

LETTER 

---

 Communicated by Terrence Sejowski**Asynchronous Event-Based Motion Processing: From Visual Events to Probabilistic Sensory Representation****Mina A. Khoei***mina.khoei@gmail.com***Sio-hoi Ieng***sio-hoi.ieng@upmc.fr**Natural Vision and Computation Team, Vision Institute, Université Pierre et Marie Curie-Paris 6 (UPMC), Sorbonne Université UMR S968 Inserm, UPMC, CHNO des Quinze-Vingts, CNRS UMRS 7210, Paris 75012, France***Ryad Benosman***ryad.benosman@upmc.fr**Natural Vision and Computation Team, Vision Institute, Université Pierre et Marie Curie-Paris 6 (UPMC), Sorbonne Université UMR S968 Inserm, UPMC, CHNO des Quinze-Vingts, CNRS UMRS 7210, Paris 75012, France; University of Pittsburgh Medical Center, Pittsburgh, PA 15213; and Carnegie Mellon University, Robotics Institute, Pittsburgh, PA 15213, U.S.A.*

In this work, we propose a two-layered descriptive model for motion processing from retina to the cortex, with an event-based input from the asynchronous time-based image sensor (ATIS) camera. Spatial and spatiotemporal filtering of visual scenes by motion energy detectors has been implemented in two steps in a simple layer of a lateral geniculate nucleus model and a set of three-dimensional Gabor kernels, eventually forming a probabilistic population response. The high temporal resolution of independent and asynchronous local sensory pixels from the ATIS provides a realistic stimulation to study biological motion processing, as well as developing bio-inspired motion processors for computer vision applications. Our study combines two significant theories in neuroscience: event-based stimulation and probabilistic sensory representation. We have modeled how this might be done at the vision level, as well as suggested this framework as a generic computational principle among different sensory modalities.

**1 Introduction** 

---

This letter presents a hierarchical event-based model for motion detection and trajectory estimation using a neuromorphic event-based artificial retina (Posch, Matolin, & Wohlgenannt, 2008). The algorithm uses the concept of motion energy filters (Adelson & Bergen, 1985), rewritten and extended to

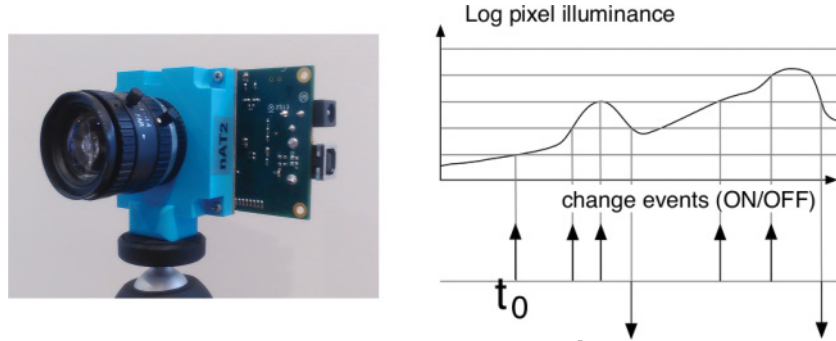


Figure 1: Operation principle of ATIS camera. The camera emits visual events when there is a significant change in the luminance of each pixel.

operate in an asynchronous time-driven framework. The goal is to explore spatial and spatiotemporal filtering of local motion signals and develop a probabilistic sensory representation of motion in the time domain, and eventually to use it for high-resolution motion estimation. During the past two decades, a vast amount of research and development has been devoted to computer vision and robotics. Many algorithms have been implemented to mimic biological vision processes, such as detecting visual signals and reconstructing basic properties of visual stimuli such as edges, shape, and motion. In general, these models and algorithms provide a structure to formulate biological vision and, consequently, apply engineering techniques toward implementing this knowledge in a machine or robot. These attempts are mainly application oriented and most often ignore important biological constraints. Biology performs asynchronous, locally independent, and frameless sampling only when and where there is a change in the scene. This results in a very fast and massively parallel sensory flow into the eye, which is to be opposed to the classical redundant in time and space-frame-based sampling used in computer vision.

The asynchronous time-based image sensor(ATIS) (Posch et al., 2008), also referred to as a silicon retina, is a neuromorphic camera that mimics luminance acquisition of the mammalian eye by recording asynchronous and parallel events only from pixels that measure significant changes in the luminance. The acquisition process of the ATIS makes it a realistic experimental platform and input layer for developing biologically inspired vision models. The camera provides streams of asynchronous visual events or artificial spikes that encode either significant increases or decreases in the visual signal, time-stamped with microsecond accuracy. This spike representation is “neuromorphic” as it is similar to biological spikes signaling (see Figure 1).

**1.1 Time-Based and Asynchronous Motion Processing.** Biological vision locates moving objects within a reasonable processing time, a process highly dependent on its spatiotemporal acuity and precision. This ability has evolved as a survival necessity for different species to fulfill basic tasks like locomotion, food hunting, and escaping predators. The visual efficiency is tightly associated with spatial and temporal resolutions by which the visual message is received in the retina and processed progressively along the neuronal hierarchy. Spatial and temporal acuities of the visual system can be quantified by various experiments, for instance, delicate catching and grasping tasks or highly interactive activities like games or sports, in which performance depends on reaction time.

The photoreceptors, as the first station of light, intricately detect changes in the luminance of visual field and pass them to the retinal ganglion cells (RGCs), whose response is known to have a high temporal precision on the order of a millisecond (Uzzell & Chichilnisky, 2004; Berry, Warland, & Meister, 1997), providing a dynamic and asynchronous spiking pattern for the downstream regions in the visual hierarchy.

There are some studies on experimental measurements of delays in different stages of visual processing (Berry, Brivanlou, Jordan, & Meister, 1999; Jancke, Erlhagen, Schöner, & Dinse, 2004; Maus, Fischer, & Whitney, 2013), as well as theoretical and modeling work to investigate various delay compensation hypotheses at single cell or population levels (Erlhagen, 2003; Kaplan, Khoei, Lansner, & Perrinet, 2014; Lim & Choe, 2008; Khoei, Masson, & Perrinet, 2017). The inputs to these models are made of frames of the stimulus sampled with an arbitrary and often slow frame rate that has no relation to the local spatiotemporal dynamics of the visual scene.

The precision of conventional frame-based motion estimation algorithms is constrained by the stimulation rate, in addition to redundant computational load, caused by the re-processing of nonchanging parts of the scene.

Data recorded with the ATIS provide an event-based stimulation layer with high-precision time stamps that put forward an asynchronous motion processing hierarchy for the visual system. The ATIS reduces drastically redundant visual inputs and encodes very transient and quick changes in the scene into so-called visual events. Each event is a computational unit defined as  $e = x, y, t, p$ , where  $[x, y]$ ,  $t$ , and  $p$ , respectively, correspond to the coordinates of the stimulated pixel, the stimulation time, and the recorded polarity (an on or off event, for either an increase or decrease of the luminance). The sensory computational framework that we introduce in this letter takes  $t$ , the *precise activation time of each pixel*, as the key input. Motion is progressively estimated from the stream of asynchronous events and updated for each incoming event. This operational principle is consistent with the response characteristics of retinal cells (with a temporal resolution of 1–10 ms), due to the independent activation of photoreceptors (Lettvin, Maturana, McCulloch, & Pitts, 1959; Berry et al., 1997; Gollisch & Meister,

2008). In addition, the ATIS generates a fast stream of visual on and off events that may be considered a simulation of retinal spikes corresponding to the on and off centered receptive fields.

**1.2 Motion Energy Filters and Spatiotemporal Filtering of Sparse Visual Events.** Photoreceptors are sensitive building blocks of the retina to capture luminance signals. Any significant change in light intensity hits the retina and elicits a response (phototransduction). This light information acquisition is then followed by a spatial filtering done by on and off center-surround retinal ganglion cells (RGCs), which have sensitivities to light and dark spots placed in their receptive fields, respectively. The output of this processing is later sent to the lateral geniculate nucleus (LGN) and cortex, where some more complex motion processing is carried on.

Motion is a spatiotemporal concept, mapped into activity of specialized motion detectors in the cortex. In this regard, the existence of Gabor-like motion-sensitive parts has been reported, and it is well known that V1 (Hubel & Wiesel, 1962, 1968) and middle temporal (MT) (Dubner & Zeki, 1971; Allman & Kaas, 1971) areas with their specialized machinery play a main role in encoding direction and speed. Classical motion energy filters were first proposed by Adelson and Bergen (1985), and since then, different variants have been widely used as key building blocks in various models of motion perception. Here, in our event-based motion detection algorithm, we have used a spatiotemporal Gabor-like filter bank similar to the one proposed in Petkov and Subramanian (2007). The stream of events from the ATIS is first spatially filtered in a layer modeling LGN cells, and then the spatiotemporal activity of this layer is passed to the Gabor kernels, simultaneously modeling speed and direction selectivity of the visual cortex. More important, our model input is (absolute) luminance free and includes only precise activation times of different locations in the visual scene. Instead of using methods like luminance conservation along with generative models to estimate velocity (Weiss, Simoncelli, & Adelson, 2002), we have used temporal sequences of events in spatiotemporal activity maps.

