Probabilistic Models for Occurrences

Modelling probabilistic dependencies (causalities) and independencies between discrete events of a scene

X_i random variable ***models uncertain propositions about a scene***

$X_i = a$ hypothesis

Decomposition of joint probabilities:

$$P(X_1, X_2, X_3, \ldots, X_n) = P(X_1 \mid X_2, X_3, \ldots, X_n) \bullet P(X_2 \mid X_3, X_4, \ldots, X_n) \bullet \ldots \bullet P(X_{n-1} \mid X_n) \bullet P(X_n)$$

Simplification in the case of statistical independence:

X independent of X_i

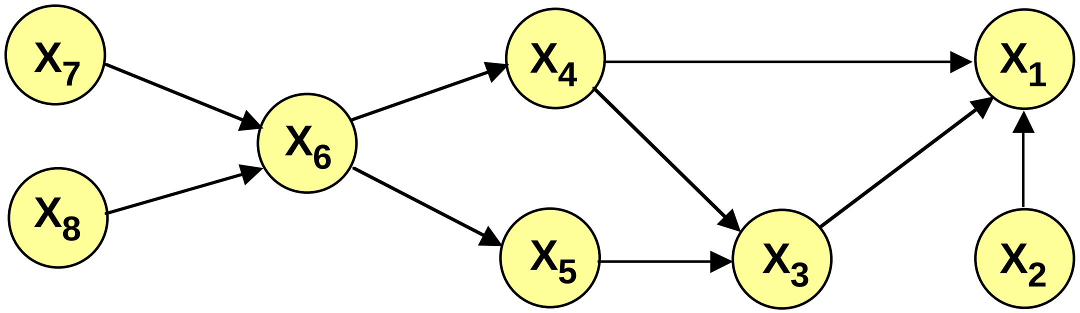
$$P(X \mid X_1, \ldots X_{i-1}, X_i, X_{i+1}, \ldots, X_n) = P(X \mid X_1, \ldots X_{i-1}, X_{i+1}, \ldots, X_n)$$

Joint probability of N variables may be simplified by ordering the variables according to their direct dependence (causality)

Causality Graph

Conditional dependencies (causality relations) of random variables define partial order.

Representation as a directed graph:



$$P(X_1, X_2, X_3, \ldots, X_8) =$$
$$P(X_1 \mid X_2, X_3, X_4) \cdot P(X_2) \cdot P(X_3 \mid X_4, X_5) \cdot P(X_4 \mid X_6) \cdot P(X_5 \mid X_6) \cdot P(X_6 \mid X_7 X_8) \cdot P(X_7) \cdot P(X_8)$$

Constructing a Bayes Net

By domain analysis:

1. **Select discrete variables X_i relevant for domain**
2. **Establish partial order of variables according to causality**
3. **In the order of decreasing causality:**
   - **(i) Generate node X_i in net**
   - **(ii) As predecessors of X_i choose the smallest subset of nodes which are already in the net and from which X_i is causally dependent**
   - **(iii) determine a table of conditional probabilities for X_i**

By data analysis:

Use a learning method to establish a Bayes Net approximating the empirical joint probablity distribution.

Computing Inferences

We want to use a Bayes Net for probabilistic inferences of the following kind:

Given a joint probability $P(X_1, \ldots, X_N)$ represented by a Bayes Net, and evidence $X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}$ for some of the variables, what is the probability $P(X_n= a_i \mid X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K})$ of an unobserved variable to take on a value a_i ?

In general this requires

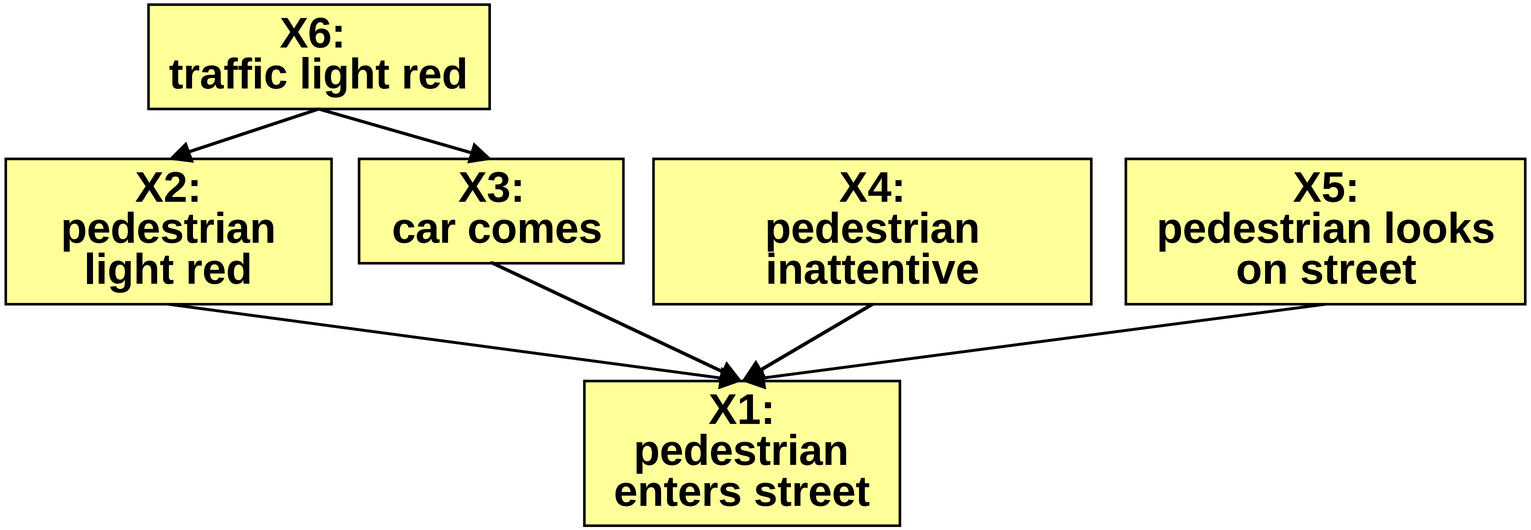
- **expressing a conditional probability by a quotient of joint probabilities**

$$P(X_n= a_i \mid X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}) = \frac{P(X_n= a_i, X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K})}{P(X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K})}$$

- **determining partial joint probabilities from the given total joint probability by summing out unwanted variables**

$$P(X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}) = \sum_{X_{n_1}, \ldots, X_{n_K}} P(X_{m_1}=a_{m_1}, \ldots, X_{m_K}=a_{m_K}, X_{n_1}, \ldots, X_{n_K})$$

Example: Traffic Behaviour of Pedestrians



Conditional probability table for each node must be known

<u>Examples</u>: **P(X1 | X2, X3, X4, X5)** **P(X2 | X6)**

X1	X2	X3	X4	X5	P
T	T	T	T	T	0.3
F	T	T	T	T	0.7
T	F	T	T	T	0.9
F	F	T	T	T	0.1
⋮	⋮	⋮	⋮	⋮	⋮

X2	X6	P
T	T	0.2
F	T	0.8
T	F	1.0
F	F	0.0

Estimating Probabilities from a Database

Given a sufficiently large database with tupels $\underline{a}^{(1)}$ ... $\underline{a}^{(N)}$ of an unknown distribution $P(\underline{X})$, we can compute maximum likelihood estimates of all partial joint probabilities and hence of all conditional probabilities.

