frontiers in Neuroscience

ORIGINAL RESEARCH
published: 20 April 2015
doi: 10.3389/fnins.2015.00137

On event-based optical flow detection

Tobias Brosch, Stephan Tschechne and Heiko Neumann *

Faculty of Engineering and Computer Science, Institute of Neural Information Processing, Ulm University, Ulm, Germany

Event-based sensing, i.e., the asynchronous detection of luminance changes, promises low-energy, high dynamic range, and sparse sensing. This stands in contrast to whole image frame-wise acquisition by standard cameras. Here, we systematically investigate the implications of event-based sensing in the context of visual motion, or flow, estimation. Starting from a common theoretical foundation, we discuss different principal approaches for optical flow detection ranging from gradient-based methods over plane-fitting to filter based methods and identify strengths and weaknesses of each class. Gradient-based methods for local motion integration are shown to suffer from the sparse encoding in address-event representations (AER). Approaches exploiting the local plane like structure of the event cloud, on the other hand, are shown to be well suited. Within this class, filter based approaches are shown to define a proper detection scheme which can also deal with the problem of representing multiple motions at a single location (motion transparency). A novel biologically inspired efficient motion detector is proposed, analyzed and experimentally validated. Furthermore, a stage of surround normalization is incorporated. Together with the filtering this defines a canonical circuit for motion feature detection. The theoretical analysis shows that such an integrated circuit reduces motion ambiguity in addition to decorrelating the representation of motion related activations.

Keywords: event-based sensor, motion detection, optical flow, address-event representation, motion integration, velocity representation, spatio-temporal receptive fields

OPEN ACCESS

Edited by:
Emmanuel Michael Drakakis, Imperial College London, UK

Reviewed by:
Anton Civit, University of Seville, Spain
Joaquin Sitte, Queensland University of Technology, Australia

***Correspondence:**
Heiko Neumann, Faculty of Engineering and Computer Science, Institute of Neural Information Processing, Ulm University, James-Franck-Ring, Ulm 89081, Germany
heiko.neumann@uni-ulm.de

Specialty section:
This article was submitted to Neuromorphic Engineering, a section of the journal Frontiers in Neuroscience

Received: *12 January 2015*
Accepted: *02 April 2015*
Published: *20 April 2015*

Citation:
Brosch T, Tschechne S and Neumann H (2015) On event-based optical flow detection. Front. Neurosci. 9:137. doi: 10.3389/fnins.2015.00137

1. Introduction

The initial stages of visual processing extract a vocabulary of relevant feature items related to a visual scene. Rays of light reach the observer's eye and are transformed to internal representations. This can be formalized as sampling the ambient optic array (Gibson, 1978, 1986). Formally, the plenoptic function $P(\theta, \phi, \lambda, t, V_x, V_y, V_z)$ describes the intensity of a light ray of wavelength λ passing through the center of the pupil of an idealized eye at every possible angle (θ, ϕ) located at the position (V_x, V_y, V_z) at time t (Adelson and Bergen, 1991). As a simplification we assume a single stationary camera sensing a single narrow band of wavelengths in the electromagnetic spectrum on its image plane (x, y), reducing the plenoptic function to $P_{\lambda, V_x, V_y, V_z}(x, y, t) = g(x, y, t)$ (the spatio-temporal gray level function). Elemental measurements are necessary to access the plenoptic structures. Conventional frame-based cameras sample the optic array by reading out measurements of all light-sensitive pixels at a fixed rate. Since the temporal sampling rate is limited through reading all pixel values in a fixed time interval, fast local luminance changes are integrated over time and cannot be differentiated in the further processing. When no changes occur in the intensity function, redundant information is generated that is carried to the subsequent processing steps. Address-event representations (AER), on the other hand, originate from image

sensors in which pixel operate at individual rates generating events based on local decisions to generate an output response, like in the mammalian retina (Mead, 1990; Liu and Delbruck, 2010).