**1.3 From Sparse Asynchronous Visual Events to a Probabilistic Sensory Map.** One of the main questions that we address in this letter is how an asynchronous visual stimulation could be translated to a probabilistic representation of the sensory input. It is equivalent to understanding how the very precise times of receptive field activation are integrated to build a global neural code for motion. It is well known that cells corresponding to speed encoding, direction selectivity, and orientation selectivity demonstrate a dominant tendency to code one aspect of stimulus, while responding moderately to the others as well (Hubel & Wiesel, 1968). A probabilistic framework is suitable to embed such a model for neural coding. We introduce a model for transition of sparse asynchronous events from the visual scene to a probabilistic sensory map of motion. Importantly, our

event-based model captures local motion signals asynchronously and conducts feedforward spatiotemporal filtering.

**1.4 Event-Based Motion Estimation: Discrete Representation of Belief with a Set of Particles.** We present an event-based motion estimation framework that relies on a Bayesian model, in which the probabilistic sensory maps of position and velocity (motion likelihoods) are integrated with prior knowledge on smoothness of trajectory. Motion estimation has been implemented with particle filtering (condensation algorithm; Isard & Blake, 1998). Compared to the Kalman filter (Kalman, 1960), this allows discrete, asynchronous, and multimodal distributions of estimated position and velocity, which is consistent with the nature of input visual events.

## 2 Model Description

**2.1 LGN Layer Model: Event-Based Spatiotemporal Filtering.** The first stage of our model reproduces the LGN spatiotemporal filtering according to the following steps:

1. The denoising of the silicon retina's output by using an activity filter. This operation removes events that do not have any neighboring events within some temporal window. It is useful to remove spontaneous activity of the sensor, and the usual spatiotemporal window corresponds to a 3-by-3 pixel neighborhood observed when the sensor outputs some hundred of events.
2. The denoised stream of visual events then goes under a spatial integration step by a layer simulating a simple model of LGN. This layer is composed of ON center gaussian receptive fields of a certain size. As shown in Figure 2A, visual events representing the length of bar are mapped to a grid layer of receptive fields, and this activity trace (active pixels displayed for some short duration) is shifted as the bar moves. Figure 3 illustrates the spatiotemporal activity map for events registered from a vertically moving bar. In a very short temporal interval (140 millisecond), collective activity caused by events leaves a continuous trace on this layer.
3. For each  $i$ th event  $e_i(x_i, y_i, t_i, p_i)$ , we can define the corresponding receptive field at time  $t_i$  as

$$\begin{cases} RF_{e_i}(x, y) = \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{\sigma^2}\right) \\ g_i(t) = \exp\left(-\frac{(t_i-t)^2}{\tau_{cl}^2}\right) \end{cases}. \quad (2.1)$$

Every registered event creates a gaussian activity around corresponding  $(x_i, y_i)$ , which decays by the factor of  $\tau_{cl}$ . As the scene changes, the set of events that will feed the motion estimation is

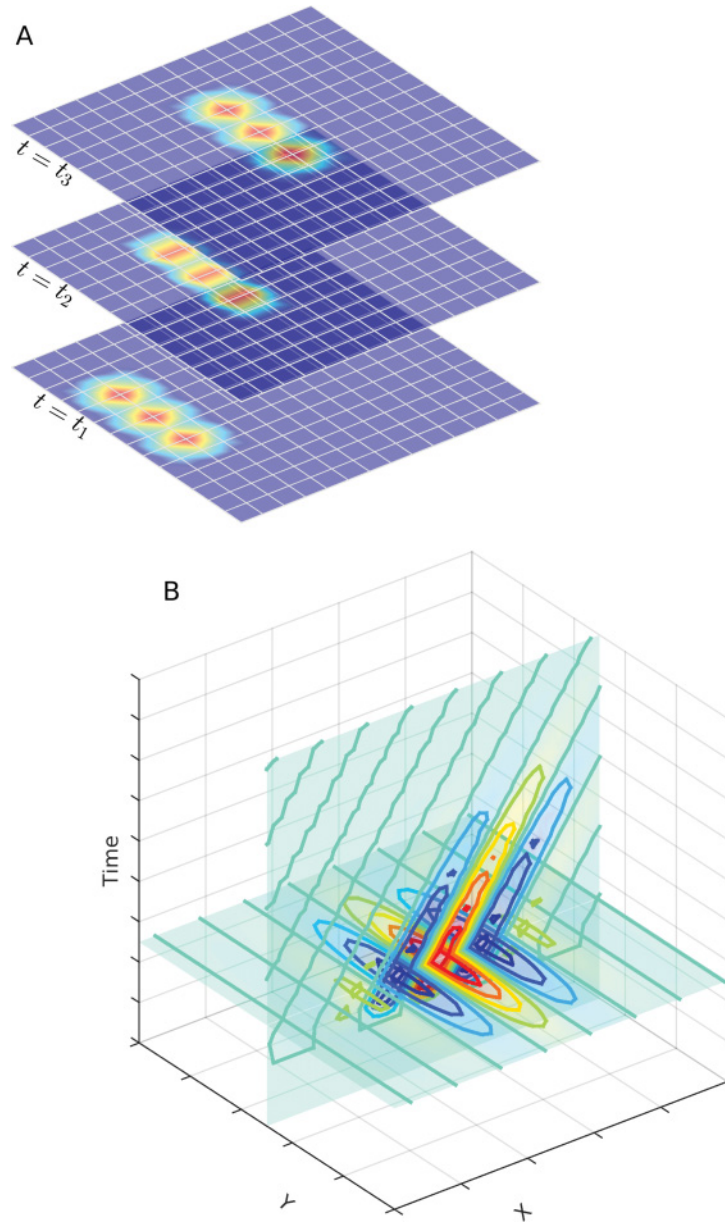


Figure 2: Schematic of the model. (A) Spatial filtering by gaussian receptive fields of LGN layers as the response to a moving bar. Three layers correspond to activated positions at three successive times. The colors from blue to red represent the normalized activity. (B) The spatiotemporal filtering (V1 + MT model).



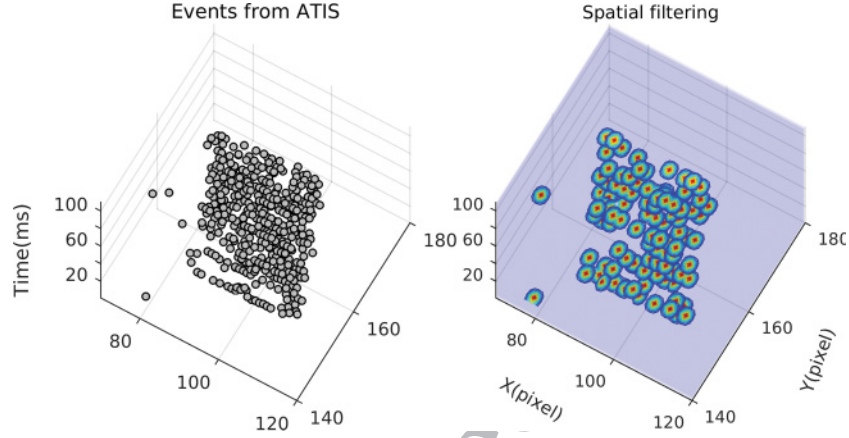


Figure 3: (Left) One thousand ATIS events during 140 milliseconds. (Right) Spatial filtering of events. Each event elicits a gaussian activity in space with standard deviation of 1 pixel.

enclosed into the volume:

$$C_{e_i} = \iiint RF_{e_i}(x, y) dx dy \int g_i(t) dt. \quad (2.2)$$

$C_{e_i}$  is a spatiotemporal volume that demonstrates the activity of the LGN layer, mainly created by a cloud of events recorded between  $(t_i - \tau_{cl}, t_i)$ , where the weights of each event are defined by the temporal gaussian. In theory, a big enough  $\tau_{cl}$  will integrate all activated spatial gaussians from  $t = 0$  until the present, where the contribution of older events (closer to  $t = 0$ ) will be smaller.

