

Face Analysis by Multi-Objective 3D Active Appearance Model

Abdul SATTAR

SUPELEC

Signal Communications & Embedded Electronics

Seminar SCEE
30 Oct. 2008

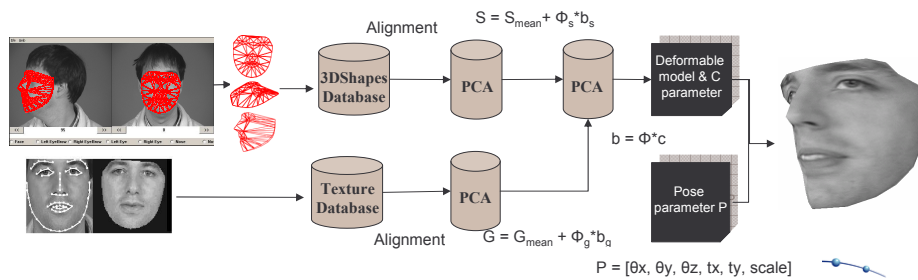
Plan

- **Introduction**
 - 2.5 D Active Appearance Model
 - Segmentation Phase
- **Single-View or Single-Objective 2.5 D Active Appearance Model**
 - Optimized Face Search Techniques
 - Gradient Descent And Nelder Mead Simplex
 - Genetic Algorithm With Gaussian Mixtures
- **Multi-View or Multi-Objective 2.5 D Active Appearance Model**
 - Multi Camera Database
 - Real Faces
 - Synthetic Faces (Autodesk Maya)
 - Optimized Face Search Techniques
 - NSGA-II
- **Conclusions**

- **Introduction**
 - 2.5 D Active Appearance Model
 - Segmentation Phase
- **Single-View or Single-Objective 2.5 D Active Appearance Model**
 - Optimized Face Search Techniques
 - Gradient Descent And Nelder Mead Simplex
 - Genetic Algorithm With Gaussian Mixtures
- **Multi-View or Multi-Objective 2.5 D Active Appearance Model**
 - Multi Camera Database
 - Real Faces
 - Synthetic Faces (Autodesk Maya)
 - Optimized Face Search Techniques
 - NSGA-II
- **Conclusions**

- **Face Analysis**
 - Recognition
 - Expression detection
 - Alignment
 - Feature extraction
 - Pose estimation
- **Problem Statement**
 - Our objective to align and extract the features of a face moving in all six degrees of freedom in real-time.
 - Systems should be
 - Efficient in time
 - Efficient in memory
 - Robust with respect to face orientation.
- **Two types of methods**
 - View based method (PCA, LDA, ICA etc.)
 - Appearance based or model based methods (AAM, 3DMM)

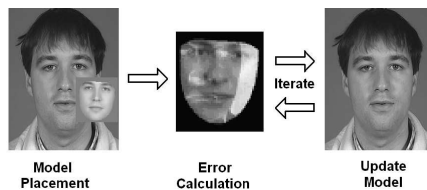
- In training phase 2.5D AAM is constructed by the combination of
 - 3D landmarks taken from frontal and profile view of a face image.
 - Texture of only frontal view of a face image.
 - We have appearance parameters C and pose parameters P.



- **Segmentation**
 - Shape model obtained by varying C and P parameters, is placed on the query image I to warp the face to mean frontal shape.
 - We apply photometric texture normalization to overcome illumination variations.
 - The objective is to minimize pixel error between image and the model by varying C and P parameters.

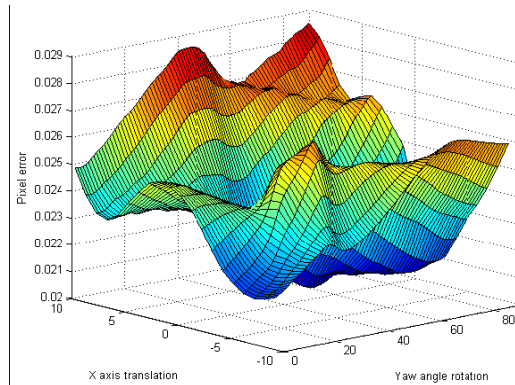
$$\text{Pixel Error} = \| I(C, P) - M(C) \|$$

where $I(C, P)$ is the segmented image and $M(C)$ is the model obtained by C parameters. To choose good parameters we need an optimization method.



