

# **Aveiro's Team in SRVC'2009 – software league**

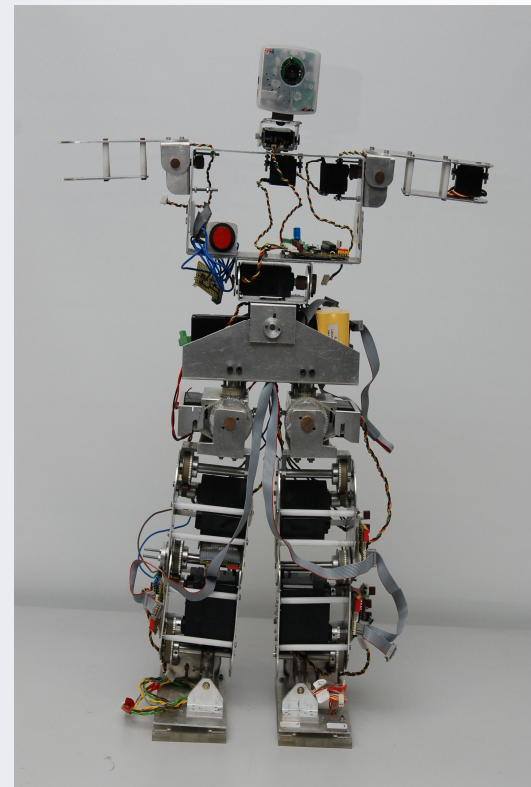
Mário Antunes, Luís Seabra Lopes

Transverse Activity on Intelligent Robotics  
IEETA/DETI - Universidade de Aveiro, Portugal

# Transverse activity of on intelligent robotics

- The mission of the Transverse Activity on Intelligent Robotics (ATRI) is to study integrated architectures for intelligent robotic systems
  - The group builds on synergies between the main research areas of our institute
  - Created in 1999, acknowledging the strategic importance of developing this field of research in the institute
  - ATRI is involved in EURON – European Robotics Research Network and its interest groups

# Interest in human-robot interaction



# Interest in semantic vision for human-robot interaction

- Contribute to the development of robots that are easy to command and instruct by naïve users
- Spoken language dialogs with robots
- Integration with basic navigation, including guided navigation
- Knowledge acquisition and question answering
- Symbol grounding and language acquisition
- Visual object learning and recognition

# Scenario in human-robot interaction

- Our robot, Carl – “Hi, how may I help you?”
- A user – “Hi Carl, please bring me the Spam in the kitchen table”
  - [ but there are several objects in the table ... ]
- Carl – “Hum ... ”
  - [ Carl searches the web to find out what Spam is ]
- Carl – “Ok, I’ll pick it up!”

# Interest in semantic vision for robotic soccer

- According to the latest rules for the Middle-Size League of RoboCup, an arbitrary soccer ball can be used
  - Not necessarily orange!



# Aveiro's team in SRVC

- Luís Seabra Lopes, faculty, team leader
- Mário Antunes, student (2009)
- Frederico Valente, student (2008)
- Rui Pereira, student (2008)
- Luís Ribeiro, student (2008)

# Overview of entry

- Implemented in C/C++ and scripting languages
- A set of independent programs that store their results in the file system
  - Good for robustness of the overall system

# General versus specific categories

- General categories tend to be more heterogeneous and more dependent on shape analysis
- Specific entities tend to be more homogeneous and provide enough local features for modelling
  - particularly when they are branded and/or have elaborate graphic design

# Global versus local descriptors

- General and specific categories are handled differently, according to the following heuristic
- Instances of specific categories are represented with SIFT features
  - The name of the category contains capital letters and/or quotes
- Instances of general categories are represented by polar descriptors (“global shape context”)

