

# Dynamic Detectors of Oriented Spatial Contrast from Isotropic Fixational Eye Movements

Simone Testa<sup>1</sup>, Giacomo Indiveri<sup>2</sup> <sup>a</sup> and Silvio P. Sabatini<sup>1</sup> <sup>b</sup>

<sup>1</sup>Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genoa, Genoa, Italy

<sup>2</sup>Institute of Neuroinformatics, University of Zürich and ETH Zürich, Zürich, Switzerland

**Keywords:** Active Vision, Fixational Eye Movements, Event-based Sensor, Neuromorphic Computing, Receptive Fields, Spiking Neural Networks.

**Abstract:** Good vision proficiency and a complex set of eye movements are frequently coexisting. Even during fixation, our eyes keep moving in microscopic and erratic fashion, thus avoiding stationary scenes from fading perceptually by preventing retinal adaptation. We artificially replicate the functionalities of biological vision by exploiting this active strategy with an event-based camera. The resulting neuromorphic active system redistributes the low temporal frequency power of a static image into a range the sensor can detect and encode in the timing of events. A spectral analysis of its output attested both whitening and amplification effects already postulated in biology depending on whether or not the stimulus' contrast matched the  $1/k$  falloff typical of natural images. Further evaluations revealed that the isotropic statistics of fixational eye movements is crucial for equalizing the response of the system to all possible stimulus orientations. Finally, the design of a biologically-realistic spiking neural network allowed the detection of stimulus' local orientation by anisotropic spatial summation of synchronous activity with both ON/OFF polarities.

## 1 INTRODUCTION

Visual perception is a fundamentally active process. Humans and many other mammals are endowed with a specific and complex set of eye movements through which they incessantly scan the environment (Land, 2019). This active method has long been proven to overcome loss of vision during fixation of static objects (Ditchburn and Ginsborg, 1952) thanks to a peculiar ensemble of oculomotor mechanisms known as *Fixational Eye Movements* (or FEMs) (Martinez-Conde and Macknik, 2008). In particular, while the desensitization properties of retinal ganglion cells to unchanging stimuli would lead to perceptual fading of stationary objects during retinal stabilization, FEMs enable refreshing neural responses by inducing temporal transients (Riggs and Ratliff, 1952). The fact visual systems so strongly depend on temporal changes suggests that the still-camera model of the eye and the spatial coding idea is at least lacking. Actually, in order to extract and code spatial information, a combination of spatial sampling and temporal processing

is required (Rucci et al., 2018). The performance gap between artificial and biological visual systems could therefore depend on substantial differences about how information is acquired and encoded. Biological evidences (Gollisch and Meister, 2008) indicate that retinal ganglion cell outputs are massively parallel, data-driven (asynchronous) and with high temporal resolution. Here, a temporal encoding scheme is adopted, where information is carried in the timing of activation, as opposed to pure spatial encoding schemes, which are solely based on the identity of activated receptors.

Similar to a biological retina, a neuromorphic camera, such as the *Dynamic Vision Sensor* (DVS), only responds to temporal transients in the visual scene, by converting them into a stream of asynchronous events uniquely based on time-variations of luminance contrast (Lichtsteiner et al., 2008). In addition to the position and the timing of brightness change, each event brings information about its polarity, i.e. ON or OFF events, that represent dark-to-light or light-to-dark transitions, respectively. By interfacing these sensors to mixed signal analog-digital neuromorphic electronic processors, such as the *Dynamic Neuromorphic Asynchronous Processor* (DY-

<sup>a</sup>  <https://orcid.org/0000-0002-7109-1689>

<sup>b</sup>  <https://orcid.org/0000-0002-0557-7306>

NAP) (Moradi et al., 2017), real-time and energy-efficient vision processing systems can be implemented. However, despite the outstanding capabilities of neuromorphic sensors with respect to classical frame-based cameras, their intrinsic blindness to stationary images limits their application in several real-world scenarios, as a continuous relative motion between the scene and the sensor is required. Hence, a commonly adopted approach is to use moving stimuli displayed on a monitor, resulting however in recording artifacts due to video refreshing (Orchard et al., 2015). Yet, by emulating the FEMs of biological vision systems, we will be able to extract information from otherwise undetectable static stimuli and simultaneously avoid such artifacts.

