Neurocomputing 158 (2015) 268–280

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A biologically inspired spiking model of visual processing for image feature detection

Dermot Kerr [a,*], T.M. McGinnity [a,b], Sonya Coleman [a], Marine Clogenson [a,1]

[a] Intelligent Systems Research Centre, University of Ulster, Magee Campus, Northern Ireland BT48 7JL, United Kingdom
[b] School of Science & Technology, Nottingham Trent University, Nottingham, United Kingdom

ARTICLE INFO

Article history:
Received 6 May 2014
Received in revised form
31 October 2014
Accepted 7 January 2015
Communicated by Yongzhen Huang
Available online 7 February 2015

Keywords:
Spiking neural networks
Image feature detection

ABSTRACT

To enable fast reliable feature matching or tracking in scenes, features need to be discrete and meaningful, and hence edge or corner features, commonly called interest points are often used for this purpose. Experimental research has illustrated that biological vision systems use neuronal circuits to extract particular features such as edges or corners from visual scenes. Inspired by this biological behaviour, this paper proposes a biologically inspired spiking neural network for the purpose of image feature extraction. Standard digital images are processed and converted to spikes in a manner similar to the processing that transforms light into spikes in the retina. Using a hierarchical spiking network, various types of biologically inspired receptive fields are used to extract progressively complex image features. The performance of the network is assessed by examining the repeatability of extracted features with visual results presented using both synthetic and real images.

1. Introduction

Biological visual systems are intrinsically complex hierarchical processing systems with diverse specialised neurons, various layers and feedback loops, displaying very powerful specific biological processing functionalities that traditional computer vision techniques have not yet fully emulated. Previous research has shown that the visual system deals with visual information processing by using complicated networks of diverse specialised neurons and complex interconnections to adapt to an extensive set of dynamic visual environments [15]. Existing bio-inspired artificial vision technology has neglected the possible benefit of modelling this rich diversity of cells.

In this work we present a hierarchical spiking neural network that focuses on modelling specific types of neuronal visual circuitry, in order to emulate specific aspects of a biological vision system. We develop a biologically inspired system that is capable of extracting key points from static images. Section 2 presents the background to this research. Section 3 describes the implementation of the presented model, including detailing the specific spiking neuron model used, centre–surround receptive fields, orientation specific receptive fields, end-stopped receptive fields, and full details of all the layers in the hierarchal model. Experimental results are presented in Section 4, including example visual results with synthetic images and real images, and a comprehensive performance evaluation is executed using a well-known feature repeatability evaluation technique. Section 5 discusses the network performance with advantages and weaknesses of the approach highlighted and also explores some possible areas for future improvement.

2. Background

The retina is the only source of visual information to the brain. It is a light sensitive tissue lining within the inner surface of the eye. It is regarded as an extension of the brain and formed embryonically from neural tissue and connected to the brain by the optic nerve. Neurons within the retina are arranged in three cellular layers and are interconnected in two intervening synaptic layers. The retina is composed of approximately 50 distinct types of cells. Visual processing begins in the first cellular layer when photons stimulate the 90 million rod photoreceptors and 6 million cone photoreceptors. These cells convert the information into chemical signals and send them through intermediate networked layers of various cell types with distinct functional processing abilities. The resulting processed visual scene from this network is represented by 1.2 million retinal ganglion cells of 15 distinct

* Corresponding author. Tel.: +44 28716 75330.
E-mail addresses: d.kerr@ulster.ac.uk (D. Kerr), tm.mcginnity@ulster.ac.uk (T.M. McGinnity), sa.coleman@ulster.ac.uk (S. Coleman), marine.clogenson@epfl.ch (M. Clogenson).
[1] Present address: EPFL, MC B2 293, Rue de la Maladière 71b, CP 526, CH-2002 Neuchâtel 2, Switzerland.

http://dx.doi.org/10.1016/j.neucom.2015.01.011
0925-2312/

types in the retina's output layer. The retinal ganglion cells convey a representation of the visual scene using action potentials (or spikes) along the ganglion cell axons to the optic nerve and onwards to the lateral geniculate nucleus and the visual cortex within the brain.

Each of the 15 distinct types of ganglion cells in the retina covers the entire visual field and transmits a completely processed image of the scene to higher brain areas [52]. Recent research has identified that computations of functional visual characteristics such as texture motion detection, approaching motion detection, orientation detection, contrast detection and motion extrapolation are carried out within the retina, the first stage in visual processing, and are not restricted to higher stages in the visual system as was once previously thought. Object-motion-sensitive ganglion cells in the retina, for example, have been found to distinguish between local object motion and global, self-induced motion, thus providing a rapid information channel for disentangling complex dynamical scenes. Similarly, retinal cells sensitive to approaching motion might underlie a quick avoidance reaction, and the observed retinal mechanisms for motion extrapolation are thought to underlie real-time object tracking. These insights illustrate the important role of early vision processing (i.e. retina level) in biological systems in terms of the detection of specific features, and highlight some of the advantages that a bio-inspired vision system could bring to artificial vision applications.

