Interaction with an autonomous agent

Jim Little
Laboratory for
Computational Intelligence
Computer Science
University of British Columbia
Vancouver BC Canada

Background

When we build systems that interact with the world, with humans, and with other agents, we rely upon all of the aspects of cognitive vision:

- knowledge representation
- descriptions of the scene and its constituent objects
- models of agents and their intentions
- learning
- adaptation to the world and other agents
- reasoning about events and about structures
- interpretation of other agents' and users' interactions
- recognition and categorization

Background (cont.)

I will review the ongoing Robot Partners project at UBC which focuses on the design and implementation of visually guided collaborative agents, specifically interacting autonomous mobile robots. I will show how

- localization and mapping
- user modeling
- interpretation of gestures and actions
- interaction with human agents

are achieved in the context of the project

Cognitive Vision for Agents
Robot Partners:
Visually Guided Multi-agent Systems

Jim Little
Laboratory for
Computational Intelligence
Computer Science
University of British Columbia
Vancouver BC Canada

Personnel

- Jim Little (Project Leader), CS UBC
- James J. Clark, ECE McGill University
- Nando de Freitas, CS UBC
- David G. Lowe, CS UBC
- Alan K. Mackworth, CS UBC
- Dinesh K. Pai, CS UBC
- Stephen Se, MD Robotics

Research Focus

To develop techniques for the specification, design and implementation of collaborative robotic systems, based on rich interaction between humans and sensor-based robotic systems.

New tools:

- stochastic constraint-based design
- real-time systems for vision
- control with event-based interaction with perception
- stereo-based shape and appearance models
- robot-human communication based on gestures, sounds, and spoken commands
- precise visually guided localization for mobile agents

Benefits

- Theories and tools for modelling, developing and verifying constraint-based controllers for stochastic agent environments.
- Multi-agent algorithms for collaboration and competition in e-games with mixed initiative (human/robot) control.
- Location recognition capability for mobile agents
- Theories and implementations of human/robot communication through gestures and speech
- Mobile robot collaboration on mapping and model building
- Distributed surveillance and monitoring

Applications

Autonomous and semi-autonomous systems that assist and partner with humans:

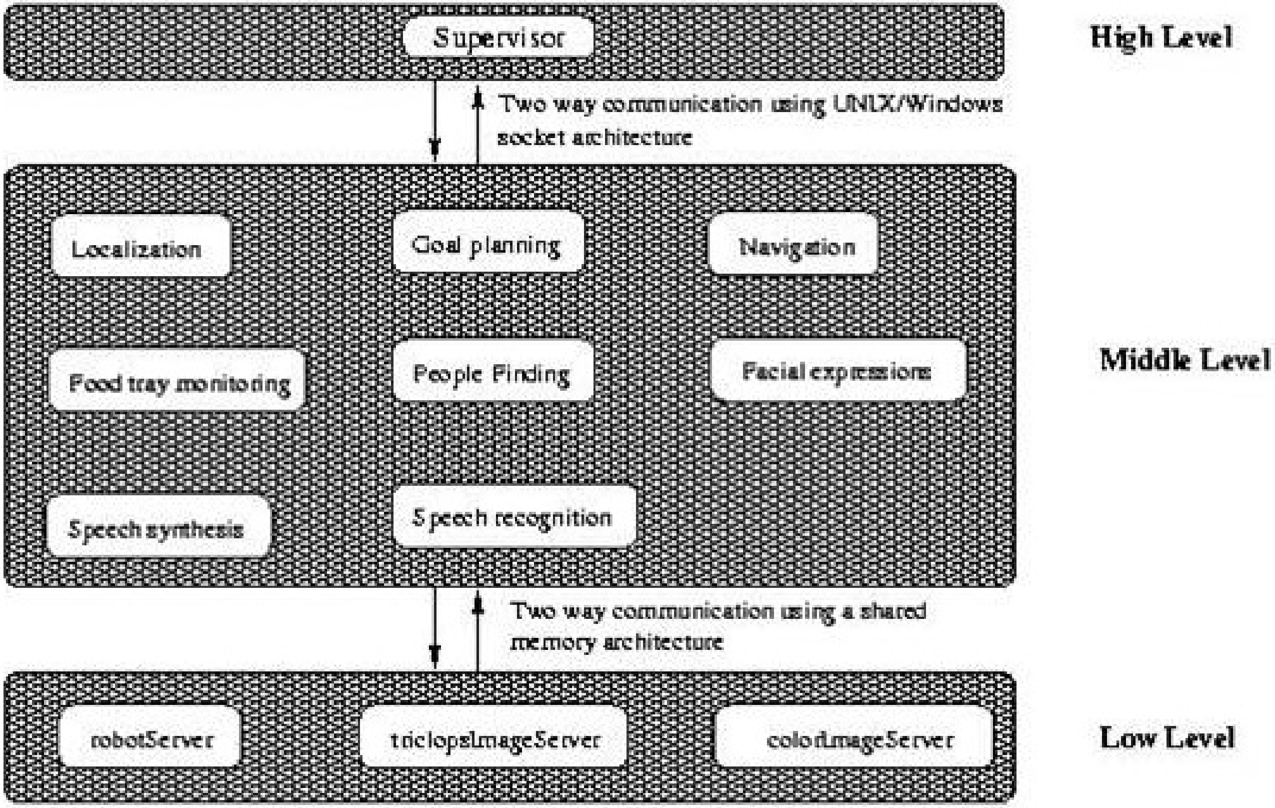
- warehouse and inventory control systems
- construction systems: teams of vehicles for surveying, collecting, and retrieval
- office assistants
- mapping and modeling; surveillance and monitoring
- remote agents for telepresence in meetings, tours and lectures

Our methods/technologies, such as pose estimation, tracking, object/world modeling, localization, and model learning, have direct application in all embedded systems, as well as space applications, including monitoring activity and autonomous and semi-autonomous exploration.

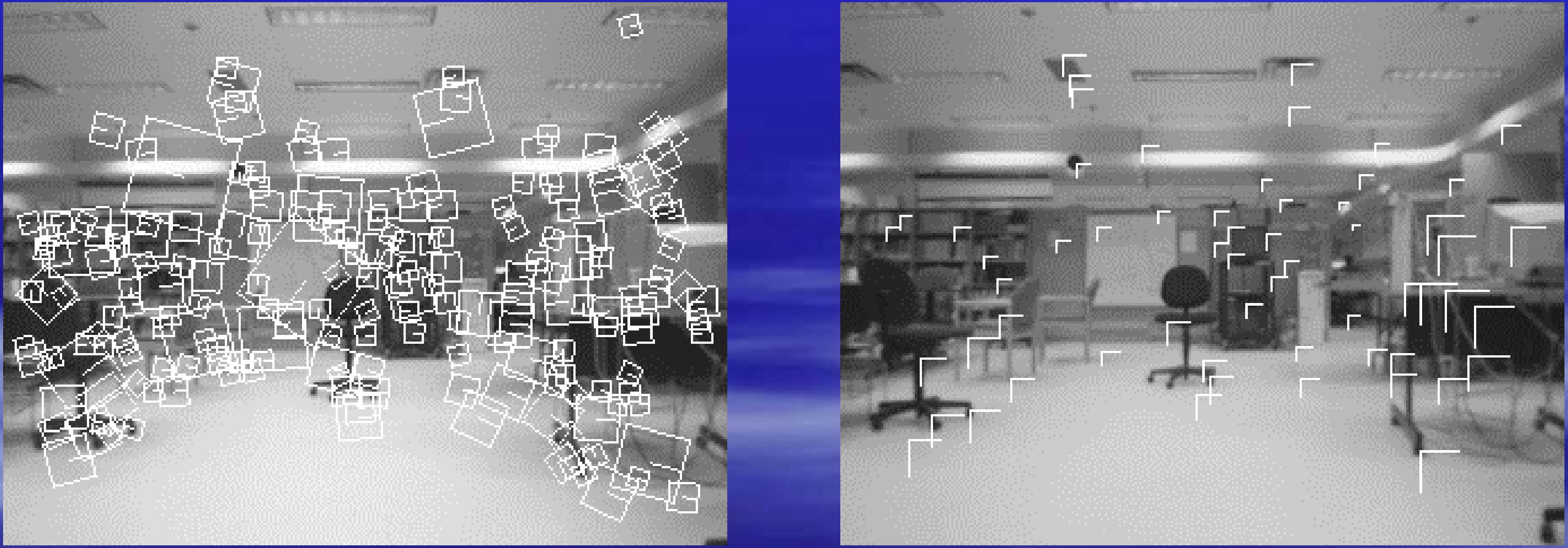
José: Autonomous service

- Robot Control Architecture
- Localization
- Navigation - Avoiding dynamic obstacles
- People Finding
- Location Decision: Where to serve next?
- Food Tray Monitoring
- Face Modeling
- Eric's persona
- Speech Recognition and Synthesis
- Interaction

