# Learning to Select Object Recognition Methods for Autonomous Mobile Robots

**Reinaldo A. C. Bianchi** [1,2] and **Arnau Ramisa**[2] and **Ramón López de Mántaras** [2]

**Abstract.** Selecting which algorithms should be used by a mobile robot computer vision system is a decision that is usually made *a priori* by the system developer, based on past experience and intuition, not systematically taking into account information that can be found in the images and in the visual process itself to learn which algorithm should be used, in execution time. This paper presents a method that uses Reinforcement Learning to decide which algorithm should be used to recognize objects seen by a mobile robot in an indoor environment, based on simple attributes extracted on-line from the images, such as mean intensity and intensity deviation. Two state-of-the-art object recognition algorithms can be selected: the constellation method proposed by Lowe together with its interest point detector and descriptor, the Scale-Invariant Feature Transform and a bag of features approach. A set of empirical evaluations was conducted using a household mobile robots image database, and results obtained shows that the approach adopted here is very promising.

## 1 INTRODUCTION

Reinforcement Learning [7] is concerned with the problem of learning from interaction to achieve a goal, for example, an autonomous agent interacting with its environment via perception and action. On each interaction step the agent senses the current state $s$ of the environment, and chooses an action $a$ to perform. The action $a$ alters the state $s$ of the environment, and a scalar reinforcement signal $r$ (a reward or penalty) is provided to the agent to indicate the desirability of the resulting state. The policy $\pi$ is some function that tells the agent which actions should be chosen, and is learned through trial-and-error interactions of the agent with its environment. Several algorithms were proposed as a strategy to learn an optimal policy $\pi^*$ when the model ($\mathcal{T}$ and $\mathcal{R}$) is not known in advance, for example, the $Q$–learning [8] and the SARSA [6] algorithms.

Some researchers have been using RL as a technique to optimize image segmentation and object recognition algorithms. For example, Peng et al. used RL to learn, from input images, to adapt the image segmentation parameters of a specific algorithm to the changing environmental conditions, in a closed-loop manner [1, 5] and Draper et al. modeled the object recognition problem as a Markov Decision Problem, and proposed a method to learn sequences of image processing operators for detecting houses in aerial images [2].

To allow a robotic agent to decide which object recognition method should be used, during on line world exploration, we propose to use RL to learn a policy that minimizes computing time, discarding an image if it is not suitable for analysis or choosing between two well known algorithms, described in the following section.

[1] Centro Universitário da FEI, São Bernardo do Campo, Brazil.
[2] Artificial Intelligence Research Institute (IIIA-CSIC), Bellaterra, Spain.

## 2 TWO OBJECT RECOGNITION METHODS

Two successful general object recognition approaches that have been widely used are the constellation method proposed by Lowe together with its interest point detector and descriptor SIFT [3] and a bag of features approach [4].

The first approach is a single view object detection and recognition system with some interesting characteristics for mobile robots, most significant of which are the ability to detect and recognize objects at the same time in an unsegmented image and the use of an algorithm for approximate fast matching. In this approach, individual descriptors of the features detected in a test image are initially matched to the ones stored in the object database using the Euclidean distance. False matches are rejected if the distance of the first nearest neighbor is not distinctive enough when compared with that of the second. Once a set of matches is found, the generalized Hough transform and Iteratively Reweighted Least Squares are used to cluster each match and to estimate the most probable affine transformation for every hypothesis.

The Bag of Features (BoF) approach to object classification comes from the text categorization domain, where the occurrence of certain words in documents is recorded and used to train classifiers that can later recognize the subject of new texts. This technique has been adapted to visual object classification substituting the words with local descriptors such as SIFT. The descriptor space is discretized in a codebook created applying hierarchical k-means to a dataset of descriptors. A histogram of descriptor occurrences is built to characterize an image. Next, a multi-class classifier – the k-NN in this implementation – is trained with the histograms of local descriptor counts. The class of the object in the image is determined as the dominant one in the k nearest neighbors.

Although both object recognition methods proved their reliability in real world applications, they have their limitations: Lowe's method performs poorly when recognizing sparsely textured objects or objects with repetitive textures, while the Bag of Features needs an accurate segmentation stage prior to classification, which can be very time consuming. Furthermore, the method depends on the quality of that segmentation stage to provide good results.

## 3 EXPERIMENTS AND RESULTS

In order to decide which algorithm should be used by the agent, the RL problem was defined as a 2 stage MDP, with 2 possible actions in each stage: In the first one, the agent must decide if the image contains an object, and thus must be recognized, or if the image does not contain objects, and can be discarded, saving processing time. In the second stage, the agent must decide which object recognition algorithms should be used: Lowe's or Bag of Features.


**Figure 1.** Images from the dataset.

At each stage the agent choses a system state $s$, composed of the stage the agent is at plus a combination of simple attributes extracted on-line from the images, for example, mean image intensity and standard deviation. Then, it selects an action to be executed, compute the reward and update the value function.

