{s}_i \in \mathbb{R}^3$ ($i = 1, \dots, n$) specified on the skull, we set up a space deformation that fits the skin and the muscle layout to the skull.

The target skull landmarks have associated tissue depth vectors \mathbf{d}_i , so corresponding skin landmark positions \mathbf{q}_i are defined as

$$\mathbf{q}_i = \mathbf{s}_i + \mathbf{d}_i.$$

The problem can now be treated as one of interpolation: we need to find a function \mathbf{f} that maps the \mathbf{p}_i to the \mathbf{q}_i :

$$\mathbf{q}_i = \mathbf{f}(\mathbf{p}_i), \quad i = 1, \dots, n.$$

The unknown function \mathbf{f} can be expressed by a radial basis function, i.e., a weighted linear combination of n basic functions ϕ_i and an additional explicit affine transformation:

$$\mathbf{f}(\mathbf{p}) = \sum_{i=1}^n \mathbf{c}_i \phi_i(\mathbf{p}) + \mathbf{R}\mathbf{p} + \mathbf{t}, \quad (1)$$

where $\mathbf{p} \in \mathbb{R}^3$ is a point in the volume, $\mathbf{c}_i \in \mathbb{R}^3$ are (unknown) weights, $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ adds rotation, skew, and scaling, and $\mathbf{t} \in \mathbb{R}^3$ is a translation component. The ϕ_i are defined by the source skin landmark points. According to BOOKSTEIN [1997], for deformation of biological solids an approach based on thin-plate splines is favorable. We thus use the simple biharmonic basic function $\phi_i(\mathbf{p}) := \|\mathbf{p} - \mathbf{p}_i\|_2$, which minimizes bending energy for the deformation [Duchon 1977].

To remove affine contributions from the weighted sum of the basic functions [Pighin et al. 1998; Carr et al. 2001], we include the additional constraints

$$\sum_{i=1}^n \mathbf{c}_i = \mathbf{0} \quad \text{and} \quad \sum_{i=1}^n \mathbf{c}_i^T \mathbf{p}_i = 0.$$

The resulting system of linear equations is solved for the unknowns \mathbf{R} , \mathbf{t} , and \mathbf{c}_i using a standard LU decomposition with pivoting, to obtain the final warp function \mathbf{f} . This function can now be used according to Eq. (1) to transform a point \mathbf{p} in the volume spanned by the landmarks. We apply \mathbf{f} to the skin and muscle components of the generic model in the following ways:

- The *skin mesh* is deformed by direct application of the function to the vertices of the mesh.
- The *muscles* are transferred to the new geometry by warping their layout grid vertices, followed by recomputation of the shape to fit the deformed skin mesh.

Since our landmark set is comprised of only 40 landmarks (see Appendix A), the computed deformation doesn't properly align the skin to the skull in all places, as can be seen in Fig. 6(a). Interactive specification of more landmarks puts an undesirable additional burden onto the user, so additional landmark pairs are computed automatically by interpolation between existing ones on the upper and back part of the cranium, as well as on the mandible, as shown in Fig. 6(b). The thickness value of an interpolated skull landmark is also interpolated, where only such skull areas are chosen for landmark interpolation where the tissue thickness is near-constant. Tissue depth interpolation would be problematic, for instance, in the mid-face area, where thickness values change drastically from the cheekbone to the mid-face region below.

4.3 Additional Reconstruction Hints

The tissue depth values at the marker positions define the basic shape of the reconstructed head, assuming depth measurements being always strictly orthogonal to the skull surface. As mentioned in Section 3, this assumption is not always valid. A number of rules are thus used in traditional facial reconstruction to help locate certain features of the face based on the skull shape, employing empirical knowledge about shape relations between skin and skull [Taylor

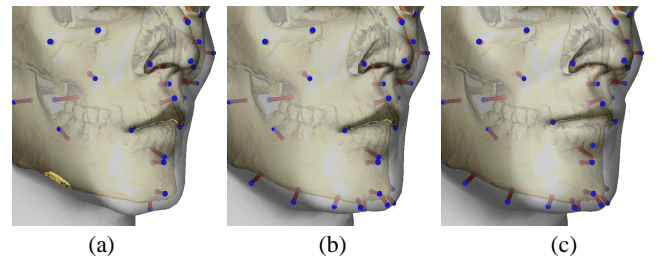


Figure 6: Fitting stages, shown on the lower face. a) Warp using only user-specified landmarks (some skull areas still intersecting the skin); b) with automatically interpolated landmarks on the mandible; c) using additional heuristics for lip and nose shaping.

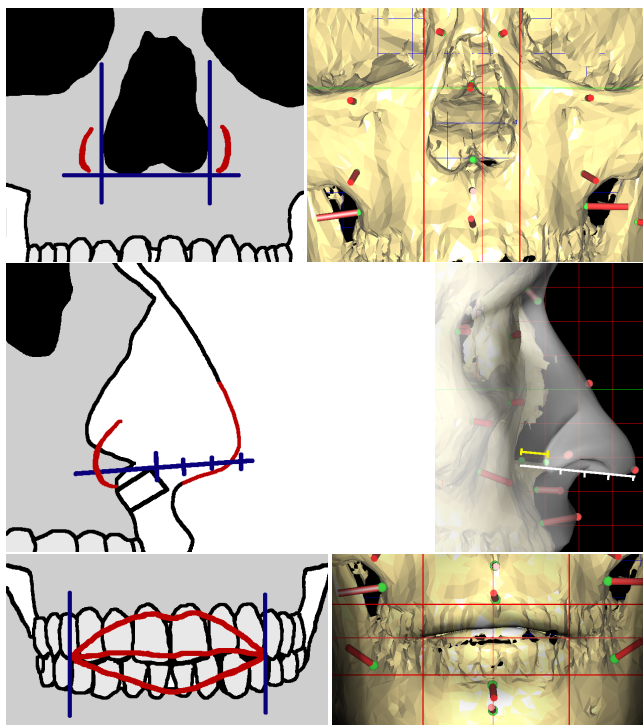


Figure 7: Comparison of heuristics used in traditional reconstruction (left) with our graphical interface (right). (Note: different skulls are used in the adjoining images.) Top: estimation of nose width; center: positioning of the nose tip; bottom: setting lip width, height, and mouth corner position.

2001]. We have translated some of these heuristics for use with the skull landmark editor: the final fitting result, as shown in Fig. 6(c), is obtained by including this additional user input.

To keep the user interface uniform, most rules are expressed by the placement of vertical and horizontal guides in a frontal view of the skull. From this user input, the placement of a few landmarks on the skin is adjusted, resulting in a new target landmark configuration. The updated landmark set is used to compute another warp function, which deforms the pre-fitted head model in the adjusted regions. Five rules influence the shape of the nose and the shape of the mouth, as shown in Fig. 7:

- The width of the nose wings corresponds to the width of the nasal aperture at its widest point, plus 5mm on either side in Caucasoids. In the editor, the user places two vertical guides to the left and right of the nasal aperture. From their position, the displacement of the two al^1 skin landmarks placed at the nose wings is computed (cf. Fig. 7, top row).
- The position of the nose tip depends on the shape of the anterior nasal spine. According to KROGMAN’s formula [Taylor 2001, p. 443], the tip of the nose is in the extension of the nasal spine. Starting from the z value of the tissue depth marker directly below the nose (mid-philtrum, see Appendix A), the line is extended by three times the length of the nasal spine (cf. the white and yellow lines in the rightmost image of Fig. 7, middle row). In the editor, begin and end points of the nasal spine are marked. The prn landmark at the nose tip is then displaced according to the formula.

¹see, e.g., [Farkas 1994] for a definition of standard facial landmarks

- The width of the mouth is determined by measuring the front six teeth, placing the mouth angles horizontally at the junction between the canine and the first premolar in a frontal view. Two vertical guides are used for positioning the ch landmarks located at the mouth angles (vertical lines in Fig. 7, bottom row).
- The thickness of the lips is determined by examining the upper and lower frontal teeth. Seen from the front, the transition between the lip and facial skin is placed at the transition between the enamel and the root part of the teeth. Two horizontal guides are placed by the user at the upper and lower transition, respectively. This determines the vertical position of the id and sd landmarks marking the lip boundary (top and bottom horizontal lines in Fig. 7, bottom row).
- The parting line between the lips is slightly above the blades of the incisors. This determines the vertical placement of the ch landmarks (middle horizontal line in Fig. 7, bottom row).

Using these heuristics, a better estimate of the mouth and nose shapes can be computed. The effect is strongest on the lip margins, since the assumption of an orthogonal connection between corresponding skin and skull landmarks is in fact not correct at these sites, as the right part of Fig. 4 shows. The initial deformation thus gives a good estimate of the tissue thickness of the lips while the second deformation using the information provided by interactive guide adjustment refines the vertical placement of the lip margins.

5 Facial Expressions and Rendering

In manual facial reconstruction, a neutral pose of the face is preferred as the most “generic” facial expression. Other expressions could be helpful for identification purposes, but the cost of modeling separate versions of the head model is prohibitive. In our virtual reconstruction approach, this does not pose a problem. Since the fitted head model has the animatable structure of skin and muscles, different facial expressions can be assumed by setting muscle contractions, as in other physics-based facial animation systems [Kähler et al. 2001; Lee et al. 1995]. Fig. 8 shows how muscles are used to form different facial expressions.

For a completely animatable head model, it is necessary to include a separately controllable mandible, a tongue, rotatable eyeballs, and eye lids into the head model. We have decidedly left them out of the reconstruction approach since these features are not particularly useful in this application: while a modest change of expression such as a smile or a frown might aid identification, rolling of eyes, blinking, and talking would probably not. It is also nearly impossible to correctly guess details such as a specific way of speaking—errors in this respect would produce rather misleading results in a real identification case. The effort of placing tongue, eye, and potentially teeth models thus does not offset the benefits.



Figure 8: Expressions on the generic head model and the corresponding muscle configurations.

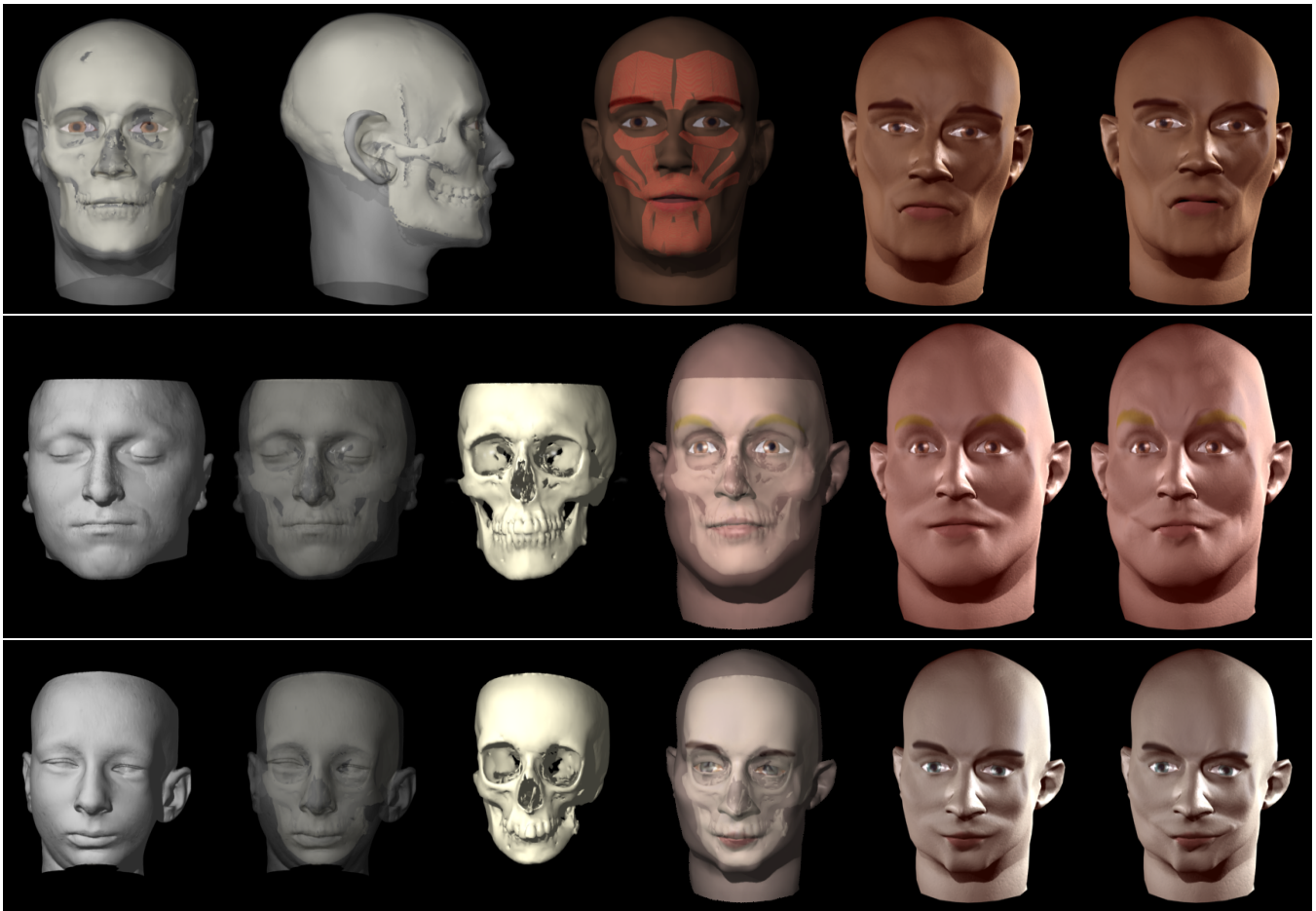


Figure 9: Examples of facial reconstructions created with our system. Top: model created from a scanned real skull, showing fit of skin to skull, transferred muscles, and two facial expressions. Middle: Reconstruction from a volume scan of a male, showing the actual face as contained in the data, superimpositions of the actual and the reconstructed face with the skull, and the reconstruction with neutral and “worried” expression. Bottom: Reconstruction from volume scan of a female with strong skull deformations. The CT data sets don’t contain the top and bottom of the heads, thus the source skull and face models are cut off. The actual head height had to be guessed in these cases.

If additional information about the modeled person is available, for instance, from remnants of hair found with the skull, the resulting mesh can be colored correspondingly. Our system includes basic capabilities for coloring the parts associated with skin, lip, and eyebrows in the model’s texture map. Colors can be adjusted interactively in HSV space on the reconstructed head model. Finally, the color adjustments are merged into a neutral base texture and saved as a new texture map. The fitted, texture-mapped triangle mesh can be easily imported into various rendering packages for display. The examples shown in Fig. 9 show three different skin colorations created in this way.

6 Results

We have tested our technique on a real skull that was made available to us by a forensic institute and on two medical volume scans. All data pertains to individuals of Caucasian type. Each reconstruction required approximately an hour of interactive work, excluding time for data acquisition.

The real skull, depicted on the first page of this paper, was unearthed on a construction site and belongs to an unidentified male, approximately 35 years of age. As can be seen from the hole in the frontal bone, he was killed by a head shot—the owner of this skull

probably was a war victim or a soldier. After scanning the skull, the resulting mesh was simplified to 100k triangles. Interactive placement of skull landmarks and facial feature guides was relatively easy in this case since the skull is complete and in good condition. Due to its war-time origin, we assumed the face to be rather skinny, so we selected the “slender” tissue thickness table. Fitting results can be seen in Fig. 9, top row. Since the actual appearance of the individual is unknown, the accuracy of the reconstruction can only be guessed. Nonetheless, our reconstruction seems plausible. Notably, the shape of the chin, which can be predicted from the corresponding region on the skull, has been reproduced well.

To show examples utilizing other data sources, and also for validation, we extracted skull and skin surfaces from medical volume scans. The first data set, shown in the middle row of Fig. 9, pertains to a male subject of roughly 30 years. The subject’s face is rather bulky, so we chose the “obese” tissue thickness data set (in a real case, this choice would have to be made based on other available information such as the size of clothes, if present). Our first reconstruction attempts showed a consistent emphasis on prominent cheek bones and hollow cheeks: no matter which data set we picked, the face would become more bulky, but not show the expected general roundness of the face. This effect is demonstrated in Fig. 10 on variations of our first model. A closer examination

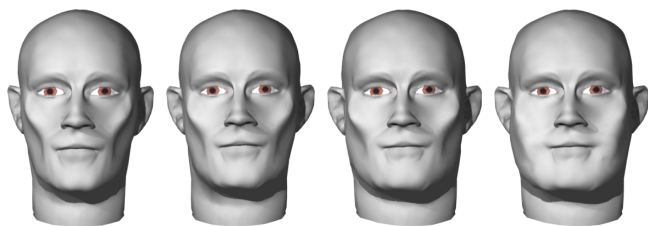


Figure 10: Left to right: RHINE’s traditional “slender”, “average”, and “obese” tissue depth tables (cf. [Taylor 2001, p. 350 ff.]) often result in hollow cheeks and prominent cheekbones (see also Fig. 9). Rightmost image: the shape can be improved by “bulging out” the affected mesh areas.

revealed that the reason lies in the relatively low thickness values RHINE assigned to the landmarks defining the cheek region (*sbm2* and *spm2* in Table 1). After excluding these two landmarks, we obtained the results shown in Fig. 9. The rightmost image in Fig. 10 shows how simple mesh modeling techniques could be used at this point to improve and individualize the reconstruction.

The second volume data set shows a female patient with strong skull deformations. We produced a reconstruction of this face to test the method with a decidedly non-average skull shape. The result can be seen in the bottom row of Fig. 9. Since our automatic landmark interpolation scheme (see Section 4.2) is designed to handle the normal range of skull variations, the unusual shape of the mandible resulted in very sparse sampling of the chin area. Another prominent feature of the skull data is the protrusion of one incisor, pushing the upper lip to the front. We modeled this effect by moving the *sd* landmark a few millimeters down onto the blade of the incisor, thus pushing the associated skin landmark forward as well. This did not impair the positioning of the upper lip boundary since this is adjusted separately by the mouth guides (cf. Fig. 7).

7 Conclusion and Future Work

The face reconstruction approach presented in this paper mirrors the manual tissue depth method and thus has essentially the same prediction power. Our results show overall good reproduction of facial shape and proportions, and some surprisingly well-matched details. It should be noted that our examples were produced by computer scientists with no training in forensic reconstruction.

The advantages of the computerized solution are evident: instead of weeks, it takes less than a day to create a reconstructed face model, including scanning of the skull. Once the scan data is marked with landmarks, different varieties such as slimmer or more obese versions can be produced within seconds at the push of a button, which is practically impossible with the manual method due to the vast amount of time needed for production of a single model. Slight variations in facial expression can also be obtained quite easily by animating the muscle structure underlying the model.

Since the virtual reconstruction is based on 3D scans, which can be acquired contact-free, the risk of damage to the original skull is reduced. On the other hand, the scanning process has inherent limitations: depending on the maximum resolution of the digital scanner, much of the finer detail on the skull is lost. The delicate structure of, for instance, the nasal spine cannot be fully captured with current scanning technology. For this reason, it is necessary to consult the original skull from time to time for reference.

In our experiments, we often found that surface normals on the scanned skull geometry do not always behave the way they should, reflecting the orientation of the surface only very locally. It might be useful to consider an average of normals in a larger area around

the landmark position to solve this. Sometimes, it would be desirable to adjust the orientation manually.

