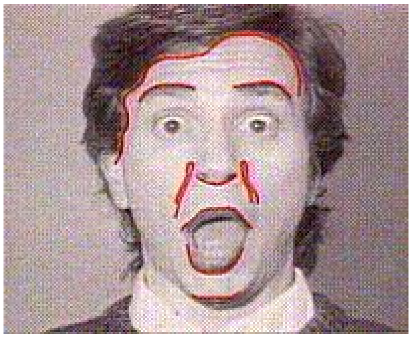
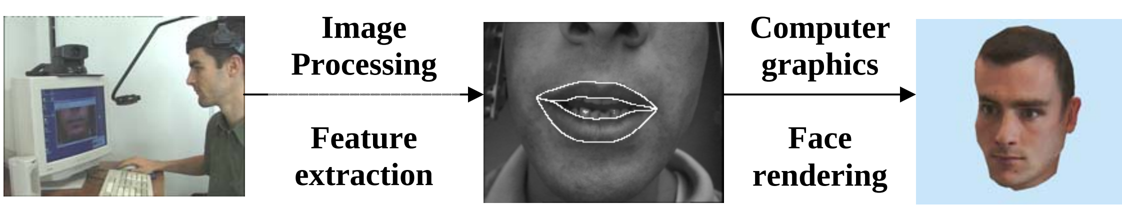
3D face analysis and synthesis overview



D. Terzopoulos

Parke's model

Within the last 10 years, fast increase in performance of memory, display and processor speed has allowed the expansion of Computer Graphics. It has now overcome Image Processing in its achievement. In the 3D face field, the CG-generated faces are almost indiscernible from real faces. Still it requires manual drawing for each image and artistic skills. It may take hours to reproduce a particular face and much more to render expressions.



These lectures will focus not on the 3D face rendering by means of Computer Graphics but on 3D faces analysis and synthesis from real images or videos, acquired either from 2D or 3D acquisition systems, of a human character.

3D vs 2D

There is a growing need for more interactive (remote) communication devices between humans and machines (avatars), humans (hyper-realistic faces: clones).
There is a strong belief that 3D face representation caries more information than 2D face representation.

2 schools:

2D: the 2D images are the ones measured by the cameras and eyes, available to computers and animals. If we could model these signals, we could have an elementary and general explanation for perception tasks. Still 2D human face recognition results are degraded when the face rotation is more than 30 % from its frontal view. The 2D approach usually does not solve the illumination variation problems

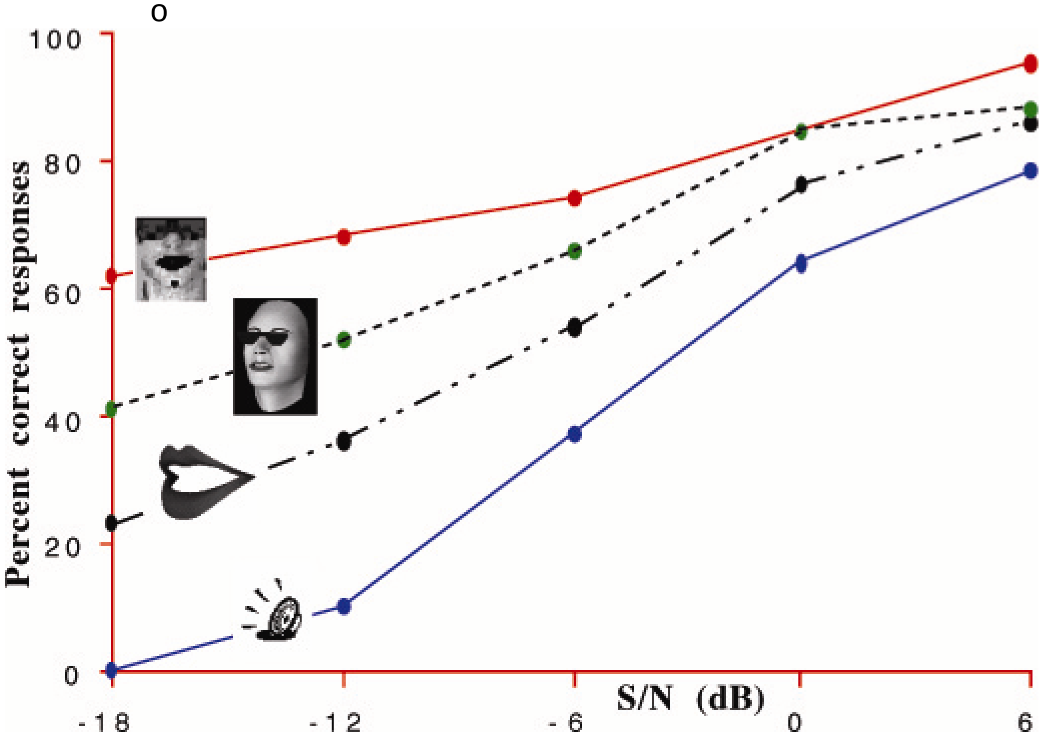
3D: it explains the patterns of faces accurately using the physics of the world. It may solve the illumination condition problem (at least for range data obtained from 3D scans).