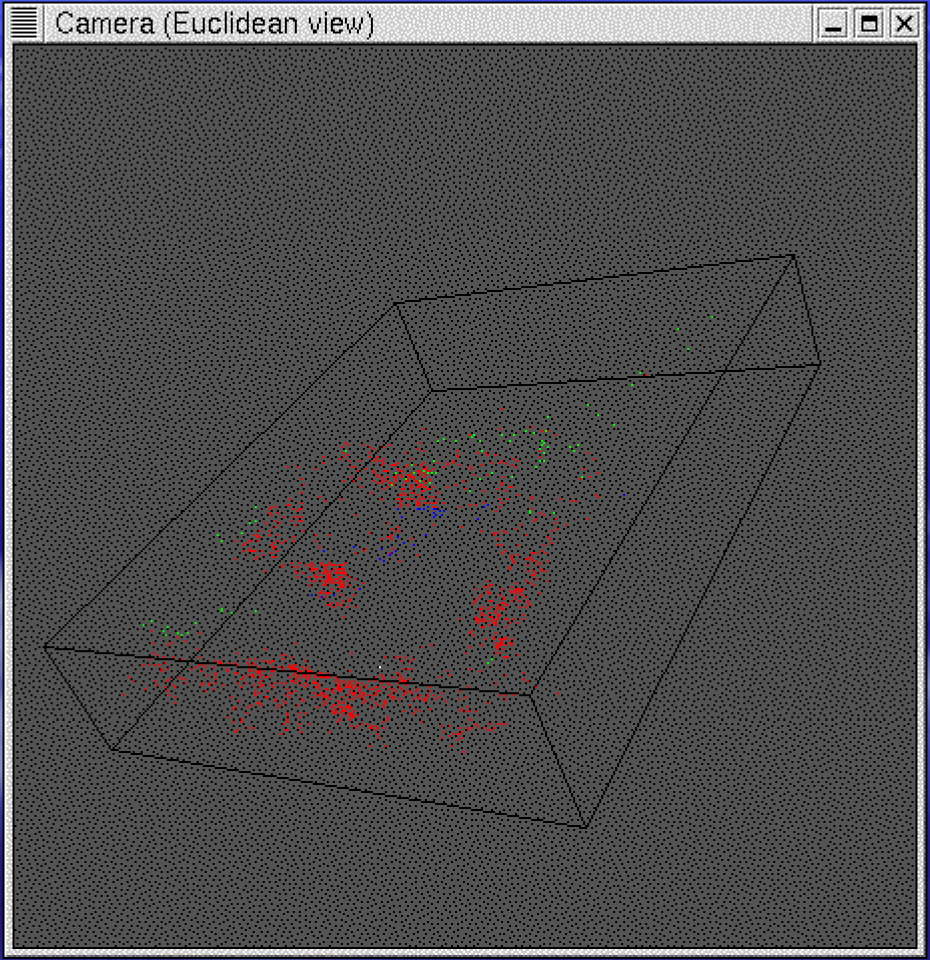
Control Architecture



SIFT-base localization



3D Map



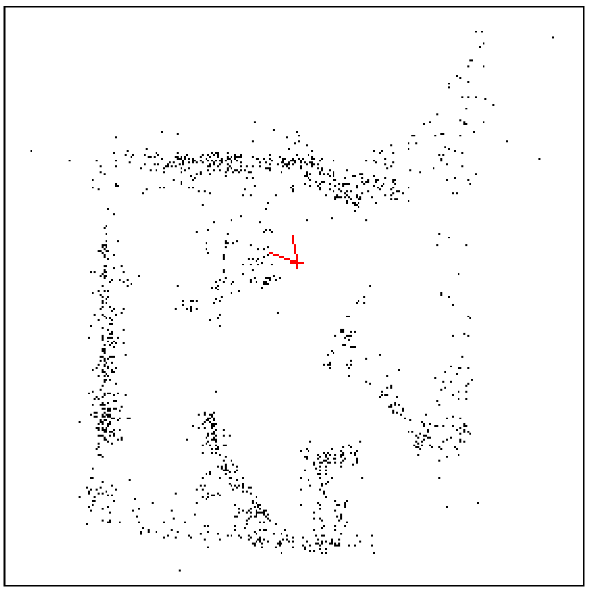
Red: Features on the Walls

Blue: Features on the Floor

Green: Features on the Ceiling

Global Localization

- Kidnapped robot problem
- Recognize robot pose relative to a pre-built map
- Hough Transform matching approach
- Vote bins for pose of each potential match
- Peak : pose with most matches



Measured Pose :
(70, 300, -40°)
Estimated Pose :
(75.8, 295.9, -41.1°)

Face

Neutral

Surprised

Angry

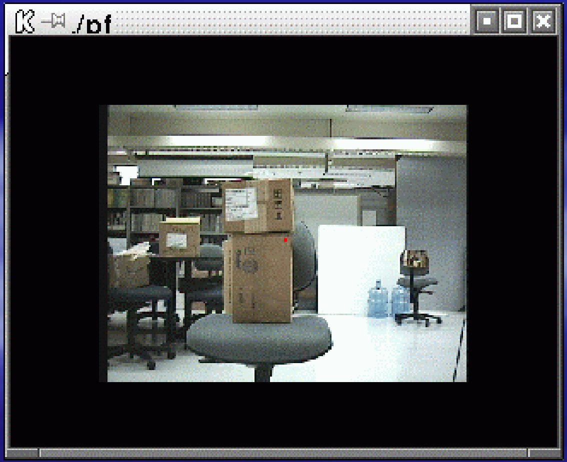
Sad

Finding People

- Construct occupancy grid probability map of where people are standing
- Use the map to decide where to serve next
- Detect people using skin color segmentation
- Use stereo data to compute 3D position of people
- Project locations to floor plane
- Decrease the probabilities over time because people move around

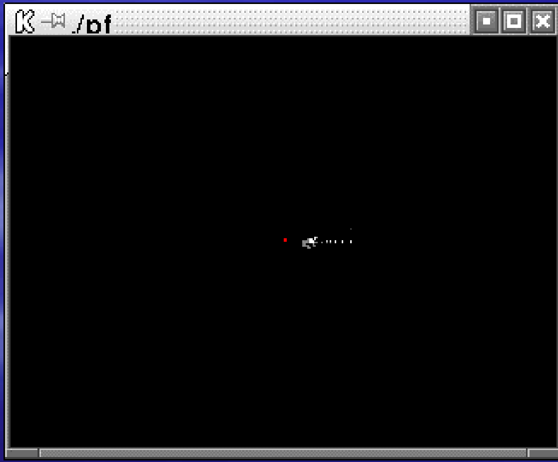
Finding People

Use occupancy grid probability map to decide where to serve
Detect people using skin color
Use stereo data to compute 3D position of people
Decrease the probabilities over time because people move



Color Image

Skin Regions



Probability Map

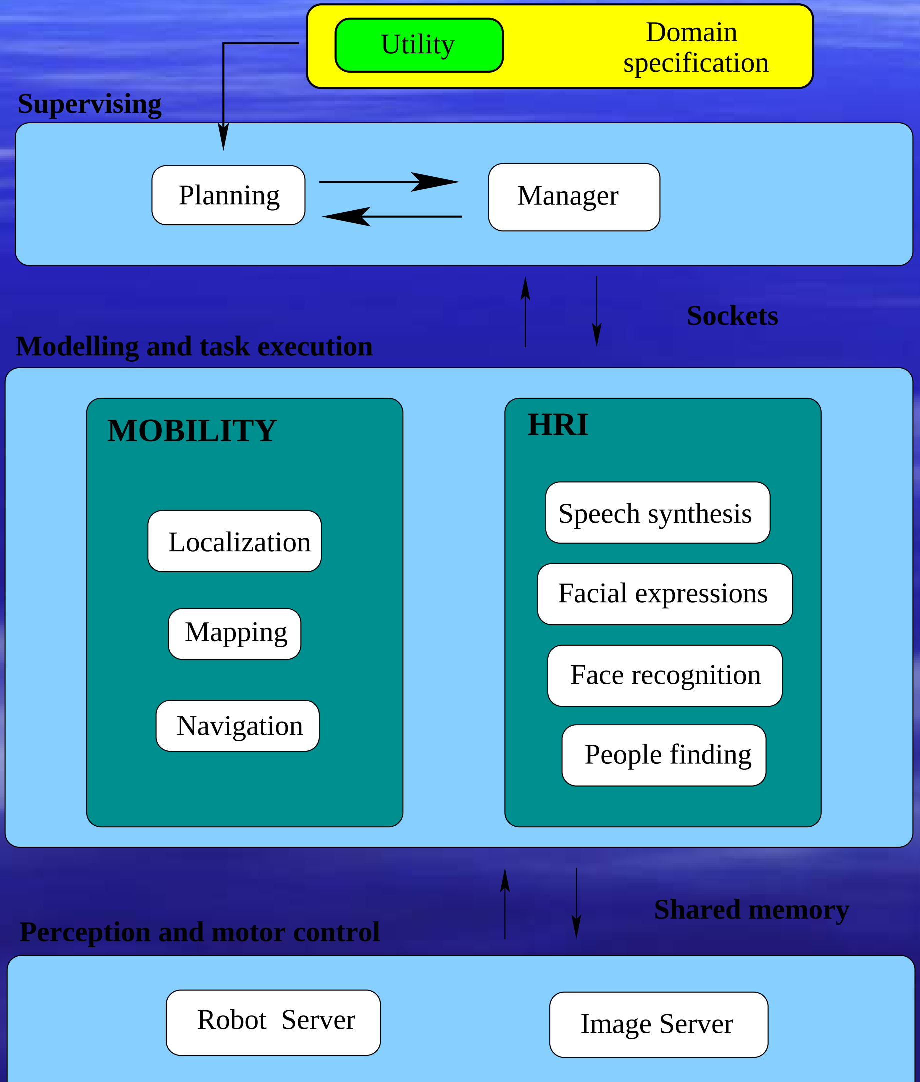
Homer:
Human Oriented Messenger Robot

Homer:
Human Oriented Messenger Robot

- stereo-vision guided mobile robot for performing human-interactive tasks.
  - navigation, localization, map building and obstacle avoidance
  - human interaction capacities
    - person recognition
    - speech,
    - facial expression and gesture recognition
    - human dynamics modeling.
- capabilities
  - modular and independent,
  - integrated in a consistent and scalable fashion
  - controlled by a decision-theoretic planner to model the uncertain effects of the robot's actions
- planner uses factored Markov decision processes,

  allowing for simple specification of tasks, goals and state spaces.
- Task: message delivery task