The detection of these specific types of features is facilitated through a number of powerful biological functionalities. For example, the diverse range of retinal and visual cortex neurons exhibit strong nonlinear processing steps and the specific connectivity between neurons incorporates varying delays and facilitates lateral inhibition through the use of complex synaptic interconnections. In particular, these complex synaptic interconnections form the basis of receptive fields of which various types have been identified [22]. Simple on–off centre–surround receptive fields have been identified that respond to contrast change [22], and to sharpen the image in space and also in time. Other types of receptive fields have been identified [22,29] where each type responds to different stimuli, for example orientated features. In addition, Shapley and Tolhurst [48] illustrated through psychophysical experiments that particular features, specifically edges, contours and corners are very important for visual perception [48]. Thus, with the biological and psychophysical evidence indicating such a strong focus on the detection of specific features in biological vision systems, then surely artificial vision systems should take a similar approach?

In terms of image processing, a local feature may be regarded as an image pattern which differs from its immediate neighbourhood. Normally this difference is connected to a change of image property such as intensity, colour or texture. Commonly extracted features from images include edges, corners or interest points. Once features have been identified a region surrounding the feature is normally used to identify the feature with a descriptor which may be used for various applications. The image processing community has proposed many feature detection operators in the past 30 years, in particular a number of different approaches for the detection of edge features have been proposed. Some of the earliest methods of enhancing edges in images used small convolution masks to approximate the first derivative of the image intensities to enhance edges [44]. Marr and Hildreth [37], used of zero crossings of the Laplacian of a Gaussian (LoG). Canny developed a multi-stage approach by incorporating first derivative approximation, non-maximal suppression and hysteresis suppression. Alternative feature detectors to detect edge junctions and corners have been proposed. Moravec [41] developed a corner detector that shifted a small square window in vertical, horizontal, and diagonal directions. Harris and Stephens [19] expanded the Moravec operator, removing the limitation of discrete window shifts, to develop a combined corner and edge detector. The operator response determines whether the detected feature is a corner, edge, or a flat region. Smith and Brady's SUSAN corner detector [50] is based on brightness comparisons over neighbourhoods and the detector can distinguish between corner and edge pixels. Shen and Wang [49] have expanded a local edge detector so that corners may also be detected. A combined edge and corner detector was presented in [5] that was capable of detecting both feature types concurrently.

Whilst these techniques have been somewhat successful for the detection of particular features in images, when comparisons are drawn between the performance of such artificial vision feature detectors and the processing capabilities of biological vision systems it becomes apparent that current computer vision approaches suffer serious weaknesses. To try to overcome these failings of conventional artificial vision techniques research has started to examine, take inspiration, and emulate aspects of biological vision systems. This process of simulating biological information processing in engineering is termed neuro-engineering [42] and such techniques are typically used for various artificial intelligent systems. For example, in [55] a feedforward second generation neural network is used to detect corner points using a model of end-stopped cortical cells and corner features are extracted using Gabor filter responses in [36]. In [16] a model that discovered characteristic features spontaneously was introduced; this differs from the work presented here in that we pre-define the feature detection structures taking inspiration from retinal circuitry. However, as precise knowledge of the complete retinal and cortex neuronal circuits is still not available, it is difficult to implement detailed exact models of biological visual processing. Thus, most current artificial models are based on specific assumptions, and simplified biological processes using variations of second generation neural networks that lack aspects of biological realism [6].

Spiking neural networks (SNNs) are the third generation class of neural networks that use a temporal coding scheme. This coding scheme enables spiking neural networks to more accurately mimic the biological information processing in the brain and visual system, and can be used to increase computational processing power and speed when compared with traditional second generation neural networks enabling real-time processing [34]. SNNs use simple neuronal models and communicate using spikes in a manner similar to action potentials found in biological neurons. There has been some research investigating the application of SNNs to visual processing. In [51] scene categorisation is performed and this work is then expanded in [38] to perform object recognition. In [25] contours are detected in images through the synchronisation of integrate and fire neurons. SNN approaches have also recently been applied for the purpose of image segmentation, in [53] which has proven to be fast and efficient and in ([53], Wu, et al., 2012) a SNN was proposed that detected right angle corners.

A SNN is used to model two areas of the brain concerned with motion with the aim of performing action recognition [8] and a distributed SNN is proposed for extracting saliencies in an image [4]. In [3] a SNN is used to perform difference of Gaussian filtering. Additionally, spiking neural networks have been previously used as controllers in evolutionary robotics to perform vision based obstacle avoidance [9], and for laser-based retinal model robot vision [39]. In [12,35] a robot's sensory information is converted into spikes and a spiking neural network is used to process the information and control the robot. A biologically inspired flying robot is developed in [10] that uses a spiking neural network to convert visual information into motor commands and in [18] a spiking neural network is used to control a mobile robot using sonar sensors.

This paper presents an approach to feature detection using biologically inspired spiking neural networks to develop an artificial vision system that reflects a stronger correlation with biological visual systems than second generation neural networks. This stronger correlation is achieved via the use of a spiking neural network that mimics the biological visual processing system and the use of biologically