ECE 417 Lecture 23: 3D Face Modeling

Slides by Vuong Le, Spring 2013

Modified by Mark Hasegawa-Johnson, Fall 2018

Contents

- 3D Facial Geometry Modeling
- 3D Facial Geometry Acquisition
- 3D Facial Deformation Modeling
- Applications

3D Facial Geometry Modeling

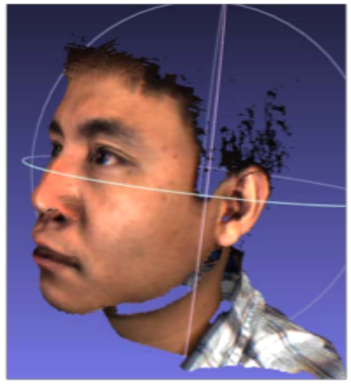
- Polygonal Mesh
- 3D Morphable Model
- Implicit Surface

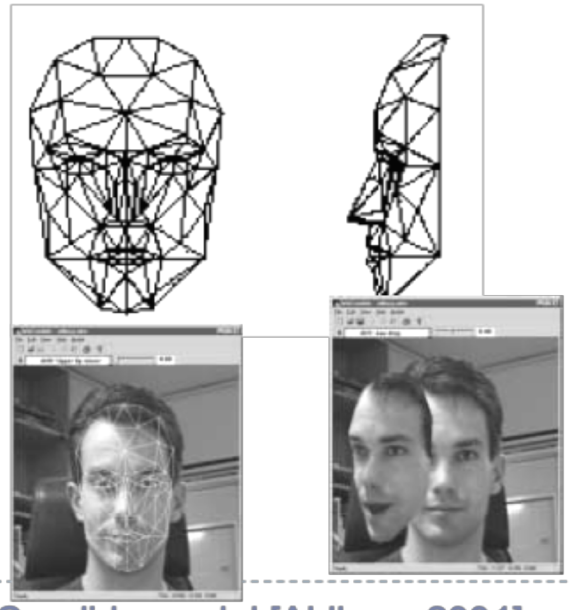
Point cloud / Polygonal mesh

- Consisting
  - Vertices
    - sample point of the human facial surface
    - can be in correspondence
  - Triangles
  - Normal vectors
  - Texture
    - mapped on points/flat triangles

→ Easy for rendering and low level operations

→ Very high dimensional space for detail models



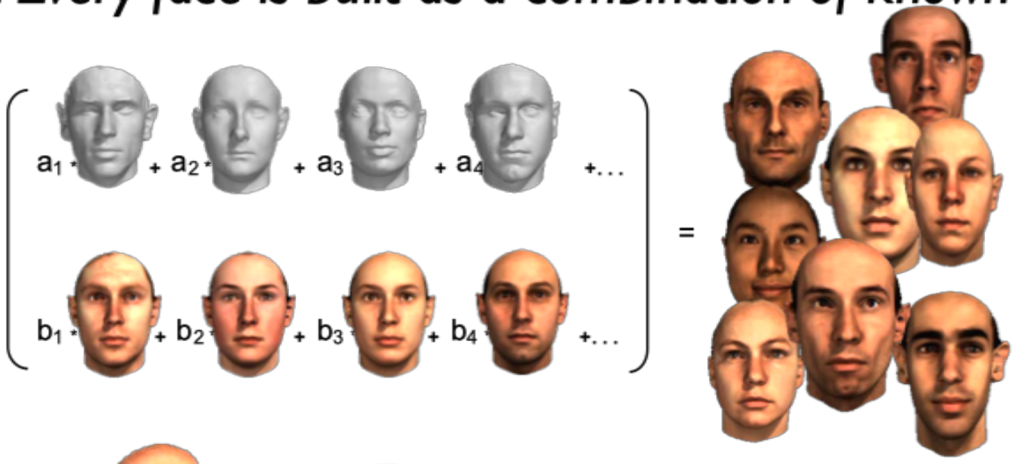


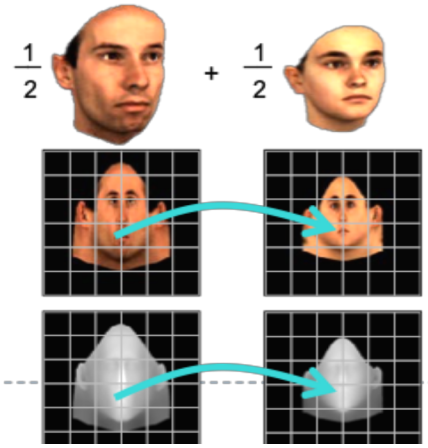
Candide model [Ahlberg 2001]

3D Morphable Model (1)

A human face is not an arbitrary 3D shape

Assumption: Every face is built as a combination of known faces







Without correspondence



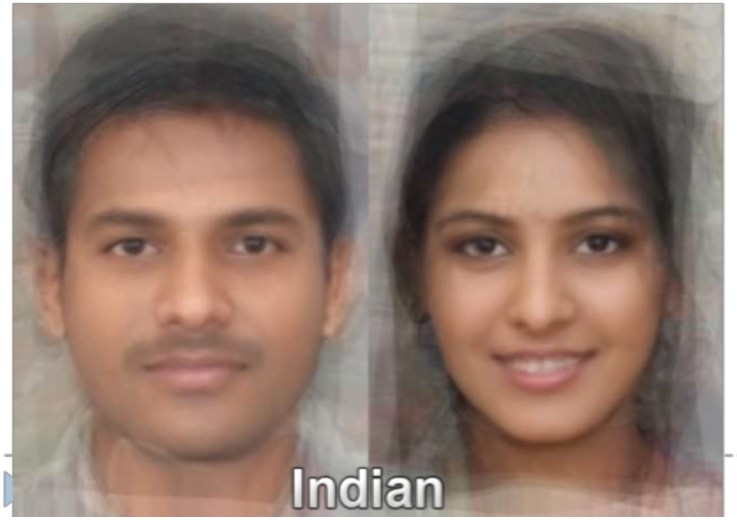
With correspondence

3DMM [Vetter 2009]