$X_{m_1}, \dots , X_{m_K}$ = subset of $X_1, \dots X_L$ with $K \le L$

$w_{\underline{a}}$ = number of tuples in database with $X_{m_1}=a_{m_1}, \dots , X_{m_K}=a_{m_K}$

N = total number of tuples

Maximum likelihood estimate of $P(X_{m_1}=a_{m_1}, \dots , X_{m_K}=a_{m_K})$ is

$$P'(X_{m_1}=a_{m_1}, \dots , X_{m_K}=a_{m_K}) = w_{\underline{a}} / N$$

If a priori information is available, it may be introduced via a bias $m_{\underline{a}}$:

$$P'(X_{m_1}=a_{m_1}, \dots , X_{m_K}=a_{m_K}) = (w_{\underline{a}} + m_{\underline{a}}) / N$$

Often $m_{\underline{a}} = 1$ is chosen for all tupels $\underline{a}$ to express equal likelihoods in the case of an empty database.

Expectation Maximization (1)

Recommended reading: Borgelt & Kruse, Graphical Models, Wiley 2002

Often databases are neither complete (insufficient samples, missing attributes) nor precise (ambiguous or uncertain values). In this case Expectation Maximation (EM) provides an iterative procedure to estimate probabilities.

1. <u>Imprecise data</u>

Given a tuple with ambiguous attributes

$$\underline{a}^T = [\ \{a_{11}, a_{12}, \ldots\}, \{a_{21}, a_{22}, \ldots\}, \ldots, \{a_{K1}, a_{K2}, \ldots\}\ ]$$

and number of occurrence $w_{\underline{a}}$, redistribute $w_{\underline{a}}$ equally among all combinations of attribute values.

2. <u>Incomplete database</u>

Execute iterative 2-step procedure:

A Compute sample frequencies from estimated probabilities

B Estimate probabilities from samples, maximizing likelihood of data (see previous slide)

Expectation Maximization (2)

Expectation step of EM:

Use current (initial) probability estimates to compute probability $P(\underline{a})$ for all attribute combinations $\underline{a}$.

For Bayes Nets, this requires computing $P(\underline{a})$ from the conditional probabilities assigned to the nodes.

At the initial step, absolute frequencies of missing attribute tuples $\underline{a}^*$ are completed:

$$\underline{a}_1 = [X_1=a_1 , X_2=a_{m2}, X_3=a_{m3}, \ldots] \quad w_{\underline{a}^*} \bullet P(\underline{a}_1)$$

$$\underline{a}_2 = [X_1=a_2 , X_2=a_{m2}, X_3=a_{m3}, \ldots] \quad w_{\underline{a}^*} \bullet P(\underline{a}_2)$$

$$\underline{a}^* = [\ * , X_2=a_{m2}, X_3=a_{m3}, \ldots] \quad w_{\underline{a}^*}$$

$$\vdots$$

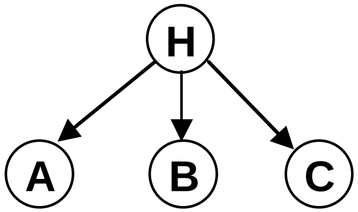
$$\underline{a}_M = [X_1=a_M , X_2=a_{m2}, X_3=a_{m3}, \ldots] \quad w_{\underline{a}^*} \bullet P(\underline{a}_M)$$

missing attribute — **absolute frequency** — **completed database**

Example for Expectation Maximization (1)

(adapted from Borgelt & Kruse, Graphical Models, Wiley 2002)

Given 4 binary probabilistic variables A, B, C, H with known dependency structure:



Given also a database with tuples [* A B C] where H is a missing attribute.

H	A	B	C	w
*	T	T	T	14
*	T	T	F	11
*	T	F	T	20
*	T	F	F	20
*	F	T	T	5
*	F	T	F	5
*	F	F	T	11
*	F	F	F	14

absolute frequencies of occurrence

Estimate of the conditional probabilities of the Bayes Net nodes !

Example for Expectation Maximization (2)

Initial (random) probability assignments:

H	P(H)
T	0.3
F	0.7

A	H	P(A|H)
T	T	0.4
T	F	0.6
F	T	0.6
F	F	0.4

B	H	P(B|H)
T	T	0.7
T	F	0.8
F	T	0.3
F	F	0.2

C	H	P(C|H)
T	T	0.8
T	F	0.5
F	T	0.2
F	F	0.5

With $$P(H \mid A,B,C) = \frac{P(A \mid H) \cdot P(B \mid H) \cdot P(C \mid H) \cdot P(H)}{\sum_{H} P(A \mid H) \cdot P(B \mid H) \cdot P(C \mid H) \cdot P(H)}$$

one can complete the database:

H	A	B	C	w
T	T	T	T	1.27
T	T	T	F	3.14
T	T	F	T	2.93
T	T	F	F	8.14
T	F	T	T	0.92
T	F	T	F	2.37
T	F	F	T	3.06
T	F	F	F	8.49

H	A	B	C	w
F	T	T	T	12.73
F	T	T	F	7.86
F	T	F	T	17.07
F	T	F	F	11.86
F	F	T	T	4.08
F	F	T	F	2.63
F	F	F	T	7.94
F	F	F	F	5.51

Example for Expectation Maximization (3)

Based on the modified complete database, one computes the maximum likelihood estimates of the conditional probabilities of the Bayes Net.

Example: $P(A = T \mid H = T) \approx \dfrac{1.27 \bullet 3.14 \bullet 2.93 \bullet 8.14}{1.27 \bullet 3.14 \bullet 2.93 \bullet 8.14 \bullet 0{,}92 \bullet 2.73 \bullet 3.06 \bullet 8.49} \approx 0.51$

This way one gets new probability assignments:

H	P(H)
T	0.3
F	0.7

A	H	P(A|H)
T	T	0.51
T	F	0.71
F	T	0.49
F	F	0.29

B	H	P(B|H)
T	T	0.25
T	F	0.39
F	T	0.75
F	F	0.61

C	H	P(C|H)
T	T	0.27
T	F	0.60
F	T	0.73
F	F	0.40

This completes the first iteration. After ca. 700 iterations the modifications of the probabilities are less than 10^{-4}. The resulting values are

H	P(H)
T	0.5
F	0.5

A	H	P(A|H)
T	T	0.5
T	F	0.8
F	T	0.5
F	F	0.2

B	H	P(B|H)
T	T	0.2
T	F	0.5
F	T	0.8
F	F	0.2

C	H	P(C|H)
T	T	0.4
T	F	0.6
F	T	0.6
F	F	0.4

Hidden Markov Models

A sequence of observations may be governed by underlying probabilistic state transitions.

Example: A person laying a table may plan to first place the plates, then the cups, then the cutlery in a cyclic order (with a chance to deviate from this order).

As usual in vision, observations may be disturbed and may provide uncertain evidence about the current state.

Such phenomena may be modelled by a Hidden Markov Model (HMM).

