Review of Gesture Recognition Techniques

By François Rioux,

Feb. 23, 2004

Presentation Outline

- Overview of the problem
- Human gesture representation
- Features selection
- Recognition techniques
  - HMM and « improved » HMMs
  - Other techniques

Gesture Recognition = Complex Task

- Motion modeling
- Motion analysis
- Pattern recognition
- Machine learning
- Psycholinguistic studies
- …

Human Gesture Representation

- Psycholinguistics research by Stokoe:
  - Hand shape
  - Position
  - Orientation
  - Movement

Human Gesture Representation(cont)

- Application scenarios of gestures
  - Conversational
  - Controlling
    - eg: vision-based interfaces
  - Manipulation
    - eg: Interact with virtual objects
  - Communication
    - eg: Sign language → Highly structured

More on Communicative Gestures

- Highly structured
- 3 phases
  - Preparation
  - Stroke (dynamic part)
  - Retraction
- The dynamic part contains the information

Features Extraction

- Good feature extraction is CRUCIAL
- Static hand posture
  - Fingertips
  - Finger direction
  - Hand contour
  - …

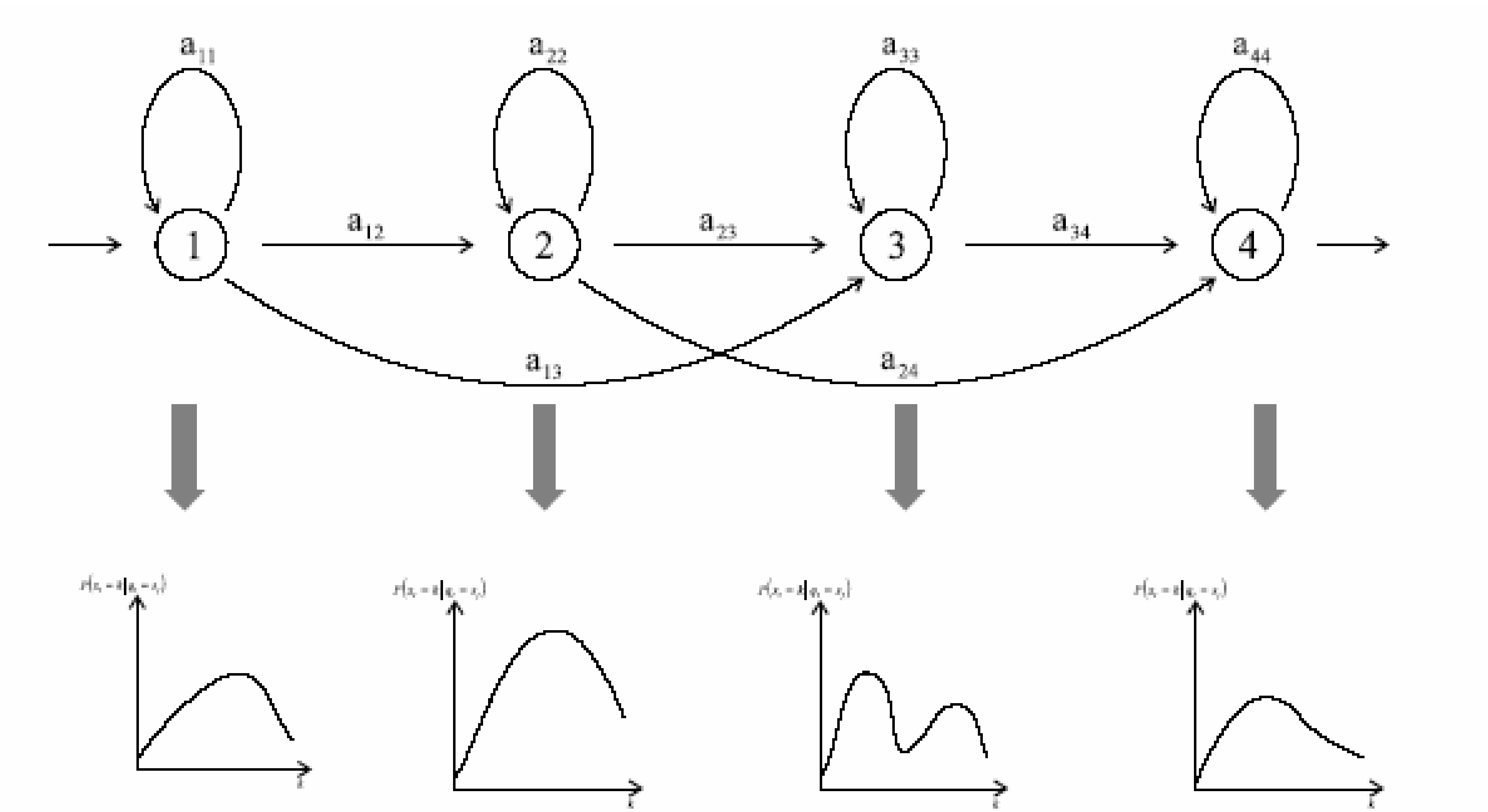
Features Extraction (cont)