Average faces across ethnicities

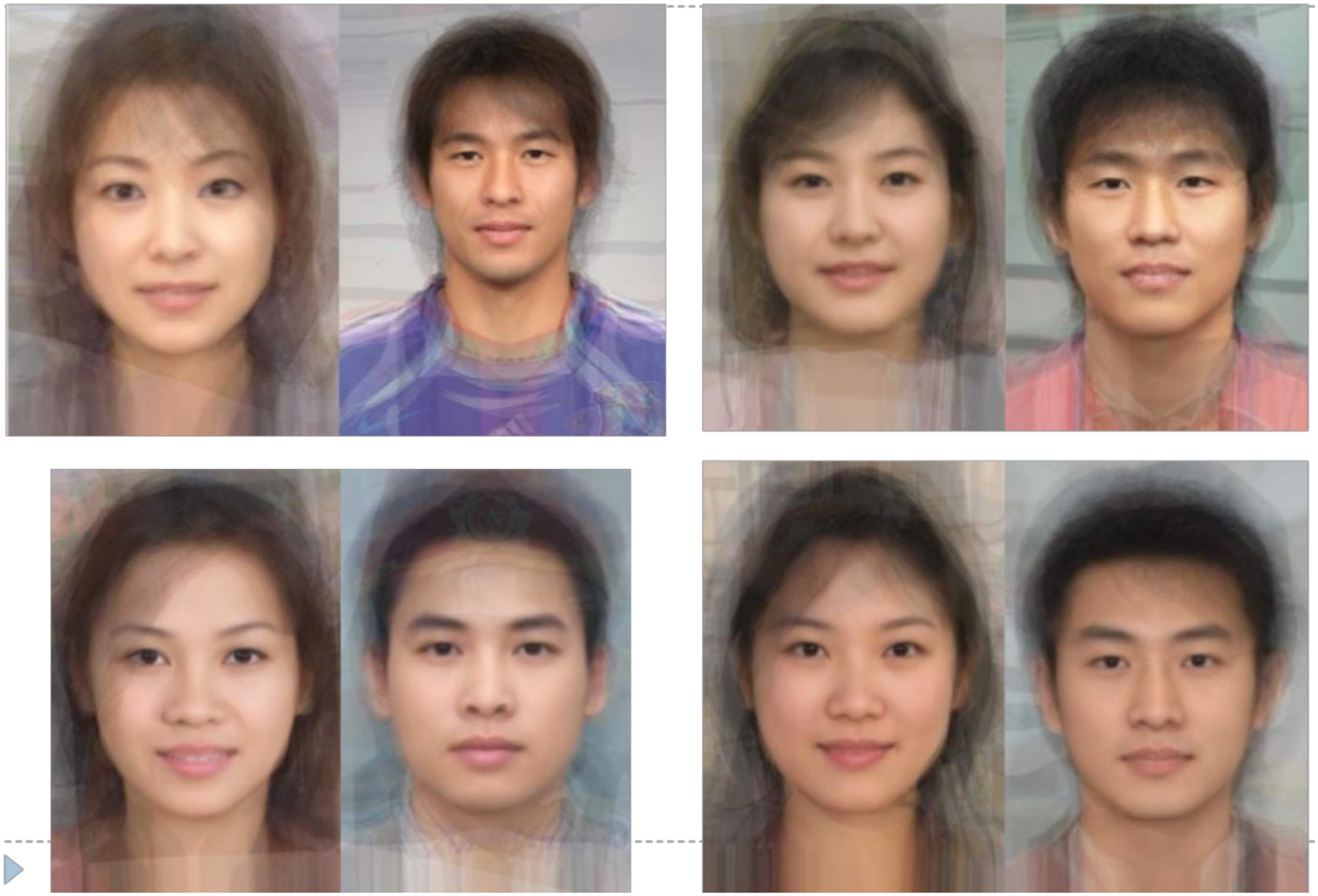






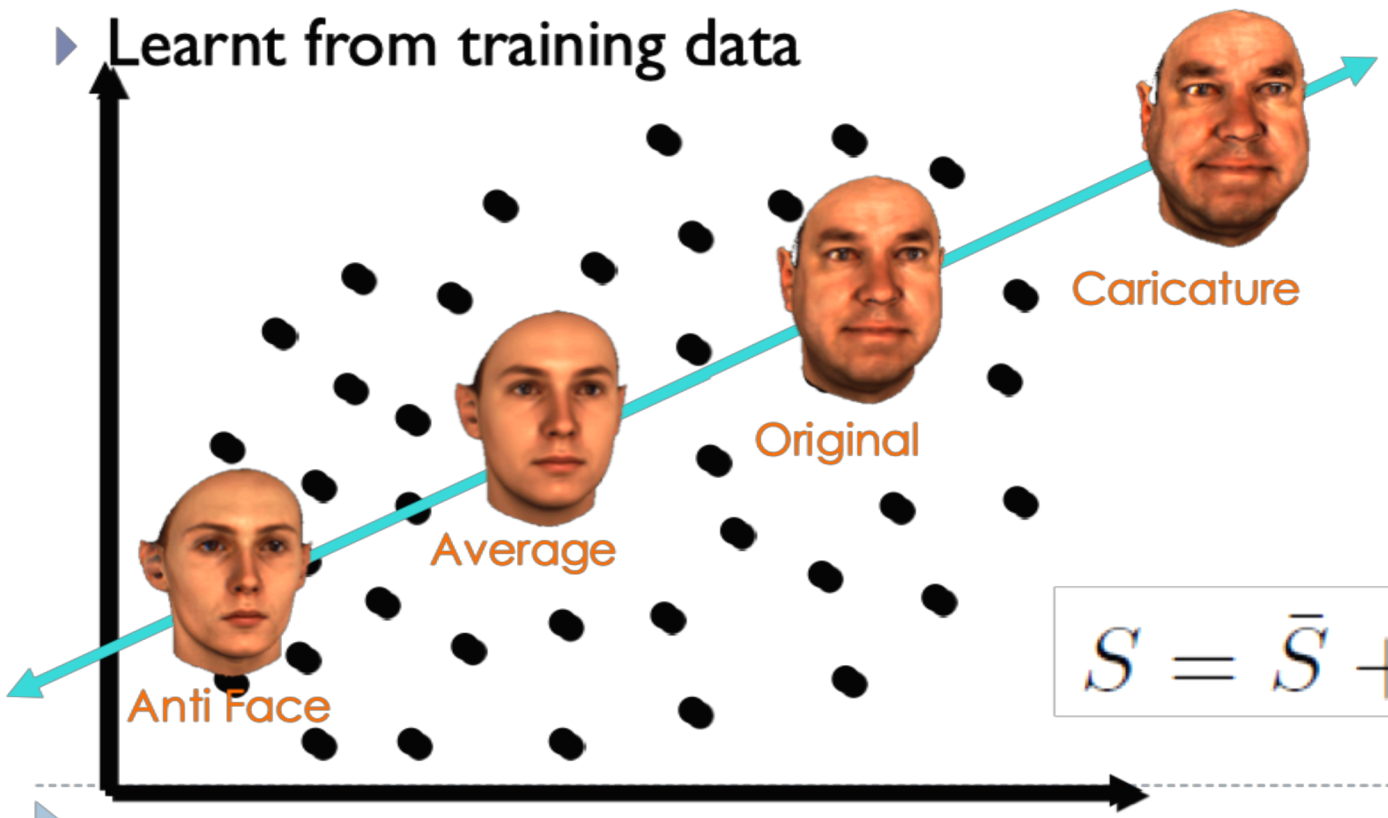


Average faces across nations



3D Morphable Model (2)

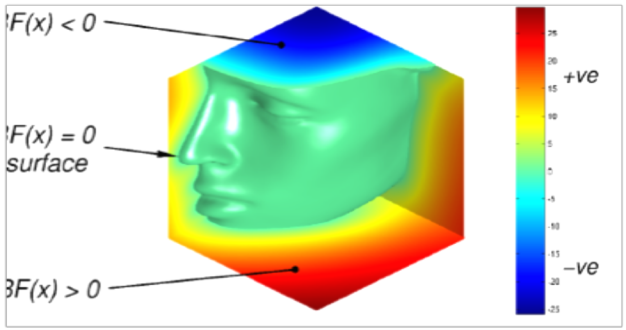
- A vector space of 3D faces
- Spanned by a PCA basis
- Learnt from training data

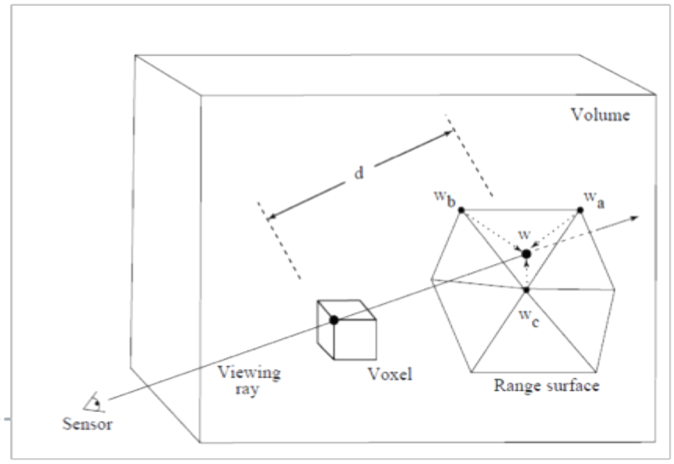


$$S = \bar{S} + P\vec{\sigma}$$

Implicit surface
Signed distance function volume

- Represented as a function has domain in R^3, giving the signed distance to surface
  - surface as zeros
  - free space as positive values
  - occupied space as negative values
- Discretized, limited into voxels of a cube
- Can be updated incrementally from pieces of point cloud
- Processing operations can be highly parallelized
- Converted to point cloud by ray tracing or marching cubes
- Example system: KinectFusion, Newcombe 2011





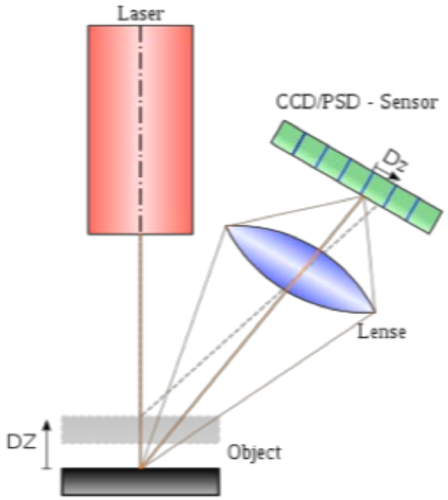
3D Facial Geometry Acquisition

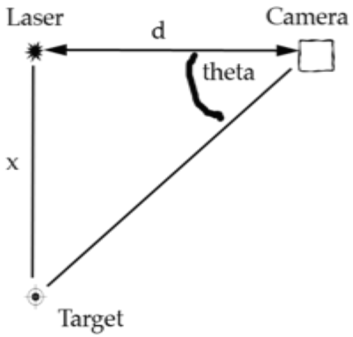
- Artist's Design (skipped)
- Active Acquisition
- Passive Acquisition
- Reconstruction from 2D Images

Active acquisition with Laser radiation

- Triangulation
  - Laser emitter shines a laser on the subject
  - Camera looks for the location of the laser dot.
  - Depending on how far away the laser strikes a surface, the laser dot appears at different places in the camera's field of view.