We will focus on silicon retinas that generate an AER, namely the dynamic vision sensor (DVS; Delbrück and Liu, 2004). Whenever the change in the log-luminance function exceeds a predefined threshold ϑ, events $e_k \in \{-1, 1\}$ are generated at times t_k that emulate spike sequences of on- and off-contrast cells in the retina, respectively (**Figure 1**). We discuss what kind of information is accessible from the initial stages of event-based visual sensing and compare different approaches to estimate optical flow from the stream of on- and off-events visualized in **Figure 1**. We identify weaknesses, suggest improvements, propose a novel biologically inspired motion detector and conduct experiments to validate the theoretical predictions of flow estimation. The proposed detector is then further extended by incorporating an inhibitory pool of activation over a neighborhood in the space-time-feature domain that leads to contextual modulation and response normalization. Together with the initial filtering stage the scheme defines a canonical circuit model as suggested in Kouh and Poggio (2008); Carandini and Heeger (2012); Brosch and Neumann (2014a). This competitive mechanism is investigated from an information-theoretic point of view, shown to accomplish decorrelation, and linked to radial Gaussianization of the input response distribution (Lyu and Simoncelli, 2009b). Finally, we investigate whether motion transparency encoding (Snowden and Verstraten, 1999), i.e., the percept of two competing motions at a single location, like flocks of birds flying in front of passing clouds, can be supported.

2. Materials and Methods

2.1. Theoretical Aspects of Event-Based Visual Motion Detection

2.1.1. Nomenclature and Principal Problems

We describe the stream of events by the function

$$e : \mathbb{R}^2 \times \mathbb{R} \to \{-1, 0, 1\} \tag{1}$$

which is always zero except for tuples $(x_k, y_k; t_k) = (p_k; t_k)$ which define the location and time of an event k generated when the luminance function increases or decreases by a significant amount. In other words, the function that defines the event generation $e(p_k; t_k) = e_k$, generates 1 if the log-luminance changed more than a threshold ϑ, i.e., an ON event, and -1 if it changed more than $-\vartheta$, i.e., an OFF event. This sampling of the lightfield essentially represents the temporal derivative of the luminance function g

$$\frac{d}{dt} g(p; t) = g_t(p; t) \approx \frac{\vartheta}{\Delta t} \sum_{k:\, t_k \in (t-\Delta t, t]} e_k \,, \tag{2}$$

with ϑ the sensitivity threshold of the event-based sensor.

To estimate local translatory motion we assume throughout the paper that the gray level function remains constant within a small neighborhood in space and time, i.e., $g(x, y; t) = g(x + \Delta x, y + \Delta y; t + \Delta t)$ (gray level constancy; c.f. Horn and Schunck, 1981). Note that due to the low latency of 15 µs of the event-based sensor (Lichtsteiner et al., 2008), this assumption is more accurate than for conventional frame based sensors.

