Applications of Hidden Markov Models

Narayanan Ramanathan

Outline

- A brief introduction to Hidden Markov Models
- Three applications of HMMs
  - Human identification using Gait
  - Human action recognition from Time Sequential Images
  - Facial expression identification from videos
- Discussions & Conclusions

Hidden Markov Models (HMMs) – A General Overview

- HMM : A statistical tool used for modeling generative sequences characterized by a set of observable sequences.
- The HMM framework can be used to model stochastic processes where
  - The non-observable state of the system is governed by a Markov process.
  - The observable sequences of system have an underlying probabilistic dependence.

Hidden Markov Model

$\lambda = (A, B, \pi)$ → HMM Model Parameters

Initial state probabilities

$$\pi_i = P[q_1 = S_i], \quad 1 \le i \le N.$$

State Transition Matrix

$$A = \{a_{ij}\}$$

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i],$$

$$1 \le i, j \le N.$$

Observation Probability Matrix

$$B = \{b_j(k)\}$$

$$b_j(k) = P[v_k \text{ at } t | q_t = S_j]$$

$$1 \le k \le M \quad 1 \le j \le N$$

Three Basic Problems in HMMs

- Given a set of observation sequences $O = O_1 O_2 \cdots O_T$ and the HMM parameters $\lambda = (A, B, \pi)$, computing the probability $P(O|\lambda)$

- Given a set of observation sequences $O = O_1 O_2 \cdots O_T$ and the HMM parameters $\lambda = (A, B, \pi)$, computing the optimal state sequences

- Given a set of observation sequences $O = O_1 O_2 \cdots O_T$ adjusting the HMM parameters $\lambda = (A, B, \pi)$ to maximize the probability $P(O|\lambda)$

Things you'll need to be familiar with……

- Forward Algorithm / Backward Algorithm
- Viterbi Decoding
- Baum Welch Algorithm (Expectation Maximization)
- K-means clustering
- Vector Quantization etc.

Human Identification using Gait

- USF Gait dataset

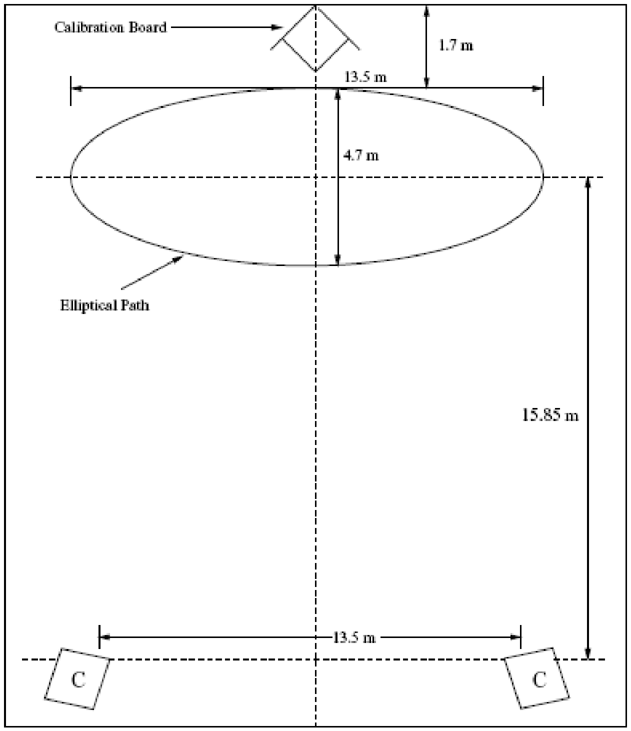




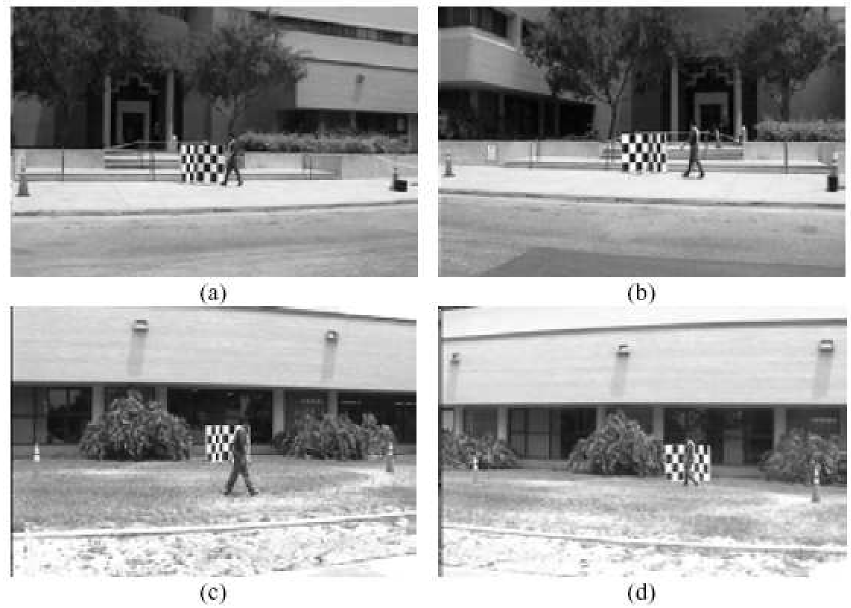
  - Dataset comprises of 122 individuals
  - 12 different Probe sets (different sessions, walking surfaces, shoe type, w/o briefcases, camera orientation)

Can we characterize Human gait using Hidden Markov Models ?