A (discrete) HMM is defined by

- **a finite number of states $a_1, a_2, \ldots, a_K$**
- **a sequence of state transition events $t_0, t_1, \ldots, t_n$ (not necessarily times)**
- **probabilities of state transitions p_{ij} from state i to state j depending only on the past states**
- **observations $b_1, b_2, \ldots, b_M$ probabilistically related to each state**
- **probabilities q_{km} which map states into observations**

Notation for HMM

- sequence of random variables $X^{(1)}, \dots, X^{(n)}$ (state variables) with values from $\{a_1, \dots, a_K\}$
- Markov Chain property of $X^{(1)}, \dots, X^{(n)}$: $P(X^{(n)}|X^{(n-1)} \dots X^{(1)}) = P(X^{(n)}|X^{(n-1)})$
  - if $P(X^{(n)}|X^{(n-1)})$ is independent of n, the Markov Chain is homogeneous
  - transition probabilities $P(X^{(n}=a_i|X^{(n-1)}=a_j)$ are represented by the state transition matrix

$$W^{(n)} = \begin{bmatrix} p_{11} & \dots & p_{1K} \\ \vdots & & \\ P_{K1} & \dots & p_{KK} \end{bmatrix}$$

- random variables $Y^{(1)}, \dots, Y^{(n)}$ (observations) with values from $\{b_1, \dots, b_M\}$
- observation probabilities $P(Y^{(n)}|X^{(n)})$ are represented by the matrix

$$Q = \begin{bmatrix} q_{11} & \dots & q_{1M} \\ \vdots & & \\ q_{K1} & \dots & q_{KM} \end{bmatrix}$$

- initial probabilities $\underline{\pi}^T = [\ P(X^{(1)}=a_1)\ \ P(X^{(1)}=a_2) \dots P(X^{(1)}=a_K)\ ]$

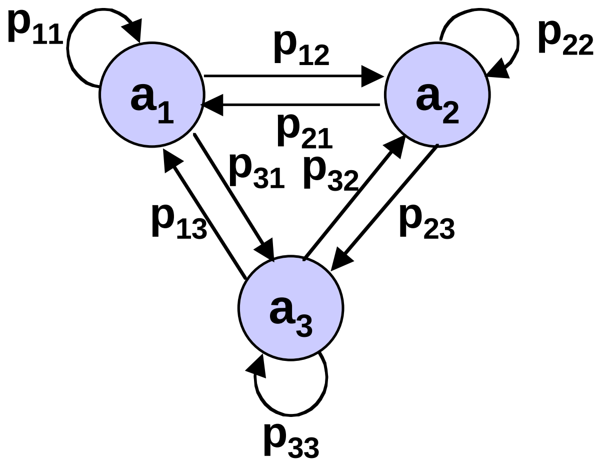
Properties of a Homogeneous HMM

Probability vector for state $X^{(2)}$: $\underline{\pi}^{(2)} = W^T \underline{\pi}$

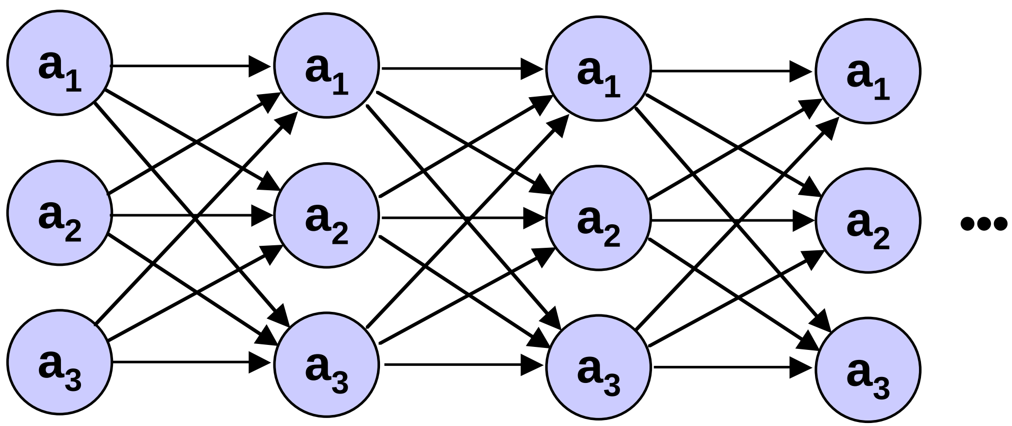
Probability vector for state $X^{(n)}$: $\underline{\pi}^{(n)} = (W^T)^{n-1} \underline{\pi}$

There is always a <u>stationary distribution</u> $\underline{\pi}_s$ such that $\underline{\pi}_s = W^T \underline{\pi}_s$

Graphical representation:



<u>Trellis</u> ("Spalier") representation:



- each (directed) path corresponds to a legal sequence of states
- the probability of a path is equal to the product of the transition probabilities

Paths through a HMM

Given a sequence of N observations, we want to find the most probable sequence of states which may have led to the observations.

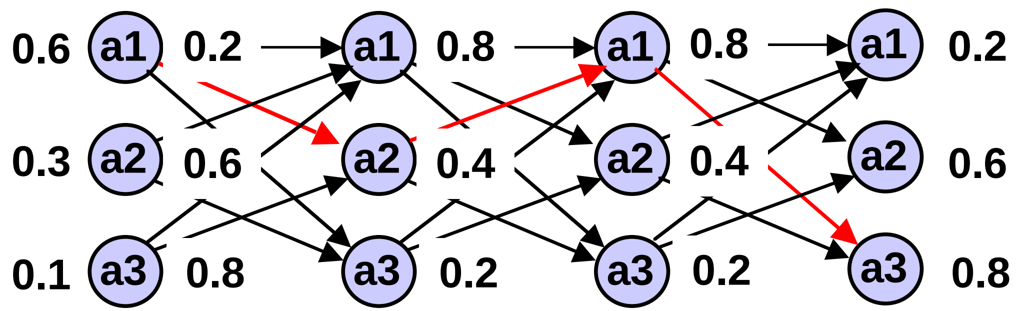
Extension of trellis representation

- **arc weights leading into states $X^{(n)}$:** transition probabilities p_{ij}
- **node weights of states $X^{(n)}$:** observation likelihoods q_{jm} for given observations $Y^{(n)} = b_{m_n}$
- **product of initial probability and node and arc probabilities along path:** $P(Y^{(1)}=b_{m_1}, \ldots , Y^{(N)}=b_{m_N}, X^{(1)}=a_{k_1}, \ldots , X^{(N)}=a_{k_N})$ probability of observations and states

Example:

$$W = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.0 & 0.9 \\ 0.4 & 0.6 & 0.0 \end{bmatrix} \quad Q = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} \quad \pi = \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix}$$

observations b_2, b_1, b_1, b_2



probability of observations along path are

$P(Y^{(1)}=b_2, Y^{(2)}=b_1, Y^{(3)}=b_1, Y^{(4)}=b_2,$ states of path$) =$

$0.6 \cdot 0.2 \cdot 0.2 \cdot 0.4 \cdot 0.1 \cdot 0.8 \cdot 0.5 \cdot 0.8$

Finding Most Probable Paths

The most probable sequence of states is found by maximizing

$$\max_{k_1 \dots k_N} P(X^{(1)}=a_{k_1}, \dots, X^{(N)}=a_{k_N} \mid Y^{(1)}=b_{m_1}, \dots, Y^{(N)}=b_{m_N}) = \max_{\underline{a}} P(\underline{a} \mid \underline{b})$$

Equivalently, the most probable sequence of states follows from

$$\max_{\underline{a}} P(\underline{a}\ \underline{b}) = \max_{\underline{a}} P(\underline{a} \mid \underline{b})\, P(\underline{b})$$

Hence the maximizing sequence of states can be found by exhaustive search of all path probabilities in the trellis. However, complexity is $O(K^N)$ with K = number of different states and N = length of sequence.

The Viterbi Algorithm does the job in O(KN)!