In this paper, non-saccadic FEMs were modeled by a Brownian motion and physically induced on the sensor by using a pan-tilt unit (PTU). The resulting active approach proved to be effective for making event-based sensors responsive to static scenes, thus closely matching a strategy that biology has evolved to perform for similar tasks. The spatial frequency characterization of our sensing system confirmed behaviors postulated in biological studies (Rucci and Victor, 2015). Further analysis revealed that the spatial statistical distribution of micro arbitrary movements (particularly their isotropy) equalizes sensor activity with respect to the orientation of visual stimuli, preserving the efficacy of subsequent feature extraction stages. Specifically, a *Spiking Neural Network* (SNN) was designed to properly discriminate stimulus orientation at a local scale. The detection mechanism exploits highly synchronized events of both polarities, emitted by jittery pixels aligned with oriented edges, collected by means of a push-pull configuration of anisotropic spatial kernels, that resemble biological receptive fields of simple cells in the primary visual cortex.

The rest of the paper is organized as follows. Section 2 introduces both the set-up considered for FEM reproduction on the DVS and the experiments conducted, providing a description on how information is encoded by the system. Section 3 details the spectral analysis of the overall behavior of the active sensing system and the influence of isotropic random-like motion on the response to differently oriented stimuli. Section 4 presents the SNN for the detection of local orientation and, finally, in Section 5 we draw the conclusions.

## 2 MATERIALS AND METHODS

### 2.1 Active Vision System and Stimuli

Inducing bio-inspired fixational eye movements on a sensor requires, first of all, a mathematical model of such motion based on its characteristics in natural viewing. To this aim, we used a Brownian motion model to approximate some peculiar components of FEMs (Kuang et al., 2012). Four different random seeds of Brownian motions (with number of steps varying between 30 to 50) have therefore been generated through Python simulations. The resulting paths should keep unchanged the erratic aspect, mean frequency and size of biological FEMs, adapting the latter to the characteristics of neuromorphic sensors.

The overall experimental set-up is shown in Fig. 1a. It is mainly composed of a neuromorphic sensing hardware and a remotely controlled motorized unit for the generation of precise pan and tilt rotations of the camera. Specifically, we used a neuromorphic sensor DVS128 (having a  $128 \times 128$  pixel array) with a TV-lens C-mount F1.4-16 (6 mm focal length) and a PTU-D46-17 from Directed Perception Inc. (chosen for its resolution, as low as 3.1 arcmin). In order to obtain reproducible results, all acquisitions were conducted in controlled lighting conditions, by executing them in a specifically-dedicated dimly lit room. Stimuli were generated using PsychoPy and displayed on a 19" LCD Philips monitor with a resolution of  $1440 \times 900$  pixels, and refresh rate of 60 Hz. The distance between sensor and screen was kept at 50 cm for all the experiments. The event-based data, provided by the sensor in the Address-Event Representation (AER) protocol, have been visualized on a computer's screen and saved to disk for off-line processing.

The same custom Python script simulating the Brownian motion model converts the whole FEM-like path in a specific set of commands encoded in ASCII format. These commands were sent to the servo-motors of the PTU by means of a terminal emulator, which serves as interface for the controller, in order to drive the position of the PTU over time, according to the simulated path. As the field of view (FOV) of a single DVS pixel ( $\sim 22.9$  arcmin, leading to a maximum discernible spatial frequency of  $\sim 1.3$  cyc/deg) strongly differs from that of human photoreceptors ( $\sim 0.6$  arcmin in the fovea), typical sizes of natural FEMs have been scaled up accordingly. Thus, the maximum step amplitude for artificial FEM induced on the DVS is  $\sim 190$  arcmin (corresponding to 5 arcmin of biological motion), while the minimum is  $\sim 3.1$  arcmin (corresponding to 5 arcsec for biologi-