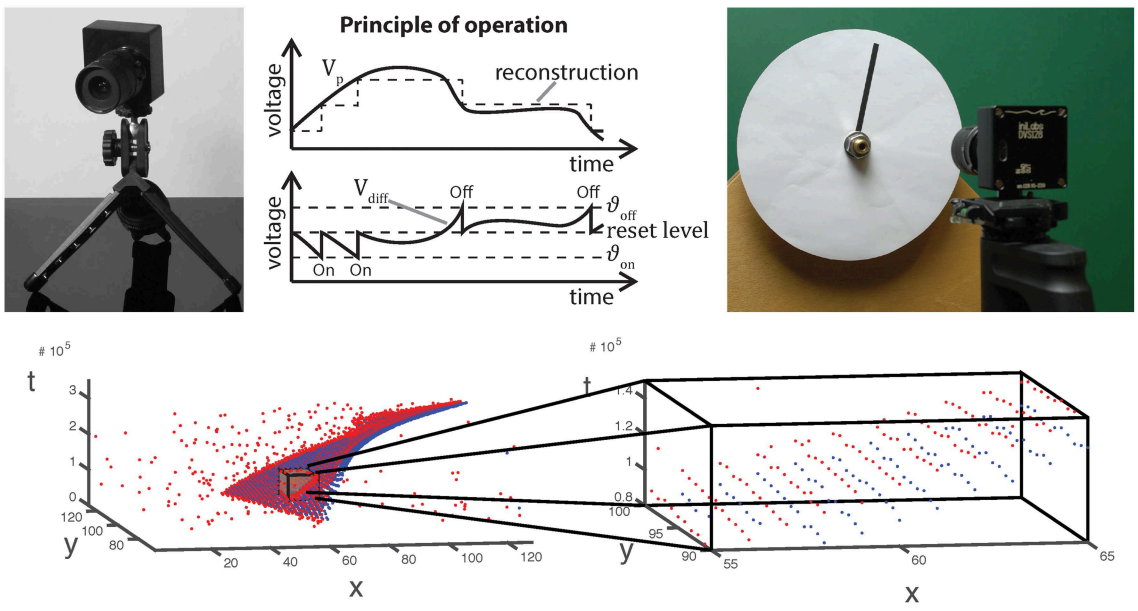

FIGURE 1 | Top from left to right: Image, operating principle and stimulus generation of/for an asynchronous event sensor. Luminance changes exceeding a given threshold evoke ON and OFF events for positive and negative changes, respectively. The very low latency of the dynamic vision sensor (15 µs) requires analog stimulus generation as illustrated on the right. **Bottom:** Visualization of the spatio-temporal event cloud generated by the rotating stimulus in the upper right. A small volume has been zoomed in. As can be seen, only few events have been generated at a single location.

Local expansion up to the second order yields the constraint $\mathbf{\Delta}x^T\nabla_3 g + 1/2\mathbf{\Delta}x^T H_3 \mathbf{\Delta}x = 0$. Here, $\mathbf{\Delta}x = (\Delta x, \Delta y, \Delta t)^T$, $\nabla_3 g = (g_x, g_y; g_t)^T$ is the gradient with the 1st order partial derivatives of the continuous gray-level function, and H_3 denotes the Hessian with the 2nd order partial derivatives of the continuous gray-level function that is defined in the x–y–t-domain. If we further assume that the 2nd order derivative terms are negligible (linear terms dominate) we arrive at the spatio-temporal constraint equation that has been used for least-squares motion estimation. The least-squares formulation is based on a set of local constraint measures over a small neighborhood under the assumption of locally constant translations (Lucas and Kanade, 1981), i.e., $g_x u + g_y v + g_t = 0$ given that $\Delta t \rightarrow 0$ and $u^T = (u, v) = (\Delta x/\Delta t, \Delta y/\Delta t)$. Note that this motion constraint equation can also be represented in the frequency domain in which $f_x u + f_y v + f_t = 0$ holds with f denoting the frequency with subindices referring to the respective cardinal axes and assuming a non-vanishing energy spectrum for the gray-level luminance signal, i.e., $\|\hat{G}\| \neq 0$. The local image motion u of an extended contrast can only be measured orthogonal to the contrast (normal flow, Wallach, 1935; Barron et al., 1994; Fermüller and Aloimonos, 1995; Wuerger et al., 1996). For simplicity, we assume a vertically oriented gray level edge ($g_y = 0$). Then the motion can be estimated along the horizontal directions (left or right with respect to the tangent orientation of the contrast edge). When the edge contrast polarity is known (light-dark, LD, $g_x < 0$ or dark-light, DL, $g_x > 0$) the spatio-temporal movements can be estimated without ambiguity. For an DL edge if $g_t < 0$ the edge moves to the right, while for $g_t > 0$ the edge moves to the left (c.f. **Figure 2**).

For an LD edge the sign of the temporal derivatives g_t changes for both respective movement directions, i.e., only the ratio of gray-level derivatives yields a unique direction selector orthogonal to the oriented luminance contrast. This means that, $\mathrm{sgn}(g_x/g_t) = -1$ implies rightward motion while $\mathrm{sgn}(g_x/g_t) = 1$ implies leftward motion, irrespective of the contrast polarity. Note, however, that an estimate of g_x is not easily accessible from the stream of events of an asynchronous event sensor. Thus, a key question is to what extend the required spatio-temporal derivative information is available and can be estimated.

2.1.2. Moving Gray-Level Edges and the Spatio-Temporal Contrast Model

We describe the luminance function g for a stationary DL transition by convolving a step edge $\mathcal{H}(\cdot)$ with a parameterized Gaussian,

$$g_\sigma(x) = \frac{c}{\sqrt{2\pi}\sigma} \cdot \mathcal{H}(x) * \exp\left(-\frac{x^2}{2\sigma^2}\right) + g_0 = c \cdot \mathrm{erf}_\sigma(x) + g_0, \tag{3}$$

with c denoting the luminance step height, g_0 the basic luminance level, and "$*$" denoting the convolution operator (since we only study the derivatives, we assume $g_0 = 0$). The parameter σ controls the spatial blur of the luminance edge with $\sigma \rightarrow 0$ resulting in the step-function. Different contrast polarities are defined by $g_\sigma^{DL}(x) = c \cdot \mathrm{erf}_\sigma(x)$ and $g_\sigma^{LD}(x) = c \cdot (1 - \mathrm{erf}_\sigma(x))$, respectively (Neumann and Ottenberg, 1992).

