Acting under Uncertainty, Opponents, Dynamics…

Manuela Veloso
Reid Simmons

Carnegie Mellon University
Computer Science Department

Planning – Fall 2008

A Few More Final Ideas

- Multifidelity plans

Acting Under Uncertainty

- Conditional plans
- MDPs
- POMDPs
- Action uncertainty
- State uncertainty

Multi-Fidelity Plans

Winner & Veloso, AAAI-2000

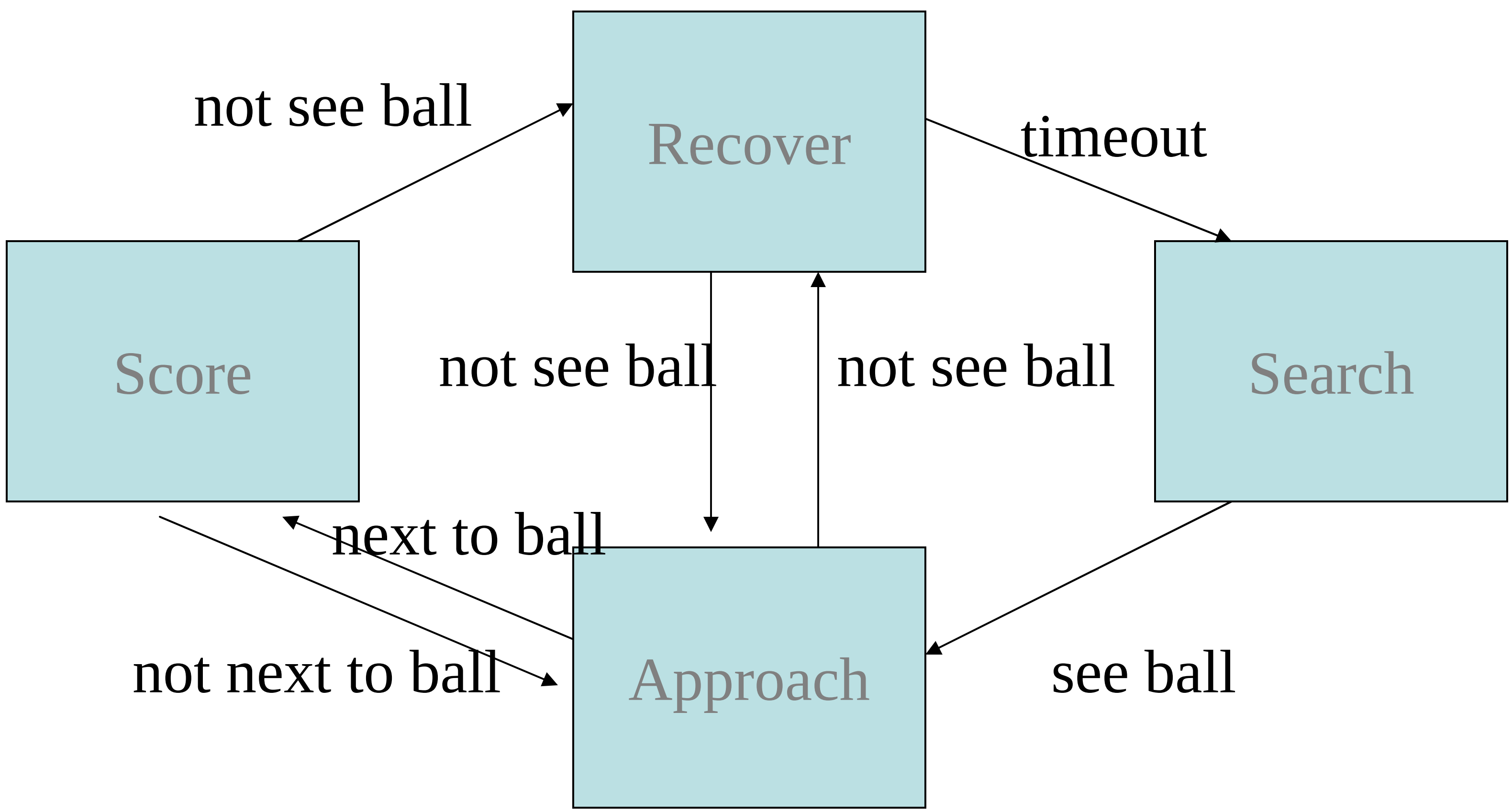
- Acting with state information of *different quality*
  - General modes of behavior
  - Different levels of behaviors as a function of the ``*level of accuracy*'' of the processed sensory data: visual, localization

