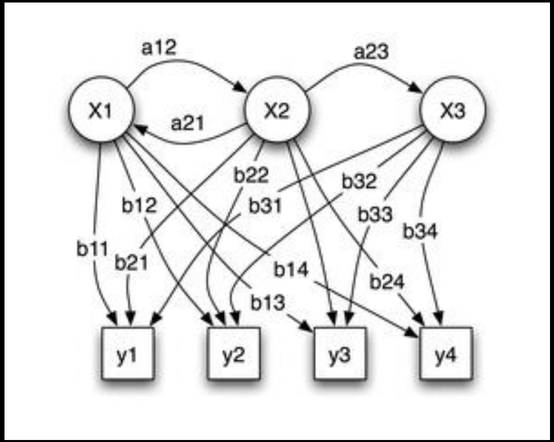
CS4495/6495
Introduction to Computer Vision

8D-L3 *Hidden Markov Models*

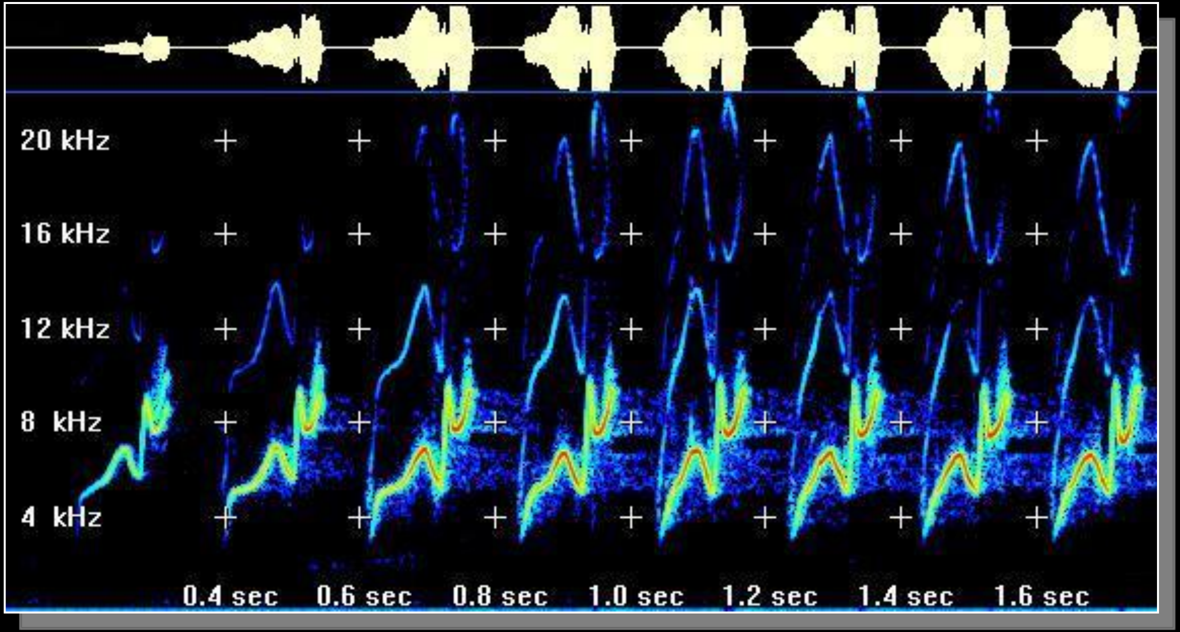


Outline

- Time Series
- Markov Models
- Hidden Markov Models
- 3 computational problems of HMMs
- Applying HMMs in vision – Gesture Recognition

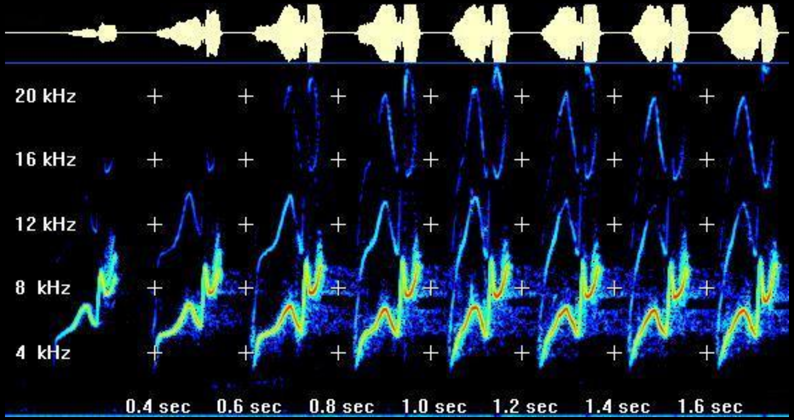
Audio Spectrum

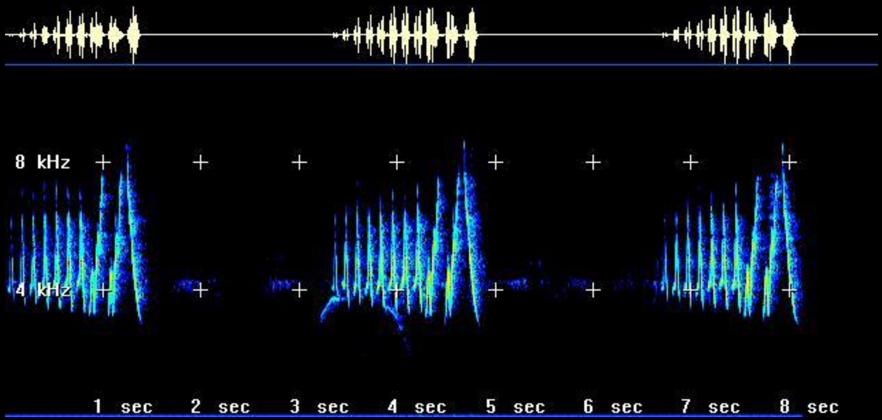
Audio Spectrum of the Song of the Prothonotary Warbler



Prothonotary Warbler

Chestnut-sided Warbler

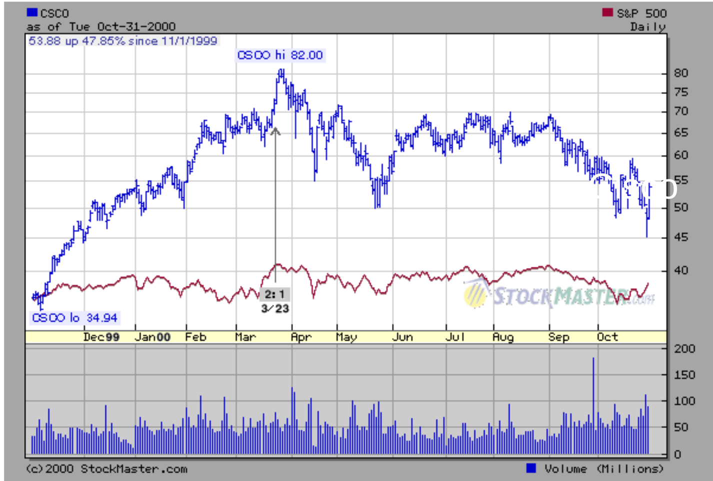


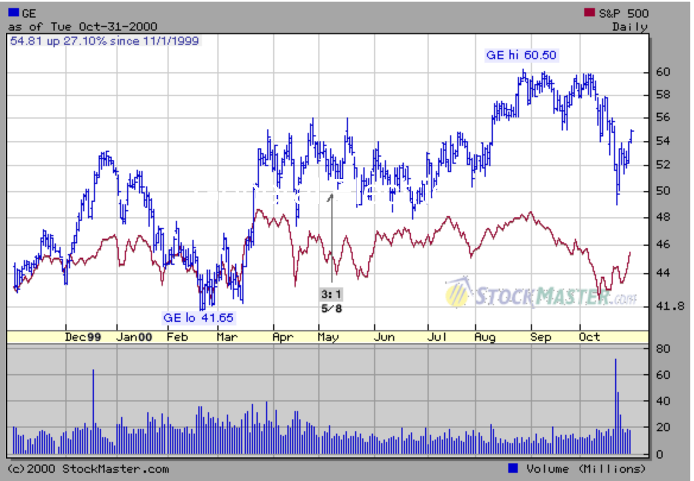


Questions One Could Ask

- What bird is this?
- How will the song continue?
- Is this bird sick?
- What phases does this song have?

