InfoMax Control for Social Robots

Ian Fasel\textsuperscript{1}, Paul Ruvolo\textsuperscript{2,3}, Tingfan Wu\textsuperscript{2,3}, Javier Movellan\textsuperscript{3,4}

\textsuperscript{1}Department of Computer Science\textsuperscript{2,3} University of Arizona\textsuperscript{3} CalIT2,\textsuperscript{4} Institute for Neural Computation University of California San Diego

Problem

Robots with multiple directional sensors
Integrating multiple unreliable sensors with very different characteristics to produce intelligent behavior. It is very difficult to scale up heuristic, hand-coded approaches. A principled approach: Stochastic Optimal Control + Infomax.

Goal: A Socially Aware Robot
Create a dynamic map of where people are located and what their state is (identity, facial expression).

Sensors:
- Foveal vision: Oriented camera with 22 degree field of view - face detection and facial expression recognition.
- Peripheral vision: Motion camera with 90 degree field of view.
- Audio localization: Based on inter-aural delays in multi-microphone system.

Actuators:
- Orientation of foveal vision (the eyes in Einstein robot).

InfoMax control

POMDP problem
State: Location and expression of people (25 possible locations)
Observations: Vision and Audio sensors
Actions: Move head to any orientation (discretized)
Reward: Negative entropy of the belief distribution
Environment: Room in which people randomly enter, move, leave.

Sensors are integrated using conditional random field to produce beliefs about locations and expressions of all people in the room.

Learning the policy
Classic dynamic programming approaches did not scale up. Therefore used Natural Actor Critic (Peters, Schaal, 2008) to directly optimize a convolutional neural network-based policy.

Controller is a convolutional RBF neural network:
- Input: beliefs about location and expression
- Output: Move head to a target orientation

Enforced translation invariance and symmetry about x-axis reflects:
- Only relative positions of people matter, not absolute positions
- Beliefs “flipped” about x-axis should produce identical (but similarly “flipped”) actions

Learned Policy
The policy takes the form of a non-linear filter applied to the probability maps.
Intuitive interpretation:
- Look for regions of high uncertainty
- Look for regions with faces
- Avoid regions where there are probably not any faces

Emergent Behaviors
System learned intuitive behaviors automatically:
- Track individuals
- Alternate gaze between people
- Turn to look in the direction of sudden noises or motions