Overall maximization may be decomposed into a backward sequence of maximizations:

$$\max_{\underline{a}} P(\underline{a}\ \underline{b}) = \max_{k_1 \dots k_N} \pi_{k_1} q_{k_1 m_1} \prod_{n=2 \dots N} p_{k_{n-1} k_n} q_{k_{n-1} m_n}$$

$$= \max_{k_1} \pi_{k_1} q_{k_1 m_1} \Big(\max_{k_2} p_{k_1 k_2} q_{i_2 m_2} \big(\dots (\max_{k_N} p_{k_{N-1} k_N} q_{k_{N-1} m_N}) \dots\big)\Big)$$

Step N　　Step N-1　　Step 1

Example for Viterbi Algorithm

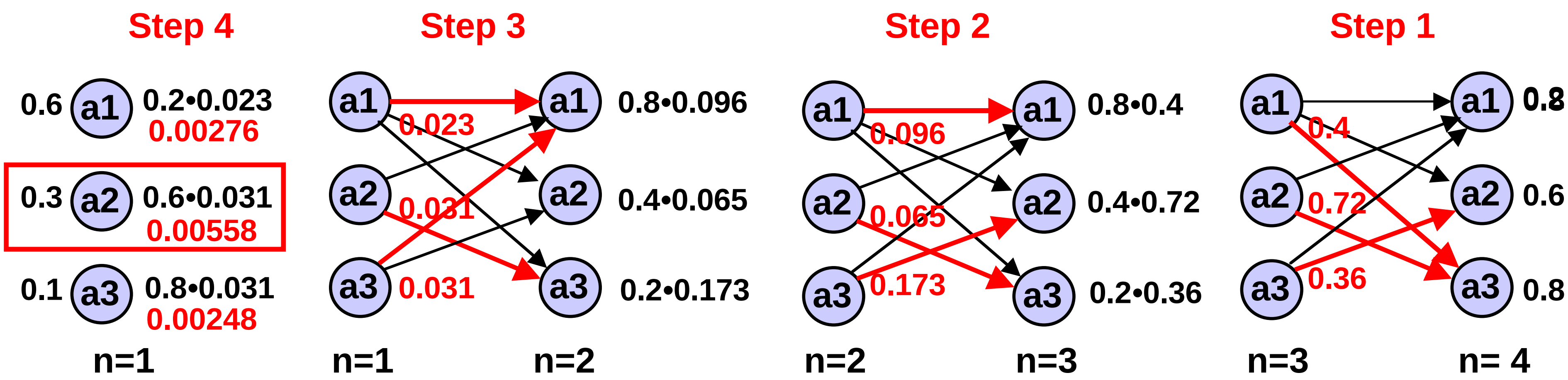
Typical maximization step of Viterbi algorithm:

$$\max_{k_n} \{ p_{k_{n-1}k_n} \cdot q_{k_{n-1}m_n} \cdot \langle \text{result of previous maximization step} \rangle \}$$

Example as earlier:

$$W = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.0 & 0.9 \\ 0.4 & 0.6 & 0.0 \end{bmatrix} \quad Q = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} \quad \pi = \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix}$$

observations
b_2, b_1, b_1, b_2



red numbers show maximization results, red arrows maximizing transitions

Model Evaluation for Given Observations

What is the likelihood that a particular HMM (out of several possible models) has generated the observations?

Likelihood of observations given model:

$$P(Y^{(1)}=b_{m_1}, \dots, Y^{(N)}=b_{m_N} \mid \text{model}) = P(\underline{b}) = \sum_{\underline{a}} P(\underline{a}\ \underline{b})$$

Instead of summing over all $\underline{a}$, one can use a forward algorithm based on the recursive formula:

$$\underline{P(a_j^{(n+1)}, b_{m_1}, \dots, b_{m_n}, b_{m_{n+1}})}$$

$$= P(a_j^{(n+1)}, b_{m_1}, \dots, b_{m_n}) \bullet P(b_{m_{n+1}} \mid a_j^{(n+1)})$$

$$= \sum_i [\, P(a_j^{(n+1)}, P(a_i^{(n)}, b_{m_1}, \dots, b_{m_n})] \bullet P(b_{m_{n+1}} \mid a_j^{(n+1)})$$

$$= \sum_i [\, P(a_j^{(n+1)} \mid P(a_i^{(n)}, b_{m_1}, \dots, b_{m_n}) P(a_i^{(n)}, b_{m_1}, \dots, b_{m_n})] \bullet P(b_{m_{n+1}} \mid a_j^{(n+1)})$$

$$= \sum_i [\, P(a_j^{(n+1)} \mid P(a_i^{(n)}) \bullet P(a_i^{(n)}, b_{m_1}, \dots, b_{m_n})] \bullet P(b_{m_{n+1}} \mid a_j^{(n+1)})$$

$$= \sum_i [\, p_{ij} \bullet \underline{P(a_i^{(n)}, b_{m_1}, \dots, b_{m_n})}] \bullet q_{j\, m_{n+1}}$$

Finally: $P(b_{m_1}, \dots, b_{m_N}) = \sum_i P(a_i^{(n+1)}, b_{m_1}, \dots, b_{m_N})$

Example for Model Evaluation (1)

Computing the probability of observations stepwise as they come in.

Example as earlier:

$$W = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.0 & 0.9 \\ 0.4 & 0.6 & 0.0 \end{bmatrix} \quad Q = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} \quad \pi = \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix}$$

observations
b_2, b_1, b_1, b_2

Step 1

$$P(a_j^{(1)}, b_{m_1}) = \pi_j \bullet q_{j\, m_1}$$

$$P(a_1^{(1)}, b_2) = 0.6 \bullet 0.2 = 0.12$$
$$P(a_2^{(1)}, b_2) = 0.3 \bullet 0.6 = 0.18$$
$$P(a_3^{(1)}, b_2) = 0.1 \bullet 0.8 = 0.08$$

Note that $P(b_{m_1}, \ldots, b_{m_n})$ can be computed after each step by summing out the dependency on the state $X^{(n)}$.

Step 2

$$P(a_j^{(2)}, b_{m_1}, b_{m_2}) = \Sigma\ [\, p_{ij} \bullet P(a_i^{(1)}, b_{m_1})] \bullet q_{j\, m_2}$$

$$P(a_1^{(2)}, b_2, b_1) = [\ 0.3 \bullet 0.12 + 0.1 \bullet 0.18 + 0.4 \bullet 0.08\ ] \bullet 0.8 = 0.0314$$
$$P(a_2^{(2)}, b_2, b_1) = [\ 0.2 \bullet 0.12 + 0.6 \bullet 0.08\ ] \bullet 0.4 = 0.0288$$
$$P(a_3^{(2)}, b_2, b_1) = [\ 0.5 \bullet 0.12 + 0.9 \bullet 0.18\ ] \bullet 0.2 = 0.0072$$

Example for Model Evaluation (2)

Example continued:

$$W = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.1 & 0.0 & 0.9 \\ 0.4 & 0.6 & 0.0 \end{bmatrix} \quad Q = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} \quad \pi = \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix}$$

observations
b_2, b_1, b_1, b_2

Step 3

$$P(a_j^{(3)}, b_{m_1}, b_{m_2}, b_{m_3}) = \Sigma \; [\, p_{ij} \bullet P(a_j^{(2)}, b_{m_1}, b_{m_2})] \bullet q_{j\, m_3}$$

$$P(a_1^{(3)}, b_2, b_1, b_1) = [\, 0.3 \bullet 0.0314 + 0.1 \bullet 0.0288 + 0.4 \bullet 0.0072 \,] \bullet 0.8 = 0.01214$$
$$P(a_2^{(3)}, b_2, b_1, b_1) = [\, 0.2 \bullet 0.0314 + 0.6 \bullet 0.0072 \,] \bullet 0.4 = 0.00424$$
$$P(a_3^{(3)}, b_2, b_1, b_1) = [\, 0.5 \bullet 0.0314 + 0.9 \bullet 0.0288 \,] \bullet 0.2 = 0.00832$$