- Dynamic of gestures
  - Spatial features
    - eg: 2D location of hands. More general: 3D
  - Temporal features
    - Need of statistical method independent of time
- Choosing only most relevant features
  - MEF: Most Expressive Features
    - Karhunen-Loeve projection
  - MDF: Most Discriminative Features

Temporal Gesture Modeling and Recognition

- Similar to speech recognition
  - HMMs can be used for low level recognition
  - However more complicated than speech…
- Gesture semantics for high level recognition

(Hidden Markov Model Overview)



Modeling the Dynamics

- Low-level dynamics of human motion
  - Useful for human motion recognition
  - Quantitative representation of simple motion
  - Not sufficient for complex motion

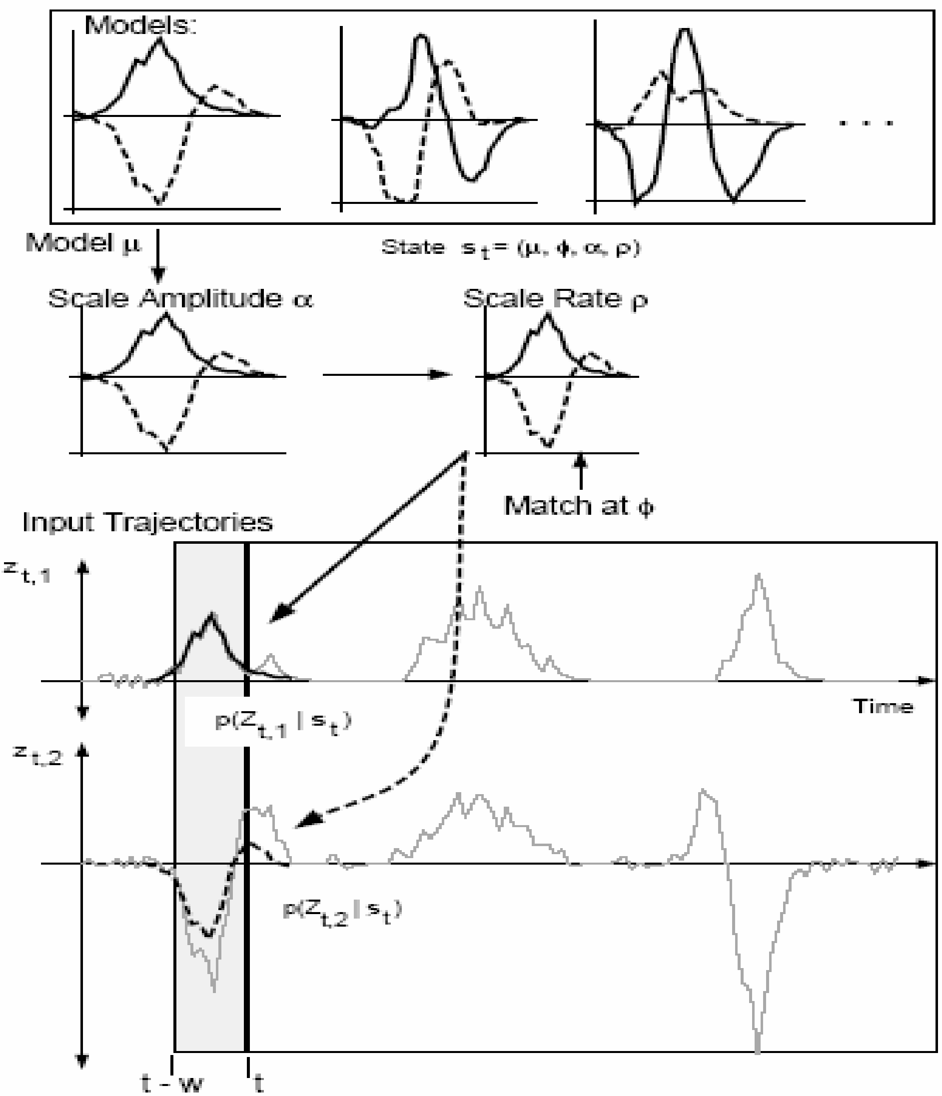
- Kalman filter:
  - Estimate, interpolate and predict motion
  - Not sufficient, Gaussian assumption

