Biologically Inspired Cognitive Architectures (2013) 6, 8–11

Available at www.sciencedirect.com

SciVerse ScienceDirect

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

RESEARCH ARTICLE

Self-organisation of motion features with a temporal asynchronous dynamic vision sensor ☆

F. Koeth [a,b,*], H.G. Marques [a], T. Delbruck [c]

[a] AI Lab, University of Zürich, Zürich, Switzerland
[b] Institute for Cognitive Science, University of Osnabrück, Osnabrück, Germany
[c] Institute for Neuroinformatics, University of Zürich, Zürich, Switzerland

Received 14 March 2013; received in revised form 23 May 2013; accepted 23 May 2013

KEYWORDS
Self-organisation;
Motion perception;
Dynamic vision sensor;
Kohonen-network

Abstract
Neural circuits closer to the periphery tend to be organised in a topological way, i.e. stimuli which are spatially close tend to be mapped onto neighbouring processing neurons. The goal of this study is to show how motion features (optic-flow), which have an inherent spatio-temporal profile, can be self-organised using correlations of precise spike intervals. The proposed framework is applied to the spiking output of an asynchronous dynamic vision sensor (DVS), which mimics the workings of the mammalian retina. Our results show that our framework is able to form a topologic organisation of optic-flow features similar to that observed in the human middle temporal lobe.

1. Introduction

Over the past decades, neuroscience research has shown the immense impact of plasticity on the mammalian brain (Sur et al., 1988; Clark et al., 1988). This discovery has decreased the need to explain brain circuits in terms of hard-wired connections defined solely by innate processes. In spite of the progress made, there are still many open questions about how exactly the brain develops.

This study presents a self-organising approach to model the development of (visual) motion features, for which receptive fields can also be found in the mammalian cortex.

The proposed system is based on spatio-temporal correlation between precise spike time intervals. The input to the system is obtained from an asynchronous dynamic vision sensor (DVS), in which each photoreceptor fires asynchronously in response to changes in illumination contrast. When

* Corresponding author at: Institute for Cognitive Science, University of Osnabrück, Osnabrück, Germany.
E-mail address: fkoeth@uos.de (F. Koeth).

2212-683X/$ - see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bica.2013.05.010

a change is detected above (or below) a given threshold value, the receptor triggers a spike in a way similar to the action potentials generated in the mammalian retina.

The remainder of this paper is organised as follows. In the Materials section, the DVS and the robot platform are explained. The method section describes the self-organising network, the data collection procedure and data analysis. Experiments and results are presented in the following section. In the end the conclusion and a future outlook of the work are given.

2. Materials

The *silicon retina* is a vision sensor in which each pixel records local illumination changes independently and continuously with microsecond precision (Lichtsteiner et al., 2008). Since the sensor records discontinuities dynamically, it reduces the redundancy of conventional frame-based intensity images. One output event s_i of the camera is defined by its location (x, y), the timing t and the event-type, which is either ON or OFF for an increase or decrease of illumination respectively and can be defined by the derivative of the illumination $l : s = \{t, x, y, \mathrm{sgn}(\frac{\partial l}{\partial t})\}$. Since all of the 128×128 pixels work independently of one another, the system is not affected by over- or under-exposure.

As there is no output from the sensor when the image is constant, changes in the illumination have to be present, which can be caused either by moving external stimuli, or by self-induced movements (or both). In this study, we use self-induced movements (see Lungarella et al., 2003; O'Regan and Noë, 2001). The camera is mounted on top of two servo motors which allow it to pan and tilt. The target position of each servo is given by γ_h for the pan (horizontal direction) and γ_v for the tilt (vertical direction). The whole setup is placed in front of three different stimuli printed on an A4 sheet of paper: a bar, a checkered board and a filled circle (see Fig. 1 right).

3. Methods

The algorithm used in this study is a variation of a *Kohonen-network*, also called *self-organising map* (SOM) (Kohonen, 1982). It is an unsupervised learning method, which is based on competitive and Hebbian learning (Hebb, 1949). Our approach is based on temporal difference codings of the spiking input events (see below).

