A Vision-based Approach for Facial Expression Cloning by Facial Motion Tracking

J.C Chun

http://giplab.kyonggi.ac.kr

Computer Graphics & Image Processing Lab.

Dept. of Computer Science, Kyonggi University, South Korea

Agenda

- Motivation
- Previous Works
- Research Goals
- Application Area
- Related Works
- Proposed System
  - Head Pose Estimation
  - Facial Expression Control
- Experimental Results
- Concluding Remarks

Motivation

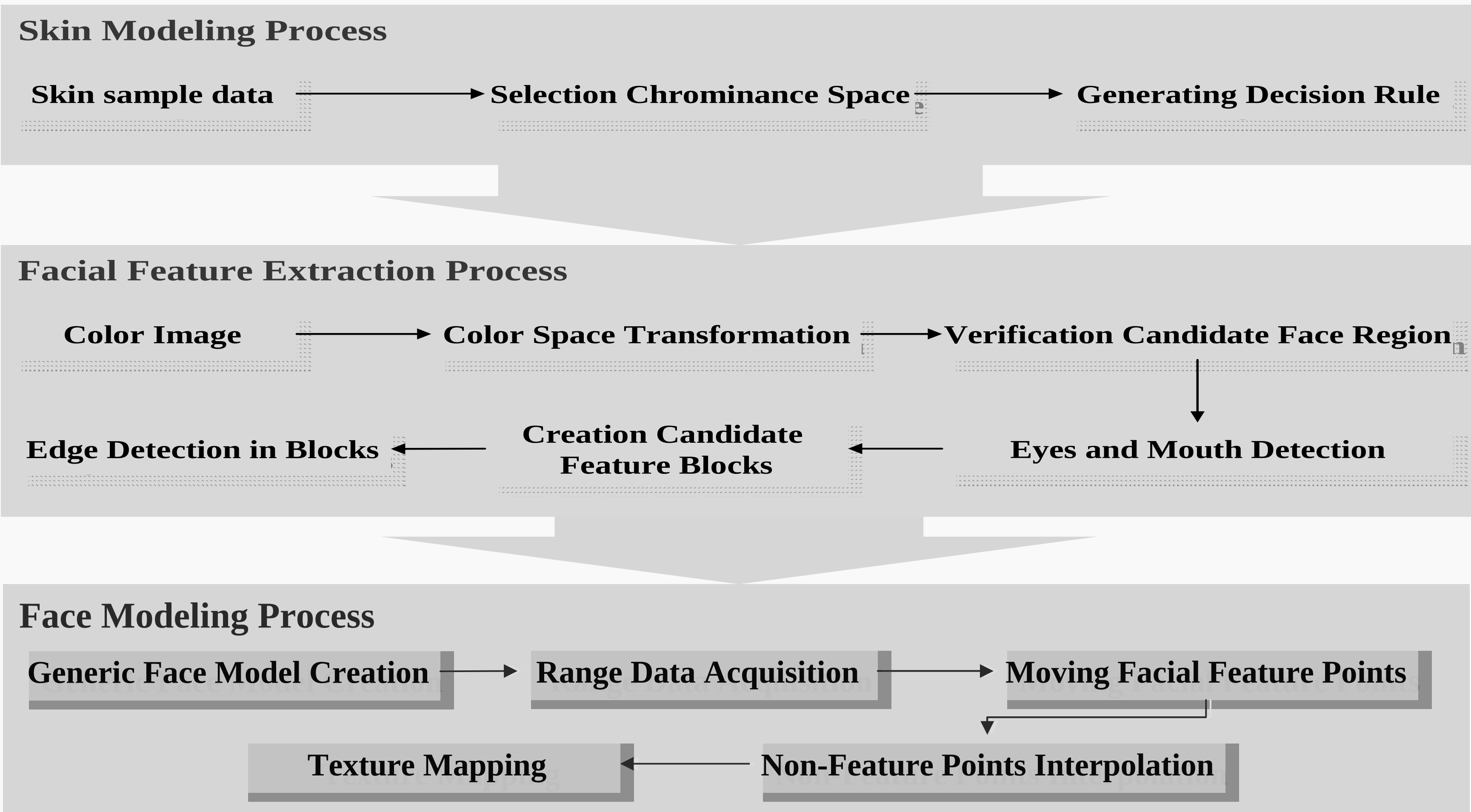
- The Convergence Technologies of Computer Graphics, Computer Vision and Human Computer Interaction are widely used.

- Several research woks have been done and processed for years in GIP at Konggi University especially in the applications of ``human face”.

  - Face Recognition using Wavelet transform and Eigenface.
  - Automatic 3D Face modeling from 2D images.
  - Facial Expression Recognition from input video images for emotion care services.
  - Vision-based facial motion tracking and 3D face animation.
  - Markerless augment reality using facial and hand motion tracking.

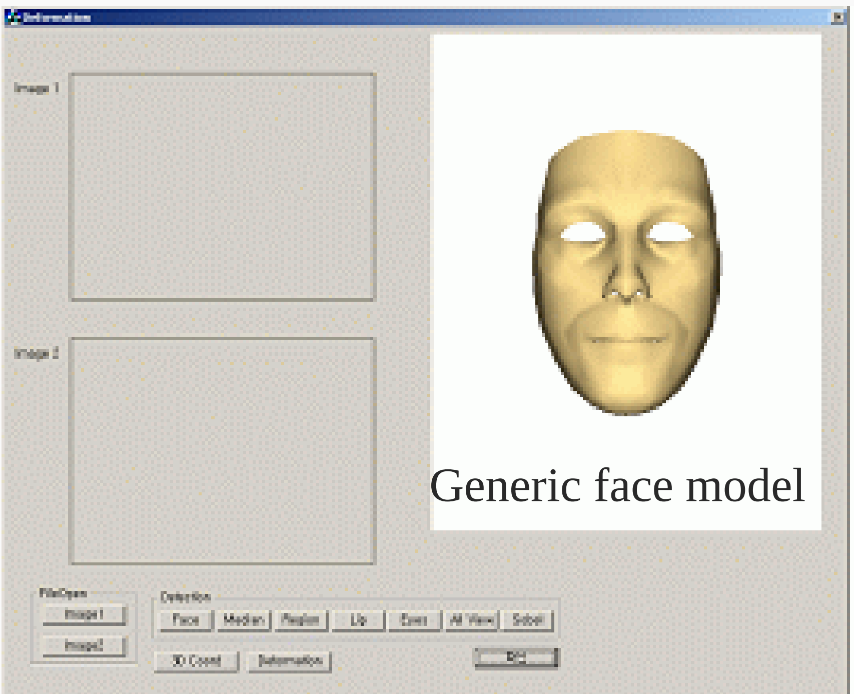
Previous Work(1)

- 3D Face modeling from 2D images.

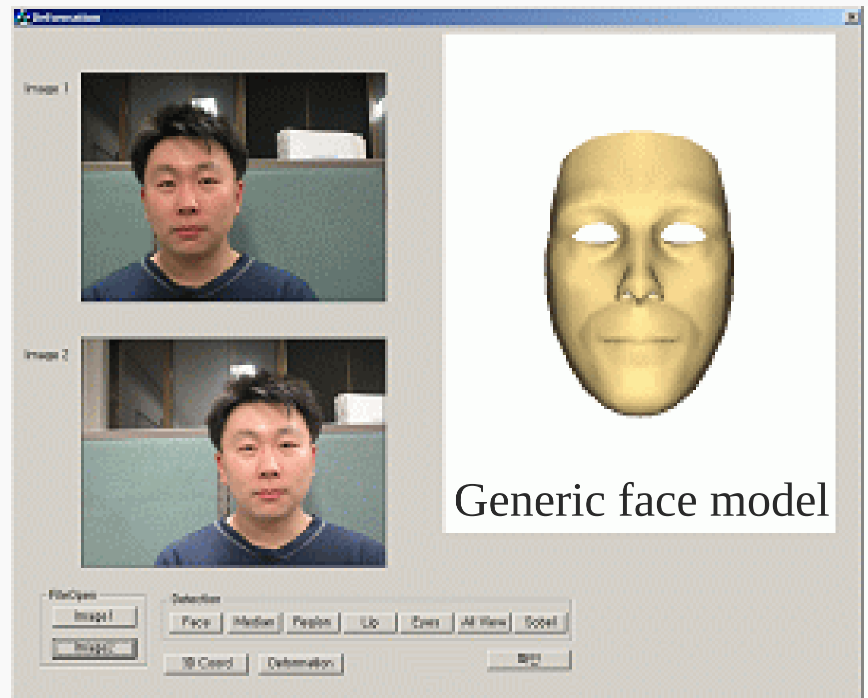


Previous Work(2)

- Use two frontal face images and a generic 3D model.

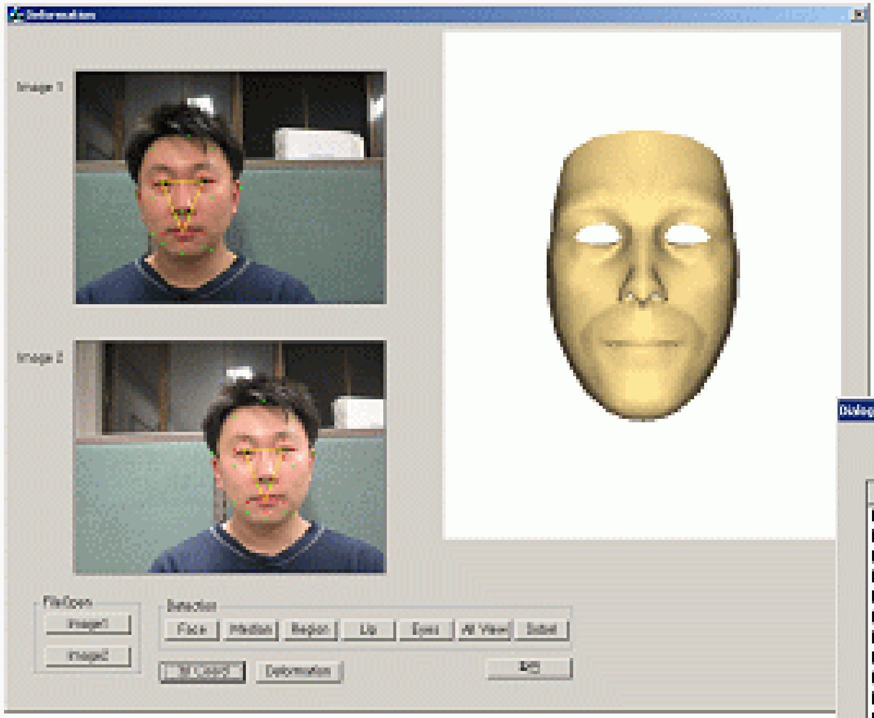


[User interface]