Modeling the Dynamics (cont)

- Condensation Algorithm (Black & Jepson, 1998)
  - Combine *Dynamic Time Warping* (DTW) and HMM
    - Capture detailed information of the motion
    - Has a probabilistic framework
  - Goal: match trajectory models to input data
    - Extension to the standard condensation algorithm

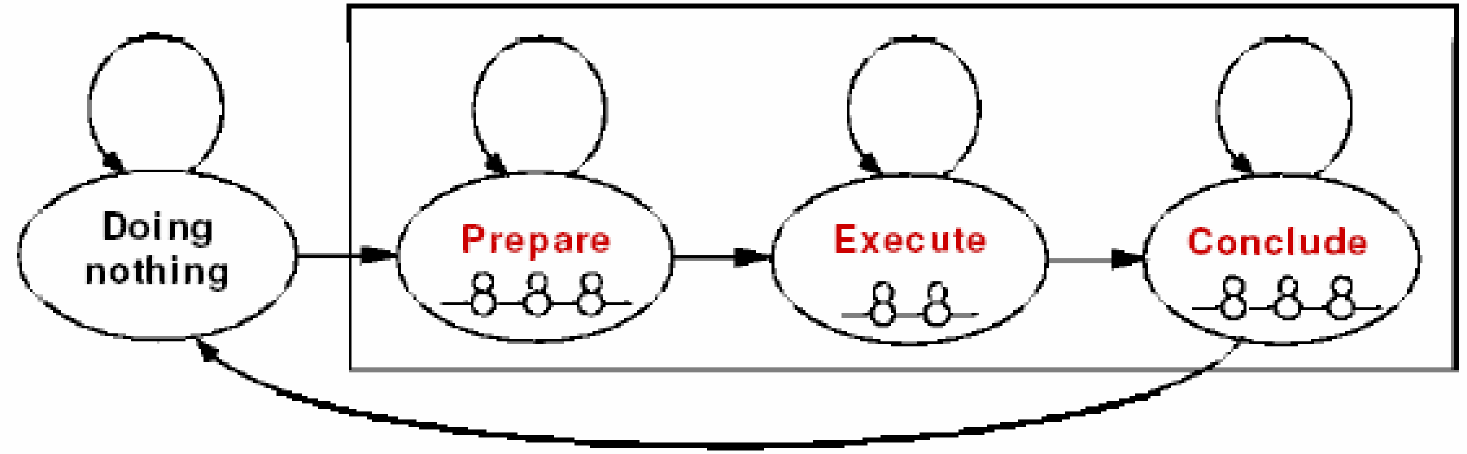
Modeling the Dynamics (cont)

- Condensation Algorithm (Black & Jepson, 1998)
- Tests on whiteboard
  - Slower than real time
  - Manual evaluation of transition probabilities
  - Training data pre-segmented



Modeling the Dynamics (cont)

- Modeling and Prediction of Human Behaviour
  - Pentland & Liu (1999)
  - Multiple dynamic models sequenced by Markov chain
    - Kalman filter + HMM (observations are innovations of Kalman filter)