- ➢ Time series classification
- ➢ Time series prediction
- ➢ Outlier detection
- ➢ Time series segmentation


CSCO
S&P 500
as of Tue Oct-31-2000
Daily
53.88 up 47.85% since 11/1/1999
CSCO hi 82.00
2:1
3/23
CSCO lo 34.94
80
75
70
65
60
55
50
45
40
Dec99 Jan00 Feb Mar Apr May Jun Jul Aug Sep Oct
200
150
100
50
0
(c)2000 StockMaster.com
Volume (Millions)


GE
S&P 500
as of Tue Oct-31-2000
Daily
54.81 up 27.10% since 11/1/1999
GE hi 60.50
3:1
5/8
GE lo 41.65
60
58
56
54
52
50
48
46
44
41.8
Dec99 Jan00 Feb Mar Apr May Jun Jul Aug Sep Oct
80
60
40
20
0
(c)2000 StockMaster.com
Volume (Millions)


INTC
S&P 500
as of Tue Oct-31-2000
Daily
45.00 up 18.42% since 11/1/1999
INTC hi 75.83
2:1
7/31
INTC lo 34.75
75
70
65
60
55
50
45
40
35.0
Dec99 Jan00 Feb Mar Apr May Jun Jul Aug Sep Oct
350
300
250
200
150
100
50
0
(c)2000 StockMaster.com
Volume (Millions)


MSFT
S&P 500
as of Tue Oct-31-2000
Daily
68.88 down 25.44% since 11/1/1999
MSFT hi 119.94
MSFT lo 48.44
110
100
90
80
70
60
50
Dec99 Jan00 Feb Mar Apr May Jun Jul Aug Sep Oct
175
150
125
100
75
50
25
0
(c)2000 StockMaster.com
Volume (Millions)

Questions One Could Ask

- Will the stock go up or down?
- What type stock is this (eg, risky)?
- Is the behavior abnormal?

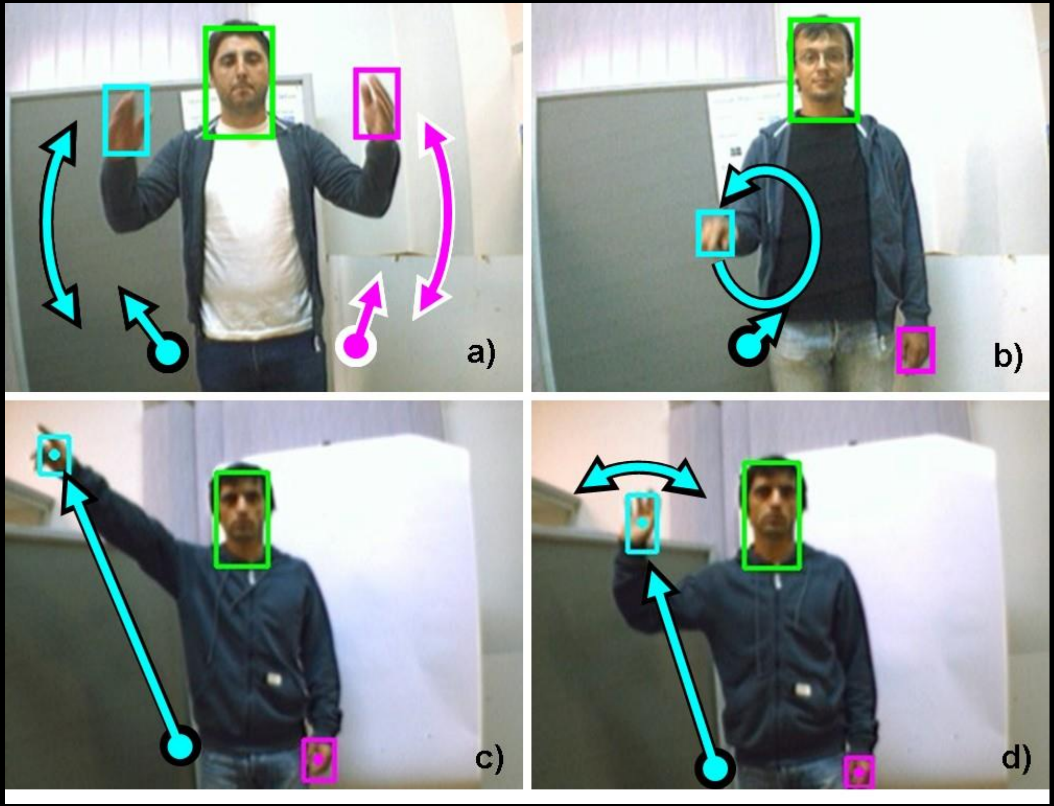
- Time series prediction
- Time series classification
- Outlier detection

Music Analysis

Questions One Could Ask

- Is this Beethoven or Bach?
  - ➢ Time series classification
- Can we compose more of that?
  - ➢ Time series prediction/generation
- Can we segment the piece into themes?
  - ➢ Time series segmentation

For vision: Waving, pointing, controlling?



The Real Question

- How do we model these problems?
- How do we formulate these questions as a inference/learning problems?