When this gray-level transition moves through the origin at time $t = 0$ it generates a slanted line with normal n in the x–t-space (c.f. **Figure 3**). The speed s of the moving contrast edge is given by $s = \sin(\theta)/\cos(\theta)$, where θ is the angle between n and the x-axis (this is identical to the angle between the edge tangent and the t–axis). For a stationary gray-level edge (zero

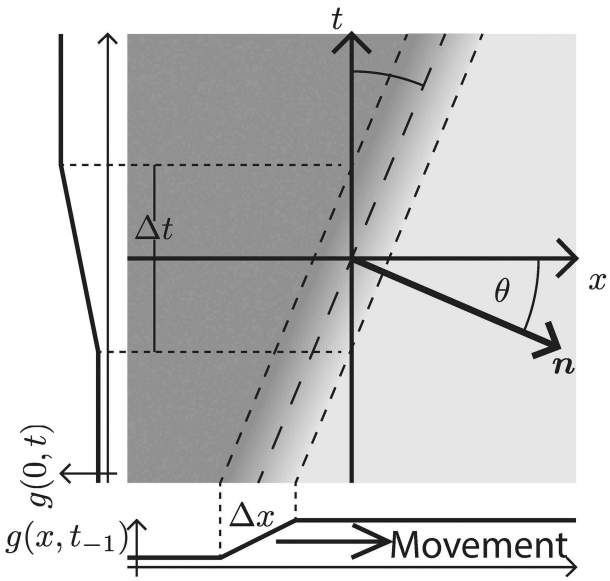

FIGURE 3 | Rightward moving 1D edge illustrated in the *x*–*t*-domain. The velocity is defined by the direction and the speed of the spatio-temporal change. In the case depicted here, the direction is to the right and the speed is encoded by the angle θ between the x-axis and the normal vector n along the spatio-temporal gradient direction (measured in counter-clockwise rotation). Alternatively, for a contrast edge of known finite transition width Δx, the speed can be inferred from the time Δt, it takes the contrast edge to pass a specific location on the x–axis.

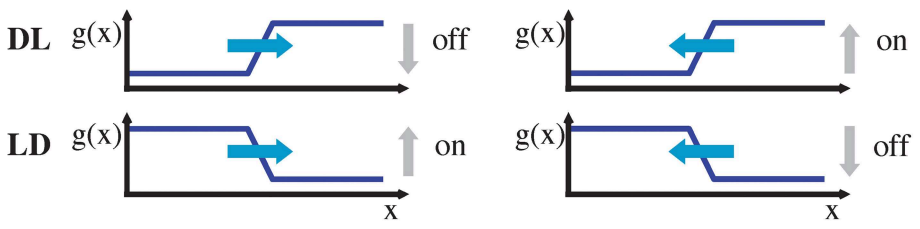

FIGURE 2 | Moving DL (dark-light) and LD (light-dark) edge, either to the left or to the right (denoted by blue arrows), have an associated temporal on/off signature. Note that without knowledge about the edge type (DL vs. LD), an on/off event alone is insufficient to determine the motion direction.

speed) we get $\theta = 0$ (i.e., the edge generated by the DL transition in the x–t-domain is located on the t-axis). Positive angles $\theta \in (0°, 90°)$ (measured in counterclockwise direction) define leftward motion, while negative angles define rightward motion. For illustrative purposes, we consider an DL contrast that is moving to the right (c.f. **Figure 3**). The spatio-temporal gradient is maximal along the normal direction $n = (\cos\theta, \sin\theta)^T$. The function $g(x; t)$ describing the resulting space-time picture of the movement in the x-t-space is thus given as

$$g_{\sigma\theta}(x; t) = \frac{c}{\sqrt{2\pi}\sigma}\mathcal{H}(x_\perp) * \exp\left(-\frac{x_\perp^2}{2\sigma^2}\right), \tag{4}$$

with $x_\perp = x \cdot \cos\theta - t \cdot \sin\theta$. The respective partial temporal and spatial derivatives are given as

$$\frac{\partial}{\partial t}g_{\sigma\theta}(x; t) = \frac{-c}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x_\perp^2}{2\sigma^2}\right) \cdot \sin\theta\,, \tag{5}$$

$$\frac{\partial}{\partial x}g_{\sigma\theta}(x; t) = \frac{c}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x_\perp^2}{2\sigma^2}\right) \cdot \cos\theta\,. \tag{6}$$