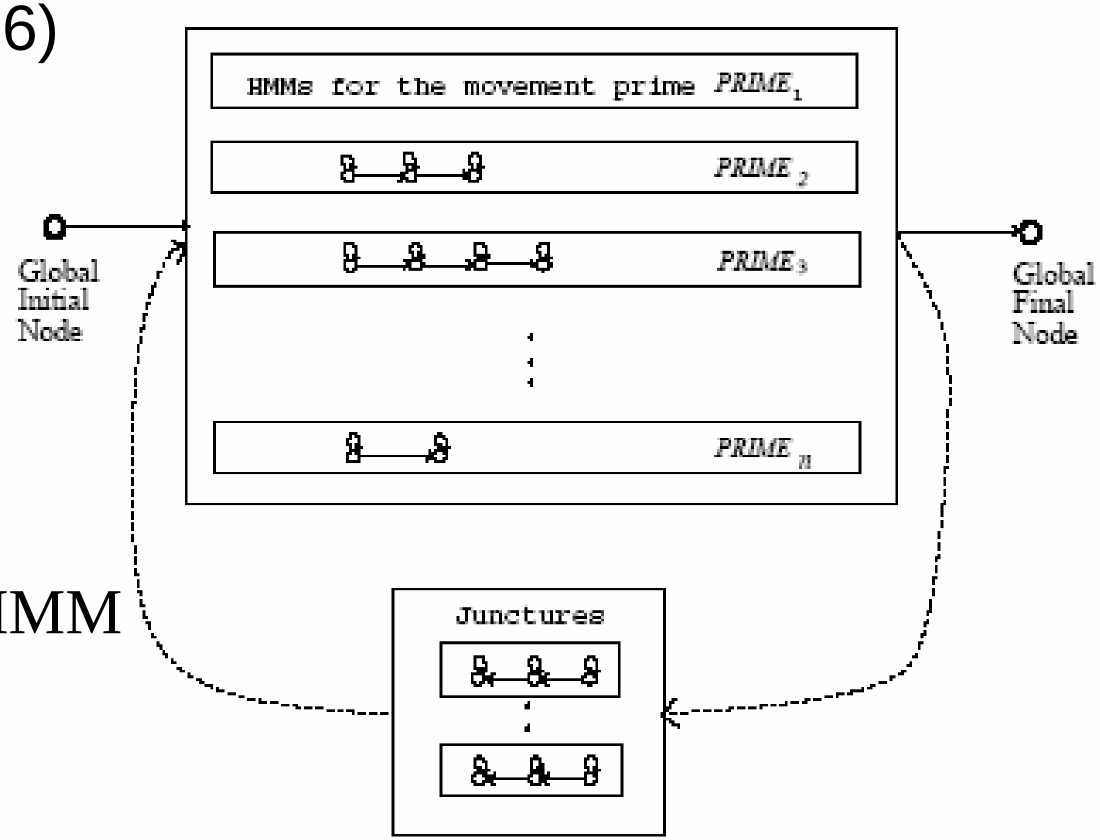
  - Application: prediction of driver's behaviour

Modeling the Semantics

- When modeling the dynamics is not enough
  - More complex gestures
- *Finite State Machine* (FSM)
- Extended variable-valued logic
  - Rule-based induction algorithm
- PNF network {past, now, fut}
  - Constraints on states based on previous
- Bayesian network