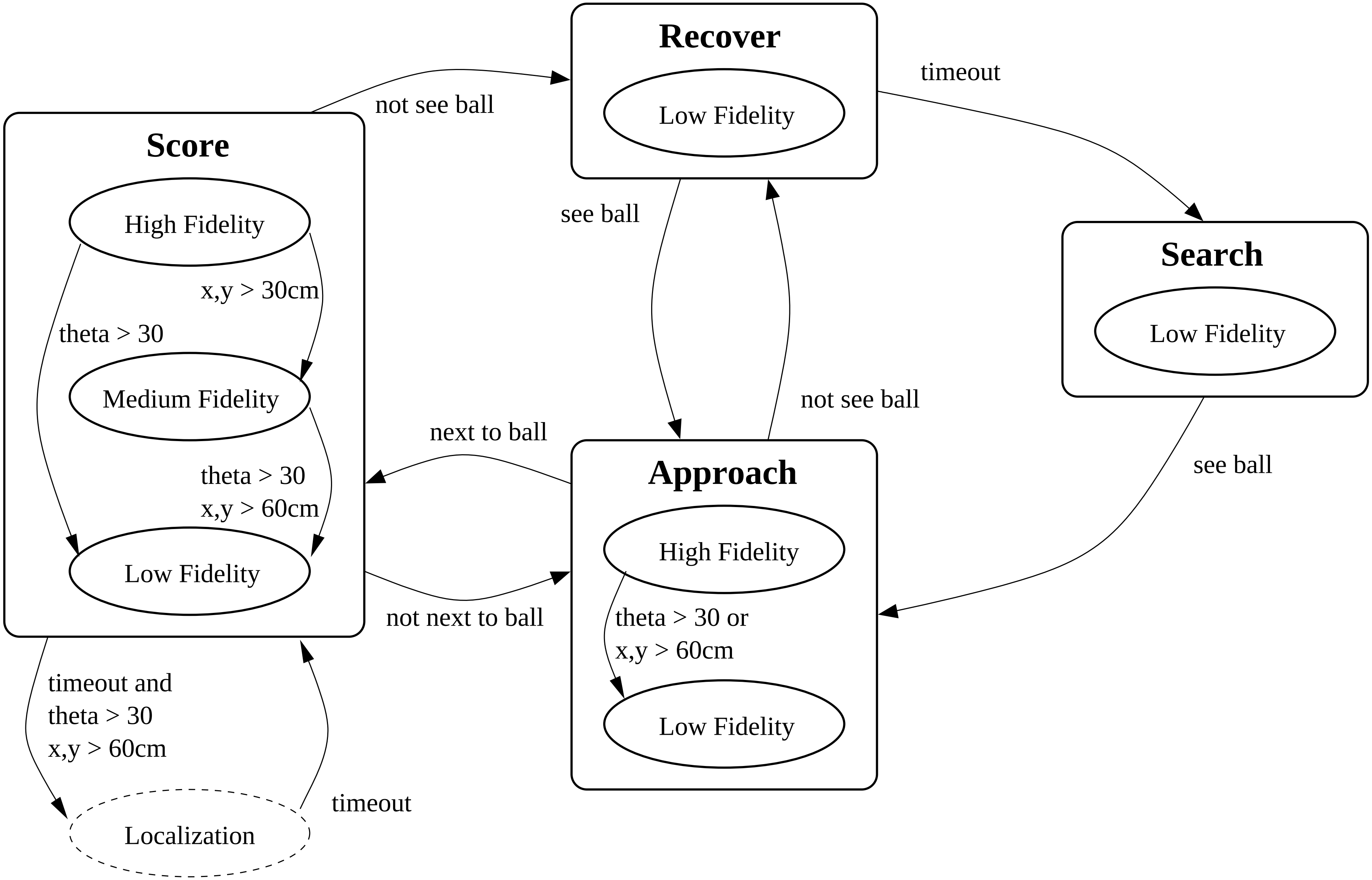
- Empirical setting of fidelity thresholds