Human Identification using Gait



The set-up environment



Two views : Camera L & R
Two different surfaces : Grass and Concrete

Motivation

- Human Gait is often studied as a collection of gait cycles. A Gait cycle corresponds to : rest position – right foot forward – rest position – left foot forward – rest position.
- The two inherent components of human gait :
  - Structural component : One's physical features
  - Dynamic component : The body's motion dynamics (the coordinated hand and leg movements)

Can the structural and dynamical aspects of Human gait be captured using a Hidden Markov Model framework ?

Motivation (contd)

- From a human gait recognition perspective, what is the physical significance of the HMM parameters ?

$$\lambda = (A, B, \pi)$$

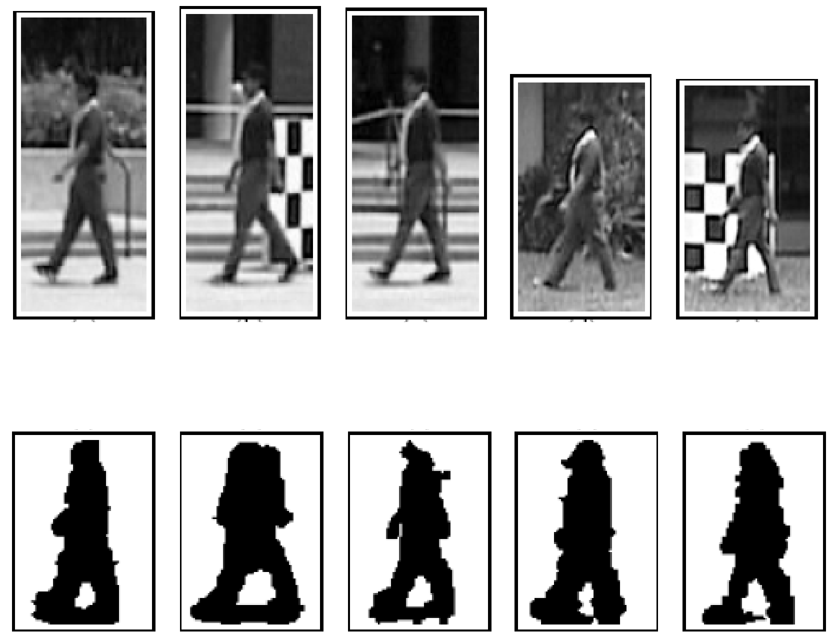
$\pi \rightarrow$ Initial state probabilities

$A \rightarrow$ State Transition matrix

$B \rightarrow$ Observation Probability matrix

Silhouette Extraction :

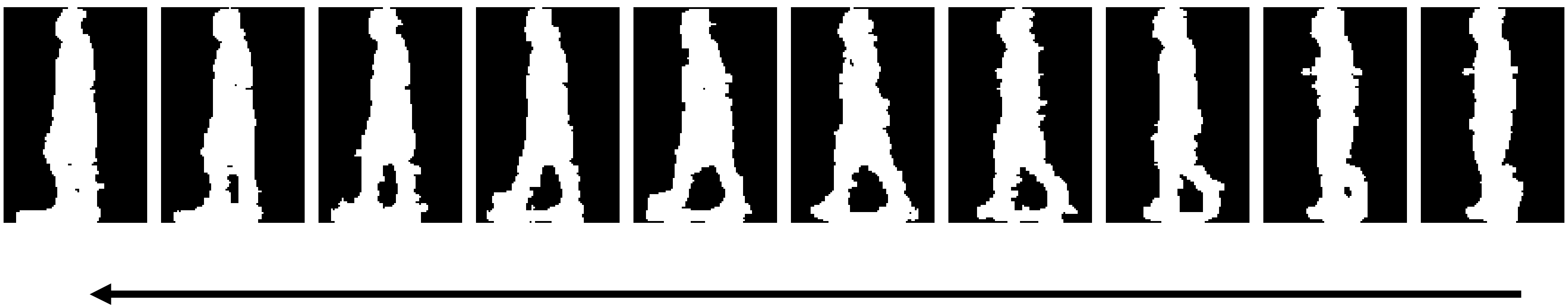
- For each frame in the gait video, a bounding box was defined manually.
- A background model is built using statistics of pixels outside the bounding box
- Having learnt the background distribution, the pixels within the bounding box are classified as foreground or background pixels.