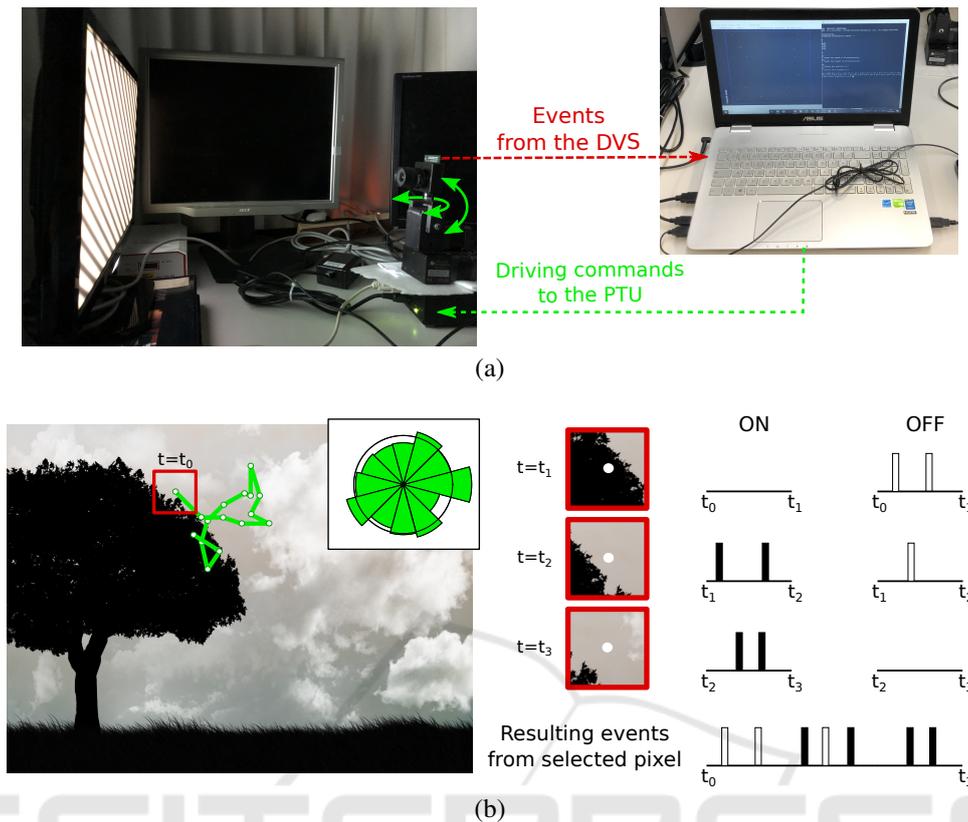


Figure 1: Set-up and consequence of fixational movements. (a) The experimental set-up: the DVS-PTU system is placed in front of a monitor for active scanning of artificial visual stimuli. Driving commands (in ASCII code) are sent from a host computer’s serial port to the PTU controller for generation of pan and tilt rotations. Conversely, data coming from the sensor, in AER protocol, are sent to the computer via USB interface. (b) Example of a natural visual scene and relative ON/OFF events generated by a single pixel of the DVS during fixational eye movements (green path). The inset shows results of the isotropic analysis of the movement (green polar histogram) with respect to a perfectly isotropic motion (black circle, representing the reference probability value of  $\sim 8.3\%$ , as 12 directions are considered). Red boxes represent the image portions falling in the FOV of the pixel at three different times.

cal motion). Lastly, the speed of both rotations was also controlled, in order to keep the whole motion at a mean temporal frequency between 40 and 50 Hz.

By varying the seed of the random process, different FEMs have been induced on the sensor while it is exposed to various stimuli, and the relative data streams have been recorded. Visual experiments have been conducted with artificial grating stimuli with varying contrast, spatial frequency and orientation. A first set of stimuli was composed of 180 gratings with: unitary contrast, spatial frequency  $k$  evenly spaced between 0.2 and 1.6 cyc/deg (15 discrete values), and orientation  $\theta$  evenly spaced between 0 and 165 deg (12 discrete values, with a 15 deg step). An additional set of 180 gratings was considered, having same values of  $k$  and  $\theta$  but an adjusted contrast for every tested spatial frequency, according to the  $1/k$  falloff of natural image spectrum amplitude (Field, 1987).

## 2.2 Spatiotemporal Coding

Figure 1b shows an example of how static spatial information is temporally encoded in the activity of the neuromorphic sensor as a result of the induced fixational movements. As a matter of fact, simulated FEMs ensure that the spatial structure of a static image is encoded both in space and time, as both the position and the timing of an activated pixel is informative. Specifically, camera movements transform a stationary spatial scene into a spatiotemporal luminance flow, thus redistributing its power in the (non-zero) temporal frequency domain which is capable of activating sensors pixels. Spatial luminance discontinuities are now converted into synchronous firing activity due to the combined effect of microscopic camera movements and the event-based (i.e., non-redundant) acquisition process in DVS pixels (see also Section 4). It is worth noting that a similar role of natural FEMs