Example

- Localization: crucial input for robot motion
- Uncertainty in localization

Planning with a FSM




Recover
Low Fidelity
timeout
not see ball
Score
High Fidelity
x,y > 30cm
theta > 30
Medium Fidelity
theta > 30
x,y > 60cm
Low Fidelity
see ball
not see ball
Search
Low Fidelity
next to ball
Approach
High Fidelity
theta > 30 or
x,y > 60cm
Low Fidelity
see ball
not next to ball
timeout and
theta > 30
x,y > 60cm
timeout
Localization

Approach

- **Low:**
  Run straight towards the ball
- **High:**
  Skew approach to ball to get behind it,
  when closer to its goal position.

Score

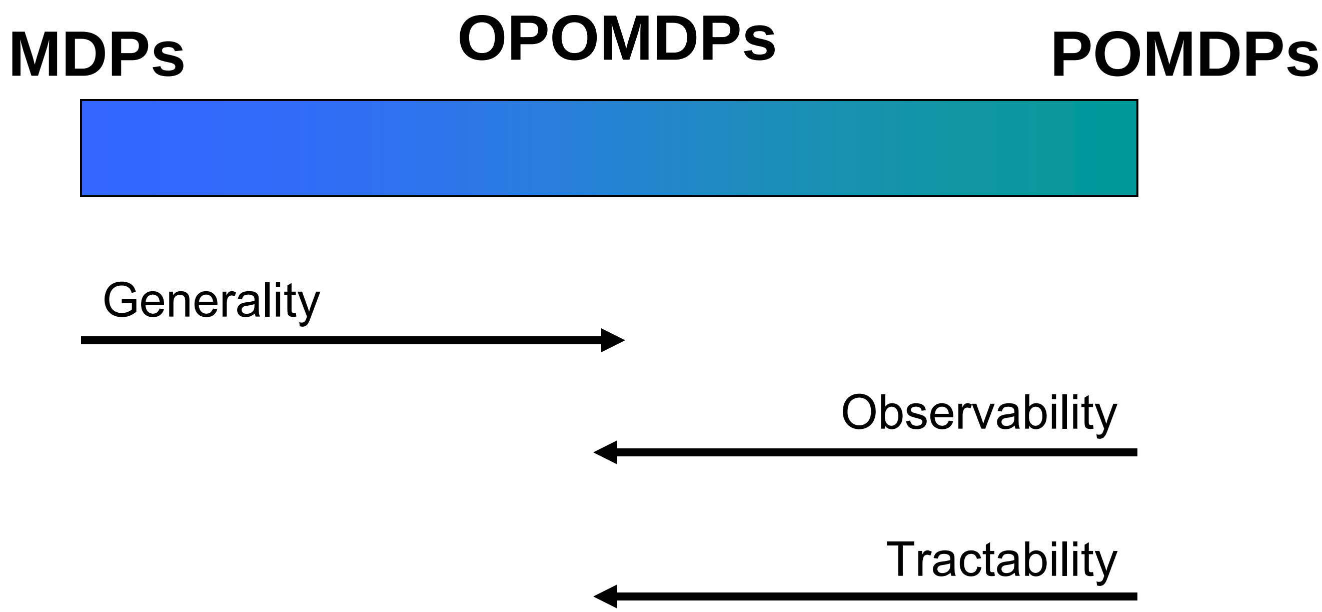
- **Low:**
  Until the robot sees the goal,
  Walk sideways around the ball.
  Walk forward pushing the ball.
- **Medium-High:**
  Circle ball using shortest distance
  If facing goal, push ball forwards.