The interactive system allows for an iterative reconstruction approach: a model is produced quickly from a given landmark configuration, so landmarks can be edited repeatedly until the desired result is obtained. The emphasis on the interaction component makes the speed of the fitting process an important issue. While the actual calculation of the warp function and the deformation of the mesh are performed instantaneously, about five seconds are needed in our test setting on a 1.7 GHz Pentium Xeon to examine skull and skin for potential insertion of additional landmarks. This time is for the largest part used for ray intersections of the skull and skin meshes, which are done in a brute force manner. We expect a big speed-up through the use of space partitioning techniques.

For practical use, the facial reconstruction system should provide more editing facilities for skin details and hair. Useful additions include, for instance, a choice of templates for haircuts and facial features such as eyebrow shapes, beards, and wrinkles. At this point, large-scale validation of the system would be necessary to evaluate the usability of the system.

As TAYLOR writes in her book, the tissue depth values should not be taken at face value in three-dimensional facial reconstruction, but rather act as guides for the final facial reconstruction, which still relies heavily on artistic skills and intuition. Our tests confirm that strict adherence to RHINE’s data for the solution of the interpolation problem is too limiting. This indicates not a weakness in our method, but reflects the low number of samples (between 3 and 37 in each group) and the technical limitations at the time RHINE assembled his data tables. Given the current state of technology, more samples of higher precision could be acquired, resulting in much more comprehensive and usable data. Ultimately, computer-based facial reconstruction could then even become superior to the traditional approach.

8 Acknowledgements

The authors would like to thank Dr. D. Buhmann from the Institute of Forensic Medicine, Saarland University, for his valuable comments and for providing the CT data sets.

References

- ARCHER, K. M. 1997. *Craniofacial Reconstruction using hierarchical B-Spline Interpolation*. Master’s thesis, University of British Columbia, Department of Electrical and Computer Engineering.
- BLANZ, V., AND VETTER, T. 1999. A Morphable Model for the Synthesis of 3D Faces. In *Proc. ACM SIGGRAPH 1999*, ACM Press / ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series, 187–194.
- BOOKSTEIN, F. L. 1997. *Morphometric Tools for Landmark Data*. Cambridge University Press.
- CARR, J. C., BEATSON, R. K., CHERRIE, J. B., MITCHELL, T. J., FRIGHT, W. R., MCCALLUM, B. C., AND EVANS, T. R. 2001. Reconstruction and Representation of 3D Objects With Radial Basis Functions. ACM Press / ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series, 67–76.
- CHADWICK, J. E., HAUMANN, D. R., AND PARENT, R. E. 1989. Layered Construction for Deformable Animated Characters. In *Computer Graphics (Proc. ACM SIGGRAPH 89)*, 243–252.
- DUCHON, J. 1977. Spline minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller, Eds., vol. 571 of *Lecture Notes in Mathematics*, 85–100.
- FARKAS, L. G., Ed. 1994. *Anthropometry of the Head and Face*, 2nd ed. Raven Press.
- FEDOSYUTKIN, B. A., AND NAINYS, J. V. 1993. *Forensic Analysis of the Skull*. Wiley-Liss, ch. 15: The Relationship of Skull Morphology to Facial Features, 199–213.

- GARLAND, M., AND HECKBERT, P. S. 1997. Surface simplification using quadric error metrics. In *SIGGRAPH 97 Conference Proceedings*, 209–216.
- GRÜNER, O. 1993. *Forensic Analysis of the Skull*. Wiley-Liss, ch. 3: Identification of Skulls: A Historical Review and Practical Applications.
- HELMER, R. P., RÖHRICHT, S., PETERSEN, D., AND MÖHR, F. 1993. *Forensic Analysis of the Skull*. Wiley-Liss, ch. 17: Assessment of the Reliability of Facial Reconstruction, 229–246.
- KÄHLER, K., HABER, J., AND SEIDEL, H.-P. 2001. Geometry-based Muscle Modeling for Facial Animation. In *Proc. Graphics Interface 2001*, 37–46.
- KÄHLER, K., HABER, J., YAMAUCHI, H., AND SEIDEL, H.-P. 2002. Head shop: Generating animated head models with anatomical structure. In *ACM SIGGRAPH Symposium on Computer Animation*, ACM SIGGRAPH, 55–64.
- KOCH, R. M., GROSS, M. H., AND BOSSHARD, A. A. 1998. Emotion Editing using Finite Elements. In *Computer Graphics Forum (Proc. Eurographics '98)*, vol. 17, C295–C302.
- KROGMAN, W. M. 1946. The reconstruction of the living head from the skull. *FBI Law Enforcement Bulletin* (July).
- LEBEDINSKAYA, G. V., BALUEVA, T. S., AND VESELOVSKAYA, E. V. 1993. *Forensic Analysis of the Skull*. Wiley-Liss, ch. 14: Principles of Facial Reconstruction, 183–198.
- LEE, Y., TERZOPOULOS, D., AND WATERS, K. 1995. Realistic Modeling for Facial Animations. In *Proc. ACM SIGGRAPH 1995*, ACM Press / ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series, 55–62.
- MICHAEL, S., AND CHEN, M. 1996. The 3D reconstruction of facial features using volume distortion. In *Proc. 14th Eurographics UK Conference*, 297–305.
- MIYASAKA, S., YOSHINO, M., IMAIZUMI, K., AND SETA, S. 1995. The computer-aided facial reconstruction system. *Forensic Science Int.* 74, 1-2, 155–165.
- MOORE, W. J., AND LAVELLE, C. L. B. 1974. *Growth of the Facial Skeleton in the Hominoidea*. Academic Press, London.
- PIGHIN, F., HECKER, J., LISCHINSKI, D., SZELISKI, R., AND SALESIN, D. H. 1998. Synthesizing Realistic Facial Expressions from Photographs. In *Proc. ACM SIGGRAPH 1998*, ACM Press / ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series, 75–84.
- PLATT, S. M., AND BADLER, N. I. 1981. Animating Facial Expressions. In *Computer Graphics (Proc. ACM SIGGRAPH 81)*, 245–252.
- RHINE, J. S., AND CAMPBELL, H. R. 1980. Thickness of facial tissues in American blacks. *Journal of Forensic Sciences* 25, 4, 847–858.
- RHINE, J. S., AND MOORE, C. E. 1984. Tables of facial tissue thickness of American Caucasoids in forensic anthropology. *Maxwell Museum Technical Series 1*.
- SCHEEPERS, F., PARENT, R. E., CARLSON, W. E., AND MAY, S. F. 1997. Anatomy-Based Modeling of the Human Musculature. In *Proc. ACM SIGGRAPH 1997*, ACM Press / ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series, 163–172.
- SEDERBERG, T. W., AND PARRY, S. R. 1986. Free-Form Deformation of Solid Geometric Models. *Computer Graphics (Proc. ACM SIGGRAPH 86)* 20, 4 (Aug.), 151–160.
- SZELISKI, R., AND LAVALLÉE, S. 1996. Matching 3-D Anatomical Surfaces with Non-Rigid Deformations using Octree-Splines. *International Journal of Computer Vision* 18, 2, 171–186.
- TAYLOR, K. T. 2001. *Forensic Art and Illustration*. CRC Press LLC.
- TERZOPOULOS, D., AND WATERS, K. 1990. Physically-based Facial Modelling, Analysis, and Animation. *Journal of Visualization and Computer Animation* 1, 2 (Dec.), 73–80.
- VANEZIS, P., VANEZIS, M., MCCOMBE, G., AND NIBLETT, T. 2000. Facial reconstruction using 3-D computer graphics. *Forensic Science Int.* 108, 2, 81–95.
- WATERS, K. 1987. A Muscle Model for Animating Three-Dimensional Facial Expression. In *Computer Graphics (Proc. ACM SIGGRAPH 87)*, 17–24.
- WEIJS, W. A., AND HILLEN, B. 1986. Correlations between the cross-sectional area of the jaw muscles and craniofacial size and shape. *Am. J. Phys. Anthropol.* 70, 423–431.
- WILHELMS, J., AND VAN GELDER, A. 1997. Anatomically Based Modeling. In *Proc. ACM SIGGRAPH 1997*, ACM Press / ACM SIGGRAPH, Computer Graphics Proceedings, Annual Conference Series, 173–180.
- Y'EDYNAK, G. J., AND İŞCAN, M. Y. 1993. *Forensic Analysis of the Skull*. Wiley-Liss, ch. 16: Anatomical and Artistic Guidelines for Forensic Facial Reconstruction, 215–227.
- ZOLLIKOFER, C. P. E., PONCE DE LEÓN, M. S., AND MARTIN, R. D. 1998. Computer-Assisted Paleoanthropology. *Evolutionary Anthropology* 6, 2, 41–54.

A Landmark Set used for Reconstruction

Table 1 lists the paired landmarks on skin and skull that are used for the facial reconstruction approach described in this paper. Most skull landmark names and descriptions are taken from [Taylor 2001, page 350 ff.]. Short skull landmark names are listed in the *id* column. We have tried to adhere to naming conventions used in the forensic and anthropometric literature as much as possible [Taylor 2001; y'Edynak and İşcan 1993; Farkas 1994]. For simplicity, corresponding landmarks on skull and skin have the same short name in our system, which is not generally the case in the literature. In a few cases, marked by * in the table, we invented short names. Not all skull landmarks have an “official” counterpart on the skin, so we placed the corresponding skin markers using our own judgment. The *mp* landmark pair is not part of the standard set. We added it to improve the alignment of skin to skull in the region behind the ears, where the mastoid process adds a bulge to the skull.

<i>name</i>	<i>id</i>	<i>description</i>
Midline		
Supraglabella	tr	Above glabella, identified with the hairline
Glabella	g	The most prominent point between the supraorbital ridges in the midsagittal plane
Nasion	n	The midpoint of the suture between the frontal and the two nasal bones
End of nasals	na	The anterior tip or the farthest point out on the nasal bones
Mid-philtrum	a	The mid line of the maxilla (east and west), placed as high as possible before the curvature of the anterior nasal spine begins
Upper lip margin (Supradentale)	sd	Centered between the maxillary (upper) central incisors at the level of the Cementum Enamel Junction (CEJ)
Lower lip margin (Infradentale)	id	Centered between the mandibula (lower) central incisors at the level of the Cementum Enamel Junction (CEJ)
Chin-lip fold (Supramentale)	b	The deepest mid line point of indentation on the mandible between the teeth and the chin protrusion
Mental eminence (Pogonion)	pog	The most anterior or projecting point in the mid line on the chin
Beneath chin (Menton)	me	The lowest point on the mandible
Bilateral		
Frontal eminence	fe*	Place on the projections at both sides of the forehead
Supraorbital	sci	Above the orbit, centered on the upper most margin or border
Suborbital	or	Below the orbit, centered on the lower most margin or border
Endocanthion	en	point at the inner commissure of the eye fissure; the landmark on the skin is slightly lateral to the one on the bone
Exocanthion	ex	point at the outer commissure of the eye fissure; the landmark on the skin is slightly medial to the one on the bone
Inferior malar	im	The lower portion of the maxilla, still on the cheekbone
Lateral orbit	lo	Drop a line from the outer margin of the orbit and place the marker about 10 mm below the orbit
Zygomatic arch, midway	zy	Halfway along the zygomatic arch (generally the most projecting point on the arch when viewed from above)
Supraglenoid	sg	above and slightly forward of the external auditory meatus
Gonion	go	The most lateral point on the mandibular angle
Supra M ²	spm2*	Above the second maxillary molar
Occlusal line	ol	On the mandible in alignment with the line where the teeth occlude or bite
Sub M ₂	sbm2*	Below the second mandibular molar
Mastoid process	mp*	Most lateral part on the mastoid process behind and below the ear canal

Table 1: Landmark set used for face reconstruction.



Universal Capture: Image-based Facial Animation and Rendering for *The Matrix* sequels

George Borshukov,
Dan Piponi, Oystein Larsen, J.P. Lewis,
Christina Tempelaar-Lietz
ESC Entertainment



The Challenge

- Our task was to produce photorealistic animated renditions of known actors:
 - Keanu Reeves, Laurence Fishburne, Hugo Weaving
- The synthetic reproductions needed to intercut seamlessly with footage of the real actor



A Daunting Task

- Photorealistic human faces are the ultimate challenge for computer graphics
- Faces are particularly scrutinized by human observers
 - We grow up and then spend most of our lives looking at faces
 - Incredible variety, richness & subtlety of human facial movement
 - Human viewer's extreme sensitivity to facial nuances
- No examples of believable human face at the time



Motivation

- Traditional facial animation (blendshapes, muscle deformers) would not produce realistic results re-creating a real actor
- Believable facial rendering requires textures that change over time
 - Color changes due to blood flow, skin strain
 - Fine wrinkles form and disappear
 - Microscopic self-shadowing effects



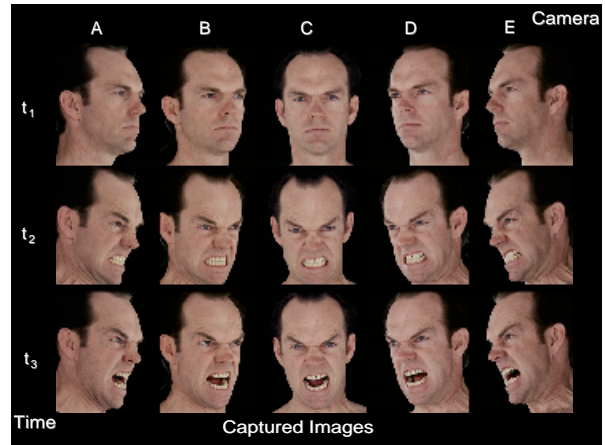
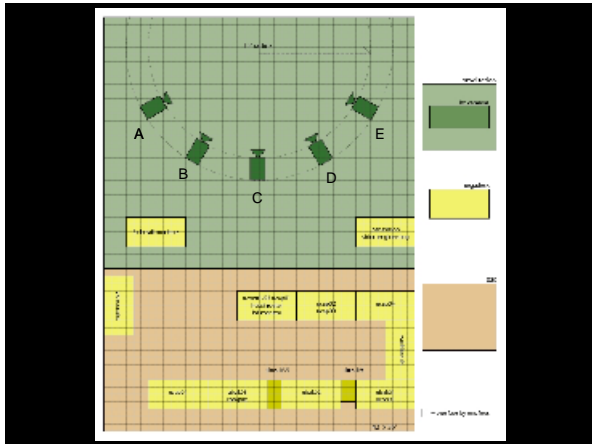
Universal Capture

- Our previous experience with image-based and computer vision approaches (*What Dreams May Come*, *Matrix I*) suggested a "non-traditional" approach
- Capture a 3-D recording of an actor's performance
- Play it back with different camera and lighting
- Combine two powerful vision techniques: optical flow and photogrammetry



Hi-Definition Capture

- Five synchronized cameras capture the actor's performance in ambient lighting
- Sony/Panavision HDW-F900 cameras
 - Portrait mode 1080x1920 resolution
 - 60i for maximal temporal information
 - 1/500th sec shutter to minimize motion blur
- Real-time capture/storage
 - Computer workstations with HD capture boards
 - 21 terabyte disk arrays
 - Tape robot for overnight data backup



Optical Flow + Photogrammetry

- Optical flow in each camera view
- Photogrammetric reconstruction of camera locations
- Core algorithm "warps" a neutral face model:
 - Project vertices into each camera
 - Find 2-D motion of each vertex
 - Project back into 3-D
 - Triangulate to obtain 3-D motion

Optical Flow + Photogrammetry

- Result: *markerless* capture of the complete deforming face geometry

Optical Flow + Photogrammetry



- Result: *markerless* capture of the complete deforming face geometry

Optical Flow + Photogrammetry



- Result: *markerless* capture of the complete deforming face geometry

Optical Flow + Photogrammetry



- Result: *markerless* capture of the complete deforming face geometry

Optical Flow + Photogrammetry



- Result: *markerless* capture of the complete deforming face geometry

Optical Flow + Photogrammetry



- Result: *markerless* capture of the complete deforming face geometry

Optical Flow “drift”

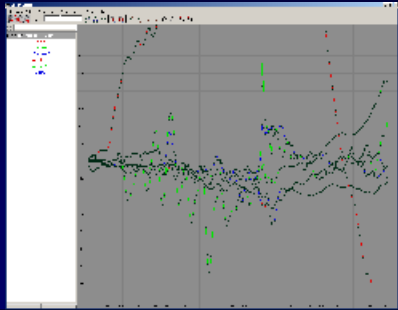


- Optical flow errors accumulate over time
- Partially address by reverse optical flow
- After a visible error has accumulated
 - Manually correct using keyshapes
 - Algorithmically interpolate and propagate the correction back through the performance

Rigid vs. Deformable Motion



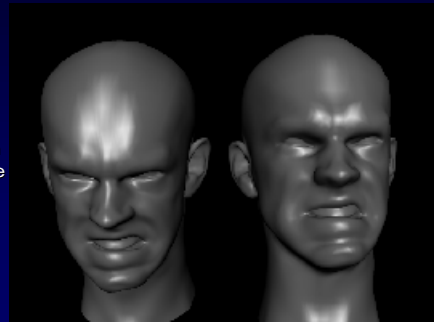
- Underlying rigid (skull) transformation
- Recovered curves estimated using a least squares procedure
- Can apply signal processing to preserve nuance



Rigid vs. Deformable Motion



- Subtracting the rigidly transformed neutral face from the 3-D reconstruction (left) gives the animated facial deformation (right)



Rigid vs. Deformable Motion



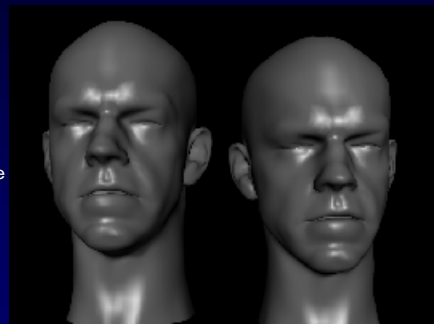
- Subtracting the rigidly transformed neutral face from the 3-D reconstruction (left) gives the animated facial deformation (right)



Rigid vs. Deformable Motion



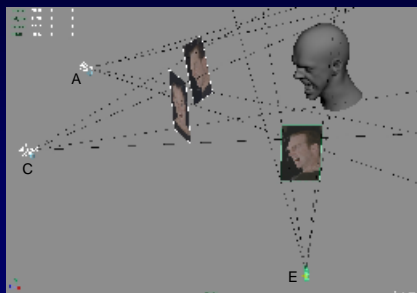
- Subtracting the rigidly transformed neutral face from the 3-D reconstruction (left) gives the animated facial deformation (right)



Animated Texture Map Extraction



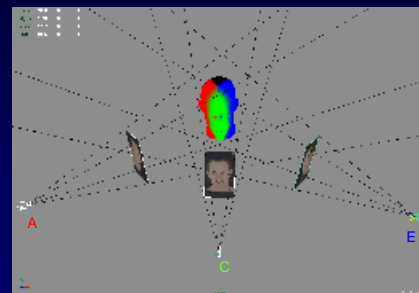
- Image re-projection
- Merge ambient images from multiple camera views over time to produce seamless animated UV color maps