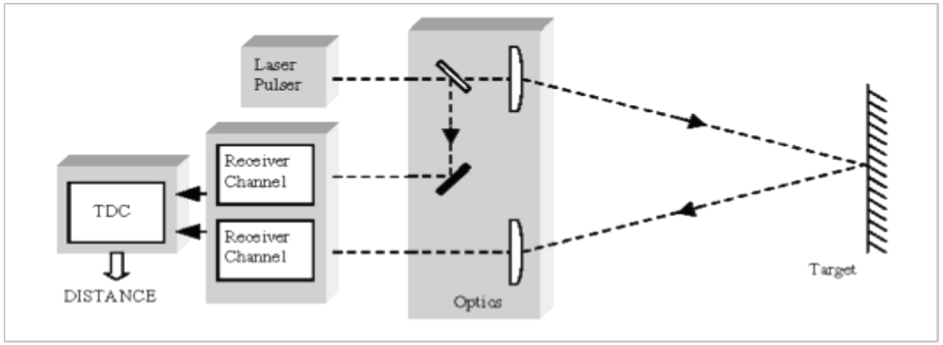






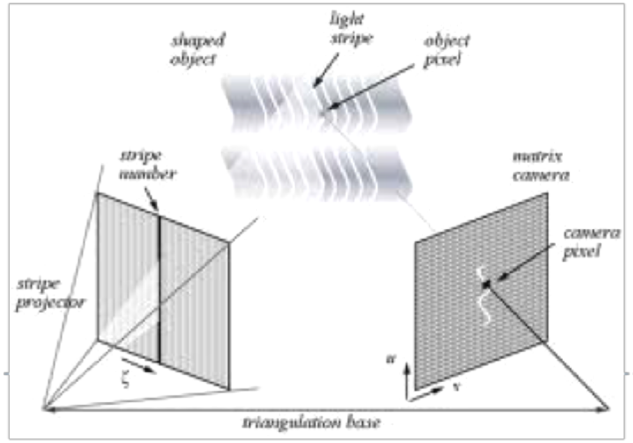
Active acquisition with Infra-red radiation

- Time of flight
  - Examples: DepthSense , PMD, Creative/Intel perceptual kit

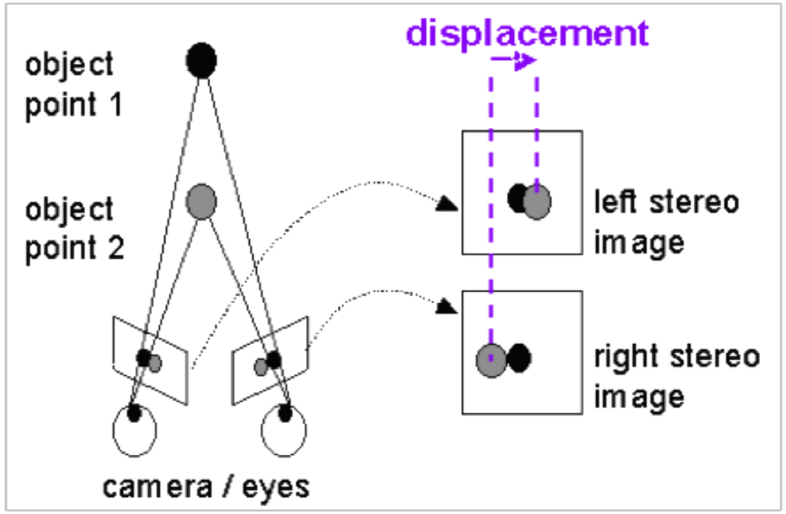




- Structured light
  - Example: Kinect, Asus Xtion Pro



Passive Stereo Photogrammetry



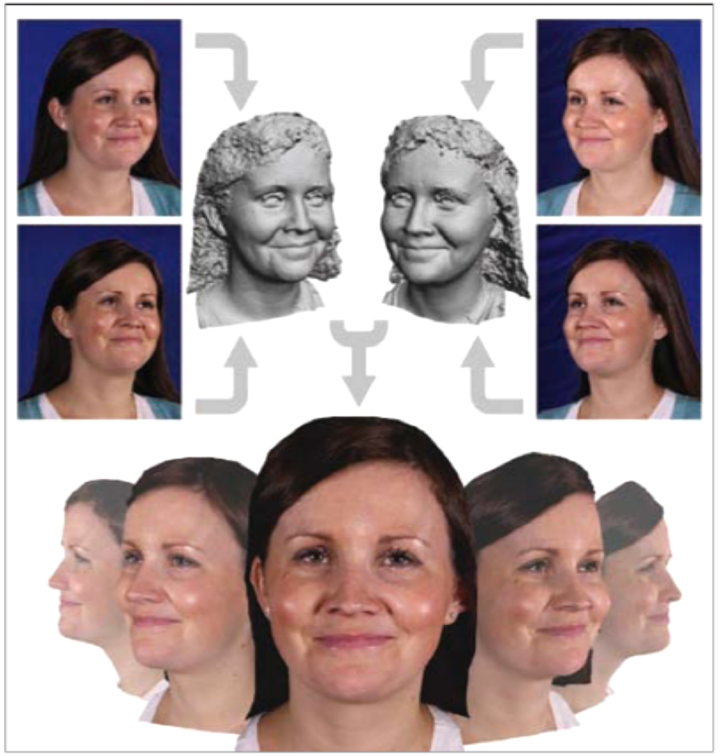
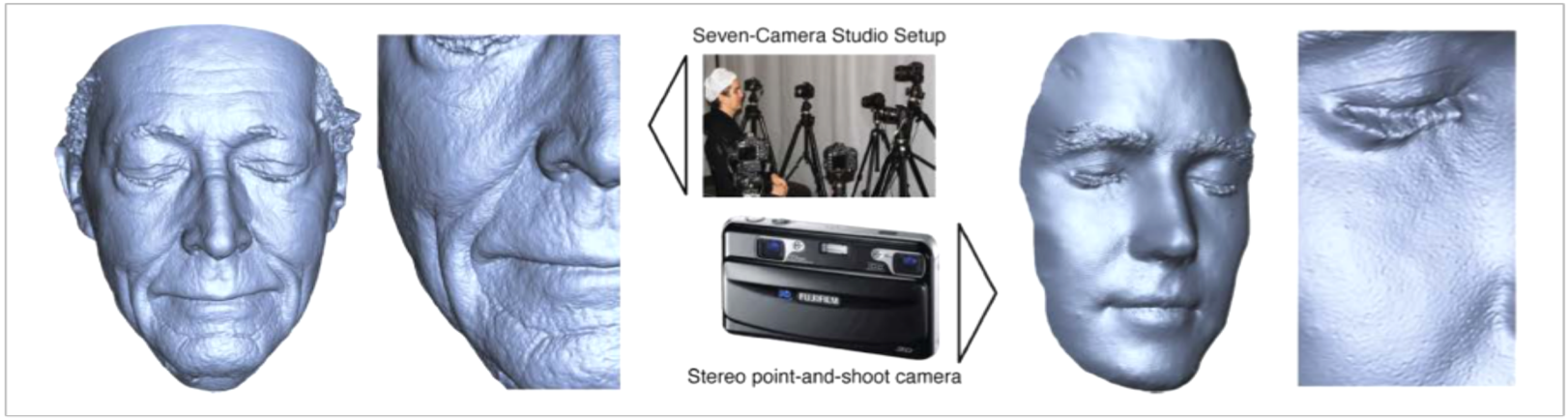
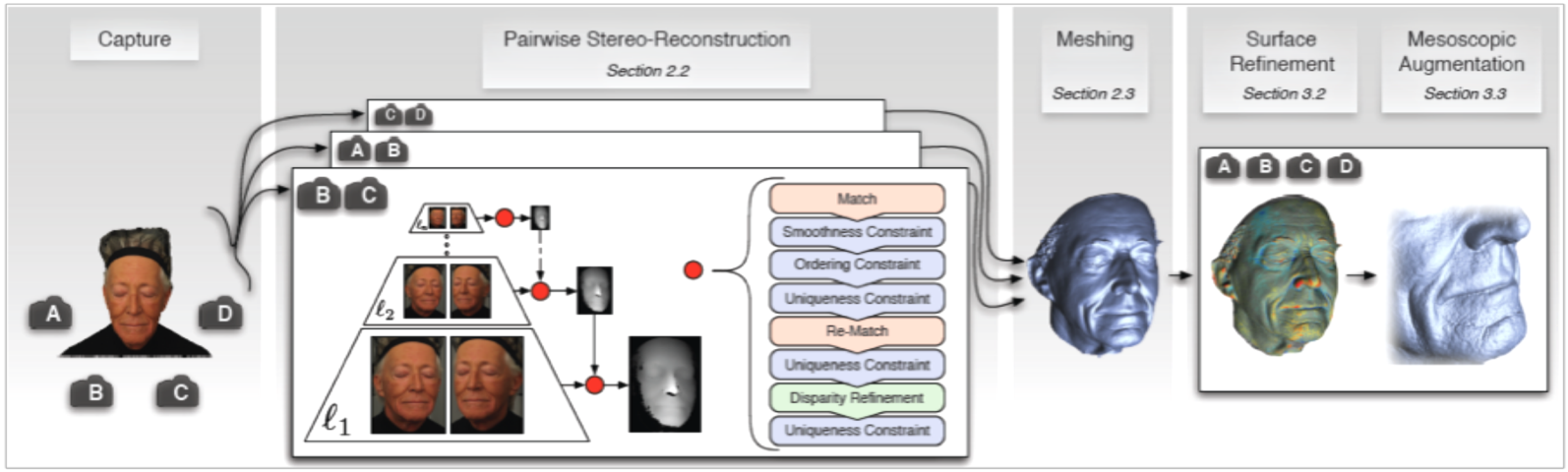