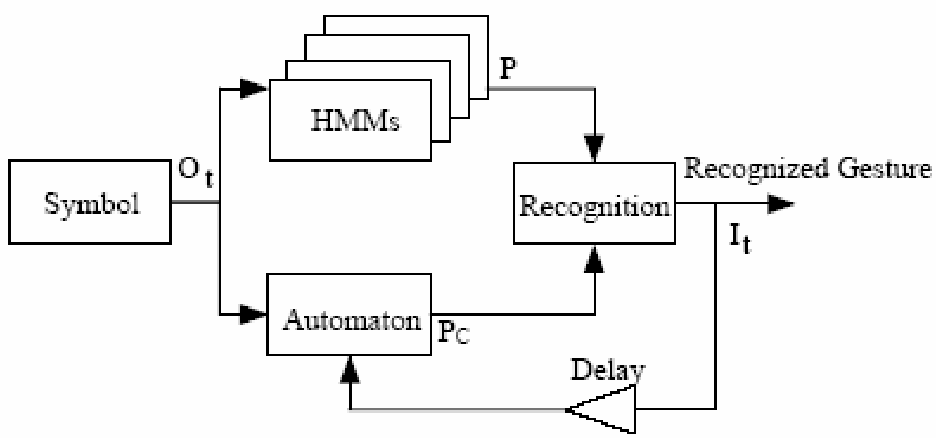
HMM Framework

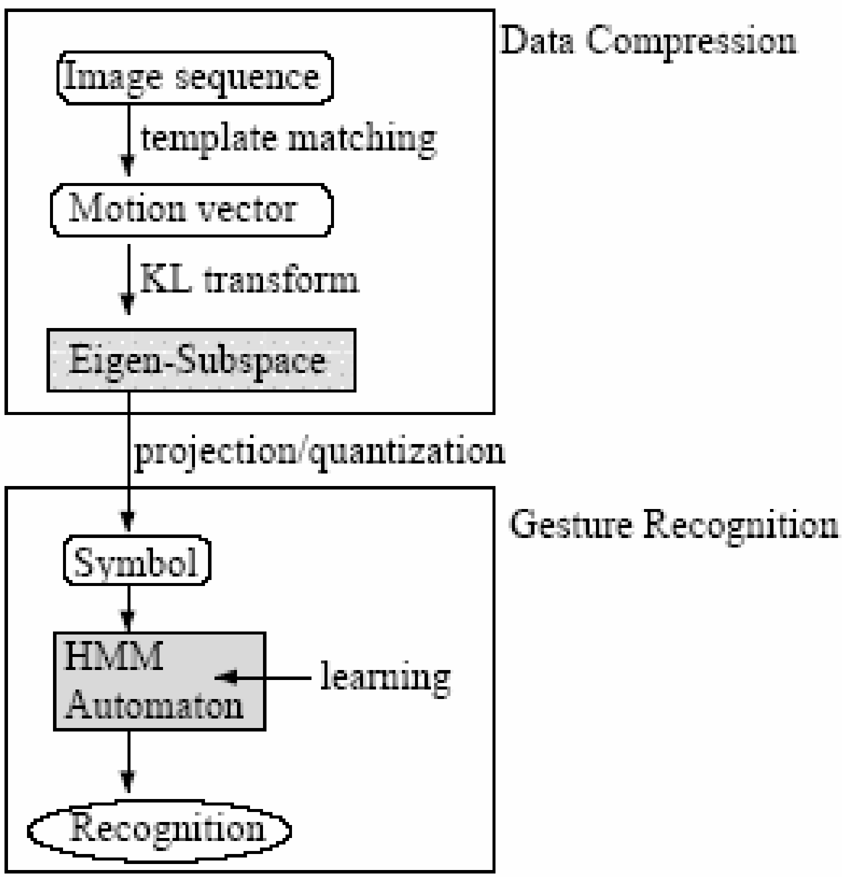
- Capacity to model low level and semantics
  - Nam & Wohn (1996)
- Problems:
  - P(O|λ)
  - max P(S|O,λ)
  - Training => λ
- Variation:
  - Multi dimensional HMM



HMM Framework (cont)

- **Real-Time Context-based Gesture Recognition Using HMM and Automaton**
  - Iwai, Shimizu & Yachida (1999)
  - KL-transform to compress input data
  - Context-based HMM