A Few More Final Ideas

- Multifidelity plans
- Oracular POMDPs

"Uncertainty Continuum"

OPOMDPS – Armstrong-Crews et al, ICRA'07, ICRA'08, forthcoming 2010 PhD thesis



Q-MDP

- Solve the underlying MDP, state transitions that result of agent's actions, assuming that at policy-generation time, the world state is fully observable.
- The MDP solution produces a set of value functions $Q_a(s)$, as the expected future reward for taking action a from s.
- $Q_{MDP}(b) = \text{argmax sum } b(s) \times Q_a(s)$

The Oracular POMDP

- "Oracle" action gives perfect state info
- Fixed cost to consult the oracle
- No other observations

$$\mathcal{S} \equiv \mathcal{S}^{\mathrm{MDP}}$$

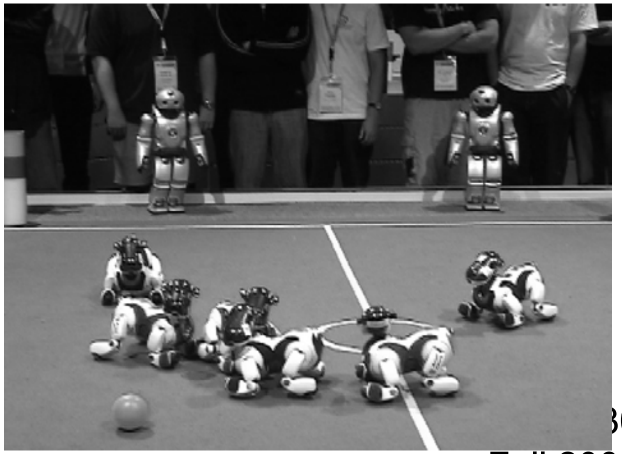
$$\mathcal{A} \equiv \mathcal{A}^{\mathrm{MDP}} \bigcup \{o\}$$

$$\tau(b,a)|_{s'} = \sum_{s} b(s) \mathcal{T}^{\mathrm{MDP}}(s,a,s') \quad \forall\, a \in \mathcal{A}^{\mathrm{MDP}}$$

$$\rho(b,a) = \begin{cases} \sum_s b(s) \mathcal{R}^{\mathrm{MDP}}(s,a) & a \in \mathcal{A}^{\mathrm{MDP}} \\ \sum_s b(s) \mathcal{R}^{\mathrm{MDP}}(s, \mathrm{NO\_OP}) & a = o \end{cases}$$

Examples of OPOMDPs

- Human/robot interaction
- Heterogeneous multirobot team
- Covert operations
- Active learning



JIV Algorithm – J_{MDP} Information Value

- Greedy approximation of the value function
- Solve the underlying MDP during pre-planning, then use MDP solution at execution time.
- Choice:
  - take domain-level action that maximizes long-term expected reward – use Q_{MDP}
  - Consult and pay oracle to reduce uncertainty

The JIV Heuristic (Armstrong-Crews and Veloso '07)