There is no true 3D image display system
Late studies show that 3D information as well as 2D information is used by human observers (Liu and Kersten 1998) for some recognition tasks. Moreover, being trained on a given set of 3D objects helps the recognition task (for the same class of objects) under unusual viewing conditions.

What is it with face?

It is the vector of communication (with hands):

- Through expressions
- Through visual speech (mainly mouth area)
- Through features (mouth, eyebrows, eyes, shape, skin texture).
    - Provide recognition
    - Speech augmented
    -



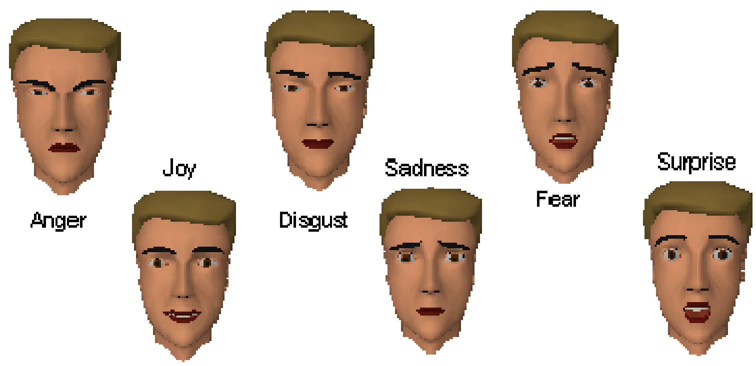
Speech intelligibility vs S/N ratio (**ICP**)

Variability
Highly deformable area: The face is able to generate around 55 000 different expressions, involving more than 200 distinct muscles, accounting for about 30 semantic distinction. Facial action coding system (FACS) has been developed (Ekman 77) to provide a quantification of facial expressions. Six primary expressions common to almost all the cultures: anger, disgust, fear, happiness, sadness and surprise.
Haig (1984) did psychological experiments altering the relative position of facial features.

Results showed that small changes in the spatial relationships between face features (eyebrows, eyes) create great variation in facial recognition ratio. Changing eye distance affects the recognition of human faces.

Emotion: Sequence of expressions in time

- Psychic and physical reactions of person to precept.
- Involves body response, facial expression and gesture response
- Different emotions may arouse from different persons when subjected to the same percept
- For the face, it can be described as a set of expressions
-



Expressions (based on Mpeg 4 FAPS and AU's)

Illumination

Studies on the degree of variability of images of male faces (without glasses or beards and excluding hair) demonstrated that 3 main factors were involved: illumination, viewpoint and different individuals.

- Difference in individuals account only for a least portion of variability.
- The viewpoint was responsible for 20% of the variability
- Illumination 150 % of the variability.
-

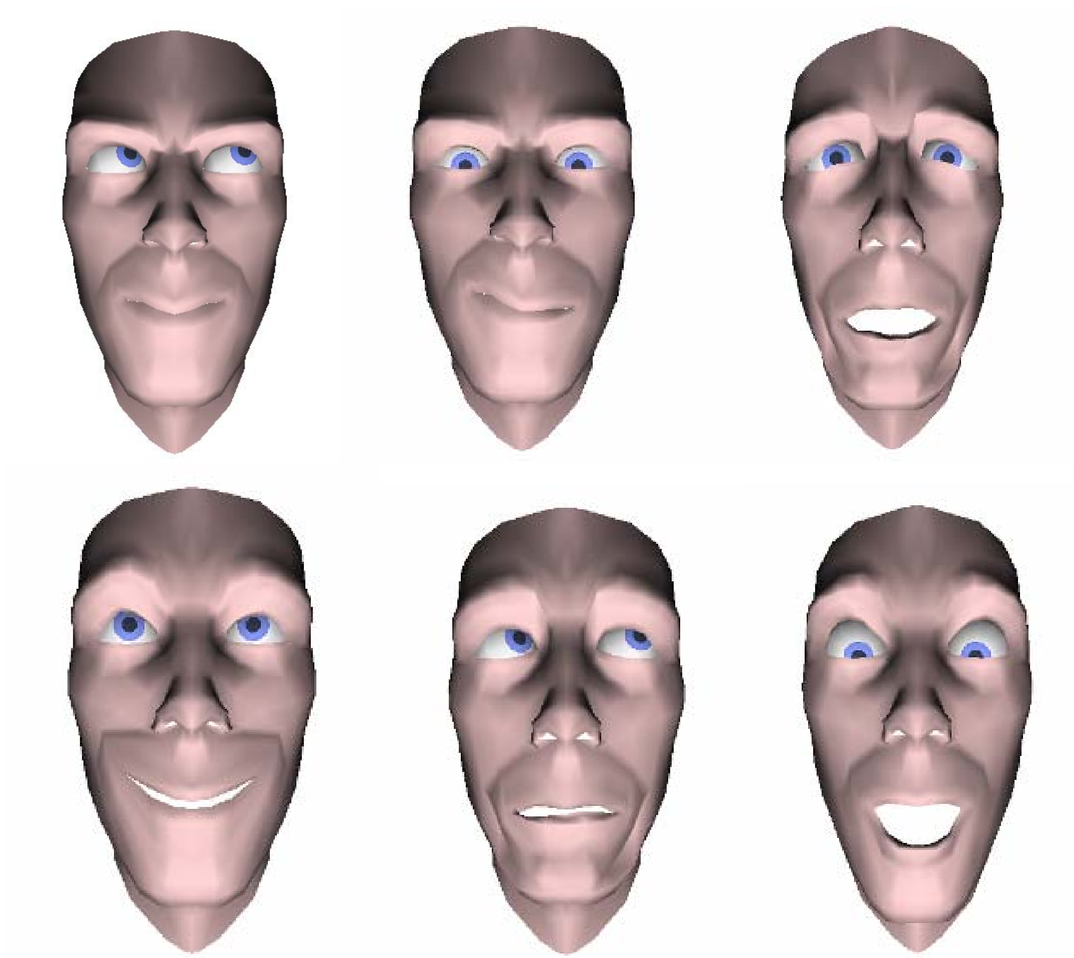
Face analysis systems usually proceed under constant (and/or well defined) lighting conditions