Silhouette extraction results. The top row illustrates the bounding boxes defined over each frame

HMM : Observation symbols

- Using Background subtraction, the binarized video sequences are extracted

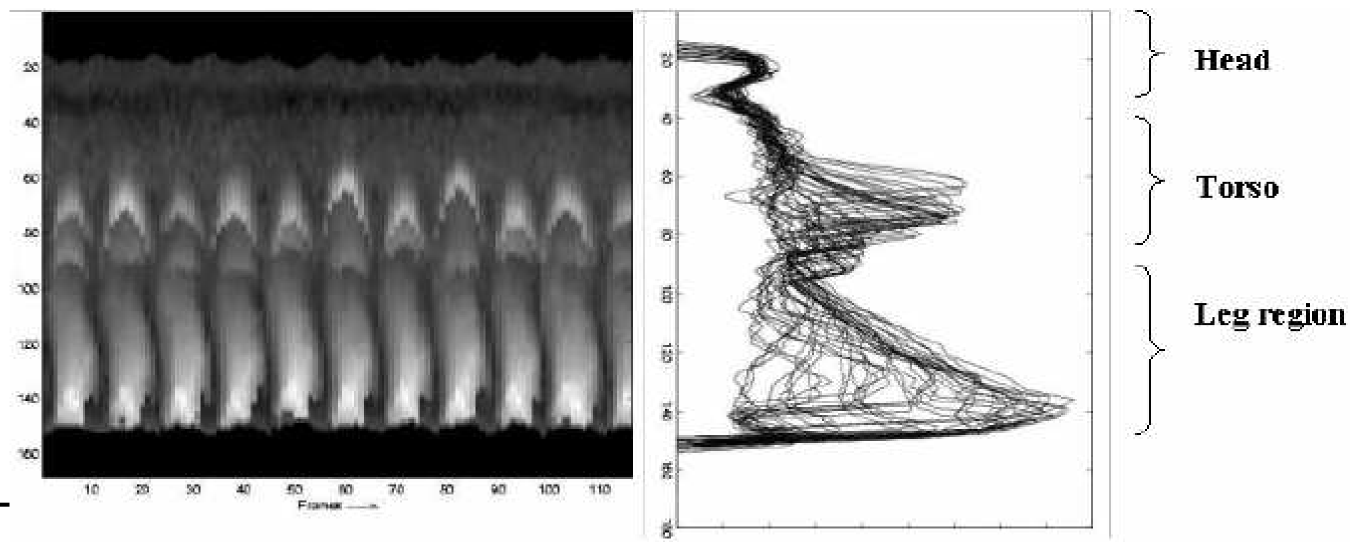


This corresponds to a one half of a gait cycle :
Rest position – right foot forward – rest position – left foot forward – rest position

The observation symbols for the HMM problem :
they are functions of such binarized silhouettes :

HMM : Observation symbols

- Kale et al. 2004, define two interpretations to the observation symbols for the HMM framework :
  - In the first case, the entire background subtracted silhouette is taken as the observation symbol.
  - In the second case, the width vector is extracted from each frame. Frame-to-Exemplar distance (FED) is defined over each frame and this is taken as the observation symbol.



HMM : System State identification

- System state identification is often seen analogous to the design of a code book
  - Criterion : Minimizing the overall distortion in such a representation

$$\mathbf{e}_i = \arg\min_{\mathbf{e}} E\left(d(\mathbf{x},\mathbf{e})|\mathbf{x} \in C_i\right)$$

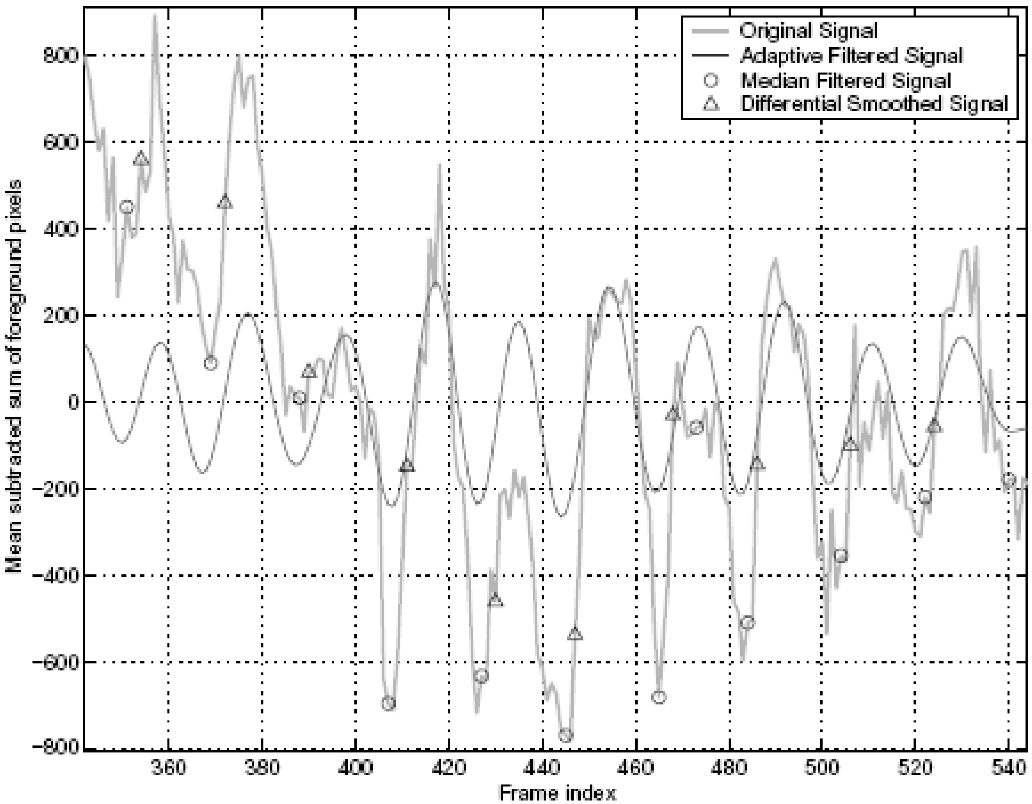
exemplar (system state)

Distance measure

Vornoi cells

Model order (no: of exemplars) is selected based on a plot of the overall distortion

HMM : System State identification



A plot of sum of the foreground pixels across each frame. The boundaries of gait cycles can be identified.