Now, recall that the event-based DVS sensor provides an estimate of g_t at a specific location [c.f. Equation (2)]. For a moving contrast profile this leads to a changing luminance function along the t-axis (side graph $g(0, t)$ in **Figure 3**). The temporal derivative of this profile is formally denoted in Equation (5). Given a *known velocity* specified by θ, we can combine equations (5) and (6) to determine g_x as

$$\frac{\partial}{\partial x}g_{\sigma\theta}(x; t) = -\frac{\partial}{\partial t}g_{\sigma\theta}(x; t) \cdot \tan\theta\,. \tag{7}$$

In sum, the temporal edge transition can be reconstructed in principle from a (uniform) event sequence at the edge location for a specific motion direction, given that

- a reliable speed estimate is available to infer a robust value for θ, and
- reliable estimates of temporal changes have been generated as an event cloud over an appropriately scaled temporal integration window Δw_t.

Note, that both parameters, θ and Δw_t, need to be precisely estimated to accomplish robust estimates of contrast information of the luminance edge. In Sections 2.1.4 and 2.1.5, we will briefly outline the necessary steps in such an estimation process. Alternatively, one can try to directly estimate the partial derivatives used in the motion constraint equation from the stream of events. The construction of this approach and its related problems are described in the following Section 2.1.3.

2.1.3. Estimating Spatio-Temporal Continuity using Event-Sequences

The local spatio-temporal movement of a gray-level function can be estimated by least-squares optimization from a set of local contrast measurements which define intersecting motion constraint lines in velocity space (Lucas and Kanade, 1981). Given a dense temporal sampling the spatio-temporal gray-level function can be reasonably well captured by a first-order approximation (as summarized in Section 2.1.1). The key question remains how one could estimate the spatial and temporal derivatives in the constraint equations, $g_x u + g_y v + g_t = 0$ from event sequences generated by the DVS. Events only encode information about the temporal derivative g_t [c.f. Equation (2)]. Thus, without additional information it is impossible to reliably estimate g_x or g_y, as outlined in the previous Section 2.1.2. The derivative of a translatory moving gray level patch, however, generates a unique response in h: $= g_t$. Thus, we can apply the motion constraint equation to the function h and solve $h_x u + h_y v + h_t = 0$, instead. Using two temporal windows $\mathcal{T}_{-2} = (t - 2\Delta t, t - \Delta t]$ and $\mathcal{T}_{-1} = (t - \Delta t, t]$, we can approximate h_t, for example, by a backward *temporal* difference

$$h_t(p; t) = g_{tt}(p; t) \approx \frac{\vartheta}{\Delta t^2}\left(\sum_{t' \in \mathcal{T}_{-1}} e(p; t') - \sum_{t' \in \mathcal{T}_{-2}} e(p; t')\right), \tag{8}$$

with $p = (x, y)^T$ and ϑ denoting the event-generation threshold. The *spatial* derivatives h_x and h_y can be approximated by central difference kernels $[-1, 0, 1]$ and $[-1, 0, 1]^T$, respectively. These can be applied to the function h estimated by integrating over the temporal window $\mathcal{T}$ (e.g., $\mathcal{T} = \mathcal{T}_{-2} \cup \mathcal{T}_{-1}$)

$$h_x(p; t) = g_{tx}(p; t) \approx \sum_{t' \in \mathcal{T}} e(x + 1, y; t') - \sum_{t' \in \mathcal{T}} e(x - 1, y; t')\,, \tag{9}$$

$$h_y(p; t) = g_{ty}(p; t) \approx \sum_{t' \in \mathcal{T}} e(x + 1, y; t') - \sum_{t' \in \mathcal{T}} e(x, y - 1; t')\,. \tag{10}$$

Consequently, the resulting flow computation results in a sparsification of responses since stationary edges will not be represented in h. This approach is similar to that of Benosman et al. (2012) but consistently employs the second derivative instead of mixing the first and second derivatives which leads to inconsistencies in general.