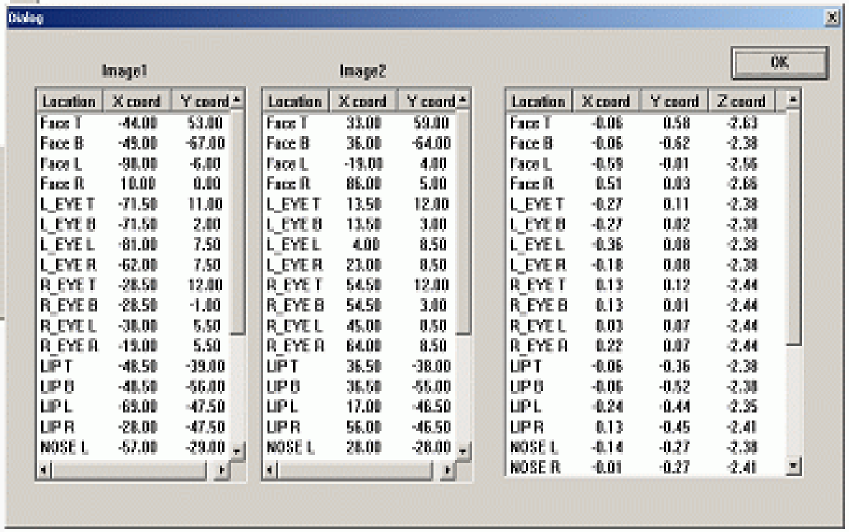
[Stereo images input]

Previous Work(3)



Facial feature detection

[3D coordinates extraction]



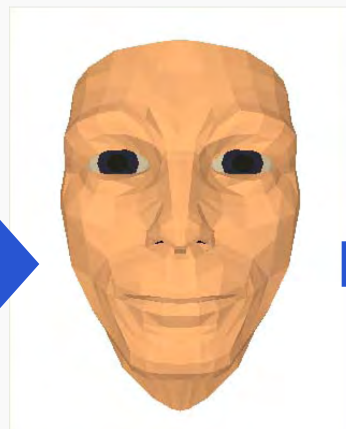
Previous Work(4)

[3D model generation]

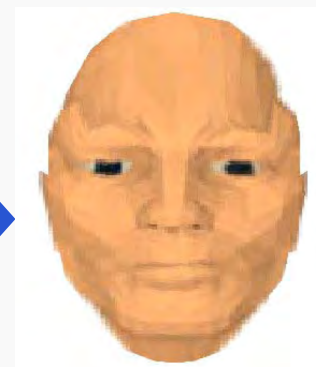
Previous Work(5)

Front image

Side image



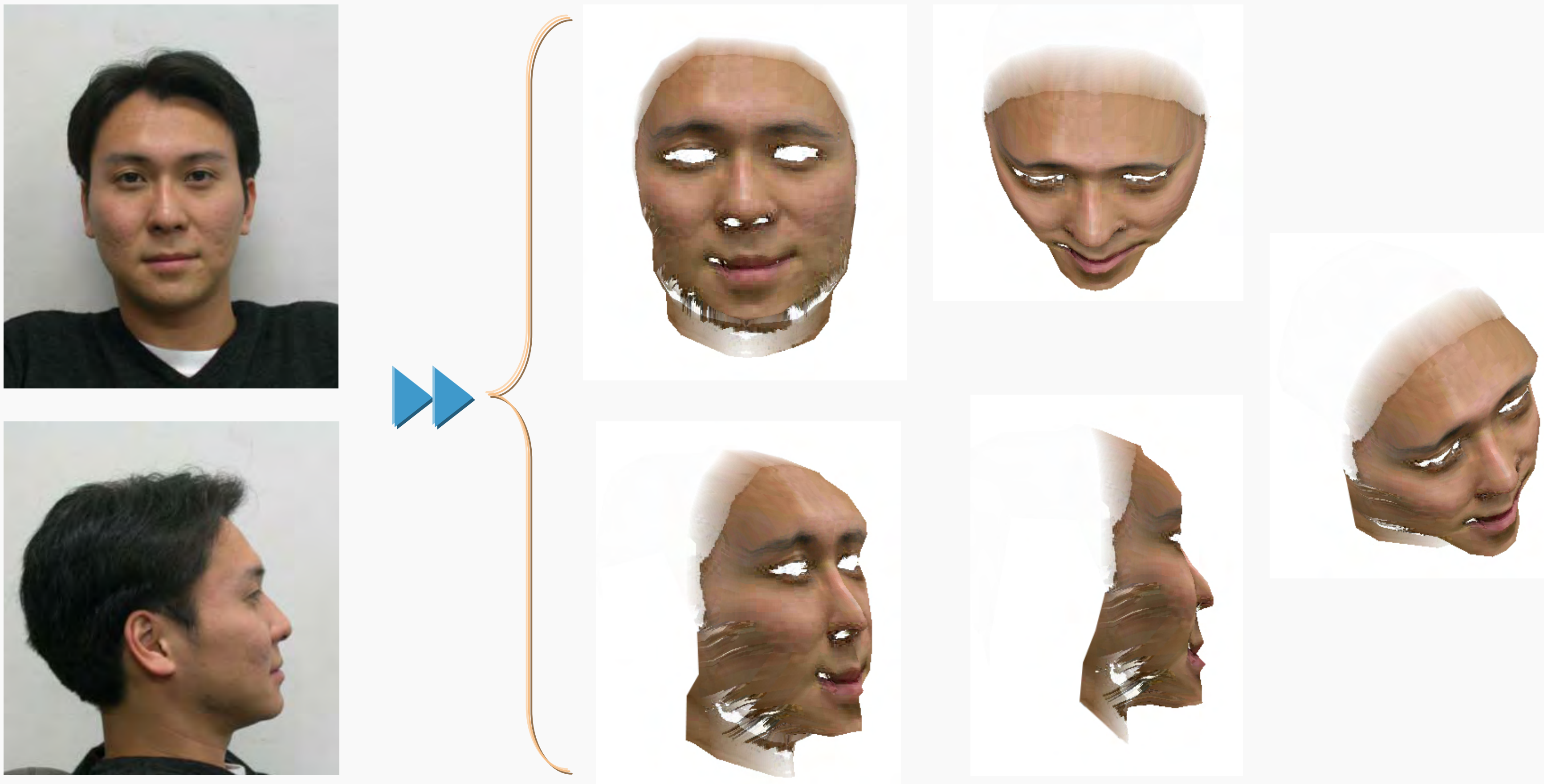
Generic face model



Deformed face model

Generated 3D face model

Previous Work(6)



Previous Works(7)

- Facial expression recognition for emotion care services.

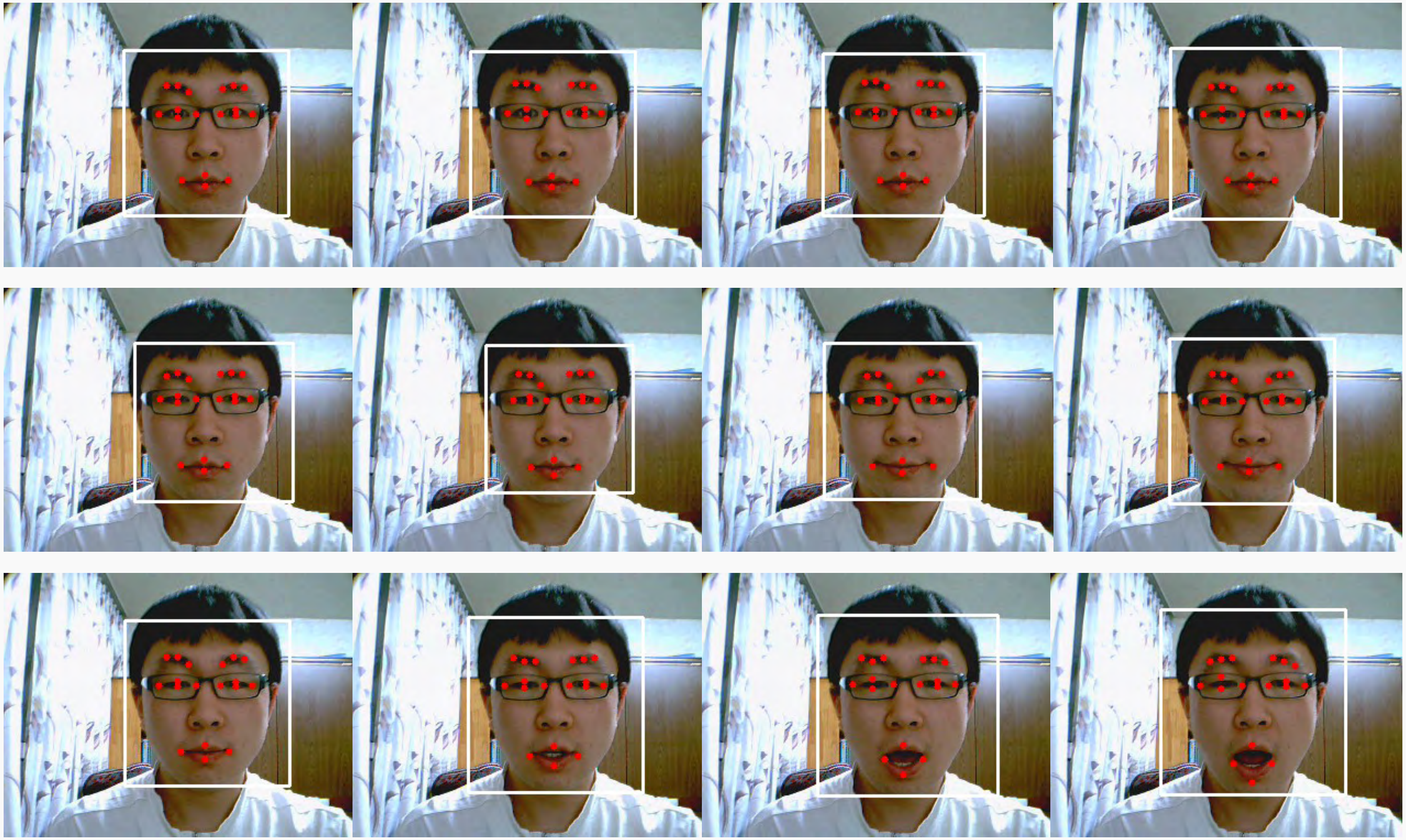
Sad → Surprise → Happy

Previous Works(8)

Happy → Surprise → Neutral

Previous Works(9)

Facial Feature Tracking



Research Goals

- The goal of the work is to develop a vision-based facial model animation system.
- The sub-goals are to estimate 3D head pose from a sequence of input images and to retarget facial expression from the video input to 3D face model for human computer interaction (or vision-based animation).
- The exact head pose estimation and facial motion tracking are critical problems to be solved in developing a vision based human computer interaction system.

