SCHOOL OF COMPUTER ENGINEERING

Final Year Project Plan

Brain-inspired Object Recognition

SCE13-0343

by

Chris Stephen Naveen Ranjit

NTU Supervisor: Prof Tan Ah Hwee

Industry Supervisor: Tian Bo

Student ID: U1022364G

AY 2013-14

1. Introduction

1.1 Background and Motivation

Object recognition, the ability to discern objects in a visual field despite variance in appearance, has been a hot topic for research for at least the past 40 years in neuroscience and computer vision. In recent years, a growing body of corroborating evidence in neuroscience suggests that the ‘core circuitry’ of object recognition in the mammalian brain has been mostly uncovered to be the ventral pathway[1]. However, current understanding of the algorithm employed in this circuitry is still limited at best.

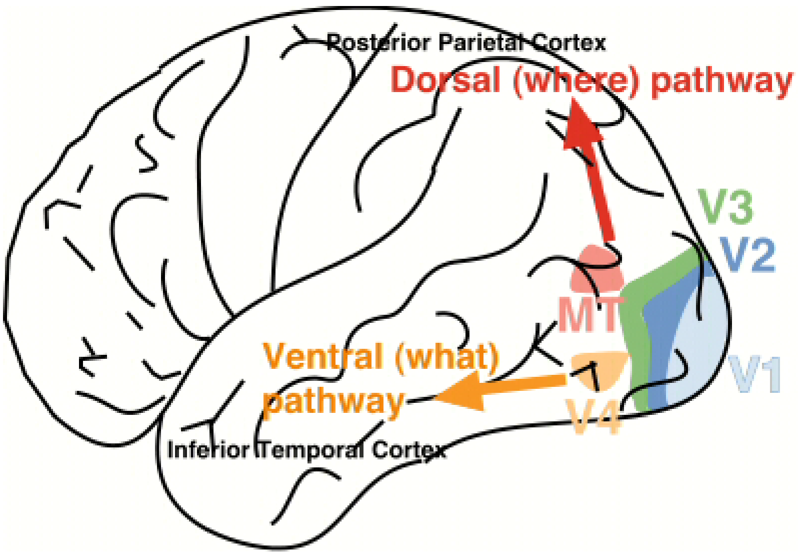

Figure 1: An illustration of the two-streams hypothesis in the neural processing of vision

Source: http://www.uwosh.edu/departments/psychology/Vreven/Lab/Images/brain.jpg

With the advent of modern computing, efforts have been made to reproduce in machines what Mother Nature has developed over eons. Advances in computational neuroscience over the years have yielded the HMAX model, which mimics known data of the ventral pathway in the mammalian visual cortex. It uses alternating layers of feature extraction and pooling for object recognition. Multiple improvements have been made to the HMAX model since its inception, and one of the more recent versions exhibit comparable performance to that of computer vision approaches to the problem, such as Bag-of-Words architectures, and outperforms other deep biologically inspired models [2].

1.2 Project Aims and Objectives

This project aims to employ an extended HMAX model [2] to complete an object recognition task for an in-house robot at the Institute for Infocomm Research, I^2R. The tasks designed to achieve this goal increase in difficulty and complexity as the project progresses.

Initial objectives are to familiarize with the extended HMAX model and the theory it is based upon, followed by application of the model and image pre-processing techniques to achieve simple object detection as well as segmentation of the region of interest within an input image.

Further along the project timeline, the system would be required to be able to recognize and classify multiple objects within the input image fed to it. The ability to recognize objects is necessary for future development plans for the robot to be able to grab objects which it has recognized.

Towards the end of the project, the project focus will shift to making an attempt to improve the system, including but not limited to making modifications to the algorithms and code optimization. If significant results are achieved, plans to publish a paper can be factored into the schedule.

1.3 Scope and Limitations

1.3.1 Scope

- Familiarization with the brain-inspired image recognition model HMAX
- Familiarization with existing image pre-processing techniques
- Implementation of HMAX model to solve object detection/recognition tasks
- Research on alternative approaches and possible improvements to the system

1.3.2 Limitations

- Project is visual processing-based, and does not extend to visually-guided motion

2. Related Work

2.1 Computer Vision

In computer vision, several approaches for object categorization have been developed. One of the simplest architectures [3] consists of a number of modules, each storing one view of an object. These modules output a value based on the similarity of an input image to their respective stored views. The outputs are weighted, summed and compared against some threshold value to classify the input image.

Other, more complicated, models have been created to tackle the problem of object categorization in images of the real world. These models are functionally composed of separate classifiers, each of which is trained to classify a region of interest in the input image for a set viewpoint. The classifiers are then aggregated together for multiple viewpoints of the same object [4]. These models vary in the features used in the classification process. Some use raw pixel values [5], while others create overcomplete measurement sets using wavelet filters [6], which are approximations of receptive fields in V1 of the visual cortex [4].

Some newer approaches have a hierarchical design wherein components of an object are distinguished in an image and later combined to represent the object itself [7, 8]. These approaches have been shown to be adept at recognizing faces, people and cars.