- Difficulty to recognize negative images
  - It has been demonstrated that bottom lighting (apparent direction of lighting for negative images) generate harder recognition task for familiar faces
- Under unknown but simple illumination, illumination estimation problem has been solved for range data obtained from 3D scanner (Eurecom 2000)
-

Still today, the exact process of face recognition by the brain is not precisely known.

- It integrates high and low frequencies recognition task
  - Sex judgment may be performed using only LF components
  - Identification requires HF components
- Color is not necessarily needed
- Importance of the pose

Other model - Same expressions



Human Factors

- Gender
- Body weight
- Facial expressions
- Shape
- Facial features
- Skin texture

Environment factors

- Orientation in space (pose)
- Illumination conditions
- Background

Face modelling

The realistic modelling of the human face is one of the most elusive goals in computer animation.
Two of the most difficult challenges:

- Modelling changes in skin colour as a figure moves from light to dark areas of a virtual world
- Mapping the facial movements of one person onto another.

Main developments
Facial rendering
Animating speech
Animating expressions
Modeling hair
Rendering hair
Modeling skin (texture)
Rendering skin

Eye gaze
First model:

- Fixed eyes or random movements for the Parke's model and its successors

Next generation:

- Eyes express emotions and intentions and help direct attention.
- Will pay attention to the look and behaviour of the clones and avatars' eyes.

Applications:

Medical

1. Facial reconstruction from skull (medical-investigation)
2. Medical Imaging
3. Modeling wound closure (simulation of skins layers)
4. Simulating facial muscles

5. Simulating facial surgery

Creating caricatures

Compression of videos with faces

Telecommunication:

1. Speech animation (from voice to face animation)
2. Virtual avatars
3. Morphing expression

Virtual teleconferencing

- Realistic Face: Virtual Reality
- Telecommunication a low bandwidth

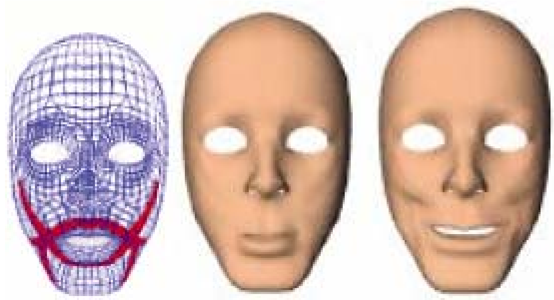
Human-machine communication

Interactive environments

Main research categories

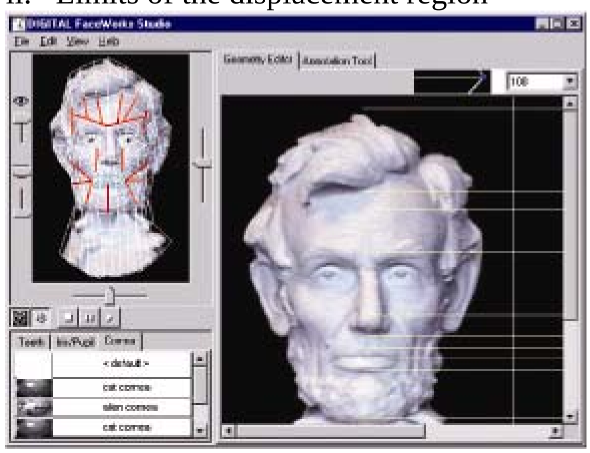
1. Physical, Muscle Models (Waters/Terzopoulos)
    - Multi-layered face model, which combines physics-based model of facial tissue with anatomically based facial muscle control.
    - 3 layers (epidermal, dermal and subcutaneous) to simulate the skin tissue.
    - Skin properties and muscle actions are simulated using an elastic spring mesh and forces organized in networks and connected layers. Any deformation of a muscle is propagated through skin layers.
    - Skin deformation is propagated at the surface as well.
    - 6 levels of representation
        a. Expression
        b. Control
        c. Muscles
        d. Physics
        e. Geometry
        f. Rendering
    - Create realistic facial animation (based on real physic observation).
    - Drawback:
        a. Difficulty to drive (although only 16 muscles involved)
            i. Coordination of several muscles for a given movement
        b. Computationally intensive.
        c. Redundancy: subspace of feasible displacement is far higher than the space of visemes needed to reproduce the overall face displacement.