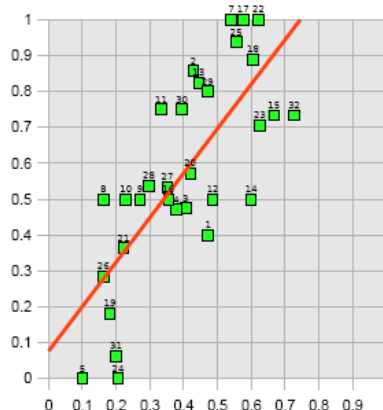
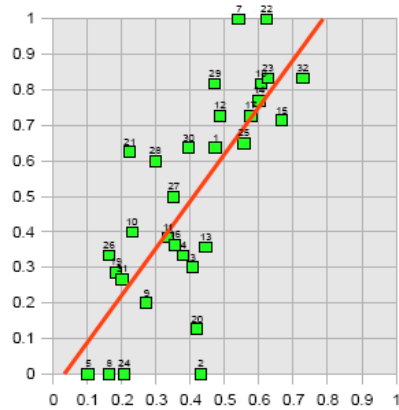
# Web search and model building

- Search is done via Google only using English language only
- Unsupervised subset selection
- Category models are simply collections of instance descriptors

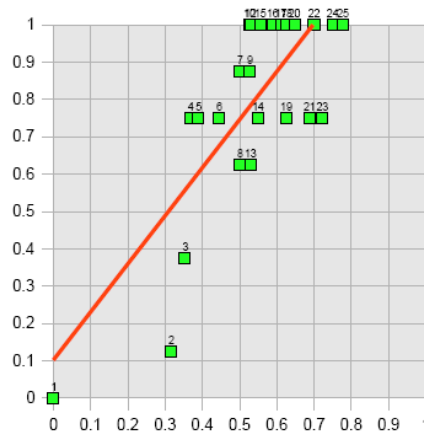
# Unsupervised subset selection

- K-means clustering of fetched images repeated several times
- Relevance of fetched images rated by the number of times they occur in the largest cluster
- Number of instances included in the model depends on an evaluation of category heterogeneity
  - More instances are stored for more heterogeneous categories

# Unsupervised subset selection



- 1-bicycle helmet;
- 2-spoon;
- 3-trashcan;
- 4-pear;
- 5-orange;
- 6-scientific calculator;
- 7-fork;
- 8-banana;
- 9-desk;
- 10-table;
- 11-eyeglasses;
- 12-staple;
- 13-coffee cup;
- 14-whiteboard;
- 15-compact disc;
- 16-ice skates;
- 17-telephone;
- 18-calendar;
- 19-hammer;
- 20-clothes hanger;
- 21-electric iron;
- 22-green apple;
- 23-red bell pepper;
- 24-rolling suitcase;
- 25-red plastic cup;
- 26-blue dry eraser;
- 27-tape dispenser;
- 28-computer mouse;
- 29-chair;
- 30-nokia cellphone;
- 31-baseball hat.



- 1-Coca-Cola can;
- 2-Apple iPhone;
- 3-Butterfinger candy bar;
- 4-Dvd "300";
- 5-CD "Final Straw" by Snow Patrol;
- 6-Cadbury Caramilk;
- 7-Twinings Earl Grey Tea;
- 8-Nintendo Wii box;
- 9-CD "Look-Alikes Jr." by Joan Steiner;
- 10-Ritter Sport Dark Chocolate;
- 11-Aerius Allergy;
- 12-CD "Introducing Joss Stone";
- 13-DVD "Madagascar";
- 14-Colgate Total;
- 15-Dasani water;
- 16-book "Artificial Intelligence: A Modern Approach";
- 17-The Economist magazine;
- 18-cd "Hey Eugene";
- 19-Kellogg's Corn Flakes;
- 20-Pringles Can;
- 21-Fanta Orange Can;
- 22-Sony AIBO;
- 23-Ryobi drill;
- 24-book "Probabilistic Robotics";
- 25-Crayola Crayons box