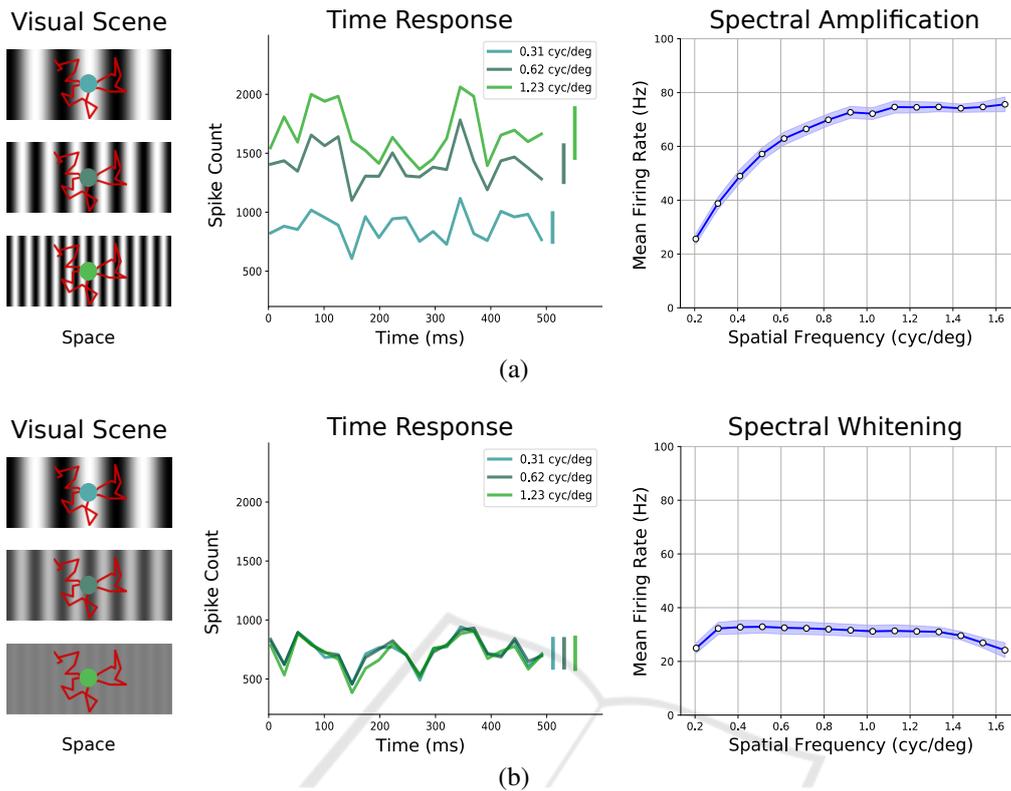


Figure 2: Sensors spectral characterization, for comparison with the results in (Rucci and Victor, 2015). (a) On the left, examples of unit-valued contrast gratings with 3 different spatial frequencies. FEM-like movements, induced on the sensor, are superimposed (red path). The central plot presents the resulting activity fluctuations exhibited by the  $30 \times 30$  central pixels of the DVS during 500 ms of recording: response is measured in number of events occurring in time windows of  $\sim 24$  ms. On the right, the amplification of the system to high-frequency grating stimuli with unitary contrast; standard deviation for different orientations of the gratings is represented by light-blue shaded regions. (b) On the left, gratings having same frequencies as in (a) but with contrast adjusted according to the  $1/k$  falloff of natural images. Resulting activity fluctuations of the neuromorphic active system are shown in the central plot, and spectral whitening effect to all frequencies on the right (shaded regions as in (a), right plot).

in biological early visual processing has been postulated both in (Greschner et al., 2002) and (Kuang et al., 2012), for initiating edge extraction and providing redundancy reduction of spatial information towards an economical representation of the image signal. Furthermore, phase shifts of activation in nearby pixels, that are subject to the same movement, should reflect spatial variations of luminance discontinuities impinging on nearby receptors, as proposed in the dynamic theory of vision presented in (Ahissar and Arieli, 2001).

It is worth noting that, as the DVS is sensitive to temporal changes of luminance only, anisotropic movements are undesirable, as they put into an adverse condition all stimuli having the same orientation of the directional bias of FEMs. Hence, to better investigate the effects of the erratic motion on the perception of such variously oriented stimuli, an isotropic analysis was conducted. To this aim, the

possible angles that each step of the Brownian motion could take was limited to multiples of 15 deg (as for stimuli). Results of the analysis (further described in section 3.2) for one particular realization of Brownian motion are shown in the polar plot in the inset of Fig. 1b. We anticipate that oriented luminance transitions would be the feature that benefit of such a property of FEM-like movements, supplying the system with an effective method for unbiased perception of local elements in the image signal.

### 3 SYSTEM RESPONSE CHARACTERIZATION

By taking inspiration from neuroscience studies (Rucci and Victor, 2015), we characterize the behavior of our active sensing system with respect to stimulus' spatial frequency - i.e. we examine sen-

sensor response to individual spatial gratings for different contrast conditions (either constant or matching the statistics of natural images). Certainly, the process for event generation in the DVS, the structure of the visual stimulus, and the specific camera movement adopted, collectively affect such a behavior at any moment. Therefore, we expect that the activity elicited in each receptor of our artificial retina complies with the statistical properties of natural images in the same way it happens in the biological one (Kuang et al., 2012). Furthermore, isotropic FEMs (i.e., for which all directions are visited with equal probability at each step) should provide an unbiased orientation information in the output signal. Accordingly, we have analyzed the mean response of the sensor with respect to this visual feature by systematically using gratings of different orientations. It is worth noting that all the results shown in the following are relative to one particular Brownian motion sequence, but same results were achieved independently of the random seed we considered.