**2.2 Spatiotemporal Energy Model.** Inspired by Hubel and Wiesel (1962, 1968), Adelson and Bergen (1985) proposed several spatiotemporal energy models for motion perception. Since then, a large number of studies have used various versions of these models to interpret neurophysiological data or for purely theoretical purposes in computer vision applications. For instance, in Hubel and Wiesel (1962), the function of simple cells in cat striate cortex is expressed by a mathematical formulation of 2D Gabors. This has been validated by reasonable accordance of data from receptive fields, with multiple tests to characterize the spatial frequency properties of receptive fields. The spatiotemporal filtering method used in this work is adopted from Petkov and Subramanian (2007), originally proposed as an efficient motion computation algorithm with frame-based inputs. We use three-dimensional spatiotemporal Gabor kernels, described by equation 2.3, to detect the direction and speed of the moving stimulus from a stream of

visual events recorded by an ATIS camera,

$$\begin{aligned}
 K_{v,\theta,\phi}(x, y, t) = & \frac{\gamma}{2\pi\sigma^2} \exp\left(\frac{-((\bar{x} + v_c t)^2 + \gamma^2 \bar{y}^2)}{2\sigma^2}\right) \\
 & \cdot \cos\left(\frac{2\pi}{\lambda}(\bar{x} + vt) + \phi\right) \\
 & \cdot \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(t - \mu_t)^2}{2\tau^2}\right) \cdot U(t), \tag{2.3}
 \end{aligned}$$

where

$$\begin{cases} \bar{x} = x \cos(\theta) + y \sin(\theta) \\ \bar{y} = -x \sin(\theta) + y \cos(\theta) \\ \lambda = \lambda_0 \sqrt{1 + v^2} \\ 0.3 < \frac{\sigma}{\lambda} < 0.6 \end{cases}.$$

This formula generates filters like the ones illustrated in Figure 2B as a simple model of cortical motion processing.

Here we give an intuitive description of parameters of equation 2.3. This equation describes a family of spatiotemporal Gabor kernels to simulate velocity detection in the visual system, and it includes multiplication of two gaussian functions (one spatial and one temporal) and a cosine function. The kernel is selective to a specific direction of motion represented by  $\theta$ , and the preferred spatial orientation of the filter is perpendicular to that. The spatial gaussian is defined in  $(x, y)$  coordinates, and its center is moving with time at the speed of  $v_c$ . The ratio of its extension between  $x$  and  $y$  is controlled with  $\gamma$ , and  $\sigma$  determines its size (standard deviation of the spatial gaussian). The second term of equation 2.3 is a cosine function moving with the speed  $v$ , its spatial frequency and phase are defined by  $\lambda$  and  $\phi$ , respectively, to add the excitatory and inhibitory lobes to the receptive field. Finally a temporal gaussian (with  $\mu$  and  $\tau$  as mean and standard deviation) is multiplied in the kernel to simulate temporal integration of the receptive field. To remain in a biologically plausible range, as suggested by Petkov and Subramanian (2007), we have chosen the spatial frequency and the spatial size of filters in order to satisfy the ratio ( $0.3 < \frac{\sigma}{\lambda} < 0.6$ ).

The motion energy of each kernel to a certain stimulus is calculated as the convolution of  $C_{e_i}$  with the energy of a kernel, where  $C_{e_i}$  is the spatiotemporal volume created by the activity of a set of events as described in equation 2.2.

Depending on the speed and dynamics of the scene, the time constant  $\tau_{cl}$  is adjusted to gather smoothly and continuously events that are 1 to tens of milliseconds old with regard to  $e_i$ . Then the event-based response is defined as

$$E_{eb}(v, \theta, t, \phi) = |C_{e_i} * K_{v,\theta,\phi}(x, y, t)|^+, \tag{2.4}$$



where  $|\cdot|^+$  is a rectification operator keeping only positive lobes of the energy, and  $*$  represents convolution. The final motion energy at a certain speed and direction is calculated as the root of sum squared energy values for filter phases of 0 and  $\pi/2$ :

$$\bar{E}_{eb}(v, \theta, t) = \sqrt{(E_{eb}^2(v, \theta, t, 0) + E_{eb}^2(v, \theta, t, \pi/2))}. \quad (2.5)$$

$\bar{E}_{eb}(v, \theta, t)$  is called phase-insensitive motion energy and is considered a model of the response of complex cells in the early visual cortex (Adelson & Bergen, 1985).  $\bar{E}_{eb}$  is updated with every coming event, integrating the sensory information carried by the events enclosed in  $C_{e_i}$ . For a stimulus moving with a certain direction and speed, at arbitrary time  $t_i$ , the energy of one filter in the filter bank (defined in a range of  $\theta$  and  $V$ ) is significantly higher than others. As such, the detected speed and direction are defined as

$$\begin{cases} s_d = \operatorname{argmax}_v \bar{E}_{eb}(v, \theta, t) \\ \theta_d = \operatorname{argmax}_\theta \bar{E}_{eb}(v, \theta, t) \end{cases}. \quad (2.6)$$

**2.3 Bayesian Motion Estimation.** Generic probabilistic frameworks have been successfully used for motion estimation in various contexts (Burgi, Yuille, & Grzywacz, 2000; Perrinet & Masson, 2012; Weiss et al., 2002). As such, planar motion of an object in the visual space at a given time can be described by the probability distribution, which reflects the distribution of our value of belief about its position and velocity. For instance, noise and other perturbations add uncertainty to the measurement of motion and shift the probability from 1 toward 0, leading to a wider distribution around the peak.

This representation is compatible with Bayesian models, which optimally integrate the prior information with measured motion, to compute posterior probability. In this framework, likelihoods are defined as a measure of belief, knowing that the measurements (sensory data) and prior distribution include any information about the motion before measurement. Ultimately, in a Bayesian model, a decision-making step is often relied on to choose the maximum of posterior probability. The majority of previously proposed models for Bayesian motion estimation rely on the luminance conservation equation to generate an approximated likelihood distribution as a gaussian distribution. The key difference of our Bayesian inference is that there is no luminance information in the sensory data; therefore, luminance-based methods cannot be applied for likelihood generation. Instead, the spatiotemporal structure of the stimulus is captured by a simple model of LGN + V1 (motion energy detectors), and motion likelihood is created based on normalized activity of this layers.

**2.3.1 Event-Based Detection of Motion.** As described, the stream of asynchronous events after spatial and spatiotemporal filtering is converted to probabilistic maps, reflecting the belief on the position and the velocity of the stimulus (see Figure 3). For each stimulus, a spatiotemporal volume ( $C_{e_i}$  in equation 2.2) is convolved with a bank of Gabor kernels covering a range of speeds = {12.5, 25, 50, 100, 200} (pixel/second) and directions = {0°, 45°, 90°, 135°}. The filter bank used for the spatiotemporal filtering is composed of Gabor kernels described by equation 2.3. Different kernel sets are used for speed and direction measurements. For speed measurement, stimuli are filtered with a filter bank with preferred direction matched to the physical direction of stimulus and a range of preferred speeds. For direction measurement, the preferred speed of the filters is kept constant, and filters are sensitive to a range of directions.

The motion energy response of each filter is a spatiotemporal volume  $E_{eb}$ , the result of the three-dimensional convolution shown by equation 2.4, and is updated with every event. For an arbitrary  $C_{e_i}$  at  $t = t_i$ , there will be one filter with the highest response, encoding the instantaneous speed and direction of the stimulus. Figures 5 and 6 illustrate the integral of filter energies over their activity time, leading to a 2D activity in  $x$ - $y$  space).

Measured motion by filter banks (sensory information) reflects the belief in the speed and direction of the stimulus, represented by  $\mathbf{S}_t$  in the Bayesian framework. In order to generate a probabilistic distribution out of sensory information (motion likelihoods), the response of each filter is represented by the integral value of  $E_{eb}$  for values higher than a threshold. The choice of this threshold is critical for speed decoding and, as illustrated in Figure 5, filters corresponding to speeds higher than the speed of the stimulus may have a comparable sum of energy. This is caused by the larger size of these filters, and despite having a small overlap with  $C_{e_i}$ , the motion energy can be a moderate response distributed over a larger area. As such, defining a threshold value can eliminate any unfair contribution of these filters in final probabilistic maps. In this letter, we have used a threshold of 60%, meaning that for each filter, response values higher than 60% of maximum (of normalized response of filter bank) are summed and the value represents that speed in probability distribution functions (pdf) for speed. The threshold's value has been chosen empirically. As we have studied direction coding over a filter bank with the same preferred speed (same spatial size; see Figure 6), however, using a threshold is not necessary.