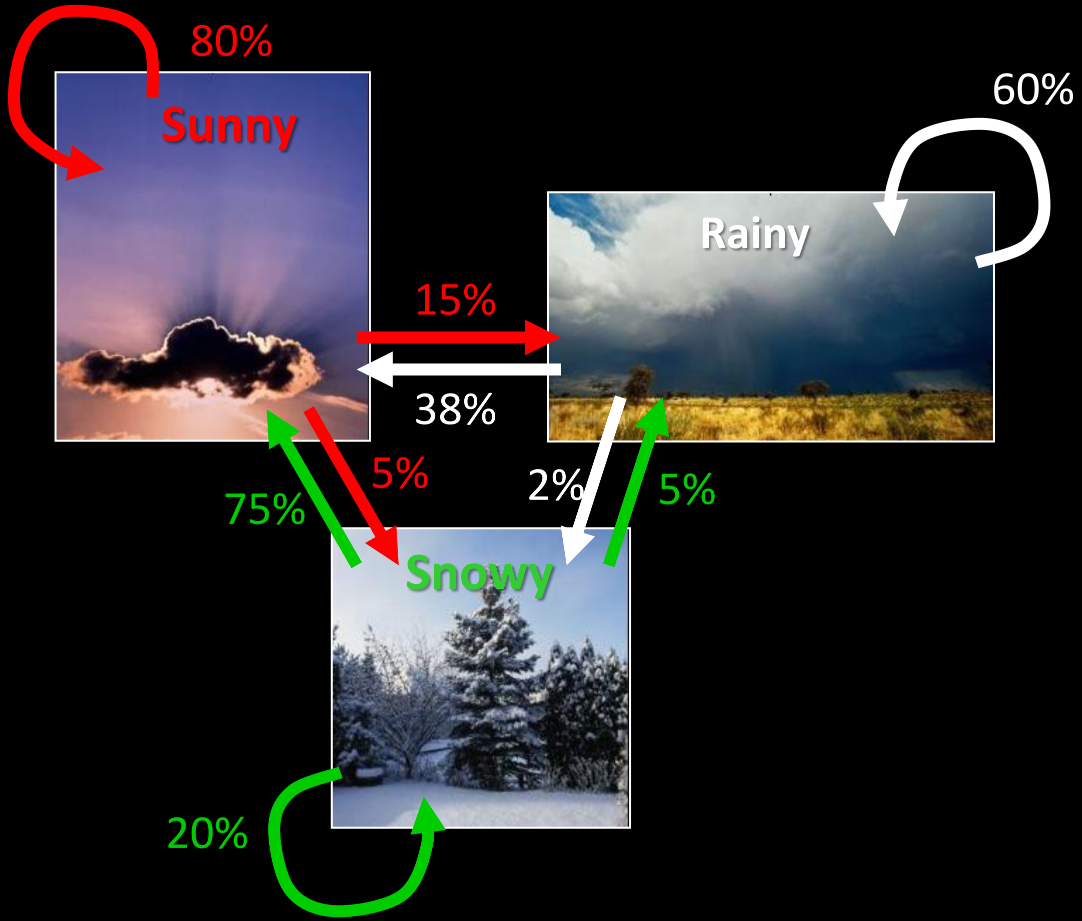
Outline For Today

- Time Series
- Markov Models
- Hidden Markov Models
- 3 computational problems of HMMs

Weather: A Markov Model *(maybe?)*

Probability of moving to a given state depends only on the current state:

1st Order Markovian



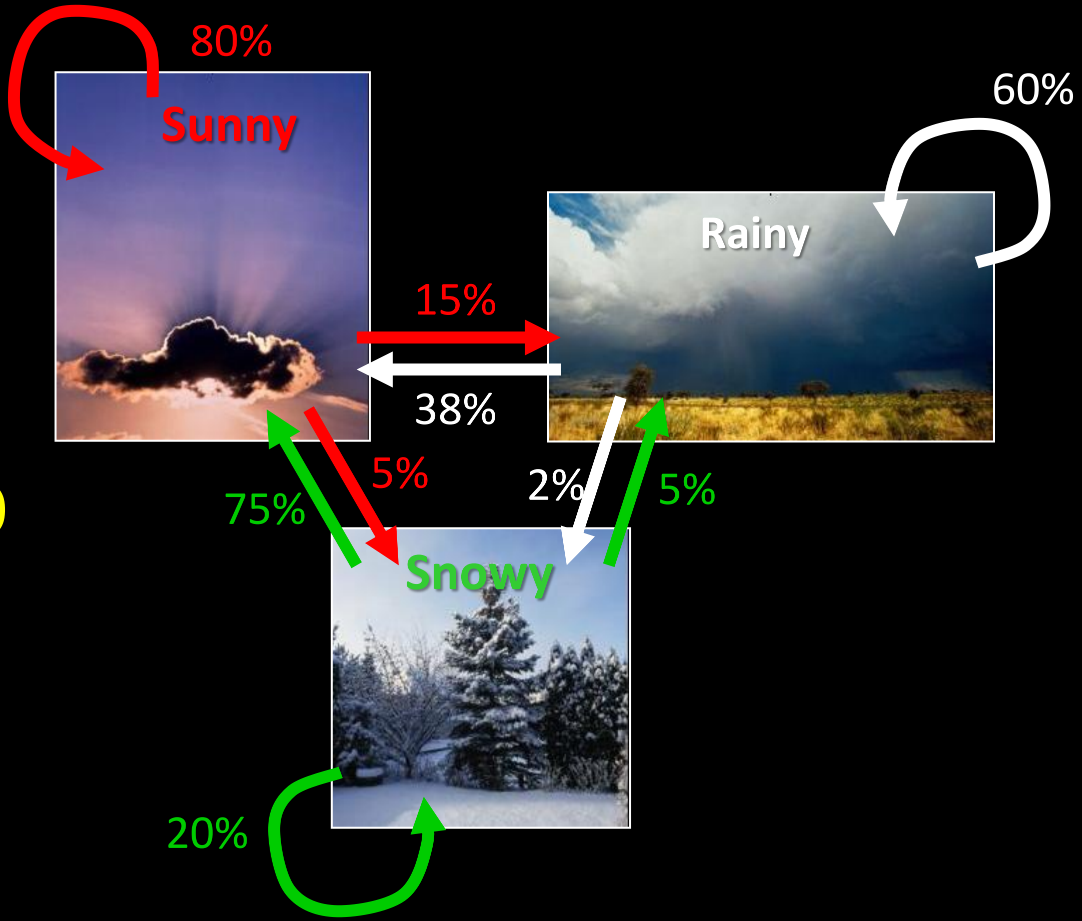
Ingredients of a Markov Model

- States:

  $\{S_1, S_2, ..., S_N\}$
- State transition probabilities:

  $a_{ij} = P(q_{t+1} = S_i \mid q_t = S_j)$
- Initial state distribution:

  $\pi_i = P[q_1 = S_i]$



Ingredients of a Markov Model

- States:

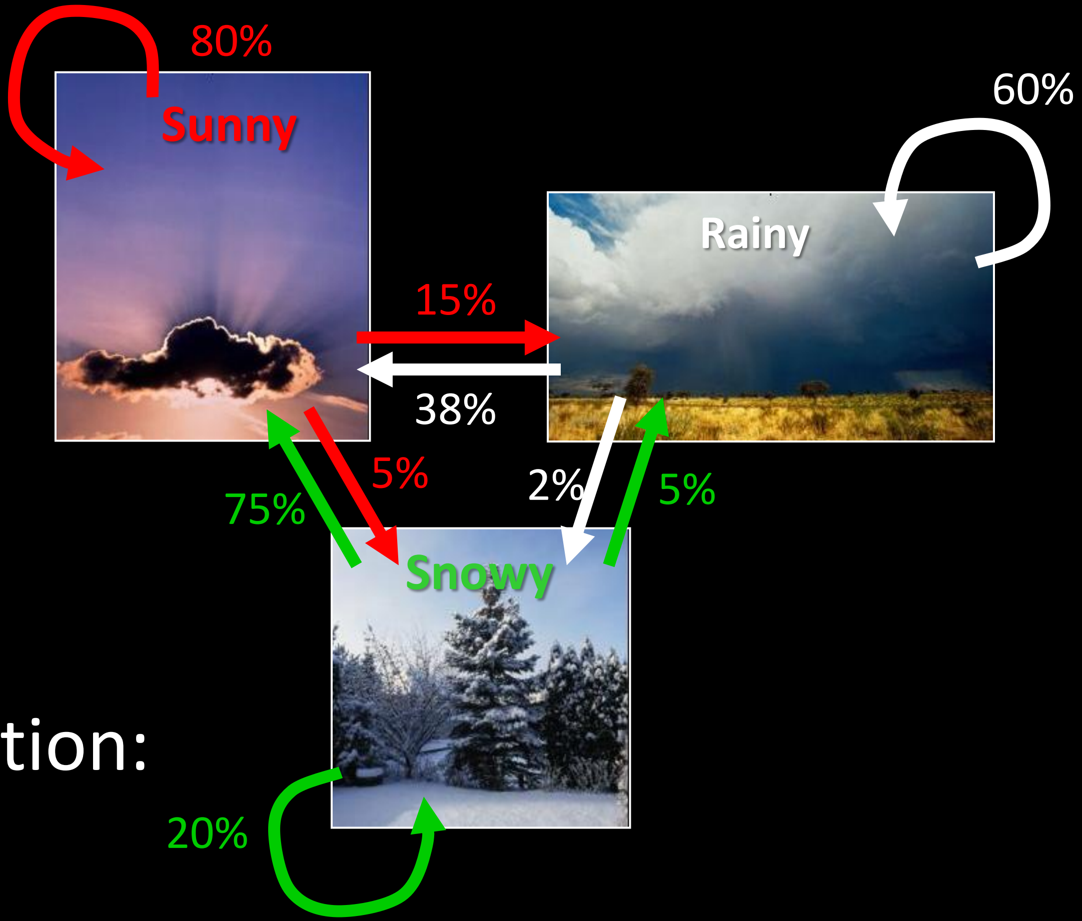
  $$\{S_{sunny}, S_{rainy}, S_{snowy}\}$$

- State transition
- probabilities:

  $$A = \begin{pmatrix} .8 & .15 & .05 \\ .38 & .6 & .02 \\ .75 & .05 & .2 \end{pmatrix}$$

- Initial state distribution:

  $$\pi = (.7 \quad .25 \quad .05)$$



Probability of a Time Series

- Given: 
- What is the probability of this series?

$$P(S_{sunny}) \cdot P(S_{rainy} \mid S_{sunny}) \cdot P(S_{rainy} \mid S_{rainy}) \cdot P(S_{rainy} \mid S_{rainy})$$

$$\cdot P(S_{snowy} \mid S_{rainy}) \cdot P(S_{snowy} \mid S_{snowy})$$

$$= 0.7 \cdot 0.15 \cdot 0.6 \cdot 0.6 \cdot 0.02 \cdot 0.2 = 0.0001512$$

$$A = \begin{pmatrix} .8 & .15 & .05 \\ .38 & .6 & .02 \\ .75 & .05 & .2 \end{pmatrix} \qquad \pi = (.7 \quad .25 \quad .05)$$

Outline For Today

- Time Series
- Markov Models
- Hidden Markov Models
- 3 computational problems of HMMs

Hidden Markov Models: Intuition

- Suppose you can't observe the state
- You can only observe some evidence…

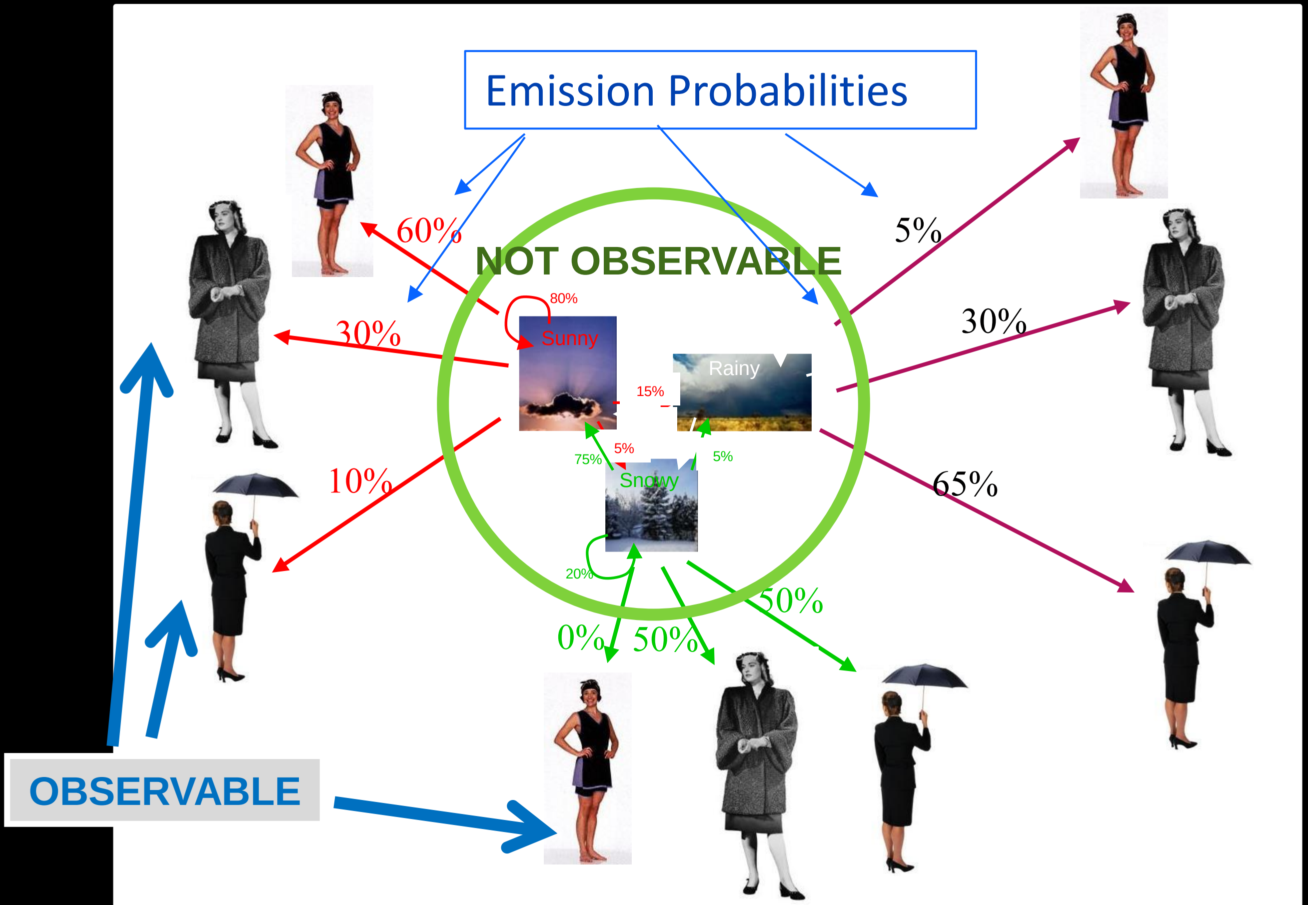
Hidden Markov Models: Weather Example

Observables:

Emission probabilities:

$$b_j(k) = P(o_t = k \mid q_t = S_i)$$

Hidden Markov Models: Weather Example



Probability of a Time Series

- Given: 
- What is the probability of this series?