### 3.1 Spectral Analysis

If we compute the mean firing rate of sensors pixels, for each spatial frequency  $k$  tested and by averaging over all the orientations, we get a measure of the overall spectral response of the system. In case of stimuli with maximum contrast, we can notice that pixels' mean activity increases with spatial frequency (see Fig. 2a rightmost panel): stimuli with higher frequencies elicit stronger responses in the system and a plateau value is reached approximately when spatial frequency  $k$  approaches the spatial resolution limit of the pixel array. As a matter of fact, by randomly moving around, as  $k$  increases, each receptor scans an increasing number of light-dark transitions, thus eliciting an increasing number of events. Remarkably, by adjusting the gratings' contrast according to the  $1/k$  falloff of natural image amplitude spectrum (Field, 1987), the frequency response with the same drift trajectory gives a roughly constant profile over the whole range of discernible spatial frequencies (see Fig. 2b rightmost panel). Hence, the active system tends to intrinsically oppose to the  $1/k$  trend of natural image distribution across spatial frequencies, thus counterbalancing the latter and enabling a whitened response.

### 3.2 Orientation Analysis

First, we computed the circular statistics of the Euclidean distances travelled in each of the 12 directions considered for the stimuli, relative to the whole path length. The aim was to quantify the isotropy of

FEM-like movements of our active system, in relation with the effect that each step of the whole movement causes on the perception of a specific orientation of the stimulus (that is why the isotropy was computed in the  $0 - \pi$  range rather than on the full circle). Results for a FEM-like motion sequence of 35 steps (e.g., see the inset of Fig. 1b, where the 12 probability values were remapped on a full circle for a more intuitive graphical representation) indicate that the maximum deviation from circularity is of 3%, thus evidencing a good approximation of an isotropic behavior. It is worth noting that the longer is the duration of active fixations the better is the isotropy we achieve. As a matter of fact, by considering 80 Brownian motion sequences with increasing number of steps (up to 400), the maximum discrepancy from isotropy decreases with this number (reaching a minimum of  $\sim 1\%$ ), in line with the fact that vision improves for longer fixational periods. By analogy with previous analysis, we have then computed the mean firing rate of sensor's pixels as a function of stimulus' orientation  $\theta$ , by averaging across spatial frequencies. As the sensor moves in isotropic fashion, we expect that the mean response of the system does not change with respect to different oriented gratings, as evidenced in Fig. 3, which refers to  $1/k$  falloff contrast gratings. Similar behavior has been obtained for constant contrast gratings as well (not shown). This equalized activity, in the form of spike sequences from each DVS pixel, provides a convenient input to a subsequent stage for unbiased detection of such orientation. Thus, we have designed and tested a bio-inspired spiking neural network for the detection of static objects' orientations on a local scale during active fixation.

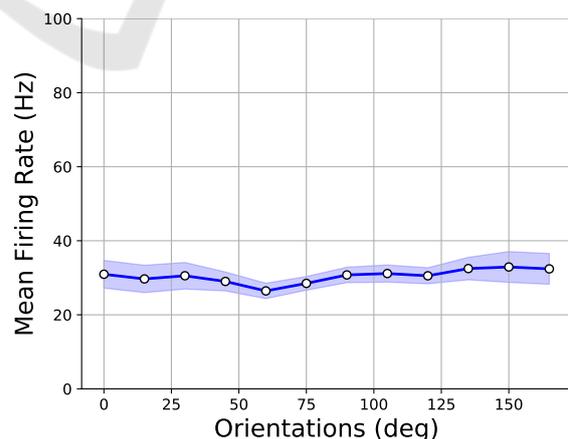


Figure 3: Average sensor response (mean pixel firing rate) to different oriented gratings (blue curve). The light-blue shaded region represents the standard deviation of pixels firing rate for different spatial frequencies.

## 4 ORIENTATION DETECTION

Orientation detectors consist of spiking neurons organized in a specific network architecture and fed with the event-based data provided by the neuromorphic camera. A spiking neural network is a biologically-realistic computational primitive where neurons and synapses are modeled as differential equations, representing the dynamics of membrane and synaptic currents (or voltages) respectively (Maass, 1997). Therefore, SNNs are powerful instruments for exploiting both space and time information of events from a jittery sensor. Besides, they can be implemented onto neuromorphic electronic processors (namely, the DYNAP) for bio-inspired, highly-efficient and real-time computing systems. To this purpose, we had to take into account hardware limitations both in the design and simulation phases. Accordingly, the neuron model used for software simulations of the network is the *Differential Pair Integrator* (DPI) (Indiveri et al., 2011), a variant of the generalized integrate-and-fire model, which is implemented by silicon neurons in the DYNAP. Likewise, the DPI model of synapses (Bartolozzi and Indiveri, 2007) was considered.