Data collection. In our experiments we use only a small subset of pixels located at the centre of the DVS. These pixels form a central patch containing a total of 4×4 pixels (see Fig. 1 left). The input x to the SOM is calculated from the temporal difference between the firing of each pixel x_i in the central patch and a given reference time t_0. The time t_0 is given by the time at which the first pixel fires in the central patch. In total the input to the SOM consists of a vector of 16 elements (one for each pixel in the patch). All the pixels must fire at least once in a time interval $[0, T_{max}]$ (where $T_{\max} = 50$ ms), otherwise the input is neglected (i.e. not fed into the SOM).

SOM architecture. The SOM architecture is shown in Fig. 1. The SOM consists of a fully connected network, in which all the weight vectors are initialised with small values taken from a uniform distribution. The output of the SOM consists of $p \times q$ nodes which form the feature map encoding the motion features (i.e. optic-flow). For all the experiments $p = q = 8$.

The SOM works as follows. At every input sample $\mathbf{x_j}$ presented to the network a winning node is identified as the node whose euclidean distance to the input vector is the smallest. In a circular neighbourhood around the winning node $\mathbf{w_{k,l}}$, the nodes $\mathbf{w_{h,i}}$ are updated according to the learning rule (Kohonen and Honkela, 2007), which is a variation of the original ruled proposed in (Kohonen, 1982)

$$\mathbf{w_{h,i}} = \mathbf{w_{h,i}} + (\alpha(j)\Theta_{k,l}(h, i, j)(\mathbf{w_{h,i}} - \mathbf{x_j})) \tag{1}$$

where $\alpha(j)$ is the time-varying learning rate and $\Theta_{k,l}$ is the *neighbourhood-function*, which represents the competitive part of the learning. The learning rule is proportional to the difference from the input vector to the current node, which is the Hebbian part of the learning.

The neighbourhood-function as well as the learning rate are both functions of time. To save computational time, the neighbourhood-function is first realised as a decreasing cut-off radius $r(j)$, where the euclidean distance in the grid has to be smaller than $r(j)$. The cut-off radius $r(j), \alpha$ and Θ functions used in this study are defined by:

$$r(j) = ae^{-bj}, \quad \alpha(j) = e^{-cj}, \quad \Theta_{k,l}(h, i, j) = e^{2 \cdot ||(k,l)^\top - (x,y)^\top||^2 \cdot r(j)^{-2}} \tag{2}$$

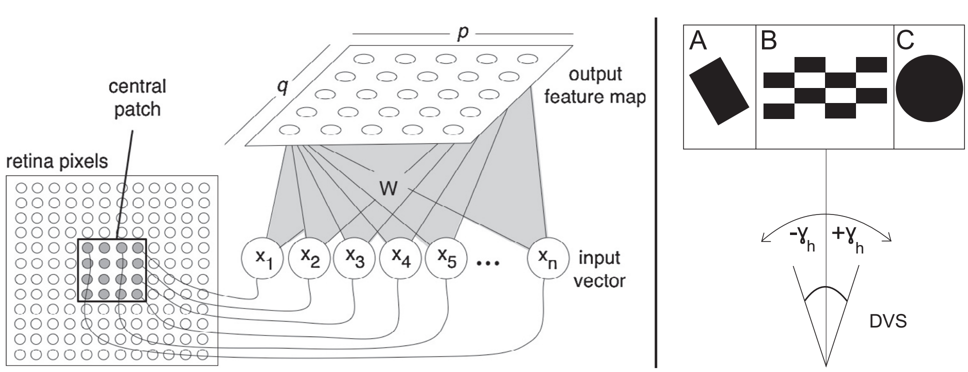

Fig. 1 **Left:** Schematic diagram of the network architecture. **Right:** Experimental setup with the three different stimuli, i.e. either a black bar, a chequerboard or a circle.