Photo courtesy: Dimensional imaging

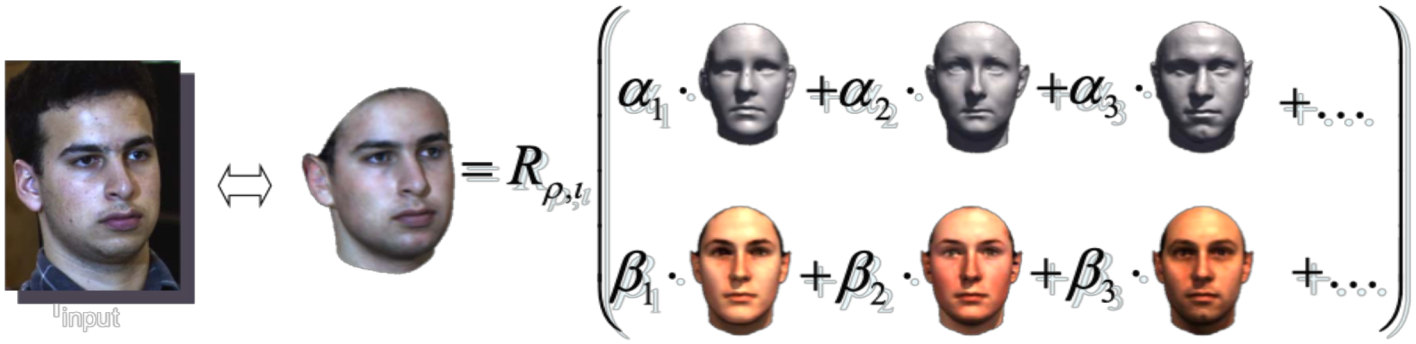
Beeler - Siggraph 2010





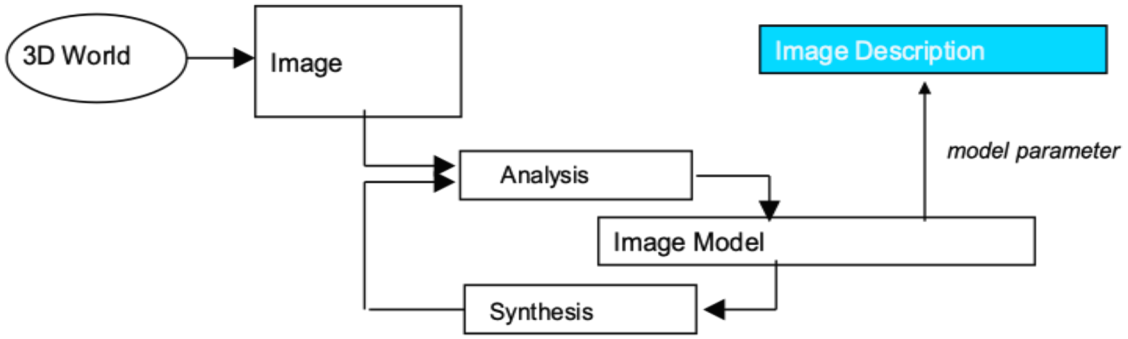
http://www.youtube.com/watch?v=JX5stsU6xfE

3D Face reconstruction from a single 2D image (1)
Fitting 3D morphable model



R = Rendering Function

To match a face, find optimal α, β, ρ, ι → An optimization problem



Analysis by synthesis strategy

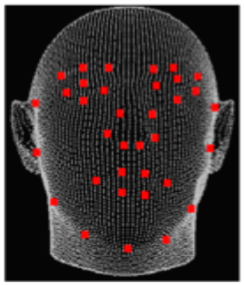
Vetter 2009

3D Face reconstruction from a single 2D image (2)
Fast feature point based method



Neutral Frontal Face

with 2D feature points



Generic model

$$P(s_{3D}, t_x, t_y, f, \alpha, \beta, \gamma \mid S_{2D})$$



Shape model



Model with Texture map

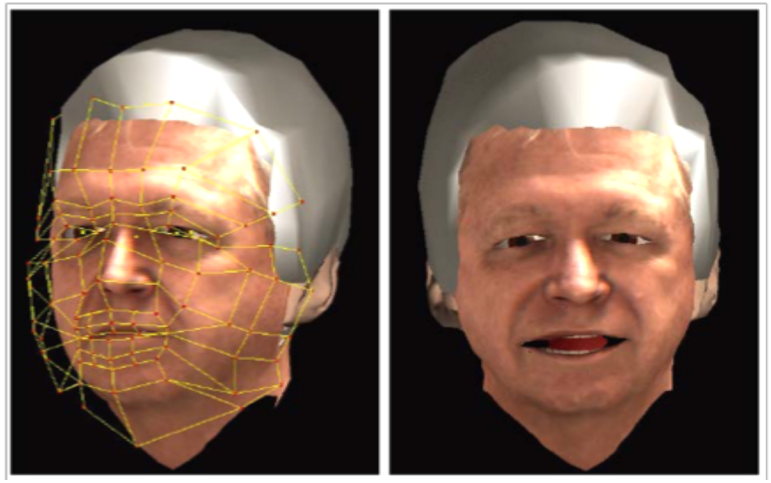
Hu. et. al. FGR2004

3D Facial Deformation Modeling

- Free-form
- Bio-kinetics inspired
- Facial Action Coding System (Ekman & Friesen, 1978)
- Data-driven methods

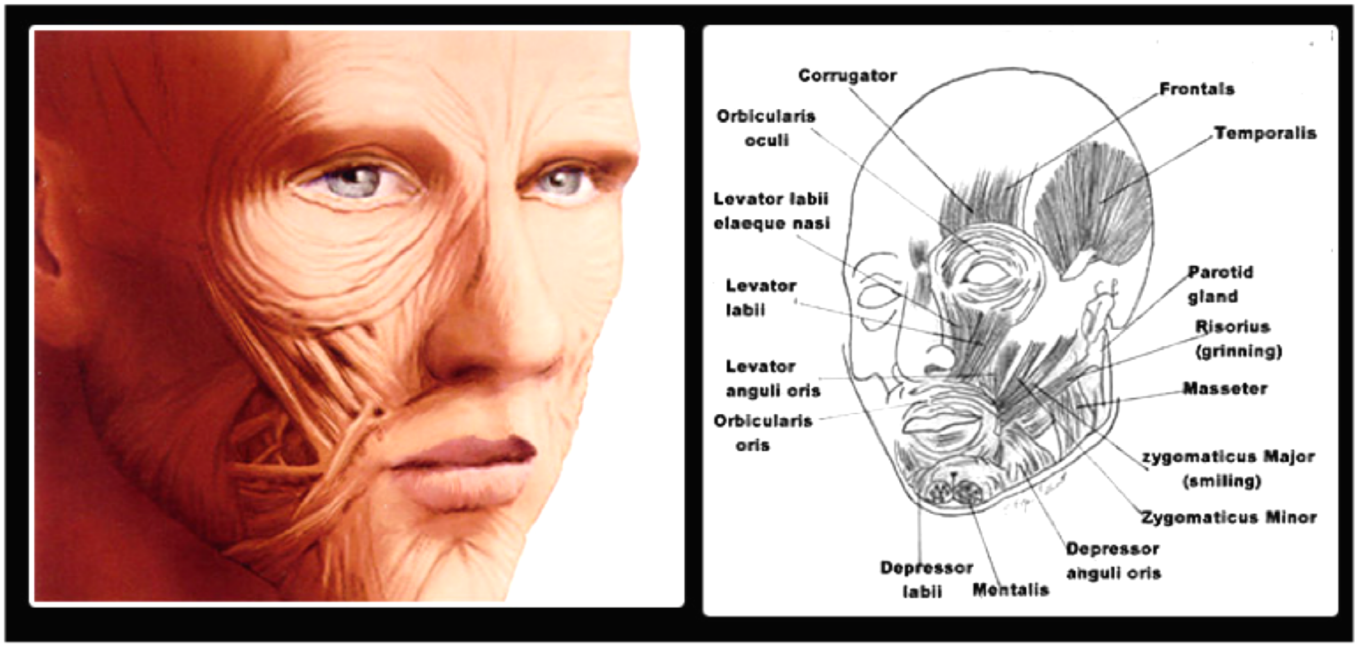
Free-form deformation model

- The coordinates of the mesh vertices can be deformed in a free-form manner by changing the positions of some control points
- Control points can either belong to the mesh vertices or not



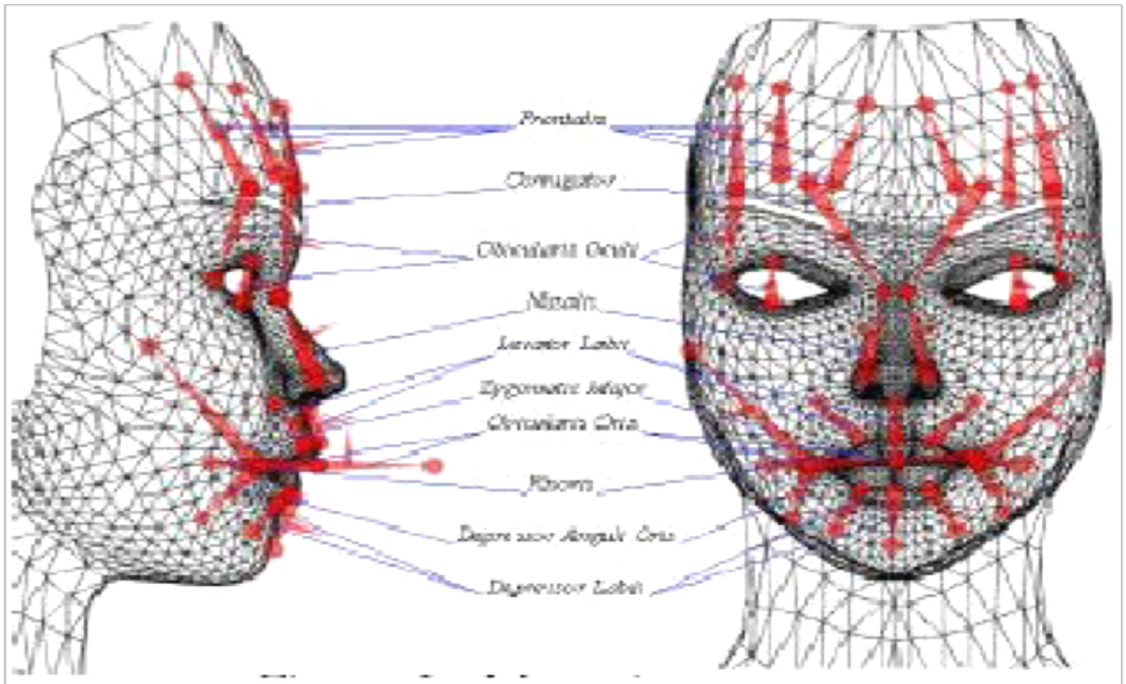
Muscle-based deformation model (1)