Kale et al (2004) define 6 states for the gait recognition system

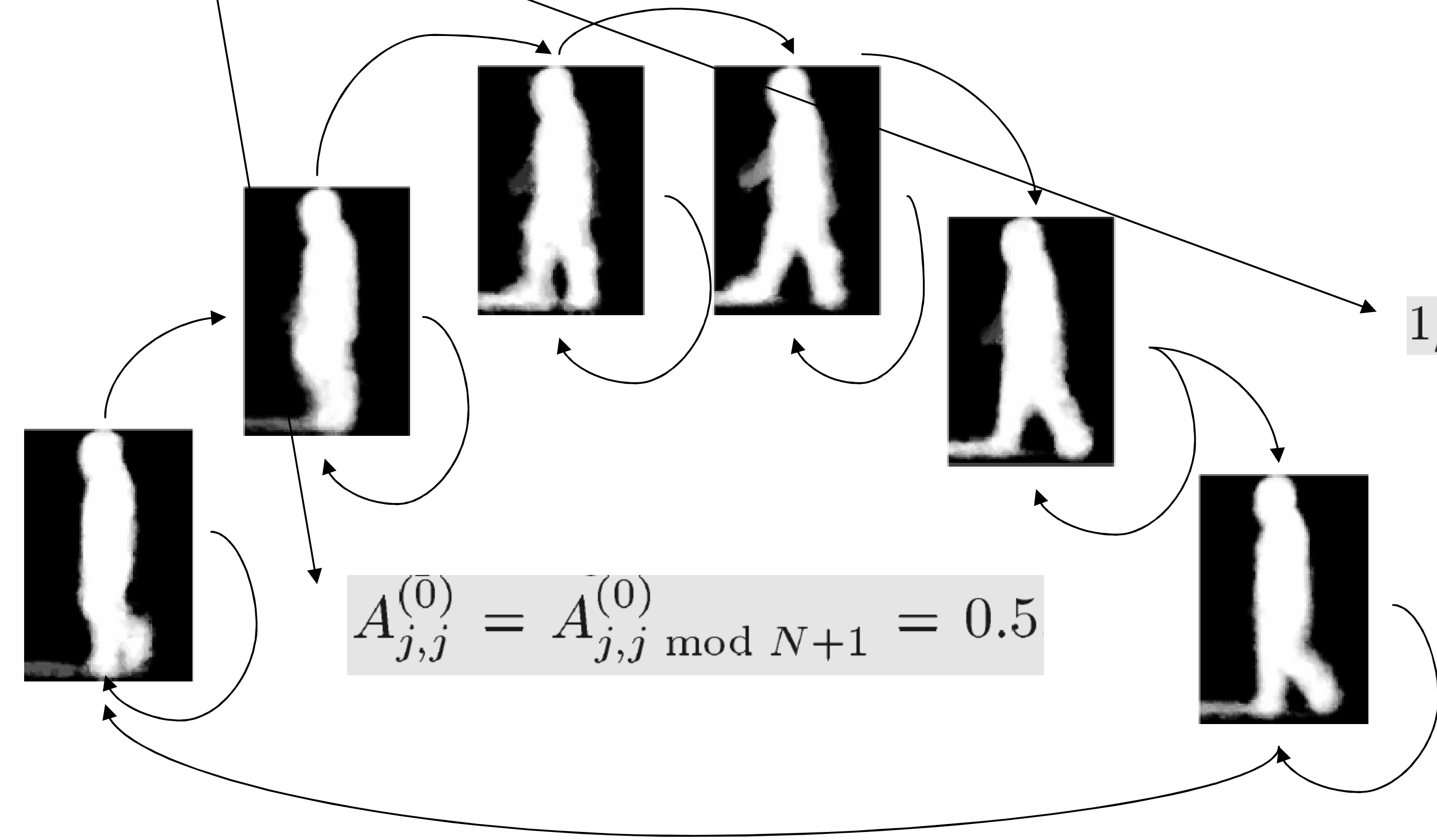
HMM : System State

- The initial exemplars for a walking sequence are computed as follows :
  - The gait cycle boundaries are identified for a walking sequence
  - Each gait cycle is partitioned in 6 groups of temporally adjancent stances.
  - The averages of all stances that belong to a particular partition is computed & hence exemplars (or system states) are identified.