Control Architecture



Utility

Domain specification

Supervising

Planning

Manager

Sockets

Modelling and task execution

MOBILITY

Localization

Mapping

Navigation

HRI

Speech synthesis

Facial expressions

Face recognition

People finding

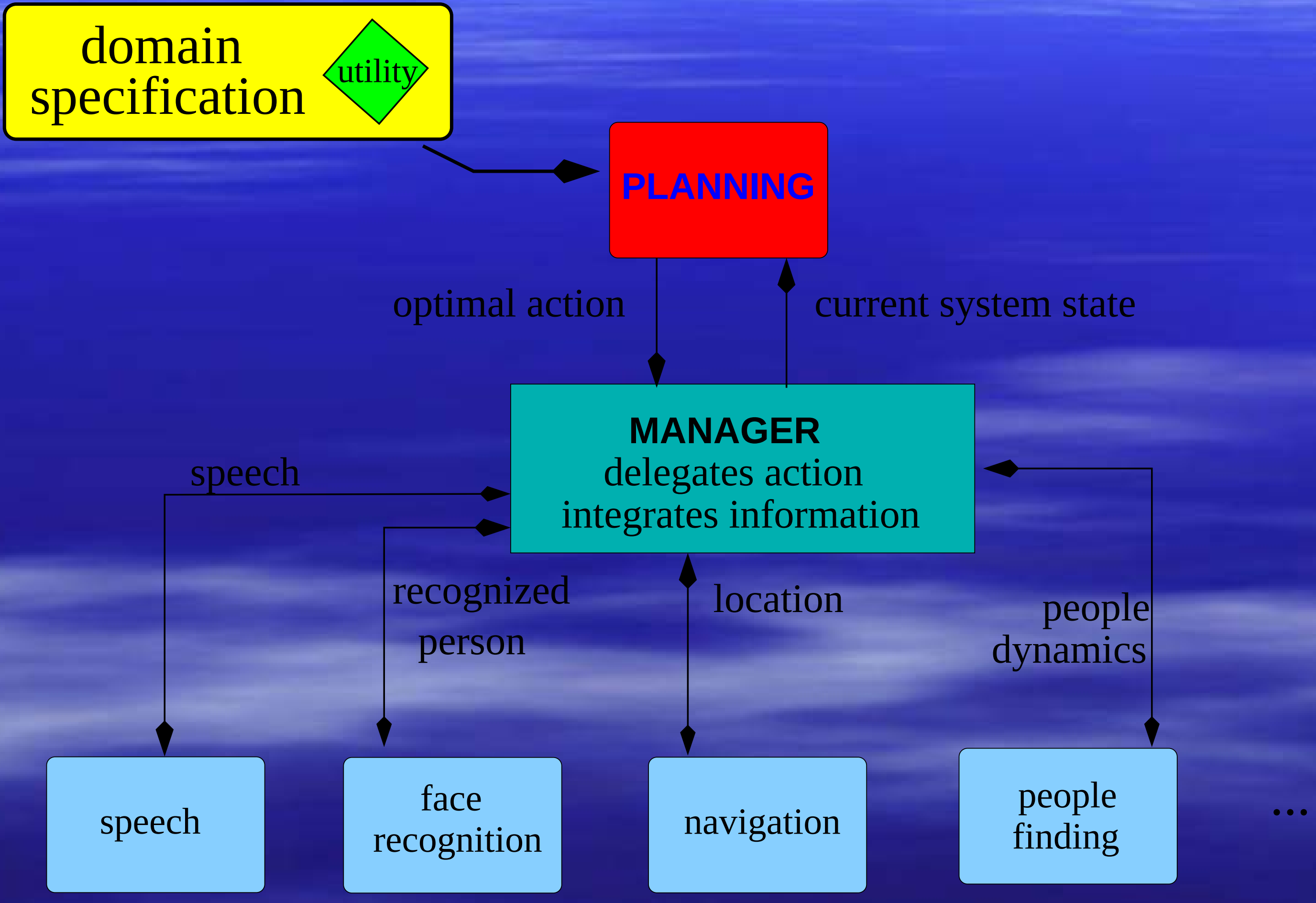
Shared memory

Perception and motor control

Robot Server

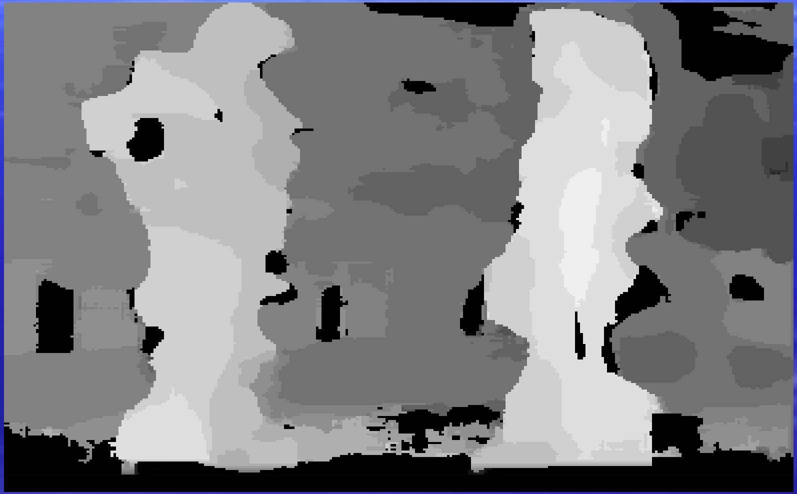
Image Server

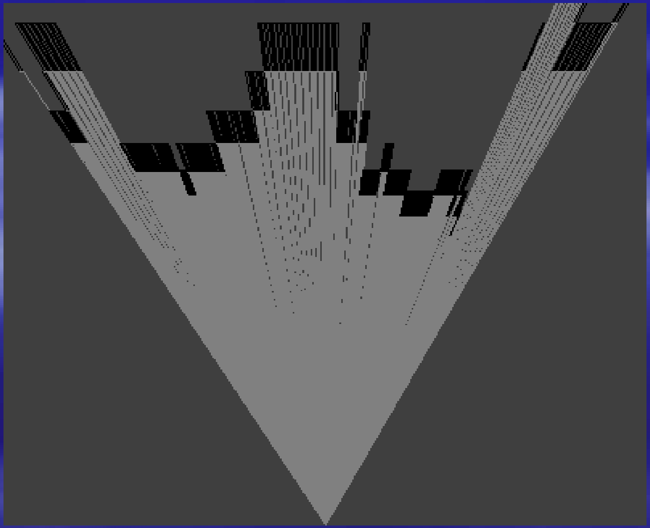
Task Organization

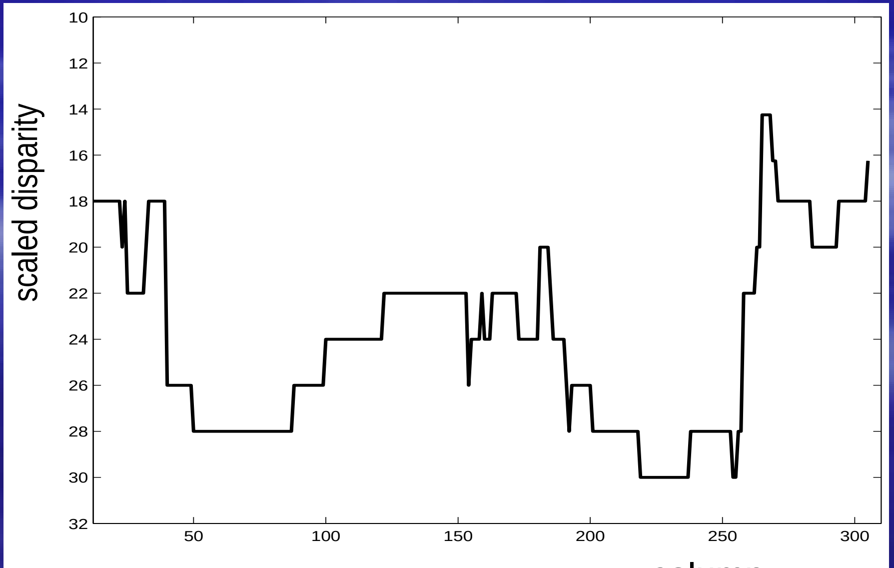


modules control actuators on robot

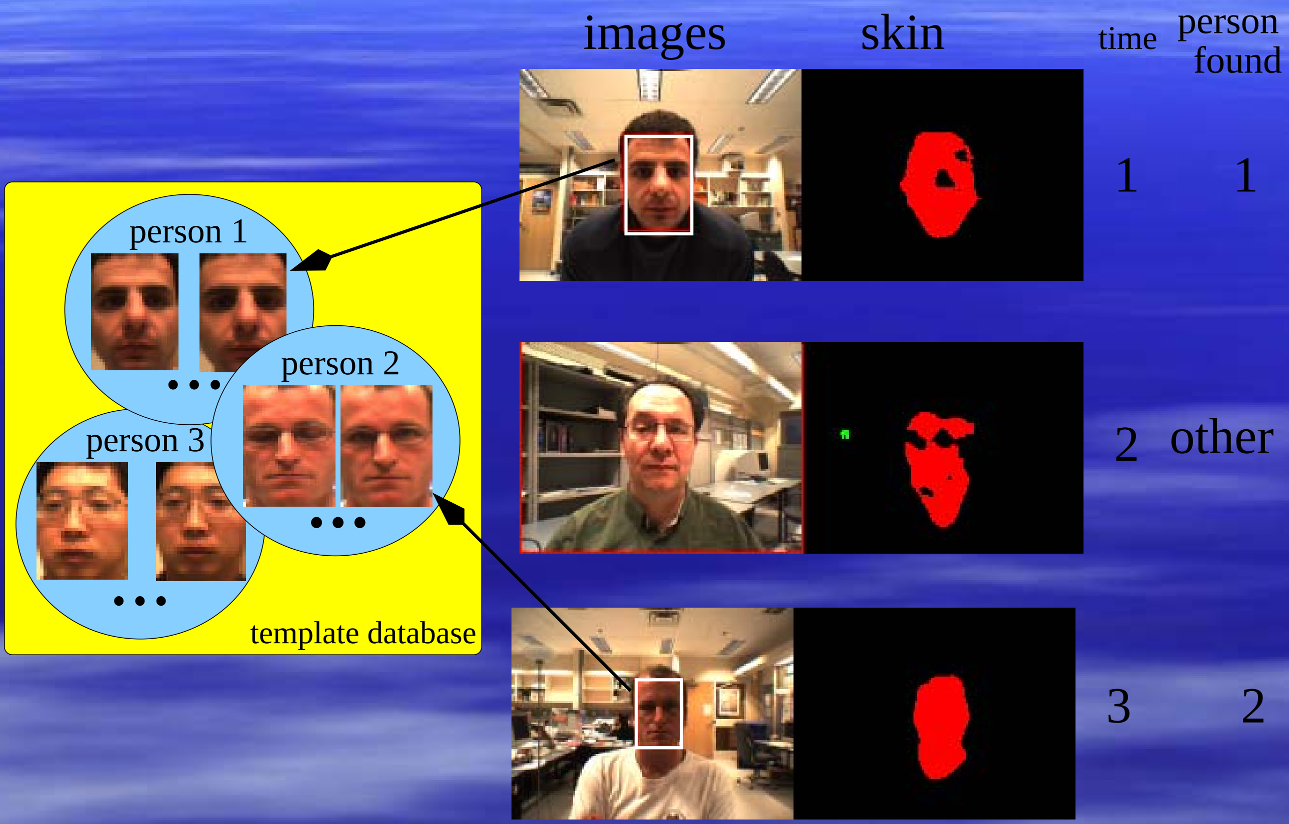
From stereo to maps







Face recognition



person 1

person 2

person 3

template database

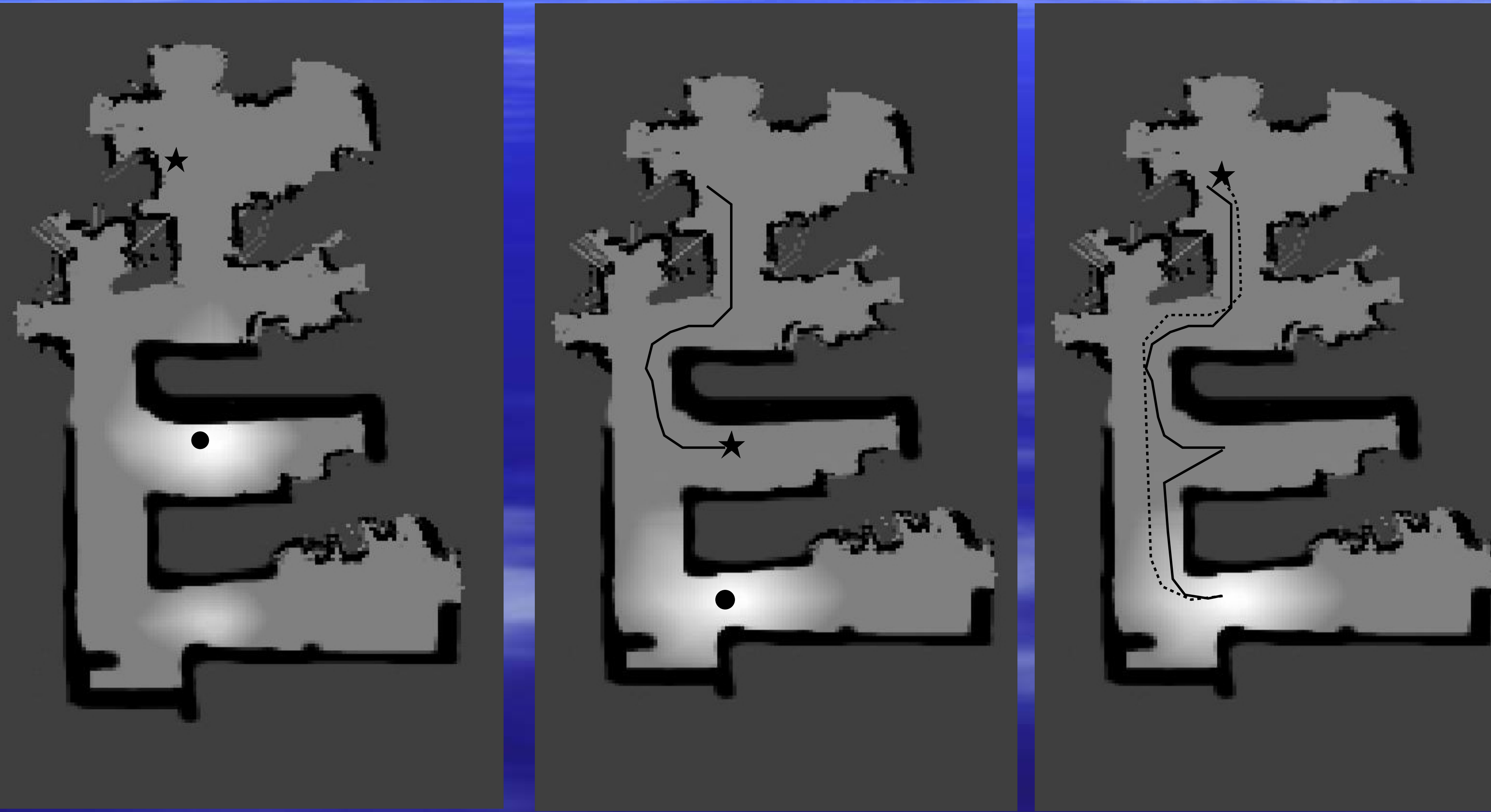
images

skin

time	person found
1	1
2	other
3	2