- Facial movement are generated by muscles of the face



Muscle-based deformation model

- Build simplified mathematical model that simulates muscle actions on the facial skin



Linear Muscle Models

Facial Action Coding System (FACS)

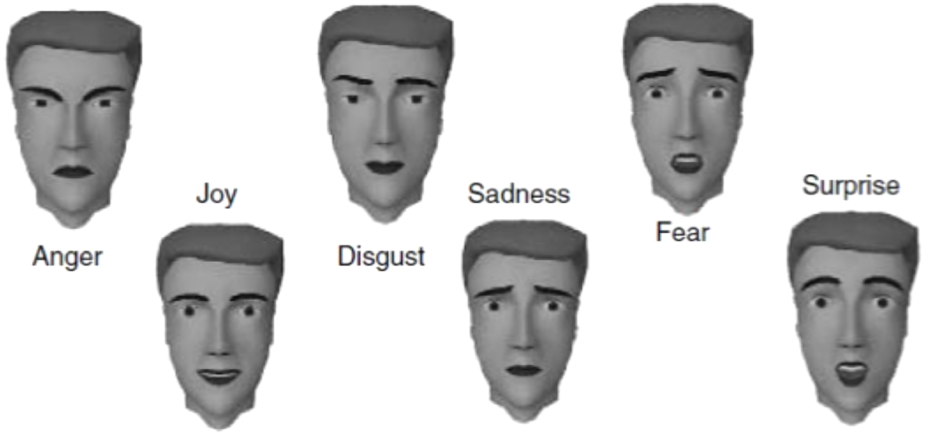
- Developed by Ekman and Friesen, in 1978
- FACS describes facial deformations in terms of "Action Units" (AUs)
  - correspond directly to actions of facial muscles
  - involve to other motions

Such as the movement of the tongue or air filling the cheeks

AU	Description	Facial muscle	Example image
1	Inner Brow Raiser	*Frontalis, pars medialis*	
2	Outer Brow Raiser	*Frontalis, pars lateralis*	
4	Brow Lowerer	*Corrugator supercilii, Depressor supercilii*	
5	Upper Lid Raiser	*Levator palpebrae superioris*	
6	Cheek Raiser	*Orbicularis oculi, pars orbitalis*	
7	Lid Tightener	*Orbicularis oculi, pars palpebralis*	
9	Nose Wrinkler	*Levator labii superioris alaquae nasi*	
10	Upper Lip Raiser	*Levator labii superioris*	
11	Nasolabial Deepener	*Zygomaticus minor*	

Facial Action Coding System (FACS)

AUs may be combined to describe *facial expressions*



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

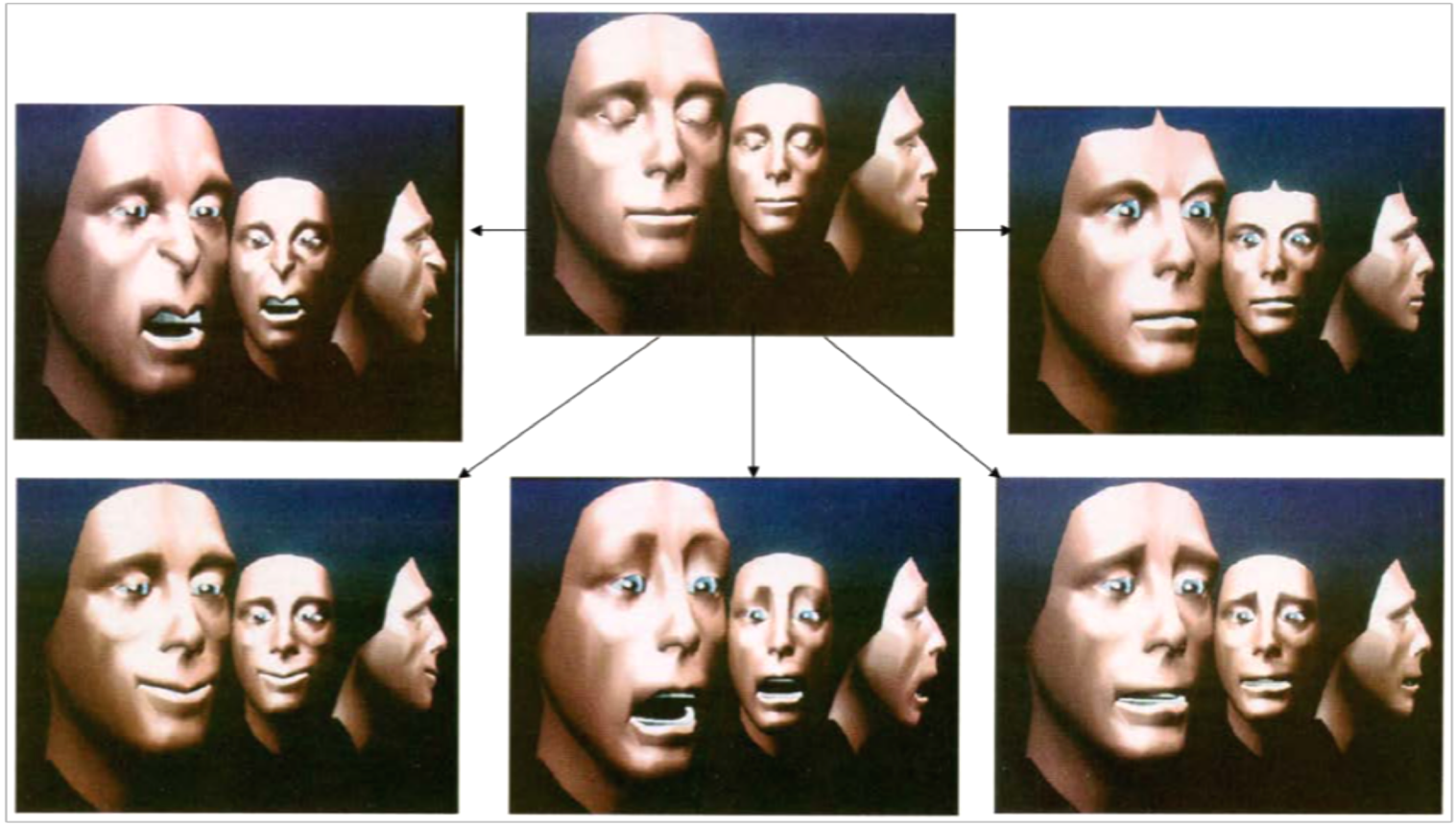
MPEG4 Facial Animation Parameters (FAPs)

- MPEG4 defines 68 FAPs, categorized into 10 groups

Table 2.2 FAP groups

Group	Number of FAPs
1. Visemes and expressions	2
2. Jaw, chin, inner lowerlip, cornerlips, midlip	16
3. Eyeballs, pupils, eyelids	12
4. Eyebrow	8
5. Cheeks	4
6. Tongue	5
7. Head rotation	3
8. Outer-lip positions	10
9. Nose	4
10. Ears	4