3D reconstruction of a facial expression (Surprise? Fear?) with Terzopoulos' model



Muscles involved (and area of influence) for the deformation of the face from neutral position to smile expression

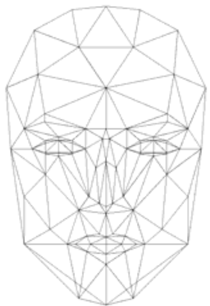
2. Parametric model (Parke, Marriot, Cohen and Massaro)
    - A (3) multilayered model which relies on a reduced set of meshes to represent the face geometry, a reduced number of muscle to represent face feature displacements and a 2D or 3D texture mapped onto the meshes surface.
    - A set of 3D meshes which described the surface geometry
    - A limited (simplified) set of muscles, which is connected to a single feature point of the face and a single point of the skull (mesh surface).
    - Each muscle has an area of influence via its connected feature point.
        i. Vertices moved by the feature points
        ii. Limits of the displacement region



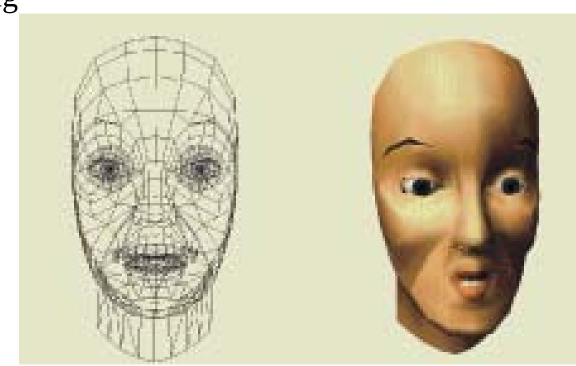
Faceworks interface with Parke's like model

- 3D structure of the face can be modified and deformed by the action of parameters
- Two different sets of parameters:
    i. Conformation parameters (control the topology of the face)
    ii. Expression parameters (eyebrow shapes, mouth shapes,..).
        1. Move vertices of the of the face (and connected meshes) through simple geometric transformations (rotation and translation) in one or more dimensions
- Facial animation:
    i. Change the values of control parameters

    ii. Redraw the face by using the new values.
- Advantages
    i. Simple (no elasticity model)
    ii. Efficient (requires low data storage)
- Drawbacks:
    i. Cannot cope with the complexity of face expressions
    ii. Poor rendering



Candide meshes

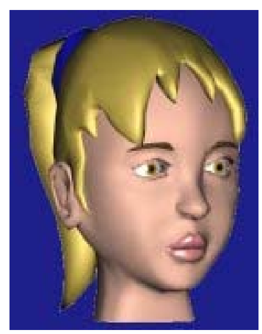


A modified Parke-Beskow meshes model

Outcome of CG's technique improvement:

- More vertices and meshes
- Finer interpolation of surface for better rendering via polynomial interpolation.

Mike (Pockaj) 750 polygons
408 vertices



Asia (Pockaj) 7726 polygons
4237 vertices

3. Hierarchical control. (Moving regions as opposed to points) (Object interaction occurs at the regional level as opposed to the point level)
4. Key framing
    - Extract a subset of key face images
    - Apply morphing techniques to move from one key frame to another one
    - It is used indirectly today by the MPEG 4 FACS system
5. Human actor tracked. (Puppet masters or markers)
    - Used mainly for entertainment applications
    - Experiments
6. Using deformations of Texture maps (optical flow) and geometry of the face.
    - OF: resulting apparent motion between 2 successive images
    - Minimization of quadratic difference between 2 images

- Optical flow is highly dependent on noise level
- Aperture problem: OF cannot be determined along the direction of the brightness pattern

7. Extraction of 3D model from 2D pictures.
   - Photometric stereo:
     - One camera + different light sources orientation
     - Depth map + texture mapping
     - Calibration
     - Not accurate but maybe enough for face rendering
   - Stereo vision:
     - 2 cameras + well defined (and constant) illumination
     - (Feature) Point correspondence
     - Feature points of the face are difficult to follow
   - Structured light
     - Projection of a pattern on the face and reconstruction of a 3D object by comparing the deformation of the pattern with its generic configuration
     - Calibration
     - With faces, several images needed to render the front face
8. Extraction of 3D model from 3D scans.
   - Accurate depth map
   - Accurate RGB color map
   - Expensive
   - Does not solve the feature point extraction problem
9. Facial recognition techniques.

Optotrak LED Markers (DWM)

Photometric stereo Yen Chen (CITR)