Step 4

$$P(a_j^{(4)}, b_{m_1}, b_{m_2}, b_{m_3}, b_{m_4}) = \Sigma \; [\, p_{ij} \bullet P(a_j^{(2)}, b_{m_1}, b_{m_2}, b_{m_3})] \bullet q_{j\, m_4}$$

$$P(a_1^{(4)}, b_2, b_1, b_1, b_2) = [\, 0.3 \bullet 0.01214 + 0.1 \bullet 0.00424 + 0.4 \bullet 0.00832 \,] \bullet 0.2 = 0.001479$$
$$P(a_2^{(4)}, b_2, b_1, b_1, b_2) = [\, 0.2 \bullet 0.01214 + 0.6 \bullet 0.00832 \,] \bullet 0.6 = 0.004452$$
$$P(a_3^{(4)}, b_2, b_1, b_1, b_2) = [\, 0.5 \bullet 0.01214 + 0.9 \bullet 0.00424 \,] \bullet 0.4 = 0.003954$$

Final step

$$P(b_{m_1}, b_{m_2}, b_{m_3}, b_{m_4}) = \Sigma \; P(a_j^{(4)}, b_{m_1}, b_{m_2}, b_{m_3}, b_{m_4}) = \boxed{0.009885}$$

Learning Models for High-level Image Interpretation

What parts of a scene constitute "meaningful occurrences" and should be recognized?

Basic engineering applications:

Fixed recognition tasks, determined by the application context.

=> handcrafted models

Advanced engineering applications:

Flexible recognition tasks, determined by user.

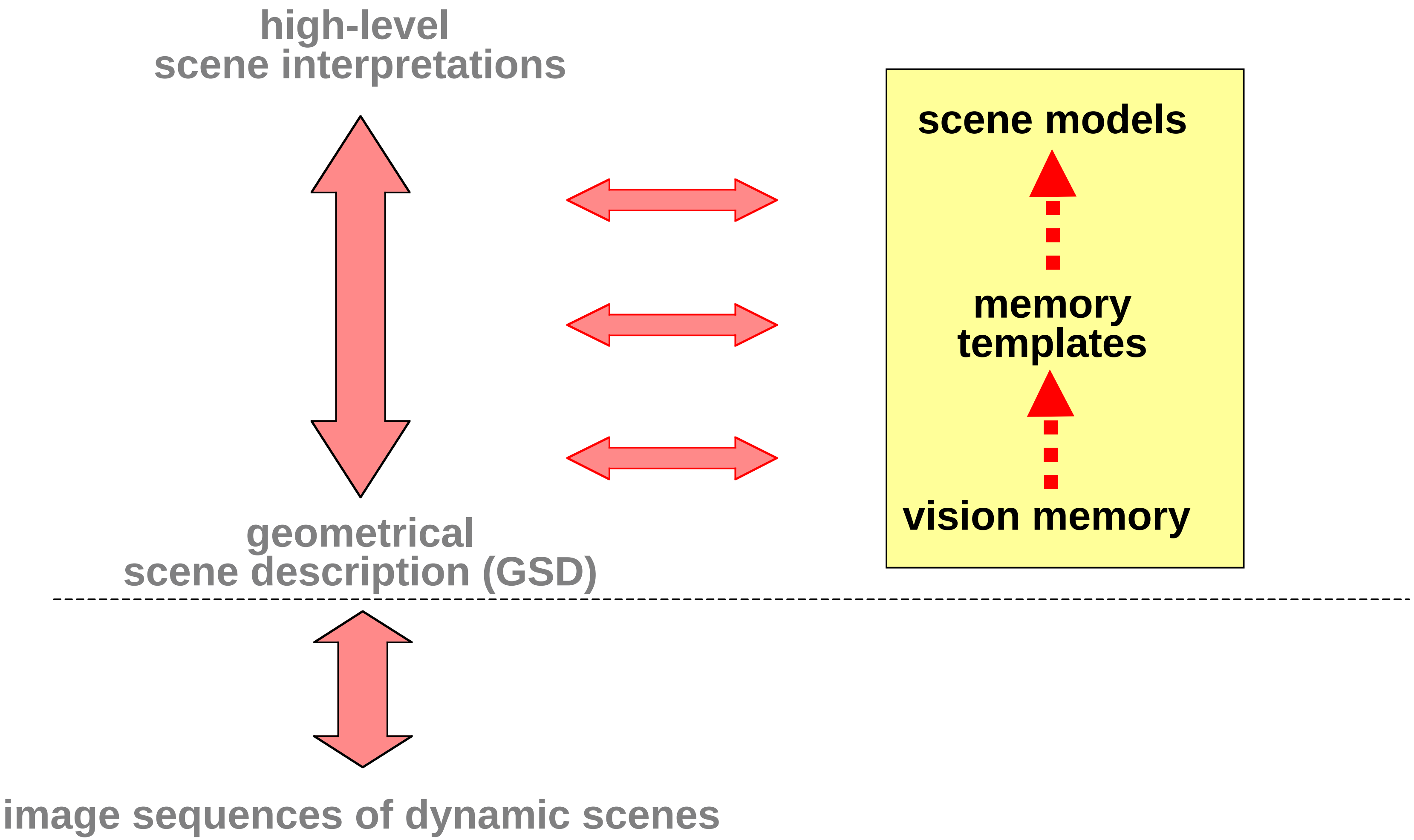
=> models result from supervised learning

Biological vision:

Recognition should support expectation generation and hence survival.

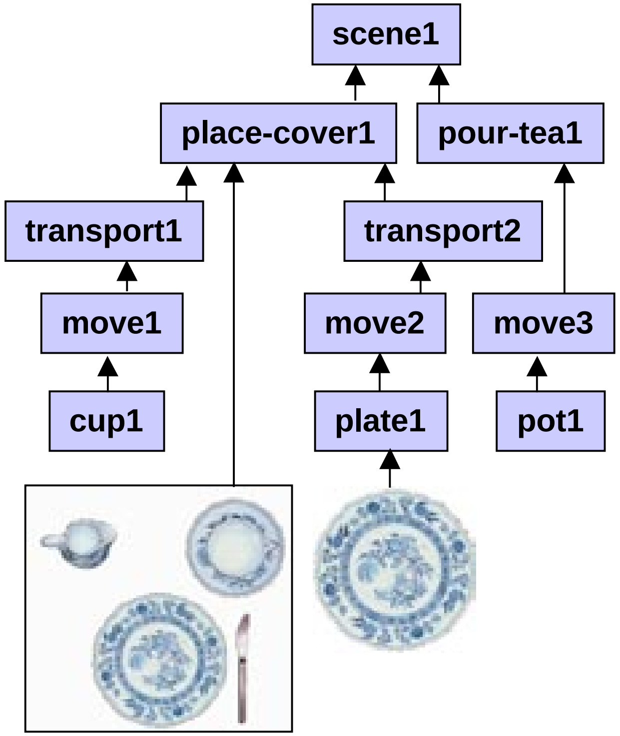
=> models result from unsupervised learning