$$A = \begin{pmatrix} .8 & .15 & .05 \\ .38 & .6 & .02 \\ .75 & .05 & .2 \end{pmatrix} \quad \pi = (.7 \quad .25 \quad .05) \quad B = \begin{pmatrix} .6 & .3 & .1 \\ .05 & .3 & .65 \\ 0 & .5 & .5 \end{pmatrix}$$

Probability of a Time Series

- Given:

$$P(O) = P(O_{coat}, O_{coat}, O_{umbrella}, \ldots, O_{umbrella})$$

$$= \sum_{\text{all } Q} P(O \mid Q) P(Q) = \sum_{q_1,\ldots,q_7} P(O \mid q_1,\ldots,q_7) P(q_1,\ldots,q_7)$$

$$= (0.3^2 \cdot 0.1^4 \cdot 0.6) \cdot (0.7 \cdot 0.8^6) + \ldots$$

(All sun!)

$$A = \begin{pmatrix} .8 & .15 & .05 \\ .38 & .6 & .02 \\ .75 & .05 & .2 \end{pmatrix} \quad \pi = (.7 \quad .25 \quad .05) \quad B = \begin{pmatrix} .6 & .3 & .1 \\ .05 & .3 & .65 \\ 0 & .5 & .5 \end{pmatrix}$$

Specification of an HMM

N - number of states

- S $= \{S_1, , S_2, \dots S_N\}$ – ***set*** *of states*
- $Q = \{q_1;\ q_2; \dots ; q_T\}$ – ***sequence*** *of states*

Specification of an HMM: $\lambda = (A,B,\pi)$

A - *the state transition probability matrix*

$$a_{ij} = P(q_{t+1} = j | q_t = i)$$

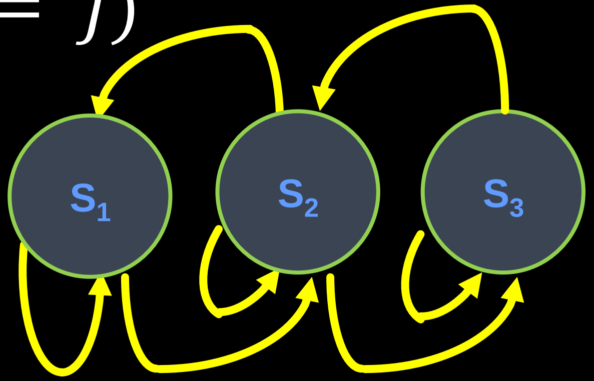
B- *observation probability distribution*

Discrete: $b_{j(k)} = P(o_t = k \,| q_t = j) \; 1 \le k \le M$

Continuous: $b_j(x) = p(o_t = x \,|\, q_t = j)$

π - *the initial state distribution*

$$\pi\,(j) = P(q_1 = j)$$



Specification of an HMM

Some form of output symbols

- *Discrete* – finite vocabulary of symbols of size M. One symbol is "emitted" each time a state is visited
- *Continuous* – an output density in some feature space associated with each state where a output is emitted with each visit

Specification of an HMM

Considering a given observation sequence O

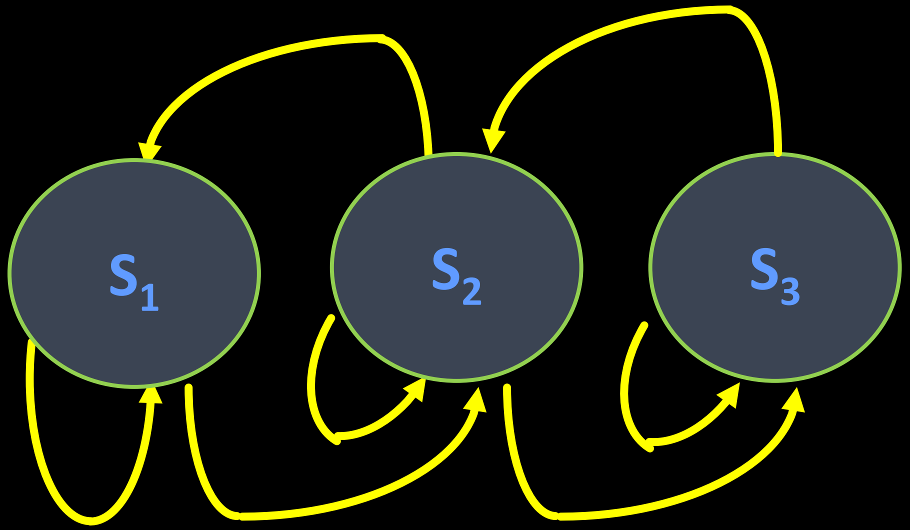
- $O = \{o_1;\ o_2; \ldots ; o_T\}$ – o_i *observed symbol or feature at time i*

(sometimes a set of them)

Specification of an HMM: $\lambda = (A, B, \pi)$

A - the state transition probability matrix

$$a_{ij} = P(q_{t+1} = j | q_t = i)$$

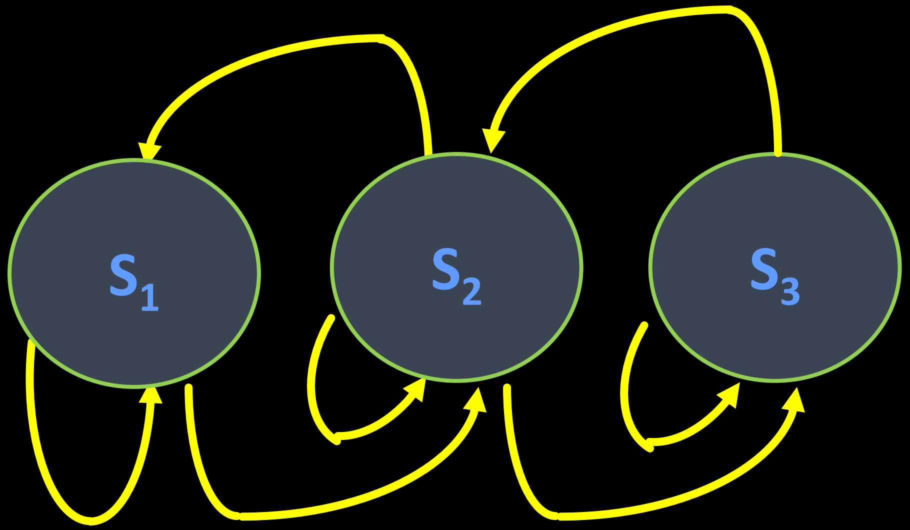


Specification of an HMM: $\lambda = (A, B, \pi)$

B- *observation probability distribution*

Discrete: $b_{j(k)} = P(o_t = k \mid q_t = j)\ 1 \leq k \leq M$

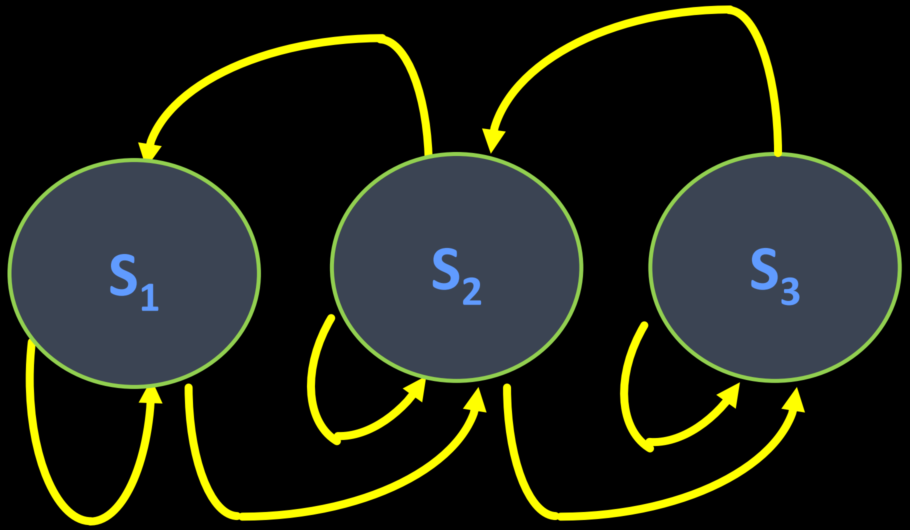
Continuous: $b_j(x) = p(o_t = x \mid q_t = j)$