HMM : Parameter initialization

$\lambda = (A, B, \pi)$

$$b_n\left(\mathbf{x}(t)\right) = P\left(\mathbf{x}(t)|\mathbf{e}_n\right) = \beta e^{-\alpha D(\mathbf{x}(t),\mathbf{e}_n)}$$



$1/N$

$$A_{j,j}^{(0)} = A_{j,j \bmod N+1}^{(0)} = 0.5$$

HMM : Training Phase

- Iterative refining is performed in two stages :
  - Using current values of Exemplars (Eo) and Transition Matrix (Ao) , Viterbi decoding is performed on the input sequence and the most probable sequence of states is obtained :
    $\mathcal{Q} = \{q_1^{(i)}, q_2^{(i)}, \ldots, q_T^{(i)}\}$
  - The corresponding observation index (set of all time instants when a particular state was observed) is provided by
    $$\mathcal{T}_j^{(i)} = \{t : q_t^{(i)} = j\}$$
  - The new set of exemplars are re-esimated using the above observation indices $\mathbf{E}_j^{(i+1)} = \arg_{\mathbf{E}} \min \sum_{t \in \mathcal{T}_j^{(i)}} D(\mathbf{Y}_t, \mathbf{E})$

HMM : Training Phase

- Using estimated exemplars E (at time t+1) and state transition matrix A (at time t), we estimate A (at time t+1) using Baum Welch algorithm.

- Computing E (at t+1) => computing B (at t+1)

- Kale et al (2000) re-initialize the initial state probabilities to 1/N at every time instant.

HMM : Testing Phase

- Given a test sequence, we compute the probability

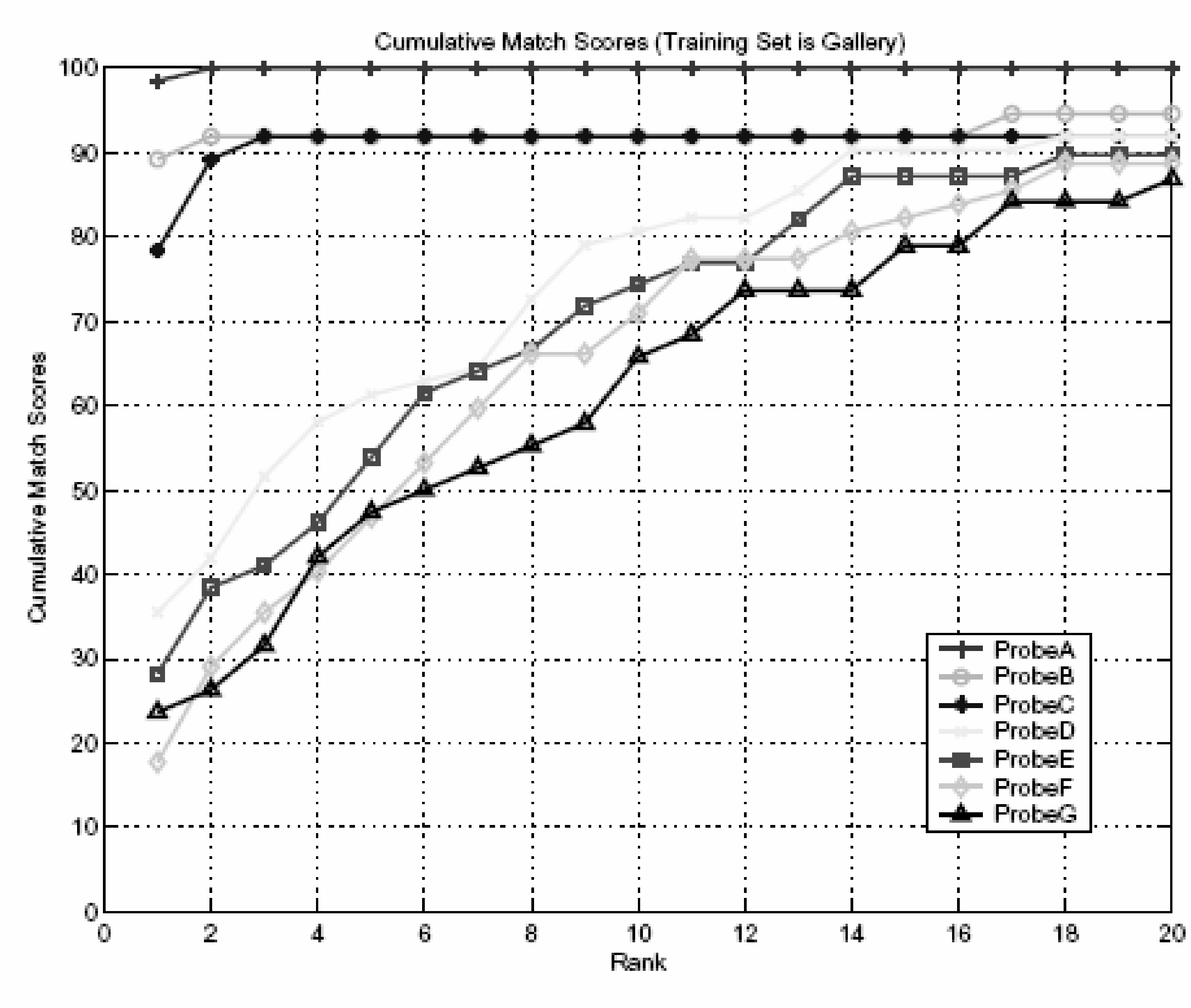
$$P_j = \log\left(P(\mathcal{Y}|\lambda_j)\right)$$

