

ADAPTING COPYCAT TO CONTEXT-DEPENDENT VISUAL OBJECT RECOGNITION

SCOTT BOLLAND

*Department of Computer Science and Electrical Engineering The University of Queensland
Brisbane, Queensland 4072 Australia
E-mail: scottb@csee.uq.edu.au*

JANET WILES

*Department of Computer Science and Electrical Engineering, Department of Psychology The
University of Queensland Brisbane, Queensland 4072 Australia
E-mail: janetw@csee.uq.edu.au*

Spatial context provides an important role in facilitating the visual recognition of objects in humans, but remains a cue largely unutilised by artificial systems. This deficit is due primarily to the complexity of the processing issues involved which have proven difficult to model from a primarily symbolic or connectionist approach. Success at addressing these issues has however been achieved by Copycat, a computational model of analogy-making which is capable of representing complex hierarchical structures in a flexible manner. The project described by this paper is aimed at modifying this model to the domain of context-dependent visual object. This paper provides an overview of a partial implementation of the system that has demonstrated the viability of the approach, being able to recognise a set of objects over a range of change in illumination, pose and scale. The use of spatial context in the recognition process as well as the implications for future work are also discussed.

1 Introduction

Visual object recognition is a seemingly effortless and natural activity in many animal species, however, robust recognition of familiar objects still lies beyond the capabilities of artificial systems. One factor that has been demonstrated to have an effect on human object recognition performance that has been relatively unexplored in artificial systems is the influence of context. It is well known through empirical investigations that contextual information either internal (such as prior knowledge or expectations) and external (such as the position of the object in congruent settings) facilitates object recognition in humans both in speed and accuracy [1-4]. Many approaches to object recognition such as Biederman's Recognition-by-Components theory [5] assume that objects can be uniquely defined by a relatively small number of shape models. Other approaches such as Mel's SEEMORE [6] assume that objects have characteristic local features (such as texture and colour) that are reliable indicators of the object's identity. However, there are many conditions under which both of these assumptions fail, so that the object cannot be uniquely identified in isolation, but in which context can be used as a cue for disambiguation.

Due to the complexity of the issues involved, although attempts have been made to utilize spatial context, success has been limited to narrow domains such as in the disambiguation of handwritten letters using word context [7]. Such issues however, including the nature of the representations that are required, and the processes of representation formation and analogical mapping, have been investigated in other areas of cognitive science such as the fluid analogy models proposed by Hofstadter and his research group [8]. The aim of this project is to apply the ideas of one such model, Copycat [9], to address the representational and constructional issues involved in the implementation of a context-dependant visual object recognition system.

The facilitating role of spatial context in object recognition involves many issues fundamental to high-level perception tasks in general that have proven challenging to model. Firstly, not only do the structures that are needed to represent the context involve complex hierarchical relations that are difficult to model from a connectionist framework, they must be modeled in a flexible manner that precludes the use of a purely symbolic approach. Furthermore, as with the identification of individual objects, the recognition of a context cannot be achieved using direct matching strategies due to the many potential sources of variability. Rather, matching should be viewed as a form of analogy making in which the current situation is compared to stored contexts at a high level of abstraction in order to equate lower level non-identical percepts. Most models of analogy making however, such as Gentner's "structure-mapping" theory [10], require that only the raw data that is central to the analogy be given in the initial problem representation. However, in real world domains, the selection of the relevant data is an integral part of the analogy-making process. For example, in recognizing a bottle as a potential rolling pin or a potential vase, different properties need to be attended to. In contrast to other such models, in Copycat the properties of the situation attended to as well as the mappings that are made across source and target analogues are co-dependent and evolve together.

The fluid analogies framework for modeling creative analogies was explicitly designed to address the interplay between the derivation of representations and the correspondences between them (see [8], for a detailed description of the underlying issues). The Copycat model [8, 9] utilises a hybrid architecture that combines the representational power of the symbolic approach with the flexibility and robustness of connectionism to allow for the equating of complex situations. The aim of the project described in this paper is to adapt Copycat to the implementation of a context-dependent visual object recognition system that utilises the flexibility of Copycat in the recognition of both objects and contexts. The goal of the system is to robustly recognize a set of objects under a range of viewing conditions including a degree of change in object setting, scale, pose and illumination. Objects are represented in terms of deformable grids of Gabor wavelet probes, which has proven efficacious in the recognition of rigid objects as well as a range of distortable

objects such as faces [11, 12]. Unlike other systems utilizing this method which apply a serial search through all stored object models, Copycat instantiates a more efficient “parallel terraced scan” in which object models are explored to a degree based on the amount of evidence for their presence acquired from the environment. Figures 1-4 from this paper describe a partial implementation of the system that demonstrates the viability of the adaptation of Copycat to this domain. This paper also describes the main modifications to Copycat that were required and concludes with a discussion of future work.