Specification of an HMM: $\lambda = (A, B, \pi)$

π - the initial state distribution

$$\pi(j) = P(q_1 = j)$$



What does this have to do with Vision?

- Given some sequence of observations, what “model” generated those?
- Using the previous example: given some observation sequence of clothing:

- Is this Philadelphia, Boston or Newark?
- Notice that if Boston vs. Arizona would not need the sequence!

Outline For Today

- Time Series
- Markov Models
- Hidden Markov Models
- 3 computational problems of HMMs

The 3 great problems in HMM modelling

1. Evaluating $P(O|\lambda)$: Given the model $\lambda = (A, B, \pi)$ what is the probability of occurrence of a particular observation sequence

$$O = \{o_1, \dots, o_T\}$$

- *Classification/recognition problem: I have a trained model for each of a set of classes, which one would most likely generate what I saw.*

The 3 great problems in HMM modelling

2. Decoding: Optimal state sequence to produce an observation sequence

$$O = \{o_1, \dots, o_T\}$$

- *Useful in recognition problems – helps give meaning to states.*

The 3 great problems in HMM modelling

3. Learning: Determine model λ, given a training set of observations

- *Find λ, such that $P(O|\lambda)$ is maximal*

Problem 1 $P(O|\lambda)$: Naïve solution

Assume

- We know state sequence $Q = (q_1, \ldots q_\mathrm{T})$
- Independent observations:

then

$$P(O \mid q, \lambda) = \prod_{i=1}^{T} P(o_t \mid q_t, \lambda) = b_{q1}(o_1) b_{q2}(o_2) \ldots b_{qT}(o_T)$$

Problem 1 $P(O|\lambda)$: Naïve solution

- But we know the probability of any given sequence of states:

$$P(q \mid \lambda) = \pi_{q1} a_{q1q2} a_{q2q3} \cdots a_{q(T-1)qT}$$

Problem 1 $P(O|\lambda)$: Naïve solution

- Given $$P(O\,|\,q,\lambda) = \prod_{i=1}^{T} P(o_t\,|\,q_t,\lambda) = b_{q1}(o_1)b_{q2}(o_2)\ldots b_{qT}(o_T)$$
$$P(q\,|\,\lambda) = \pi_{q1}a_{q1q2}a_{q2q3}\ldots a_{q(T-1)qT}$$
- We get: $$P(O\,|\,\lambda) = \sum_{q} P(O\,|\,q,\lambda)P(q\,|\,\lambda)$$

*But this is summed over **all** paths. There are N^T states paths, each 'costing' $O(T)$ calculations, leading to $O(TN^T)$ time complexity.*

Problem 1 $P(O|\lambda)$: Efficient solution

Define auxiliary *forward* variable α:

$$\alpha_t(i) = P(o_1, \ldots, o_t, q_t = i \mid \lambda)$$

$\alpha_t(i)$ is the probability of observing a partial sequence of observables $o_1, \ldots o_t$ **AND** at time *t*, state $q_t = i$

Problem 1 $P(O|\lambda)$: Efficient solution

Forward Recursive algorithm:

- Initialise: $\alpha_1(i) = \pi_i b_i(o_1)$
- Each time step: $\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(o_{t+1})$
- Conclude: $P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$
- Complexity: $O(N^2 T)$

Can reach *j* from any preceding state

Probability of the entire observation sequence is just sum of observations and ending up in state i, for all i.

Rest of HMMs (in brief)

- The *forward* recursive algorithm could compute the likelihood of being in a state *i* at time *t* and having observed the sequence from the start until t, *given HMM* λ

Rest of HMMs (in brief)

- A *backward* recursive algorithm could compute the likelihood of being in a state *i* at time *t* and observing the remainder of the observed sequence, *given HMM* λ

So… or hmmmm…

1. If we know HMM λ then with the forward and backward algorithm we can get an ***E***stimate of the distribution over which state the system is in at time t.
2. With those distributions and having actually observed output data, I can determine the emission probabilities $b_j(k)$ that would ***M***aximize the probability of the sequence.

So... or hmmmm...

3. Given distribution about state can also determine the transition probabilities a_{ij} to ***M***aximize probability.
4. With the new a_{ij} and $b_j(k)$ I can get a new estimate of the state distributions at all time. (Go to 1)

HMMs: General

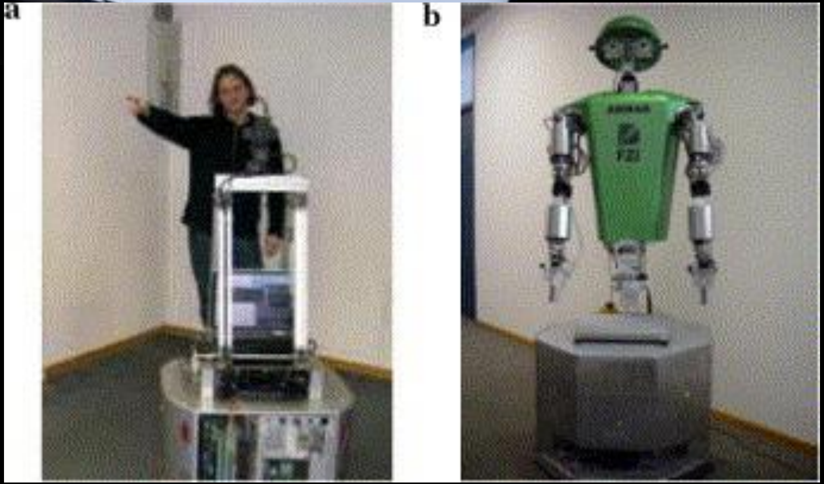
- HMMs: Generative probabilistic models of time series (with hidden state)
- Forward-Backward: Algorithm for computing probabilities over hidden states
  - Given the forward-backward algorithms you can also train the models.
- Best known methods in speech, computer vision, robotics, though for really big data CRFs winning.