Animated Texture Map Extraction



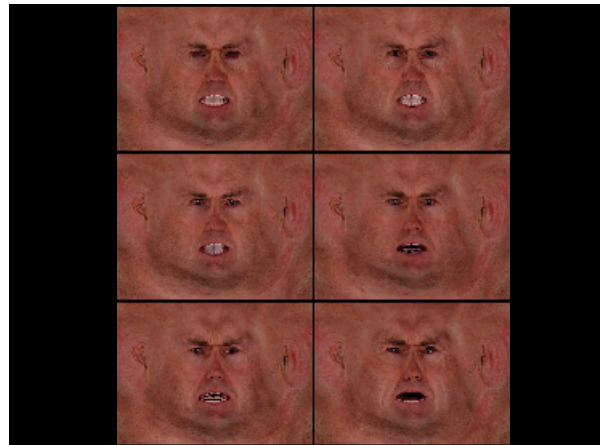
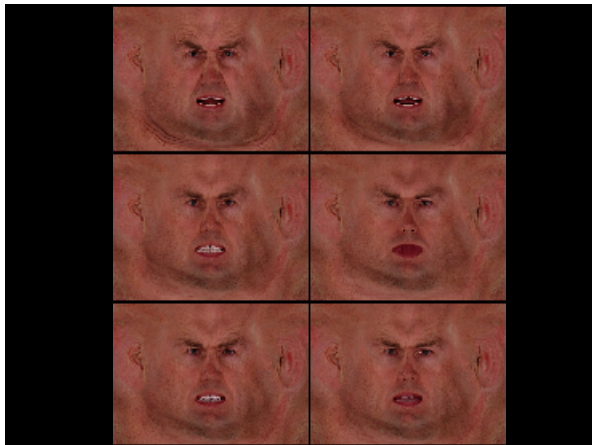
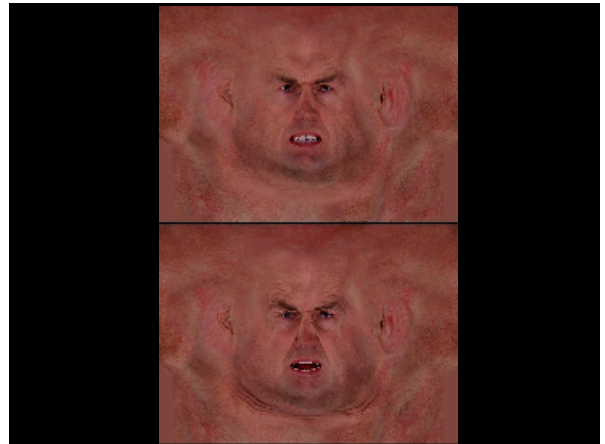
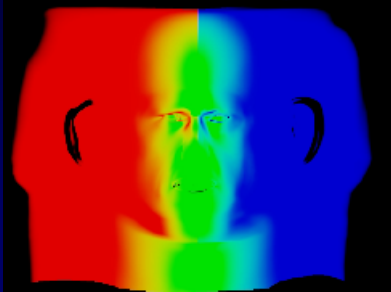
- Image re-projection and blend regions



Animated Texture Map Extraction



- Regions of seamless blend in UV space



Early Observations – Winter/Spring 2000



- Surface detail on the face (of human skin)
 - Unique: pores, wrinkles, moles, scars, etc.
 - Highly variable spatially
 - Very small scale ~100 micron features
 - Has extremely complex pattern
 - Hard to paint
 - Even harder to generate procedurally
- Color texture detail also unique and complex
 - We knew we can address that thanks to our photogrammetry and image-based rendering

Early Observations – Winter/Spring 2000

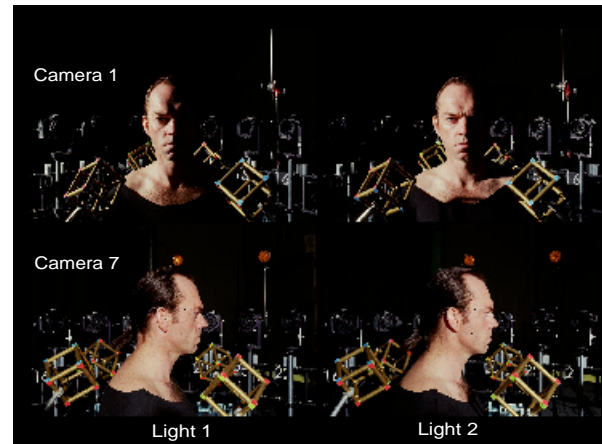


- Reflectance – is BRDF enough or ...
- Skin is *translucent*, conventional shaders will not work
- Lighting is just as important
 - Area lights with ray traced shadows
 - Lights from every direction of the environment
 - Does “global illumination” play a role?

BRDF Capture



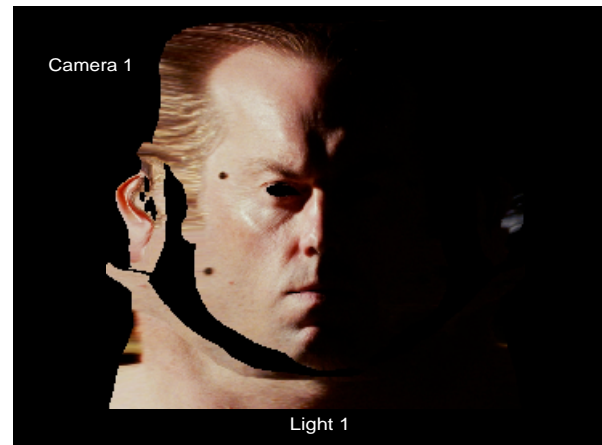
- Marschner et. al. – Image-based BRDF measurement
- Had lots of cameras lying around from the Bullet Time rig and a way to trigger them simultaneously
- Capture actor illuminated from various lighting directions with 30 cameras around the head



BRDF Image Alignment



- Photogrammetry used to reconstruct the camera positions
- Color calibrated, image space aligned images from each camera brought into a common UV space by projection onto cyberscan model
- The registered images implicitly contain skin reflectance for various incoming and outgoing light directions



Data-Derived Analytical BRDF



- Due to imperfections in our color calibration, image alignment, and cyberscan it was hard to fit a model automatically
- Parameters for an approximate analytical BRDF are derived from this data:
 - Lambert-like diffuse component
 - Phong-like specular with Fresnel effect (acknowledgement: Matthew Landauer)

Surface Detail



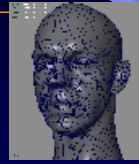
- Applying BRDF to existing model without bump map detail
 - Images were disturbingly fake
 - Tried procedural cellular texture approach – dismal failure
 - Tried extracting bump detail from color map (already had access to UCap color texture maps) – better result but hardly photorealistic
- Convinced that we had to scan the real actor's facial detail somehow

Raw Facial Geometry

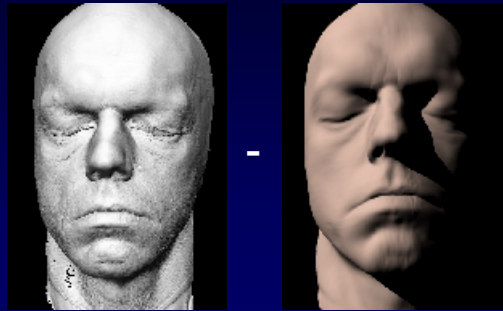
- Plaster casts of the actors
 - Acquired through the movie production
- Aruis3d scanning technology
 - ~\$20 million of government funding over 10+ years
 - Service provided by XYZRGB
 - 100-micron scan of actors' faces
 - Highest resolution model: 10 million triangles
 - Provided multiple resolutions

Detail Extraction Approach

- Base resolution quad mesh (constructed with Paraform)
- Use as subdivision surface
- Residual displacement obtained with mental ray lightmapping and custom shader
 - Ray trace from the subdivision surface to the raw scan
 - Store distance to intersection in a UV map



Agent Smith Detail Extraction

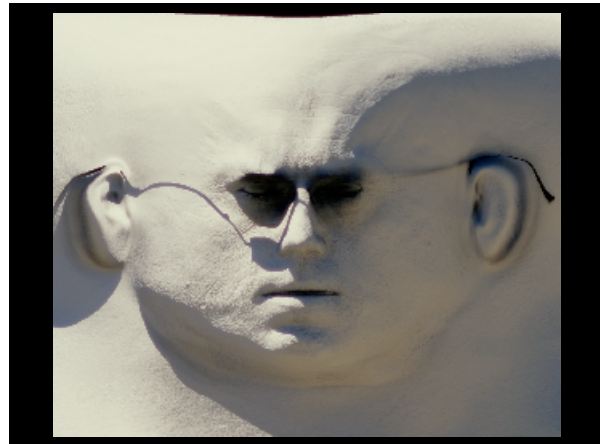
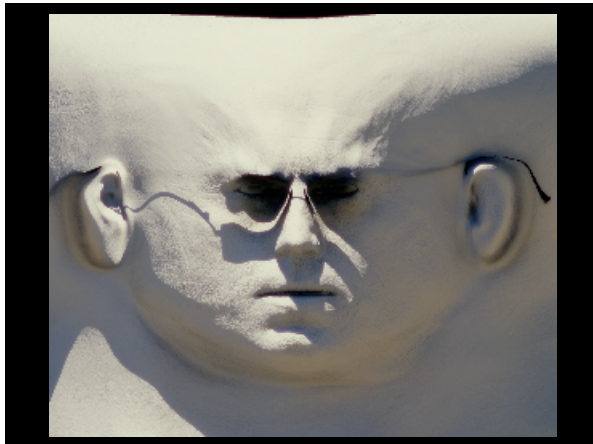


Agent Smith Bump





Subsurface Scattering

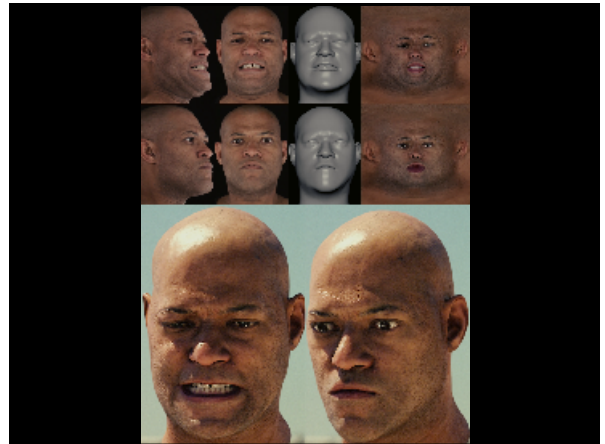
- We were very close but the renders looked more like granite than skin – Henrik was right!
 - Existing subsurface models: complex, also not 100% convincing
- Instead, approximately simulate light diffusion in the image map domain
 - Different diffusion length for different colors
 - Heavily translucent areas (ears) handled by ray tracing



Rendering

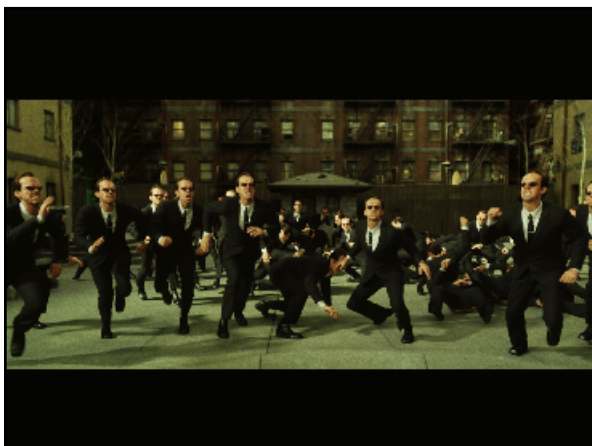
- Lighting Reconstruction Toolkit
 - image-based lighting approach gave us the realistic lighting from an environment
- Renderer: mental ray
 - Lightmapping
 - Ray traced shadows
- Reference photo shoot with the actors to verify our results






The Matrix Reloaded & Revolutions

- (Stills)



Video



Acknowledgements



- Oystein Larsen – pipeline software design & implementation
Performance processing co-supervision
- Dan Piponi – optical flow & 3-d motion reconstruction algorithms
- J.P. Lewis – subsurface scattering algorithm
- Christina Tempelaar-Lietz – additional software design & implementation

Acknowledgements



- Steve Avoujageli – additional pipeline tools development
- Ken Faiman, Steve Rembuskos, Mike Dalzell – for their incredible artistry demonstrated during the keyshaping stage of performance processing
- Paul Ryan, John Llewellyn, and Ryan Schnizlein – for their work on the HiDef capture setup
- Dan Piponi – for suggesting the detail extraction approach
- Brian Freisinger – for modeling, UVing the heads

Acknowledgements



- Matthew Landauer – for his contributions to the image-based skin shader derivation
- Rene Garcia – for his amazing paint work
- Ryan Todd (RT) – for tackling all the head replacement shots in the film
- Haarm-Pieter Duiker, Tadao Mihashi, and Ben Gunsberger
- Thomas Driemeyer/mental images – for making important extensions to their lightmapping feature
- John Jack, Kim Libreri, and John Gaeta – for embracing and believing in this “unorthodox” approach

Universal Capture – Image-based Facial Animation and Rendering for “The Matrix” Sequels

George Borshukov, et. al., ESC Entertainment
Appendix to SIGGRAPH’2004 course notes on Facial Modeling and Animation

Introduction

The VFX R&D stage for *The Matrix* sequels was kicked off in January 2000 with the challenge to create realistic human faces. The ultimate challenge in photorealistic computer graphics is rendering believable human faces. We are trained to study the human face since birth, so our brains are intimately familiar with every nuance and detail of what human skin is supposed to look like. The challenge of rendering the appearance of human skin is further complicated by some technical issues such as the fact that skin is a highly detailed surface with noticeable features in the order of ~100 microns and the fact that skin is translucent. For animation, we believed that traditional approaches like muscle deformers or blend shapes would simply never work, both because of the richness of facial movement and because of the human viewer’s extreme sensitivity to facial nuances. Our task was further complicated as we had to recreate familiar actors such as Keanu Reeves, Laurence Fishburne, and Hugo Weaving. Our team had been very successful at applying image-based techniques for photorealistic film set/location rendering, so we decided to approach the animation problem from the image-based side again. We wanted to produce a 3-d recording of the real actor’s performance and be able to play it back from different angles and under different lighting conditions. Just as we can extract geometry, texture, or light from images, we are now able to extract movement. Universal Capture combines two powerful computer vision techniques: optical flow and photogrammetry.

HiDef Capture Setup

We used a carefully placed array of five synchronized cameras that captured the actor’s performance in ambient lighting. For the best image quality we deployed a sophisticated arrangement of Sony/Panavision HDW-F900 cameras and computer workstations that captured the images in uncompressed digital format straight to hard disks at data rates close to 1G/sec.

Optical Flow + Photogrammetry

We use optical flow to track each pixel’s motion over time in each camera view. The result of this process is then combined with a cyberscan model of a neutral expression of the actor and with photogrammetric reconstruction of the camera positions. The algorithm works by projecting a vertex of the model into each of the cameras and then tracking the motion of that vertex in 2-d using the optical flow where at each frame the 3-d position is estimated using triangulation. The result is an accurate reconstruction of the path of each vertex though 3-d space over time.

Keyshaping, Adapt, Removing Global Motion

Optical flow errors can accumulate over time, causing an undesirable drift in the 3-d reconstruction. To minimize the drift we make use of reverse optical flow. On this production the problem was eliminated by introducing a manual keyshaping step: when the flow error becomes unacceptably large the geometry is manually corrected and the correction is then algorithmically propagated to previous frames.

The reconstructed motion contains the global “rigid” head movement. In order to attach facial performances to CG bodies or blend between different performances this movement must be removed. We estimate the rigid transformation using a least squares fit of a neutral face and then subtract this motion to obtain the non-rigid deformation.

Texture Map Extraction

No believable facial rendering can be done without varying the face texture over time. The fact that we did not use any markers on the face to assist feature tracking gave us the important advantage that we could combine the images from the multiple camera views over time to produce animated seamless UV color maps capturing important textural variation across the face, such as the forming of fine wrinkles or changes in color due to strain, in high-res detail on each side of the face.

Facial Surface Detail

Although the extracted facial animation had most of the motion nuances it lacked the small-scale surface detail like pores and wrinkles. The geometry used for our rendering was based on a 100-micron resolution scan of a plaster cast mold of the actors’ faces. Arius3d provided the scanning technology. These scans had extremely high polygonal counts (10 million triangles; see Fig. 1). To use these models in production and preserve the detail we deployed the following technique. A low-res ~5K quad model was constructed using Paraform software. The model was given a UV parameterization and then used as a subdivision surface. The high resolution detail was extracted using the lightmapping feature of the mental ray renderer combined with custom shaders that performed ray tracing from the low-res subdivision surface model to the high-detailed 10M triangle raw scan; the distance difference is stored in a displacement map. We applied the low frequency component of this map as displacement; the high frequency component was applied using bump mapping. Dynamic wrinkles were identified by image processing on the texture maps; these are then isolated and layered over the static bump map.

Image-based Derivation of Skin BRDF

Our skin BRDF was derived using an image-based approach. In Summer 2000 as part of the early stages of *Matrix Reloaded* R&D we had a setup, which consisted of 30 still cameras arranged around the actor’s head. Actors were photographed illuminated with a series of light sources from different directions (see Fig. 2). The setup was carefully color calibrated and photogrammetry was used to precisely reconstruct the camera positions and head placement with respect to each camera for each image. The collected image data from each camera was brought into a common UV space through reprojection using a cyberscan model of the actor. This convenient space (see Fig. 3) allowed us to analyze the skin reflectance properties for many incident and outgoing light directions. We derived parameters for an approximate analytical BRDF that consisted of a Lambertian diffuse component and a modified Phong-like specular component with a Fresnel-like effect.

Subsurface Scattering of Skin

As production progressed it became increasingly clear that realistic skin rendering couldn’t be achieved without subsurface scattering simulation. There are a number of published methods for rendering translucent materials however they are all fairly complex, require large amounts of CPU power and produce somewhat disappointing results. To address this we developed a technique for producing the appearance of subsurface scattering in skin that is computationally inexpensive and fairly easy to implement. The result of the diffuse illumination reflecting off the face in the camera direction is stored in a 2-d light map (see Fig.

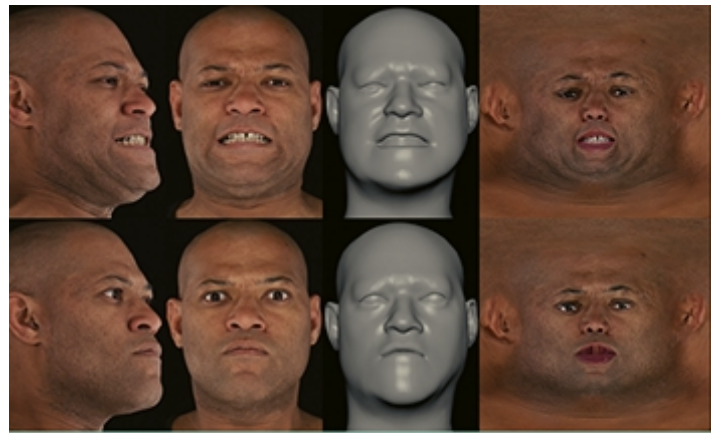
4). We then approximately simulate light diffusion in the image domain. To simulate the different mean free path for different light colors we vary the diffusion parameters for each color channel. For animations the lightmap needs to be computed at every frame, so our technique computes an appropriate lightmap resolution depending on the size of the head in frame. For objects like ears where light can pass directly through, we employed a more traditional ray tracing approach to achieve the desired translucency effect.