Learning in Support of High-level Scene Interpretation

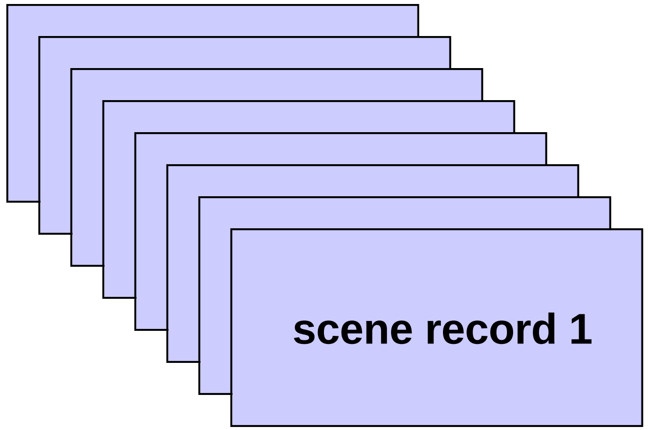


Basic Structure of Vision Memory

scene record



vision memory

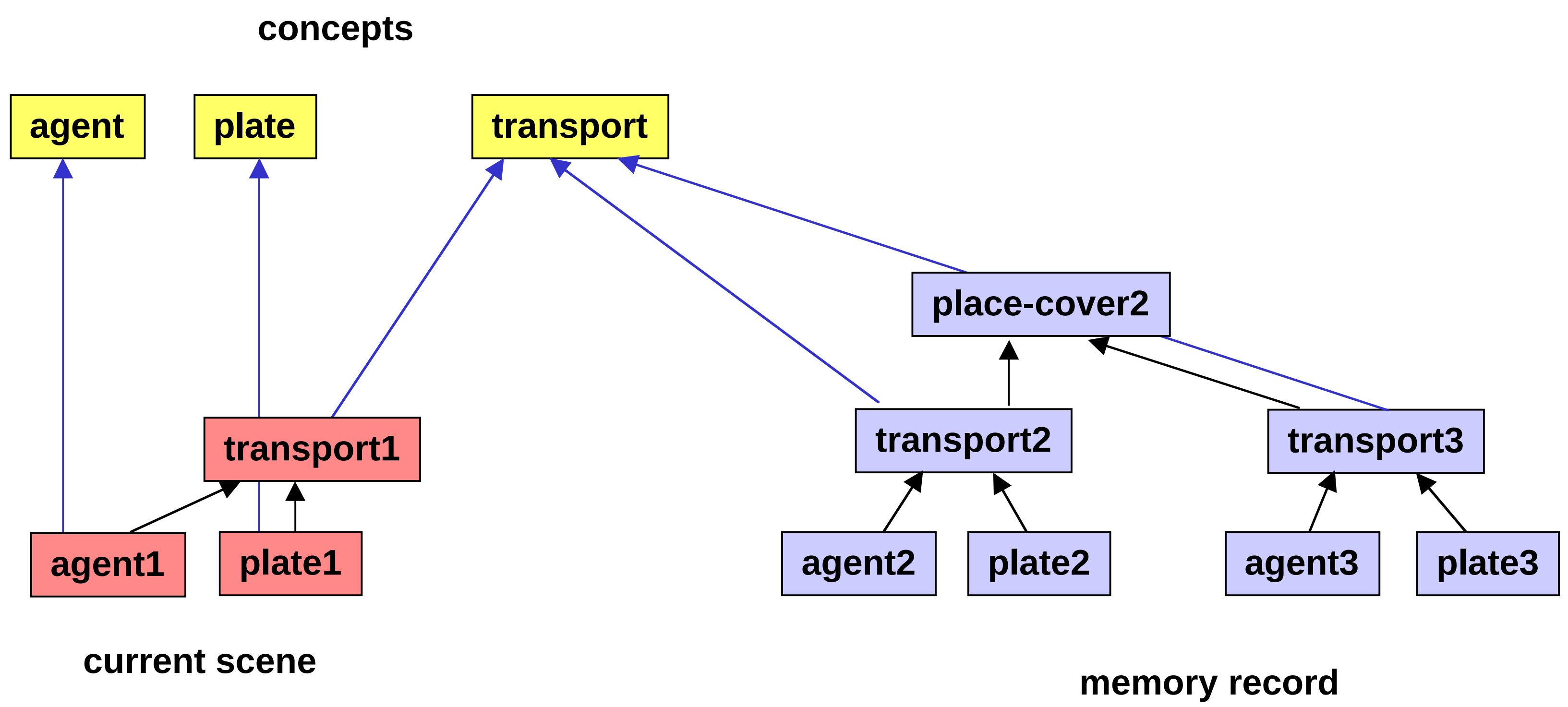


It is an open research question, how much imagery should (can) be preserved in a vision memory.

Case-based Expectation Generation from Memory Records

Memory records are "cases" which may provide missing information for an ongoing scene:

- **identify memory records which partially match current scene**
- **adapt memory information to current scene**
- **provide expectations about current scene**



Basic Learning Tasks

Michalski 86: Learning is the construction or modification of representations of experiences.

<u>Unsupervised learning</u>

determine reoccuring patterns in scene records

=> conceptual clustering

<u>Supervised learning</u>

determine description covering several examples

=> inductive generalization

Example of Supervised Learning

1. "This is how you lay a table"

2. "This is how you lay a table"

•
•
•

42. "This is how you lay a table"

determine
covering
description

Unsupervised Learning: The Baby Scenario

Given memory records and primitive occurrence models, discover higher-level occurrence models

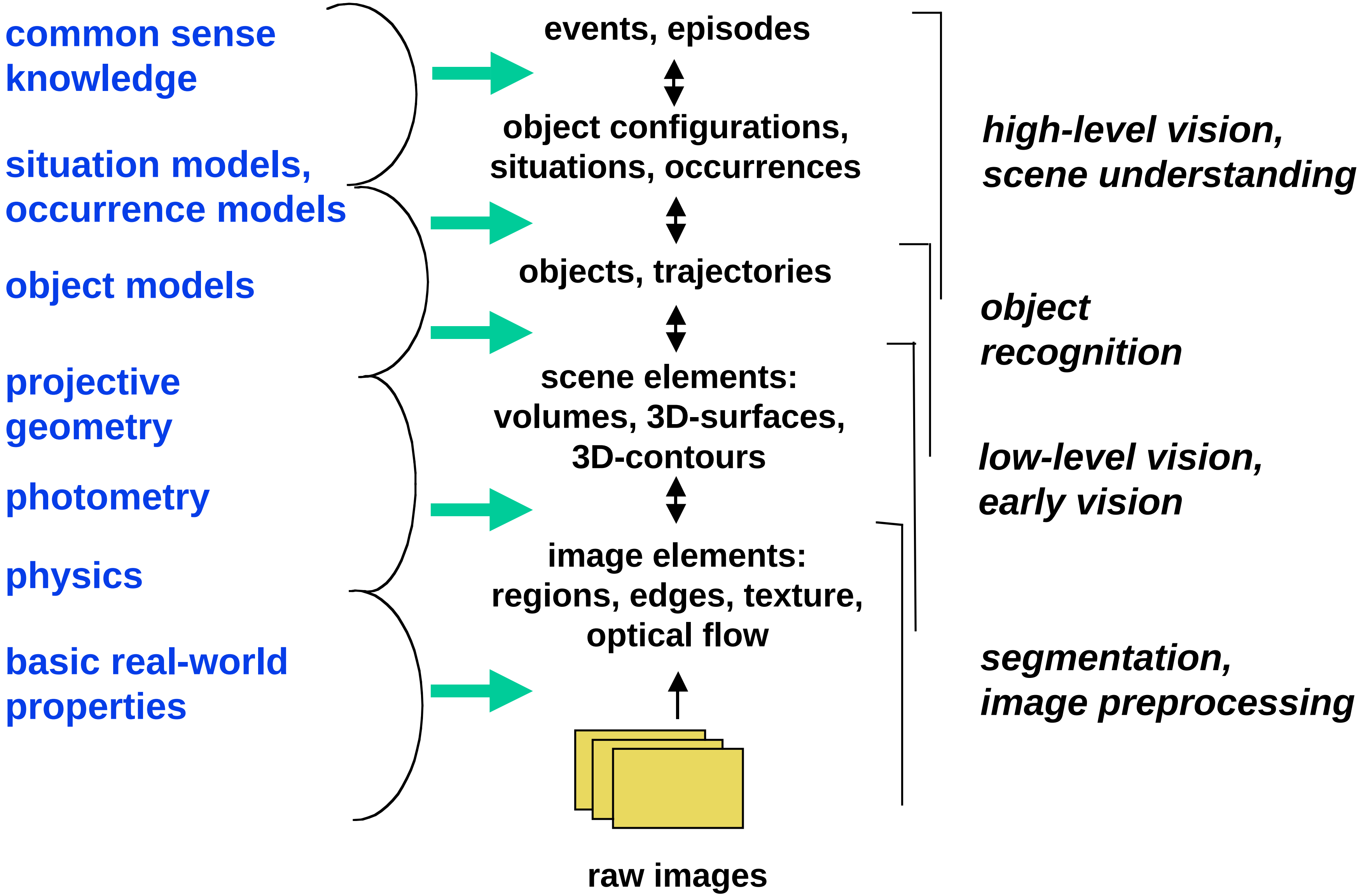
Example:
Discover "transport" = simultaneous motion of hand touching object

