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

Conference Paper

Author(s):
Koeth, F.; Marques, Hugo Gravato; Delbrück, Tobias (iD)

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

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

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 $l : s = \left\{ t, x, y, \text{sgn}(\frac{\partial l}{\partial t}) \right\}$. 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:

$$r(j) = a e^{-bj}, \quad \alpha(j) = e^{-cj}, \quad \Theta_{k,l}(h,i,j) = e^{2.||(k,l)^\top - (x,y)^\top||^2.r(j)^{-2}} \tag{2}$$
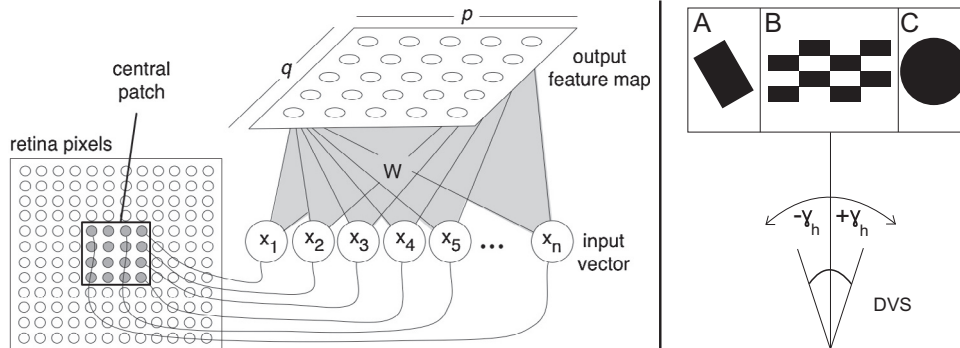
Fig. 1 **Left:** Schematic diagram of the network architecture. **Right:** Experimental setup with the three different stimuli, i.e. either a black bar, a chequerboard or a circle.

(a) 30°   (b) 90°
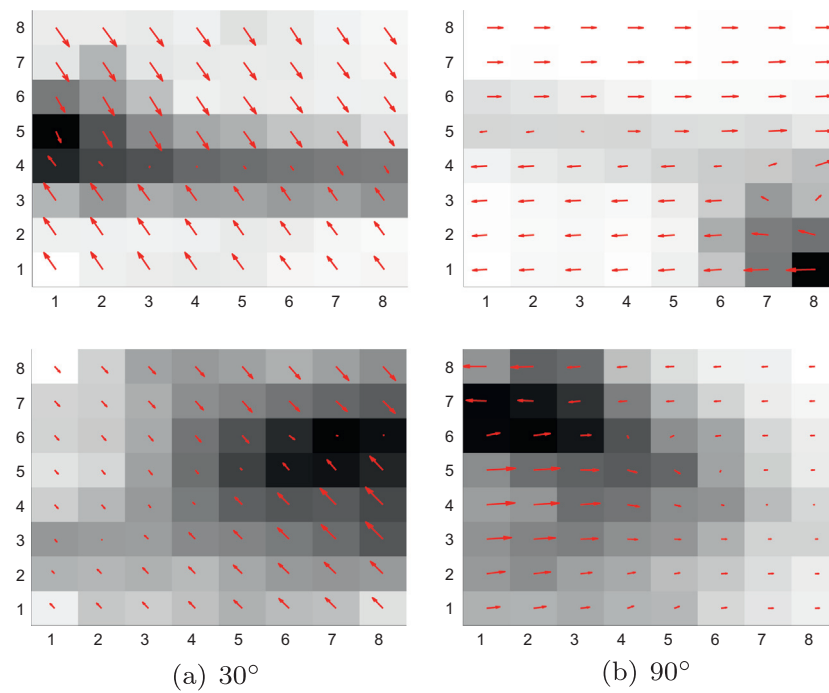
Fig. 2 For four different orientations of the bar, the SOM was trained with the data samples obtained from the video. The preferred direction of each node is plotted as a red arrow. The average distance to the neighbouring nodes is shown by the background for the node, where a bright background indicates small values and a dark background high values. **Top row:** condition (a): fixed target angles ($\gamma_h = \pm 6$); **bottom row:** condition (b): random target angles ($\gamma_h \in [-15°, +15°]$).