Exemplars for three persons, input images linked to their most likely and most likely match scores and reported person.

HOMER test run

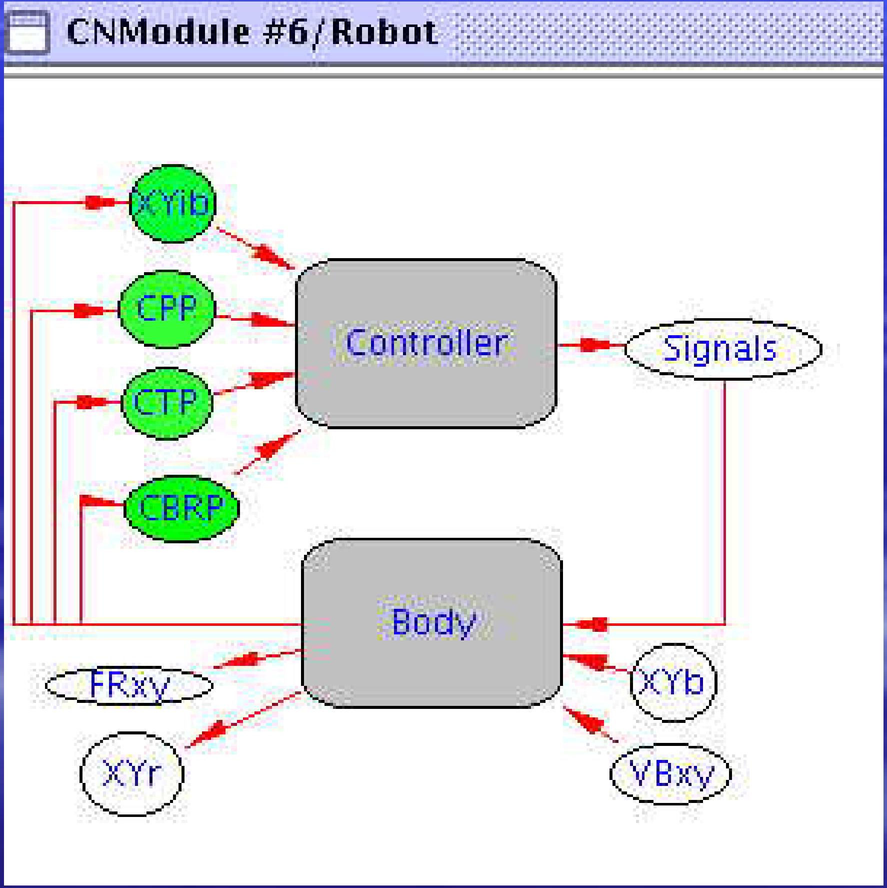


Constraint Nets: theory, tools, applications

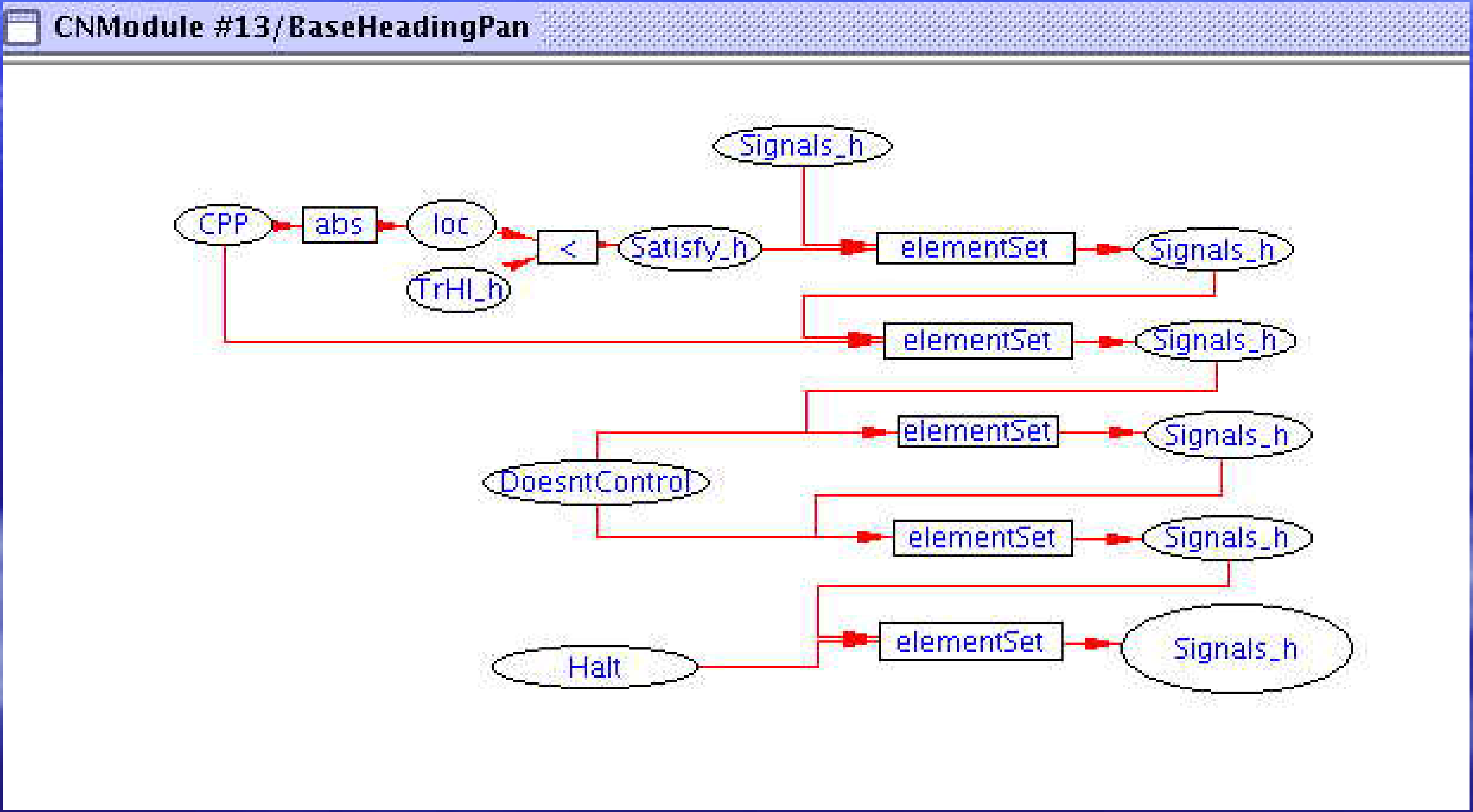
- Robert St-Aubin & Mackworth are designing and building Probabilistic Constraint Nets (PCN) for representing uncertainty in robotic systems.
- Song & Mackworth have designed and implemented CNJ, a visual programming environment for Constraint Nets, implemented in Java.

  The system includes a specification and implementation of CNML, an XML environment for Constraint Nets.
- CNJ is now being used as a tool by Pinar Muyan to develop a constraint-based controller for a simple robot soccer player.