The RL algorithm used is the Q-learning [8], because it directly approximates the optimal policy independently of the policy being followed (it is an off-policy method), allowing the state and the action to be executed by the agent to be selected randomly.

The rewards used during the learning phase are computed using a set of training images. If the state in which the agent is corresponds to a training image, and the action taken results in a correct classification, the agent receives a reward. Otherwise it is zero. For example, if we have a training image that does not contain an object, with mean intensity value of 50, standard deviation of 10, the reward given to the state Q(stage = 1, mean = 50, std = 10, action = discard) is 100.

Several experiments were executed using a dataset consisting of approximately 150 images of objects occurring in typical household environments plus 30 background images. The objects, that can be textured, untextured or with repetitive textures, are mugs, books, trashcans, chairs and computer monitors (Figure 1). The images includes occlusions, illumination changes, blur and other typical nuisances that can be encountered while navigating with a mobile robot.

To evaluate the result of the learning process statistical validation method called Leave-One-Out was used. Six different experiments were conducted, using three different combinations of image attributes as space state and two different image sizes (the original size and a 10 by 10 pixels reduced size image). The combinations of image attributes used as space state are: mean and standard deviation of the image intensity (MS); mean and standard deviation of the image intensity plus entropy of the image (MSE); and mean and standard deviation of the image intensity plus the number of interest points detected by the Difference of Gaussians operator (MSI). The parameters used in the experiments were: the learning rate $\alpha = 0.1$ and the discount factor $\gamma = 0.9$. Values in the Q table were randomly initiated.

Tables 1 and 2 present the results obtained. The first line of Table 1 shows the percentage times that the agent correctly choose to discard a background image, and the second line shows the percentage of times the agent correctly choose to use the Lowe algorithm, instead of the BoF. The columns in this table presents the results for the six experiments, the first three using the original image and, from the fourth to sixth column, showing the results for the reduced size image. The last column shows the percentage of times a human expert takes the correct action. Table 2 is similar to Table 1, but shows the classification error. The first line shows the percentage of images discarded as background, when they should be analyzed, and line two presents the number of times the Lowe algorithm is chosen, when the correct one is the BoF.

**Table 1.** Correctly classified images (percentage).

| | Full Img | | | Small Img | | | Expert |
|---|---|---|---|---|---|---|---|
| | MS | MSE | MSI | MS | MSE | MSI | |
| Back | 80.4 | 100.0 | 100.0 | 82.6 | 100.0 | 100.0 | 100.0 |
| Lowe | 52.3 | 93.2 | 22.7 | 63.6 | 93.2 | 11.4 | 93.2 |

**Table 2.** Incorrect classification (percentage).

| | Full Img | | | Small Img | | | Expert |
|---|---|---|---|---|---|---|---|
| | MS | MSE | MSI | MS | MSE | MSI | |
| Back | 4.8 | 0.0 | 1.4 | 3.4 | 0.7 | 1.4 | 8.2 |
| Lowe | 25.5 | 0.0 | 6.9 | 18.6 | 0.0 | 6.9 | 10.8 |

These results shows that the use of the MSE combination presented very good results, for original size images as well as reduced size ones. On the other hand, the use of the number of interest points detected by the Difference of Gaussians operator as space state did not produce good results.

## 4 CONCLUSION

The results obtained shows that the use Reinforcement Learning to decide which algorithm should be used to recognize objects yields good results, performing better than a human expert in some cases. To the best of our knowledge, there is no similar approach using automatic selection of algorithms for object recognition.

Future works includes testing other image attributes that can be used as the system's state, other RL algorithms and applying RL techniques to the image segmentation problem.

## ACKNOWLEDGEMENTS

This work has been partially funded by the FI grant and the BE grant from the AGAUR, the 2005-SGR-00093 project, supported by the Generalitat de Catalunya, the MID-CBR project grant TIN 2006-15140-C03-01 and FEDER funds. Reinaldo Bianchi is supported by CNPq grant 201591/2007-3.

## REFERENCES

[1] B. Bhanu, Y. Lin, G. Jones, and J. Peng, 'Adaptive target recognition', *Machine Vision and Applications*, **11**(6), 289–299, (2000).

[2] B. A. Draper, J. Bins, and K. Baek, 'ADORE: Adaptive object recognition', in *International Conference on Vision Systems*, pp. 522–537, (1999).

[3] D. Lowe, 'Distinctive image features from scale-invariant keypoints', *International Journal of Computer Vision*, **60**(2), 91–110, (2004).

[4] D. Nister and H. Stewenius, 'Scalable recognition with a vocabulary tree', in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pp. 2161–2168, (2006).

[5] J. Peng and B. Bhanu, 'Closed-loop object recognition using reinforcement learning', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(2), 139–154, (1998).

[6] G. A. Rummery and M. Niranjan, 'On-line Q-learning using connectionist systems', Technical Report CUED/F-INFENG/TR 166, Cambridge University Engineering Department, (1994).

[7] R.S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 1998.

[8] C. J. C. H. Watkins, *Learning from Delayed Rewards*, PhD Thesis, University of Cambridge, 1989.