Biologically inspired Vision on a Modular Reconfigurable System (BMV)

Jon Binney      RESL
Lior Elazary     ILAB
Nadeesha Ranasinghe PRL
Overview of Presentation

- BMV
  - Combination of research ideas from our labs

- Search and Rescue (S&R) Scenario
  - Locate injured people
  - Identify hazards for the safety of the rescue crews
  - Obstacle avoidance and exploration

- Modular Reconfigurable Robot

- Stereo Vision & Structure From Motion

- Saliency Based Identification & Tracking
Robots & Vision

- Why should we use robots for S&R?
  - Dangers present in a disaster area
    - Collapsing structures
    - Elemental hazards (fire, water, electricity, gas)
    - Minimize risking lives
  - Cheaper
  - Faster
  - Tolerant to Elemental hazards to some extent

- Why should we use Vision?
  - Very powerful sensor for high-level task-based sensing
  - More noise tolerant than most other sensors
Modular Reconfigurable Robot (SuperBot)

- Standard Module’s Capabilities
  - 3 DOF (x - yaw, y - pitch, z - yaw)
  - IR Communication & Proximity Sensing
  - 3D Accelerometer
  - One-way Radio Communications
  - 6 Docks for Reconfiguration

- Additional WiFi Communications & Wireless Camera Module
Go Anywhere with Shapes & Gaits

- **Reconfigurable Shape**
  - Changing the global shape to maneuver through various obstacles
  - Reposition cameras
  - Track, Snake, Spiral, Biped, Quadruped, Hexapod etc.

- **Gaits**
  - Ways of moving
  - Restrictions imposed by Shape
  - Rolling, Sidewinder, Caterpillar, Walkers, Climbers etc.
What will the robot see?
Stereo Vision & Structure From Motion

- Calibration
- Dense Stereo
- Structure from Motion
- Fitting a Model to Data
Calibration

- Each camera creates images which are warped in slightly different ways.

- Calibration uses a known target to calculate these parameters.

Dense Stereo

- **Goal:** Find the 3D depth of each pixel
- **Input:** A left and a right image
Steps in Dense Stereo

- 1. For each pixel in the left image, find the corresponding pixel in the right image (disparity)
- 2. Use knowledge of the relative positions of the two cameras to triangulate the 3D position of the point
Use of Dense Stereo for Robotics

- Provides a TON of information about the depth of 3D points in front of the robot, which can be used for obstacle avoidance

- Can be combined with a SLAM technique to provide estimates of robot position over time


Structure from Motion

- Goal: Use multiple images to build a 3D model of the environment (and possibly calibrate the camera at the same time)

- How is this different from stereo reconstruction?
SfM for Robotics

- SfM works naturally with robotics because the robot gets a sequence of images as it moves through its environment.
- Many SfM techniques have the advantage of not needing calibrated cameras.
- Solves for the positions of the cameras (and robot) as it solves for the structure of the environment.

Fitting a Model to Data

- Dense Stereo and Structure from Motion (SfM) result in a set of points in 3D. How do we turn this into a more useful model of the environment?
Fitting a Model to Data

- Assume some basic structure for the environment
- Assume some error distribution for points
- Find the most 'Likely' model using a technique like EM


Saliency Based Identification & Tracking

- Bottom-Up Saliency
- Top-Down Saliency
- Navigation
Saliency Based Identification & Tracking
Saliency Based Identification & Tracking
Visual Search

- Free examination
- Estimate material circumstances of family
- Give ages of the people
- Surmise what family has been doing before arrival of “unexpected visitor”
- Remember clothes worn by the people
- Remember position of people and objects
- Estimate how long the “unexpected visitor” has been away from family
Attention

- Given an input image, predict which location in the image will automatically attract your attention.
- Vision is expensive and ambiguous.
  - Requires a large amount of processing to compute high-order information, i.e. object recognition.
  - Too much information at one time can hinder the system. Categorizing two objects at one time as appose to one at a time.
- Don’t want to search the whole image.
  - Saliency gives clues of where the object of interest would be.
- Need to find likely interest positions based on simple features.
Natural scenes
Many applications

- Including...

- Video compression
- Automatic target detection
- Driver alerting & monitoring
- Surveillance
- Robotics
- Animation of virtual agents
- Analysis of satellite imagery
- Star Wars binoculars

- ... many more.
Alumni Advisors
- Vien Ho (Lockheed Martin Corporation - KAPL, Inc.)
- Khoo Yit Phang (University of Maryland)

Officers

President
- Zack Gussman

Software Lead
- Kevin Jones
- Randolph Voorhies

Hardware Lead
- Asim Bhalera
- Brian Inacio

Treasurer
- To be determined

Public Relations Representative
- Andre Rosa

Webmaster
- Elliot Lee

Continuing Members
- Nicholas Kiswanto
- Andrew Chambers
- Han-Nung Lin
- Charmy Chhichhia
- Harold Lim

New Members
- Mohd Yusuf Abdul Hamid
- Cheng-Yuan Wu
- David Li
- Randolph Voorhies
- Seung Gun Cheong
- Chun-Yi Leu
- Brent Deets
- Yu-Chi Hsiao
- Alex Stram
- Nidhi Jain
- Harmeet Bhutani
- Christina Carter-Brown
- Ronakumar Patel
- Don Hainesworth
- Brett Kelford
- Donko Krajencik
- Andrew Hart
- Robert Alland
- Taiko Kinoshita
- Tao Ma

Former Members
- Tasha Drew
- Harris Chu
- Syamsul Hasan
Example - Beobot
Top Down Attention
Where is Waldo?

- Knowing the target in a visual space leads to faster search (Vickery et al. 2005, Wolfe 1994)
- What features from the object do we learn for biasing.
- How can biases be applied in the most efficient manner.
Selecting Features Using Saliency

- Salient location within an object would remain the same under various transformations.
- Get the raw center-surround features from the most salient location.
  - Choose the most salient location within each submap.
- Using Bayesian decision theory to decide object classification (Richard et al. 2001)
- Bias image using learned likelihood function.
Results: Search Task for houses
Results: Search Task for houses
Results: Search Task for houses and roads
Landmark Navigation Using Biased Attention

- Toy jeep fitted with a wireless (1.2GHz) camera and a standard RC remote control.
- The camera receiver was connected to a capture card.
- The RC remote control was connected to a device (sc8000) which allowed the computer to control the robot.
Landmark Navigation Using Biased Attention

- **Left Image**
  - Yellow box and dot represent the landmark found using SIFT.
  - Blue dot represents the current tracking location using biased attention.

- **Right Image** is the resulting saliency map.
  - Saliency map is reversed.

- **Left bottom text (from left to right)**.
  - Image capture rate frame per second.
  - Biased saliency map frame per second (tracking).
  - Current landmark id the robot is navigating toward.
    - Can be thought of as the current leg in the path.
Landmark Navigation Using Biased Attention
Conclusion

- Modular reconfigurable robots in S&R
- Saliency to identify possible targets
- Reconstruct the structure around the target