3D Scanner acquisition DWM

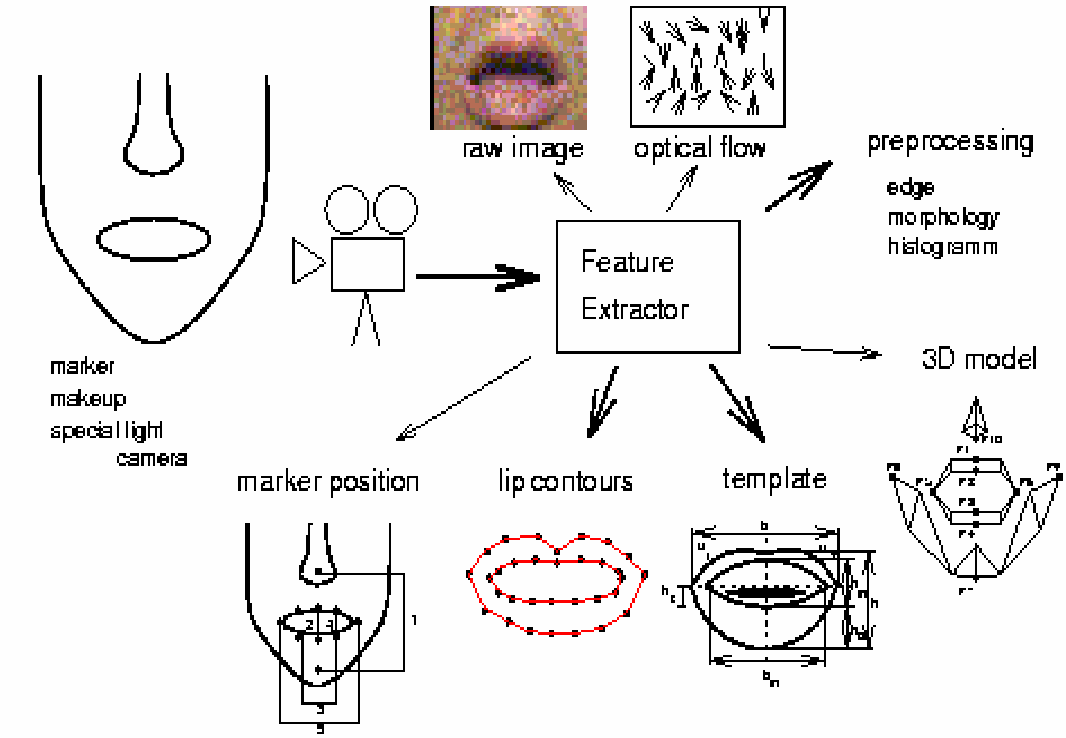
Inter-disciplinary aspects

1. Computer graphics
2. Computer vision, image processing
3. Animation
4. Art and design
5. Physics
6. Psychology
7. A.I.