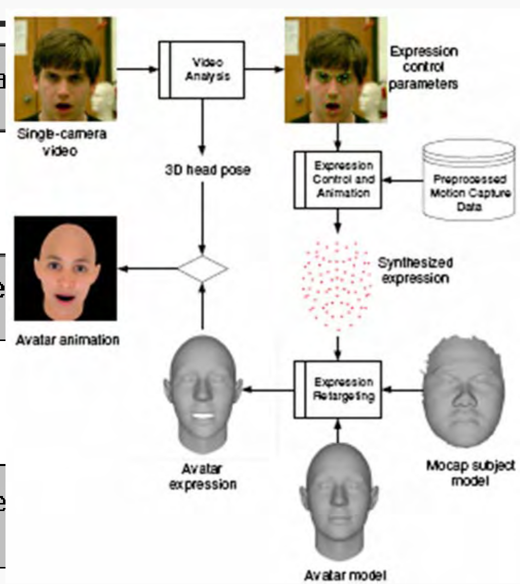
Application Areas

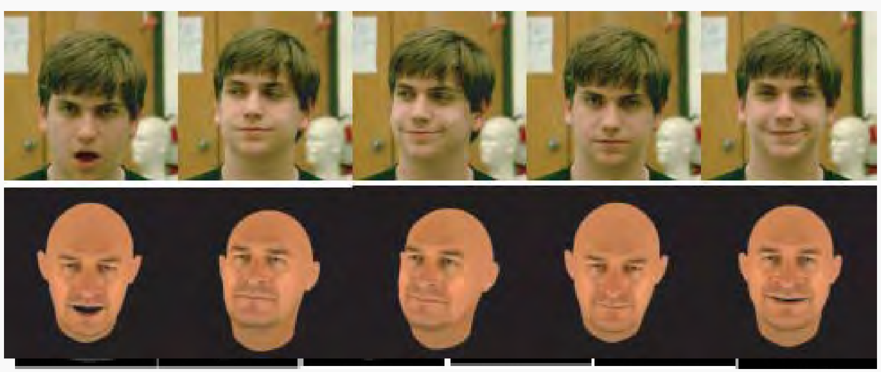
- **R**ealistic animated facial modeling and facial expression control of a 3D face model have been important research fields for diverse application areas such as virtual character animation for entertainment, 3D avatars in the internet, and 3D teleconferencing.

  - Vision-Based HCI
  - Virtual Reality (Augmented Reality)
  - 3D Animation & Game

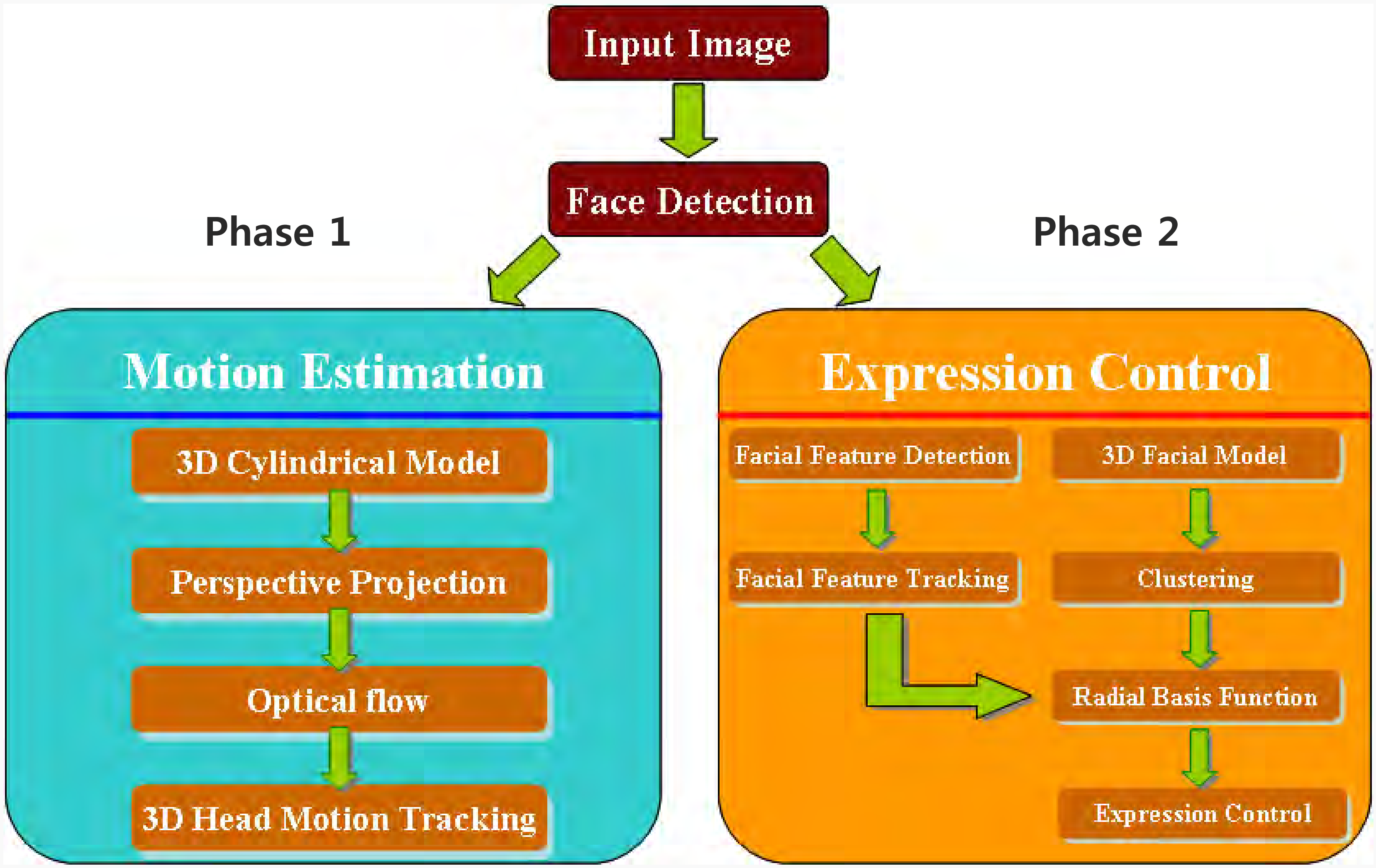
Related Work

Title	Method
Vision-Based Control of 3D Facial Animation [J.X. Chai '2003] CMU	Optical flow & Motion Capture
A Real Time Face Tracking And Animation System [X. Wei '2004] NYU	Geometric Feature & Interpolation





Proposed System



Phase 1: Motion Estimation

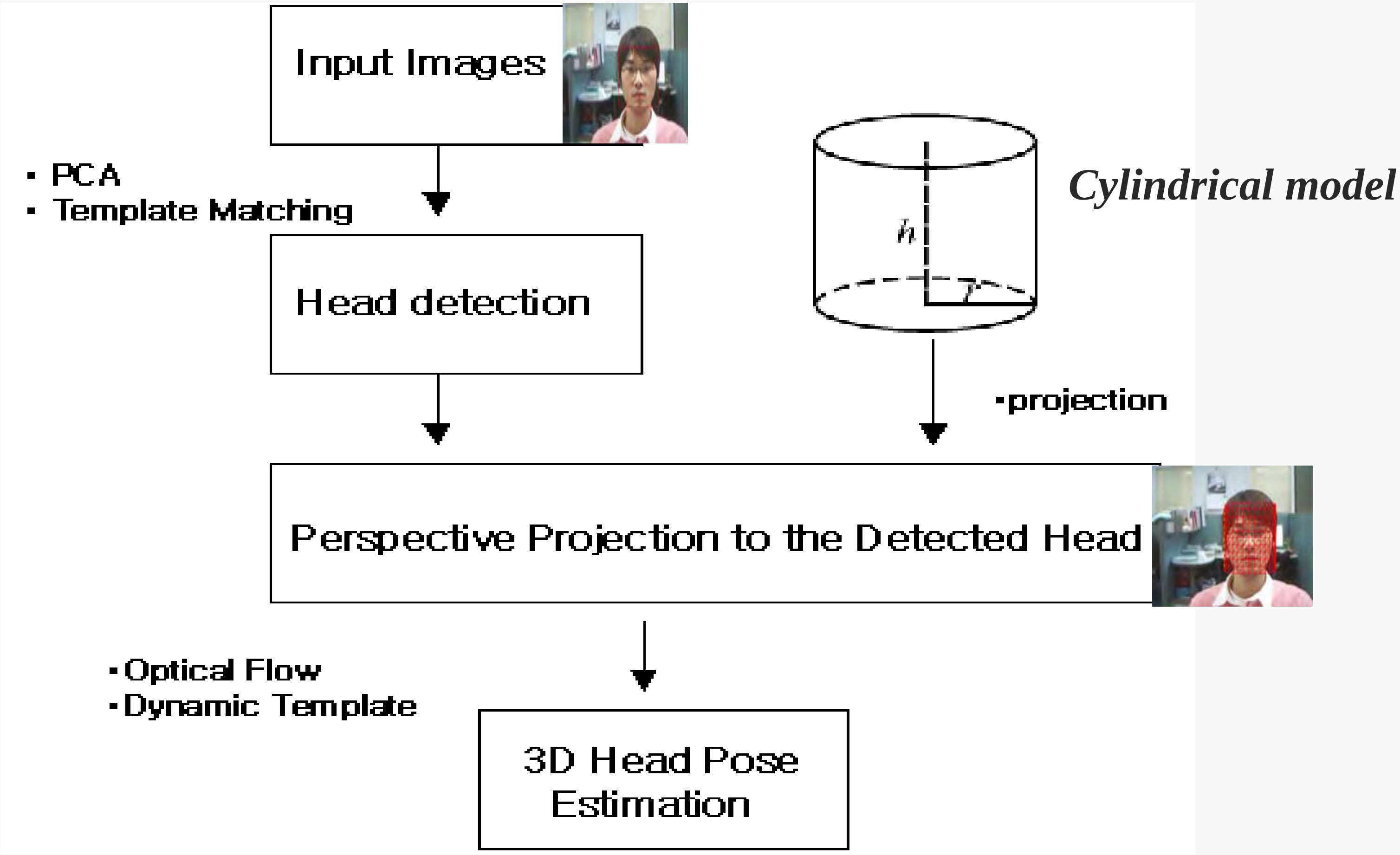
- Why estimate motion?
  - Track object behavior
  - Correct for camera jitter (stabilization)
  - Align images (mosaics)
  - 3D shape reconstruction
  - Special effects

- Given an initial reference template of head image and corresponding head pose, full the head motion is recovered by using a cylindrical head model.