### 4.1 Network Architecture

The custom-designed detectors aim to mimic the functionality of simple cells in the primary visual cortex (V1). At any given spatial scale, detection is achievable by means of a 2D spatial kernel - i.e. a receptive field (RF) profile that defines the (feedforward) synaptic weights between the DVS pixels and a single V1-like spiking neuron. Neurons' selectivity to specifically-oriented edges relates to the direction of kernel's elongation ( $\theta$ ), i.e. by the anisotropic spatial summation of events provided by the sensor. In particular, we used difference of 2D offset elongated Gaussians that model two sub-fields organized in a push-pull configuration (i.e. adjacent excitatory and inhibitory regions) for antagonistic effects of ON and OFF events, as postulated by experimental evidence - e.g., see (Hirsch and Martinez, 2006). We specifically used RF sub-fields with approximate size of  $10 \times 3$  pixels and orientations ( $\theta$ ) evenly spaced between 0 and 165 deg (as we have done for the stimuli). The kernel was defined as:

$$h(x_\theta, y_\theta) = e^{-\frac{(x_\theta + \sigma)^2 + (y_\theta/p)^2}{2\sigma^2}} - e^{-\frac{(x_\theta - \sigma)^2 + (y_\theta/p)^2}{2\sigma^2}} \quad (1)$$

where  $x_\theta, y_\theta$  define a rotated reference frame, with respect to  $x, y$ ,  $\sigma$  represents the standard deviation (in number of pixels) along  $x_\theta$ , and  $p > 1$  defines the elongation of the RF along  $y_\theta$ . The specific values of the parameters are  $\sigma = 1.08$  and  $p = 4$ . It is worth

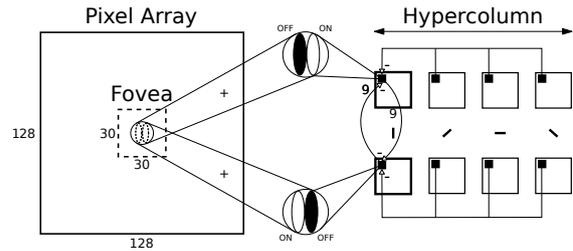


Figure 4: Schematic representation of network organization. The  $30 \times 30$  central region (fovea) of the whole DVS pixel array provides input to the SNN. Each neuron of a  $9 \times 9$  orientation column takes excitatory synaptic connections from the fovea through specific anisotropic spatial kernels acting as receptive fields (RFs). These RFs are characterized by two adjacent sub-regions selective to ON and OFF events respectively, and such distinction is interchanged for neurons in the two hypercolumns. The push-pull configuration is achieved via reciprocal inhibitory connections between corresponding neurons in the hypercolumns. Within a given hypercolumn, recurrent inhibition between different orientation columns is also considered.

noting that, in order to restrict the number of neurons in the network to a size that can be contained on the DYNAP board, only the central  $30 \times 30$  pixel region of the DVS was considered as input of the network, which is comparable to the biological fovea. Furthermore, in order to satisfy the constraints imposed by the DYNAP on the maximum number of connections available for each neuron, synaptic weights having absolute values lower than 0.2 were set to 0, thus limiting the actual size of the kernel.

A schematic (non-exhaustive) overview of network organization can be found in Fig. 4. The whole SNN consists of two neuronal populations, which we call “hypercolumns”, each one comprising 12 “orientation columns” (for the sake of clarity, only 4 are shown for each hypercolumn in Fig. 4). A single column constitutes a group of neurons having same tuning selectivity to a specific orientation of the stimulus as a matter of fact, the name relates to the columnar-like structures by which cortical orientation-selective cells are known to be organized in V1 (Hubel and Wiesel, 1974). The center of each RF was shifted by 3 pixels for nearby neurons belonging to the same column (i.e. their RFs share the same  $\theta$  but have different centers on the receptor array). Thereby, each orientation column includes 81 neurons, i.e.  $9 \times 9$ , starting from position (3,3) of the DVS fovea and ending in (27,27). The set of all orientation columns, internal to a same hypercolumn, constitutes a total of 972 neurons ( $81 \times 12$ ), which can then be contained in a single chip of the DYNAP board. Finally, the distinction between the two hypercolumns is given by the relative organization of their neuronal receptive fields. Particularly, if one sub-region of a RF is selective to

ON events, the adjacent sub-field will be selective to OFF events only, and this selectivity is interchanged between neurons in the two hypercolumns. The push-pull configuration is then achieved by means of reciprocal inhibition between neurons sharing the same location in the two hypercolumns (therefore sensitive to opponent contrast polarity). Furthermore, recurrent connectivity within both hypercolumns was considered, with the aim of optimizing discrimination by lateral inhibition (Blakemore et al., 1970) between neurons having different orientation selectivity but RFs centered in the same location of the pixel array. The weight of such connections was, in absolute value, 10 times higher than that used for feedforward synapses. Note that, in order to optimize network stabilization, an inhibitory neuron (not shown in Fig. 4), spiking according to a Poisson process (with a 80 Hz firing rate), was also connected to all other neurons.