Results

The above components are combined with our real world Lighting Reconstruction technology, and a ray tracer such as mental ray to produce the highly realistic synthetic images in Fig. 5 and 6. For comparison Fig. 7 shows a photograph of Keanu Reeves (Neo). The bottom image is a fully virtual frame from *The Matrix Reloaded*.



The first two rows of images in the next group show captured views from two of the five HiDef cameras, the recovered model, and color texture maps for two different moments in time for a performance by Laurence Fishburne. The next row shows a rendering of this performance from novel viewpoints and under different lighting conditions. The last row shows renderings of a performance captured from Hugo Weaving and a frame for *The Matrix Reloaded*.



Acknowledgments

Special thanks and credits go to the co-authors of this work - **Dan Piponi** for work on the optical flow, 3-D reconstruction, and global transform estimation algorithms and for suggesting the detail extraction approach, **Oystein Larsen** for pipeline software design and implementation and performance processing supervision, **J.P. Lewis** for the subsurface scattering algorithm idea, **Christina Tempelaar-Lietz** for additional software design and implementation, **Steve Avoujageli** for writing additional pipeline tools, **Ken Faiman**, **Steve Rembuskos**, and **Mike Dalzell** for demonstrating incredible artistry in processing the performances, **John Llewellyn**, **Ryan Schnizlein**, and **Paul Ryan** for designing the HiDef capture setup, **Brian Freisinger** for modeling and UV mapping the heads, **Rene Garcia** for his paint work, **Haarm-Pieter Duiker** and **Tadao Mihashi** for their invaluable help in the color, lighting, and rendering area, **Matthew Landauer** for contributing to the images-based skin BRDF estimation, the **team at mental images** for making important extensions to the lightmapping feature of mental ray and in general for designing a great renderer, and **John Jack**, **Kim Libreri**, and **John Gaeta** for believing and supporting this unorthodox approach.

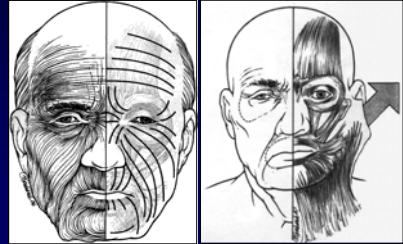


Medical Applications & Behavioral Models

Demetri Terzopoulos

New York University
University of Toronto

Craniofacial Surgery: Face Lift (Rhytidectomy)



Craniofacial Surgery: Cleft Lip and Palate



PreOp



PostOp

Facial Modeling for Surgery Simulation

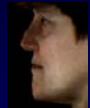
[Girod et al.] [Gross et al.] ...



PreOp



Simulation



PostOp



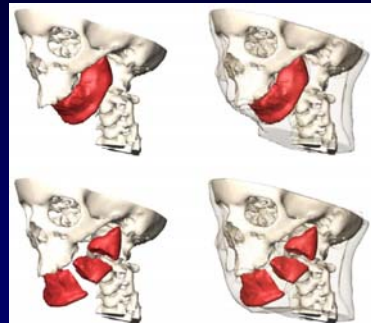
Mandibular Hypoplasia

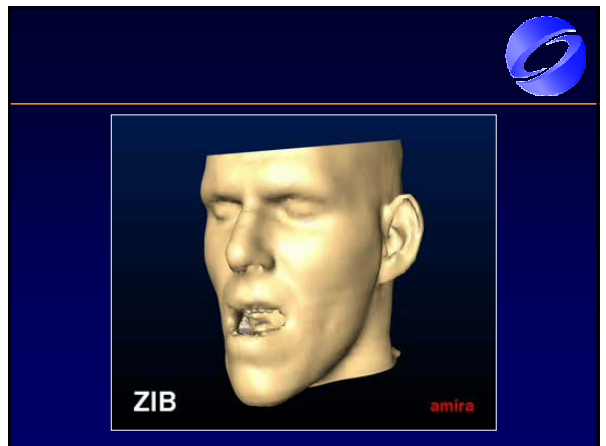
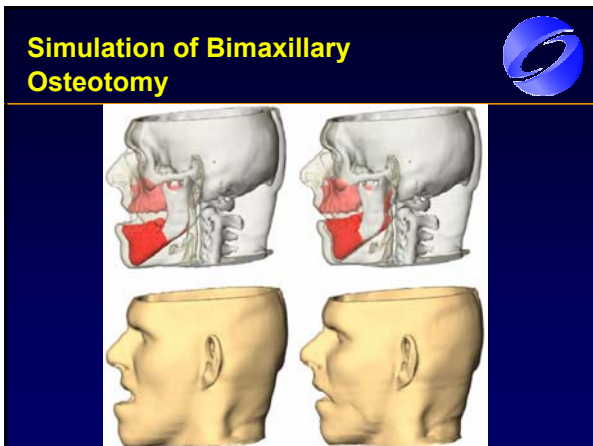
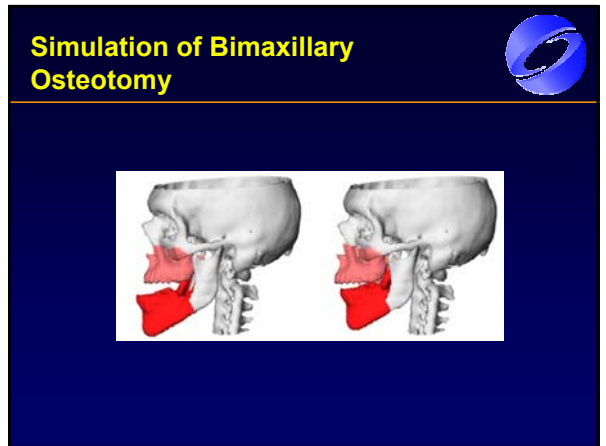
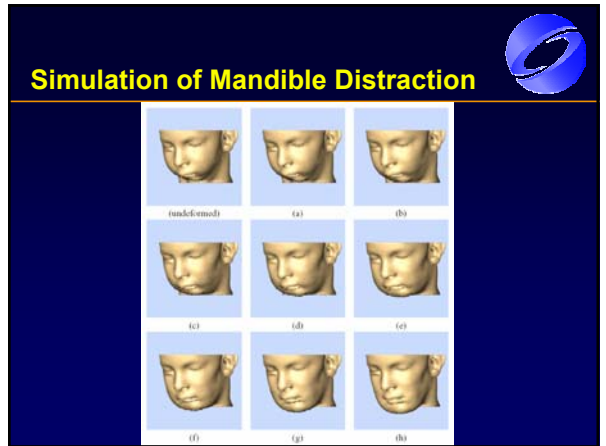
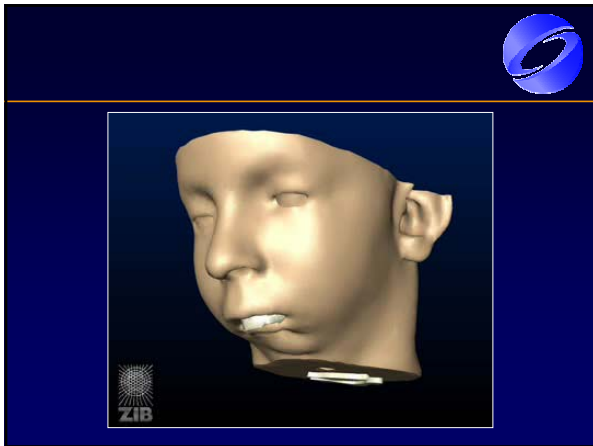


From [Gladilin 2002], Zuse Institute Berlin



Simulation of Mandible Distraction



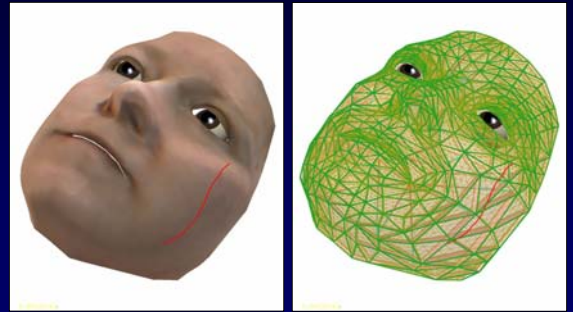


Craniofacial Surgery

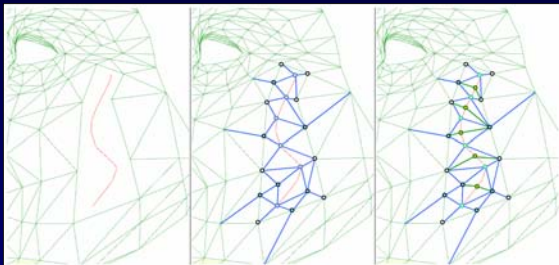
From [Gladilin 2002], Zuse Institute Berlin



Soft Tissue Surgery Simulation: Incision on Facial Mesh



Retriangulation Around Incision



Behavioral Animation of Faces

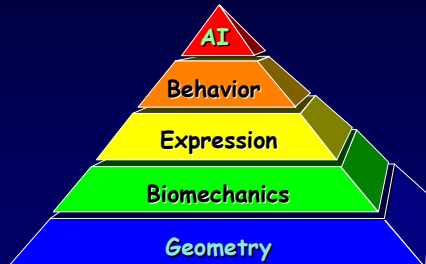
Towards fully autonomous facial models

- Behavioral animation methods
 - An ethological approach
- Part of an “Artificial Life” modeling framework
 - Previously useful for modeling biological systems
 - Plants
 - Animals
 - Humans → faces

Hierarchical Facial Model Structure

From geometry to intelligence

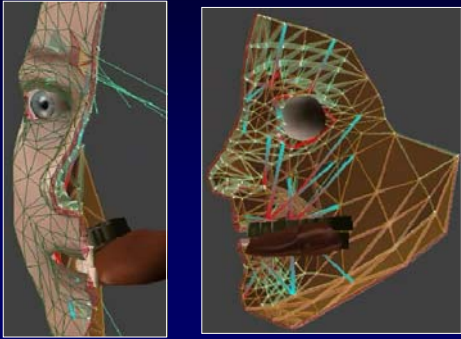
- Cognition
- Muscle Actuation
- Kinematics
- Rendering



Geometry



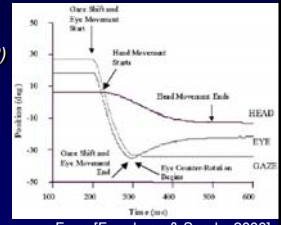
Biomechanics



Perceptual Modeling

Head-Eye-Gaze Kinematics

- Gaze-holding movements
 - Optokinetic reflex (OKR)
 - Vestibulo-ocular reflex (VOR)
- Gaze-shifting movements
 - Saccades
- Fixation movements
 - Slow drift
 - Rapid, low-amplitude tremor
 - Micro-saccades

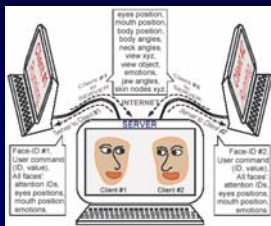


From [Freedman & Sparks 2000]

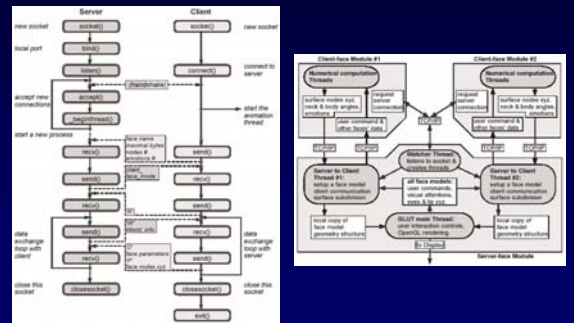
Distributed Face Simulation

Server-client architecture

- Face simulation clients
- Rendering server
- Communication between server and clients
 - Supports perception between faces



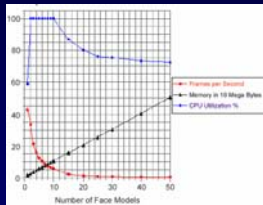
TCP/IP Server-Client Communication



Distributed Simulation Performance Data



Dual PIII 1GHz CPUs, nVIDIA GeForce 3



Autonomous Expressive Behavior



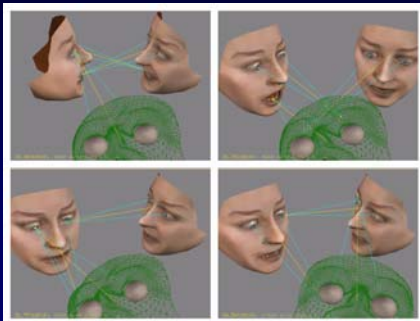
Initial behavioral repertoire

- Attentive behavior routine
- Snubbing behavior routine
- Visual search behavior routine
- Expressive behavior routine
- Mimicking behavior routine
- Interactive behavior routine

Mental state

- "Leader" or "follower"
- Fatigue

Autonomous, Interacting Faces



Autonomous, Interacting Faces



Behavioral Animation of Faces: Parallel, Distributed, and Real-Time

Demetri Terzopoulos^{1,2} and Yuencheng Lee^{2,1}

¹ Courant Institute of Mathematical Sciences, New York University, New York, NY 10003, USA

² Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4, Canada

Abstract

Facial animation has a lengthy history in computer graphics. To date, most efforts have concentrated either on labor-intensive keyframe schemes, on manually animated parameterized methods using FACS-inspired expression control schemes, or on performance-based animation where facial motions are captured from human actors. As an alternative, we propose the fully automated animation of faces using behavioral animation methods. To this end, we employ a physics-based model of the face, which includes synthetic facial soft tissues with embedded muscle actuators. Despite its technical sophistication, this biomechanical face model can nonetheless be simulated in real time on a high-end personal computer. The model incorporates a motor control layer that automatically coordinates eye and head movements, as well as muscle contractions to produce natural expressions. Utilizing principles from artificial life, we augment the synthetic face with a perception model that affords it a visual awareness of its environment, and we provide a sensorimotor response mechanism that links percepts to meaningful actions (i.e., head/eye movement and facial expression). The latter is implemented as an ethologically inspired behavioral repertoire, which includes a rudimentary emotion model. We demonstrate a networked, multi-computer implementation of our behavioral facial animation framework. Each of several faces is computed in real time by a separate server PC which transmits its simulation results to a client PC dedicated to rendering the animated faces in a common virtual space. Performing the appropriate head/eye/face movements, the autonomous faces look at one another and respond in a natural manner to each other's expressions.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Three-Dimensional Graphics and Realism]: Animation I.3.5 [Computational Geometry and Object Modeling]: Physically based modeling

1. Introduction

Facial modeling and animation has a lengthy history in computer graphics. The area was pioneered over thirty years ago by Frederic Parke at the University of Utah [Par72]. A survey of the field is presented in the volume [PW96]. Briefly, realistic facial models have progressed from keyframe (blend-shape) models [Par72], to parameterized geometric models [Par74], to muscle-based geometric models [Wat87], to anatomically-based biomechanical models [TW90, LTW95, KHYS02]. In parallel with the model-based approaches, a variety of successful facial data driven technologies have recently been developed for facial modeling and animation [WT91, GGW*98, PHL*98, BV99, BBPV03]. To date, most efforts in production facial animation have con-

centrated either on labor-intensive (blendshape) keyframe schemes often involving manually-animated parameterized schemes [Pix88], or on performance-based animation where facial motions are captured from human actors [Wil90, fac04]. With regard to facial motion capture, it remains a challenge to modify the captured facial motions.

As an alternative, it would be desirable to have a fully automated face/head model that can *synthesize* realistic facial animation. Such a model would be of value both for the production animation and especially in the interactive computer games industries. Ultimately, this model should be an *intelligent* one, which would possess both nonverbal and verbal facial communications skills and would be able to interact autonomously in a virtual environment with other such intelligent face/head models. To this end, we have been inspired

by the Artificial Life framework advocated by Terzopoulos and his group [Ter99], which prescribes biomechanical, perceptual, behavioral and, ultimately, learning and cognitive modeling layers. In this paper, we begin to tackle the challenge of applying this framework to the modeling and animation of the human face.

1.1. Background and Contributions

As an initial step, we propose the goal of fully automated facial animation synthesis through the use of behavioral animation methods. We achieve this goal through an ethologically inspired behavioral repertoire for human faces, which includes a rudimentary emotion model. Behavioral animation was introduced to computer graphics by Reynolds in his seminal work on “booids” [Rey87]. It was further developed and applied to artificial animals by Tu and Terzopoulos [TT94]. In the context of character animation, Cassell et al. [CVB01] presented a behavior toolkit which converts from typed sentences to synthesized speech and synchronized nonverbal behaviors, including gestures and some facial expressions.

Behavior ties perception to action in meaningful ways. Our approach is focused on behavior-controlled dynamics of all aspects of the human head and face. We employ a biomechanical model of the face, which includes synthetic facial soft tissues with embedded muscle actuators. Our model is a significantly improved version of the one published by Lee et al. [LTW95]. Despite its technical sophistication, our biomechanical face model has been optimized such that it may be simulated in real time on a high-end personal computer.

An important component of our work is the simulation of head-eye movements. The role of eye movements in conversational characters is discussed by Vertegaal et al. [VSDVN01] who present interesting empirical observations about gaze control during conversations. Lee et al. [LBB02] describe a statistical model that, from tracked eye movements in video, can synthesize believable ocular motion for an animated face. Our model incorporates a novel motor control layer that automatically coordinates synthetic eye and head movements, as well as muscle contractions to produce natural expressions.

Our work builds a repertoire of facial behaviors that are driven by perception. Utilizing principles from artificial life, we augment the synthetic face with a perception model that affords it a visual awareness of its environment, and we provide a sensorimotor response mechanism that links percepts to sensible reactions (i.e., head/eye movement and facial expression). Active, foveated perceptual modeling for virtual humans using computer vision techniques was discussed by Terzopoulos and Rabie (see, e.g., [Ter99]). Although the use of computer vision techniques may be the ultimate goal of our work, for the sake of efficiency we currently employ a “synthetic vision” scheme [Rey87, RMTT90, TT94].

As a final contribution, we demonstrate a networked, multi-computer implementation of our behavioral facial animation framework. Each of several faces is computed in real time by a separate server PC which transmits its simulation results to a client PC dedicated to rendering the animated faces in a common virtual space. Performing the appropriate head/eye/face movements, the autonomous faces look at one another and respond naturally to each other’s expressions in a multiway nonverbal communication scenario.

1.2. Overview

The remainder of this paper is organized as follows: Section 2 summarizes our human face model. Section 3 describes how this model is simulated in real time in parallel on multiple processors. In Section 4 we detail the distributed simulation of multiple faces on multiple networked computers, as well as the mechanism for exchanging perceptual information between multiple heads. Section 5 presents the head/eye movement coordination model. Section 6 develops our behavioral model for human faces and presents an experiment demonstrating the autonomous behavioral interaction among multiple heads. Section 7 concludes the paper and presents an outlook on our future work.