Discover commonalities of memory records in terms of

- parts of joint occurrence (e.g. obj1, obj2, motion1, motion2)
- type constraints (e.g. obj1 instance hand)
- temporal constraints (e.g. tb1 = tb2, te1 = te2)
- spatial constraints (e.g. obj1 dc obj2)

Active research area!

Review of Image Understanding as a Knowledge-based Process



Review Week 1

Computer Vision
Contents
Literature
Website
Exercises
Why study image processing, image
    analysis and machine vision?
What is "Image Processing"?
What is "Image Analysis"?
What is "Image Understanding"?
Image Understanding is
    Silent Movie Understanding
What is "Pattern Recognition"?
What is "Computer Vision"?
Computer Vision vs. Biological Vision
Geometry in human vision
Human object perception
Human character recognition
Human face recognition

Complexity of natural scenes
The computer perspective on images
Greyvalues of the section
Street scene containing the section
Computer Vision as an
academic discipline
Important conferences
Important Journals
Important application areas
Example-based image retrieval
Example: Medical image analysis
Example: Driver assistance
History of Computer Vision (1)
History of Computer Vision (2)

Review Week 2

Definition of Image Understanding
Illustration of Image Understanding
Image Understanding as a
Knowledge-based Process
Abstraction Levels for the Description
of Computer Vision Systems
Example for Knowledge-level Analysis
Image Formation
Formation of Natural Images
Intensity of Sensor Signals
Multispectral Images
Spectral Sensitivity of Human Eyes
Dimensions of Colour
RGB Images of a Natural Scene
Discretization of Images
Spatial Quantization
Reconstruction from Samples
Sampling Theorem
Aliasing
Reconstructing the Image Function
from Samples
Sampling TV Signals

Quantization of Greyvalues
Uniform Quantization
Nonlinear Quantization Curves
Optimal Quantization (1)
Optimal Quantization (2)
Binarization
Threshold Selection by Trial and Error
Distribution-based Threshold Selection
Threshold Selection Based on Reference
Image Capturing for Thresholding
Perspective Projection Transformation
Perspective Projection in Independent Co
3D Coordinate Transformation (1)
3D Coordinate Transformation (2)
Perspective Projection Geometry
Perspective and Orthographic Projection
From Camera Coordinates to
Image Coordinates
Complete Perspective Projection Equatio
Homogeneous Coordinates (1)
Homogeneous Coordinates (2)
Homogeneous Coordinates (3)

Review Week 2 (continued)

Inverse Perspective Equations
Binocular Stereo (1)
Binocular Stereo (2)
Distance in Digital Images
Connectivity in Digital Images
Closed Curve Paradoxon
Geometric Transformations
Polynomial Coordinate Transformations
Translation, Rotation, Scaling, Skewing
Example of Geometry Correction
    by Scaling
Minimizing the MSE
Principle of Greyvalue Interpolation
Nearest Neighbour
    Greyvalue Interpolation
Bilinear Greyvalue Interpolation
Bicubic Interpolation

Review Week 3

Global Image Properties
Empirical Mean and Variance
Greyvalue Histograms
Example of Greyvalue Histogram
Histogram Modification
Projections
Cross-sections
Noise
Noise Removal by Averaging
Example of Averaging
Simple Smoothing Operations
Bimodal Averaging
Averaging with Rotating Mask
Median Filter
Local Neighbourhood Operations
Example of Sharpening
Spectral Image Properties
Illustration of
    1-D Fourier Series Expansion
Discrete Fourier Transform (DFT)
Basic Properties of DFT

Illustrative Example of
    Fourier Transform
Examples of Fourier Transform Pairs
Fast Fourier Transform (FFT)
Convolution
Filtering in the Frequency Domain
Filtering in the Spatial Domain
Low-pass Filters
Discrete Filters
Matrix Notation for Discrete Filters
Avoiding Wrap-around Errors
Convolution Using the FFT
Convolution and Correlation
Correlation and Matching
Principle of Image Restoration
Image Restoration by Minimizing
    the MSE

Review Week 4

Image Data Compression
Run Length Coding
Probabilistic Data Compression
Huffman Coding
Statistical Dependence
Karhunen-Loève Transform
Illustration of Minimum-loss
    Dimension Reduction
Compression and Reconstruction with
    the Karhunen-Loève Transform
Example for Karhunen-Loève
    Compression
Predictive Compression
Example of Linear Predictor
Discrete Cosine Transform (DCT)
Principle of Baseline JPEG
YUV Color Model for JPEG
Illustrations for Baseline JPEG
JPEG-compressed Image
Problems with Block Structure of JPEG

Progressive Encoding
MPEG Compression
Quadtree Image Representation
Quadtree Image Compression

Review Week 5

Segmentation
Problems with Segmentation
Primary Goal of Segmentation
Secondary Goals of Segmentation
Thresholding
Representing Regions
Component Labelling
Boundaries
Chain Code
Chain Code Derivatives
k-Slope and k-Curvature
Digital Straight Lines
Uniformity Assumption
Region Growing
Segmentation into Regions
    Using Histograms
Region Segmentation by
    Split-and-merge
Maximum-likelihood Edge Finding
Greyvalue Discontinuities
Are Edges Object Boundaries?
Robert´s Cross Operator
Sobel Operator
Example for Sobel Operator
Kirsch Operator
Laplacian Operator
Marr-Hildreth Operator
Difference of Gaussians (DoG)
Canny Edge Detector (1)
Canny Edge Detector (2)
Examples for Canny Edge Detector

Review Week 6

Grouping
Cognitive Grouping
Fitting Straight Lines
Straight Line Fitting by
    Iterative Refinement
Straight Line Fitting by
    Eigenvector Analysis (1)
Straight Line Fitting by
    Eigenvector Analysis (2)
Straight Line Fitting by
    Eigenvector Analysis (3)
Example for Straight Line Fitting by
    Eigenvector Analysis
Grouping by Search
Dynamic Programming (1)
Dynamic Programming (2)
Grouping by Relaxation
Contexts for Edge Relaxation
Modification Rule for Edge Relaxation
Example of Edge-finding by Relaxation

Histogram-based Segmentation
    with Relaxation (1)
Histogram-based Labelling
    with Relaxation (2)
Relaxation with a Neural Network
Hough Transform (1)
Hough Transform (2)
Hough Transform (3)
Generalized Hough Transform