### 4.2 Simulation Results

Python simulations of the SNN were performed on the event-based data corresponding to the second set of grating stimuli, having a spatial frequency of  $\sim 0.5$  cyc/deg and all possible orientations. It is worth noting that the movement of the sensor induces synchronized activity of DVS receptors, as also postulated for biological FEMs (Greschner et al., 2002; Kuang et al., 2012), and the spatial structure of the static image will therefore be temporally encoded in the relative activity of nearby pixels. Hence, by exploiting this attribute thanks to the designed kernels

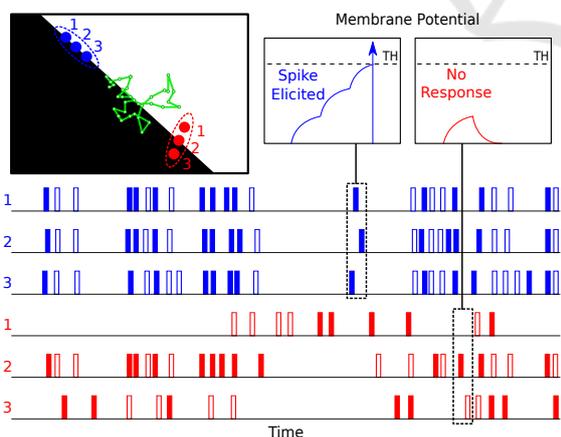


Figure 5: (Top) A cartoon illustrating the spiking dependent spatial summation of ON and OFF events. A simple visual scene is shown on the left, where 3 pixels aligned (blue) or not (red) with the edge are displayed. The FEM-like path is superimposed (green). (Bottom) An example of membrane potential in neurons taking inputs from blue or red retinal neurons is also shown: only blue pixels, thanks to their synchronized activity (for both ON/OFF polarities), are able to elicit a response on the target V1 model neuron.

and the timing properties of neurons, we can achieve optimal discrimination performances. Hence, at any given time, only the membrane potential of neurons that integrate synchronized events (i.e. from pixels that are aligned with the edge) will exceed the threshold for generating a spike, and thus an output response. The same behavior cannot be observed in neurons collecting events from pixels that are not aligned with the edge, as these activities are not synchronous. Figure 5 illustrates such a mechanism.

To present detection results of our network, we have computed the mean activity of all neurons within a single column, when exposed to a set of differently oriented gratings. Accordingly, we derived the tuning curve of such neuronal group. Figure. 6 shows, on 6 different plots, the simulation results for the same number of distinct orientation columns in both hypercolumns. As expected, the tuning curves have a peak in correspondence of the RFs  $\theta$  value. Hence, we can infer that the designed SNN is capable of a reliable estimation of local orientations. Finally, it is worth noting that the trend is very similar for the two curves in all plots. This could be due to the fact that random-like movements of a receptor across an edge trigger almost the same amount of ON and OFF events.

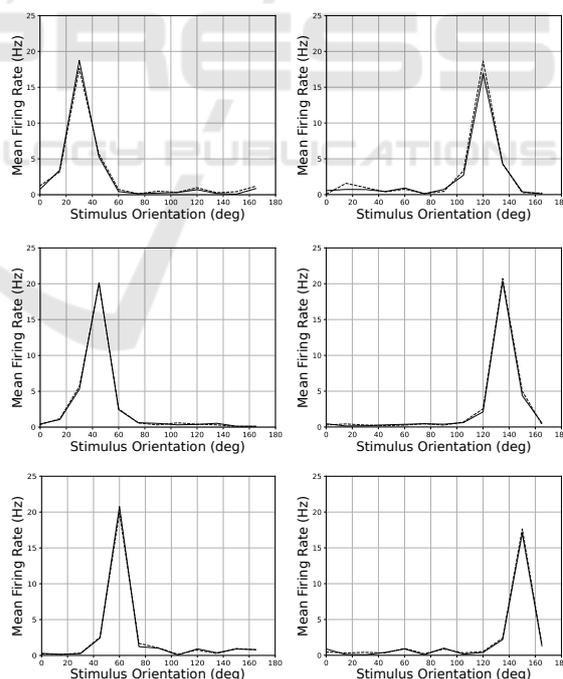


Figure 6: Resulting tuning curves. They describe, for both hypercolumns (dashed and solid black curves), the firing activity, with respect to all tested orientations of the stimuli, corresponding to 6 different orientation columns (with tuning of 30, 45 and 60 deg from top to bottom on the left, and 120, 135 and 150 deg on the right).