- By updating the template dynamically, it is possible to recover head pose robustly regardless of light variation and self-occlusion.

Head Pose Estimation (1)

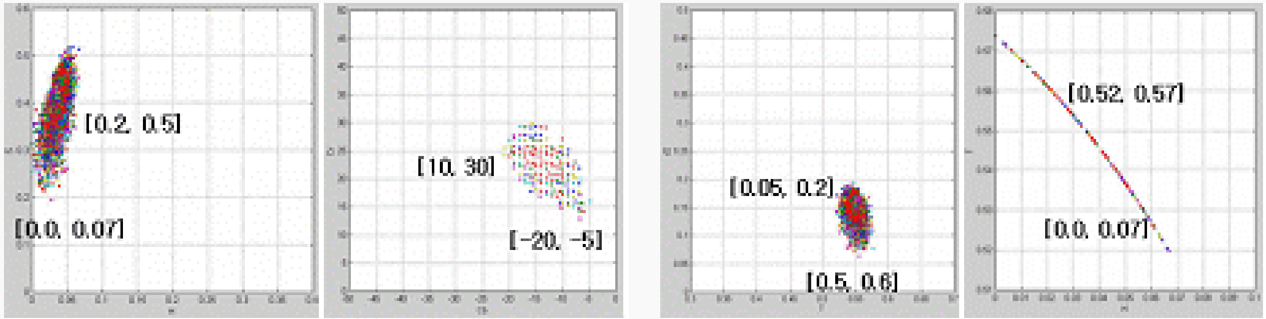
Overall Procedure



Head Pose Estimation (2)

Face Detection(1)

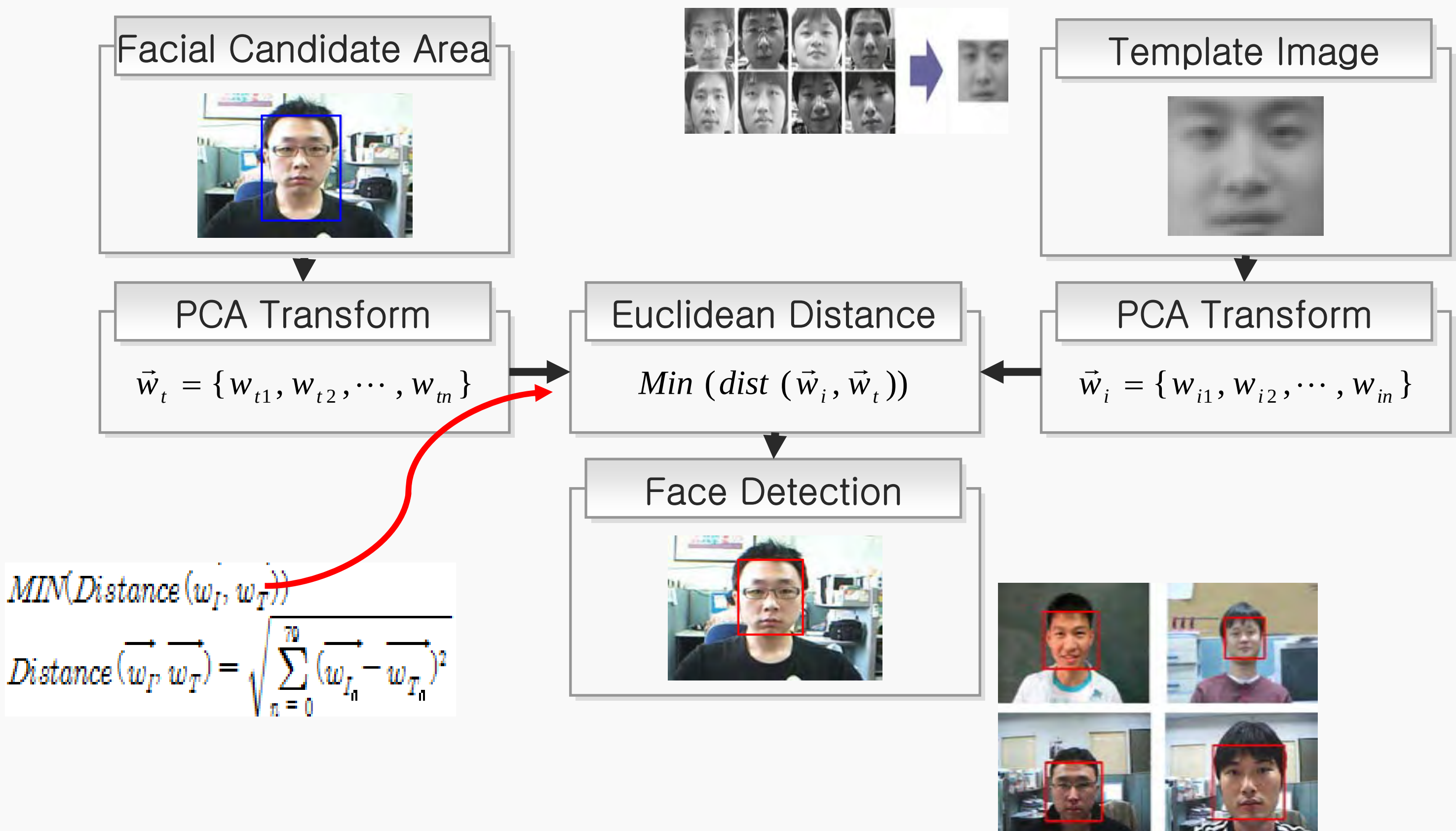
- Color information is efficient for identifying skin region.
- We propose nonparametric HT skin color model to detect facial area efficiently rather than using other parametric skin color model



Facial Color Distribution(H-S, Cb-Cr, T-S, H-T)

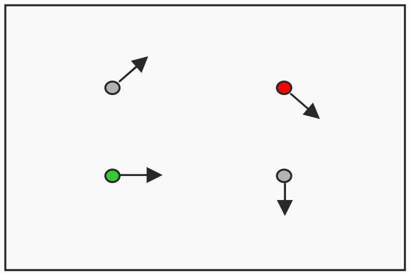
Head Pose Estimation (3)

- Face Detection(2): by template matching.
- Compare the eigen vectors of template image and those of candidate facial region are compared to detect face.

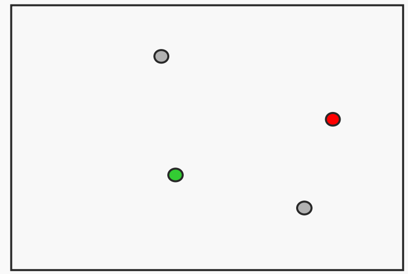


Head Pose Estimation (4)

Optical Flow(1)



$H(x, y)$

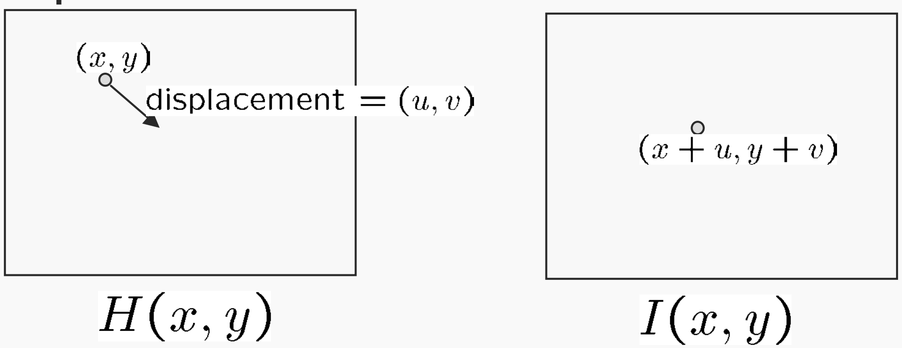


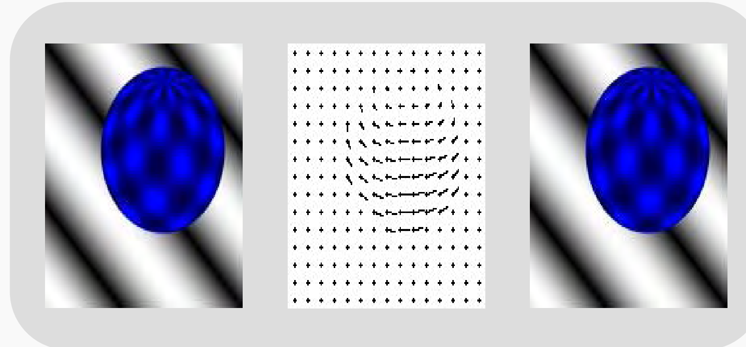
$I(x, y)$

- How to estimate pixel motion from image H to image I?
  - Solve pixel correspondence problem
    - given a pixel in H, look for nearby pixels of the same color in I
- Key assumptions
  - color constancy: a point in H looks the same in I
  - **small motion**: points do not move very far
- This is called the **optical flow** problem

Head Pose Estimation (5)

Optical Flow(2)





- Let's look at these constraints more closely
  - brightness constancy: Q: what's the equation?

    H(x,y)=I(x+u, y+v)
  - small motion: (u and v are less than 1 pixel)
    - suppose we take the Taylor series expansion of I:

$$I(x+u, y+v) = I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order terms}$$

$$\approx I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

Head Pose Estimation (6)

Optical Flow(3)

- If an image I(u,t) at time t where u=(x,y) is a pixel in the image is given, at t+1 , u moves to u'=(F, μ) , where μ is the motion parameter vector and (F, μ) is the parametric model, which maps u to the new position u' .

- The motion vector μ can be obtained by minimizing following function when the illumination condition is unchanged.