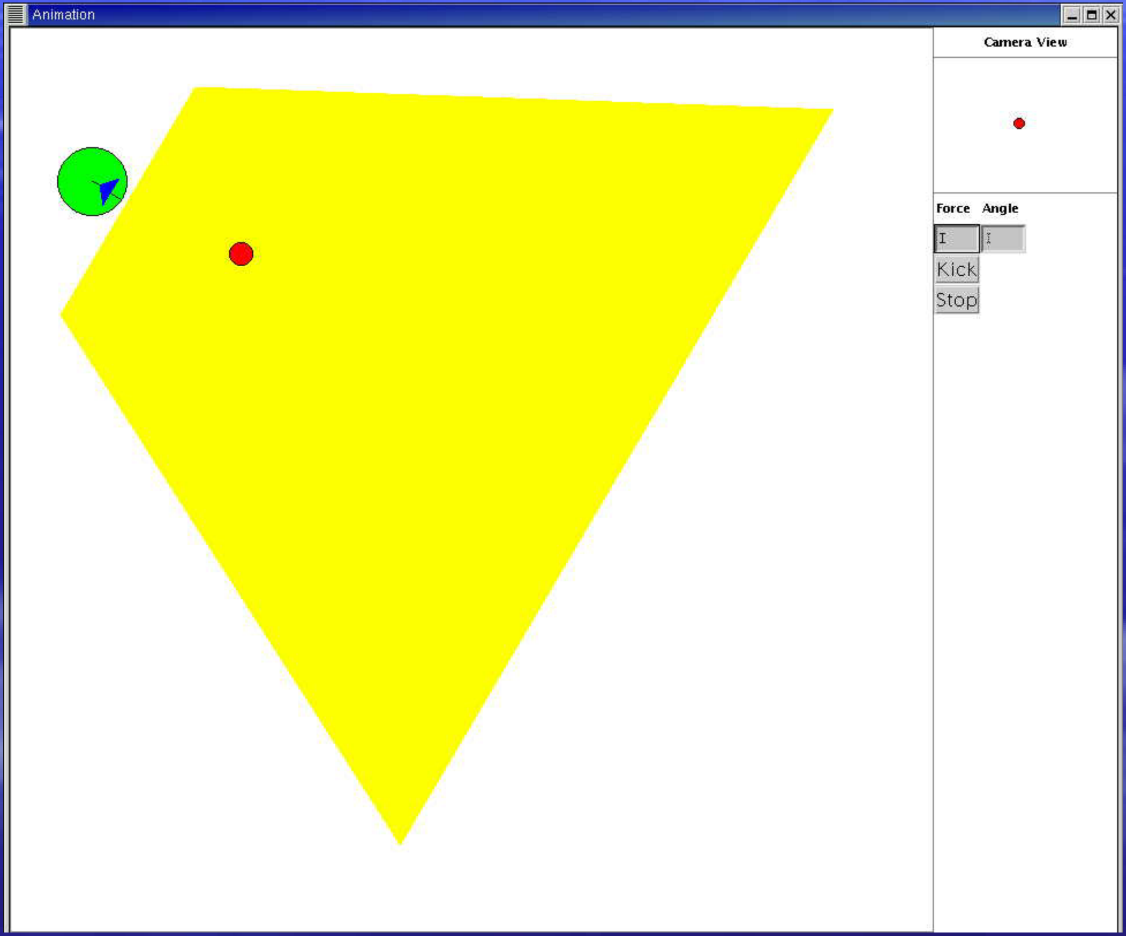
CNJ: Robot in pursuit of ball



CNJ: Controlling Rotation/Pan/Tilt

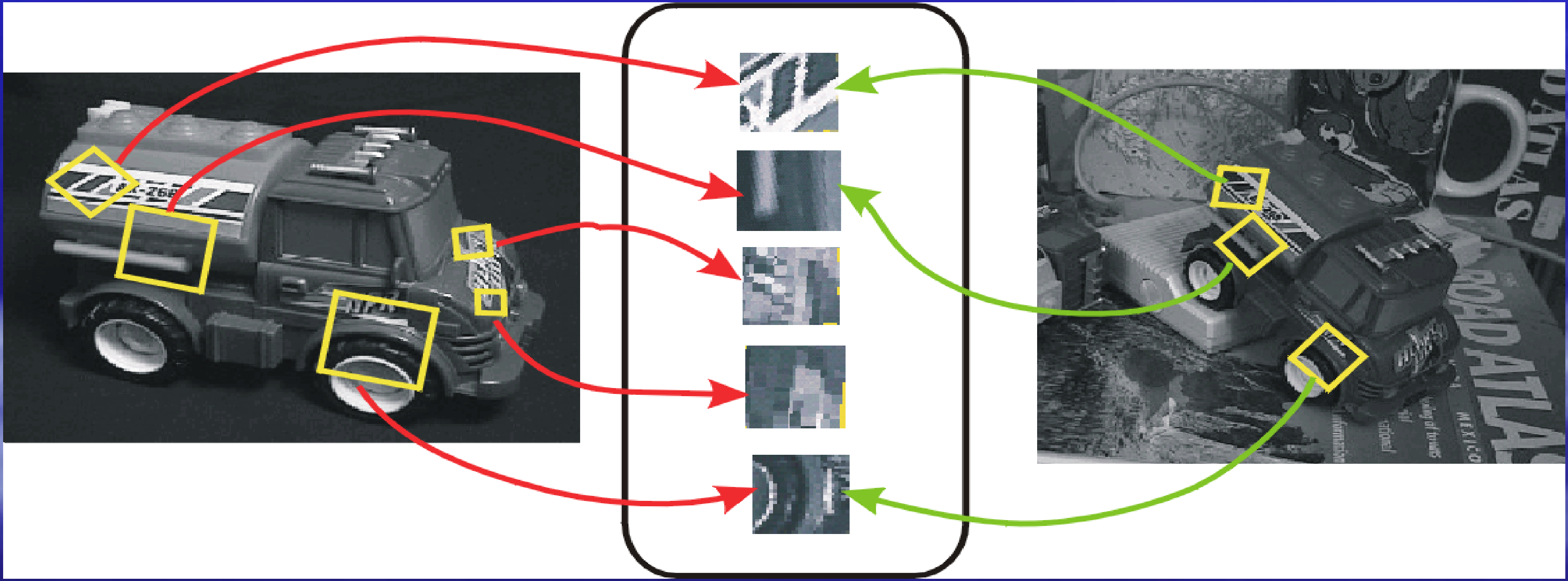


CNJ: Animation of controller



Scale Invariant Feature Transform (SIFT)

- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters

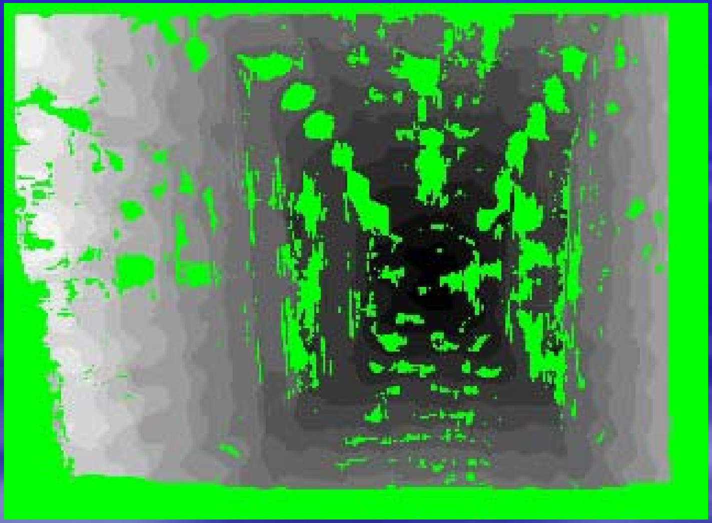


SIFT Features

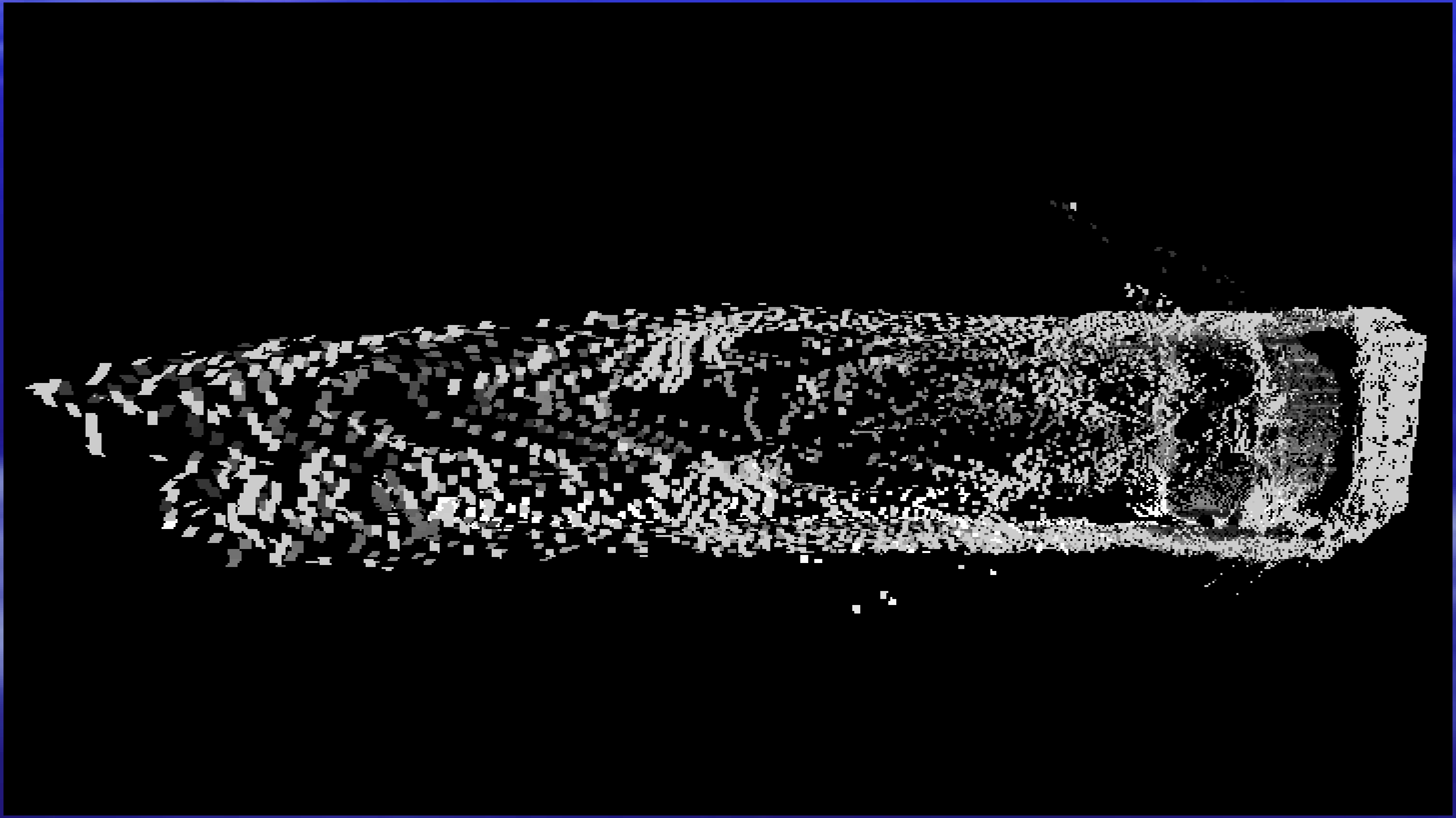
Patchlet Surface Representation

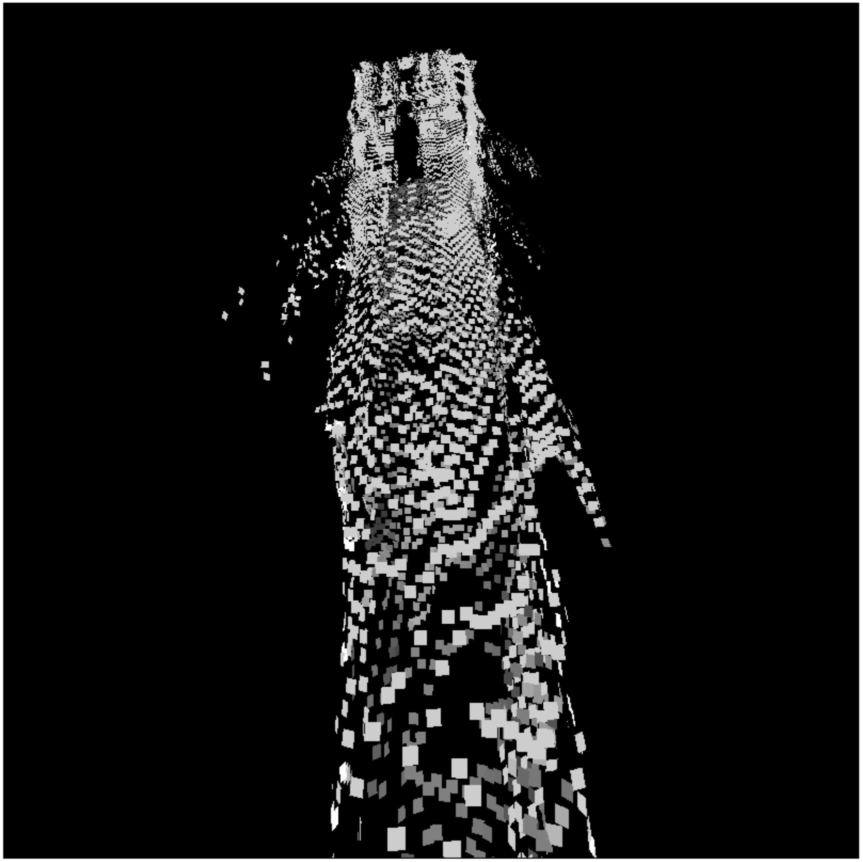
- Goal: to properly interpret the uncertainty of stereo measurements in surface reconstruction.
- The sensor elements considered are local patches in the stereo image that create patchlets.
- These patchlets are fit to a plane and the uncertainty of the plane in orientation and position is determined from the stereo 3d points.

