Face Recognition in Video Using Gabor Wavelet Networks

Rajkiran Gottumukkal and Vijayan Asari

Old Dominion University

Overview

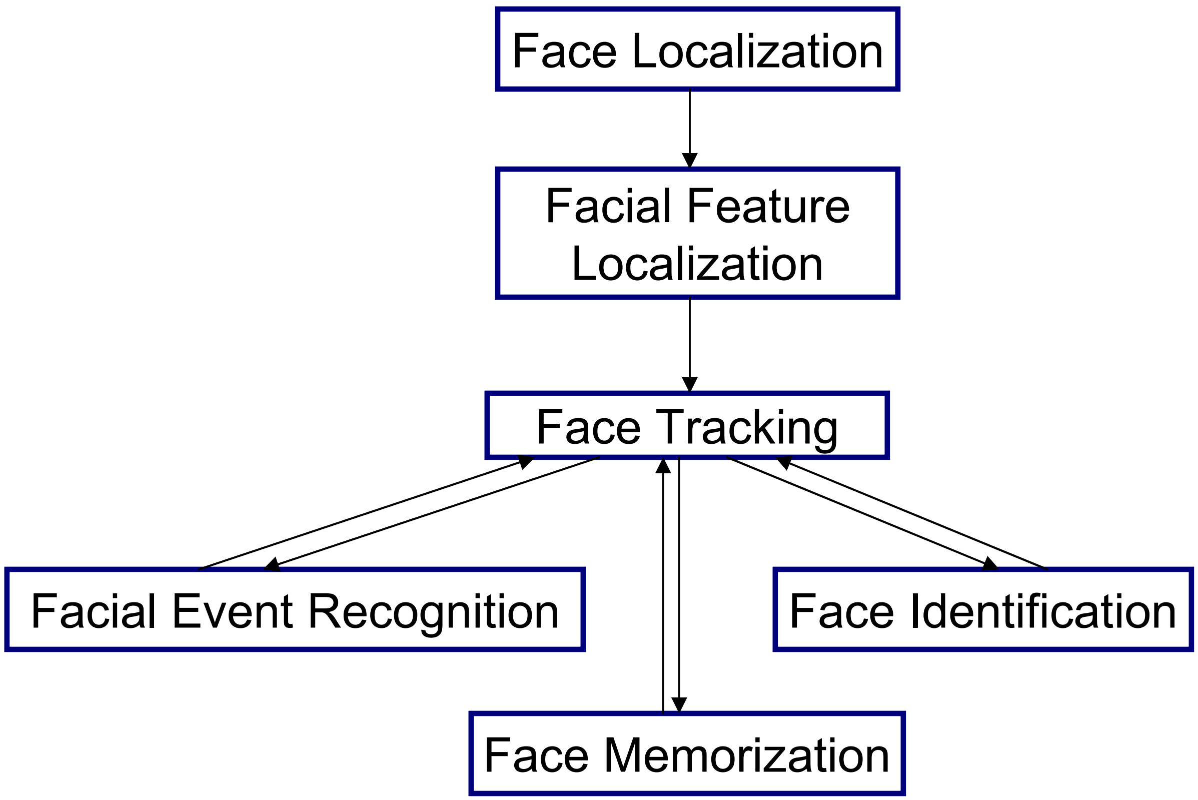
- Introduction
- Face Processing Tasks
- Face Localization
- Facial Feature Localization
- Face Tracking & Recognition
- Video Quality
- Conclusion

Introduction

- Video based recognition is less intrusive
- Video provides soft biometrics and identification at a distance
- Still a new field, lot needs to be done
- Face not frontal, poor illumination, blurriness, bad focus
- Not mature enough for intelligent video surveillance

11/9 Hijackers as seen on surveillance cameras

Face Processing Tasks



Face Localization

- Use methods to detect faces quickly for real time processing
- Optical flow to find changes in scene
- Viola-Jones face detector to detect faces
- Can obtain real time face localization