2. A Functional Facial Model

We have developed a sophisticated, functional model of the human face and head that is efficient enough to run at interactive rates on high-end PCs. Conceptually, the model decomposes hierarchically into several levels of abstraction, which represent essential aspects related to the psychology of human behavior and facial expression, the anatomy of facial muscle structures, the histology and biomechanics of facial tissues, facial geometry and skeletal kinematics, and graphical visualization:

1. *Behavior.* At the highest level of abstraction, the synthetic face model has a repertoire of autonomous behaviors, including reactive and intentional expressive behaviors with coordinated head/eye movements.
2. *Expression.* At the next level, the face model executes individual expression commands. It can synthesize any of the six primary expressions (joy, sadness, anger, fear, surprise and disgust) within a specific duration and degree of emphasis away from the neutral face. A muscle control process based on Ekman and Friesen’s FACS [EF86] translates expression instructions into the appropriately coordinated activation of actuator groups in the soft-tissue model. This coordination offers a semantically rich set of control parameters which reflect the natural constraints of real faces.
3. *Muscle Actuation.* As in real faces, muscles comprise the basic actuation mechanism of the face model. Each muscle submodel consists of a bundle of muscle fibers. The action of the contractile fibers is modeled in terms of a

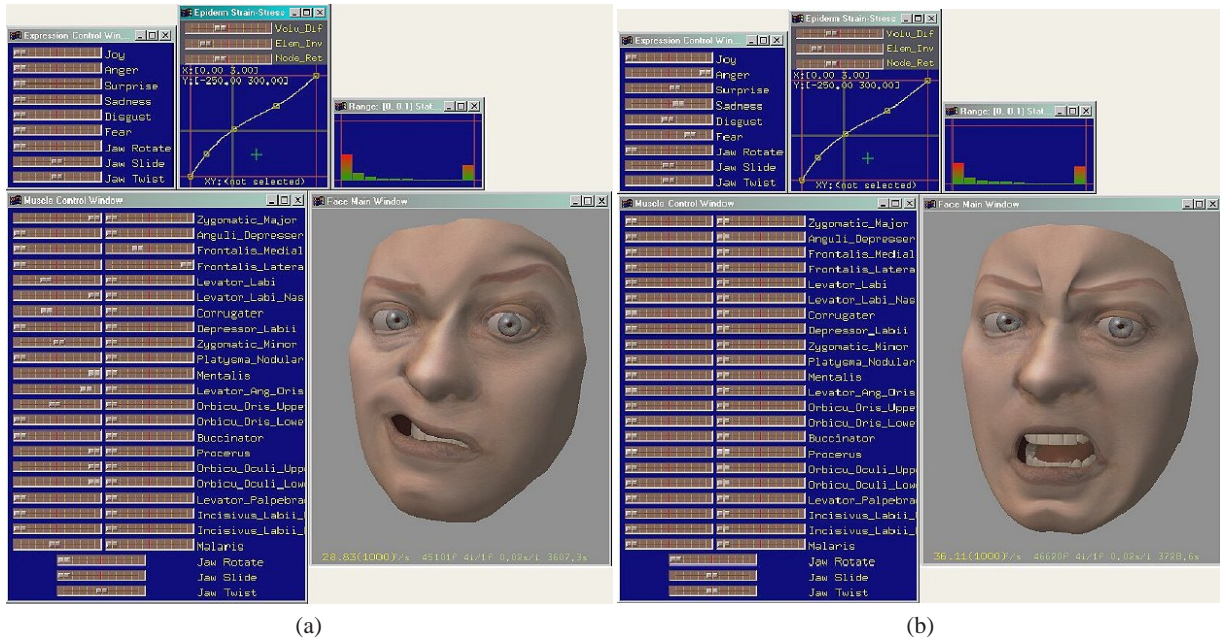


Figure 1: Panels for manually adjusting expressions, muscle contractions, and stress-strain curves at the expression, muscle actuation, and tissue biomechanics levels of the facial model. (a) Adjusting the muscle panel. (b) Adjusting the expression panel.

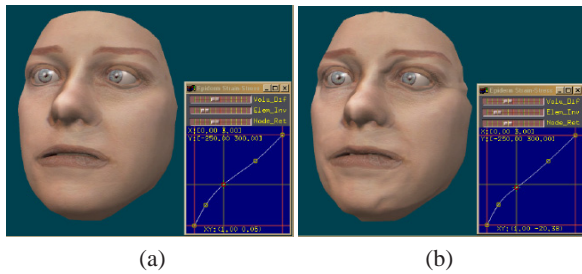


Figure 2: Skin model with interactively adjustable stress-strain curves. (a) Normal epidermal/dermal stress-strain curve. (b) Curve adjusted to simulate looser skin, resulting in an aged appearance.

force profile longitudinally along the central vector of the muscle and laterally from the vector (see [LTW95] for the details). In our model, there are 42 muscles of facial expression in the synthetic face, which augments the musculature of its predecessor model described in [LTW95].

4. *Biomechanics.* When muscles contract, they displace their points of attachment in the facial tissue or the articulated jaw. The face model incorporates a physical approximation to human facial tissue, a nonhomogeneous and nonisotropic layered structure consisting of the epidermis, dermis, subcutaneous fatty tissue, fascia, and muscle layers. The tissue model [LTW95] is a lattice of point

masses connected by nonlinear viscoelastic springs, arranged as layered prismatic elements that are constrained to slide over an impenetrable skull substructure. Large-scale synthetic tissue deformations are numerically simulated by continuously computing the response of the assembly of volume-preserving elements to the stresses induced by activated muscle fibers.

5. *Geometry/Kinematics.* 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 synthetic tissue deformations distort the neutral geometry into an expressive geometry. The epidermal display model is a smoothly-curved subdivision surface [DKT98] (in our case a Loop subdivision surface [Loo87]) that deforms in accordance with the simulated tissue elements. In addition, the complete head model includes functional subsidiary models, including a skull with articulated jaw, teeth, tongue/palate, eyes, and eyelids.
6. *Rendering.* After each simulation time step, standard visualization algorithms implemented in the PC OpenGL graphics pipeline render the deforming facial geometry in accordance with viewpoint, light source, and skin reflectance (texture) information to produce the lowest level representation in the modeling hierarchy, a continuous stream of facial images.

The hierarchical structure of the model appropriately encapsulates the complexities of the underlying representa-

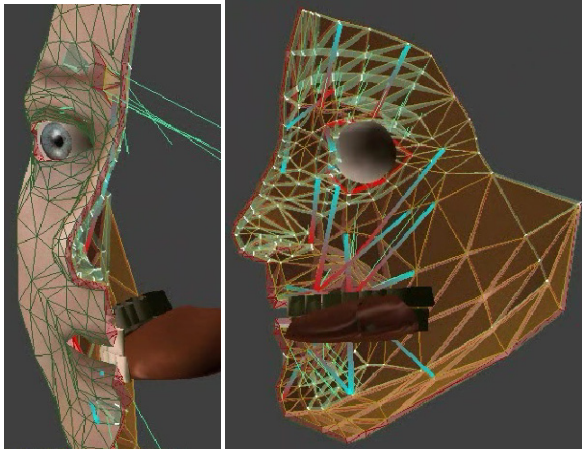


Figure 3: Cross section through the biomechanical face model, showing multilayer skin and underlying muscle actuators (represented as red-blue vectors). The epidermal triangles indicate the triangular prism element mesh.

tions, relegating the details of their simulation to automatic procedures.

3. Parallel, Real-Time Simulation of the Face Model

The biomechanical simulation of our face model yields realistic tissue deformations, but it is computationally expensive relative to conventional geometric facial models. We have made significant effort to make our model computable in real time, as we describe in this section.

3.1. Biomechanical Soft Tissue Model

The biomechanical soft tissue model has five layers. Four layers—epidermis, dermis, sub-cutaneous fatty tissue, and fascia—comprise the skin, and the fifth consists of the muscles of facial expression [FH91]. In accordance with the structure of real skin, and following [LTW95], we have designed a synthetic tissue model composed of the triangular prism elements which match the triangles in the adapted facial mesh. The elements are constructed from lumped masses interconnected by uniaxial, viscoelastic units. Each uniaxial unit comprises a spring and damper connected in parallel. The springs have associated stress-strain curves which can be manually adjusted from interactive panels.

The individual muscle model is the same as that in [LTW95]. Fig. 3 shows the face model in cross-section, revealing the muscle actuators underlying the multilayer, biomechanical skin model. Fig. 1 illustrates various interactive panels that a user can employ to make manual adjustments at the expression, muscle actuation, and biomechanics levels of the model. As a point of interest, Fig. 2

Number of face models	Number of threads per face model	Frame rates per second	Memory in MB	CPU Utilization %
1	2	50.80	17.4	93.4
1	1	42.90	17.3	58.6
2	1	33.58	27.2	100.0
3	1	21.48	37.1	100.0
4	1	16.20	47.0	100.0
5	1	12.45	57.9	100.0
6	1	10.31	66.9	100.0
7	1	8.59	76.9	100.0
8	1	7.44	86.8	100.0
9	1	6.50	96.8	100.0
10	1	5.76	106.8	100.0
15	1	2.40	156.5	86.9
20	1	1.42	206.0	80.0
25	1	1.03	255.7	76.2
30	1	0.80	305.2	75.5
40	1	0.56	404.5	73.3
50	1	0.43	503.7	72.4

Table 1: Simulation rates of the physics based face model with 1078 nodes, 7398 springs, 42 muscles, 1042 elements, and 1042 facets using 4 iterations of numerical computation per rendered frame (with base level surface subdivision at each frame) on a dual Intel Pentium III 1 GHz CPU system with 1 GB of PCI33 memory and an nVIDIA GeForce3 AGP2X graphics card with 64 MB of graphics memory.

shows two different settings of the epidermal/dermal stress-strain curves, the first is normal, while the second has a negative residual strain which simulates looser skin, giving the face an aged appearance. Note, however, that although these various interactive panels are available, it is unnecessary to make any adjustments whatsoever through them during normal operation of the facial model, as automatic controllers at the behavior, expression, muscle actuation, and biomechanics modeling levels control the various parameters.

3.2. Parallel Simulation

In general, a physics-based simulation model makes intensive CPU usage for numerical computations to simulate dynamics. The biomechanical tissue model is simulated numerically using an explicit Euler time-integration method. As described in [LTW95], the method computes the velocities and positions of each nodal mass at the next time step from quantities that are computed at the current time step. This enables us to perform the numerical simulation of the tissue model in parallel. Parallelization is achieved by evenly

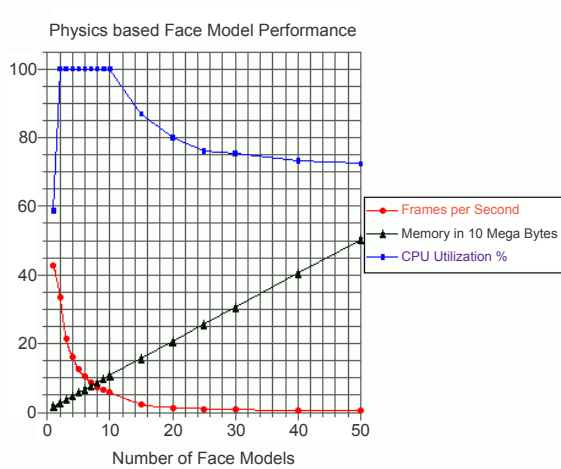


Figure 4: Facial simulation performance.

distributing calculations at each time step to all available processors using multiple execution threads. This increases the simulation speed, enabling our system to animate facial expressions at real-time rates on a dual Intel Pentium-III 1 GHz CPU computer workstation with an nVIDIA GeForce3 graphics card. Table 1 and Fig. 4 document the performance figures on this system. Note that it can simulate and render no more than two face models simultaneously in real time. We have also verified that our approach enables the face model to evenly distribute its numerical computations on a quad Intel Xeon 2.4 GHz CPU system.

We conclude that in order to simulate and render a greater number of faces in real time, we must resort to distributed facial simulation on multiple computers.

4. Distributed Face Simulation and Rendering

Our approach to distributed facial simulation is to simulate multiple instances of the face models on multiple computers (face simulation clients) networked to a dedicated graphics workstation (the rendering server) whose task is to render the updated geometric face models together in a common virtual space. For sensory perception, any client can sense the state of any other client only via the server. Fig. 5 illustrates the architecture and perceptual data flow. In accordance with our goal to separate the numerical computation from the graphics rendering, Table 2 compares the responsibility of the rendering server and simulation client when sharing the single computer simulation/rendering workload.

To maximize flexibility across different computing platforms, we decided to use the TCP/IP (Transmission Control Protocol / Internet Protocol) standard to support the distributed computation of our face models. An IP comprises a packet header which contains the originating address and

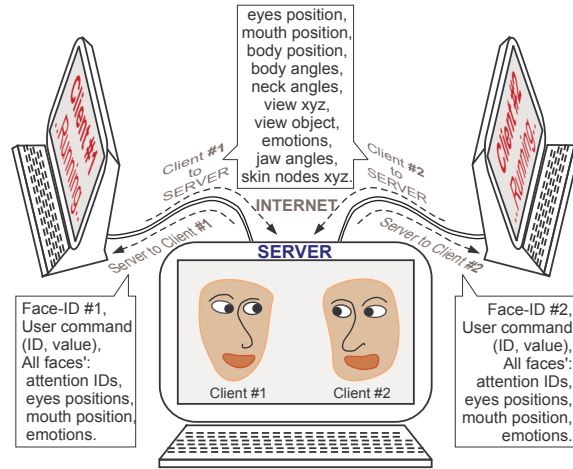


Figure 5: Data exchange in the distributed face simulation system between the server and two clients.

	Single Computer	Face-Server	Face-Client
Acting Alone	+	+	-
Direct User Interaction	+	+	-
3D Skin Rendering	+	+	-
3D Tissue Rendering	+	-	-
3D Force Rendering	+	-	-
Surface Subdivision	+ / -	+ / -	-
Numerical Computation	+	-	+
Control Events Handling	+	-	+
Networking	-	+	+
Peers Interaction	-	-	+

Table 2: Comparison between our face model modules.

the destination address, and a packet body consisting of data. The TCP is a connection-oriented 3-way handshaking communication protocol using sync, sync/ack, and ack. With TCP/IP, both the sender and the receiver can be synchronized to ensure the successful arrival of each IP using its address information of host names and port numbers. Figure 6 outlines the 3-way handshaking scheme of our client/server processes using TCP/IP. The face rendering computer acts as the server. Face simulation clients connect to this server via the internet. Once the server accepts a connection request from a client, it continuously sends and receives data with the client until this client-server connection is closed either by the client or server.

After the rendering server has received an initial connection request from a new simulation client, it will start a new thread to handle all communications with this new client (see Figure 7). This new thread will wait for the client to send the

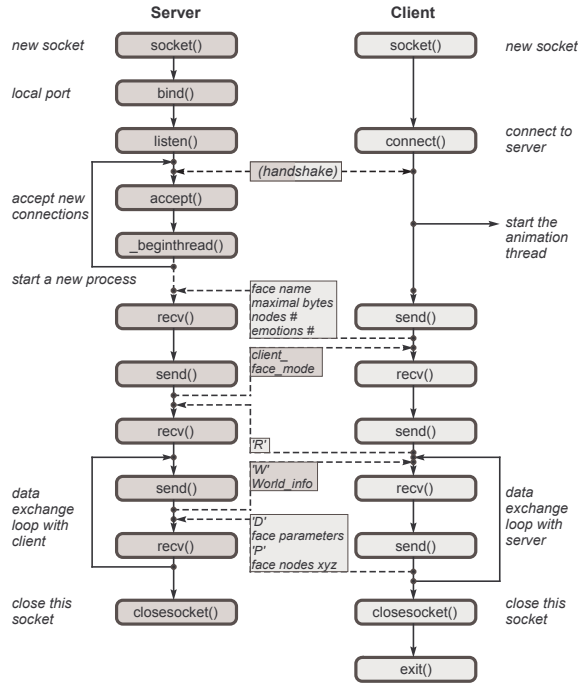


Figure 6: Connection scheme of TCP/IP sockets and data exchange between a faces rendering server and face simulation client.

name of the face model, the number of nodes in the face model, and the number of different emotion templates in the face model. After it receives this information, it will use the name of the face model to load the geometry structure definition files of this face model from its local storage into its system memory. After this face model is successfully loaded into memory, its number of nodes and number of emotion templates will be verified with the information sent from the client.

On the other hand, after a simulation client has received the handshaking acknowledgement from the rendering server, it will send the server the aforementioned information and start an animation thread to handle all the numerical calculations within the client. It will also handle all the communications with the server.

For our current face model of 539 surface nodes, at each rendering frame, each simulation client will send to the rendering 539 x, y, z floating point values, approximately 30 more geometry related floating point values, and 6 emotion related floating point values. The size of the total communicated data is roughly 8600 bytes. On the other hand, the server will send to the clients 1 integer value to identify the client, 9 geometry related floating point values, and 6 emotion related floating point values, multiplied by the number of active clients (See Fig. 5).

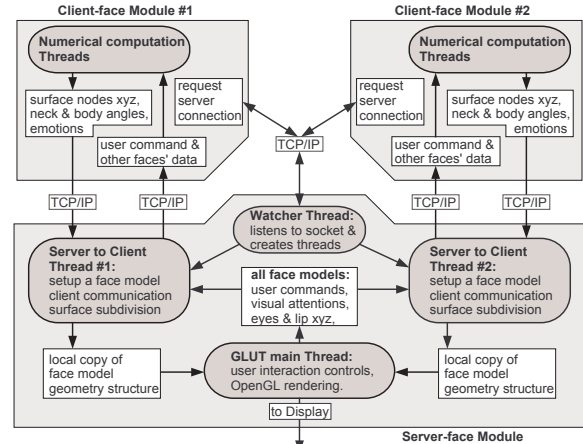


Figure 7: The multi-threads scheme within the server and the clients of our distributed face models system

4.1. Sensory Perception Between Simulated Faces

With our data exchange loop between the server and the clients in Fig. 6, at every simulation step each simulated face can sense perceptually relevant information about other simulated faces that share the same space.

The perceptual information available includes the position of a face and the locations of its relevant parts, such as the eyes and mouth, as well as the emotional state of the face. The emotional state is represented as a point in “expression space”, a 6-dimensional unit hypercube, each of whose dimensions is associated with a primary expression (joy, sadness, anger, fear, surprise and disgust). The neutral expression is at the origin of expression space. For a symbolic interpretation of expression, the continuous expression space is partitioned into a number of subregions that define qualitative “emotion templates”, which are recognizable by the observer.

5. Eye-Head Coordination

The oculomotor system, whose output is the position of the eyes relative to the head, has been the subject of much research (see, e.g., the treatise [Car88]), because it is a closed, well-defined system that is amenable to precise, quantitative study. The direction of the eye in space is called the *gaze*. There are three types of eye movements:

- *Gaze-holding movements.* Because gaze is the sum of head position and eye position, these eye movements compensate for the movement of the head (and body) in order to maximize the stability of the retinal image. Gaze-holding movements are either optokinetic reflexes (OKR), which are driven by retinal image motion (a.k.a. optical flow), or vestibulo-ocular reflexes (VOR), which are driven by the balance organs in the inner ear.