To estimate the instantaneous state of motion (position and velocity of the stimulus), we are using the following Bayesian motion inference:

Estimation:

$$P(\mathbf{M}_t | \mathbf{S}_{0:t}) = \frac{P(\mathbf{S}_t | \mathbf{M}_t) P(\mathbf{M}_t | \mathbf{S}_{0:t-\delta t})}{P(\mathbf{S}_t | \mathbf{S}_{0:t-\delta t})} \quad (2.7)$$

Prediction:

$$P(\mathbf{M}_t | \mathbf{S}_{0:t-\delta t}) = \int P(\mathbf{M}_t | \mathbf{M}_{t-\delta t}) P(\mathbf{M}_{t-\delta t} | \mathbf{S}_{0:t-\delta t}) d\mathbf{M}_{t-\delta t}$$

$$\text{where } \begin{cases} \mathbf{M}_t = (x_t, y_t, s_t, \theta_t) \\ \mathbf{S}_t = (\hat{x}_t, \hat{y}_t, \hat{s}_t, \hat{\theta}_t) \\ \mathbf{S}_{0:t} = \{\mathbf{S}_0, \dots, \mathbf{S}_t\} \end{cases},$$

in which the estimated and measured motion states are represented by  $\mathbf{M}_t$  and  $\mathbf{S}_t$ , respectively.  $\mathbf{M}_t$  is estimated from the last occurring event  $e_i$ , and  $\mathbf{M}_{t-\delta t}$  is estimated from the precedent event  $e_{i-1}$ , while  $\delta t$  is the elapsed time between the two. Thus,  $P(\mathbf{S}_t | \mathbf{M}_t)$  is the event-based probabilistic representation of sensory motion (likelihoods) that includes the normalized probabilistic responses of filters in position and velocity spaces.

Equation 2.7 is the master equation of our model to implement the event-based predictive coding in a generic probabilistic framework. Similar frameworks have been used for motion estimation with emphasis on the temporal coherency (Burgi et al., 2000), the aperture problem (Perrinet & Masson, 2012) or to explain visual illusions (Khoei et al., 2017). The main difference of our implementation is that the current motion estimation framework is asynchronous and time based and relies on luminance-free computations. So instead of  $I_t$  as sensory information, our framework builds and uses sensory maps based on asynchronous activation of spatial and spatiotemporal filters. In other words, the registration of a single event in a position elicits a trace of probabilistic activity, independent of the absolute luminance quantity. Therefore, at every millisecond, depending on the binary status of the visual scene (presence or absence of event somewhere), the motion state is either updated or preserved.

**2.3.2 Generative Model of Trajectory.**  $P(\mathbf{M}_t | \mathbf{M}_{t-\delta t})$  in equation 2.7 represents the probabilistic transition between two arbitrary motion states in adjacent times, described by a Markov chain. According to our prior knowledge on the rigidity of physical objects and smoothness of motion trajectories in natural scenes, one can build a generative model of motion transport in the visual scene. This model will make the prediction on state transitions:

$$\begin{aligned} \mathbf{X}_{t+\delta t} &= \mathbf{X}_t + \mathbf{V}_t \cdot \delta t + \mathbf{v}_X, \\ \mathbf{V}_{t+\delta t} &= \mathbf{V}_t + \mathbf{v}_V, \\ \text{where } \mathbf{X} &= (x, y), \quad \mathbf{V} = (s, \theta), \\ \mathbf{v}_X &= (v_x, v_y), \quad \mathbf{v}_V = (v_s, v_\theta). \end{aligned} \tag{2.8}$$

$\mathbf{v}_X$  and  $\mathbf{v}_V$  are blur factors or the standard deviation of the gaussian noise in the smooth trajectory model; they control the precision of motion prediction. This generative model is updated with every occurring event

and feeds the predictive term into motion integration defined by equation 2.7.

*2.3.3 Implementation with the Condensation Algorithm.* We have used the condensation algorithm (Isard & Blake, 1998) as a discrete approximation method to implement the Bayesian motion estimation introduced by equation 2.7. It is a particular type of particle filter in which probabilistic distributions are replaced with  $N$  particles, where each particle  $p_i$  contains a position  $(x_i, y_i)$ , a speed  $s_i$ , and a direction  $\theta_i$  and represents a hypothesis on the motion state.  $P(\mathbf{M}_t | \mathbf{S}_{0:t})$  can be approximated by  $\hat{P}(\mathbf{M}_t | \mathbf{S}_{0:t})$ , where

$$\hat{P}(\mathbf{M}_t | \mathbf{S}_{0:t}) = \sum_{i=1}^{i=N} w_i \cdot p_i,$$

$$p_i = \delta(\mathbf{X} - (x_i, y_i), \mathbf{V} - (s_i, \theta_i)). \quad (2.9)$$

$\delta$  is the Dirac function, and  $w_i$  is a positive weight associated with each particle, reflecting the value of belief for that hypothesis based on sensory measurements and  $\sum_{i=1}^N w_i = 1$ .

The algorithm has these steps:

1. *Initialization.*  $N = 2^{11}$  particles are initialized at time  $= 0$  when  $x, y, s$ , and  $\theta$  are randomly distributed in a defined range and a normalized equal weight is given to all particles. The distribution range for  $x$  and  $y$  corresponds to the resolution of ATIS after calibration, while direction and speed are randomly sampled from  $(0^\circ, 135^\circ)$  and  $(0, 3\hat{V})$ , respectively, where  $\hat{V}$  is the vertical speed of the stimuli (see section 3).
2. *Drift.* Particles are drifted with their velocity, as stated in equation 2.8. In the first step, they will be drifted with their randomly attributed speed and direction; in the next steps, the drift is done after evaluation and weighting of particles based on measured motion.
3. *Diffusion.* The position and velocity of particles are diffused according to equation 2.8. Blur factors have been set to  $v_X = 5$ ,  $v_s = 0.015$ , and  $v_\theta = 10$ . At the end of this step, particles still have equal weights.
4. *Measurement and weighting.* Each single hypothesis carried by particles is evaluated based on motion likelihoods (measured position and velocity from the LGN layer and motion energy filter banks). The position, speed, and direction of each particle are matched with normalized probability distributions  $P(\mathbf{S}_t | \mathbf{M}_t)$  in equation 2.7, and relative weights are updated accordingly.  $P(\mathbf{S}_t | \mathbf{M}_t)$  includes position and velocity likelihoods, normalized activity of the LGN layer, and normalized response of motion energy filters, respectively. For instance, if the position and velocity of a particle are not supported by sensory observations, the dedicated weight will be zero and the corresponding hypothesis will be eliminated.

5. *Resampling.* Finally, updated particles are resampled. To keep the total number of particles constant, particles carrying highly probable hypotheses are replicated to replace discarded ones.

### 3 Results

In this section, we present the results of an event-based model for trajectory processing. We have tested our model with two experiments using the following stimuli. The first set is composed of four moving bars with orientations  $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ , moving orthogonally to their orientations, with a ground-truth speed of 50 pixels per second. The second data set is obtained while the ATIS rotates clockwise about a fixed axis, scanning a stationary planar pattern. For this setup, the ground-truth speed is provided by a motion capture system.

The input to the model is a continuous stream of events recorded by the ATIS. After preprocessing of ATIS events (described in the previous section), the events are integrated in space briefly (see equation 2.1, the layer model of the LGN). As illustrated in Figure 3B, receptive fields are defined as gaussian blobs with  $\sigma = 1$  pixel and 1 ms of resolution. Some successive events occupy the same position as previous ones, and this can be seen in the accumulative trace of activated receptive fields.

#### 3.1 Data Set 1

*3.1.1 Stimulus Preparation and Recording.* Our stimuli are composed of a set of moving bars with the following orientations,  $0^\circ, 45^\circ, 90^\circ$ , and  $135^\circ$ , where each bar is moving in the direction perpendicular to its orientation. Bars are displayed by a digital micromirror device (DMD) projector at 1000 frames per second and recorded by ATIS camera, as depicted in Figure 4.

*3.1.2 Calibration and Ground Truth of the Stimuli.* Stimuli have been generated with 1 ms temporal resolution and were projected with a DMD and recorded with the ATIS ( $304 \times 240$  pixel array). The speed of projected bars on the screen was  $V \approx 62$  (pixel/second). To find the characteristics of the motion perceived by the ATIS, we did a calibration by mapping four corners and the center of the screen on the camera. The calibration was done by finding the homography transform between the scene and the ATIS's focal plane. We were able to map the start and end points of the motion trajectory and calculate the speed of the stimuli on the ATIS:  $\hat{V} \approx 50$  (pixel/second). Both speeds here are rounded to the lower integers.