Brightness and Depth

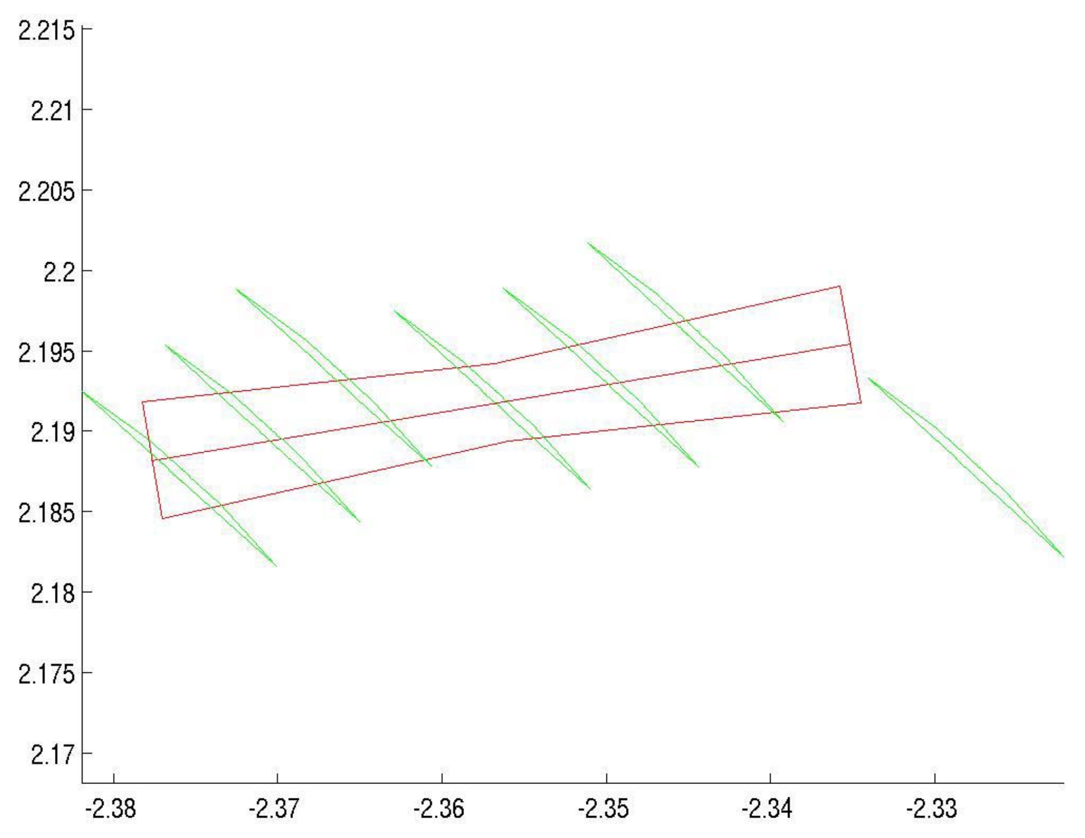


Depth Pixels and Scale

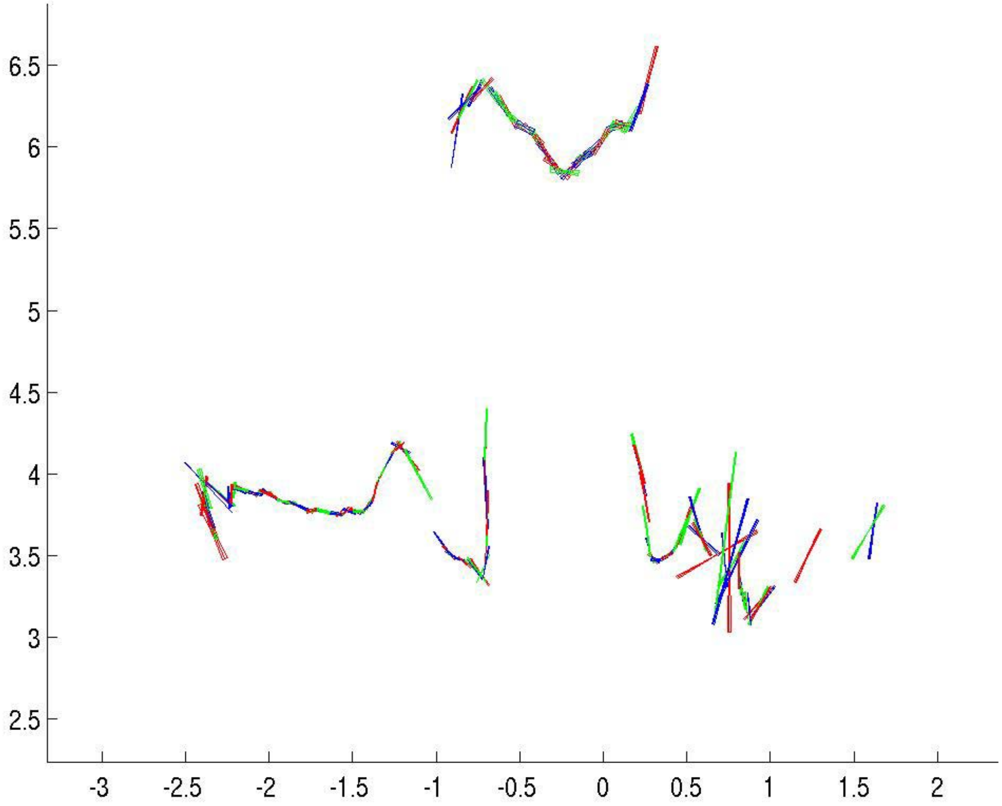




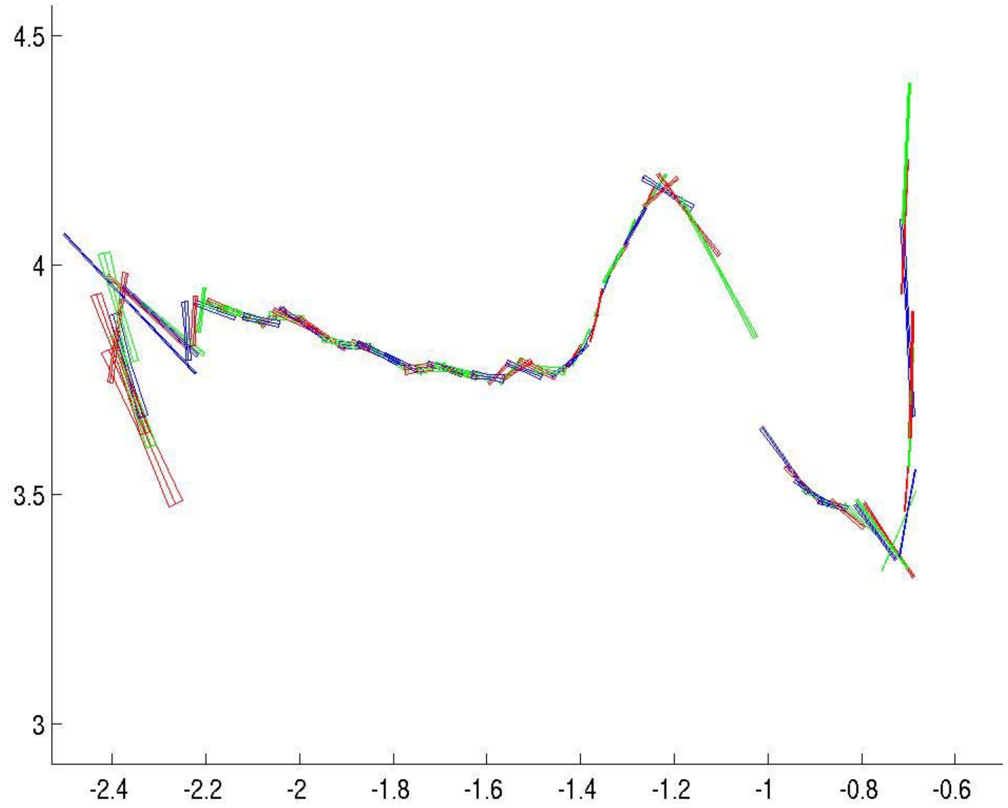
Patchlet Uncertainty



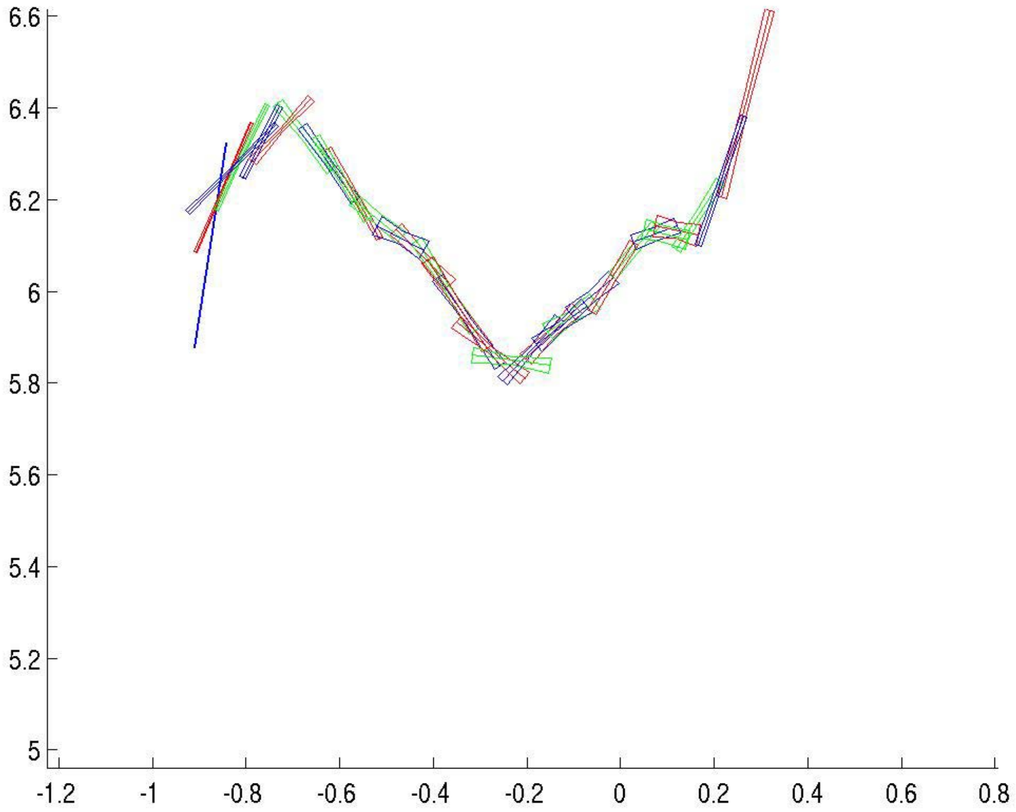
Patchlets



Near



Far



Representation and Recognition of Complex Human Motion

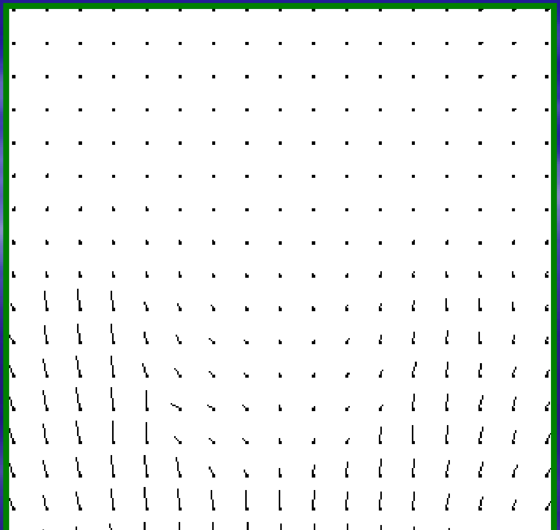
- Psychological research
- Video coding, search
- Human-computer interaction
- Articulated, non-rigid motion at many scales
- Find a general representation for any type of motion at any scale
- Show Zernike polynomials as ideal for this purpose

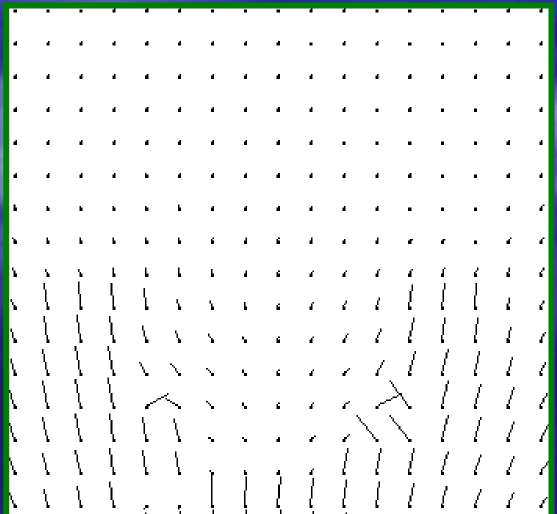
Optical Flow

t=0

t=1

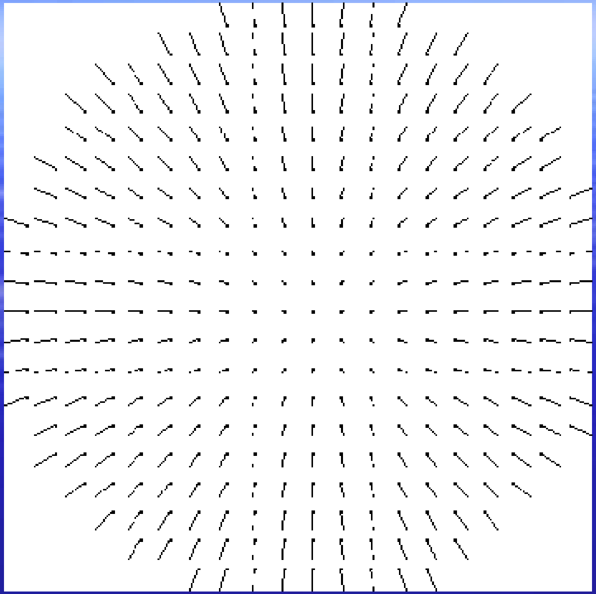
t=2



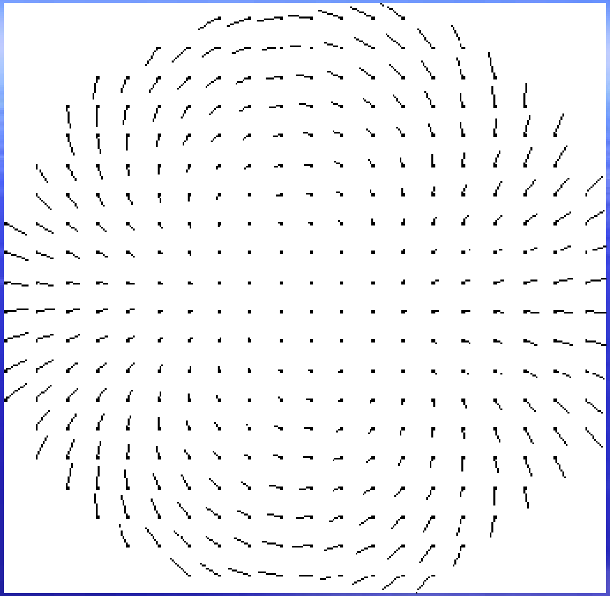


Example Flows

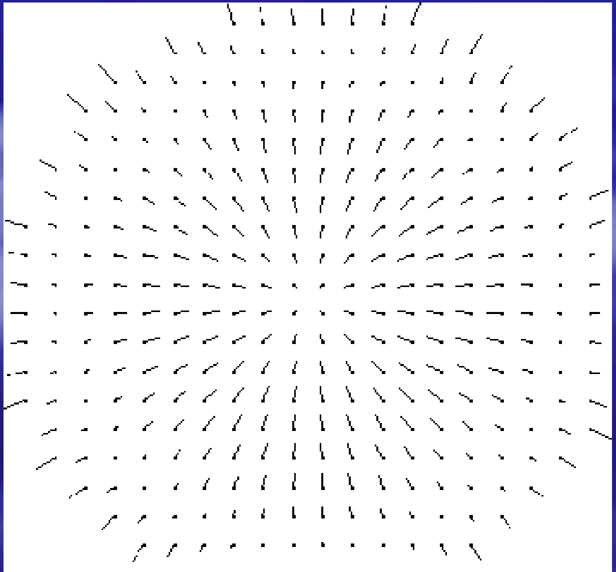
n=1 m=1 (affine)



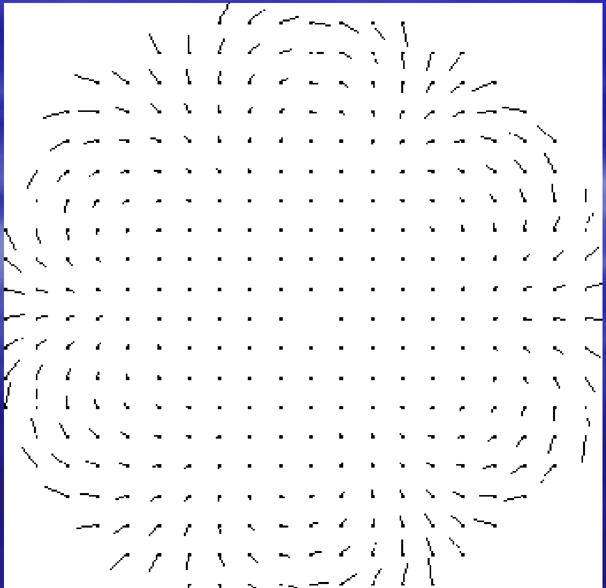
n=2 m=2



n=4 m=0 (radial only)