Image processing techniques in face feature extraction

2D face feature extraction techniques

Lips and mouth area



Eyes

- Energy color based
- Deformable template (Yuille)
- Reflection techniques

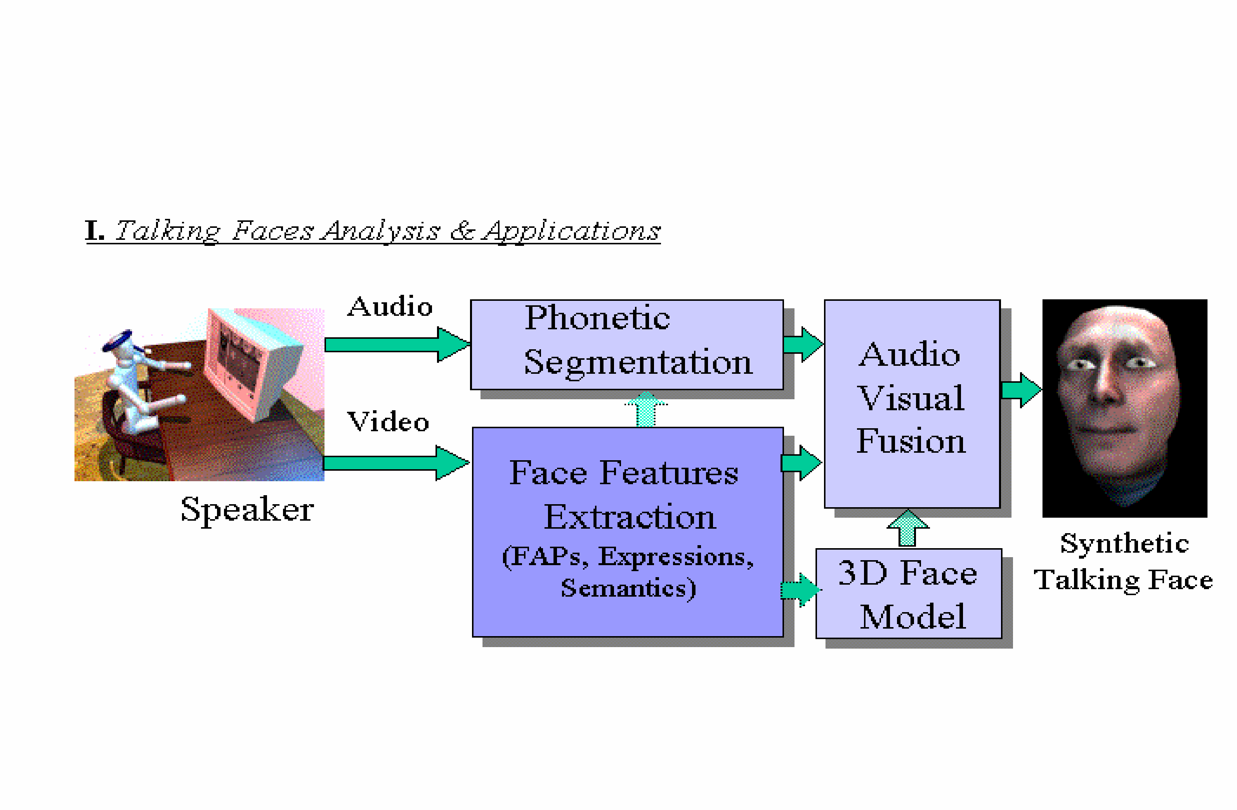
Eyebrows

- PCA
- Image training
- Active contours

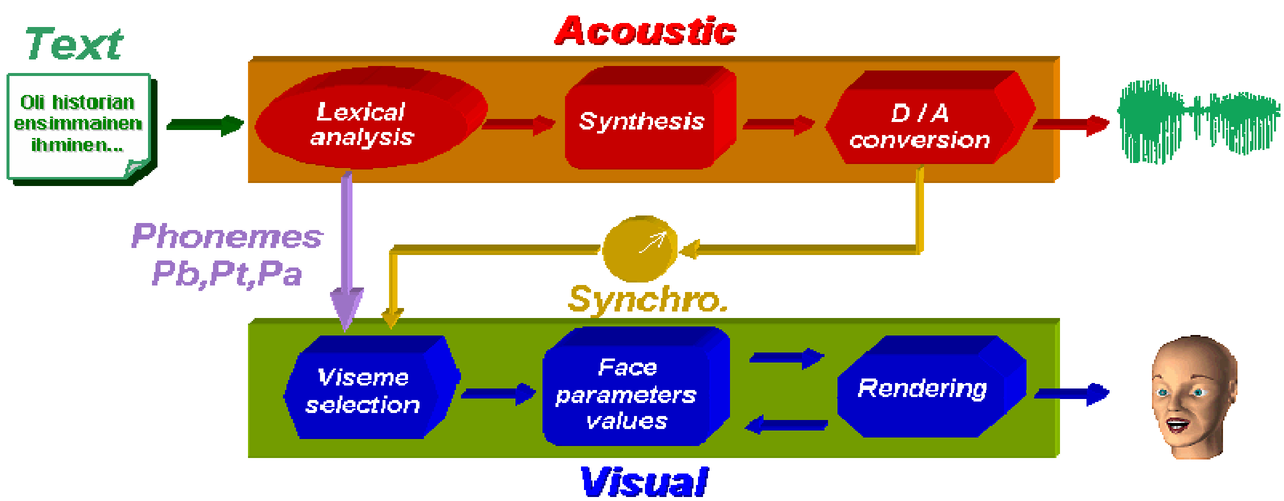
Expressions

- PCA
- Optical flow

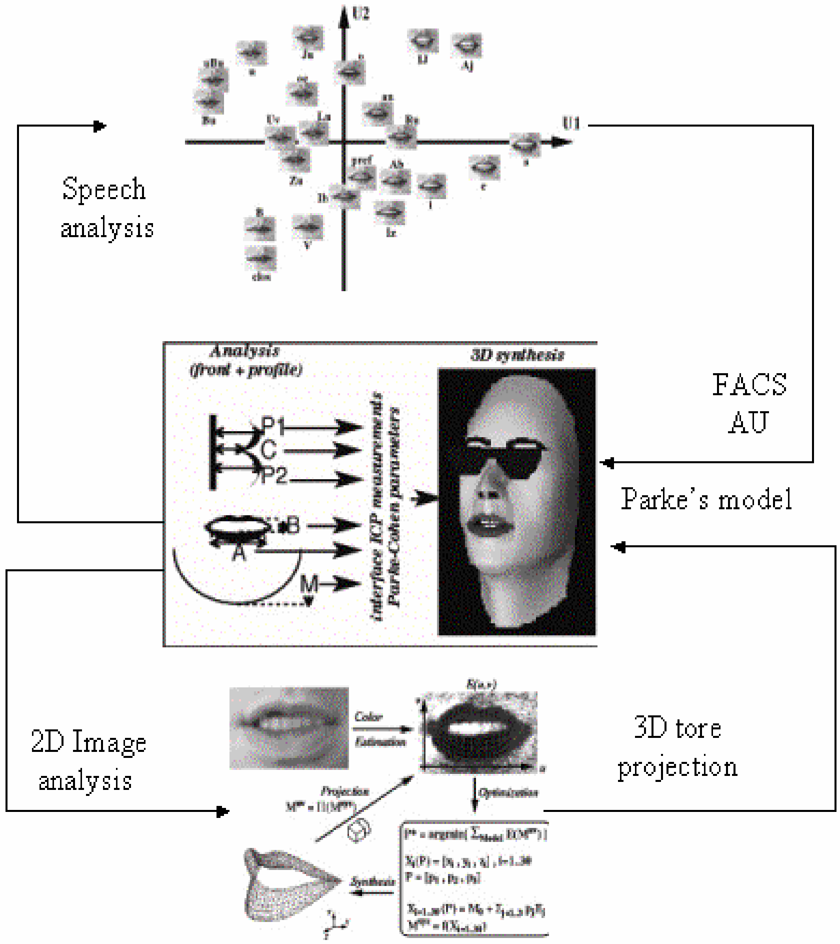
From image analysis to 3D face synthesis (an example)



Face synthesis from text analysis (the Finnish model)



Speech Analysis to infer 3D face synthesis (ICP-Grenoble)



A little bit of history:

Using technology to produce human faces and animating human faces started in the early 1970s with 2D and 3D methods. Parke (1972) creates the first animation of a 3D face by hand digitizing expressions and facial geometry then defining key positions. Animation of the face results from interpolation between the key frames. To overcome the complexity he developed (1974) a parametric facial model based on empirical and traditional hand-drawn animation methods. The parameters, which define a linear interpolation between two extremes, come in two flavors: conformation parameters and expression parameters. The conformation parameters define the shape of the face and the expression parameters are used to create the animation. Refinements of the Parke's model include more meshes.