## 5 CONCLUSIONS

Bio-inspired fixational eye movements can transform a static scene into a spatiotemporal input luminance signal to the event-based camera. As a consequence, the low temporal frequency power of a static scene is shifted into a range that the DVS can properly detect. Besides preventing “perceptual fading” of static scenes, we show that FEMs can play a central role in event-based vision by providing an efficient strategy for acquiring and processing information from natural images, both enhancing the perception of fine spatial details in the scene, and facilitating or improving the extraction of important features. Particularly, due to camera motion, edges in the visual scene will provoke highly time-correlated activity of nearby pixels. Due to the randomness of such motion, events with both polarities can be elicited over time in each pixel as a result of a same spatial luminance discontinuity. Therefore, synchronized events with both polarities eventually encode the spatial structure of the underlying static image. The push-pull configuration, at which the network operates, exploits the distinction between events’ polarities, inducing appropriate excitation or inhibition of ON and OFF events, for optimizing detection performances. The whole artificial neural architecture proposed is fully bio-inspired, both at single unit (neuron model) and at network level, and is entirely conceived to satisfy the constraints imposed by ultra low-power mixed signal analog-digital neuromorphic processors for a future hardware implementation.

## ACKNOWLEDGEMENTS

This project has received funding from the European Research Council under the Grant Agreement No. 724295 (NeuroAgents).

## REFERENCES

- Ahissar, E. and Arieli, A. (2001). Figuring space by time. *Neuron*, 32(2):185–201.
- Bartolozzi, C. and Indiveri, G. (2007). Synaptic dynamics in analog vlsi. *Neural computation*, 19:2581–2603.
- Blakemore, C., Carpenter, R. H., and Georgeson, M. A. (1970). Lateral inhibition between orientation detectors in the human visual system. *Nature*, 228:37–39.
- Ditchburn, R. and Ginsborg, B. (1952). Vision with a stabilized retinal image. *Nature*, 170(4314):36.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Opt. Soc. of America*, 4:2379–2394.
- Gollisch, T. and Meister, M. (2008). Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866):1108–1111.
- Greschner, M., Bongard, M., Rujan, P., and Ammermüller, J. (2002). Retinal ganglion cell synchronization by fixational eye movements improves feature estimation. *Nature neuroscience*, 5(4):341.
- Hirsch, J. A. and Martinez, L. M. (2006). Circuits that build visual cortical receptive fields. *Trends in neurosciences*, 29(1):30–39.
- Hubel, D. H. and Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, 158(3):267–293.
- Indiveri, G., Linares-Barranco, B., Hamilton, T. J., Van Schaik, A., Etienne-Cummings, R., Delbruck, T., Liu, S.-C., Dudek, P., Häflicher, P., Renaud, S., et al. (2011). Neuromorphic silicon neuron circuits. *Frontiers in neuroscience*, 5:73.
- Kuang, X., Poletti, M., Victor, J. D., and Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology*, 22(6):510–514.
- Land, M. (2019). Eye movements in man and other animals. *Vision research*, 162:1–7.
- Lichtsteiner, P., Posch, C., and Delbruck, T. (2008). A  $128 \times 128$  120 dB  $15\mu\text{s}$  latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2):566–576.
- Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671.
- Martinez-Conde, S. and Macknik, S. L. (2008). Fixational eye movements across vertebrates: comparative dynamics, physiology, and perception. *Journal of Vision*, 8:1–18.
- Moradi, S., Qiao, N., Stefanini, F., and Indiveri, G. (2017). A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs). *IEEE transactions on biomedical circuits and systems*, 12(1):106–122.
- Orchard, G., Jayawant, A., Cohen, G. K., and Thakor, N. (2015). Converting static image datasets to spiking neuromorphic datasets using saccades. *Frontiers in neuroscience*, 9:437.
- Riggs, L. and Ratliff, F. (1952). The effects of counteracting the normal movements of the eye. 42:872–873.
- Rucci, M., Ahissar, E., and Burr, D. (2018). Temporal coding of visual space. *Trends in cognitive sciences*, 22(10):883–895.
- Rucci, M. and Victor, J. D. (2015). The unsteady eye: an information-processing stage, not a bug. *Trends in neurosciences*, 38(4):195–206.