- J^{MDP} Information Value
- Solve underlying MDP beforehand
- At runtime, greedily choose the action with maximum estimated value:

$$\widehat{Q}^{\,\mathrm{JIV}}(b,a) \equiv \begin{cases} \rho(b,a) + \gamma \widehat{J}^{\,\mathrm{QMDP}}(b') & \text{if } a \neq o \\ \rho(b,o) - \lambda + \gamma \widehat{J}^{\,\mathrm{JIV}}(b) & \text{if } a = o \end{cases}$$

$$\widehat{J}^{\,\mathrm{JIV}}(b) \equiv \sum_{s} b(s) J^{\mathrm{MDP}}(s)$$

$$\widehat{J}^{\,\mathrm{QMDP}}(b) \equiv \max_{a} \sum_{s} b(s) Q^{\mathrm{MDP}}(s,a)$$

JIV Benefits

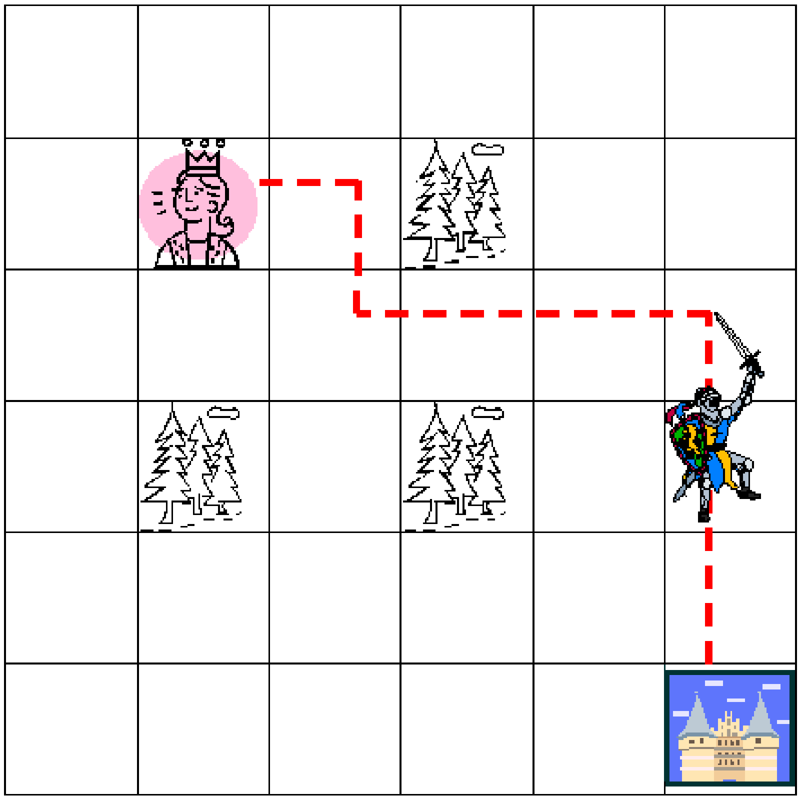
- Poly-time in |S| and |A|
- “Fixes” Q-MDP to include information-gathering actions
- Simple and easy to implement
- Performs well empirically

- But is a greedy heuristic

LA-JIV (Armstrong-Crews and Veloso ‘08)

- Lookahead J^{MDP} Information Value
- Uses heuristic search, as HSVI
- Performs this search starting at each *pure belief*
- Only descends action nodes; recurses at observation nodes, as Hansen’s policy iteration
- Starts with good upper and lower bounds

The Wizard's Curse



$\lambda = .75$

$b_0(s) = b^{pure}$

N	S	W	E	STAY
$\begin{pmatrix} .1 & .7 & .1 \\ 0 & .1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & .1 & 0 \\ .1 & .7 & .1 \end{pmatrix}$	$\begin{pmatrix} .1 & 0 & 0 \\ .7 & .1 & 0 \\ .1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & .1 \\ 0 & .1 & .7 \\ 0 & 0 & .1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

OPOMDP Empirical Results

TABLE I

COMPARISON OF EMPIRICAL SOLVE-TIMES

	36 states	900 states
LA-JIV	2.4	3219
JIV	.03	6.5
PBPI	74.5	$>$ 21000
HSVI	23	7200