The filter bank used for the spatiotemporal filtering is composed of Gabor kernels described by equation 2.3. Different kernel sets are used for speed and direction measurements. For speed measurement, stimuli are filtered with a filter bank with preferred direction matched to the physical

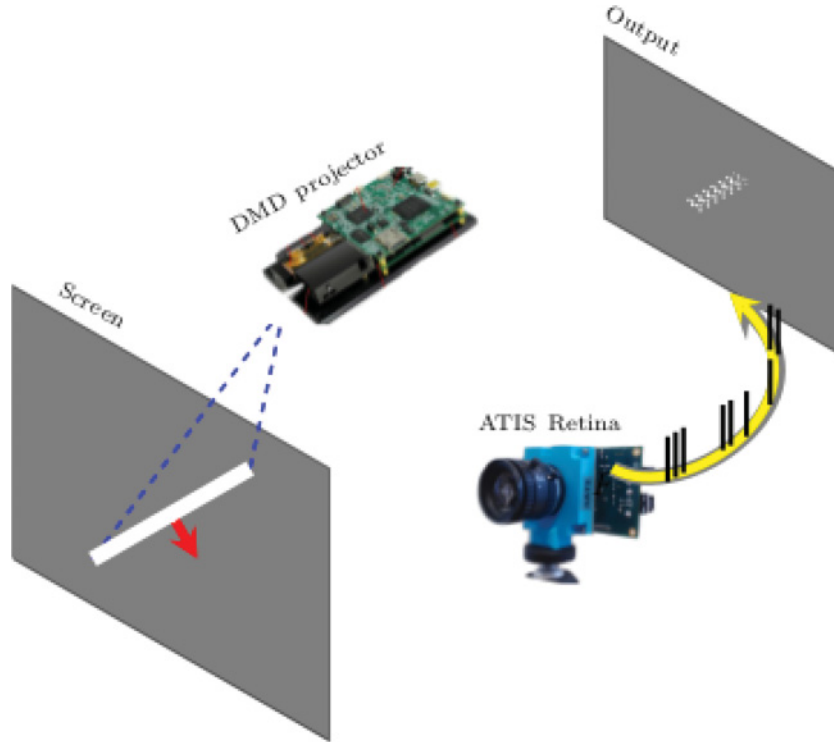


Figure 4: Experimental setup. Moving bars are projected on a screen in a dark room by a DMD projector with 1 ms resolution and recorded by ATIS camera.

direction of stimulus and a range of preferred speeds. For direction measurement, the preferred speed of the filters is kept constant, and filters are sensitive to a range of directions.

**3.1.3 Event-Based Speed Selectivity.** The filter bank covers a range of speeds = {12.5, 25, 50, 100, 200} (pixels/second), which are selected around the ground-truth speed of the stimuli ( $\approx 50$  pixels/second), and the direction of filter is matched to the physical direction of the stimulus. Figure 5 illustrates the responses of the filter bank in five preferred speeds and for all stimuli. Each row shows the resultant motion energy of the filter bank to one stimulus ( $\tau_{cl} = 50$  ms). The filters' responses are updated by every event. According to equation 2.3, the filters' spatial frequency,  $\lambda$ , grows with its preferred speed. Consequently, to respect the physiological constraints on the ratio of  $\sigma$  and  $\lambda$  (see equation 2.3), the filter needs to be more spatially elongated (larger  $\sigma$ ). This is visible in Figure 5, where larger energy traces are generated by kernels preferring higher speeds. For all stimuli,



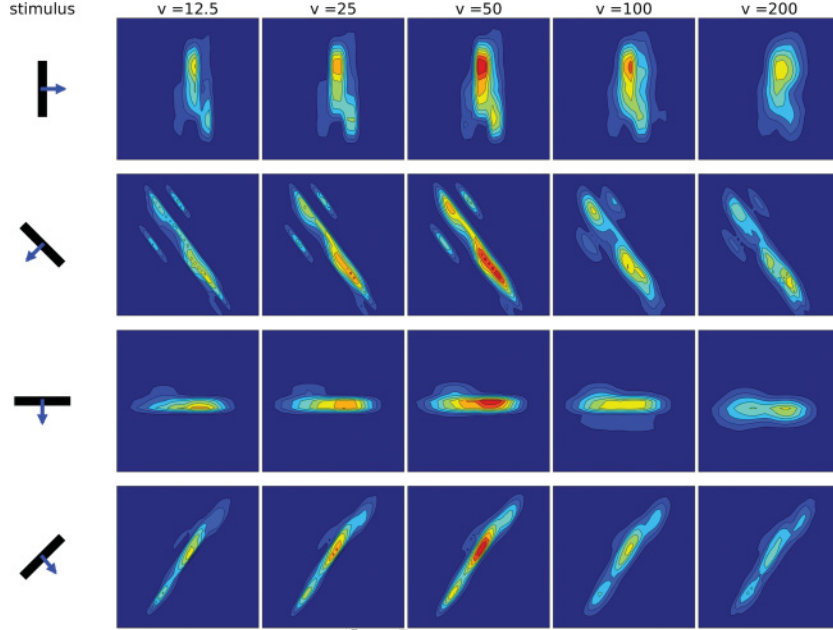


Figure 5: Detected speed by motion energy filters. Each row illustrates the response of filters with a range of preferred speeds, where preferred direction is matched to the motion direction of the stimulus. Stimuli in all directions move at about 50 pixels per second, and the illustrated motion energy is accumulated for a part of trajectory ( $\tau_{cl} \approx 50$  ms). The response has been normalized to the highest response of filter bank to each stimulus.

the filter with the preferred speed matched to the physical speed of the stimulus (middle column) demonstrates significantly higher and more robust responses. The geometrical parameters of the filter bank have been set according to the size and the physical properties of the bar seen by the visual field of the ATIS, after calibration explained in section 2, ( $\gamma = 0.2$ ,  $\lambda_0 = 1.4$ ,  $\tau = 20$ ,  $t_0 = 10$ ), respecting biological restrictions discussed in (Petkov and Subramanian, 2007).

**3.1.4 Event-Based Direction Selectivity.** For direction measurements, the filter bank used covers directions =  $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ , where the preferred speed is fixed to the physical speed of the stimuli ( $\approx 50$  pixels/second). Similar to speed computations, here, as well the spatial filtering, is performed in an event-based manner ( $\tau_{cl} = 50$  ms). Figure 6 illustrates the responses of the filter bank to all stimuli. For each stimulus, the filter with matched direction shows stronger and more robust motion energy response.

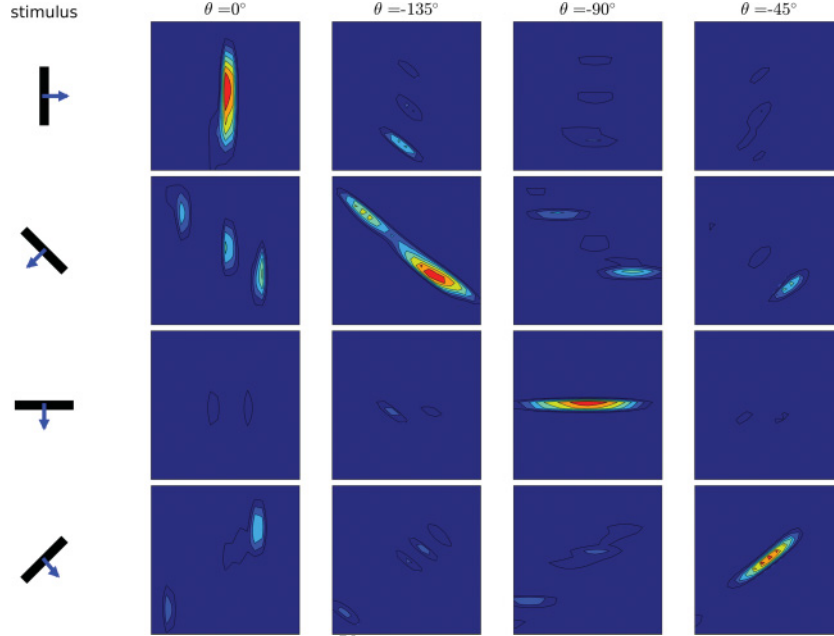


Figure 6: Detected direction by motion energy filters. Each row illustrates the responses of filters tuned to a range of directions and preferred speed matched to the speed of the stimulus. Responses are normalized to the highest response of the filter bank for each stimulus, and the illustrated motion energy is accumulated for a part of trajectory ( $\tau_{cl} \approx 50$  ms).