➤ Pixel Error

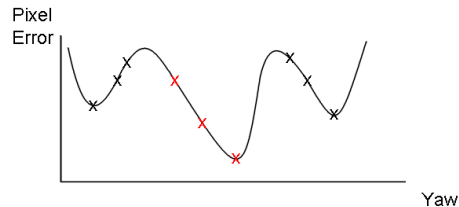
- Pixel error w.r.t X-axis translation and Yaw angle rotation.
- Contains many local minimum values.
- Need an efficient optimization technique for face analysis in a query image.



- Introduction
 - 2.5 D Active Appearance Model
 - Segmentation Phase
- Single-View or Single-Objective 2.5 D Active Appearance Model
 - Optimized Face Search Techniques
 - Gradient Descent And Nelder Mead Simplex
 - Genetic Algorithm With Gaussian Mixtures
- Multi-View or Multi-Objective 2.5 D Active Appearance Model
 - Multi Camera Database
 - Real Faces
 - Synthetic Faces (Autodesk Maya)
 - Optimized Face Search Techniques
 - NSGA-II
- Conclusions

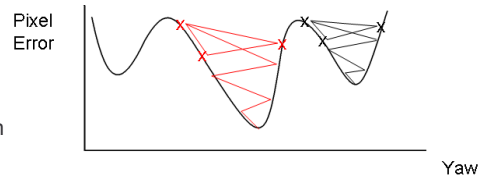
➤ Gradient Descent or Linear Regression

- Higher number of faces in learning phase in pre-computation requires high memory.
- Lack of generality (inefficient in aligning faces not included in the database).
- Requires various initializations for better convergence.
- Tendency of falling in local minimum value.



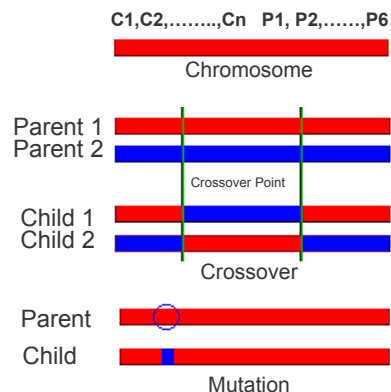
➤ Nelder Mead Simplex

- Good for generality.
- Require higher number of initialization of pose parameters.
- Poor initialization tends the system to fall in local minimum value.



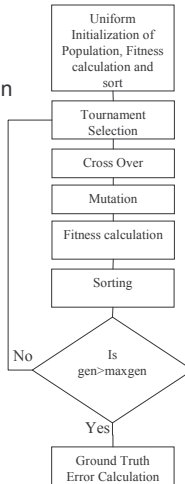
➤ Genetic Algorithm

- Good for generality.
- Population based method
- Capable of both exploitation and exploration.
- Do not require various initializations due to exploration property.
- We have used it to optimize appearance parameters C and pose parameters P of 2.5 D AAM model.
- Population is composed of chromosomes
- Chromosome is a combination of parameters and single parameter is called a gene.
- Chromosomes evolves towards better solutions.
- Its evolution depends upon selection, crossover and mutation of chromosomes.
- It terminates after certain number of generations.



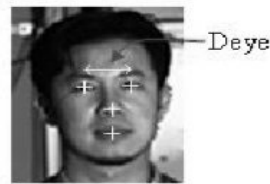
➤ **Steps of GA**

- Initialization
- Tournament Selection
- Two point Crossover
- Mutation
- Fitness Calculation
- Sorting



➤ **Ground Truth Error Calculation**

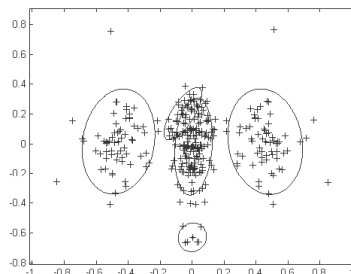
- It is the mean pixel distance between the features localized by the system and real location of these features.
- It is normalized by D_{eye} , distance between the eyes.
- We take $D_{eye} = 0.8 * D_{face}$
- D_{face} is the distance between eyes and nose