Waters, SIGGRAPH 1987



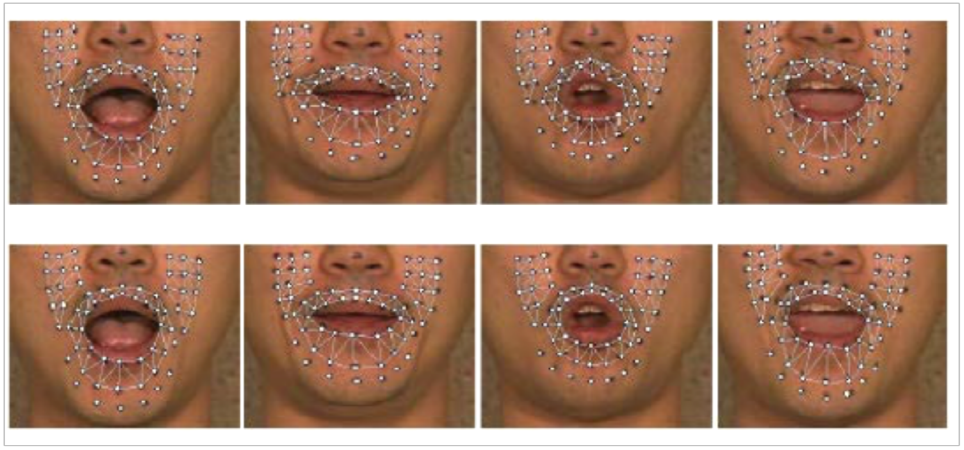


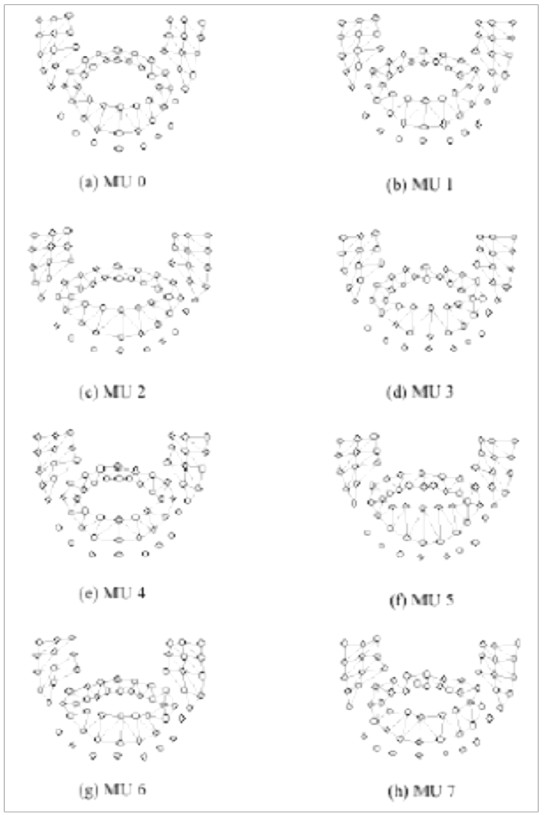
Piecewise Bezier Volume Deformation Model (Tao and Huang, 1998)



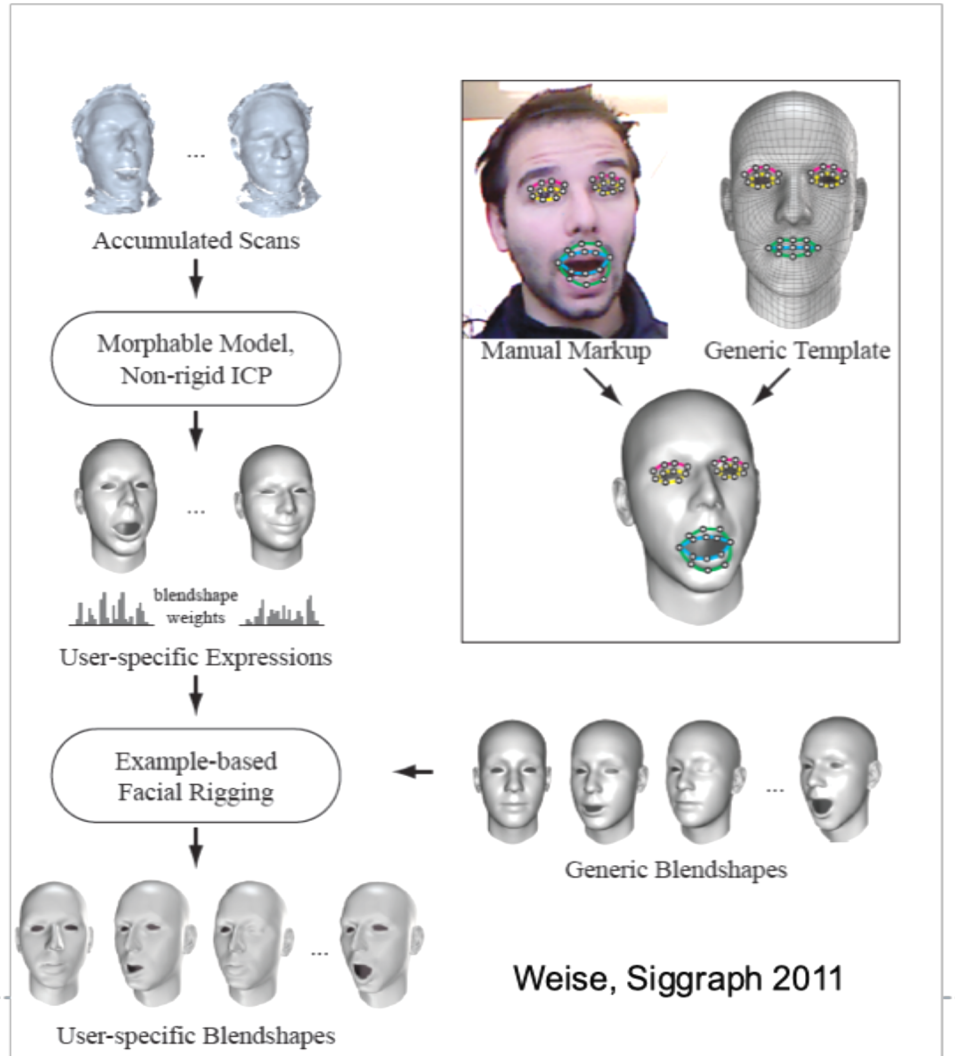
Data driven methods - Motion Units

Learn the basic facial deformations from motion capture data (Hong, Wen, and Huang, 2001)