Note, however, that this approach has multiple issues regarding any real implementation. The most important observation is that when a luminance edge passes a pixel's receptive field of the DVS sensor, the amount of events is in the range of about 10 events (often even less, depending on the contrast, speed and luminance conditions; c.f. zoomed display of the event cloud in **Figure 1**). Thus, huge approximation errors occur for h_x, h_y and especially in h_t (since this now represents the second derivative of the original gray-level function g). Furthermore, we can only estimate h_t accurately, if the temporal windows are small enough such that the gray-level edge has not already passed through the receptive field of a target cell at position p. This limits the number of events to even less and leads to magnifying the outlined problems even further. Alternatively, one could try to directly approximate the temporal derivative for each event by incorporating the time-span since the last event, i.e.,

$$\frac{d}{dt}g(p;t) = g_t(p;t) \approx \frac{\vartheta}{\Delta_W t}e(p,t)\,, \tag{11}$$

with $\Delta_W t$ representing the time that has passed since the last event generated at p. This assumes a constant intensity change since the last event. This, however, is certainly not true for the first event because first nothing happens for a long period and then occasionally some change occurs that causes the event, i.e., the estimate will be too small, because $\Delta_W t$ is too big.

2.1.4. Least-Squares Velocity Estimation

The short temporal window in which events of a briefly passing contrast edge are generated makes it difficult to reliably estimate the derivatives required in the motion constraint equation (c.f. previous section). An alternative approach is to consider the distribution of events (the "event cloud") in a small volume of the x-y-t-space. The cloud that results from a moving contrast edge generates a locally plane-like cloud of on- and/or off-events (with on- and off-events in the case of a line, for example, and only on- or off-events in the case of a transition from one homogeneous region to another) to which a velocity tangent plane can be fitted (Benosman et al., 2014). The thickness of the event cloud orthogonal to the velocity tangent plane depends on the sharpness of the contrast edge, the speed with which the gray-level discontinuity moves through the spatial location of a pixel, and its local neighborhood (the receptive field, RF, of a cell at this position). In Benosman et al. (2014) a function $\Sigma_e : \mathbb{N}^2 \rightarrow \mathbb{R}$ is defined that maps the location p of an event e to the time $\Sigma_e(p) = t$ when the event was generated. This mapping may be used to describe the cloud of events. However, care should be taken since the mapping is non-continous in principle: it is either defined for each event in which case the mapping is not differentiable, or it is defined for all events in which case the mapping is not injective (because for a given t, there are multiple events at different locations). In any case, the inverse function theorem of calculus (as employed in Benosman et al., 2014) *cannot* be applied here to derive a speed estimate. This insight might explain, why in the velocity-vector-field of a rotating bar illustrated in Figure 7b of Benosman et al. (2014) the velocity vectors at the outer parts are shorter (instead of longer) compared to the velocity vectors at the inner ones. We suggest an alternative solution in which the speed is estimated from the regression plane by solving the orthogonal system of the velocity vector $v = (u, v, 1)^T$ (defined in homogeneous coordinates), the orientation of the moving luminance edge $l = (l_x, l_y, 0)^T$, and the normal vector $n = (a, b, c)^T$ of the plane. These three vectors form an orthogonal system that spans the x-y-t space:

$$\begin{aligned}
n \perp v \quad &\Rightarrow \quad n^T \cdot v = au + bv + c = 0 && (12)\\
n \perp l \quad &\Rightarrow \quad n^T \cdot l = al_x + bl_y = 0 && (13)\\
l \perp v \quad &\Rightarrow \quad l^T \cdot v = ul_x + vl_y = 0 && (14)\\
(13)\cdot u - (14)\cdot a \quad &\Rightarrow \quad bu - av = 0 && (15)\\
(12)\cdot b - (15)\cdot a \quad &\Rightarrow \quad v\cdot(a^2+b^2) + bc = 0 && (16)\\
(12)\cdot a + (15)\cdot b \quad &\Rightarrow \quad u\cdot(a^2+b^2) + ac = 0\,. && (17)
\end{aligned}$$

The resulting velocity components u and v are then given as (with $n = (a, b, c)^T$)

$$\begin{pmatrix} u \\ v \end{pmatrix} = -\frac{c}{a^2+b^2}\begin{pmatrix} a \\ b \end{pmatrix}\,, \tag{18}$$

with the speed component $s = \sqrt{u^2+v^2} = c\cdot(a^2+b^2)^{-1/2}$. Note, that for slow or moderate velocities, a reliable estimate of the velocity tangent plane requires a spatial as well as a temporal neighborhood such that the event cloud is fully covered within the spatio-temporal window (or RF) considered for the LS regression. In particular, the neighborhood support must cover the event cloud illustrated in the bottom right of **Figure 1**. If this condition is not fulfilled, i.e., if the window is smaller than the extent of the cloud, then the principal axes are arbitrary and cannot be estimated reliably.