INSTITUT D'ÉLECTRONIQUE ET DE TÉLÉCOMMUNICATIONS DE RENNES
UMR CNRS 6184

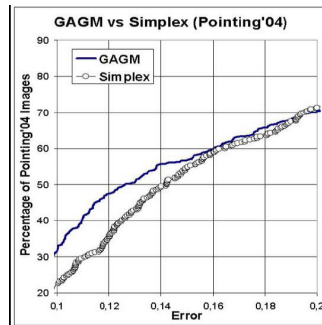
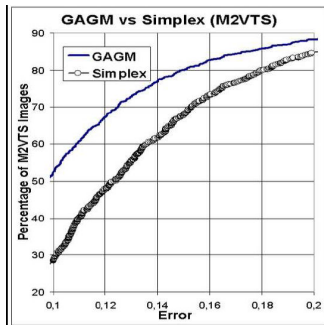
➤ **Genetic Algorithm With Gaussian Mixtures**

- A single person with different expressions, glasses, beard or mustache is regarded as totally a different person if seen without them.
- It can be observed that appearance of that person lies in multiple far apart clusters in the face space.
- Thus avoiding search in these empty spaces results in efficient Genetic Algorithm



IETR
INSTITUT D'ÉLECTRONIQUE ET DE TÉLÉCOMMUNICATIONS DE RENNES
UMR CNRS 6184

Results of comparison of simplex and GAGM (Genetic Algorithm with Gaussian Mixtures)



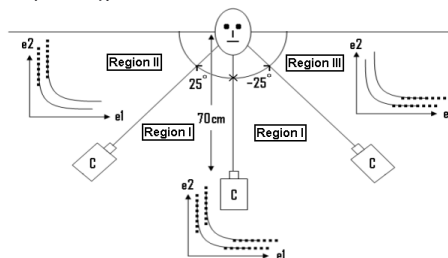
- **Introduction**
 - 2.5 D Active Appearance Model
 - Segmentation Phase
- **Single-View or Single-Objective 2.5 D Active Appearance Model**
 - Optimized Face Search Techniques
 - Gradient Descent And Nelder Mead Simplex
 - Genetic Algorithm With Gaussian Mixtures
- **Multi-View or Multi-Objective 2.5 D Active Appearance Model**
 - Multi Camera Database
 - Real Faces
 - Synthetic Faces (Autodesk Maya)
 - Optimized Face Search Techniques
 - NSGA-II
- **Conclusions**

➤ **Why we felt the need of multiple cameras**

- Two cameras instead of one camera.
- Data from two cameras make the system more robust
- Increase the capacity to extract the features of a face even with large lateral movements.
- Multi Objective optimization has less tendency to fall in local minima.

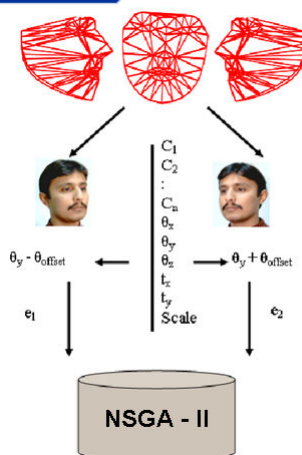
➤ **Multiple Camera Database**

- The community lacks database which involve lateral motion of a face captured by more than one camera.
- We have developed a multi-view scenario.
- Integrating two off the shelf webcams placed on the extreme edges of display screen facing towards the user as shown in figure.
- Third camera is placed in the center to compare the algorithm of MOAAM with SVAAM.
- Test images of 7 individuals are acquired with the rotation of their faces from -90 to +90 of yaw angle.