(a)   (b)

(c)   (d)
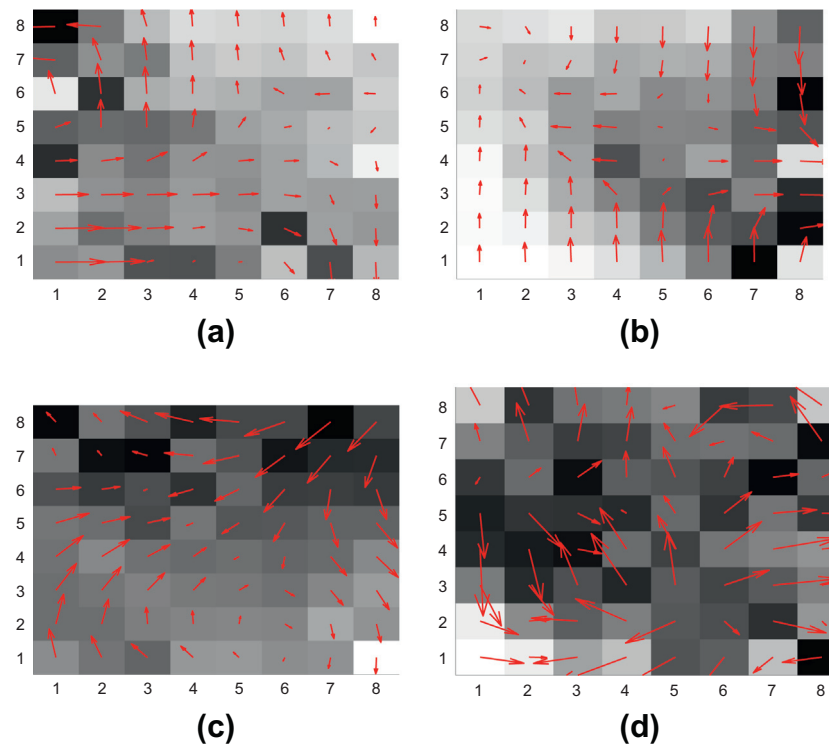
Fig. 3 2D movements over a chequerboard stimulus (a and b) or a circular stimulus (c). In (a), horizontal and vertical movements were done separately one after another, similar results for the same stimulus when movements were done in all possible directions. In (c), a circular stimulus with simultaneous movements in both directions was used. In (d) random data samples were created.

where n is the total number of input samples, $a = \max\{p, q\}, b = 2.5n^{-1}$, and $c = 1.6n^{-1}$.

To be able to visualise the data we need to project the feature map onto a two dimensional space encoding the horizontal h and vertical v components of the optic flow. These are given by:

$$h = \frac{1}{n}\left(\sum_{i=1}^{n}(w_{i,m}) - \sum_{i=1}^{n}(w_{i,1})\right) v = \frac{1}{m}\left(\sum_{i=1}^{m}(w_{1,i}) - \sum_{i=1}^{m}(w_{n,i})\right)$$ (3)

where n and m represent the row and column of the pixel associated with a given element in the weight vector.

Furthermore, to show separation properties of the SOM, one can compute the average euclidean distance to the neighbouring nodes in the grid and show this as a greyscale image (Ultsch, 1990), where dark shades represent large distances and bright shades small distances. Nodes which are in the same bright area can be considered to belong to the same cluster, whereas dark areas separate clusters. The computation of the weight distance is done on all 16 weights of the nodes, and for each node a vector representing the horizontal and vertical component is drawn.