OPOMDP Empirical Results

TABLE I

COMPARISON OF EMPIRICAL SOLVE-TIMES

	36 states	900 states
LA-JIV	2.4	3219
~~**JIV**~~	~~.03~~	~~6.5~~
PBPI	74.5	$>$ 21000
HSVI	23	7200

A Few More Final Ideas

- Multifidelity plans
- Oracular POMDPs

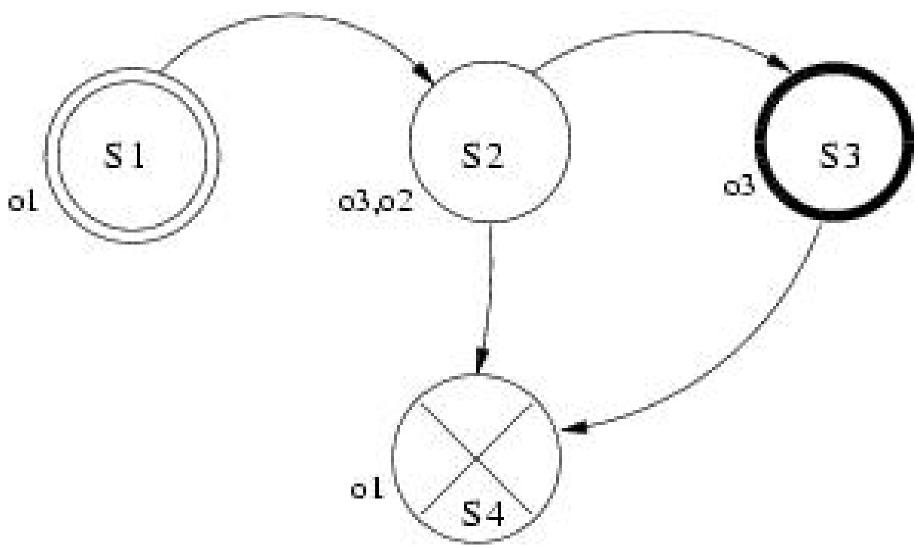
- Observation and recognition

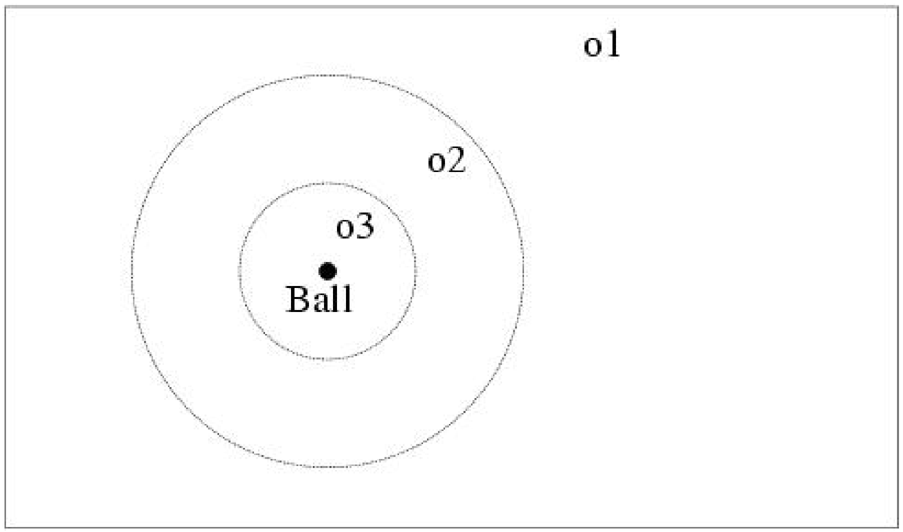
Inference from Observation

Han and Veloso, 1999, Balch 99,2000, lots of work

HMM: initial, intermediate, accepting and rejecting states; probabilistic transitions.