Test sequence

HMM parameters corresponding to the j'th individual in the gallery

$$ID = \arg_p \max_{\mathcal{Q},p} \Pr[\mathcal{Q}|\mathcal{X},\lambda_p] \longrightarrow \text{Match ID}$$

Recognition results on USF dataset



Recognizing Human Action using HMMs

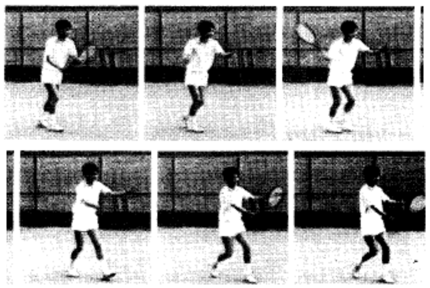
- Yamato et al (1992) use HMM to classify human actions in time-sequential images (in our case, sports activities).
  - Backhand volley
  - Backhand stroke
  - Forehand volley
  - Forehand stroke
  - Smash
  - Service

What is the intuition behind using a HMM framework to perform action classification ?

Each of the aforementioned activities can be characterized by a set of stances that are temporally related.

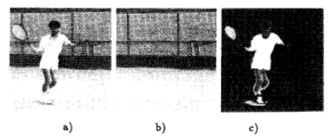
Further, each activity can be associated with a certain number of observable symbols that are associated with each characteristic stance

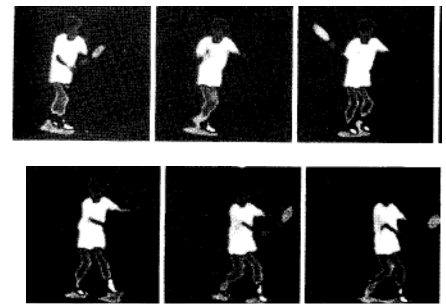
HMM : Training Phase



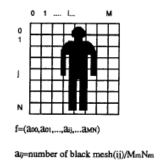
Input Video sequence

Background subtraction





Binarized sequence



Mesh representation

Feature Vector sequence

Vector quantization

Observable symbols fed to HMM for parameter estimation

HMM : Recognition

For a C class classification problem : Given the HMM parameters
For each of the C classes and the Observation sequence, the class
Is determined as :

$$c^* = \arg\max_i(\Pr(\lambda_i|O)) \Longrightarrow$$

Evaluating $\Pr(O|\lambda_i)$
(forward algorithm)

Baum – Welsh method to estimate parameters :

$$\begin{aligned}\gamma_t(i) &\equiv P(s_t = q_i|O_1,\ldots,O_T,\lambda) \\ &= \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)}.\end{aligned}$$

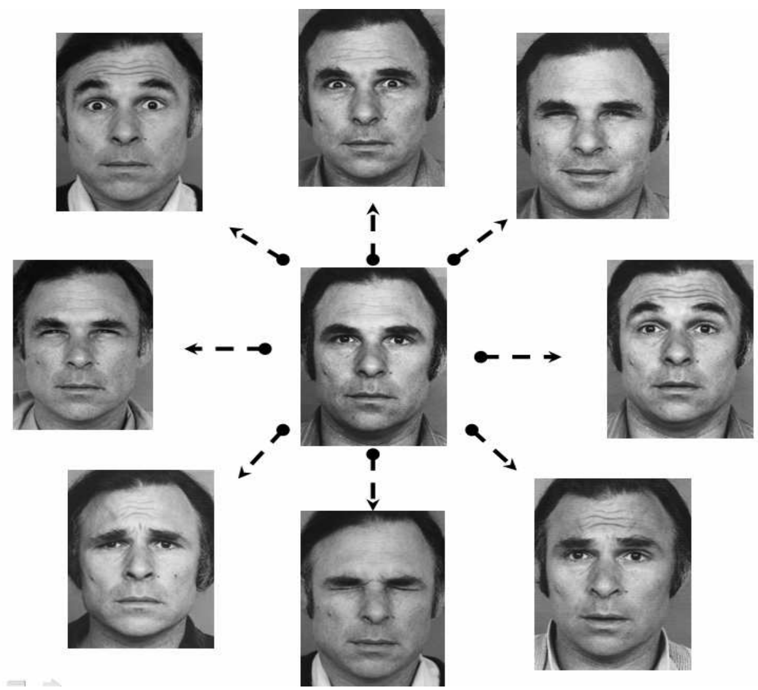
$$\begin{aligned}\xi_t(i,j) &\equiv P(s_t = q_i, s_{t+1} = q_j|O_1,\ldots,O_T,\lambda) \\ &= \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)}.\end{aligned}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)};$$

$$\bar{b}_i(k) = \frac{\sum_{t\in\{t|O_t=v_k\}}\gamma_t(i)}{\sum_{t=1}T\,\gamma_t(i)};$$

$$\bar{\pi}_i = \gamma_1(i).$$

Face Expression Characterization – Using HMM framework



FACS : Facial Action Coding Systems – prescribes one of the most comprehensive means to characterize Facial Expressions

- Discrete deformations of face regions are referred to as Action Units (AUs.)

