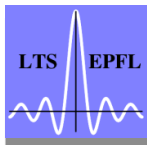


Fast Multi-view Face Tracking With Pose Estimation

Javier Cruz Mota, Julien Meynet, Taner Arsan, Jean-Philippe Thiran*

*Kadir Has University, Computer Engineering Department,
Istanbul, Turkey



Fast Multi-view Face Tracking With Pose Estimation

Julien Meynet , Taner Arsan , Javier Cruz
Gata, Jean-Philippe Thiran*

*Kadir Has University, Computer Engineering
Department, Istanbul, Turkey

Overview

- Introduction to face detection and tracking
- Face class modeling
- Multi-view face modeling
- Fast tracking and pose estimation
- Experiments and results
- Conclusions

Introduction to face detection

- Face detection:

- Feature-based approaches

- Geometrical models
- Skin colour,...

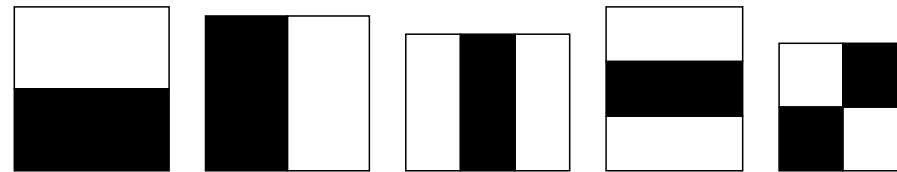
- Example based approaches:

Use a sliding window to scan an image and perform binary classification:

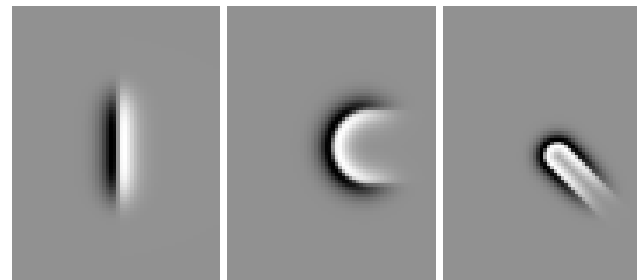
- Template matching
- SVMs
- AdaBoost,...

Face class modeling

- Asymmetric AdaBoost is used for learning face vs. non face classifiers
- Haar-like filters: computationally efficient