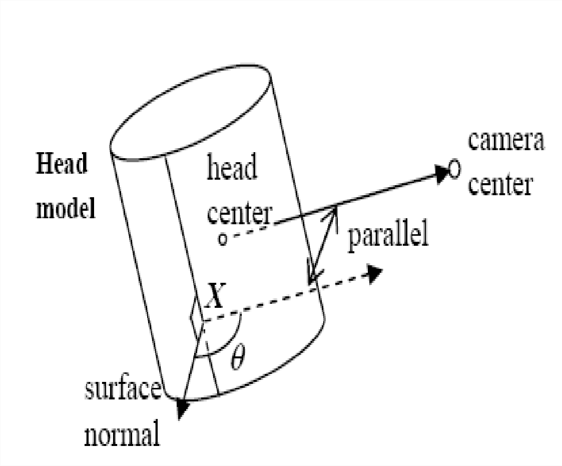
$$\min E(\mu) = \sum_{u \in \Omega} (I(F(u,\mu), t+1) - I(u,t))^2$$

- Where Ω is the region of template at t. By using Lucas-Kanade method, the problem of equation above can be solved as follows:

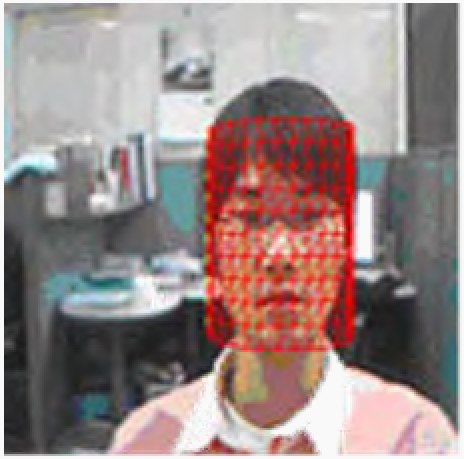
$$\mu = -\left(\sum_{\Omega} (I_u F_\mu)^T (I_u F_u) \right)^{-1} \sum_{\Omega} (I_t (I_u F_u)^T)$$

Head Pose Estimation (7)

- Perspective Projection



Cylindrical Model
to be projected



Model Projected to
Face Image

Head Pose Estimation (8)

- To present the geometry of the entire head, 3D cylindrical model is projected to the input face and the head pose is estimated using the projected face model.

- If the location of the head pose at t is $X = [x, y, z]^T$ then the locations of the head pose at $t+1$ become

$$X(t+1) = M \bullet X(t) = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \bullet X(t)$$

$$M = \begin{bmatrix} 1 & -\omega_z & \omega_y & t_x \\ \omega_z & 1 & -\omega_x & t_y \\ -\omega_y & \omega_x & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Head Pose Estimation (9)

- Then the image projection μ of $X=[x,y,z]^T$ at *t+1* can be defined

$$u(t+1)=\begin{bmatrix} x-y\omega_z+z\omega_y+t_x \\ x\omega_z+y-z\omega_x+t_y \end{bmatrix}\cdot\frac{f_L}{-x\omega_y+y\omega_x+z+t_z}$$

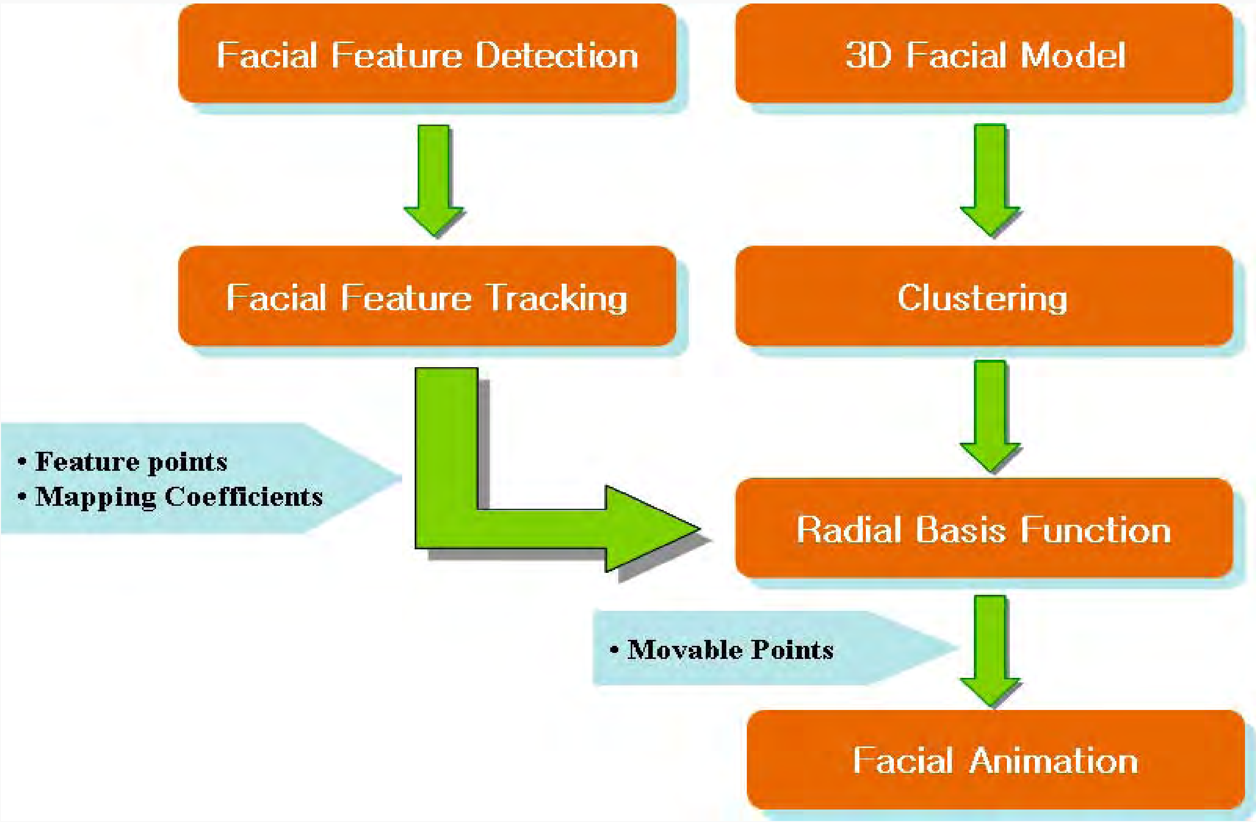
- The motion model $F(u,\mu)$ with the parameter $\mu=\lfloor\omega_x,\omega_y,\omega_z,t_x,t_y,t_z\rfloor$ can be defined by

$$F_\mu\Big|_{\mu=0}=\begin{bmatrix} -xy & x^2+z^2 & -yz & z & 0 & -x \\ -(y^2+z^2) & xy & xz & 0 & z & -y \end{bmatrix}\cdot\frac{f_L}{z^2}(t)$$

Phase 2: Facial Expression Control(1)

Facial Expression Cloning

- Retarget detected facial feature variation to 3D face model

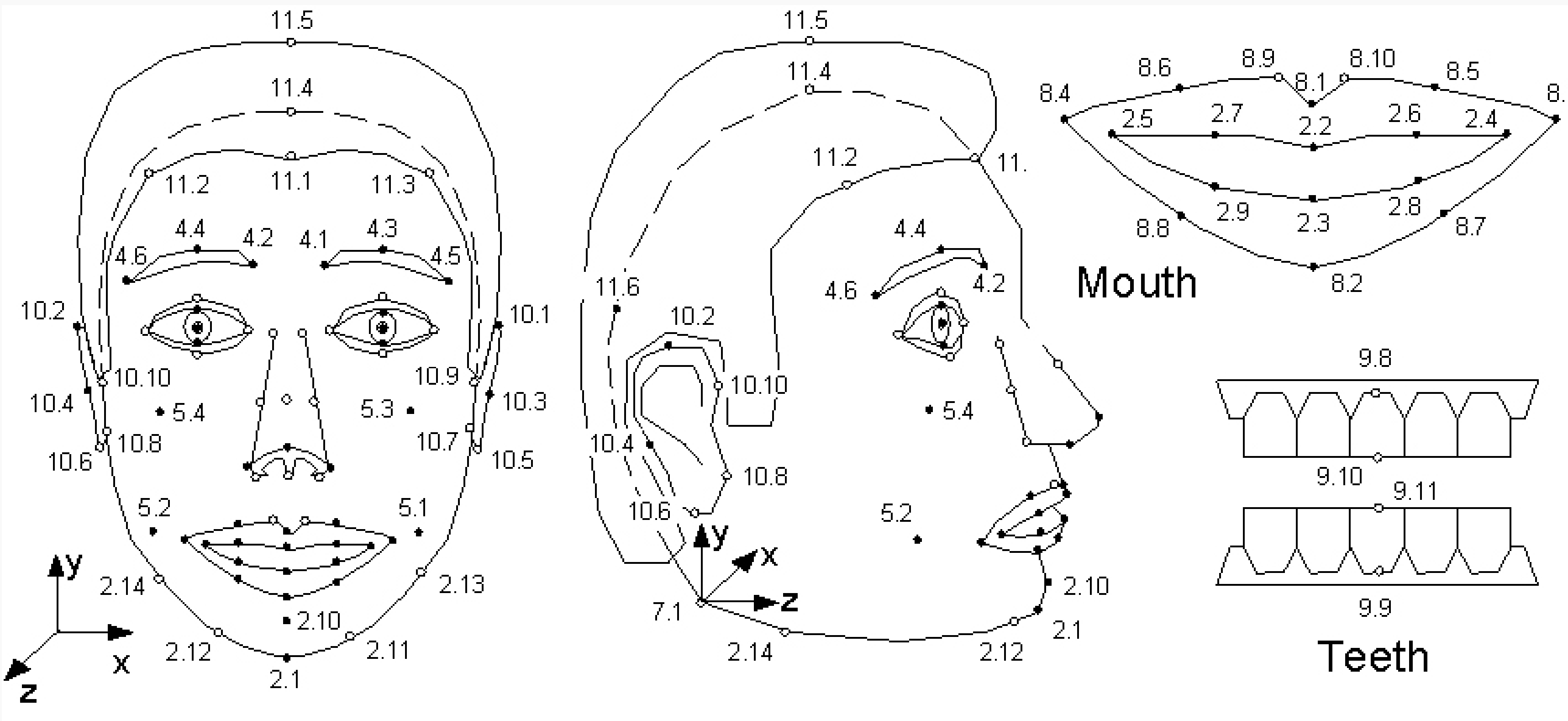


Facial Expression Control(2)

- The variations of the major facial feature points of the face images are tracked by using optical flow and the variations are retargeted to the 3D face model.
- At the same time, we exploit the RBF (Radial Basis Function) to deform the local area of the face model around the major feature points.
- Consequently, facial expression synthesis is done by directly tracking the variations of the major feature points and indirectly estimating the variations of the regional feature points.