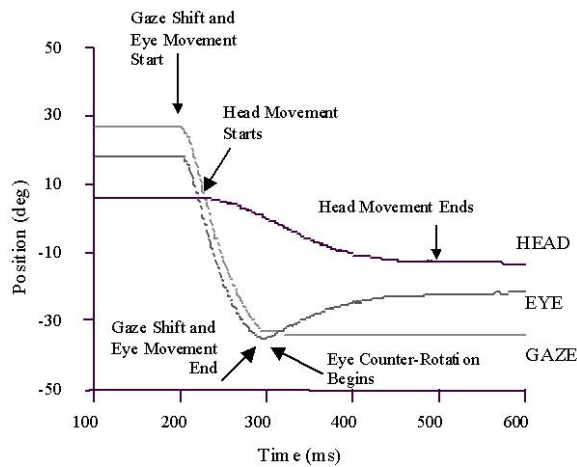


Figure 8: Typical head/eye/gaze kinematics (from [FS00]). Head, eye, and gaze position are plotted as functions of time during a 60° gaze shift composed of coordinated movements of the eyes and head. The head contributed approximately 10° to the overall change in gaze direction during this movement. The remaining 50° of the gaze shift were accomplished by the saccadic eye movement. Note that when the line of sight achieves the desired gaze, the head continues to move, but gaze remains constant due to equal and opposite eye counter-rotation mediated through the VOR.

- *Gaze-shifting movements.* Human vision is foveated. The foveal region, which spans roughly 2 degrees of visual arc, is specialized for high-acuity, color vision. To see an object clearly, gaze-shifting movements deliberately shift, directing the eye to the target. Since the resulting eye motion disrupts vision, these movements are as fast as possible and are called *saccades*. As a target object moves closer, the two eyes must converge onto the target; these are called *vergence* movements.
- *Fixation movements.* Even when fixating a stationary object, the eyes are making continual micro-movements of three types: Slow drift, rapid small-amplitude tremor, and micro-saccades that recover the gaze when the drift has moved it too far off target.

In view of the significantly greater mass of the head relative to the eye, head dynamics are much more sluggish than eye dynamics. As is documented in [Car88], when a subject voluntarily moves the head and eye(s) to acquire an off-axis visual target in the horizontal plane, the eye movement consists of an initial saccade in the direction of the head movement, presumably to facilitate rapid search and visual target localization, followed by a slower return to orbital center, which compensates for the remaining head movement. During target acquisition, head velocity is normally correlated with the amplitude of the visual target offset.

Typical head/eye/gaze kinematics are plotted in Fig. 8. We have implemented a head-eye coordination behavior that accounts for the observed phenomena as reported in the literature. Our scheme uses exponential functions with different time constants for the head and eye to approximate the empirically observed kinematic curves shown in the figure. The model which supports gaze-shifting and gaze-holding functionalities, implements the head-eye motor control layer of our synthetic face. In order to prevent the head from remaining absolutely still in an unnatural manner, we perturb the head rotation angles with some low-level Perlin noise.

6. Autonomous Expressive Behavior

As stated earlier, behavior ties perception to action. Our relatively modest goal in this initial effort is to demonstrate autonomous nonverbal behavior of a basic sort. Following the approach of [TT94], we have implemented the rudiments of a behavioral subsystem for our synthetic face model that comprises mental state variables and a repertoire of behavioral routines mediated by an action selection mechanism. The thus far rather limited repertoire includes the following behavior routines, which are ordered in terms of increasing complexity:

1. *Attentive Behavior Routine.* The face will gaze at a specific face.
2. *Snubbing Behavior Routine.* The face will not gaze at a specific face or faces.
3. *Visual Search Behavior Routine.* The autonomous face will visually scan nearby faces to acquire relevant perceptual information about them.
4. *Expressive Behavior Routine.* The face will attempt to lead an expressive exchange by deliberately performing a sequence of random expressions of some random magnitude and duration.
5. *Mimicking Behavior Routine.* The face will attempt to follow an expressive exchange by sensing the expression of a target face and mimicking that expression. This makes use of attentive behavior.
6. *Interactive Behavior Routine.* The face will take turns engaging one or more other faces in an expressive interchange. This behavior potentially makes use of all the other behaviors.

The mental state so far contains a single variable that determines whether a face will behave as a “leader” or a “follower”. The action selection mechanism includes timers that monitor how long a particular behavior is engaged. The intention generator is programmed not to sustain any particular behavior for too long a time, thus exhibiting a behavior “fatigue” effect.

6.1. Experiment

Fig. 9 illustrates our real-time, self-animating faces engaged in a 3-way interchange involving expression mimicking. In

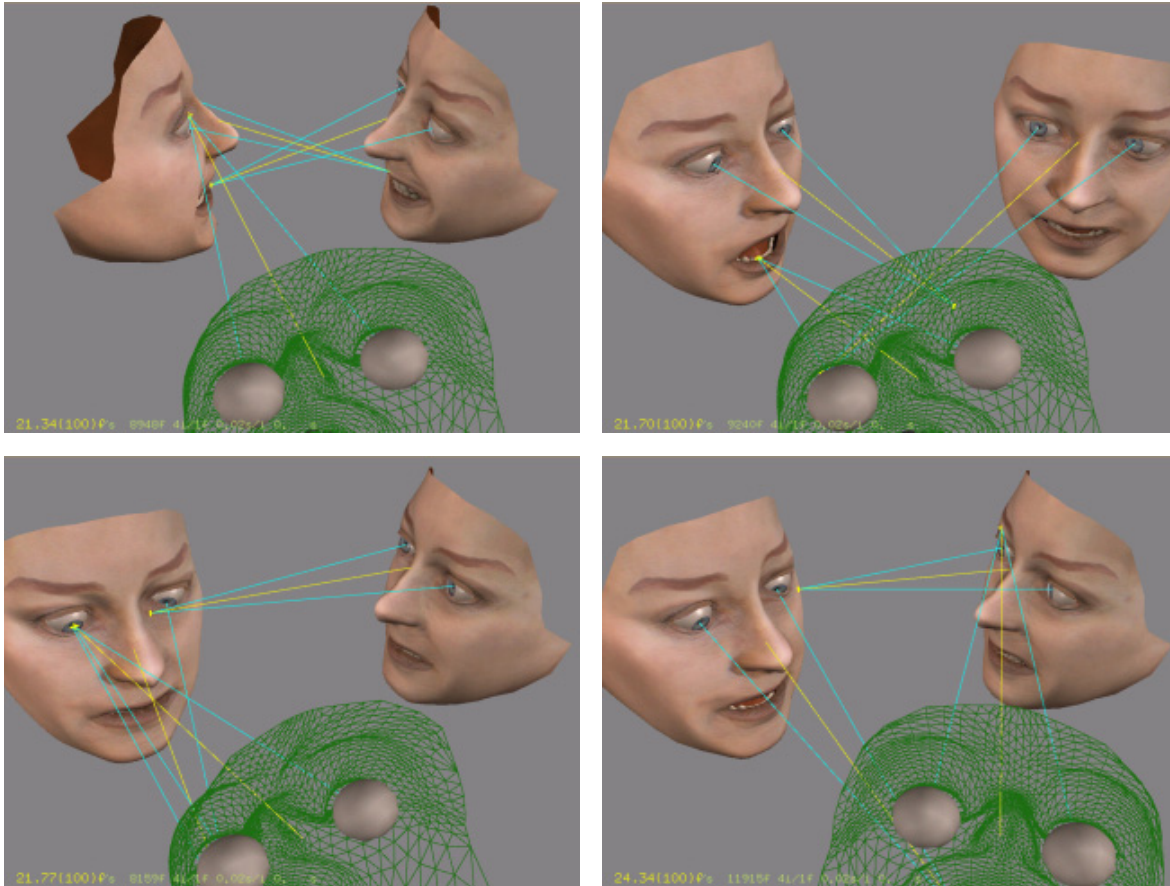


Figure 9: Autonomous behavioral-based interaction between three face simulation clients

this experiment, the first face simulation client that establishes contact with the rendering server will behave as a leader and engage Visual Search Behavior until it detects the presence of another face, then switch to Attentive Behavior and Expressive Behavior. Other face clients joining the interactive exchange behave as followers, engaging in Visual Search Behavior and Attentive Behavior with the leader when the leader attends to them. Once a follower has the leader's attention the follower will engage in Mimicking Behavior. Eventually, behavior fatigue will compel the follower to disengage the leader and attend to a different face. When confronted by more than one face, the leader engages in Interactive Behavior with the various faces. This autonomous behavioral animation results in a highly dynamic exchange, with the server acting as a medium for the transmission of perceptual information between the multiple face simulation clients.

7. Conclusion

We have introduced a behavioral animation approach for faces. Although rudimentary, the ethologically inspired model can support basic non-verbal, expressive behaviors among multiple interacting faces. This capability was demonstrated using a biomechanical model of the face exhibiting muscle-induced dynamic expressions mediated by a FACS muscle control and coordination layer. A subsidiary kinematic model provides the requisite head and eye movements in accordance with empirically measured head/eye curves reported in the literature. Finally, the self-animating, multi-head/face simulation is computed in real time in a distributed manner on multiple, dedicated face simulation clients networked to a rendering server. Using TCP/IP, the clients supply the server with dynamically updated facial geometry data for rendering and exchange the perceptual information needed to sustain the interactive behavior model.

In future work, we plan to implement a dynamic model of head/eye movements to replace the current kinematic one and to expand the breadth of the currently limited behav-

ioral repertoire. As was indicated in the introduction, our ultimate goal is to implement learning and cognitive modeling layers, thereby realizing an intelligent model of the human face/head.

References

- [BBPV03] BLANZ V., BASSO C., POGGIO T., VETTER T.: Reanimating faces in images and video. In *Proceedings of the EUROGRAPHICS 2003 Conference* (Granada, Spain, 2003), Brunet P., Fellner D., (Eds.). 1
- [BV99] BLANZ V., VETTER T.: A morphable model for the synthesis of 3D faces. In *Computer Graphics (SIGGRAPH '99 Proceedings)* (Los Angeles, CA, 1999), pp. 187–194. 1
- [Car88] CARPENTER R.: *Movements of the Eyes*, 2nd ed. Pion, London, 1988. 6, 7
- [CVB01] CASSELL J., VILHJÁLMSSON H., BICKMORE T.: Beat: The behavior expression animation toolkit. In *Proceedings of SIGGRAPH '01* (Los Angeles, CA, August 2001), pp. 477–486. 2
- [DKT98] DEROSE T., KASS M., TRUONG T.: Subdivision surfaces in character animation. In *Computer Graphics (SIGGRAPH '98 Proceedings)* (Orlando, FL, 1998), pp. 85–94. 3
- [EF86] EKMAN P., FRIESEN W.: *Manual for the Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, 1986. 2
- [fac04] FACE2FACE I.: The art of performance animation, 2004. www.f2f-inc.com. 1
- [FH91] FRICK, HANS: *Human Anatomy*, vol. 1. Thieme Medical Publishers, Stuttgart, 1991. 4
- [FS00] FREEDMAN E. G., SPARKS D. L.: Coordination of the eyes and head: Movement kinematics. *Experimental Brain Research* 131, 1 (March 2000), 22–32. 7
- [GGW*98] GUENTER B., GRIMM C., WOOD D., MALVAR H., PIGHIN F.: Making faces. In *Computer Graphics (SIGGRAPH '98 Proceedings)* (Orlando, FL, 1998), pp. 55–66. 1
- [KHYS02] KAHLER K., HABER J., YAMAUCHI H., SEIDEL H.-P.: Head shop: Generating animated head models with anatomical structure. In *Proceedings ACM SIGGRAPH Symposium on Computer Animation (SCA 2002)* (July 2002), pp. 55–64. 1
- [LBB02] LEE S. P., BADLER J. B., BADLER N. I.: Eyes alive. *ACM Transactions on Graphics* 3, 21 (2002), 637–644. *Proceedings of ACM SIGGRAPH 2002 Conference*. 2
- [Loo87] LOOP C.: *Smooth Subdivision Surfaces Based on Triangles*. PhD thesis, University of Utah, August 1987. 3
- [LTW95] LEE Y., TERZOPOULOS D., WATERS K.: Realistic modeling for facial animation. In *Computer Graphics (SIGGRAPH '95 Proceedings)* (July 1995), vol. 29, pp. 55–62. 1, 2, 3, 4
- [Par72] PARKE F.: Computer generated animation of faces. In *ACM National Conference* (1972), ACM, pp. 451–457. 1
- [Par74] PARKE F.: *A Parametric Model for Human Faces*. PhD thesis, University of Utah, Salt Lake City, Utah, 1974. 1
- [PHL*98] PIGHIN F., HECKER J., LISCHINSKI D., SZELISKI R., SALESIN D. H.: Synthesizing realistic facial expressions from photographs. In *Computer Graphics (SIGGRAPH '98 Proceedings)* (Orlando, FL, 1998), pp. 75–84. 1
- [Pix88] PIXAR: Tin toy. *Computer Animation*, 1988. 1
- [PW96] PARKE F., WATERS K.: *Computer Facial Animation*, 1 ed. A. K. Peters, Ltd., 1996. 1
- [Rey87] REYNOLDS C. W.: Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics* 21, 4 (July 1987), 25–34. 2
- [RMTT90] RENAULT O., MAGNENAT-THALMANN N., THALMANN D.: A vision-based approach to behavioural animation. *Visualization and Computer Animation 1* (1990), 18–21. 2
- [Ter99] TERZOPOULOS D.: Artificial life for computer graphics. *Communications of the ACM* 42, 8 (1999), 32–42. 2
- [TT94] TU X., TERZOPOULOS D.: Artificial fishes: Physics, locomotion, perception, behavior. In *Computer Graphics Proceedings, Annual Conference Series* (July 1994), Proc. SIGGRAPH '94 (Orlando, FL), ACM SIGGRAPH, pp. 43–50. 2, 7
- [TW90] TERZOPOULOS D., WATERS K.: Physically-based facial modelling, analysis, and animation. *The Journal of Visualization and Computer Animation 1*, 2 (1990), 73–80. 1
- [VSDVN01] VERTEGAAL R., SLAGTER R., DER VEER G. V., NIJHOLT A.: Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In *Proceedings*

- of ACM CHI 2001 Conference (Seattle, WA, March 2001), pp. 301–308. [2](#)
- [Wat87] WATERS K.: A muscle model for animating three-dimensional facial expression. *Computer Graphics (Proceedings ACM SIGGRAPH 87)* 22, 4 (1987), 17–24. [1](#)
- [Wil90] WILLIAMS L.: Performance-driven facial animation. In *SIGGRAPH 24* (1990), ACM Computer Graphics, pp. 235–242. [1](#)
- [WT91] WATERS K., TERZOPOULOS D.: Modeling and animating faces using scanned data. *The Journal of Visualization and Computer Animation* 2, 4 (1991), 123–128. [1](#)

4 References

This section contains a list of publications, which are useful to learn more about facial modeling and animation. The selection of these publications represents the authors' point of view and might not be complete.

References

- [1] Irene Albrecht, Jörg Haber, Kolja Kähler, Marc Schröder, and Hans-Peter Seidel. “May I talk to you? : -)” — Facial Animation from Text. In Sabine Coquillart, Heung-Yeung Shum, and Shi-Min Hu, editors, *Proceedings of the Tenth Pacific Conference on Computer Graphics and Applications (Pacific Graphics 2002)*, pages 77–86. IEEE Computer Society, October 2002.
- [2] Irene Albrecht, Jörg Haber, and Hans-Peter Seidel. Automatic Generation of Non-Verbal Facial Expressions from Speech. In *Proceedings of Computer Graphics International 2002 (CGI 2002)*, pages 283–293, July 2002.
- [3] Irene Albrecht, Jörg Haber, and Hans-Peter Seidel. Speech Synchronization for Physics-based Facial Animation. In *Proceedings of WSCG 2002*, pages 9–16, February 2002.
- [4] J. Allen, M. S. Hunnicutt, and D. Klatt. *From text to speech: The MITalk system*. Cambridge University Press, Cambridge, MA, 1987.
- [5] Ken-ichi Anjyo, Yoshiaki Usami, and Tsuneya Kurihara. A Simple Method for Extracting the Natural Beauty of Hair. In Edwin E. Catmull, editor, *Computer Graphics (SIGGRAPH '92 Conference Proceedings)*, volume 26, pages 111–120. ACM SIGGRAPH, July 1992.
- [6] S. Arridge, J. P. Moss, A. D. Linney, and D. James. Three-dimensional digitization of the face and skull. *Journal of Maxial Facial Surgery*, 13:1396–143, 1985.
- [7] B. Baumgart. A polyhedron representation for computer vision. In *AFIPS National Conference Proceedings*, volume 44, pages 589–596, 1975.
- [8] P. Bergeron and P. Lachapelle. Controlling facial expressions and body movements. In *Advanced Computer Animation*, volume 2 of *SIGGRAPH '85 Tutorials*, pages 61–79. ACM, New York, 1985.
- [9] E. Bizzi, W. Chapple, and N. Hogan. Mechanical properties of muscles. *TINS*, 5(11):395–398, November 1982.
- [10] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*.
- [11] Volker Blanz, Curzio Basso, Tomaso Poggio, and Thomas Vetter. Reanimating Faces in Images and Video. In Pere Brunet and Dieter Fellner, editors, *Computer Graphics Forum (Proceedings of Eurographics 2003)*, volume 22, pages 641–650, September 2003.
- [12] Volker Blanz and Thomas Vetter. A Morphable Model for the Synthesis of 3D Faces. In Alyn Rockwood, editor, *Computer Graphics (SIGGRAPH '99 Conference Proceedings)*, pages 187–194. ACM SIGGRAPH, August 1999.
- [13] J. D. Boucher and P. Ekman. Facial areas and emotional information. *Journal of Communication*, 25(2):21–29, 1975.
- [14] G. H. Bourne. Structure and function of muscle. In *Physiology and Biochemistry*, 2nd ed., volume III. Academic Press, New York, 1973.
- [15] Matthew Brand. Voice Puppetry. In Alyn Rockwood, editor, *Computer Graphics (SIGGRAPH '99 Conference Proceedings)*, pages 21–28. ACM SIGGRAPH, August 1999.
- [16] Christoph Bregler, Michele Covell, and Malcolm Slaney. Video Rewrite: Driving Visual Speech with Audio. In Turner Whitted, editor, *Computer Graphics (SIGGRAPH '97 Conference Proceedings)*, pages 353–360. ACM SIGGRAPH, August 1997.