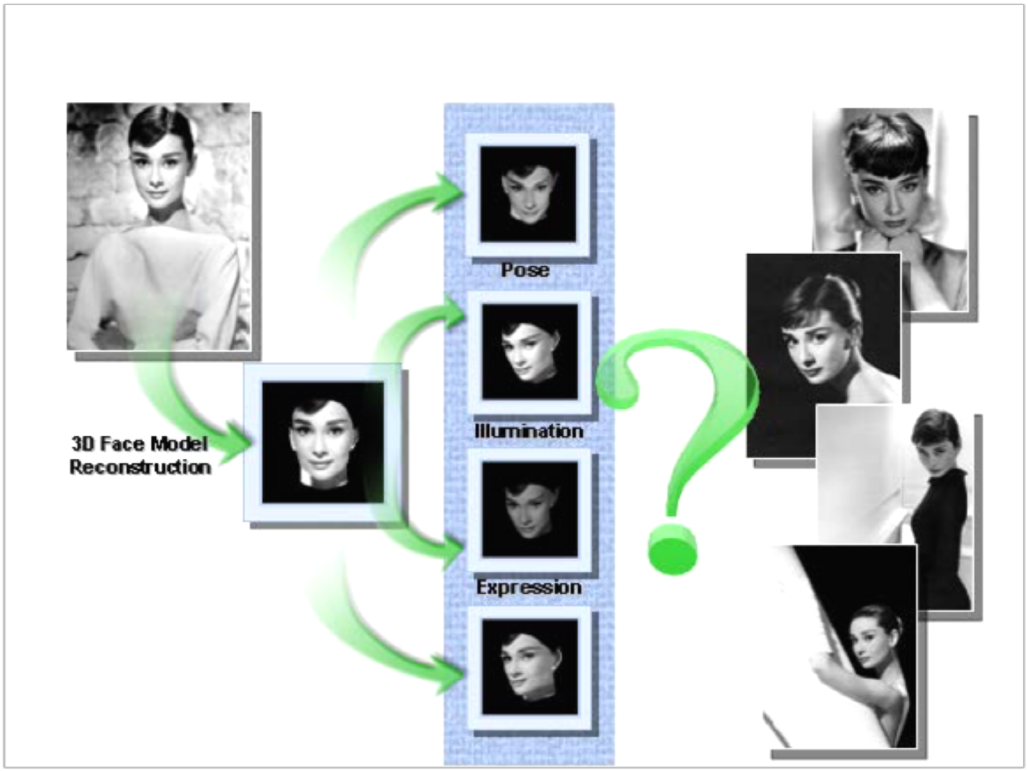
User specific deformation model



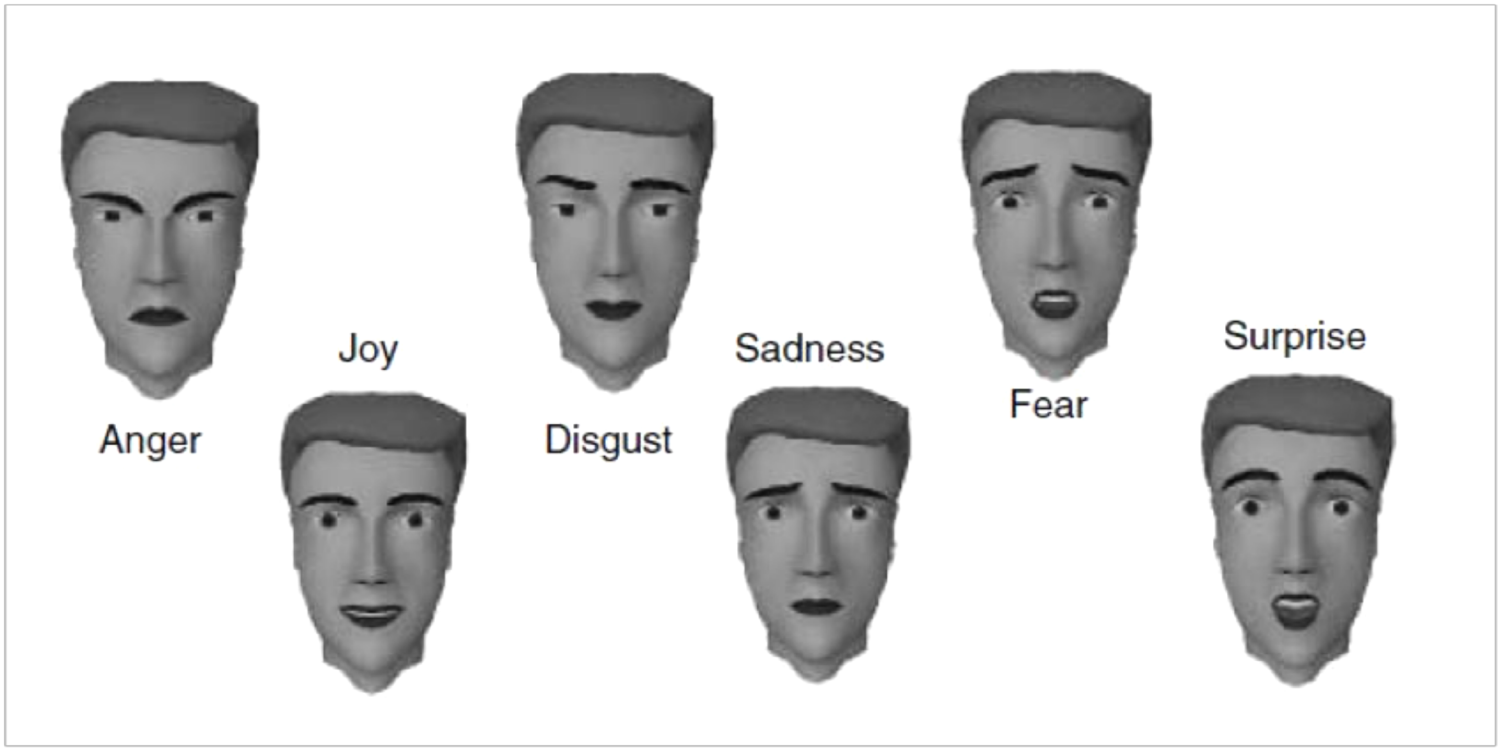
Applications

- Face recognition
- Face tracking
- Emotion recognition
- Audiovisual speech recognition
- Talking avatar

Face recognition



Face tracking and Expression recognition



Audio-Visual Speech Recognition

(Sari, Interspeech 2018)

Phonemes and *Visemes*

- A *Phoneme:* basic acoustic unit of sound
- A *Viseme:* Visual counterpart of a phoneme

Viseme #	Phonemes	Example
0	None	na
1	p, b, m	put, bed, mill
2	f, v	far, voice
3	T,D	think, that
4	t, d	Tip, doll
5	k, g	call, gas
6	tS, dZ, S	chair, join, she
7	s, z	Sir, zeal
8	n, l	Lot, not
9	r	red
10	A:	car
11	e	bed
12	I	Tip
13	Q	top
14	U	book

Audio-Visual Speech Recognition

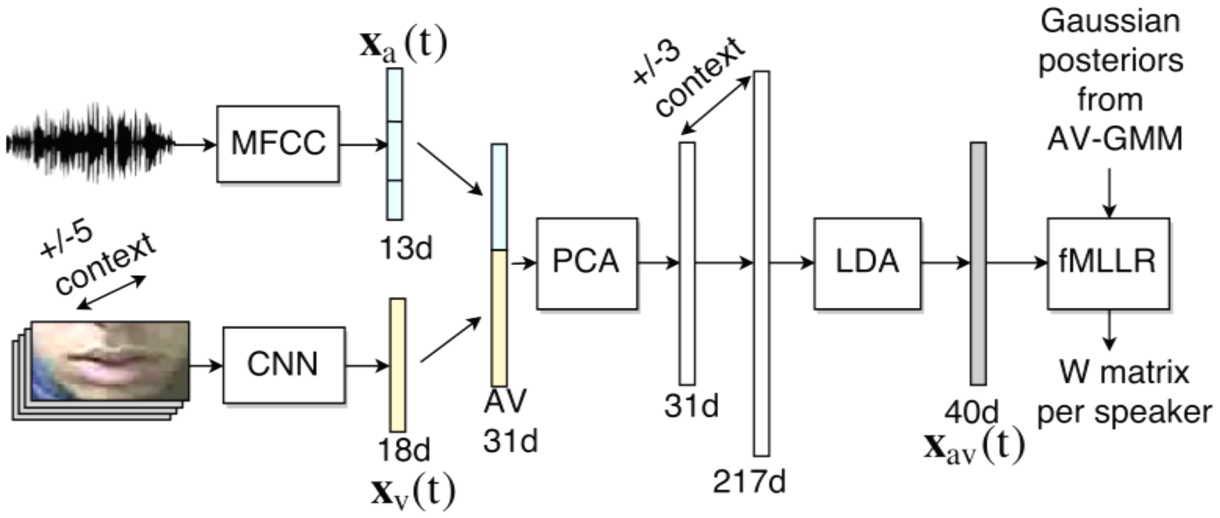
(Sari, Interspeech 2018)

Speaker Adaptation - Early Fusion

Transformed features: $\hat{\mathbf{o}}(t) = \mathbf{W}\zeta_{av}(t)$ where $\zeta_{av}(t) = [\mathbf{x}_{av}(t);\ 1]$

$$\mathbf{w}_i = (\alpha \mathbf{p}_i + \mathbf{k}_i)\mathbf{G}_i^{-1}$$

$$\mathbf{G}_i = \sum_m \frac{1}{\sigma_i^{(m)2}} \sum_t \gamma_{av}^{(m)}(t)\zeta_{av}(t)\zeta_{av}(t)^T. \quad (1)$$



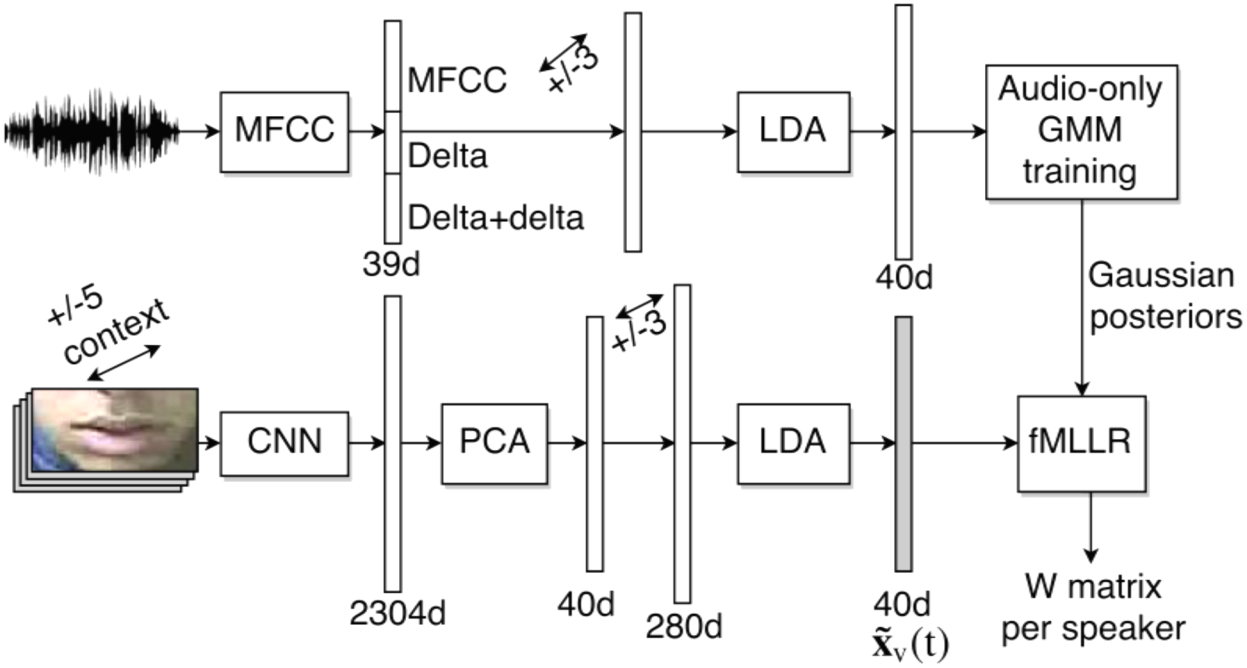
Audio-Visual Speech Recognition

(Sari, Interspeech 2018)

Speaker Adaptation - Implicit Fusion

Transformed features: $\hat{\mathbf{o}}(t) = \mathbf{W}'\zeta_v(t)$ where $\zeta_v(t) = [\tilde{\mathbf{x}}_v(t);\ 1]$