It uses pervious work of Ekman (1972), which describes the facial expressions as the result of activities of specific facial muscles.

Kanade was the first (1973) to offer a complete face recognition system, which could take an input image extract face features and classify the faces. The recognition was performed using spatial relationships between features with controlled viewing conditions.

Montgomery (1980) augments earlier work with the addition of nonlinear interpolation between frames as well as forward and backward co articulation approximation. (lip reading ability)

Terzopoulos later (1990) developed a physic-based 3D face model taking into accounts the anatomy of the face with the representation of several layers (mainly bones, muscles and skin).

Ifran Essa "solved" the inverse problem in 1995: given a sequence of images describing a facial expression, find which group of muscles caused the motion and then deduce the expression using previous knowledge about facial expressions and the corresponding patterns motion using optical flows.

Since then, tremendous amount of work but still no full 3D face analysis and synthesis system with video cameras able to **extract**, at a processing speed close to **realtime**, **facial features**, and **reproduce** a dynamic facsimile (**clone**) of the subject's face at a remote location.

1975

1995

2015

Glossary:

Viseme:
A viseme is a generic facial image that can be used to describe a particular sound. A viseme is the visual equivalent of a phoneme or unit of sound in spoken language. Using visemes, the hearing-impaired can view sounds visually, effectively "lip-reading" the entire human face.

Phoneme:
A member of the set of the smallest units of speech that serve to distinguish one utterance from another in a language or dialect phoneme posture

Coarticulation: It refers to changes in the articulation of a speech segment depending on the previous and upcoming articulation segments.

Bibliography

[1] J. Ahlberg, "*CANDIDE-3 - an updated parametrized face*," Tech. Rep. LiTH-ISY-R-2326, Department of Electrical Engineering, Linkoping University, Sweden, 2001. http://www.icg.isy.liu.se/candide/main.html

[2] J. Edge,S. Maddock, 2. Computer Facial Animation website. 2003. http://www.dcs.shef.ac.uk/~steve/openDay/page2.html

[3] Y. C. Lee, D. Terzopoulos, K. Waters. Realistic face modeling for animation. Siggraph proceedings, 1995, pp. 55-62

[4] U. Neumann, J. Noh, A survey of facial modeling and animation techniques, 1998, USC Technical Report 99-705

[5] F. I. Parke, Computer Generated Animation of Faces. MS Thesis Technical Report, UTEC-CSC-72-120, Department of Computer Science, University of Utah, Salt Lake City, Utah.

[6] F. I. Parke, K. Waters, Computer Facial Animation, 1996, ISBN 1-56881-014-8
Websites for appendices featuring example mesh and code:
http://www.crl.hpl.hp.com/publications/books/waters/Appendix1/appendix1.html
http://www.crl.hpl.hp.com/publications/books/waters/Appendix2/ap2.html

[7] Parke, F. I. (1982) Parameterized models for facial animation, IEEE Computer Graphics, 2(9), 61-68.

[8] N. Pollard, Animation Timeline
http://www.cs.brown.edu/courses/cs229/animTimeline.html

[9] B. Spitzak and others, FLTK – The Fast Light Toolkit. 2003.
http://www.fltk.org

[10] D.Terzopoulos and K. Waters (1993). Analysis and synthesis of facial Image sequences using physical and anatomical models, IEEE PAMI, vol 15 (6), June 1993

[11] K. Waters, A muscle model for animating three-dimensional facial expression. In Maureen C. Stone, editor, Computer Graphics (Siggraph proceedings, 1987) vol. 21 pp. 17-24

[12] K. Waters, A muscle model for animating three-dimensional facial expression. Website at http://www.cs.wpi.edu/~matt/courses/cs563/talks/face_anim/waters.html

Software for 3D face analysis and modeling

1. **Candide (http://www.bk.isy.liu.se/candide/)**
2. **CSU-Toolkit** (http://cslu.cse.ogi.edu/toolkit/index.html)
3. **Faceworks** (http://interface.digital.com/developers/default.htm)

Mpeg4

MPEG4 is the evolution of previous audio and video encoding decoding standards MPEG1 and MPEG2, which were introduced to offer efficient audio and video applications through the Internet.

- MPEG-4 includes a standard for 3D face coding-transmission and decoding.
- Main application: critical bandwidth conditions.
- Standardization of the 3D model calibration and evolution to guarantee interoperability
- Requirements:
    - Speech should be intelligible
    - Facial expressions allow the recognition of the speaker's mood
    - Really reproduction of the speaker's face.

MPEG-4 specified two main types of facial data:

- **Facial Animation Parameters (FAPs)**
- **Facial Definition Parameters (FDPs)**

The Facial Animation Parameters (FAPS)

MPEG-4 defines:

68 Facial Animation Parameters (FAP)

- 2 hi-level FAPs: visemes and expressions
- 66 low-level FAPs.