- [17] Christoph Bregler, Lorie Loeb, Erika Chuang, and Hrishu Deshpande. Turning to the Masters: Motion Capturing Cartoons. In John F. Hughes, editor, *ACM Transactions on Graphics (SIGGRAPH 2002 Conference Proceedings)*, pages 399–407. ACM SIGGRAPH, July 2002.
- [18] S. E. Brennan. Caricature generator. Master’s thesis, Massachusetts Institute of Technology, Cambridge, MA, 1982.
- [19] N. M. Brooke and E. D. Petajan. Seeing speech: Investigations into the synthesis and recognition of visible speech movements using automatic image processing and computer graphics. In *Proceedings of the International Conference on Speech Input/Output: Techniques and Applications*, pages 104–109, 1986.
- [20] J. Bulwer. *Philocopus, or the Deaf and Dumbe Mans Friend*. Humphrey and Moseley, London, 1648.
- [21] J. Bulwer. *Pathomyotamia, or, A dissection of the signfictive muscles of the affections of the minde*. Humphrey and Moseley, London, 1649.
- [22] N. Burtnyk and M. Wein. Computer generated key-frame animation. *Journal of SMPTE*, 80:149–153, 1971.
- [23] Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Bechet, Brett Douville, Scott Prevost, and Matthew Stone. Animated Conversation: Rule-Based Generation of Facial Expression Gesture and Spoken Intonation for Multiple Conversational Agents. In Andrew Glassner, editor, *Computer Graphics (SIGGRAPH '94 Conference Proceedings)*, pages 413–420. ACM SIGGRAPH, July 1994.
- [24] Lieu-Hen Chen, Santi Saeyor, Hiroshi Dohi, and Mitsuru Ishizuka. A System of 3D Hair Style Synthesis based on the Wisp Model. *The Visual Computer*, 15(4):159–170, 1999.
- [25] H. Chernoff. The use of faces to represent points in n -dimensional space graphically. Technical Report Project NR-042-993, Office of Naval Research, December 1971.
- [26] H. Chernoff. The use of faces to represent points in k -dimensional space graphically. *Journal of American Statistical Association*, page 361, 1973.
- [27] M. Cohen and D. Massaro. Synthesis of visible speech. *Behavioral Research Methods and Instrumentation*, 22(2):260–263, 1990.
- [28] M. Cohen and D. Massaro. Development and experimentation with synthetic visual speech. *Behavioral Research Methods, Instrumentation, and Computers*, pages 260–265, 1994.
- [29] Michael M. Cohen and Dominic W. Massaro. Modeling Coarticulation in Synthetic Visual Speech. In Nadia M. Magnenat-Thalmann and Daniel Thalmann, editors, *Models and Techniques in Computer Animation*, pages 139–156. Springer-Verlag, 1993.
- [30] Agnes Daldegan, Nadia M. Magnenat-Thalmann, Tsuneya Kurihara, and Daniel Thalmann. An Integrated System for Modeling, Animating and Rendering Hair. In Roger J. Hubbold and Robert Juan, editors, *Computer Graphics Forum (Proceedings of Eurographics '93)*, volume 12, pages 211–221, September 1993.
- [31] Charles Darwin. *Expression of the Emotions in Man and Animals*. J. Murray, 1872.
- [32] Paul E. Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the Reflectance Field of a Human Face. In Kurt Akeley, editor, *Computer Graphics (SIGGRAPH 2000 Conference Proceedings)*, pages 145–156. ACM SIGGRAPH, July 2000.
- [33] Paul E. Debevec and Jitendra Malik. Recovering High Dynamic Range Radiance Maps from Photographs. In Turner Whitted, editor, *Computer Graphics (SIGGRAPH '97 Conference Proceedings)*, pages 369–378. ACM SIGGRAPH, August 1997.
- [34] Douglas DeCarlo, Dimitris Metaxas, and Matthew Stone. An Anthropometric Face Model using Variational Techniques. In Michael F. Cohen, editor, *Computer Graphics (SIGGRAPH '98 Conference Proceedings)*, pages 67–74. ACM SIGGRAPH, July 1998.
- [35] B. deGraf. Notes on facial animation. In *State of the Art in Facial Animation*, volume 22 of *SIGGRAPH '89 Tutorials*, pages 10–11. ACM, 1989.
- [36] B. DeGraph and Wahrman. Mike, the talking head. *Computer Graphics World*, July 1988.

- [37] A. K. Dewdney. The complete computer caricaturist and whimsical tour of face space. *Scientific American*, 4(?):20–28, 1977.
- [38] S. DiPaola. Implementation and use of a 3D parameterized facial modeling and animation system. In *State of the Art in Facial Animation*, volume 22 of *SIGGRAPH '89 Tutorials*. ACM, 1989.
- [39] S. DiPaola. Extending the range of facial types. *J. of Visualization and Computer Animation*, 2(4):129–131, October-December 1991.
- [40] G. B. Duchenne. *The Mechanism of Human Facial Expression*. Jules Renard, Paris, 1862.
- [41] Peter Eisert and Bernd Girod. Model-based Facial Expression Parameters from Image Sequences. In *Proceedings of the IEEE International Conference on Image Processing (ICIP-97)*, pages 418–421, 1997.
- [42] P. Ekman. *Darwin and Facial Expressions*. Academic Press, New York, 1973.
- [43] P. Ekman. The argument and evidence about universals in facial expressions of emotion. In *Handbook of Social Psychophysiology*. Wiley, New York, 1989.
- [44] P. Ekman and W. V. Friesen. *Manual for the Facial Action Coding System*. Consulting Psychologists Press, Palo Alto, 1978.
- [45] P. Ekman, W. V. Friesen, and P. Ellsworth. *Emotion in the Human Face: Guidelines for Research and a Review of Findings*. Pergamon Press, New York, 1972.
- [46] P. Ekman and H. Oster. Facial expressions of emotion. *Annual Review of Psychology*, 1979.
- [47] M. Elson. Displacement facial animation techniques. In *SIGGRAPH State of the Art in Facial Animation: Course #26 Notes*, pages 21–42. ACM, Dallas, August 1990.
- [48] I. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. Technical Report 247, MIT Perceptual Computing Section, 1994.
- [49] Tony Ezzat, Gadi Geiger, and Tomaso Poggio. Trainable Videorealistic Speech Animation. In John F. Hughes, editor, *ACM Transactions on Graphics (SIGGRAPH 2002 Conference Proceedings)*, pages 388–398. ACM SIGGRAPH, July 2002.
- [50] G. Faigin. *The Artist's Complete Guide to Facial Expressions*. Watson-Guptill, New York, 1990.
- [51] L. G. Farkas and I. R. Munro. *Anthropometric Facial Proportions in Medicine*. Charles C. Thomas, Springfield, Illinois, 1987.
- [52] D. R. Forshey and R. H. Bartels. Hierarchical B-spline refinement. In *Computer Graphics (SIGGRAPH '88)*, volume 22, pages 205–212, 1988.
- [53] L. A. Fried. *Anatomy of the Head, Neck, Face, and Jaws*. Lea and Febiger, Philadelphia, 1976.
- [54] M. Gillenson and B. Chandrasekaran. Whatisface: Human facial composition by computer graphics. *ACM Annual Conference*, 9(1), 1975.
- [55] M. L. Gillenson. *The Interactive Generation of Facial Images on a CRT Using a Heuristic Strategy*. PhD thesis, Ohio State University, Computer Graphics Research Group, Columbus, Ohio, March 1974.
- [56] Taro Goto, Marc Escher, Christian Zanardi, and Nadia Magnenat-Thalmann. MPEG-4 based Animation with Face Feature Tracking. In Nadia Magnenat-Thalmann and Daniel Thalmann, editors, *Proceedings of the Eurographics Workshop on Computer Animation and Simulation '99*, pages 89–98, 1999.
- [57] Hans Peter Graf, Eric Cosatto, and Tony Ezzat. Face Analysis for the Synthesis of Photo-Realistic Talking Heads. In *Proceedings 4th International Conference on Automatic Face and Gesture Recognition*, pages 189–194, 2000.
- [58] B. Guenter. *A System for simulating Human Facial Expression*. PhD thesis, Ohio State University, 1989.
- [59] Brian Guenter, Cindy Grimm, Daniel Wood, Henrique Malvar, and Frédéric Pighin. Making Faces. In Michael F. Cohen, editor, *Computer Graphics (SIGGRAPH '98 Conference Proceedings)*, pages 55–66. ACM SIGGRAPH, July 1998.
- [60] T. Guiard-Marigny, A. Adjoudani, and C. Benoit. A 3D model of the lips for visual speech synthesis. In *Proc. 2nd ETRW on Speech Synthesis*, pages 49–52, New Platz, New York, 1994.

- [61] Sunil Hadap and Nadia Magnenat-Thalmann. Modeling Dynamic Hair as a Continuum. In Alan Chalmers and Theresa-Marie Rhyne, editors, *Computer Graphics Forum (Proceedings of Eurographics 2001)*, volume 20, pages C329–C338, September 2001.
- [62] U. Hadar, T. J. Steiner, E. C. Grant, and F. C. Rose. The timing of shifts in head postures during conversation. *Human Movement Science*, 3:237–245, 1984.
- [63] P. Hanrahan. Reflection from layered surfaces due to subsurface scattering. *Computer Graphics (SIGGRAPH 93)*, 27:165–174, 1993.
- [64] P. Hanrahan and D. Sturman. Interactive animation of parametric models. *The Visual Computer*, 1(4):260–266, 1985.
- [65] Antonio Haro, Brian Guenter, and Irfan Essa. Real-time, Photo-realistic, Physically Based Rendering of Fine Scale Human Skin Structure. In Steven J. Gortler and Karol Myszkowski, editors, *Rendering Techniques 2001 (Proceedings 12th Eurographics Workshop on Rendering)*, pages 53–62, 2001.
- [66] C. Henton. Beyond visemes: Using disemes in synthetic speech with facial animation, May 1994.
- [67] C. Henton and P.C. Litwinowicz. Saying and seeing it with feeling: techniques for synthesizing visible, emotional speech. In *Computer Graphics Conference Proceedings of the 2nd ESCA/IEEE Workshop on Speech Synthesis*, pages 73–76, September 1994.
- [68] E. H. Hess. The role of pupil size in communication. *Scientific American*, pages 113–119, Nov. 1975.
- [69] D. R. Hill, A. Pearce, and B. Wyvill. Animating speech: An automated approach using speech synthesis by rules. *The Visual Computer*, 3:277–289, 1988.
- [70] Carl-Herman Hjortsjo. *Man's Face and Mimic Language*. Lund, Sweden, 1970.
- [71] S. A. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, October 1996.
- [72] Horace H. S. Ip and C. S. Chan. Script-Based Facial Gesture and Speech Animation Using a NURBS Based Face Model. *Computers & Graphics*, 20(6):881–891, November 1996.
- [73] Tomomi Ishii, Takami Yasuda, Shigeki Yokoi, and Jun-ichiro Toriwaki. A Generation Model for Human Skin Texture. In *Proceedings of Computer Graphics International '93*, pages 139–150. Springer-Verlag, 1993.
- [74] ISO/IEC. Overview of the MPEG-4 Standard. <http://www.cseit.it/mpeg/standards/mpeg-4/mpeg-4.htm>, July 2000.
- [75] J. Jeffers and M. Barley. *Speechreading (Lipreading)*. Charles C. Thomas, Springfield, Illinois, 1971.
- [76] Won-Ki Jeong, Kolja Kähler, Jörg Haber, and Hans-Peter Seidel. Automatic Generation of Subdivision Surface Head Models from Point Cloud Data. In *Proceedings of Graphics Interface 2002*, pages 181–188. Canadian Human-Computer Communications Society, May 2002.
- [77] Kolja Kähler, Jörg Haber, and Hans-Peter Seidel. Geometry-based Muscle Modeling for Facial Animation. In *Proceedings of Graphics Interface 2001*, pages 37–46. Canadian Human-Computer Communications Society, June 2001.
- [78] Kolja Kähler, Jörg Haber, and Hans-Peter Seidel. Dynamically refining animated triangle meshes for rendering. *The Visual Computer*, 19(5):310–318, August 2003.
- [79] Kolja Kähler, Jörg Haber, and Hans-Peter Seidel. Reanimating the Dead: Reconstruction of Expressive Faces from Skull Data. *ACM Transactions on Graphics (SIGGRAPH 2003 Conference Proceedings)*, 22(3):554–561, July 2003.
- [80] Kolja Kähler, Jörg Haber, Hitoshi Yamauchi, and Hans-Peter Seidel. Head shop: Generating animated head models with anatomical structure. In *Proceedings ACM SIGGRAPH Symposium on Computer Animation (SCA '02)*, pages 55–64, July 2002.
- [81] G. A. Kalberer, P. Mueller, and L. Van Gool. Biological Motion of Speech. In *Biologically Motivated Computer Vision*, volume 2525 of *Lecture Notes in Computer Science*, pages 199–206. Springer, 2002.
- [82] G. A. Kalberer and L. Van Gool. Face Animation Based on Observed 3D Speech Dynamics. In H-S. Ko, editor, *Proceedings of The Fourteenth IEEE Conference on Computer Animation (CA'01)*, pages 20–27. IEEE Computer Society, 2001.

- [83] P. Kalra and N. Magnenat-Thalmann. Modeling vascular expressions in facial animation. In *Computer Animation '94*, pages 50–58, Geneva, May 1994. IEEE Computer Society Press.
- [84] P. Kalra, A. Mangili, N. Magnenat-Thalmann, and D. Thalmann. Simulation of facial muscle actions based on rational free form deformations. In *Proc. Eurographics 92*, pages 59–69, Cambridge, 1992.
- [85] Prem Kalra, Angelo Mangili, Nadia Magnenat-Thalmann, and Daniel Thalmann. SMILE: A Multilayered Facial Animation System. In *Proceedings IFIP WG 5.10, Tokyo, Japan*, pages 189–198, 1991.
- [86] A. Kendon, editor. *Nonverbal Communication, Interaction, and Gesture: Selections from Semiotica*. Mouton Publishers, New York, 1981.
- [87] R. M. Kenedi, T. Gibson, J. H. Evans, and J. C. Barbenel. Tissue mechanics. *Physics in Medicine and Biology*, 20(5):699–717, February 1975.
- [88] J. Kleiser. A fast, efficient, accurate way to represent the human face. In *State of the Art in Facial Animation*, volume 22 of *SIGGRAPH '89 Tutorials*, pages 37–40. ACM, 1989.
- [89] K. Komatsu. Human skin capable of natural shape variation. *The Visual Computer*, 3(5):265–271, 1988.
- [90] T. Kurihara and K. Arai. A transformation method for modeling and animation of the human face from photographs. In N. Magnenat-Thalmann and D. Thalmann, editors, *Computer Animation '91*, pages 45–58. Springer-Verlag, 1991.
- [91] W. Larrabee and J. A. Galt. A finite element model of skin deformation: The finite element model. *Laryngoscope*, 96:419–412, 1986.
- [92] W. Larrabee and D. Sutton. A finite element model of skin deformation. an experimental model of skin deformation. *Laryngoscope*, 96:406–412, 1986.
- [93] Won-Sok Lee and Nadia Magnenat-Thalmann. Fast Head Modeling for Animation. *Image and Vision Computing*, 18(4):355–364, March 2000.
- [94] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Constructing Physics-based Facial Models of Individuals. In *Proceedings of Graphics Interface '93*, pages 1–8. Canadian Human-Computer Communications Society, May 1993.
- [95] Yuencheng Lee, Demetri Terzopoulos, and Keith Waters. Realistic Modeling for Facial Animations. In Robert Cook, editor, *Computer Graphics (SIGGRAPH '95 Conference Proceedings)*, pages 55–62. ACM SIGGRAPH, August 1995.
- [96] B. LeGoff, T. Guiard-Marigny, M. Cohen, and C. Benoit. Real-time analysis-synthesis and intelligibility of talking faces. In *Proc. 2nd ETRW on Speech Synthesis*, pages 53–56, New Platz, New York, 1994.
- [97] J. P. Lewis. Automated lip-sync: Background and techniques. *J. of Visualization and Computer Animation*, 2(4):118–122, October-December 1991.
- [98] John P. Lewis and Frederic I. Parke. Automated Lip-Synch and Speech Synthesis for Character Animation. In John M. Carroll and Peter P. Tanner, editors, *Proceedings of Human Factors in Computing Systems and Graphics Interface '87*, pages 143–147, April 1987.
- [99] A. Lofqvist. Speech as audible gestures. In W.J. Hardcastle and A. Marchal, editors, *Speech Production and Speech Modeling*, pages 289–322. Kluwer Academic Publishers, Dordrecht, 1990.
- [100] D. Lowe. Fitting parametrized three-dimensional models to images. *IEEE Transactions on Pattern Matching and Machine Intelligence*, 13(5).
- [101] D. Lowe. Solving for the parameters of object models from image descriptions. In *Proceedings of Computer Vision and Image Understanding Workshop*, pages 121–127, College Park, MD, 1980.
- [102] E. Luice-Smith. *The Art of Caricature*. Cornell University Press, Ithaca, 1981.
- [103] N. Magnenat-Thalmann and Thalmann. D. *Synthetic Actors in Computer Generated Three-Dimensional Films*. Springer Verlag, Tokyo, 1990.
- [104] N. Magnenat-Thalmann and P. Kalra. A model for creating and visualizing speech and emotion. In *Aspects of Automatic Natural Language Generation*, Trento, Italy, April 1992. Springer-Verlag.
- [105] N. Magnenat-Thalmann, H. Minh, M. deAngelis, and D. Thalmann. Design, transformation and animation of human faces. *The Visual Computer*, 5:32–39, 1989.

- [106] N. Magnenat-Thalmann, N. E. Primeau, and D. Thalmann. Abstract muscle actions procedures for human face animation. *Visual Computer*, 3(5):290–297, 1988.
- [107] Stephen R. Marschner, Brian Guenter, and Sashi Raghupathy. Modeling and Rendering for Realistic Facial Animation. In Bernard Péroche and Holly Rushmeier, editors, *Rendering Techniques 2000 (Proceedings 11th Eurographics Workshop on Rendering)*, pages 231–242, 2000.
- [108] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. Image-Based BRDF Measurement Including Human Skin. In Dani Lischinski and Greg Ward Larson, editors, *Rendering Techniques '99 (Proceedings 10th Eurographics Workshop on Rendering)*, pages 131–144, 1999.
- [109] K. Mase and A. Pentland. Automatic lipreading by optical-flow analysis. *Systems and Computers in Japan*, 22:N06, 1991.
- [110] D. W. Massaro. *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. L. Erlbaum Associates, Hillsdale, NJ, 1987.
- [111] D. W. Massaro. A precis of speech perception by ear and eye: A paradigm for psychological inquiry. *Behavioral and Brain Sciences*, 12:741–794, 1989.
- [112] D. W. Massaro and M.M. Cohen. Perception of synthesized audible and visible speech. *Psychological Science*, 1:55–63, 1990.
- [113] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:126–130, 1986.
- [114] S. Morishima. Synchronization of speech and facial expression. *IPSJ 3-9-27*, 91(4), 1991.
- [115] S. Morishima, K. Aiwaza, and H. Harashima. An intelligent facial image coding driven by speech and phoneme. In *Proc. IEEE ICASSP89*, pages 1795–1798, 1989.
- [116] S. Morishima and H. Harashima. Facial animation synthesis for human-machine communication system. In *Proc. 5th International Conf. on Human-Computer Interaction*, volume II, pages 1085–1090, Orlando, Aug. 1993.
- [117] M. Nahas, H. Huitric, M. Rioux, and J. Domey. Facial image synthesis using texture recording. *The Visual Computer*, 6(6):337–343, 1990.
- [118] M. Nahas, H. Huitric, and M. Sanintourens. Animation of a B-spline figure. *The Visual Computer*, 3(5):272–276, March 1988.
- [119] E. B. Nitchie. *How to Read Lips for Fun and Profit*. Hawthorne Books, New York, 1979.
- [120] Jun-yong Noh and Ulrich Neumann. A Survey of Facial Modeling and Animation Techniques. USC Technical Report 99-705, University of Southern California, Los Angeles, CA, 1999.
- [121] Jun-yong Noh and Ulrich Neumann. Expression Cloning. In Eugene Fiume, editor, *Computer Graphics (SIGGRAPH 2001 Conference Proceedings)*, pages 277–288. ACM SIGGRAPH, August 2001.
- [122] F. I. Parke. Computer generated animation of faces. *ACM Nat'l Conference*, 1:451–457, 1972.
- [123] F. I. Parke. *A Parametric Model for Human Faces*. PhD thesis, University of Utah, Salt Lake City, Utah, December 1974. UTEC-CSc-75-047.
- [124] F. I. Parke. Measuring three-dimensional surfaces with a two-dimensional data tablet. *Journal of Computers and Graphics*, 1(1):5–7, 1975.
- [125] F. I. Parke. A model of the face that allows speech synchronized speech. *Journal of Computers and Graphics*, 1(1):1–4, 1975.
- [126] F. I. Parke, editor. *Course notes 22: State of the Art in Facial Animation*. ACM SIGGRAPH, July 1989.
- [127] F. I. Parke, editor. *Course Notes 26: State of the Art in Facial Animation*. ACM SIGGRAPH, August 1990.
- [128] F. I. Parke. Control parameterization for facial animation. In N. Magnenat-Thalmann and D. Thalmann, editors, *Computer Animation '91*, pages 3–14. Springer-Verlag, 1991.
- [129] F. I. Parke. Techniques for facial animation. In N. Magnenat-Thalmann and D. Thalmann, editors, *New Trends in Animation and Visualization*, pages 229–241. John Wiley, 1991.