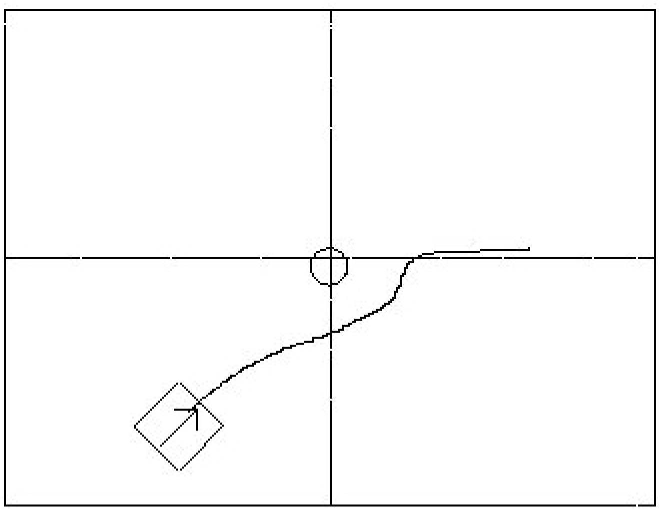
HMM observations: trace of agent activity.



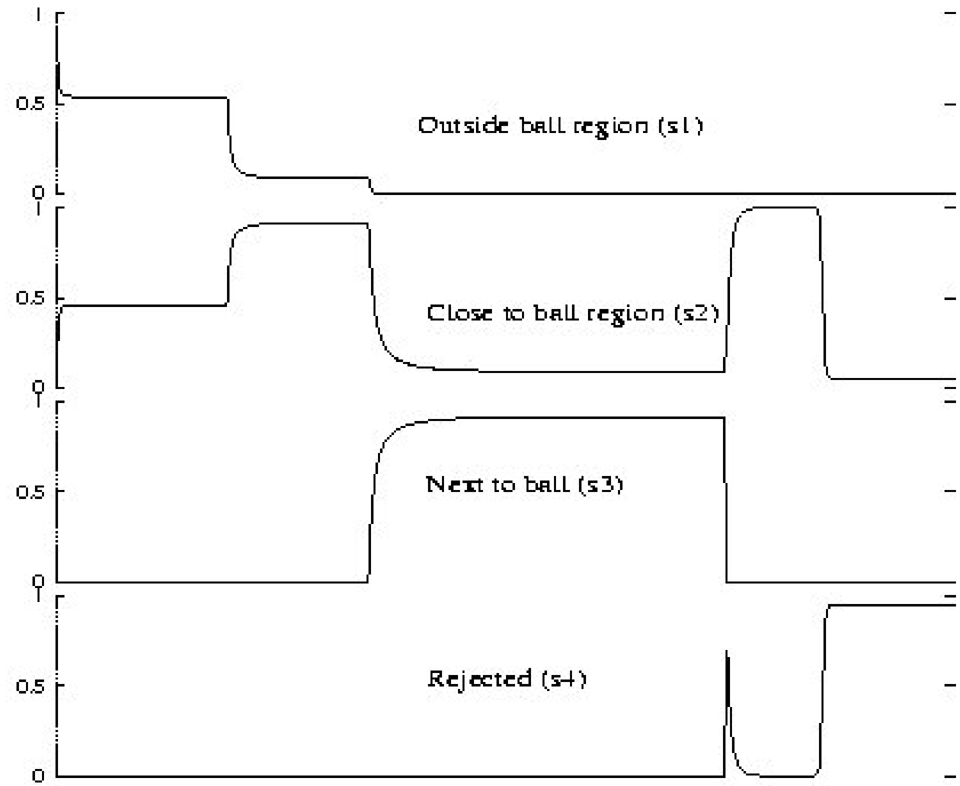


Recognition

Agent activity embeds internal states.



Computation of conditional probabilities of HMM states.



Summary

- Planning under uncertainty
- Acting under uncertainty
- Modeling dynamics