With the exception of some FAPs that control the head rotations, the eyeball rotations each low-level FAP indicates the translation of the corresponding feature point, with respect to its position in the neutral face, along one of the coordinate axes.
Hi-level FAPs represents, the most common facial expressions (joy, sadness, anger, fear, disgust, surprise) and the visemes.

- FAPs animate a 3D facial model available at the receiver.
    - The availability of the 3D model at the receiver is not relevant.
    - FAPs allow:
        - The animation of key feature points in the model
        - The reproduction of visemes and expressions

The Neutral Face

The Neutral Face is used as a reference for the interpretation of the FAP values. It is defined as follows:

- The coordinate system is right handed; head axes are parallel to the world axes and gaze is in direction of Z axis
- Face muscles are relaxed

- Eyelids are tangent to the iris with the pupil diameter is one third of the iris diameter (IRISD0)
- Lips are in contact; the line of the inner lips is horizontal and at the same height of lip corners
- The tongue is flat, horizontal, with the tip of the tongue touching the boundary between upper and lower teeth which are touching each other

Biometrics information transmitted through MPEG4 standard is related to the deviations from the neutral face feature point positions. This allows reconstruction without previous knowledge of the 3D face model chosen.

The neutral face and its MPEG4 biometric components

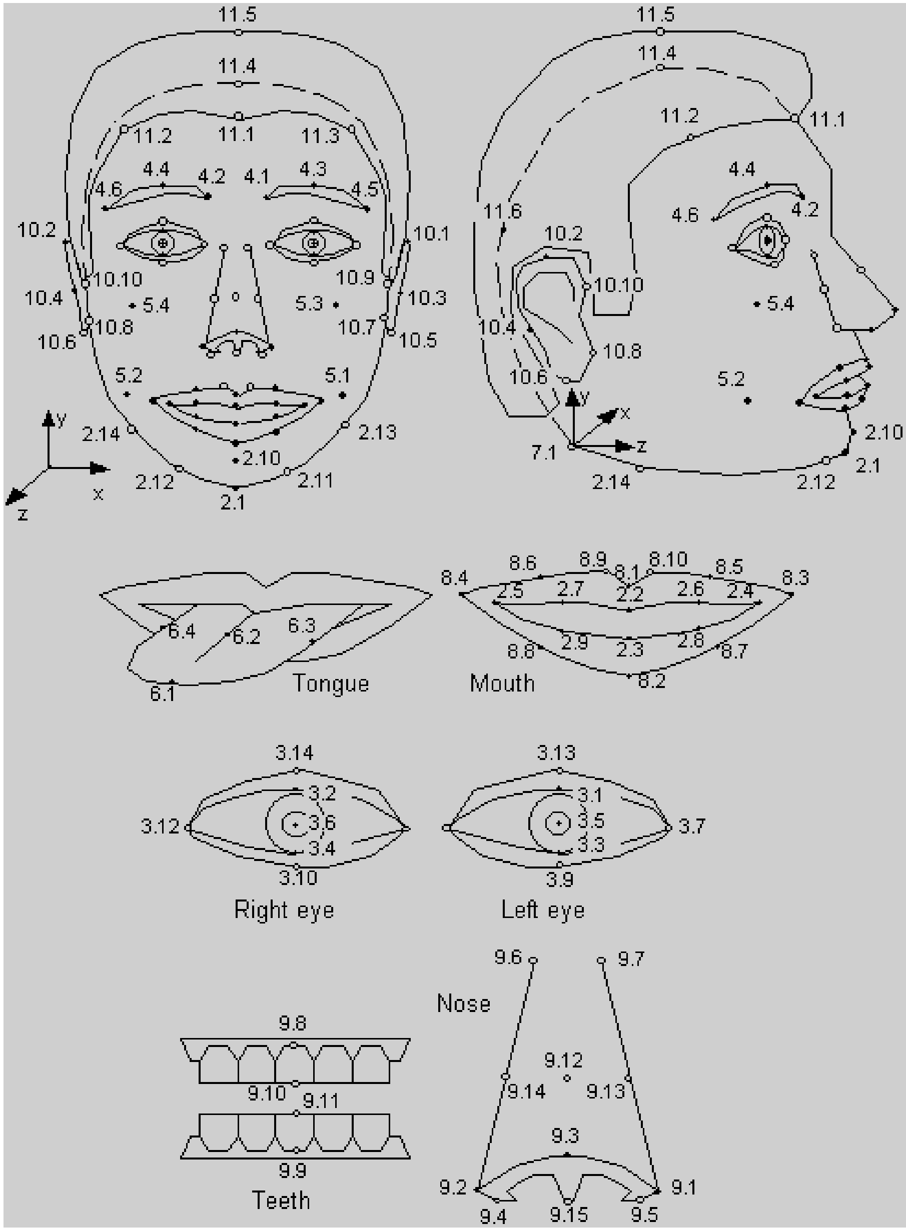
Feature Points

A feature point represents a key-point in a human face:

- The corner of the mouth
- The tip of the nose.
- Extremities of the eyes
- The corner of the eyebrow

MPEG-4 has defined a set of 84 feature points

- They are used both for the calibration and the animation of a synthetic face.



Feature points (related to FAP's). MPEG4