2.2 Brain-inspired Models

Models that adopt a brain-based approach to tackling the problem of object recognition can be categorized into two classes; object-centered models and view-based models.

Object-centered models represent objects based on the spatial relationship between parts that make up the object in three-dimensional space. The spatial relationships between parts of an object is extracted from different views of an object and matched to stored object representations. The most prominent example of such a model is recognition-by-components theory [9].

View-based models can be further broken down into feedback and feedforward variants. Most feedback models use an analysis-by-synthesis approach [4]. Essentially, these types of models make an initial guess of an object's identity, position and scale in the input image. They then create a representation of it using stored memories, compare that representation to the input image, and then correct the representation towards reducing the difference. Due to the feedback connections, these models tend to be slower than those that employ purely feedforward architectures.

Feedforward models use a hierarchical structure of neuronal units, wherein objects are represented by a combination of unit activations near the top of the hierarchy. Units at the bottom of the hierarchy are tuned to extract simplistic features, such as edges, from an input image. These low-level features are then pooled together in the next level of the hierarchy to form more complex features. This process of convolution and pooling is repeated in alternating layers, which culminates in the aforementioned high-level features at the top of the hierarchy.

3. Issues and Challenges

Due to the brain-based nature of the project, there will undoubtedly be many concepts that are foreign to me. Thus far in my literature review, some of these include:

- Unfamiliar mathematics
- Image processing techniques
- Neuroscience concepts and nomenclature

However, during my previous experience working on a cerebellum-based architecture for motor control during my Industrial Attachment, I had encountered similar obstacles and was able to cope. Although there will probably be concepts unique to the ventral pathway-based architecture used in this project, I believe I have the ability to pick it up. Additionally, I am currently enrolled in the 4th year Computer Vision and Neural Networks courses and will be able to apply what I learn to this project.

Another challenge imposed on the project is that since the extended HMAX model has undergone numerous improvements over the years, much time and effort will have to be spent understanding the theory behind them all. This is a necessary step in order to attempt to improve the model in the final stages of the project.

4. Approaches and Methodology

4.1 Tools

4.1.1 MATLAB

MATLAB is a numerical computing environment suitable for computationally intensive calculations and matrix manipulations. It possesses a vast library of built-in mathematical functions, and an extremely active user-based file repository on MATLAB Central. It is very suitable for mathematical modelling applications.

4.2 Project Stages

4.2.1 Stage 1

The first stage consists of an object detection task involving a block. One of the capabilities required of the in-house robot is to detect and recognize a block on a table using input images from a camera. This would involve familiarization with MATLAB code of the extended HMAX model, basic segmentation of region of interest, as well as training and testing of the model on the object recognition task. Also, to evaluate the robustness of the model, multiple illumination conditions will be used.

4.2.2 Stage 2

The second stage deliverables are for the robot to be able to recognize and classify multiple objects placed on the same table. The ability to recognize objects is necessary for future development plans for the robot to be able to grab objects that it has recognized, which is beyond the scope of this project. The number of recognizable objects will be increased steadily to evaluate the capacity of the model for this application.

4.2.3 Stage 3

The third stage will be an open-ended research venture to attempt to improve the system, including but not limited to making modifications to the algorithms and code optimization. Due to the difficulty and uncertainties involved with this stage of the project, nothing too concrete can be planned at this point in time. However, if fruitful, plans to publish a paper can be factored into the schedule.

5. Schedule

Task	Aug 13		Sep 13		Oct 13		Nov 13		Dec 13		Jan 14		Feb 14		Mar 14		Apr 14		May 14	
Background Reading on Object Recognition Architectures																				
Stage 1																				
Stage 2																				
Stage 3																				
Occurring Concurrently																				
Preparation & Submission of Project Plan																				
Preparation & Submission of Interim Report																				
Preparation & Submission of Final Report																				
Preparation & Submission of Amended Final Report																				
Oral Presentation																				
Exam Study Break																				

References

1. DiCarlo, J.J., D. Zoccolan, and N.C. Rust, *How does the brain solve visual object recognition?* Neuron, 2012. **73**(3): p. 415-434.
2. Thériault, C., N. Thome, and M. Cord, *Extended coding and pooling in the HMAX model.* 2013.
3. Poggio, T. and S. Edelman, *A network that learns to recognize 3D objects.* Nature, 1990. **343**(6255): p. 263-266.
4. Riesenhuber, M. and T. Poggio, *Models of object recognition.* Nature neuroscience, 2000. **3**: p. 1199-1204.
5. Brunelli, R. and T. Poggio, *Face recognition: Features versus templates.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1993. **15**(10): p. 1042-1052.
6. Oren, M., et al. *Pedestrian detection using wavelet templates*. in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. 1997.
7. Mohan, A., *Object detection in images by components.* 1999.
8. Heisele, B., T. Poggio, and M. Pontil, *Face detection in still gray images.* 2000.
9. Hummel, J.E. and I. Biederman, *Dynamic binding in a neural network for shape recognition.* Psychological review, 1992. **99**(3): p. 480.