Facial Feature Localization

- The GWN approach represents a face image through a linear combination of 2D Gabor functions.
- Considering the 2D image case, each single odd Gabor wavelet can be expressed as follows:

$$\psi_{\mathbf{n}_i}(\mathbf{x}) = \exp\left[-\frac{1}{2}(\mathbf{S}_i(\mathbf{x}-\mu_i))^T(\mathbf{S}_i(\mathbf{x}-\mu_i))\right] \times \sin\left[(\mathbf{S}_i(\mathbf{x}-\mu_i))\begin{pmatrix}1\\0\end{pmatrix}\right],$$

- GWN for an image consists of n such wavelets and a set of associated weights

$$\Psi(\mathbf{x}) = \sum_{i=1}^{n} w_i \psi_{\mathbf{n}_i}(\mathbf{x})$$

Facial Feature Localization Cont.

- GWN parameters are learned :
  - Randomly drop n wavelets within the target object
  - Perform gradient descent to minimize the difference between the GWN and the training image

$$\arg\min_{w_i,\mathbf{n}_i} \left\| \mathbf{I}^t - \sum w_i \psi_{\mathbf{n}_i}(\mathbf{x}) \right\|^2$$

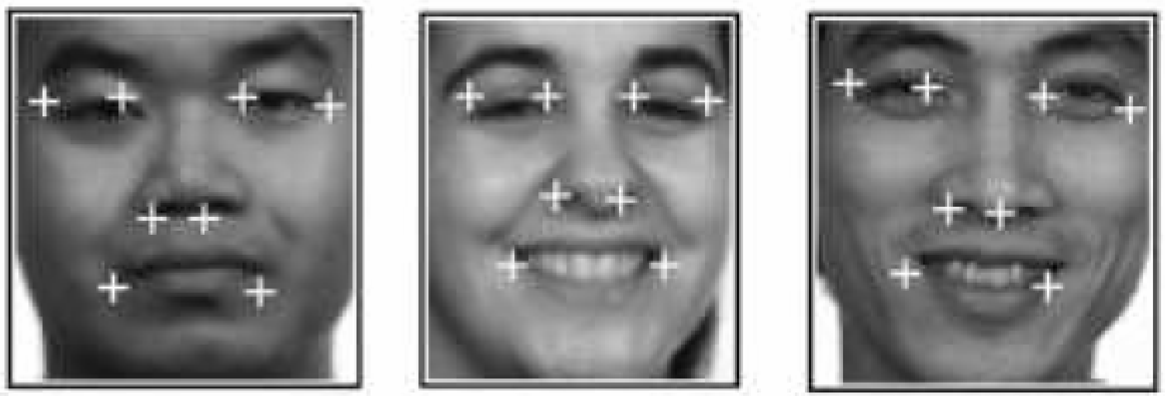
- Hierarchal wavelet networks are used to localize eight facial features.

Facial Feature Localization Cont.

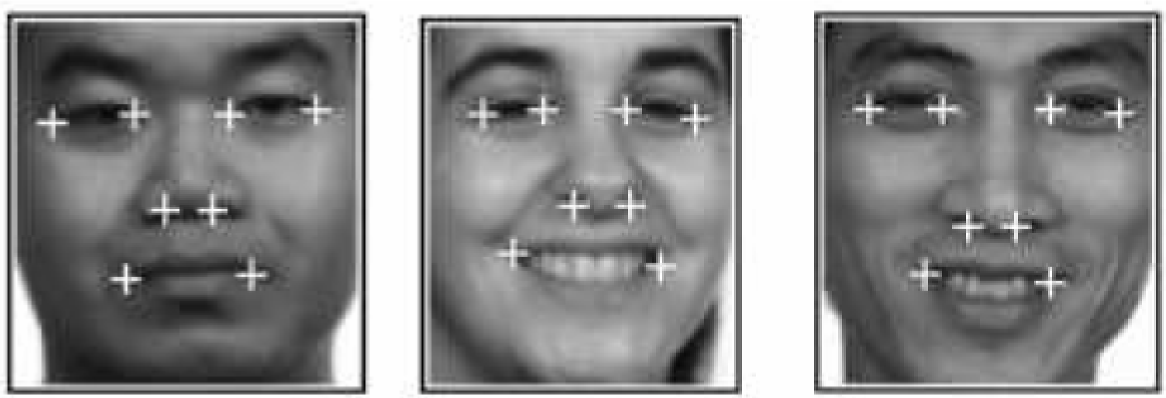
- Level One : Face Matching
  - Determine an affine transformation of the level-one GWN that registers the candidate with the target image
  - The residual score in wavelet subspace is minimized to find the best match face
- Level Two : Feature Localization
  - For each feature a brute-force search is preformed between a level-two feature GWN and the target image.
  - Candidate feature GWNs may be drawn from any of the faces in the database.

