Adaptive Visual Attention Based Object Recognition

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

Keywords: Visual attention, selective attention, tracking, object recognition, hierarchical neural networks, adaptive incremental learning.

1 Introduction

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 hierarchial 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. Afterwards tracking of unidentified as well as identified objects is performed, which can be interpreted as unconscious or conscious tracking respectively.

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 Attention Control and Object Tracking

A adaptive colour blob approach which segments the image on the basis of predefined colour ranges and evaluates the resulting binary images according to certain object specific characteristics such as width to height ratio are used to localise objects of interest within the robot’s camera image [1]. Afterwards the region of interest is determined which is the smallest rectangle containing the object.

In order to save computational time and to facilitate the tracking of objects, classification results are utilised. For each object within the visual field the classification result is calculated three times consecutively. The aggregated result is obtained by majority vote. More than
Hierarchically organized RBF networks [1] are used to classify the detected objects of interest. The hierarchy is generated using unsupervised k-means clustering and conducts a successive partitioning of the primary set of classes into disjoint subsets with coarse discrimination at higher levels of the hierarchy and stepwise refined discrimination with increasing depth of the hierarchy. The classification result is obtained similar to the retrieval process in a decision tree yielding a specific path through the hierarchy.

The incremental learning of new objects is performed in two stages. In the first stage fast but less sophisticated methods are used to obtain initial results, i.e., the novel objects are learnt but the recognition rate might be weak. In this first stage the recognition rate can be improved by using a similar method to retrain the new object with additional data. In a second stage more complex algorithms are used to adapt the system and to further improve the classification results.

At first it is necessary to identify whether the presented object is already known or whether it is a new object. This is accomplished by presenting the new data to the trained classifier and taking the strength of the classifier response into account. If an object is identified as unfamiliar it is learnt by fitting it into the hierarchy and if necessary retraining the affected nodes.

The new class associated with the unknown object is inserted into the hierarchy as a new leaf. The position of the new leaf is determined by classifying all new data and evaluating the corresponding paths through the hierarchy. The leaf will be inserted where the paths start to diverge. As complete identicalness for all data cannot be presumed even at the root node since the network has not been trained with this data a certain variance needs to be considered. Otherwise the new leaf would always almost be added at the root node.

By means of classification experiments the suitability of hierarchical neural networks for extension was examined. It could be verified that new classes which are only represented by a few samples can be learnt sufficiently well in moderate time and that it is possible to learn new classes without negatively affecting the classification performance of already learnt classes.

The adaptive incremental learning approach was evaluated against a complete redesign of the classifiers. Both approaches showed essentially the same classification quality with average classification rates of 95.63 ± 2.49% and 95.54 ± 2.56% respectively where the redesign approach features longer training time as a complete rebuild of the hierarchy and training of all nodes within the hierarchy are required.

For the evaluation the Columbia Object Image Library (COIL20) [2] consisting of 20 objects was used. For the experiments 10 classes of the 20 classes of the COIL20 data set were chosen to represent the familiar objects. The remaining 10 classes formed a pool of potentially new objects.

4 Conclusions

The proposed approach 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 adaptation. 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.

References
