Bootstrapping Visual Memories for a Humanoid Robot

Giovanni Saponaro, Alexandre Bernardino
VisLab, ISR-IST, Lisbon, Portugal

Computer Vision Laboratory

We present a cognitive memory approach based on visual saliency filters. The system creates concepts by grouping related images together in real time and starting from zero prior knowledge.

1. Motivation

Developmental robotics systems start with little or no information about the surrounding scene. The cognitive architecture of the iCub humanoid robot (53 degrees of freedom) aims at understanding the environment with an increasing degree of precision.

Within this framework, the Auto-Associative Memory component works as follows:

(i) when an image is presented to the memory, it attempts to recall that image with vision algorithms;
(ii) if a previously-stored image matches the presented image sufficiently well, then the stored image is recalled and displayed;
(iii) if no previously-stored image matches sufficiently well, then the presented image is stored in the database.

2. Approach

With regard to the matching criterion of the Auto-Associative Memory, we use a simple visual filter that assigns saliency according to Histogram Intersection in the Hue-Saturation colour space:

$$H(I, M) = \frac{\sum_{j=1}^{n} \min(I_j, M_j)}{\sum_{j=1}^{n} M_j} \quad [4]$$

where I is the input image histogram, M is the candidate image model histogram, both have n bins, the numerator is defined as the number of pixels of the same colour in the image, and the denominator is a normalization factor.

Two aspects are studied:
1. *which saliency criteria to use for selecting the candidate images to store in the visual memory;*
2. *how a single threshold on the categorization distance can be tuned to create hierarchical class representations.*

Computational cost is linear in the number of elements of the histograms, which have 16x16 bins computed in real time while receiving images from the iCub cameras.

3. Results

A low threshold parameter implies generalized object preference, while high values entail a specialized memory.

This intuition is verified while testing the algorithm in an empty room, rotating the robot head so to explore all of the environment:

threshold	≤ 0.4	0.5	0.6	0.7	0.8	0.9	0.95
# images	1	2	3	5	8	9	40

For the visual memory system to create sensible categories, the threshold parameter choice is critical, especially in initialization. We display here two classes (plants and faces) that were formed during an experiment with threshold = 0.5.

Future work includes: test visual saliency filters [3] *other than colour histogram intersection, use them concurrently with appropriate weights; observe and control the branching factor during an exploration task that iteratively classifies a scene with increasing thresholds.*

The robot's memory groups related images together into meaningful clusters.

[1] D. Figueira *et al*. *From Pixels to Objects: Enabling a Spatial Model for Humanoid Social Robots*. Proc. of ICRA 2009, Kobe, Japan, May 2009.
[2] M. Lungarella *et al*. *Developmental Robotics: A Survey*. Connection Science, 15(4):151–190, Dec. 2003.
[3] J. Ruesch *et al*. *Multimodal Saliency-Based Bottom-Up Attention*. Proc. of ICRA 2008, Pasadena, CA, USA, May 2008.
[4] M. H. Swain and D. H. Ballard. *Color Indexing*. International Journal of Computer Vision, 7(1):11–32, Nov. 1991.
[5] D. Vernon. RobotCub Deliverable D2.1 – *A Roadmap for the Development of Cognitive Capabilities in Humanoid Robots*, Oct. 2008.

This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST pluriannual funding).