Adaptive Visual Attention Based Object Recognition

Rebecca Fay, Ulrich Kaufmann, Heiner Markert, Günther Palm
Department of Neural Information Processing
University of Ulm, Germany
{fay, kaufmann, markert, palm}@neuro.informatik.uni-ulm.de

Keywords: Selective attention, tracking, object recognition, hierarchical neural networks, online learning.

1 Overview

Detecting and identifying objects, in particular the recognition of 3D objects, are important capabilities for robots performing non-trivial tasks in real world environments. In order to be able to solve object related problems the robot has to localise objects of interest in complex visual scenes and has to identify or categorise certain task-relevant objects. When performing tasks in complex environments robots are likely to encounter objects they have not seen before and consequently cannot identify. Thus next to localising and recognising objects as well as processing language and planning actions, the ability to learn novel objects during run time is an essential skill for advanced mobile service robots. This enables the robot to incrementally learn new objects shown during run time and thereby increase knowledge of its environment and adapt to new situations. In real world environments which are fairly complex and subject to numerous changes only being capable of coping with previously learnt objects might not be sufficient. Another helpful skill is the ability to track known and unknown objects since changes in the visual scene are very common due to motion of the robot and of possible objects of interest. Moreover, knowledge about the position of an already localised or classified object reduces the necessity of recalculations for every new image.

We present a multi-stage visual object recognition system that localises and identifies objects using an adaptive colour-based visual attention control algorithm and hierarchical neural networks for object recognition. Initially a window of attention is determined by means of low resolution colour and shape information. If an interesting region is found, this part of the image is analysed in more detail. High resolution features such as edges, corners or colour distribution are extracted from the determined window and used in a trained neural network for object classification. The visual scene is scanned for salient regions, where the information is processed in much more detail than in the vicinity.

This two-stage process reflects some properties of human or monkey vision such as guidance of eye-movements by visual attention and high resolution processing in the fovea versus decreasing resolution towards the periphery.

Once an object has been localised it is tracked. This tracking is a two-stage process. First the localised but not yet identified object is tracked without using model knowledge. Once the object is classified model knowledge is employed in order to facilitate the tracking of the object. Moreover the position of the object is memorised. Hence these two different tracking approaches can be interpreted as unconscious tracking and conscious tracking with position memory.

The approach is evaluated in a test scenario where a robot is located in front of a table with different kinds of fruit and other simple objects on it. The robot has to localise and identify these objects as well as to perform a set of object manipulating tasks such as grasping, showing or moving specified objects.

2 Conclusions

The proposed approached has proven functional. Although the networks were trained with only a few samples of the new classes they were able to classify the new class and no considerable deterioration of the classification results of the former classes could be observed.

The experiments conducted showed encouraging results in particular that hierarchical neural networks are suitable for incremental adaption. New objects can be learnt with reasonable classification rates and in adequate time.

The proposed approach enables the robot to deal with varying object categories in addition to the predefined categories. New objects can be learnt rather quickly with satisfactory quality. The quality can even be increased by retraining with additional samples and in an offline phase more sophisticated learning algorithms can be used to further improve the classification quality.

The tracking of the objects allows for advanced object classification even on slower computers because classification is not exerted for every image.