Some thoughts about gestures

- *There is a conference on Face and Gesture Recognition so obviously Gesture recognition is an important problem…*

- Prototype scenario:
  - Subject does several examples of "each gesture"
  - System "learns" (or is trained) to have some sort of model for each
  - At run time compare input to known models and pick one

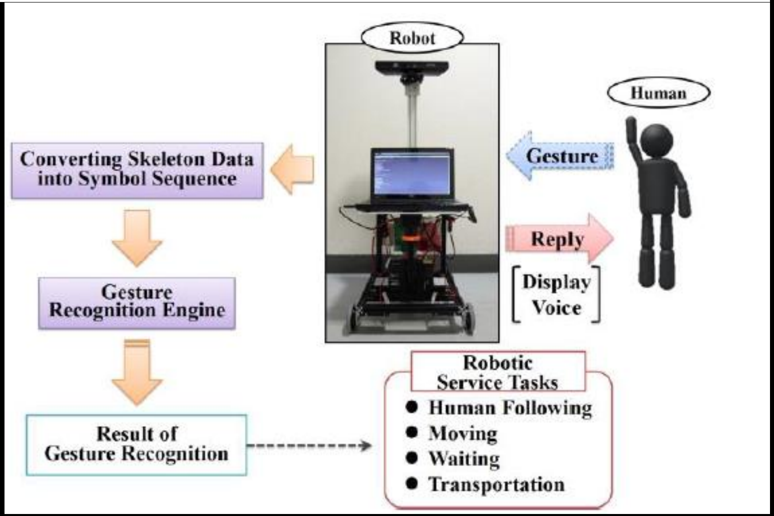
New found life for Gesture Recognition:



Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol I,
IMECS 2014, March 12 - 14, 2014, Hong Kong

Gesture Recognition System for Human-Robot Interaction and Its Application to Robotic Service Task

Tatsuya Fujii, Jae Hoon Lee, *Member, IAENG* and Shingo Okamoto



Generic Gesture Recognition using HMMs

Recognition of Space-Time Hand-Gestures using Hidden Markov Model

Yanghee Nam
Dept. of Computer Science
Korea Advanced Institute of Science and Technologies
373-1, Kusong-dong, Yusong-ku
Taejeon, Korea
+82 42 869 3572
nyh@vr.kaist.ac.kr

KwangYun Wohn
Dept. of Computer Science
Korea Advanced Institute of Science and Technologies
373-1, Kusong-dong, Yusong-ku
Taejeon, Korea
+82 42 869 3572
wohn@vr.kaist.ac.kr

ABSTRACT

The rapidly growing interest in interactive three-dimensional(3D) computer environments highly recommend the hand gesture as one of their interaction modalities. Among several factors constituting a hand gesture, hand movement pattern is spatiotemporally variable and informative, but its automatic recognition is not trivial.

In this paper, we describe a hidden Markov(HMM)-based method for recognizing the space-time hand movement pattern. HMM models the spatial variance and the time-scale variance in the hand movement. As for the recognition of

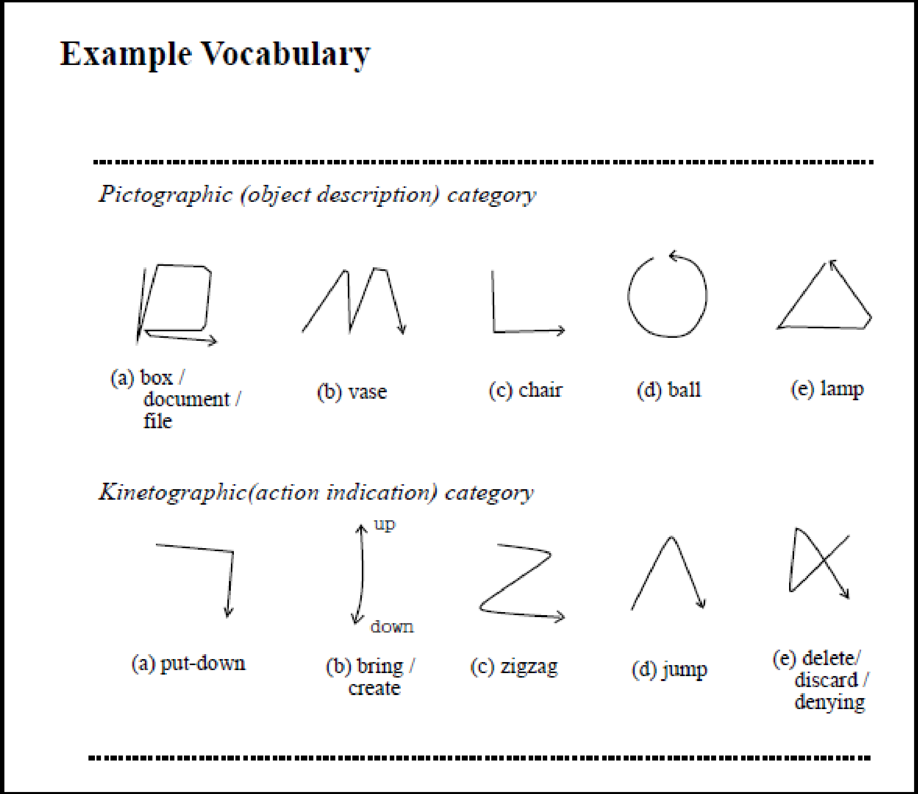
INTRODUCTION

A hand gesture is *a movement that we make with our hands to express emotion or information, either instead of speaking or while we are speaking* [1].

The use of natural hand gestures for computer-human interaction can help people to communicate with computer in more intuitive way. Moreover, recent studies on three-dimensional(3D) virtual environment and the developments of various 3D input devices encourage to add this kind of 3D interaction modality to the user interface design.

Nam, Y., & Wohn, K. (1996, July). Recognition of space-time hand-gestures using hidden Markov model. In *ACM symposium on Virtual reality software and technology* (pp. 51-58).

Generic Gesture Recognition using HMMs (1)