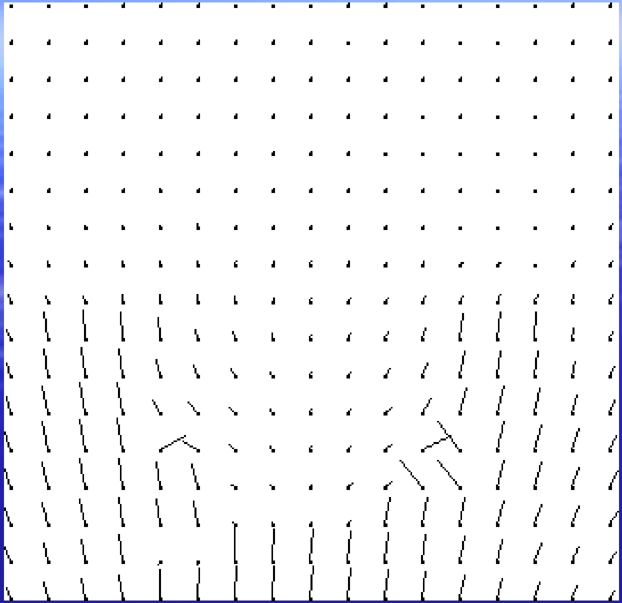


n=4 m=2

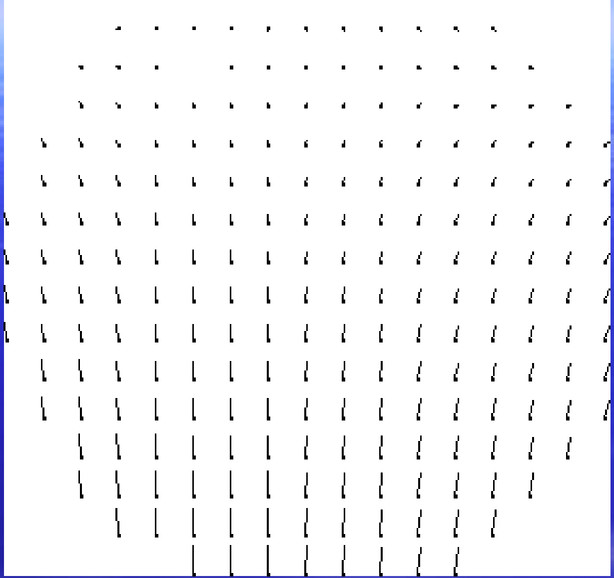


Reconstructed Flows:

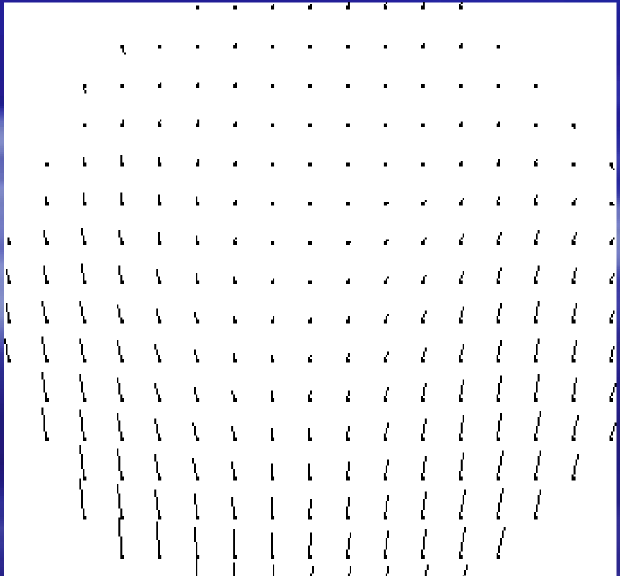
original flow field



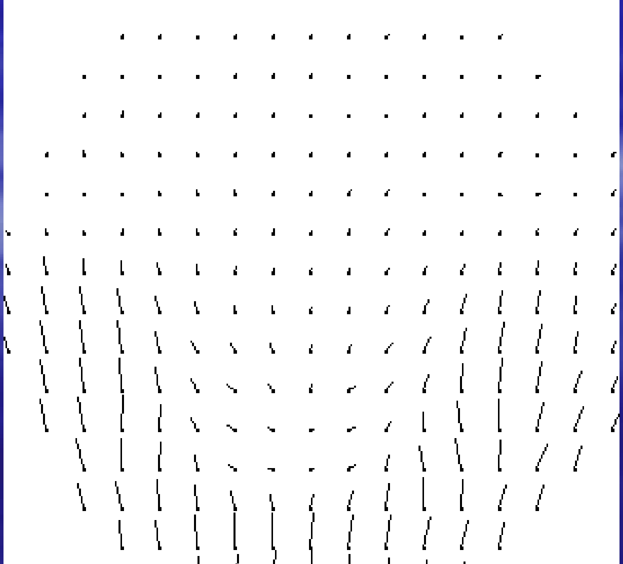
affine flow



first 7 ZPs



first 49 ZPs



Facial expression

-no rigid head motion
-72 subjects - 6 expressions*

*Cohn-Kanade Facial Expression Database

Affine (2 ZPs): **71%**

(267 sequences, 4604 frames)

First 7 ZPs **90%**

Lip-Reading

- Tulips1 database
- 12 subjects - 4 words

Affine (2 ZPs):	**66%**
First 7 ZPs	**76%**
2,4,8,9,10,14,22:	**79%**

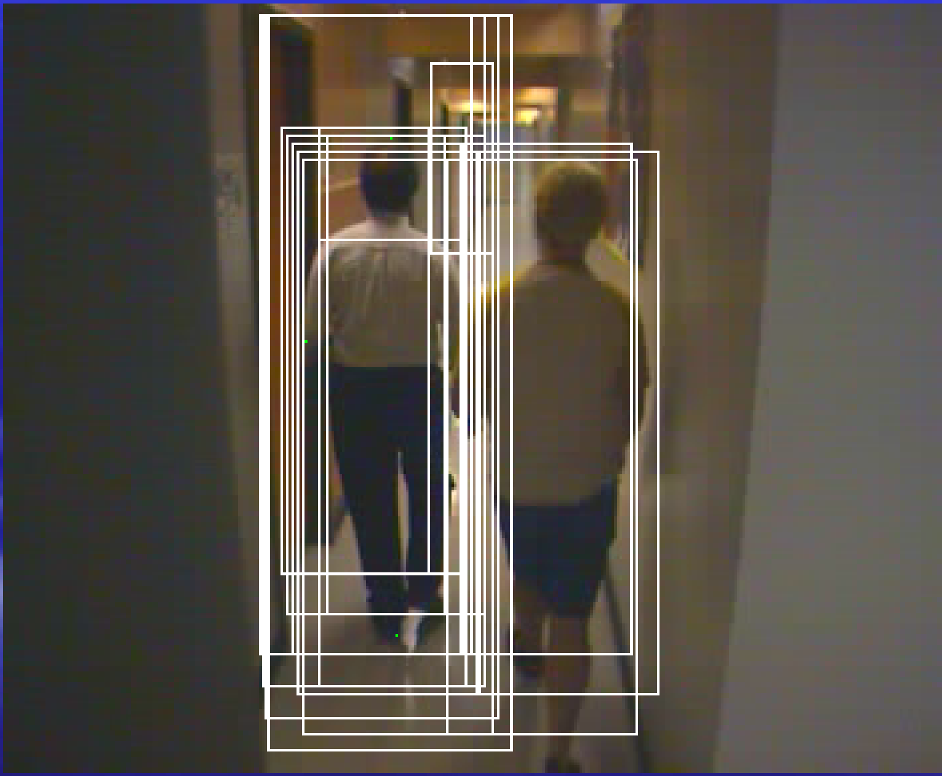
(96 sequences, 835 frames)

Multiple Camera Area Surveillance Techniques

- To distinguish "normal" people, objects, and activities from anomalous ones, and alert a security agent in the case of anomalous conditions.
- Our current projects are:
  - Fast people detection in corridor images, to be used as input for activity recognition.
  - Similarity filter development, for knowing when a given scene has been seen before.
  - Context-based object and scene recognition algorithms.

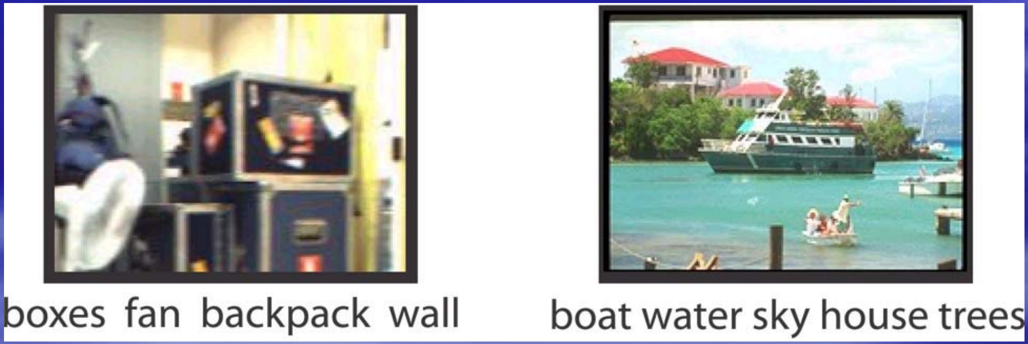
People Detection Algorithm

- uses JPEG encoded images from a network camera. A *large* set of image features is computed, *without* the need for complete JPEG decompression.
- A support-vector machine (SVM) is used to classify image regions into people or non-people regions.
- A related project is developing an FPGA hardware implementation.



Statistical Translation for Object Recognition

- statistical model for learning the probability that a word is associated with an object in a scene.
- learn these relationships without access to the correct associations between objects and words.



- a Bayesian scheme for automatic weighting of features (e.g., colour, texture, position) improves accuracy by preventing overfitting on irrelevant features.

Contextual Translation

- Poor assumption: all the objects are independent in a scene.
- Our more expressive model takes context into account.
- Use loopy belief propagation to learn the model parameters. –
- On our Corel data set (www.cs.ubc.ca/~pcarbo), we achieve almost 50% precision.

Object Recognition for Robots

- Our scheme is not real-time because of the expensive segmentation step.
- BUT: our contextual translation model + a fast, crude segmentation results in equal or better precision! Moreover, object recognition is more precise because the segments tend to be smaller.

FIN