4. Experiments and Results

Horizontal movements and bar stimulus In this experiment, we used the bar stimulus oriented in two directions: 30° and 90°. Only horizontal movements were triggered. The tests were performed in two conditions. In the first condition the pan target angle only changes in sign, $\gamma_h = \pm 6°$ (i.e. the norm of the velocity is almost constant). In the second condition, the target angles are drawn from a uniform distribution in the interval $\gamma_h \in [-15°, +15°]$, such that over time a continuum in velocity space can be found. The SOM is trained for 400 seconds for each bar orientation and condition.

Fig. 2 shows the results for both conditions and in both bar orientations. One can see that the SOM forms a topological arrangement in the motion space, i.e. neighbouring nodes usually encode similar orientations. In condition a), one can see that the distance to the neighbouring neurons increases drastically at the nodes, in which there is a change in velocity direction (dark regions in Fig. 2 top row). In the second condition, we observe smoother discontinuities in the map; nodes present higher variations in the norm of the velocity. This is because of the broader velocity spectrum induced by the controller in condition b. In both experiments it is clear that neurons that encode similar preferred directions are statistically closer to one another.

Two dimensional movements with chequerboard stimulus. In the second experiment, the movements were performed in two dimensions instead of only one. The stimulus used was the chequerboard. In the first condition, the movements in horizontal and vertical direction were done separately, such that the camera moved either horizontally or vertically. In the second condition, the movements were performed simultaneously, causing motion into all possible directions. In both conditions, the target angles are picked from a uniform distribution in the intervals $\gamma_h \in [-15°, +15°]$ and $\gamma_v \in [-15°, +15°]$).

The results are shown in Fig. 3a and b. As can be seen, the learned velocity in both conditions forms a topological arrangement according to the input space. In both conditions, the perceived motion is either up, down, left or right, i.e. independent of the direction of motion. Aside from the fact that the neurons learn the motion direction, this experiment also shows that neighbouring neurons learn similar movements.

Two dimensional movements with circle stimulus. The experimental set-up was the same as in the second condition in Section c, but with a circle stimulus instead of a chequerboard. The circle ensured that motion can be perceived motion in any direction. The results (see Fig. 3c) shows again a very clear topological arrangement of the neurons in the motion space, in which all the directions are covered. Between neighbouring neurons, the preferred movement is continuously changing.

Fig. 3d shows a SOM with randomly created spike-time differences. The incoherent structure of the input data makes it impossible for the SOM-nodes to be mapped coherently onto a horizontal and vertical direction and therefore does not yield in a topologically arranged map.

5. Conclusion

In this study we used an asynchronous vision sensor, which resembles the workings of the human retina. We have shown that using a simple algorithm, the spatio-temporal data obtained can be projected into a two dimensional space which encodes optic flow. In addition, we have shown that our self-organising principle can preserve the topology of visual spatio-temporal relations in a way similar to those in the middle temporal lobe.

References

Clark, S., Allard, T., Jenkins, W., & Merzenich, M. (1988). Receptive fields in the body-surface map in adult cortex defined by temporally correlated inputs. *Nature, 332*, 444—445.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Wiley.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 69*, 59—69.

Kohonen, T., & Honkela, T. (2007). Kohonen Network. *Scholarpedia, 2*(1), 1568.

Lichtsteiner, P., Posch, C., & Delbruck, T. (2008). A 128 times; 128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits, 43*(2), 566—576.

Lungarella, M., Metta, G., Pfeifer, R., & Sandini, G. (2003). Developmental robotics: A survey. *Connection Science, 15*(4), 151—190.

O'Regan, J. K., & Noë, a. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences, 24*(5), 939—973 (discussion 973—1031).

Sur, M., Garraghty, P. E., & Roe, a. W. (1988). Experimentally induced visual projections into auditory thalamus and cortex. *Science, 242*(4884), 1437—1441.

Ultsch, A. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In *Proceedings of the INNC'90, international neural network conference*.