Data glove

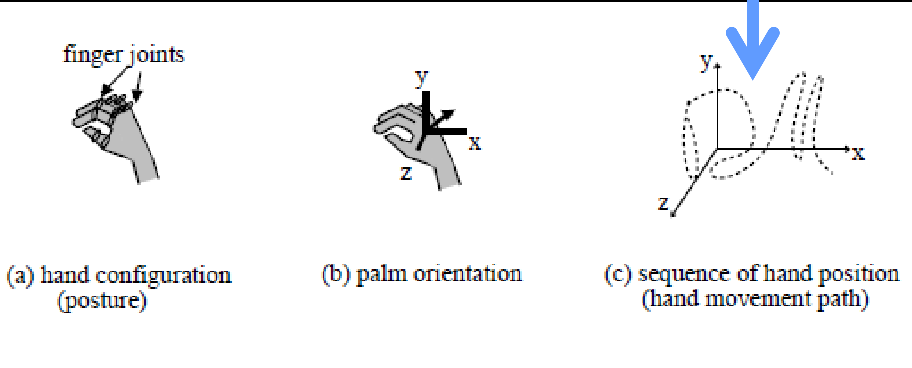
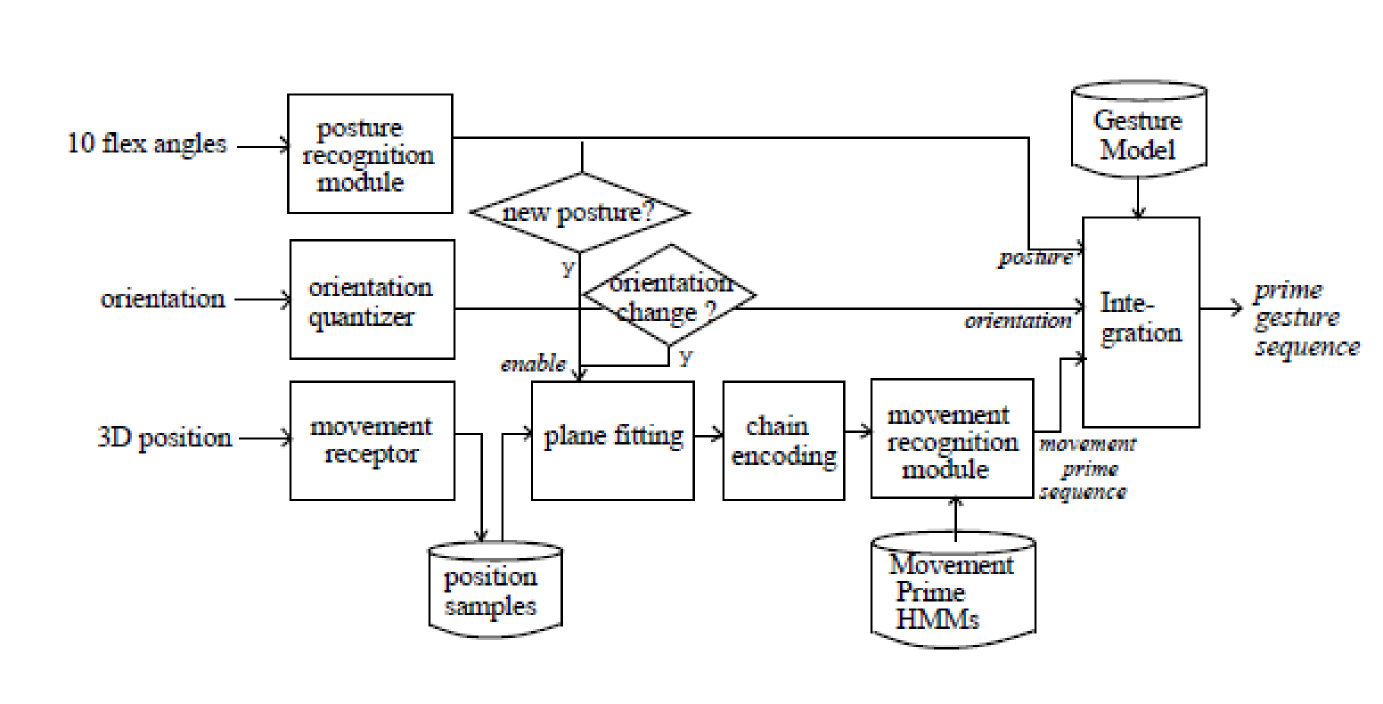


Figure 1: 3D hand gesture attributes

Generic Gesture Recognition using HMMs (2)



Generic Gesture Recognition using HMMs (3)

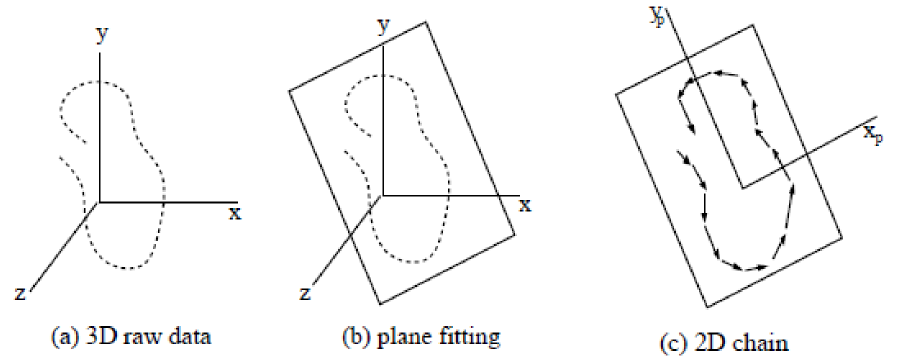


Figure 4: 3D to 2D reduction by plane fitting

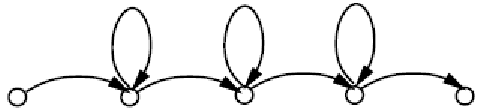


Figure 5: Simple left-to-right HMM

Generic Gesture Recognition using HMMs (4)

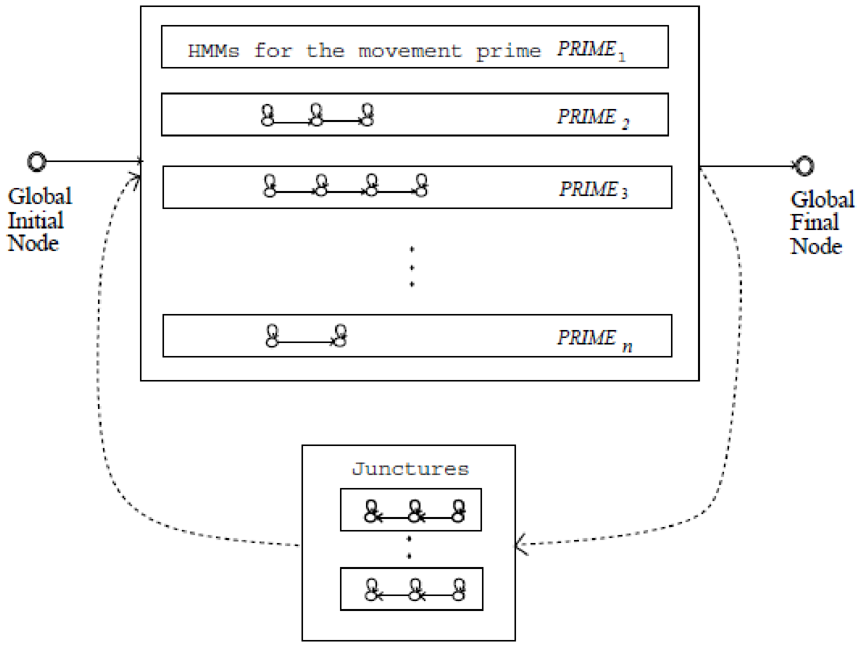


Figure 6: The HMM network for the recognition of connected hand movement pattern

Generic Gesture Recognition using HMMs (5)

movement prime	*recognition results*		
	# of tests	misses	hits
box	111	1 (0.90%)	110 (99.10%)
vase	111	0 (0.00%)	111 (100.0%)
chair	112	2 (1.79%)	110 (98.21%)
ball	111	4 (3.60%)	107 (96.40%)
lamp	112	2 (1.79%)	110 (98.21%)
put-down	111	1 (0.90%)	110 (99.10%)
bring	110	0 (0.00%)	110 (100.0%)
zigzag	111	1 (0.90%)	110 (99.10%)
jump	112	0 (0.00%)	112 (100.0%)
delete	111	0 (0.00%)	111 (100.0%)

Table 1: Recognition accuracy for the movement primes

Pluses and minuses of HMMs in Gesture

Good points about HMMs:

- A learning paradigm that acquires spatial and temporal models and does some amount of feature selection.
- Recognition is fast; training is not so fast but not too bad.

Pluses and minuses of HMMs in Gesture

Not so good points:

- Not great for on the fly labeling – e.g. segmentation of input streams. Requires lots of data to train for that – much like language.
- Works well when problem is easy. Less clear other times.