**3.1.5 Cloud Size and Intercloud Dynamics.** As described and according to equation 2.2, an event-based spatiotemporal filtering is performed on an event cloud  $C_e$ , described in equation 2.2. In a set of simulations, we investigated the role of cloud size,  $\tau_{cl}$ , on decoded speed and direction. Figures 7 and 8 illustrate the normalized response of speed and direction filter banks for  $\tau_{cl} = [12.5, 15, \dots, 50]$  ms. As shown in Figures 7 and 8, in most cases,  $\tau_{cl} = 12.5$  is enough for the emergence of decoded velocity. The minimum number of events ( $\tau_{cl}$ , equivalent to the temporal integration window in equation 2.2) for reliable motion coding depends on the speed and dynamics of the scene, and integration over bigger clouds increases the quality of velocity discrimination. In the results of this letter, we have kept  $\tau_{cl} = 50$ .

**3.1.6 Estimated Trajectory.** Figure 9 illustrates the estimated position for four moving bars with the condensation algorithm. According to equations 2.7 and 2.9, the estimated motion is approximated by a set of particles distributed spatially, with a certain velocity associated with each. In Figure 9,

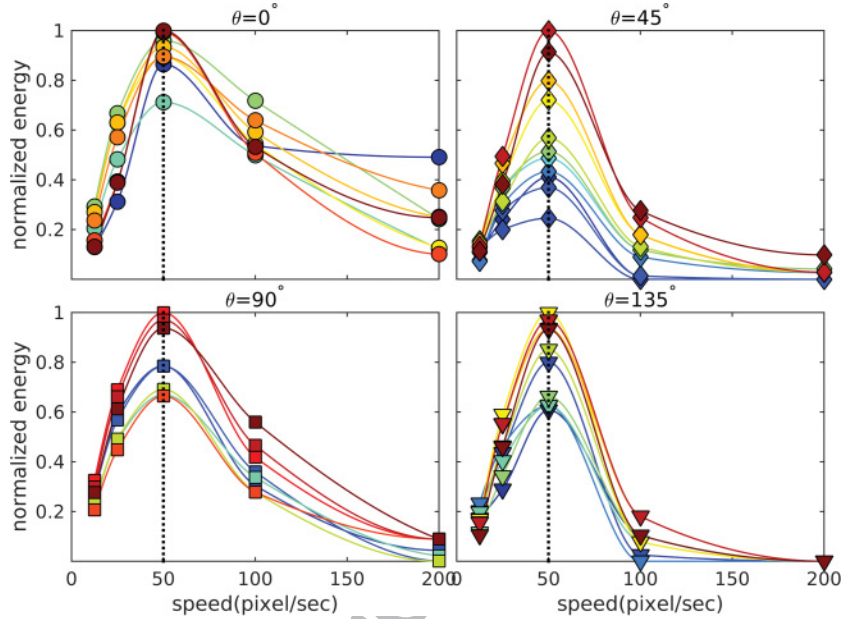


Figure 7: Intercloud dynamics of event-based speed coding for motion energy responses shown in Figure 5. Each plot has 16 traces showing the progressive emergence of detected speed by increasing  $\tau_{cl}$  from 12.5 ms to 50 ms, where the coldest and warmest colors represent the smallest and largest values, respectively. Filled points are normalized motion energy response of filter bank with preferred speeds = 12.5, 25, 50, 100, 200 (pixels/second), and the preferred direction of each filter has been set to the physical direction of each stimulus.

the estimated trajectory has been shown for a very short displacement of bars. The trajectory's duration is approximately 50 to 75 ms for all stimuli. The estimated trajectory is updated with every event and based on position and velocity likelihoods on a measured cloud  $C_{e_i}$  ( $\tau_{cl} = 50$ ). For visualization purposes we have illustrated only 1 out of every 20 estimations in Figure 9.

**3.1.7 Estimated Direction.** In Figure 10 we show the development of estimated direction for a very short displacement composed of approximately 200 events. For each stimulus, the histogram of estimated directions is illustrated at 10 estimation steps between the first and the last event, represented with cold to warm colors. For all stimuli, the distributed histogram at an early trajectory is quickly converging to a coherent estimation around the physical direction of the bars.

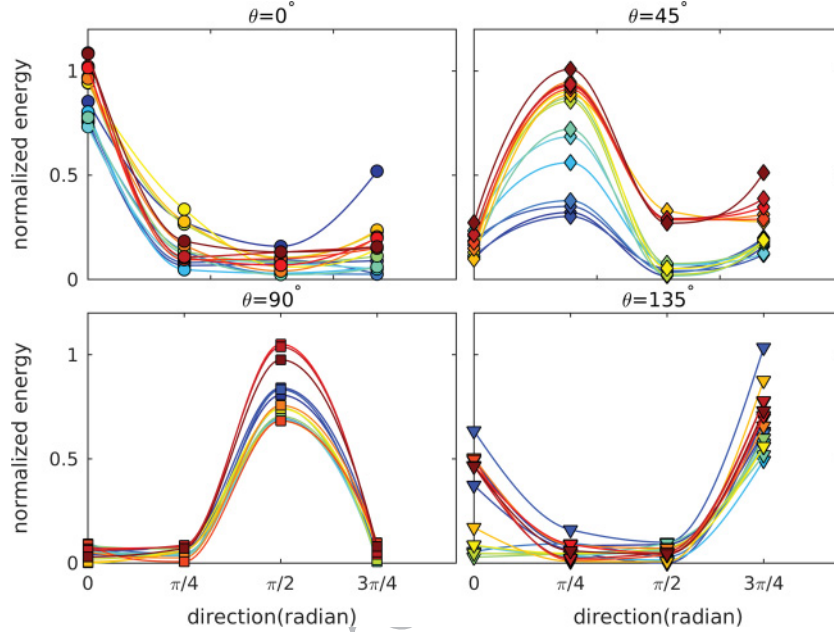


Figure 8: Intercloud dynamics of event-based direction coding for motion energy responses shown in Figure 6. Each plot has 16 traces showing the progressive emergence of the detected direction by increasing  $\tau_{cl}$  from 12.5 ms to 50 ms, where the coldest and warmest colors represent the smallest and largest values, respectively. Filled points are the normalized motion energy response of the filter bank with preferred directions  $\phi = 0, \pi/4, \pi/2, 3\pi/4$ , and the preferred speed of each filter has been set to the physical speed of each stimulus.

### 3.2 Data Set 2

**3.2.1 Stimulus Recording.** Our second data set is recorded by a clockwise rotating ATIS, scanning a stationary plane, while the camera is tracked by an OptiTrack motion capture system, consisting of eight Flex13 cameras. The motion capture has a submillimeter spatial resolution and an acquisition frequency of 120 Hz and provides the ground-truth speed. All technical specifications, including information about accuracy, can be found on the web page OptiTrack Documentation Wiki ([https://v21.wiki.optitrack.com/index.php?title=OptiTrack\\_Documentation\\_Wiki](https://v21.wiki.optitrack.com/index.php?title=OptiTrack_Documentation_Wiki)).

Tracking by the ATIS camera is done by putting some markers on its casing (see Figure 11). The world reference frame is by default the OptiTracks frame, and as the camera moves, the trajectory of casing can be updated in real time. The ground-truth measurement of the speed is based on the following information:

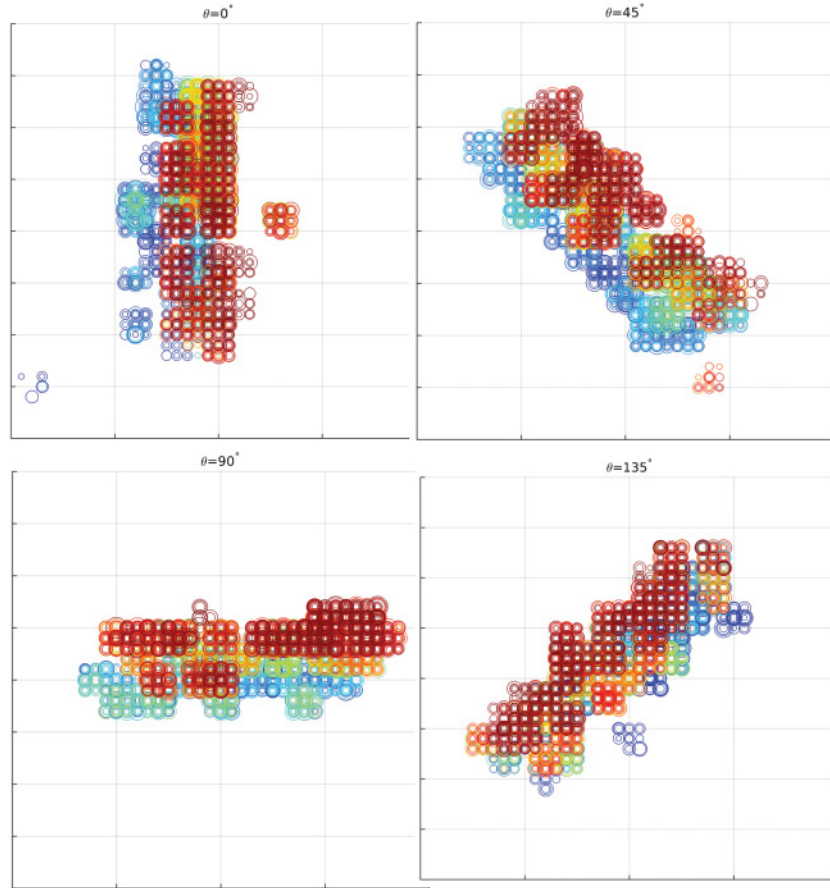


Figure 9: Event-based trajectory estimation with the condensation algorithm. The trajectory is composed of 400 events, equivalent to 75, 51, 60, and 50 milliseconds, respectively for stimuli moving in directions  $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$  and speed  $\approx 50$  pixels/second. In the motion estimation algorithm, the estimated position is updated by every event, but for visualization here, we have illustrated only 20 estimations (1 in 20 out of 400 estimations). Vertical and horizontal units are 5 and 10 pixels, and hue represents the time from the beginning of estimation, where, by advancing in the trajectory, the color of estimated positions changes from blue to red. Circle size is proportional to the number of estimated particles in the center of the circle.

- The camera casing pose,  $(R_{ca}, T_{ca})$ , provided by Optitrack
- The planar pattern pose,  $(R_p, T_p)$ , provided by Optitrack
- The camera pose,  $(R_c, T_c)$ , which is not provided by Optitrack, and has to be updated as the camera moves.

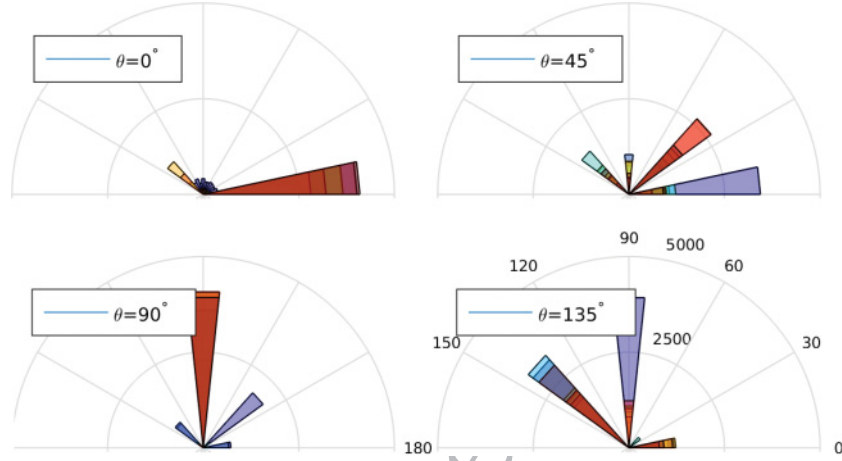


Figure 10: Event-based direction estimation with the condensation algorithm. Motion estimation is updated by every event, and each plot illustrates the improvement of estimated direction by advancing in the trajectory that corresponds to the first half of the trajectory in Figure 9 (composed of 200 events). For better visualization, we have illustrated only 10 directional histograms for each stimuli—1 in every 20 estimation. Hue represents the relative temporal order. Histograms in the coldest and warmest colors are distributed in the estimated direction by events number 1 and number 200, respectively.

A pose pair such as  $(R_{ca}, T_{ca})$  is defined relative to the world reference frame. If a point has coordinates  $X$  in the world reference frame, the coordinates of the point in the camera casing frame are given by

$$R_{ca}X + T_{ca}.$$

**3.2.2 Speed and Direction Detection.** Figure 11 illustrates the detected speed and direction for a short trajectory, including stream of events from a clockwise rotating ATIS, scanning mostly vertical edges. The filter bank used for data set 2 covers a range of speeds =  $\{0, 30, 60, 90, 120\}$  (pixels/second) which are selected around the ground-truth speed of the stimuli ( $\approx 60$  pixels/second), and the direction of filters is matched to the physical direction of the stimulus (rightward). In the figures, each trace corresponds to the normalized energy of filters updated by every coming event. The ground-truth speed provided by OptiTrack is 59 pixels per second for this part of the trajectory, and in all 400 consecutive events, the filter matched to 60 pixels per second is the most responsive one.

For direction measurements, the used filter bank covers directions =  $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ , where the preferred speed is fixed to the physical speed



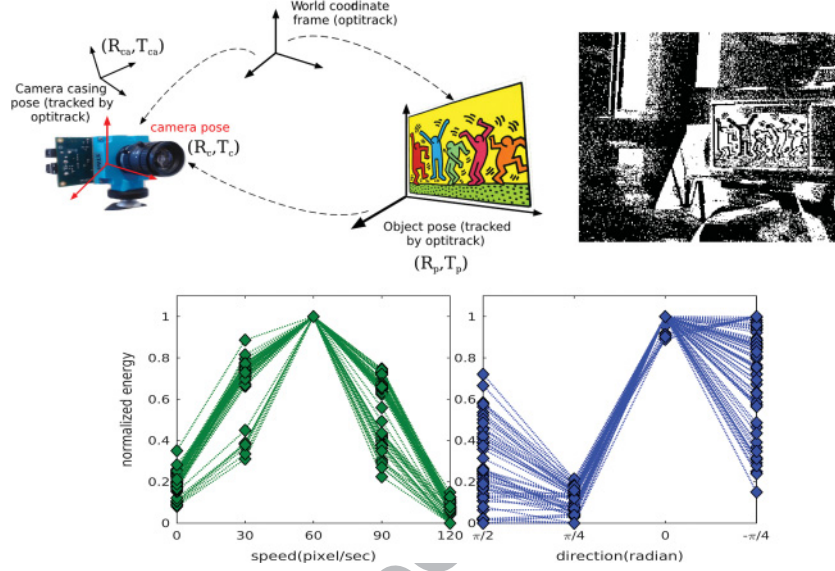


Figure 11: (Top) Tracking setup by OptiTrack is illustrated in the left panel (see section 3.2 for details). One snapshot of ATIS events, resulting from a clockwise rotating camera on a stationary pattern is shown at right. (Bottom) Detected speed and direction by motion energy filters. Each figure includes the normalized energy of filters at 400 consecutive events, while the ground-truth speed provided by OptiTrack is 59 pixels per second for that part of the trajectory. The most stable detected direction during the trajectory is rightward ( $0^\circ$ ). Because a significant edge  $\approx 45^\circ$  is also present in the scene, in some parts of the trajectory the direction  $-45^\circ$  competes with rightward motion due to the aperture problem.

of the stimuli ( $\approx 60$  pixels per second). The normalized energy of filters for direction has a persistent peak in the rightward direction ( $0^\circ$ ), decoding the motion direction of the small board (planar pattern), as a result of ATIS rotation. Note that at some parts of the trajectory, a direction of  $-45^\circ$  is also decoded with comparable energy. As shown in the ATIS snapshot, an edge of  $\approx 45^\circ$  is present in some parts of the trajectory, and because our filters are prone to an aperture problem, an orthogonal direction is detected.

The geometrical parameters of the filter bank are the same as for data set 1, explained in section 2, ( $\gamma = 0.2$ ,  $\lambda_0 = 1.4$ ,  $\tau = 20$ ,  $t_0 = 10$ ,  $\tau_{cl} = 50$ ) ms.

#### 4 Conclusion

Event-based computation introduces a novel research line to better understand vision by bringing the model of realistic stimulation into theoretical

studies. The visual events recorded by the ATIS are asynchronous and frameless and suit well as a simple model for retinal output with a high temporal resolution. In this letter, we have presented an event-based feed-forward model for motion processing using spatiotemporal Gabor kernels. Unlike most previously proposed models for spatiotemporal filtering, we have used temporally dense and spatially independent visual events as input to the model, which allows building a motion estimation algorithm with absolute time stamps. Because such image frames are absent in our methodology, it results in reduced computational expense despite having millisecond temporal resolution. In an abstract LGN-V1-MT model, we have illustrated how speed and direction selectivity may emerge out of event streams, eventually to be mapped to a continuous probabilistic sensory map.

Alternative spatiotemporal filtering methods have been used in other biologically inspired neural models of motion detection (Tschechne et al., 2014; Brosch, Tschechne, & Neumann, 2015), with event-based and address-event-representation (AER) approaches. These models have mainly focused on neuronal interactions at initial motion estimation stages and progressive motion integration, stressing modulatory MT-V1 feedback and form-motion recurrent processing, where events are buffered in windows of a certain size to contribute to motion disambiguation.