Synthetic Faces (Autodesk Maya)



Multi-View Active Appearance Model

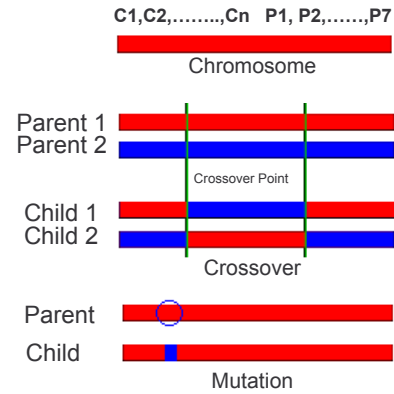
- Optimization of two errors from each camera.
- Shape model is laid on both the images with the same C and P parameters .
- Except a yaw angle offset (θ_{offset}) is introduced between the two models
- Pixel errors between the images and models are calculated.
- The objective is to minimize both pixel errors.

$$e1 = \| I(C, P) - M(C) \|$$

$$e2 = \| I(C, P) - M(C) \|$$
- Multi-objective optimization is necessary.
- Non-dominated Sorting Genetic Algorithm (NSGA-II) is implemented for multi-objective optimization.

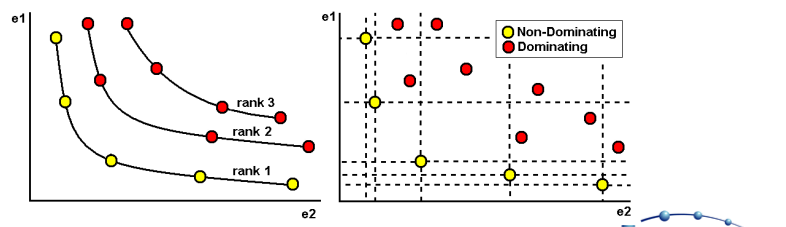
➤ NSGA-II

- Non-dominated Sorting Genetic Algorithm (NSGA-II) is the multi-objective version of Genetic Algorithm
- Chromosome is the combination of C and P parameters.
- Two point crossover and mutation is implemented for reproduction
- Selection and sorting is different from Simple GA.
- Sorting is non-dominating sort
- Selection is Pareto frontiers based



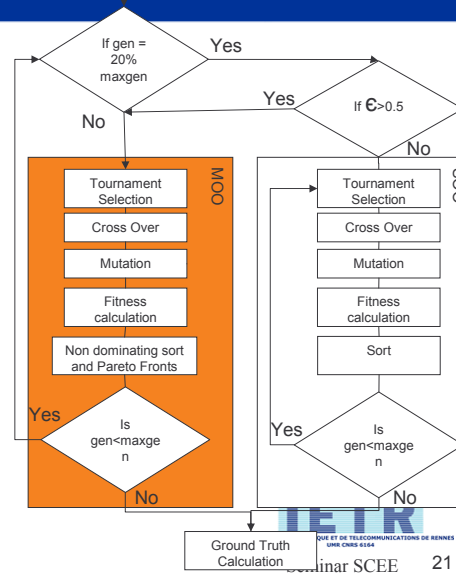
➤ Pareto Frontiers:

- Find the set of solutions in the population that are Pareto non-dominated
- These solutions are assigned the highest rank and are removed from further assignment of the ranks.
- Similarly remaining population undergoes the same process of ranking.
- The sorting is performed with respect to these frontiers instead of individual pixel errors

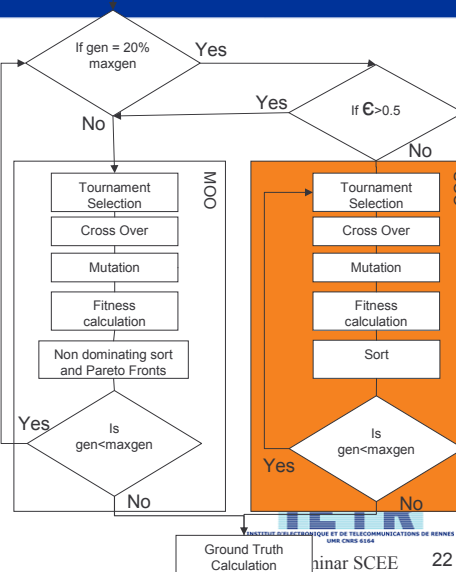
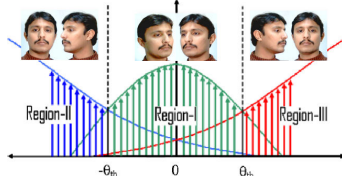