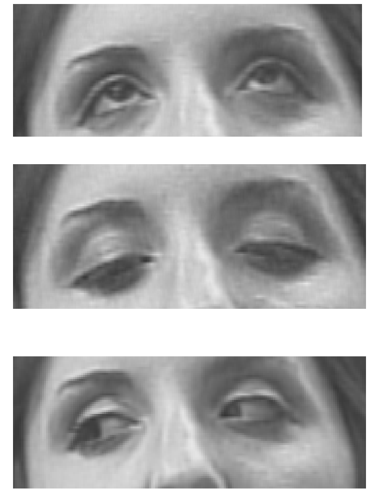
- Divides facial expression into upper facial and lower facial expressions

- 44 Basic AUs and 14 special AUs

Some illustrations…..

Upper Facial expressions

Lower Facial expressions



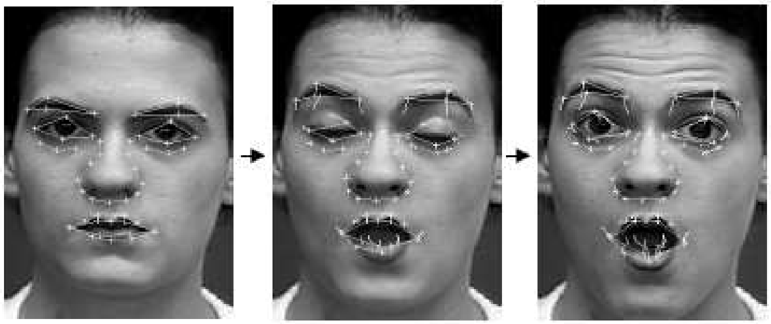
Eye Movements (special instances)

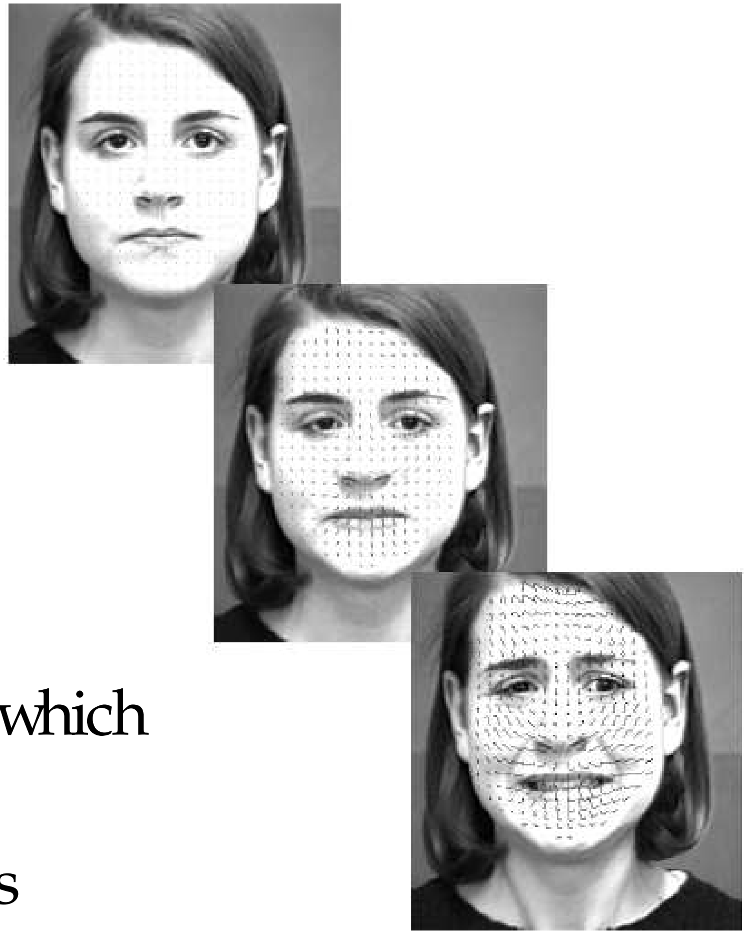
How relevant is the HMM framework to characterize facial expressions in video sequences ?

Relevance of HMM framework in characterizing Facial Expressions

- The hidden state of HMMs Ł the hidden emotional state of the individual.
- The observable symbols of HMMs Ł the feature vectors extracted from face videos
- The State Transition matrix and Observation probability matrix of HMMs Ł Dynamical information extracted from videos accompanied by observation symbols extracted using VQ.