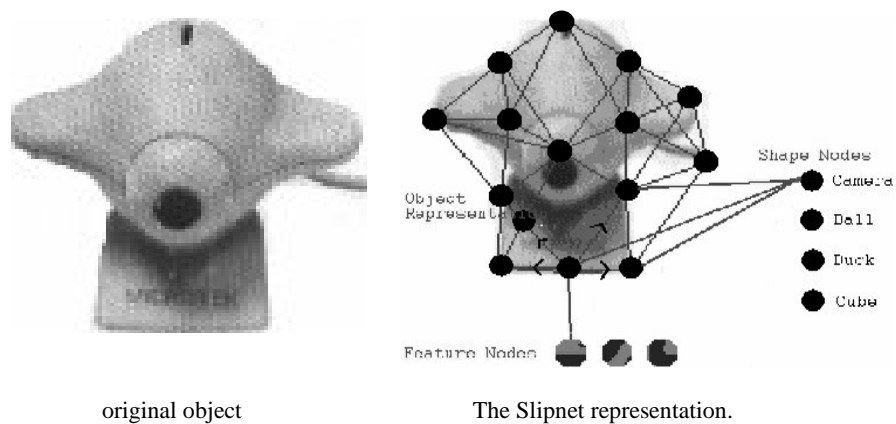


Figure 1: a portion of the Slipnet (Copycat’s long term memory that is implemented as a semantic network) showing the way in which objects are represented. Objects are represented in terms of the spatial relationship between salient local feature points which are described as a set of Gabor Wavelets at various orientations and scales. In the Slipnet, each feature is described by a single node which is linked to its four closest neighbours. The links encode the spatial relationships between the points in terms of pixel offsets.

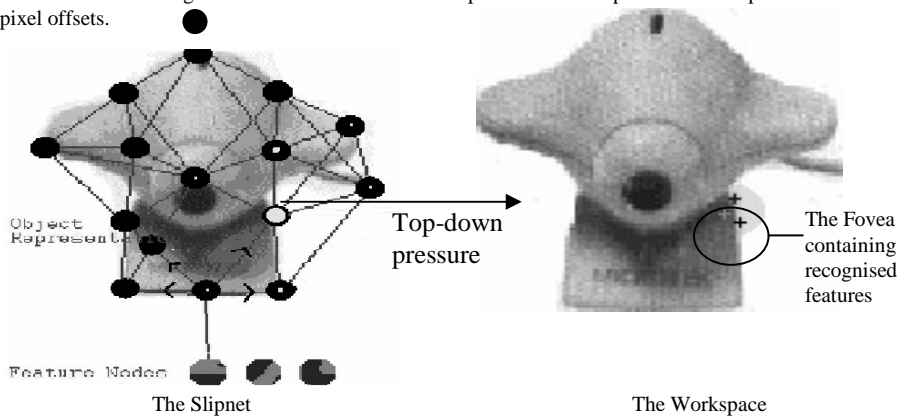


Figure 2. The recognition of feature points in the Workspace. Features are detected in the foveal area through a combination of bottom-up pressures (searching for the presence of any feature in the fovea)

and top-down pressures generated by the Slipnet (which search for the presence of concepts that are active in the Slipnet).

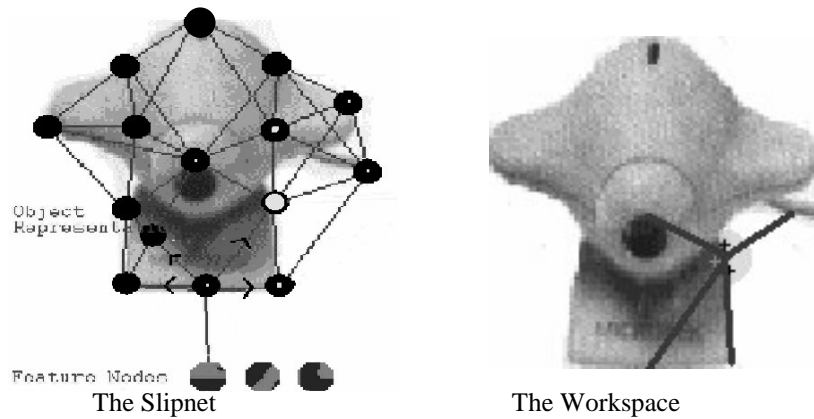


Figure 3. The proposal of bonds in the Workspace. Based on the object descriptions stored in the Slipnet, neighboring features are detected in the Workspace, strengthening the estimation of a correct feature match within the Fovea, and proposing potential saccade target sites to further explore the interpretation.