Facial Feature Localization Cont.

1-Level Matching



2-Level Matching

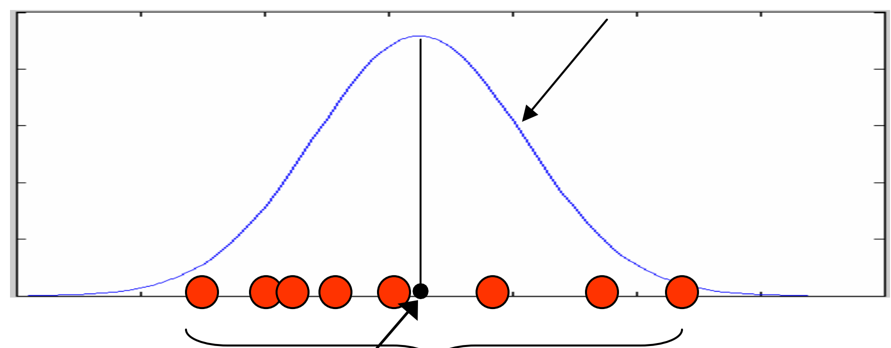


Courtesy:

R. S. Feris, J. Gemmell, K. Toyama, and V. Krueger, "*Hierarchical Wavelet Networks for Facial Feature Localization*"

Face Tracking & Recognition

- The face and its identity are simultaneously tracked using CONDENSATION filter
- Consider a normal distribution with the observation as the mean.
- Randomly select *N* samples($S_{t_0}^n$) around the mean.
- Assign a weight($w_{t_0}^n$) to each selected sample (initially $w_{t_0}^n = 1$).



- Construct a new sample set $\left\{ s_{t+1}^{(n)}, w_{t+1}^{(n)}, c_{t+1}^{(n)}, n = 1,\ldots,N \right\}$ where $c_{t+1}^{(n)}$ is the cumulative weight value such that

$$c_{t+1}^{(n)} = c_{t+1}^{(n-1)} + w_{t+1}^{(n)} \quad and \quad c_{t+1}^{(0)} = 0.$$

- The new sample set is constructed in three stages:
  - Sample, Predict, Measure

Face Tracking & Recognition Cont.

- Predict: Move particles according to deterministic dynamics (drift), then perturb individually (diffuse).

$$s_{t+1}^{(i)} = F(s_{t+1}^{'(i)}) + d$$

drift — diffusion

$$s_{t+1}^{(i)} = (-0.1 + (s_{t+1}^{'(i)} + 0.1) * 0.4) + (0.075 * r)$$

- Measure: Weigh the new samples in terms of the measured features (ex. location) from the image.

$$w_{t+1}^{(i)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-(S(s_{t+1}^{(i)}) - S(x_{t_0}))^2 / (2\sigma^2)}$$