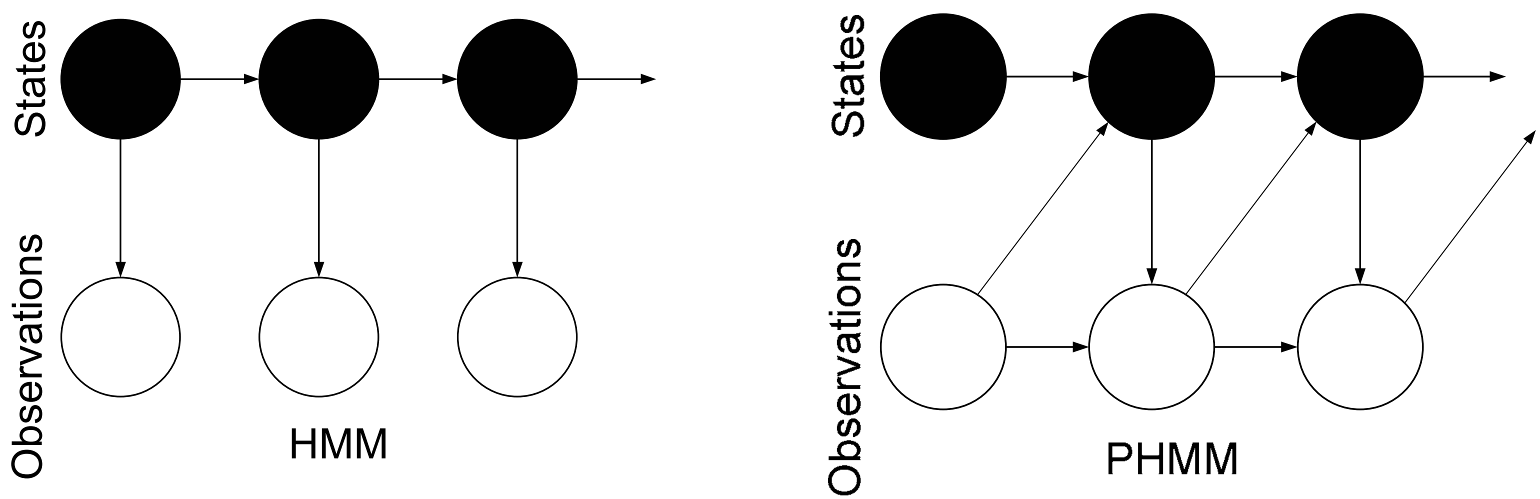


HMM Framework (cont)

- HMM: only piecewise stationary processes
  - Gestures: all parts are transient =>
    HMM Not always suitable for gesture rec.
  - Need improvements to HMM

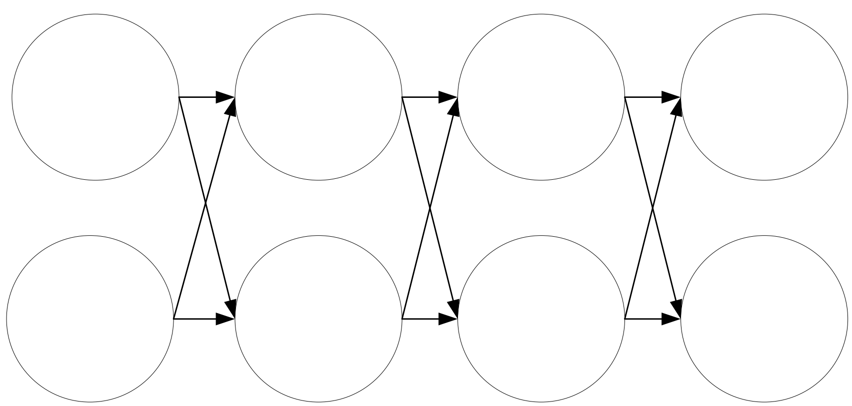
HMM Framework (cont)

- Partly Hidden Markov Model
  - Kobayashi & Haruyama (1997)
  - Second order model



HMM Framework (cont)

- Markov condition violated => HMM fails
- Coupled Hidden Markov Models (CHMM)
  - Brand, Oliver & Pentland (1997)
  - Coupling HMMs to model interactions between them



Other Techniques

- Multi-class multi dimensional discriminant analysis
  - Cui & Weng (1996)
  - Self-organizing framework
- Action Recognition using Probabilistic Parsing
  - Bobick & Ivanov (1998)
  - HMM for low level processing of features
  - Probabilistic parsing using a particular grammar and ML
- Time delay neural networks
  - Yang & Ahuja (1999)

Conclusion

- Important to model a gesture correctly
- Features extraction can be hard
- Many features => CPU intensive
- Need to keep only most relevant features
- HMMs are good, but…
- Other method???

References

- Wu & Yang, *Vision-based Gesture Recognition: A Review*, Lecture Notes in Artificial Intelligence, 1999