Facial Expression Control(2)

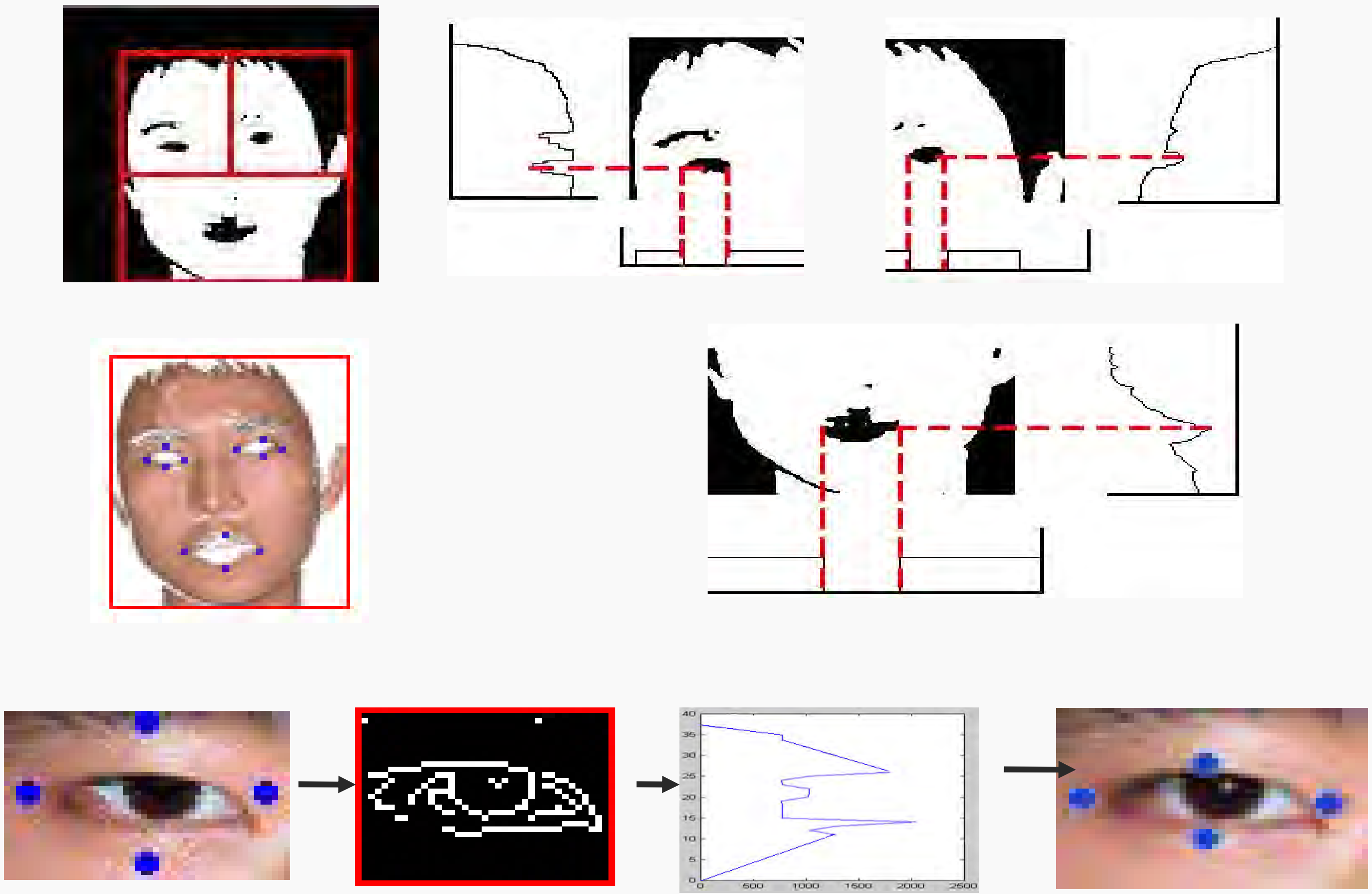
Facial Feature Point Detection



MPEG 4 Facial Definition Points

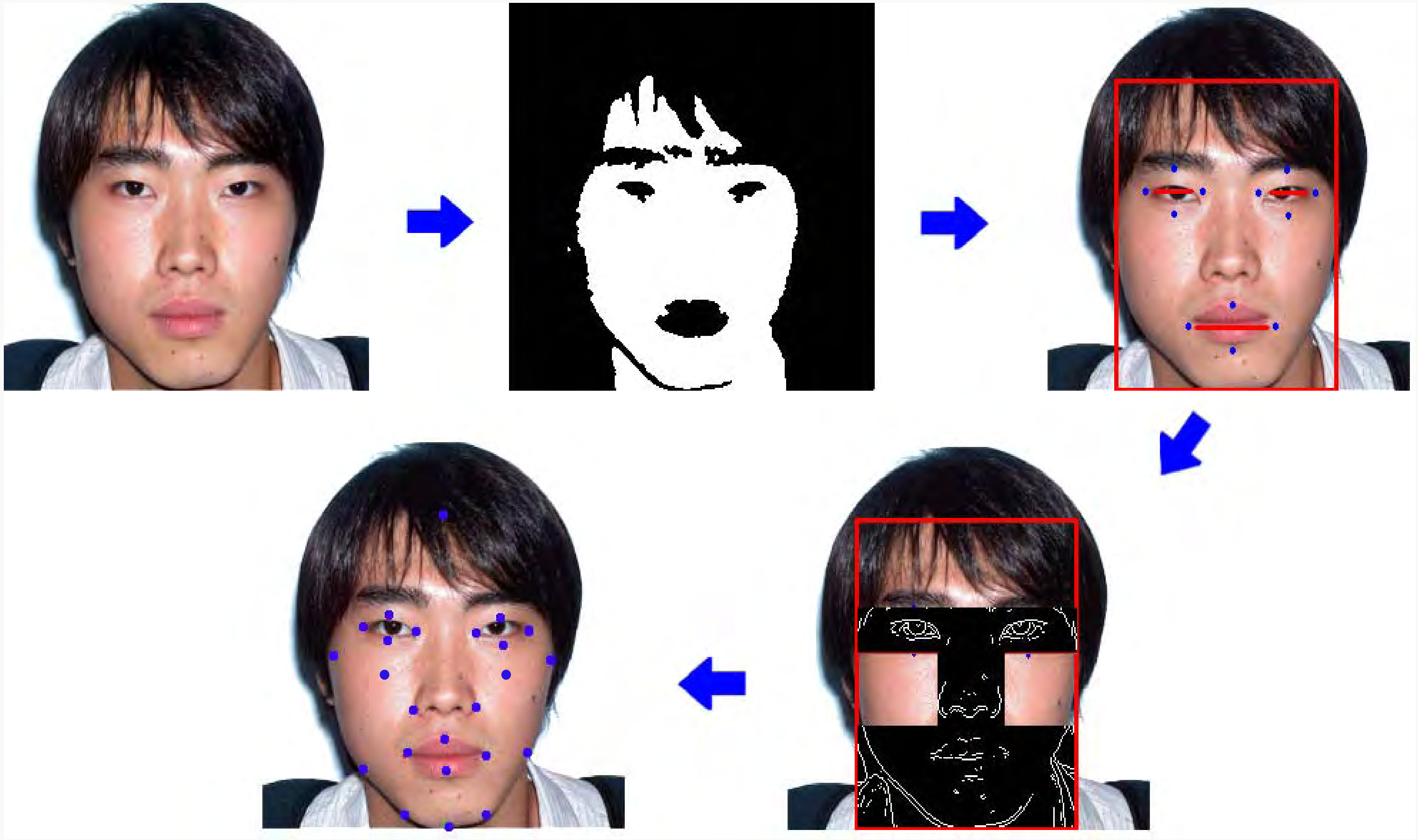
Facial Expression Control(3)

Facial Feature Point Detection



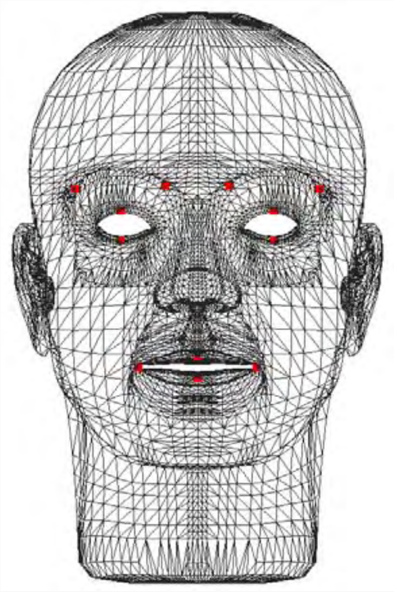
Facial Expression Control(4)

Facial Feature Point Detection

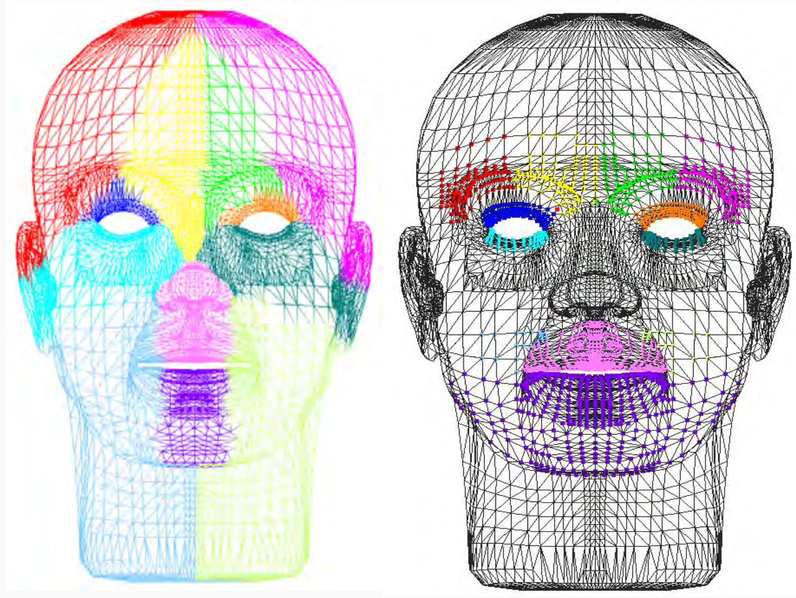


Facial Expression Control(5)

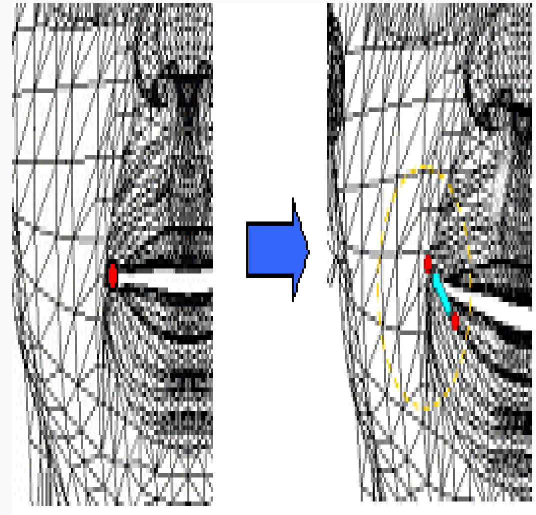
- Facial regions are classified into sub-regions by using K-mean clustering and RBF is applied to each sub-region.