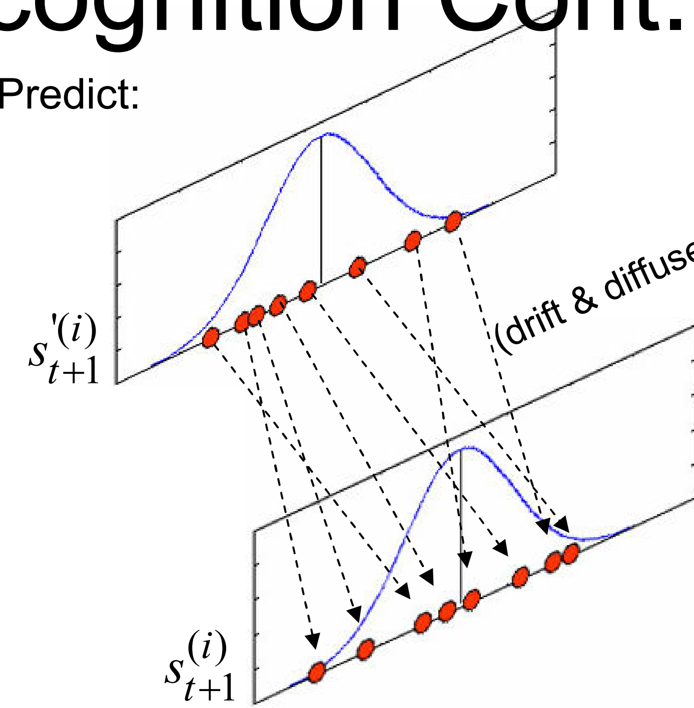
measured feature

$$\sigma = 0.03$$

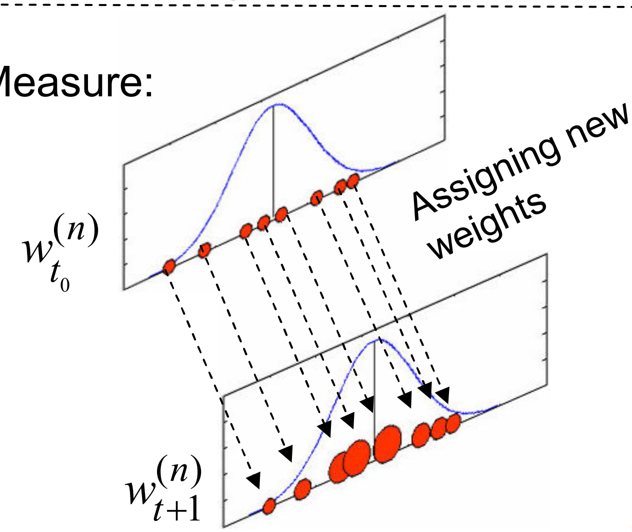
- Position can be estimated from the N samples:

$$s_{t+1} = \sum_{n=1}^{N} w_{t+1}^{(n)} s_{t+1}^{(n)} \Big/ \sum_{n=1}^{N} w_{t+1}^{(n)}$$

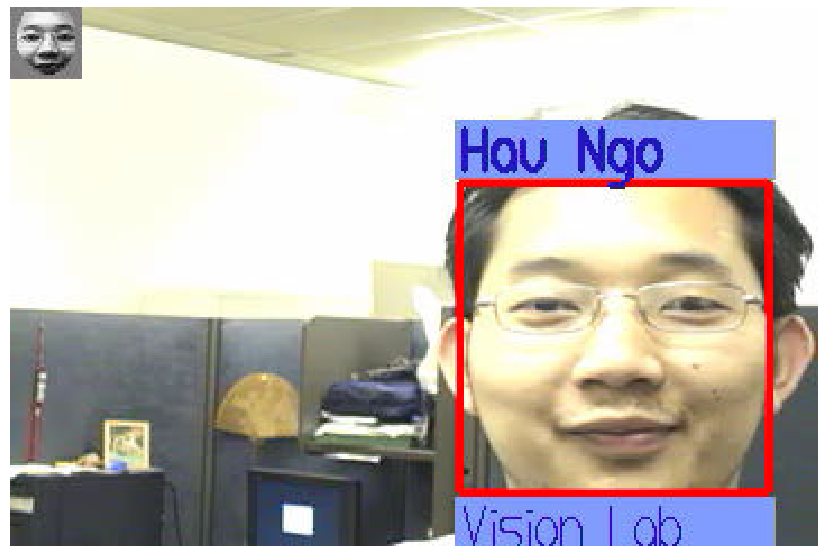
Predict:



Measure:



Demo



Video Quality

- Video quality is not good enough on current surveillance cameras
- Video quality is affected due to:
  - record to video tape
  - quality of frame grabbers
  - video compression
- Can we recognize these faces accurately?

Conclusion

- We can obtain good results in controlled conditions with good video quality
- Face recognition in video is still not practical on real world surveillance video
- We need to improve the algorithms
- Also need better quality video from surveillance cameras