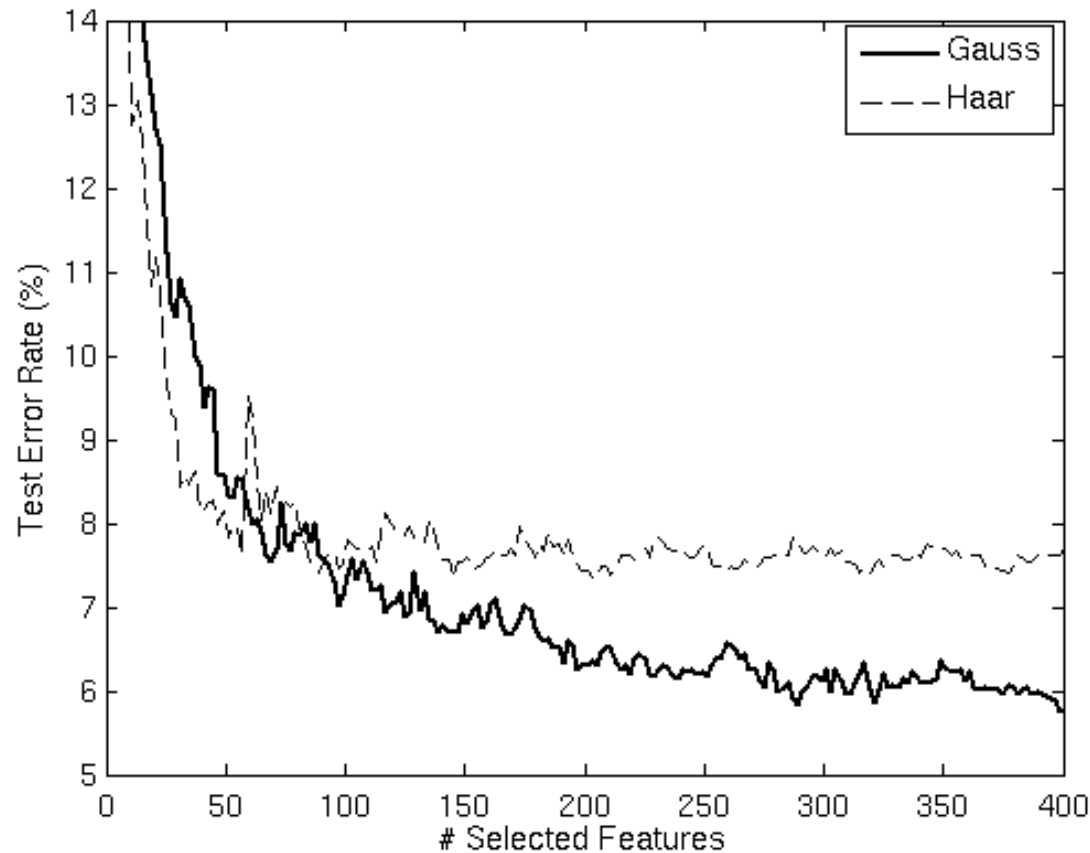


- Anisotropic Gaussian filters: more discriminant



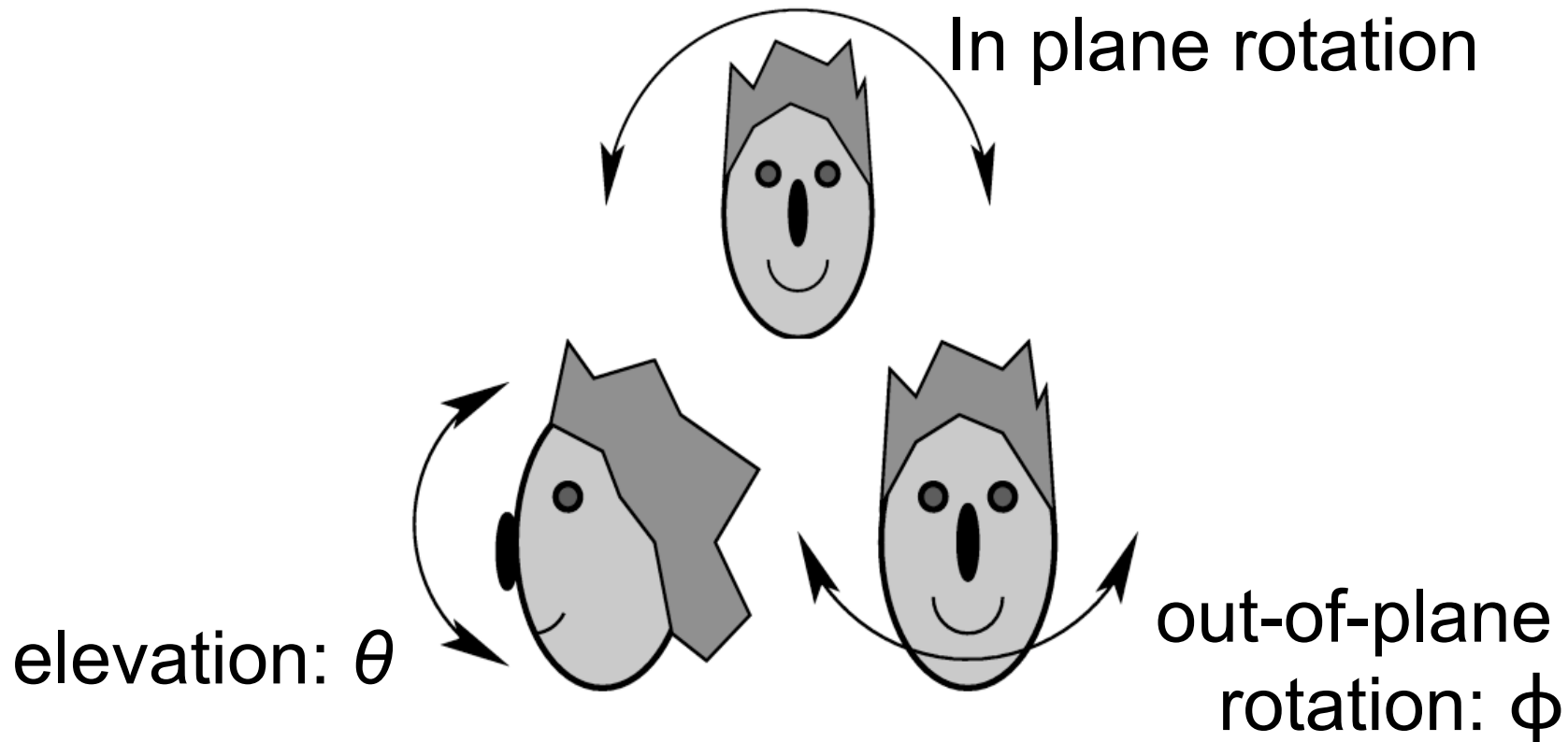
Haar-like vs. Gaussian filters