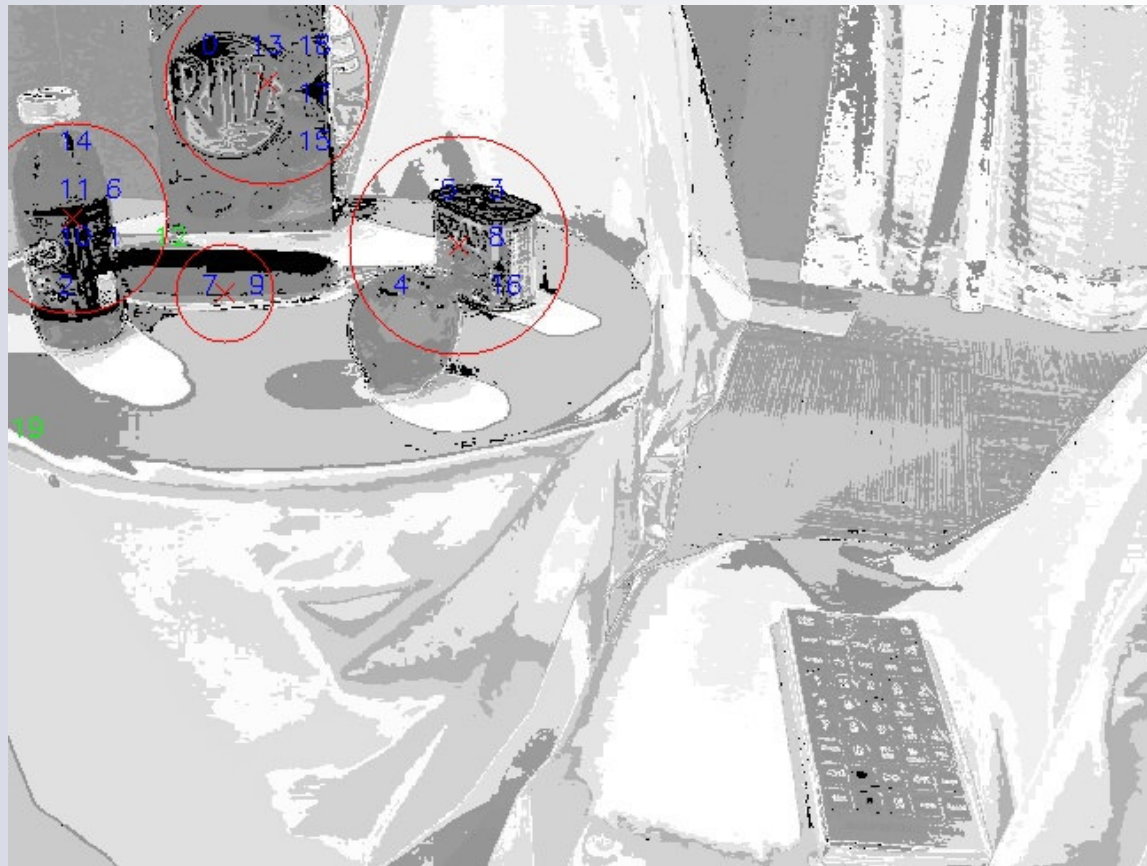
# Performance phase

- For each image in the workset ..
  - Color-based detection of salient points
  - Regions of interest formed through clustering of salient points
  - Edge detection, object segmentation and clutter removal
  - Nearest-neighbor (or maximum similarity) evaluation of category membership
  - A voting scheme was also implemented, but not used in the competition

# Color-based saliency detection

- Color histogram of the target image
- Saliency image is obtained by rating each pixel of the original image proportionally to the relative frequency of the respective color
- Sparse analysis of color saliencies of the image, providing a set of key points
- Unsupervised clustering of keypoints
- It works quite well!

# Color-based saliency detection



# Publications

- Pereira, R., and Seabra Lopes, L. Learning visual object categories with global descriptors and local features. *Progress in Artificial Intelligence, LNAI 5816*, Springer (2009).
- Pereira, R., Seabra Lopes, L., and Silva, A. Semantic image search and subset selection for classifier training in object recognition. *Progress in Artificial Intelligence, LNAI 5816*, Springer (2009).

# Results in SRVC'2009

- Objects found
  - book "I am a Strange Loop" by Douglas Hofstadter
  - DVD "Hitchhiker's Guide to the Galaxy" widescreen
  - Peperidge Farm Goldfish Baked Snack Crackers
- Time efficiency
  - 13 minutes for internet search + object modelling
  - 12 minutes for environment search

# Final comments

- Learning phase:
  - Subset selection works well
    - for initial sets with 50% of good images, the final percentage of good images varies between 63% and 75%
- Performance phase
  - Object detection provides a good basis
  - Classification system requires additional attention