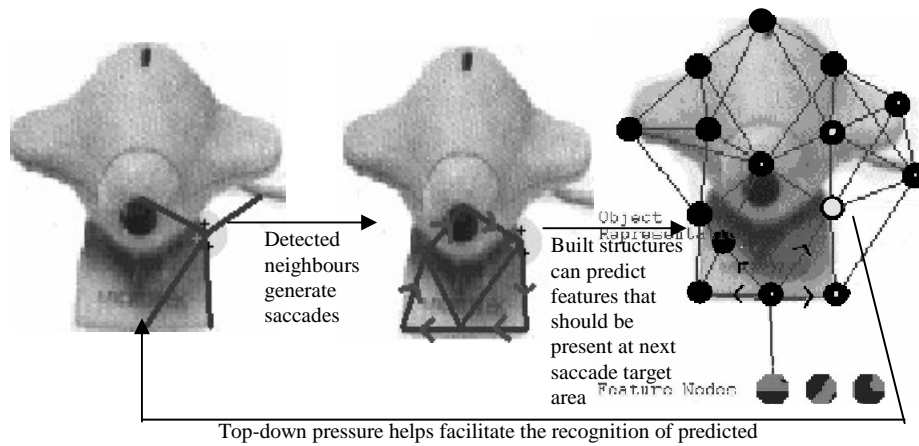


Figure 4. The detection of neighbouring concepts generates a series of saccades in the Workspace. As the fovea moves, and more features are detected, portions of the object models are reconstructed in the Slipnet. Such built structures can make predictions upon what features should be found at the next saccade target area, and activate these features in the Slipnet which place top-down pressure on the system to explore these interpretations.

2 Modifications Made to the Original Copycat Model

In adapting the original Copycat model to the domain of visual object recognition, three main modifications were required. Firstly, in the original design, pressures exerted by the system were global rather than local. That is, when a Slipnet concept became active, there was no information stored as to where in the Workspace the corresponding pressure should be applied. As object recognition concerns the spatial relationships between features, pressures to search for a particular feature in a localised region needed to be added. In the current implementation this localised pressure was achieved through the use of a moving fovea in which the Slipnet's activations could reflect the predicted features. Once part of an object model was reconstructed, predictions could be made as to what features are present in the new foveal area, with the Slipnet's activations being adjusted accordingly.

The second modification to the system was with respect to the links in the Slipnet. In the current implementation there are no "slip links" between nodes in the Slipnet that represent the potential conceptual slippages. Instead, a specific region of the image can be classified by a range of stored feature concepts with the match being defined as the angle between the representational vectors. However, as the classification of a feature point is also dependent on its spatial relationship between neighbouring features, features that have a sub-optimal match with the local region may be perceived, thus displaying conceptual slippage.

The third main form of adaptation to the Copycat model occurred in the representations that were utilised within the Workspace. In the original Copycat, objects in the Workspace (such as letters) could be allocated descriptions that represented perceived properties from which relationships between such objects could also be perceived. Such descriptions included the position of the object in the string, the letter category to which it belonged, as well as the fact that it was a letter as opposed to a group of letters. From the relationships noted between the objects, they could be grouped into higher-order chunks, such as "abc" being perceived as a successive group of letters. In the current implementation by contrast, there were only two forms of structure present (feature probes and links) both of which had a single defining feature (i.e., a single Slipnet node or link to which they corresponded). Furthermore, in searching through the possible interpretations, a region of the image was not perceived to be a particular feature in an all-or-nothing manner as in the allocation of a description, but rather several competing descriptions could be entertained, with the amount of supporting structure (i.e., bonds to other features) denoting the stronger interpretation.

3 Results

The implemented system has so far been tested on images from approximately 20 different objects, being able to fully reconstruct the original feature grids under a variety of changes in illumination, pose and scale (examples of which are given in figure 5). This result demonstrates the viability of the adaptation of Copycat to the domain of visual object recognition. Although the current system does not directly model context-dependent visual object recognition (by utilising the spatial relationships between objects to facilitate recognition), the mechanisms required in doing so are modeled at a feature level. That is, in the implemented system the spatial relationship between feature points facilitate the recognition of other locally predicted features in an analogous manner as to how spatially context-dependent object recognition may occur.