- Test error function of AdaBoost iterations



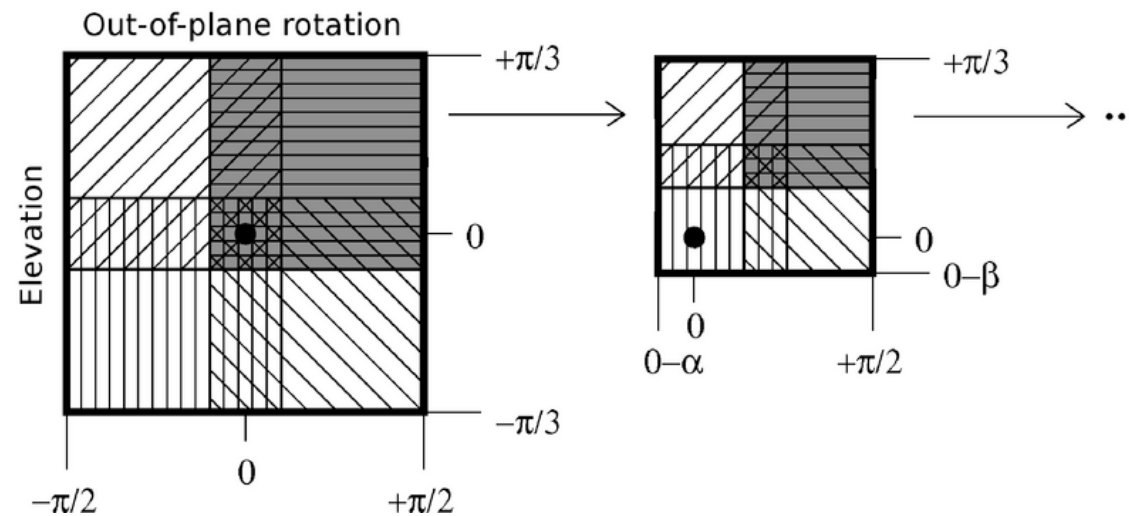
Multi-view face modeling

- Idea: use specific binary classifiers for each pose:

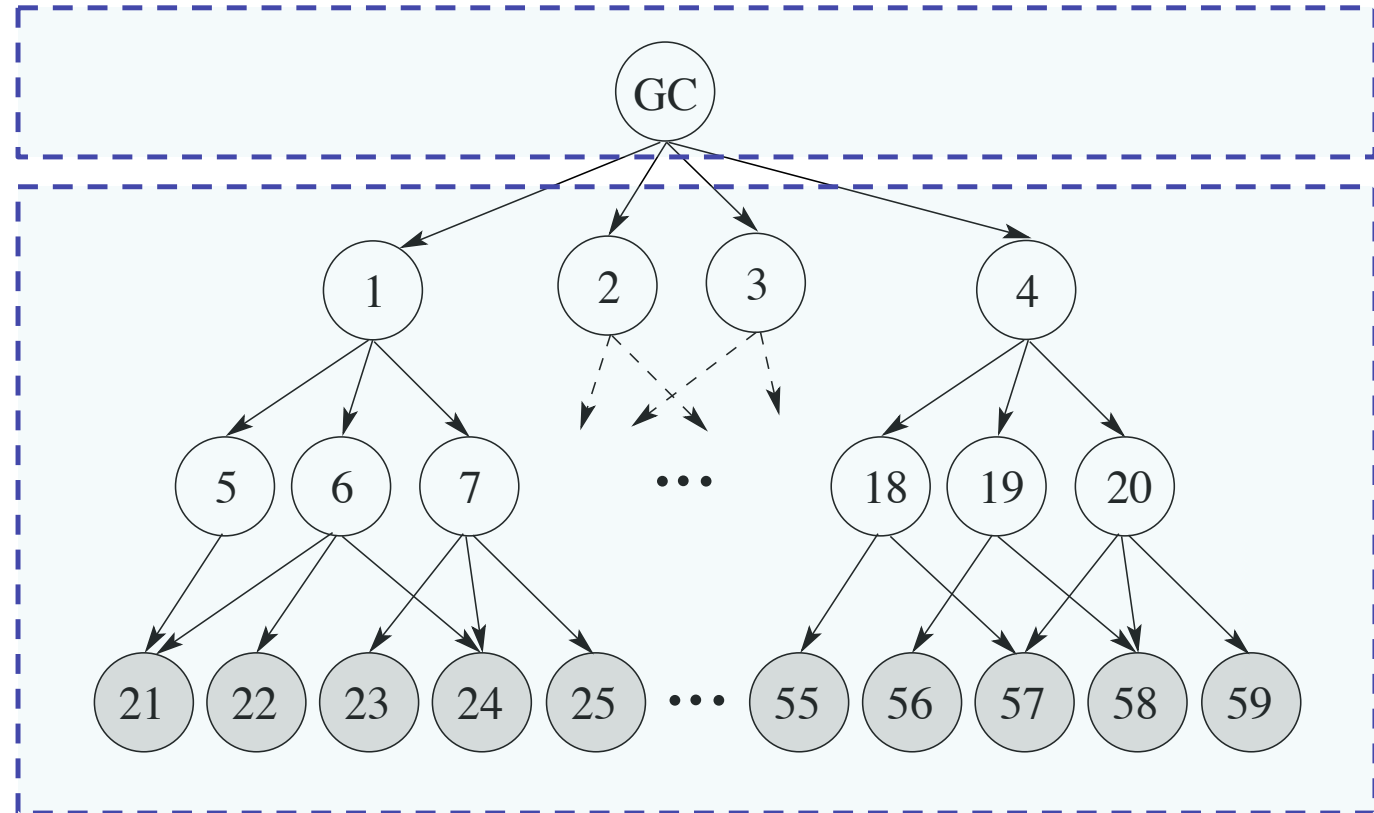
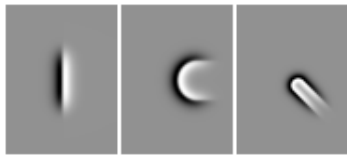
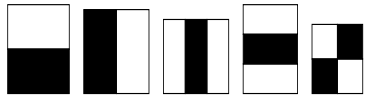


Multi-view face modeling

- A binary pose classifier:
 $\{\text{Pose}(\theta_i, \phi_i)\}$ vs. $\{\text{other poses} + \text{non faces}\}$
- Hierarchical sampling of the pose plane:



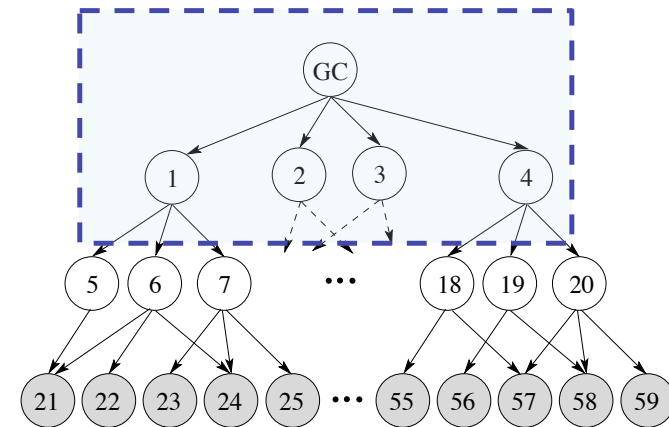
Multi-view face modeling



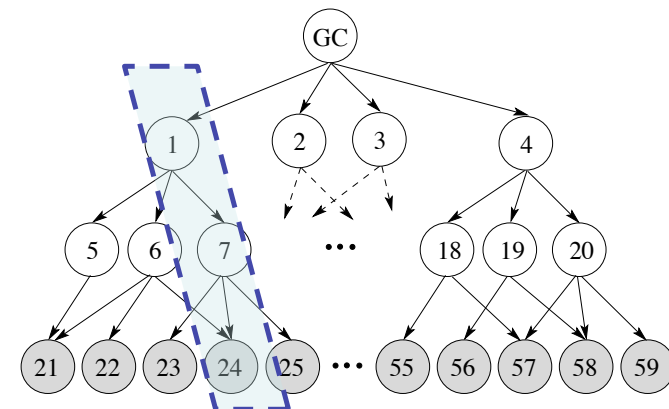
Fast face tracking

- Condensation is used for tracking detected faces
- Tracking in two steps:

1. Track the position:



2. Track the pose
at the given position



Experiments

- Datasets

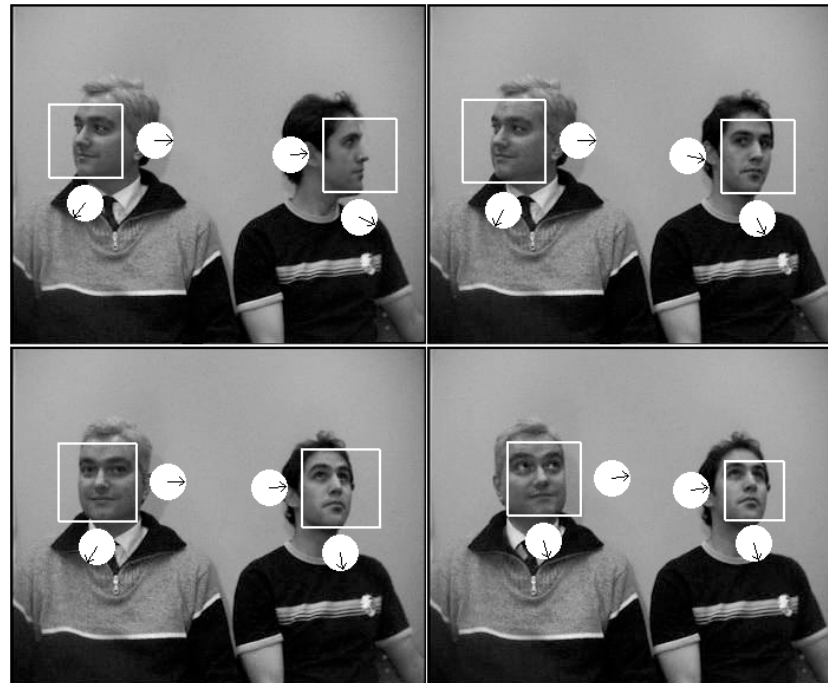
- CMU Pose, Illumination and Expression (PIE): 47954 images
- INIRIALPES: 2597 images
- Feret dataset
- Non face examples: bootstrapping on randomly selected images

- System

- 20x20 pixels images
- 1 GC: a cascade of Haar filters (≈ 150 filters)
- 59 PC: each PC is a cascade of GF with ≈ 75 filters

Results

- 93% of correct classification on complex videos
- Very precise out-of-plane estimation
- Average estimation on elevation



Results – time performances

- Tests on 1500 frames of a sequence of 320x240 pixels

Detector	fps
Detection frame by frame	6.36
Tracking	23.45

Conclusions

- Fast multi-view face tracking
- Tree of classifiers (from general to specific)
- Tracking in 2 steps: position + pose