- [130] Frederic I. Parke. Parameterized Models for Facial Animation. *IEEE Computer Graphics and Applications*, 2(9):61–68, November 1982.
- [131] Frederic I. Parke and Keith Waters. *Computer Facial Animation*. A K Peters, Wellesley, MA, 1996.
- [132] M. Patel. *Making Faces*. PhD thesis, School of Mathematics, University of Bath, UK, 1991. Technical Report 92-55.
- [133] M. Patel and P. J. Willis. FACES: Facial animation, construction and editing system. In *Proc. EURO-GRAPHICS '91*, pages 33–45. North-Holland, 1991.
- [134] E. Patterson, P. Litwinowicz, and N. Greene. Facial animation by spatial mapping. In N. M. Thalmann and D. Thalmann, editors, *Computer Animation '91*, pages 31–44. Springer-Verlag, Tokyo, 1991.
- [135] E.C. Patterson, P.C. Litwinowicz, and N. Greene. Facial animation by spatial mapping. In *Proceedings of Computer Animation '91*, pages 31–44, 1991.
- [136] A. Pearce, B. Wyvill, G. Wyvill, and D. Hill. Speech and expression: A computer solution to face animation. In *Proc. Graphics Interface '86*, pages 136–140, 1986.
- [137] A. Pearce and Y. H. Yen. A program for facial animation using computer graphics. Technical report, University of Calgary, Calgary, 1984.
- [138] D. Pearson and J. Robinson. Visual communication at very low data rates. *Proc. IEEE*, 73:795–812, April 1985.
- [139] C. Pelachaud. *Communication and Coarticulation in Facial Animation*. PhD thesis, University of Pennsylvania, Philadelphia, October 1991. Technical Report MS-CIS-91-77.
- [140] C. Pelachaud, N. Badler, and M. Steedman. Issues in facial animation. Technical Report MS-CIS-90-88, University of Pennsylvania, November 1990.
- [141] C. Pelachaud, N. I. Badler, and M. Steedman. Linguistic issues in facial animation. In N. Magnenat-Thalmann and D. Thalmann, editors, *Computer Animation '91*, pages 15–30. Springer-Verlag, 1991.
- [142] C. Pelachaud, N. I. Badler, and M. Steedman. Correlation of facial and vocal expressions in facial animation. In *Informatique '92 - Interface to Real and Virtual Worlds*, pages 95–110, Montpellier, France, March 1992.
- [143] Catherine Pelachaud, Norman Badler, and Mark Steedman. Generating Facial Expressions for Speech. *Cognitive Science*, 20(1):1–46, 1996.
- [144] A. Pentland and K. Mase. Lipreading: Automatic visual recognition of spoken words. In *Proc. Image Understanding and Machine Vision*. Optical Society of America, June 12-14 1989.
- [145] E. D. Petajan. Automatic lipreading to enhance speech recognition. In *IEEE Computer society conference on computer vision and pattern recognition*, June 1985.
- [146] E. D. Petajan, B. Bischoff, D. Bodoff, and N. M. Brooke. An improved automatic lipreading system to enhance speech recognition. In *Proceedings CHI 88*, pages 19–25, 1988.
- [147] S. D. Pieper. More than skin deep: Physical modeling of facial tissue. Master's thesis, Massachusetts Institute of Technology, Media Arts and Sciences, MIT, 1989.
- [148] S. D. Pieper. *CAPS: Computer-Aided Plastic Surgery*. PhD thesis, Massachusetts Institute of Technology, Media Arts and Sciences, MIT, September 1991.
- [149] Frédéric Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, and David H. Salesin. Synthesizing Realistic Facial Expressions from Photographs. In Michael F. Cohen, editor, *Computer Graphics (SIG-GRAPH '98 Conference Proceedings)*, pages 75–84. ACM SIGGRAPH, July 1998.
- [150] Frédéric Pighin, Richard Szeliski, and David H. Salesin. Resynthesizing Facial Animation through 3D Model-Based Tracking. In *Seventh IEEE International Conference on Computer Vision (ICCV '99)*, pages 143–150, 1999.
- [151] S. Platt, A. Smith, F. Azuola, N. Badler, and C. Pelachaud. Structure-based animation of the human face. Technical Report MS-CIS-91-15, University of Pennsylvania, Feb 1991.
- [152] S. M. Platt. A system for computer simulation of the human face. Master's thesis, The Moore School, University of Pennsylvania, Philadelphia, 1980.

- [153] S. M. Platt. *A Structural Model of the Human Face*. PhD thesis, The Moore School, University of Pennsylvania, Philadelphia, 1985.
- [154] Stephen M. Platt and Norman I. Badler. Animating Facial Expressions. In *Computer Graphics (SIGGRAPH '81 Conference Proceedings)*, volume 15, pages 245–252. ACM SIGGRAPH, August 1981.
- [155] W. T. Reeves. Simple and complex facial animation: Case studies. In *SIGGRAPH State of the Art in Facial Animation: Course 26 Notes*, pages 88–106. ACM, Dallas, August 1990.
- [156] L. G. Roberts. Machine perception of three-dimensional solids. Technical Report 315, Lincoln Laboratory, 1963.
- [157] G. J. Romanes. *Cunningham's Manual of Practical Anatomy, Vol 3: Head, Neck, and Brain*. Oxford Medical Publications, 1967.
- [158] R. Rosenblum, W. Carlson, and E. Tripp. Simulating the structure and dynamics of human hair: Modelling, rendering and animation. *J. Visualization and Computer Animation*, 2(4):141–148, October–December 1991.
- [159] M. A. Sagar, D. Bullivant, G. D. Mallinson, and P. J. Hunter. A virtual environment and model of the eye for surgical simulation. *Computer Graphics (SIGGRAPH '94)*, 28(4):205–212, July 1994.
- [160] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometry models. *Computer Graphics (SIGGRAPH '86)*, 20(4):151–160, 1986.
- [161] Y. Takashima, H. Shimazu, and M. Tomono. Story driven animation. In *CHI+CG '87*, pages 149–153, Toronto, 1987.
- [162] Marco Tarini, Hitoshi Yamauchi, Jörg Haber, and Hans-Peter Seidel. Texturing Faces. In *Proceedings of Graphics Interface 2002*, pages 89–98. Canadian Human-Computer Communications Society, May 2002.
- [163] Karen T. Taylor. *Forensic Art and Illustration*. CRC Press, 2000.
- [164] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):569–579, 1993.
- [165] Demetri Terzopoulos and Keith Waters. Physically-based Facial Modelling, Analysis, and Animation. *Journal of Visualization and Computer Animation*, 1(2):73–80, December 1990.
- [166] D. Terzopoulos and K. Waters. Techniques for realistic facial modeling and animation. In N. Magnenat-Thalmann and D. Thalmann, editors, *Computer Animation '91*, pages 59–74, Tokyo, 1991. Springer-Verlag.
- [167] K. Thorisson. Dialogue control in social interface agents. *INTERCHI (Adjunct Proceedings)*, pages 139–140, April 1993.
- [168] J. T. Todd and S. M. Leonard. Issues related to the predication of cranial facial growth. *American Journal of Orthodontics*, 79(1):63–80, 1981.
- [169] J. T. Todd, S. M. Leonard, R. E. Shaw, and J. B. Pittenger. The perception of human growth. *Scientific American*, 242:106–114, 1980.
- [170] Allen Van Gelder. Approximate Simulation of Elastic Membranes by Triangulated Spring Meshes. *Journal of Graphics Tools*, 3(2):21–41, 1998.
- [171] M. W. Vannier, J. F. Marsch, and J. O. Warren. Three-dimensional computer graphics for craniofacial surgical planning and evaluation. *Computer Graphics*, 17, 1983.
- [172] Thomas Vetter and Volker Blanz. Estimating Coloured 3D Face Models from Single Images: An Example Based Approach. *Lecture Notes in Computer Science*, 1407:499–513, 1998.
- [173] M. L. Viaud and H. Yahia. Facial animation with wrinkles. Technical Report 1753, INRIA, LeChesnay, France, September 1992.
- [174] C. T. Waite. The facial action control editor, face: A parametric facial expression editor for computer generated animation. Master's thesis, Massachusetts Institute of Technology, Media Arts and Sciences, Cambridge, MA., February 1989.

- [175] J. Walker, L. Sproull and R. Subramani. Using a human face in an interface. *ACM CHI*, pages 85–91, April 1994.
- [176] E. F. Walther. *Lipreading*. Nelson-Hall Inc, Chicago, 1982.
- [177] C. L. Wang. Automating facial gestures and synthesized speech in human character animation. In *Proceedings of the Third Annual Western Computer Graphics Symposium*, pages 39–40, Vernon, BC, April 1991.
- [178] C. L. Wang. Langwidere: Hierarchical spline based facial animation system with simulated muscles. Master's thesis, University of Calgary, Calgary, Alberta, October 1993.
- [179] T. Watanabe. Voice-responsive eye-blinking feedback for improved human-to-machine speech input. In *Proc. 5th International Conf. on Human-Computer Interaction*, volume II, pages 1091–1096, Orlando, Aug. 1993.
- [180] Y. Watanabe and Y. Suenaga. A trigonal prism-based method for hair image generation. *IEEE Computer Graphics and Applications*, 12(1):47–53, January 1992.
- [181] K. Waters. Expressive three-dimensional facial animation. *Computer Animation (CG86)*, pages 49–56, October 1986.
- [182] K. Waters. Animating human heads. *Computer Animation (CG87)*, pages 89–97, October 1987.
- [183] K. Waters. Towards autonomous control for three-dimensional facial animation. *British Computer Society*, pages 10–20, December 1987.
- [184] K. Waters. *The Computer Synthesis of Expressive Three-Dimensional Facial Character Animation*. PhD thesis, Middlesex Polytechnic, Middlesex, June 1988. Faculty of Art and Design.
- [185] K. Waters. Modeling 3D facial expression: Tutorial notes. In *State of the Art in Facial Animation, SIGGRAPH '89 Tutorial Notes*, pages 127–152. ACM, Aug 1989.
- [186] K. Waters. A physical model of facial tissue and muscle articulation derived from computer tomography data. *SPIE Proceedings of Visualization in Biomedical Computing, Chapel Hill, N.Carolina*, 1808:574–583, Oct 1992.
- [187] K. Waters and T. Levergood. An automatic lip-synchronization algorithm for synthetic faces. In *Proceedings of the Multimedia Conference*, pages 149–156, San Francisco, California, Sept 1994. ACM.
- [188] K. Waters and T. M. Levergood. DECface: an automatic lip synchronization algorithm for synthetic faces. Technical Report CRL 93/4, DEC Cambridge Research Laboratory, Cambridge, MA., September 1993.
- [189] K. Waters and D. Terzopoulos. A physical model of facial tissue and muscle articulation. *Proceedings of the First Conference on Visualization in Biomedical Computing*, pages 77–82, May 1990.
- [190] K. Waters and D. Terzopoulos. The computer synthesis of expressive faces. *Phil. Trans. R. Soc. Lond.*, 355(1273):87–93, Jan 1992.
- [191] Keith Waters. A Muscle Model for Animating Three-Dimensional Facial Expression. In *Computer Graphics (SIGGRAPH '87 Conference Proceedings)*, volume 21, pages 17–24. ACM SIGGRAPH, July 1987.
- [192] Keith Waters and Joe Frisbie. A Coordinated Muscle Model for Speech Animation. In *Proceedings of Graphics Interface '95*, pages 163–170. Canadian Human-Computer Communications Society, May 1995.
- [193] Keith Waters and Demetri Terzopoulos. Modeling and Animating Faces Using Scanned Data. *Journal of Visualization and Computer Animation*, 2(4):123–128, October–December 1991.
- [194] P. Weil. About face. Master's thesis, Massachusetts Institute of Technology, Architecture Group, August 1982.
- [195] Lance Williams. Performance-Driven Facial Animation. In Forest Baskett, editor, *Computer Graphics (SIGGRAPH '90 Conference Proceedings)*, volume 24, pages 235–242. ACM SIGGRAPH, August 1990.
- [196] V. Wright. Elasticity and deformation of the skin. In H. R. Elden, editor, *Biophysical Properties of Skin*. Wiley-Interscience, New York, 1977.

- [197] Yin Wu, Prem Kalra, Laurent Moccozet, and Nadia Magnenat-Thalmann. Simulating Wrinkles and Skin Aging. *The Visual Computer*, 15(4):183–198, 1999.
- [198] Yin Wu, Nadia Magnenat-Thalmann, and Daniel Thalmann. A Plastic-Visco-Elastic Model for Wrinkles in Facial Animation and Skin Aging. In J. N. Chen, editor, *Proceedings of the Second Pacific Conference on Computer Graphics and Applications (Pacific Graphics '94)*, pages 201–214, August 1994.
- [199] B. Wyvill. Expression control using synthetic speech. In *State of the Art in Facial Animation*, volume 22 of *SIGGRAPH '89 Tutorials*, pages 163–175. ACM, 1989.
- [200] B. Wyvill, D. R. Hill, and A. Pearce. Animating speech: An automated approach using speech synthesized by rules. *The Visual Computer*, 3(5):277–289, March 1988.
- [201] J. F. S. Yau and N. D. Duffy. A texture mapping approach to 3D facial image synthesis. In *Eurographics UK*, pages 17–30, April 1988.

5 Organizers and presenters

Jörg Haber is a senior researcher at the Max-Planck-Institute for Computer Sciences in Saarbrücken, Germany. He received his Master's (1994) and PhD (1999) degrees in Mathematics from the TU München, Germany. During the last seven years he did research in various fields of computer graphics and image processing, including global illumination and real-time rendering techniques, physics-based simulation, scattered data approximation, and lossy image compression. For the last couple of years, his major research interests concentrate on modeling, animation, and rendering of human faces. He received the Heinz-Billing-Award 2001 of the Max-Planck-Society and the SaarLB Science Award 2001 for the design and implementation of a facial modeling and animation system. Last year, he presented a paper at SIGGRAPH on the reconstruction of faces from skull data.

Demetri Terzopoulos holds the Lucy and Henry Moses Professorship in the Sciences at New York University and is a professor of computer science and mathematics at NYU's Courant Institute. He is also affiliated with the University of Toronto where he is Professor of Computer Science and Professor of Electrical and Computer Engineering. He graduated from McGill University and received the PhD in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (MIT). His published work includes more than 200 technical articles and several volumes, primarily in computer graphics and vision. He has given hundreds of invited talks around the world on these topics, among them numerous keynote and plenary addresses, including addresses at Eurographics, Graphics Interface, and Pacific Graphics. Terzopoulos is a Fellow of the IEEE. His many awards include computer graphics honors from the International Digital Media Foundation, Ars Electronica, and NICOGRAPH. The latter recognized his research on human facial animation, which stems from his pioneering work on physics-based human facial modeling and deformable models.

Frederic I. Parke is Professor of Architecture and Director of the Visualization Laboratory in the Department of Architecture at Texas A&M University. The Laboratory supports the multi-disciplinary graduate program in Visualization Sciences. Professor Parke received a PhD in computer science from the University of Utah in 1974 for his pioneering work on facial modeling and animation. In 1996 he co-authored an authoritative volume on "Computer Facial Animation", published by AK Peters Ltd. He has interests in many aspects of computer based systems, computer graphics, and visualization. Recently his research has been in the area of conversational interfaces which combines a long standing interest in facial animation with the use of speech recognition and speech synthesis to support multi-modal user interfaces and low bandwidth virtual video conferencing. Prior to joining TAMU, he was a senior member of IBM's Architecture and Technical Strategy group, Visual Systems, RS/6000 Division, and was the technical lead and chief architect of the Actor portion of IBM's Human Centered technologies. Previously, he was a professor of computer science and director of the Computer Graphics Laboratory at the New York Institute of Technology.

Lance Williams worked for Robert Haralick at the University of Kansas, and studied computer graphics and animation under Ivan Sutherland, David Evans, and Steven Coons at the University of Utah. He worked in the Computer Graphics Lab at the New York Institute of Technology (from 1976–1986) on research and commercial animation, and the development of shadow mapping and "mip" texture mapping. Subsequently Williams consulted for Jim Henson Associates, independently developed facial tracking for computer animation, worked for six years in Apple Computer's Advanced Technology Group and for three years at DreamWorks SKG. He completed a long-deferred Ph.D. at the University of Utah in August 2000, and received SIGGRAPH's Steven Anson Coons Award in 2001. In 2002, he received an honorary Doctor-

ate of Fine Arts from Columbus College of Art and Design, and a motion picture Technical Academy Award. He is currently Chief Scientist at Walt Disney Feature Animation.

Volker Blanz studied Physics at University of Tübingen, Germany, and University of Sussex, Brighton, UK. At Max-Planck Institute for biological Cybernetics, Tübingen, he wrote both his Diploma thesis (1995) on image-based object recognition, and his PhD thesis (2000) on automated reconstruction of 3D face models from images. In 1995, he worked on Multi-class Support Vector Machines at AT&T Bell Labs. In recent years, he has also worked for the Center for Biological and Computational Learning at MIT several times. From 2000 to 2003, he has been a research assistant at University of Freiburg, and since 2003, he is a researcher at Max-Planck Institute for Computer Sciences, Saarbrücken. His research interests are 3D face models, animation, face recognition, and learning theory. His Eurographics 2003 paper on facial animation in images and video received the Best Paper Award 2003.

George Borshukov is a VFX Technology Supervisor at ESC Entertainment. His work has focused on production techniques for photorealistic rendering and reality-based virtual cinematography, including human faces. He received an Academy Award for Scientific and Technical Achievement in 2000. Borshukov graduated from the University of Rochester and holds a Masters Degree from the University of California at Berkeley, where he specialized in computer graphics and computer vision. He is one of the creators of *The Campanile Movie*, shown at SIGGRAPH 1997, which presented a breakthrough in the field of image-based modeling and rendering. He entered the visual effects industry in 1997, joining the R&D team of the Academy Award winning film *What Dreams May Come* at Manex Visual Effects. He served as the Technical Designer for the “Bullet Time” sequences in *The Matrix*. At ESC Entertainment he recently completed work on *The Matrix Reloaded* and *The Matrix Revolutions* for sequences like the “Burly Brawl” and “The Superpunch”. His work has also been applied on key shots in *Deep Blue Sea*, *Mission: Impossible 2*, and the IMAX film *Michael Jordan to the MAX*.