Memory Organization for Invariant Object Recognition and Categorization

Guillermo S. Donatti, Rolf P. Würtz
Institut für Neuroinformatik, International Graduate School of Neuroscience
Ruhr-Universität Bochum
44780 Bochum, Germany
guillermo.sebastian.donatti, rolf.wuertz@neuroinformatik.rub.de

Abstract

The integration of bottom-up with top-down object processing has always been a topic of major concern in computer vision. However, while a lot is known about feature extraction, the knowledge-driven aspect of perception has been recognized as important, but hard to probe experimentally and difficult to implement in computer vision systems. How object knowledge must be organized so that it supports scene perception and can be acquired automatically is a research problem of outstanding significance for the biological, the psychological, and the computational approach to understand perception. The present work aims to develop an object memory model which can provide fast retrieval and robust recognition and categorization. The underlying data structure is inspired by the neural network structure of the human brain, connecting similar object views with excitatory synapses and using inhibitory synapses to separate different ones. The insights derived from building such a computational theory and the properties of the resulting model have implications for strategies and experimental paradigms to analyze human object memory as well as technical applications for robotics and computer vision.

1. Introduction

One fundamental aspect of perception is the processing of visual information and its relation with the accumulated knowledge of the world. This can be subsumed under the processes of “object recognition” [9], which refers to a decision about an object’s unique identity, and “object categorization”, which states the object’s kind [5]. Despite the fact that humans can determine spontaneously the correspondence of objects in a visual scene with the ones previously seen, this represents an extremely challenging task for every artificial system proposed so far [8]. Consequently, a large variety of models motivated in biology have been proposed. These models use either a featured-based or a correspondence-based approach. In both the processing of an object view is based on the extraction of image features together with the use of stored object models derived from training images.

The organization of the stored object views is crucial for having efficient object recognition and categorization. The general framework introduced in [7], the categorical basis function [6], and other models such as the neocognitron [1] use a sequence of feed-forward neural representations based on the simple-to-complex hierarchy stated in [2]. The view manifold [11] generates a two-dimensional mapping of all possible object views from a three-dimensional viewsphere. In the elastic graph matching approach [4], object views are represented by graphs whose nodes are labeled with local image features described by responses of a set of Gabor filters. Bunch graph matching [12] is a further extension of this model, which relies on the idea that objects within one class share the same landmarks with approximately identical geometrical relations, allowing the representation of a whole class of objects with one single model and reducing the number of required matching operations.

2. Motivation

Many of the proposed object recognition and categorization models focus on the sequential bottom-up processing of object views without the use of top-down feedback from the latter to the earlier stages [8]. However, the use of past experience with similar object views to overcome some of the problems encountered in visual object understanding has an indisputable importance. Recent neuropsychological and functional anatomical data indicate that object identification involves top-down activation of earlier stages of visual processing [3]; this also correlates with the reciprocity in the visual cortical connections [8]. Additionally, the results of psychopysical experiments of object recognition in noisy images suggest that object classes have associated a pictorial representation which is used during priming [8].

But how can such a process be derived? One may argue that for biological systems, given a set of image features extracted from object views, a measure of similarity can be defined between them, and this can lead to the formation of natural clusters of object models that share the same image features [8].

3. Objectives

The present work aims to develop a novel memory framework which provides a structural association to the object models stored in an object recognition and categorization model. This structure is intended to be biologically inspired and its associations will be determined automatically through a learning process.

4. Current Research

The initial stage of the present work consits of extending the object recognition and categorization model described in [10] by adding a self-organized associative structure to its stored image features. This model is currently composed of three main modules. The first one extracts image features from object views; the second one uses a pre-selection neural network to match those image features with the ones derived from training images and to generate a set of object model candidates in a feature-based manner; and the third one uses a correspondence-based technique to verify those candidates with the object view selecting the one that attains best similarity as the recognized object model.

It is expected that the introduction of a memory structure will play a key role both in the learning and recall modes of the pre-selection neural network, improving the performance and robustness of the whole model. During training mode, the image features derived from different object views are the building blocks of a self-organizing neural map which automatically establishes associations among them based on a distance measure, which determines their similarity. During recall mode, upon the presentation of a novel object view, a new set of image features is derived and matched with the ones stored in the self-organizing neural map according to their distance measure. This process is done by navigating through its structure using the previously learned information to favor the selection of some image features over others until a selection of well-matching image features is found.

5. Future Research

The next stages of the present work will increase the semantic information captured by the self-organized associative structure described in Section 4 and will introduce the means for novelty detection. The approaches to be explored include reinforcing the associations between features which have been found together in a particular kind of scene, modifying them according to the object categories in which they were observed, and assimilating novel features through the use of an incremental learning mode.

6. Acknowledgements

The present work is supported by the Marie Curie Early Stage Training Programme "NovoBrain" and the Ruhr-University Research School.

References

[1] K. Fukushima, S. Miyake, and T. Ito. Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems*, 13:826–834, 1983.

[2] D. Hubel and T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

[3] G. Humphreys, M. Riddoch, and C. Price. Top-down processes in object identification: Evidence from experimental psychology, neurophysiology and functional anatomy. *Philosophical Transactions of the Royal Society: London Series B*, 352(1358):1275–1282, 1997.

[4] M. Lades, J. C. Vorbrüggen, J. M. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Computers*, 42(3):300–311, 1993.

[5] T. J. Palmeri and I. Gauthier. Visual object understanding. *Nature Reviews Neuroscience*, 5(4):291–304, 2004.

[6] M. Riesenhuber and T. Poggio. Cbf: A new framework for object categorization in cortex. In *BMVC '00: Proceedings of the First IEEE International Workshop on Biologically Motivated Computer Vision*, pages 1–9, London, UK, 2000. Springer-Verlag.

[7] T. Serre, L. Wolf, S. Bileschi, M. Reisenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(3):441–426, 2007.

[8] S. Ullman. *High-Level Vision, Object Recognition and Visual Cognition.* Massachusetts Institute of Technology, second edition edition, 1997.

[9] J. Ward. *The Student's Guide to Cognitive Neuroscience.* Psychology Press, first edition edition, 2006.

[10] G. Westphal. *Feature-Driven Emergence of Model Graphs for Object Recognition and Categorization.* PhD thesis, Institut für Neuroinformatik, 2007.

[11] J. Wieghardt, R. P. Würtz, and C. von der Malsburg. Learning the topology of object views. In *ECCV (4)*, pages 747–760, 2002.

[12] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):775–779, 1997.