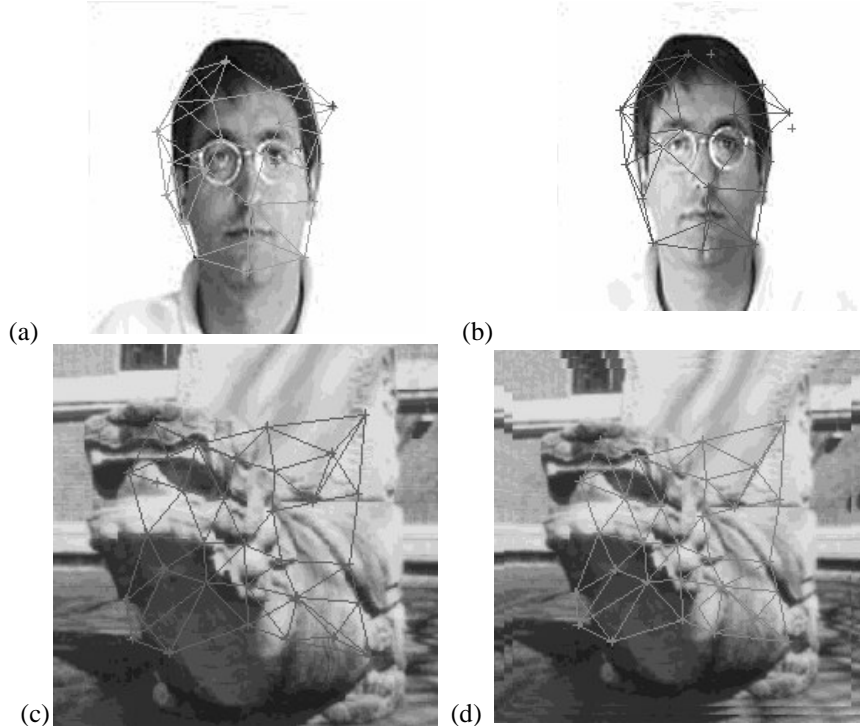


Figure 5: examples of recognised objects under varied conditions. (a) Initial object face image given to the system and the feature grid that is learned by the system. (b) The reconstructed object grid under changed illumination. (c) and (d) represent the learned and reconstructed grid of a statue under 90% image reduction.

4 Conclusion and Implications for Future Work

This report summarises the first stages of a project aimed at exploring visual object recognition through the Copycat paradigm. This approach has yielded successful recognition of several objects under conditions of varied illumination, scale, object background and occlusion, and has demonstrated the applicability of Copycat to the domain. However, to fully evaluate the performance against contemporary approaches further testing is required on common image databases to compare the efficacy of this method. The next stage of the project will involve fine tuning the parameters used by the system and comparing its performance to other approaches using databases including those of the Massachusetts Institute of Technology (MIT), the Olivetti Research Lab, the Weizmann Institute of Science and the Bern University.

5 Acknowledgements

This project has been supported by an APA to SB and an ARC grant to JW.

References

1. Biederman, I., R.J. Mazzanotte, and J.C. Rabinowitz, *Scene perception: detecting and judging objects undergoing relational violations*. Cognitive Psychology, 1982. **14**: p. 143-177.
2. Morton, J., *Interaction of information in word recognition*. Psychological Review, 1969. **76**: p. 165-178.
3. Palmer, S.E., *The effects of contextual scenes on the identification of objects*. Memory & Cognition, 1975. **3**(5): p. 519-526.
4. Potter, M.C., *Meaning in visual search*. Science, 1975. **187**: p. 565-566.
5. Biederman, I., *Recognition-by-Components: A Theory of Human Image Understanding*. Psychological Review, 1987. **94**(2): p. 115-147.
6. Mel, B.W., *SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition*. Neural-Computation, 1997. **9**(4): p. 777-804.
7. Rose, T.G., L.J. Evett, and A.C. Jobbins. *A Context Based Approach to Text Recognition*. in *Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval*. 1994. Las Vegas, Nevada.
8. Hofstadter, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanics of Thought*. 1995, New York: BasicBooks.
9. Mitchell, M., *Analogy-Making as Perception*. 1993, Cambridge Massachusetts: MIT Press.

10. Gentner, D., *Structure-mapping: A theoretical framework for analogy*. Cognitive Science, 1983. **7**(2): p. 155-170.
11. Wu, X. and B. Bhanu, *Gabor Wavelet Representation for 3-D Object Recognition*. IEEE-Transactions-on-Image-Processing, 1997. **6**(1): p. 47-64.
12. Zhang, J., Y. Yan, and M. Lades, *Face recognition: eigenface, elastic matching, and neural nets*. Proceedings-of-the-IEEE, 1997. **85**(9): p. 1423-35.