Facial feature points



Local Clustering



Movement of a Local Point

Facial Expression Control(6)

- Gaussian RBF is adopted for model deformation

$$F(\vec{x}) = \sum_{i=1}^{N} c_i \, h\left(\left\|\vec{x} - \vec{x}_i\right\|\right)$$

$$\vec{c} = H^{-1}\vec{x}^{target}$$

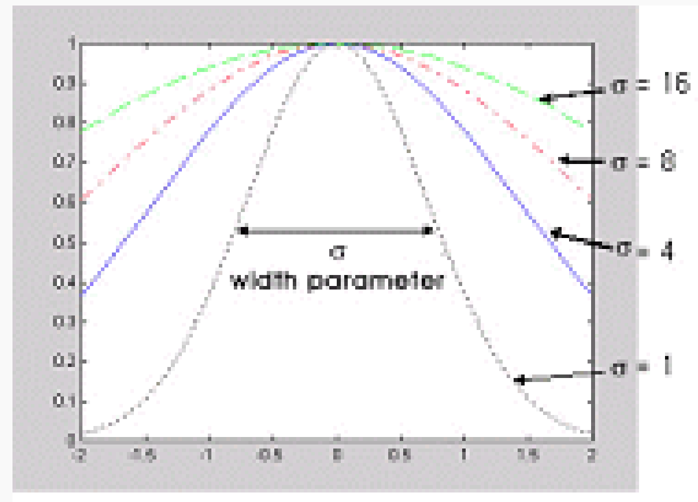
Weight value

$$h(r) = e^{(-r^2/\sigma)}$$

$$\left\|\vec{x}_j^{source} - \vec{x}_i\right\|$$

$$\sigma_i = \max_{j \neq i}\left\|\vec{x}_j^{source} - \vec{x}_i\right\| \times weight$$

Width parameter



Parameters for Animation(1)

- In the facial expression cloning of 3D face model, the geometric positions of the 3D model depend on both the facial features variation and the head pose variation from input video image.

- When the initial feature positions and the changed positions due to the head pose and facial variation are defined by $\vec{v}_0$ and $\vec{v}_p$. Then the relationship between and is defined by

$$\vec{v}_p = T \bullet R \bullet \vec{v}_f \qquad \vec{v}_f = \delta\vec{v} + \vec{v}_0$$

- Where $\vec{v}_f$ is the changed positions of the feature from the frontal face. $\delta\vec{v}$ is the animation parameter which represents the variation from an initial feature point to a changed feature point.

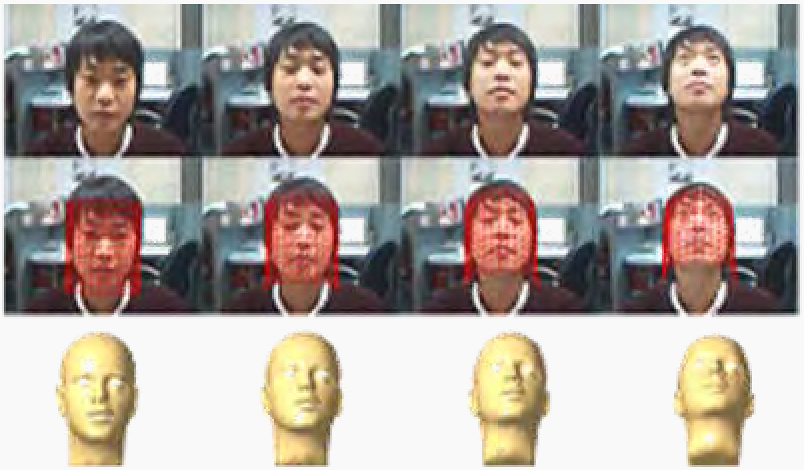
Parameters for Animation(2)

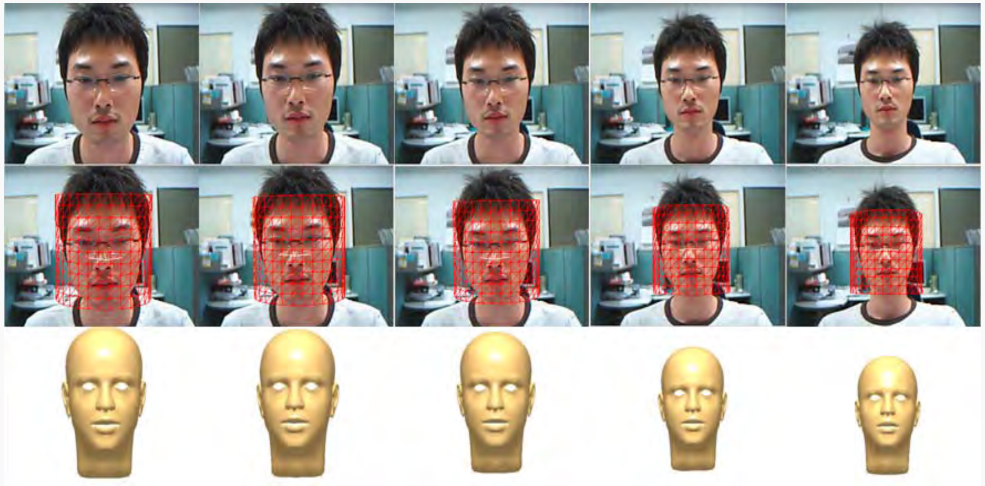
- T and R are transform and rotation matrix respectively.

$$T = \begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad R = \begin{bmatrix} u_x & u_y & u_z & 0 \\ v_x & v_y & v_z & 0 \\ w_x & w_y & w_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\delta\vec{v} = \vec{v}_f - \vec{v}_0 \qquad \vec{v}_f = R^{-1} \bullet T^{-1} \bullet \vec{v}_p$$

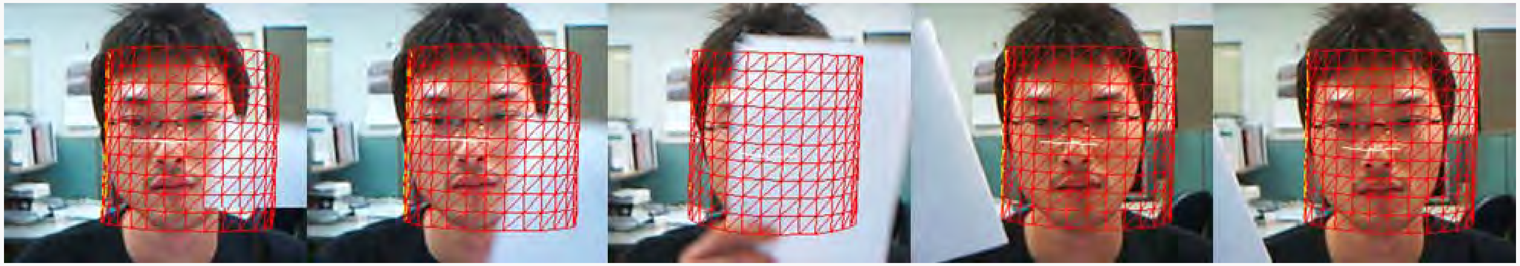
Experimental Results(1)

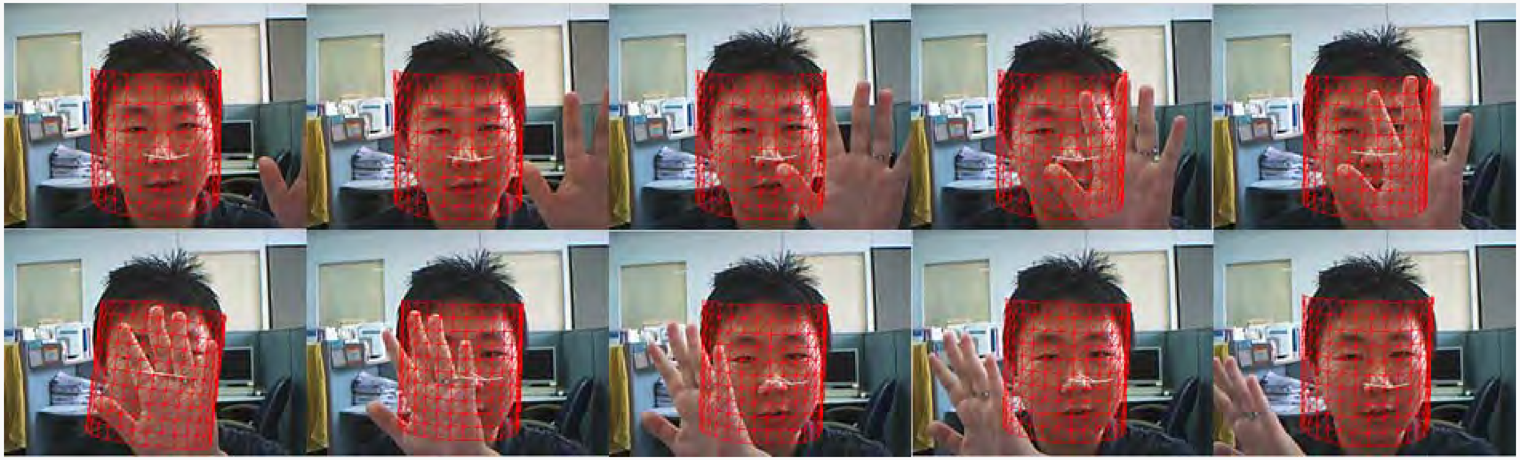




Results of Head Pose Estimation

Experimental Results(2)