- **Initialization:**
 - Both camera test images are loaded
 - Random initialization of chromosomes
 - Fitness is calculated and sorted
- **Reproduction:**
 - Tournament selection performed on frontiers.
 - Parents are crossed over and mutated.
- **Segmentation:**
 - Each chromosome corresponds to a 3D AAM shape.
 - This deformed shape is laid on both the test images with a yaw angle offset of $\pm 25^\circ$.
- **Fitness:**
 - Two sets of pixel errors (fitness) are calculated.
$$e1 = \left\| I(C, P) - M(C) \right\|$$

$$e2 = \left\| I(C, P) - M(C) \right\|$$
- **Non-Dominating Sort:**
 - Non dominating sort is performed
 - Entire population is ranked to Pareto frontiers for further evolution of generations.
- **Solution:**
 - Best chromosome from the first front is taken.
 - It represents best localization of face features



- **Switching MOO to SOO:**
- Sometimes face is oriented such that one of the camera of multi-view system do not hold a valid face information
- System discard this camera and focus on to other camera.
- Histogram of genes of entire population representing the yaw of a face is observed.
- **Region-I:** Energies from both cameras are meaningful and data from any one of them cannot be neglected.
- **Region-II and III:** Energy from one of the camera is sufficient enough to localize the face features and other camera can be discarded.
- Value of ϵ switches MOO (Multi-Objective Optimization) to SOO (Single Objective Optimization).



Mathematically, let us suppose Pop is a set of population given as

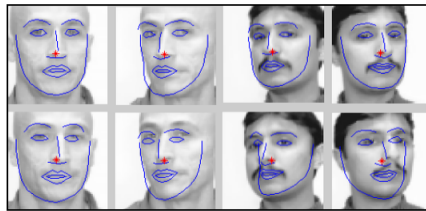
$$Pop = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1M} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2M} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{Nk} & \cdots & X_{NM} \end{pmatrix}$$

where N is the number of chromosomes X and M is the number of genes of each chromosome. Now we observe the k_{th} gene of each chromosome which represents yaw angle of the model. In order to calculate the histogram of chromosome, assign 1 to ζ such as

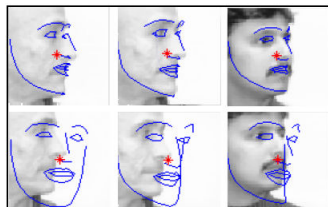
$$\zeta_i = \begin{cases} 1 & -\theta_{th} \leq X_{ik} \leq \theta_{th} \\ 0 & X_{ik} \leq -\theta_{th} \text{ or } X_{ik} \geq \theta_{th} \end{cases} \quad 1 \leq i \leq N$$

where θ_{th} is the threshold angle equals to the half of the angle between two cameras. ϵ is the ratio of number of chromosomes representing the face position in region-1 to the total number of chromosomes.

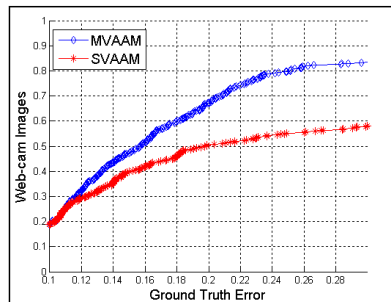
$$\epsilon = \frac{\sum_{i=1}^N \zeta_i}{N} = \begin{cases} < 0.50 & \text{Single Camera Mode} \\ \geq 0.50 & \text{Multi-View Mode} \end{cases}$$



MVAAM (top) vs. SVAAM (bottom) for region I



MVAAM (top) vs. SVAAM (bottom) for region II & III



Ground Truth Error MVAAM vs. SVAAM

- Our multi objective 2.5D AAM system of face alignment, feature extraction and pose estimation is
 - Efficient in memory because it does not require to store huge regression or jacobian matrices.
 - Genetic Algorithm outperform Simplex and classical AAM techniques of Linear regression and Gradient Descent.
 - Multi view or multi objective system is more robust with respect to pose.
 - Moreover it has the capability of switching from MOO to SOO which makes it time efficient with respect to applying multiple instances of conventional SVAAM on multiple images.
- **Future Work**
 - Genetic Algorithm no doubt has best exploration ability but its exploitation ability should be enhanced with algorithms of gradient descent or Simplex.