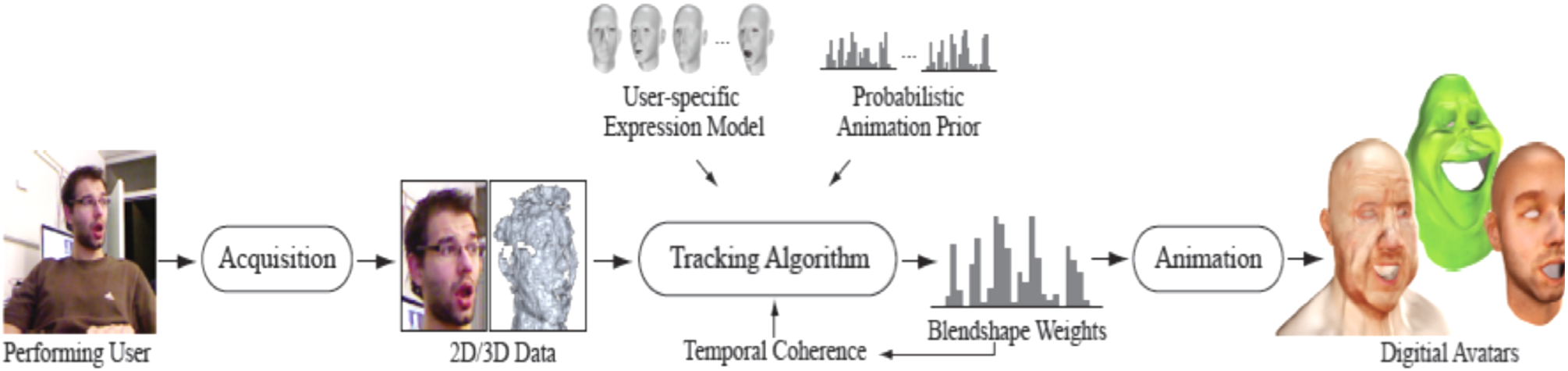
$$\mathbf{G}'_i = \sum_m \frac{1}{\sigma_i^{(m)2}} \sum_t \gamma_a^{(m)}(t)\zeta_v(t)\zeta_v(t)^T \qquad (2)$$



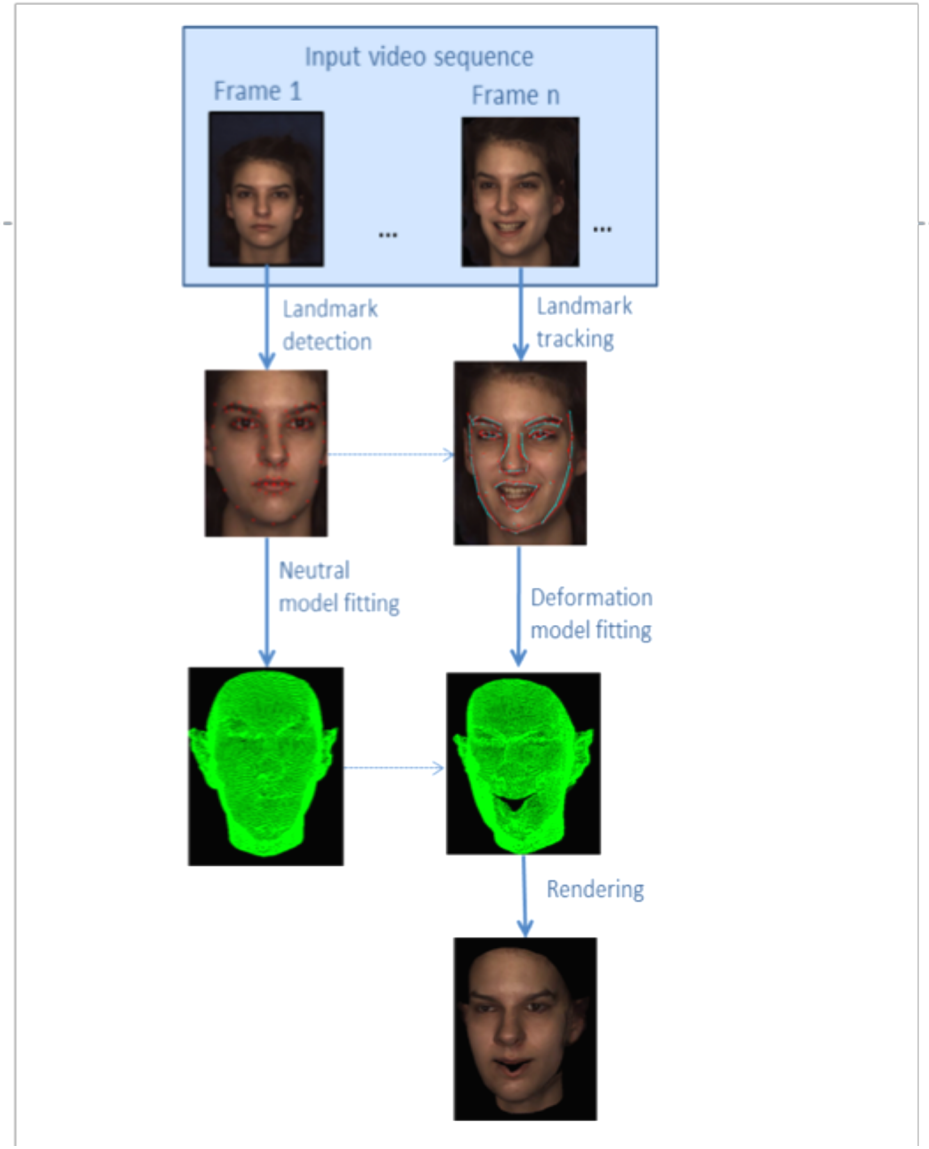
3D facial animation - Avatar

- Key-frame interpolation method
  - Place particular facial deformations at particular time instants (key-frames)
  - Facial deformations in-between key-frames are obtained by a certain interpolation scheme
- Two basic key-frame types
  - Visemes
  - Expressions

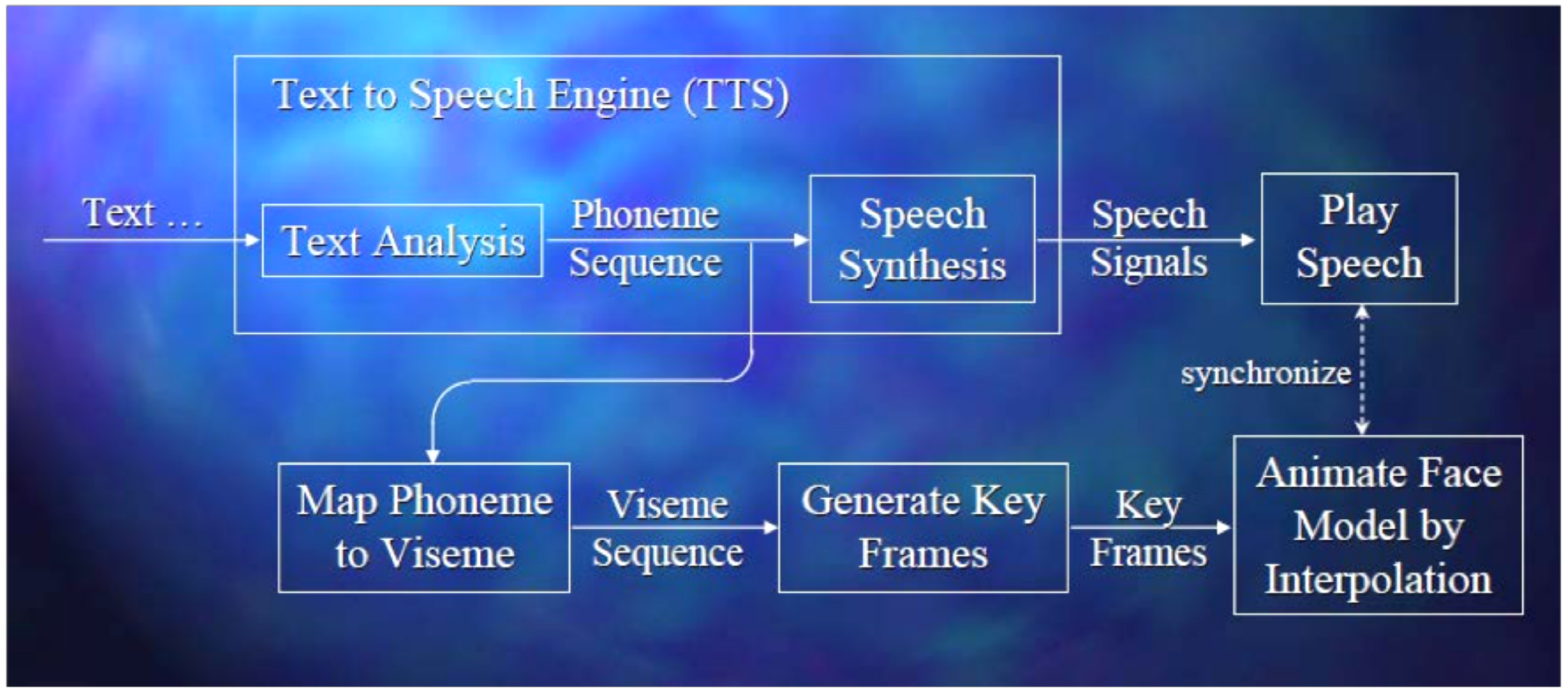
Weise, Li – Faceshift - Siggraph 2011



Performance driven avatar



Text-driven talking heads



Speech-driven talking heads