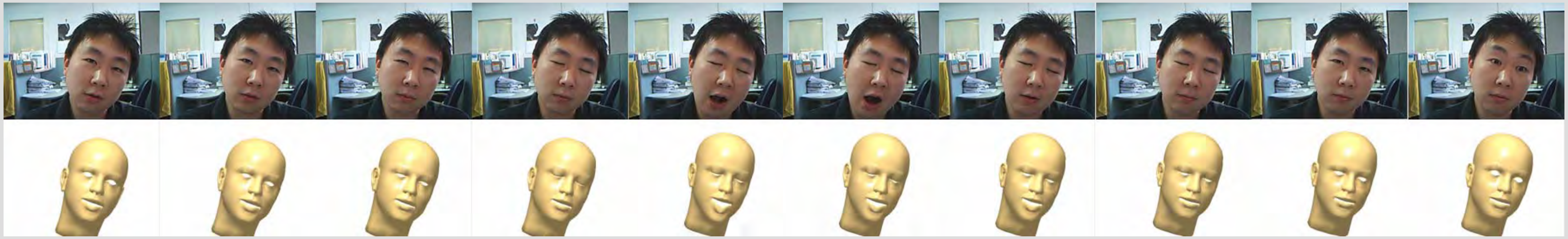
Experimental Results(3)

- Facial Feature Tracking

Experimental Results(4)

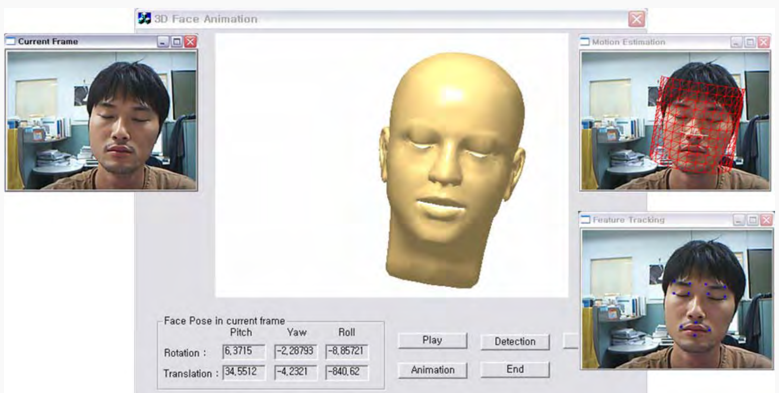
Frontal Head Position

Translated Head Position



Experimental Results(5)

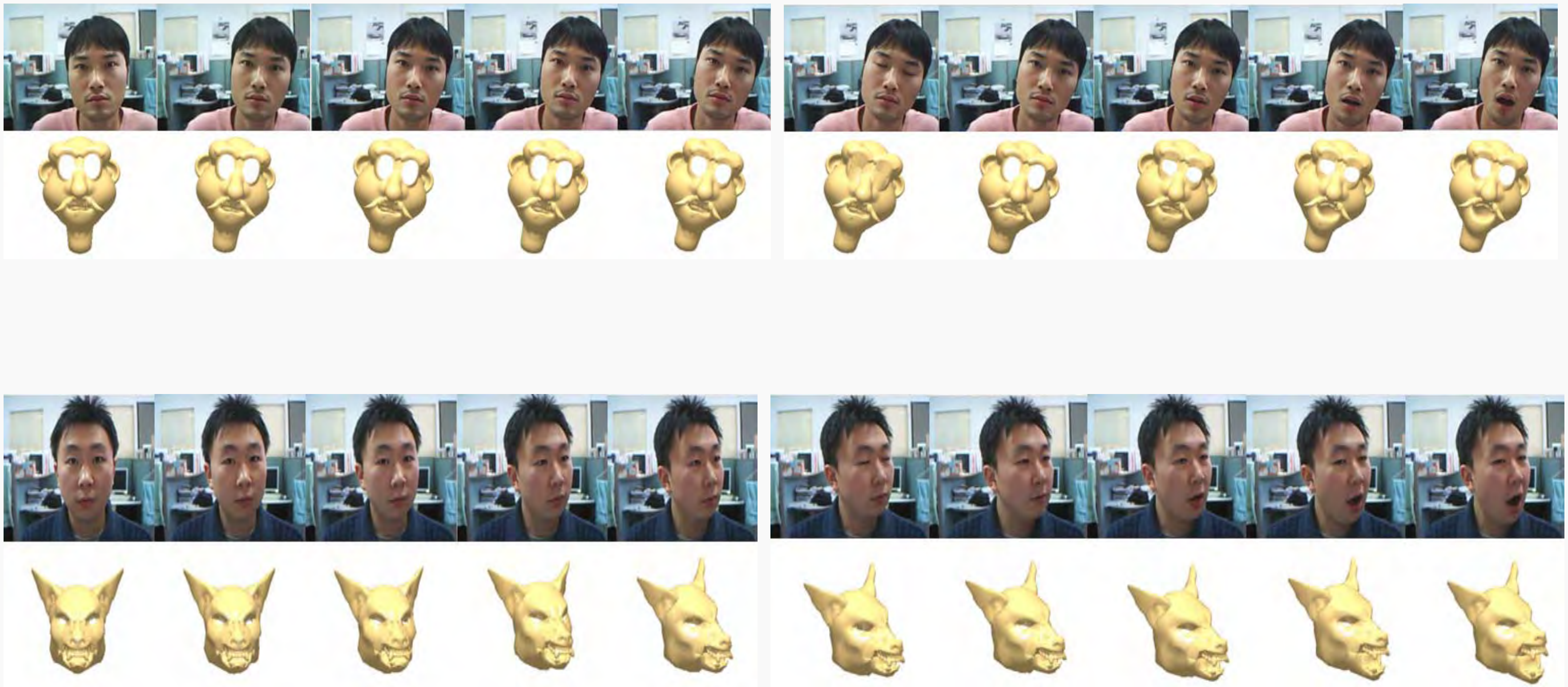
Eyes



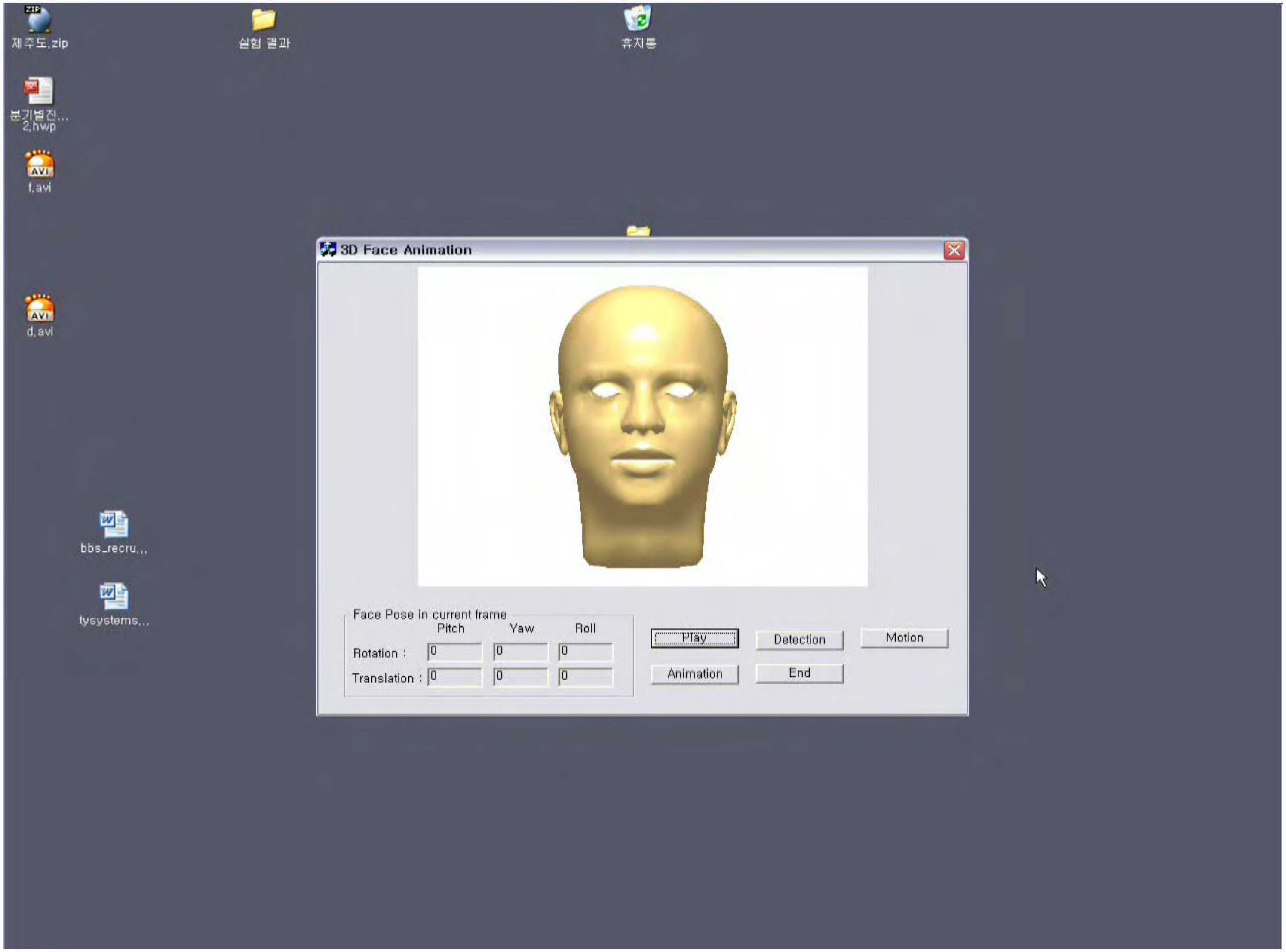
Mouth

Facial Expression Control of an avatar

Experimental Results(b)



Experimental Results-Video(7)



Concluding Remarks

- We propose a robust approach to estimate head pose and track facial features for facial expression control in real time.
- For facial expression control, the detected major facial points form the input face images are retargetted to the facial pints of the 3D avatar.
- From the experiments, we can show the proposed method can effectively recover head pose fully even when self-occlusion is occurred in the sequences of input face images.
- For real time facial expression control, the regional facial feature points around the detected major feature are rearranged by RBF.