In our model, implementing the motion estimation with particle filtering provides a framework for predictive coding of event densities on spatiotemporal maps, while treating them independently with their attached local velocity. With a set of simple and controlled stimuli, our model is the first event-based formulation of motion computations that accommodates asynchronous and temporally dense stimulation and explains the gradual buildup of motion coding in finely tuned scales and millisecond resolution. Importantly, we show that for a very short displacement of the stimulus (trajectory of 40–50 ms), advancing in the trajectory by a few events leads to significant improvement of direction and speed decoding. Our results provide insight into the potential of event-based sensory computation in explaining behavioral efficiency, despite various sources of delay and uncertainties in the sensory world.

## 5 Discussion

Precise motion processing and correct position coding are the most basic functions of the visual system. From an evolutionary point of view, different species need to be efficient in motion processing to avoid an approaching predator or to catch running or flying prey. This is not an easy task, especially when we consider various sources of noise, uncertainty, and, most important, neural delay. Despite these, there is plenty of evidence that the visual system is doing this in an almost perfect manner. This performance is demonstrated in precise tracking or catching tasks. For instance, due to

neural delays in tennis, players need to hit the ball in its physical place while having outdated access to the sensory data (with a delay  $\approx 100$  ms).

The neural system uses multiple spatiotemporal scales to encode dynamics of sensory environment. Terms like *temporal coding*, *temporal acuity*, and *temporal precision* are often used to describe either the relation between the timing of sensory events and the encoded response or the accuracy of the neural code. One may expect finer timescales to encode rapidly changing stimuli, but millisecond time precision appears to be a neural coding principle in the visual system, even though statistically, the visual world has much slower dynamics. Because the neural code is highly nonlinear and complex, it not only reflects how fast the sensory world is changing, but it is also thought that a finely varying signal with millisecond precision may encode other features of the visual stimuli (temporal coding) (Uzzell & Chichilnisky, 2004). Therefore, the existence of such a fine timescale even for a stimulus that is changing very slowly seems to be necessary to map the various aspects of the stimulus. A precise temporal response is known to be a generic property of retinal ganglion cells (RGCs), holding over a wide range of stimulus patterns and contrast in different species (Uzzell & Chichilnisky, 2004; Berry et al., 1997). Recordings from RGCs on rabbit and salamander in response to random flickers demonstrate transient and sparse spike bursts, with highly reproducible timings (Berry et al., 1997).

Millisecond precision has been observed on spiking responses of cat LGN by presenting repeated white noise patterns (Reinagel & Reid, 2000) with a very low intertrial variability. This very precise spiking activity has been shown to be consistent across all cells of the same type in LGN (Reinagel & Reid, 2002). However, the relative precision of the LGN can vary depending on the content of the stimulus (Butts et al., 2007). This stimulus-dependent temporal precision has been observed by in vivo recordings on cat LGN in response to drifting sinusoidal gratings (Reich, Victor, Knight, Ozaki, & Kaplan, 1997) (a lower precision of 5 ms for high-contrast stimulus and even less in moderate contrasts).

Timing of responses at retina appears to be a more informative neural signature than spike counts. Investigations with information theory methods on the correlated population code of RGCs of mouse while the retina is presented with natural stimuli reveals that considering RGCs as independent encoders is a reasonable strategy (Nirenberg et al., 2001), a claim that was proposed long before in a study on the frogs visual system (Lettvin et al., 1959). The efficiency of motion processing machinery is heavily relied on for its dynamic and flexible sensory sampling. The standard approach in a great number of previous models is based on optical flow computation and luminance conservancy, which is an estimation of image displacement based on spatial and temporal variations of brightness. In line with our previous work on event-based motion computations (Benosman et al., 2012, 2013), our method is luminance free and time based and contributes

to understanding the neural code of motion in the visual cortex by finely controlled and frame-free spatiotemporal sampling.

In addition, it appears that conventional frame-based stimulations at experimental setups are not able to exploit the whole excitability potential of the eye, as might be the case for natural vision. Our theory opens perspectives to investigate demanding and precise visual tasks in more realistic experimental setups and explore the perceptual consequences of enhanced neural code and potentially decreased neural delays.

## References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *JOSA A*, 2(2), 284–299.
- Allman, J. M., & Kaas, J. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (*Aotus trivirgatus*). *Brain Research*, 31(1), 85–105.
- Benosman, R., Clercq, C., Ieng, S., & Bartolozzi, C. (2013). Event-based visual flow. *IEEE Trans. Neural Networks and Learning Systems*, 25, 407–417.
- Benosman, R., Ieng, S., Clercq, C., Bartolozzi, C., & Srinivasan, M. (2012). Asynchronous frameless event-based optical flow. *Neural Networks*, 27, 32–37.
- Berry, M., Brivanlou, I., Jordan, T., & Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398(6725), 334–338.
- Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences*, 94(10), 5411–5416.
- Brosch, T., Tschechne, S., & Neumann, H. (2015). On event-based optical flow detection. *Frontiers in Neuroscience*, 9.
- Burgi, P.-Y., Yuille, A. L., & Grzywacz, N. M. (2000). Probabilistic motion estimation based on temporal coherence. *Neural Computation*, 12(8), 1839–1867.
- Butts, D. A., Weng, C., Jin, J., Yeh, C.-I., Lesica, N. A., Alonso, J.-M., & Stanley, G. B. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature*, 449(7158), 92–95.
- Dubner, R., & Zeki, S. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, 35(2), 528–532.
- Erlhagen, W. (2003). Internal models for visual perception. *Biological Cybernetics*, 88(5), 409–417.
- Gollisch, T., & Meister, M. (2008). Rapid neural coding in the retina with relative spike latencies. *Science*, 319(5866), 1108–1111.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160(1), 106–154.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 195(1), 215–243.
- Isard, M., & Blake, A. (1998). Condensation conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1), 5–28.

- Jancke, D., Erlhagen, W., Schöner, G., & Dinse, H. (2004). Shorter latencies for motion trajectories than for flashes in population responses of cat primary visual cortex. *Journal of Physiology*, 556(3), 971–982.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82 (Series D), 35–45.
- Kaplan, B. A., Khoei, M. A., Lansner, A., & Perrinet, L. U. (2014). Signature of an anticipatory response in area V1 as modeled by a probabilistic model and a spiking neural network. In *Proceedings of the 2014 International Joint Conference on Neural Networks*. Piscataway, NJ: IEEE.
- Khoei, M. A., Masson, G. S., & Perrinet, L. U. (2017). The flash-lag effect as a motion-based predictive shift. *PLoS Computational Biology*, 13(1), e1005068.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, 47(11), 1940–1951.
- Lim, H., & Choe, Y. (2008). Extrapolative delay compensation through facilitating synapses and its relation to the flash-lag effect. *IEEE Transactions on Neural Networks*, 19(10), 1678–1688.
- Maus, G., Fischer, J., & Whitney, D. (2013). Motion-dependent representation of space in area MT+. *Neuron*, 78(3), 554–562.
- Nirenberg, S., Carcieri, S. M., Jacobs, A. L., & Latham, P. E. (2001). Retinal ganglion cells act largely as independent encoders. *Nature*, 411(6838), 698–701.
- Perrinet, L. U., & Masson, G. S. (2012). Motion-based prediction is sufficient to solve the aperture problem. *Neural Computation*, 24(10), 2726–2750.
- Petkov, N., & Subramanian, E. (2007). Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal Gabor filters with surround inhibition. *Biological Cybernetics*, 97(5–6), 423–439.
- Posch, C., Matolin, D., & Wohlgenannt, R. (2008). An asynchronous time-based image sensor. In *Proceedings of the IEEE Symposium on Circuits and Systems* (pp. 2130–2133). Piscataway, NJ: IEEE.
- Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., & Kaplan, E. (1997). Response variability and timing precision of neuronal spike trains in vivo. *Journal of Neurophysiology*, 77(5), 2836–2841.
- Reinagel, P., & Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *Journal of Neuroscience*, 20(14), 5392–5400.
- Reinagel, P., & Reid, R. C. (2002). Precise firing events are conserved across neurons. *Journal of Neuroscience*, 22(16), 6837–6841.
- Tschechne, S., Brosch, T., Sailer, R., von Eglhoffstein, N., Abdul-Kreem, L. I., & Neumann, H. (2014). On event-based motion detection and integration. In *Proceedings of the 8th International Conference on Bioinspired Information and Communications Technologies* (pp. 298–305). ICST.
- Uzzell, V., & Chichilnisky, E. (2004). Precision of spike trains in primate retinal ganglion cells. *Journal of Neurophysiology*, 92(2), 780–789.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598–604.