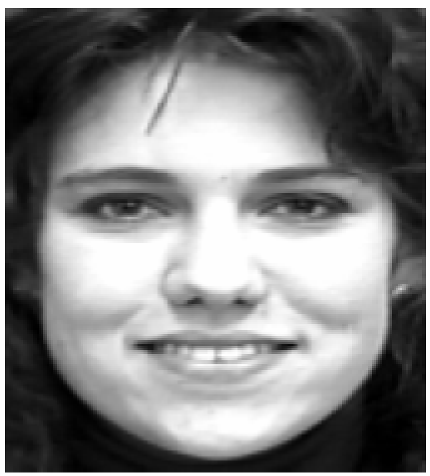
Feature selection



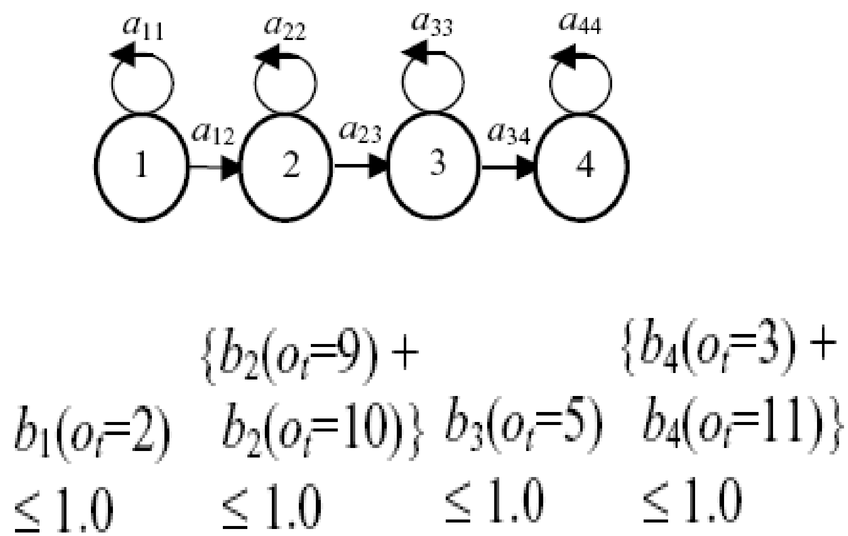


- Features selected for expression characterization could be
  - The original intensity image
  - Fiducial features from each image (which could be tracked across frames)
  - Dense optical flow of facial features

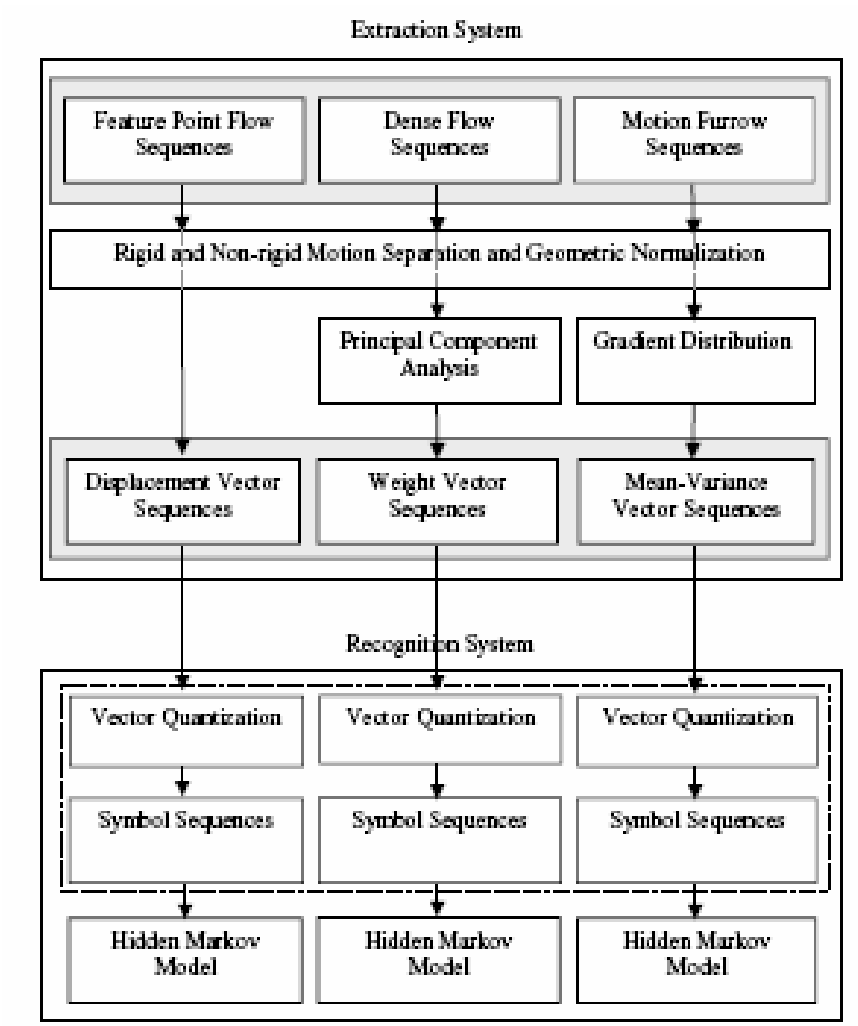
Some examples of HMMs designed for specific facial expressions :



(4) AU6+12+25



Overview & Sample Results



HMM / Human	AU4	AU1+4	AU1+2	Recognition Rate
AU4	22	3	0	88%
AU1+4	4	19	2	76%
AU1+2	0	2	23	92%