Review Week 7

Region Description for Recognition
Simple 2D shape features
Euler number
Area
Boxing rectangle
Boundary length
Compactness
Center of gravity
Second-order moments
Axis of minimal inertia
Polar signature
Object recognition using
the polar signature
Convex hull
Skeletons
Thinning algorithm
B-splines (1)
B-splines (2)
Shape Description by
Fourier Expansion (1)
Shape Description by
Fourier Expansion (2)

Templates
Cross-correlation
Artificial neural nets
Multilayer feed-forward nets
Character recognition with a neural net
Learning by backpropagation
Perceptrons (1)
Perceptrons (2)

Review Week 8

What is "Pattern Recognition"?
Basic Terminology for
Pattern Recognition
Example: Animal Footprints
A Feature Space for Footprints
Discriminant Functions for Footprints
Existence of Discriminant Functions
Linear Discriminant Functions
Class Average
Minimal Distance Classification
Nearest Neighbour Classification
Generalized Linear
Discriminant Functions
Linear Discriminant Functions for
2-Class Problems
Perceptron Learning Rule
Minimizing the Discriminant Criterion
Quadratic Criterion Function
Relaxation Rule
Minimum Squared Error
Ho-Kashyap Procedure

Discrimination with
Potential Functions
Construction of Discriminant Functions
Based on Potential Functions
Statistical Decision Theory
Example: Medical Screening (1)
Example: Medical Screening (2)
Example: Medical Screening (3)
General Framework for
Bayes Classification
Bayes 2-class Decisions
Normal Distributions
Discriminant Function for
Normal Distributions
Univariate distribution
Statistically Independent,
Equal Variance Variables
Equal Covariance Matrices
Estimating Probability Densities
Estimating the Mean in a
Univariate Normal Density

Review Week 9

Motion Analysis
Case Distinctions for Motion Analysis
Motion in Video Images
Difference Images
Counting Differences
Corresponding Interest Points
Moravec Interest Operator
Corner Models
Correspondence problem
Correspondence by Iterative Relaxation
Kalman Filters (1)
Kalman Filters (2)
Kalman Filter Example
Diagrams for Kalman Filter Example (1)
Diagrams for Kalman Filter Example (2)

Optical Flow Constraint Equation
Aperture Effect
Optical Flow Smoothness Constraint
Optical Flow Algorithm
Optical Flow Improvements
Optical Flow and Segmentation
Optical Flow Patterns
Optical Flow and 3D Motion (1)
Optical Flow and 3D Motion (2)

Review Week 10

3D Motion Analysis Based on
2D Point Displacements
Structure from Motion (1)
Structure from Motion (2)
Structure from Motion (3)
Perspective 3D Analysis of
Point Displacements
Essential Matrix
Solving for the Essential Matrix
Singular Value Decomposition of E
Nagel-Neumann Constraint
Homogeneous Coordinates
From Homogeneous World Coordinates
to Homogeneous Image
Coordinates
Camera Calibration
Calibration of One Camera
from a Known Scene
Fundamental Matrix
Epipolar Plane

Correspondence Problem Revisited
Correspondence Between Two
Mars Images
Constraining Search for Correspondence
Neural Stereo Computation
Obtaining 3D Shape from
Shading Information
Principle of Shape from Shading
Photometric Surface Properties
Lambertian Surfaces
Surface Gradients
Simplified Image Irradiance Equation
Reflectance Maps
Characteristic Strip Method
Shape from Shading
by Global Optimization
Principle of Photometric Stereo
Analytical Solution for
Photometric Stereo

Review Week 11

General Principles of
3D Image Analysis
Single Image 3D Analysis
Generality Assumption
Texture Gradient
Shape from Texture
Surface Shape from Contour
3D Line Shape from 2D Projections
3D Shape from Multiple Lines
3D Junction Interpretation
3D Line Orientation from
Vanishing Points
Object Recognition
The Chair Room
About Model-based Recognition
Model-based Object Recognition
3D Models vs. 2D Models
Holistic Models vs. Component Models
3D Shape Models
3D Space Occupancy Model
Oct-trees
Extended Gaussian Image (EGI)
Recognition with EGI Models
Illustration of
EGI Recognition Procedure
Representing Axial Bodies
Generalized Cylinders
Conditions for 3D Reconstruction
from Contours
Relational Models
Relations between Components
Object Recognition by
Relational Matching
Compatibility of Relational Structures
Example of a Relational Model (1)
Example of a Relational Model (2)
Example of a Relational Model (3)
Relational Match Using a
Compatibility Graph
Finding Maximal Cliques
Relational Matching with
Heuristic Search

Review Week 11 (continued)

Optimization Techniques
Simulated Annealing (1)
Simulated Annealing (2)
Case study: Drawing Interpretation
Partonomy of Object Parts
Specification of an Arrow
Processing Cycle
Property Spaces
Blackboard Architecture
Analysis of a Machine Drawing
Analysis of an Electrical Circuit
Qualitative Relations
Qualitative spatial relations
Combining Fuzzy Propositions

Recognition of Views by
    Qualitative 2D-Spatial Relations
Views of the Same Location
    from Different Perspectives
Views of the Same Location
    under Different Illumination
Relational Description of Views
Location Relation between Edges
Compatibility Test for
    Location Relation
Determining Offset Regions
Offset Regions for Different
    Uncertainty Intervals

Review Week 12

High-level Vision
Topics of High-Level Vision
Basic Building Blocks for
    High-level Scene Interpretation
Basic Representational Units
Temporal Decomposition of Scenes
Temporal Relations
Interval Relations in
    Allen´s Algebra
Convex Time-point Algebra
Perceptual Primitives
Qualitative Predicates for Modelling
    and Recognizing Occurrences
Example: Criminal Act Recognition
Qualitative Predicates for Occurrences
    in Traffic Scenes
Occurrence Models
Occurrence Model for Overtaking
    in Street Traffic
Occurrence Model for
    Transport Vehicle Behaviour

Occurrence Model for Placing a Cover
Parts Structure
Concept Hierarchy
Relational Structure for Placing-a-cover
Model-based Interpretation
Temporal Constraint Net for
    Convex Time-Point Algebra
Occurrence Recognition by
    Constraint Propagation (1)
Occurrence Recognition by
    Constraint Propagation (2)
Occurrence Recognition by
    Constraint Propagation (3)
Convergence and Complexity
Generalization of Temporal Relations
Recognizing Intentions and Plans
Definition of Planning
Plan Recognition
Models for Intention Recognition

Review Week 12 (continued)

From Scene Data to a
Natural-language Scene Description
Geometrical Scene Description (GSD)
Typical data of a GSD
Hierachy for object motions in
street traffic
Generating a
natural-language description
Standard plan for generating natural-
language scene descriptions
Example of an automatically generated
traffic scene description
Selecting prepositions for
trajectory location information

Review Week 13

Probabilistic Models for Occurrences
Causality Graph
Constructing a Bayes Net
Computing Inferences
Example: Traffic Behaviour of Pedestrians
Estimating Probabilities from a Database
Expectation Maximization (1)
Expectation Maximization (2)
Example for Expectation Maximization (1)
Example for Expectation Maximization (2)
Example for Expectation Maximization (3)
Hidden Markov Models
Notation for HMM
Properties of a Homogeneous HMM
Paths through a HMM
Finding Most Probable Paths
Example for Viterbi Algorithm
Model Evaluation for Given Observations
Example for Model Evaluation (1)
Example for Model Evaluation (2)

Learning Models for
High-level Image Interpretation
Learning in Support of
High-level Scene Interpretation
Basic Structure of Vision Memory
Case-based Expectation Generation
from Memory Records
Basic Learning Tasks
Example of Supervised Learning
Unsupervised Learning:
The Baby Scenario
Review of Image Understanding as
a Knowledge-based Process