2.1.5. Direction-Sensitive Filters

As an alternative to considering the LS regression in estimating the velocity tangent plane from the cloud of events, the uncertainty of the event detection might be incorporated directly. At each location, detected events define likelihood distributions $p(e|u)$ given certain velocities of the visual scene (estimated by a filter bank, for example). Using Bayes' theorem, we derive that for each event $p(u|e) \propto p(e|u)\cdot p(u)$. If each velocity is equally likely to be observed without a priori knowledge, i.e., $p(u_i) = p(u_j)$ (for arbitrary velocities i, j), it holds $p(u|e) \propto p(e|u)$ and thus, the velocity u_{est} of the movement that caused event e can be estimated as

$$u_{est} = \text{argmax}_u p(u|e) = \text{argmax}_u p(e|u)\,. \tag{19}$$

Thus, we can estimate the velocity from the responses $p(e|u_i),\ i = 1, 2, \ldots$ of a filter bank, for example. In addition, a priori knowledge could be incorporated to reduce noise and to increase coherency. Current knowledge suggests, that such distributions are represented by the filter characteristics of the spatio-temporal receptive fields of cells in area V1 which we use as inspiration for a novel filter mechanisms described in the following Section 2.2.

2.2. Event-Based Motion Estimation using Direction-Selective Filters

In this section, we define spatio-temporal filters that are fitted to the physiological findings from De Valois et al. (2000) summarized in the following Section 2.2.1.

2.2.1. Experimental Evidence

Our filter design essentially reverses the decomposition of neural responses conducted by De Valois et al. (2000) (also c.f. Tschechne et al., 2014). Based on physiological findings first described by DeAngelis et al. (1995), De Valois suggested that inseparable filters stem from a combination of various separable components (De Valois et al., 2000). In De Valois et al. (2000) cortical V1 cells were tested and strong evidence for the coexistence of two distinct types of populations of cells emerged: One population showed spatio-temporally separable weight functions of either even or odd spatial symmetry. These

have either temporally mono- or bi-phasic response characteristics which were mainly determined by a single principal component in 2D (of a singular value decomposition). The other population of cells was spatio-temporally inseparable showing a receptive field distribution of selectivity that were slanted with respect to the time axis, i.e., motion sensitive (c.f. **Figure 3**; c.f. also De Valois and Cottaris, 1998). Response characteristics of these cells were determined by *two* strong principal components in 2D. These two components of the second group were itself spatio-temporally separable with spatially out-of-phase components and always composed of pairs of mono- and bi-phasic distributions.

This main observation lead us to propose a family of spatio-temporally direction selective filters as illustrated in **Figure 4**, that are generated by superposed separable filters with quadrature pairs of spatial weighting profiles (G_{odd} and G_{even}) and mono-/bi-phasic temporal profiles (T_{mono} and T_{bi}). The details of the construction process are outlined in the following sections.

2.2.2. Spatial Gabor Filters

To construct the spatial component of the spatio-temporal filters illustrated in **Figure 4** we define Gabor filters that are fitted to the experimental results of De Valois et al. (2000). To construct multiple spatio-temporally tuned filters of different spatial orientation selectivity, we employ a filter-bank of kernels as illustrated in **Figure 5**. More precisely, we employ Gabor filters maximally selective for the spatial frequency (f_x^0, f_y^0) (with a standard deviation σ in local space) defined by (c.f. **Figure 5**)

$$G_{\sigma,f_x^0,f_y^0}(x,y) = \frac{2\pi}{\sigma^2} \cdot \exp\left[2\pi j\left(f_x^0 x + f_y^0 y\right)\right] \cdot \exp\left[-\frac{2\pi^2 \cdot (x^2+y^2)}{\sigma^2}\right], \quad (20)$$

in local space. The spatial frequencies selected by this filter can be seen by visualizing its Fourier transform (**Figure 5**, bottom left) which is given as

$$\hat{G}_{\hat{\sigma},f_x^0,f_y^0}(f_x,f_y) = \exp\left[-\frac{1}{2} \cdot \frac{\left(f_x - f_x^0\right)^2 + \left(f_y - f_y^0\right)^2}{\hat{\sigma}^2}\right], \quad (21)$$

where $\hat{\sigma} = 1/\sigma$ and the filter tuning (f_x^0, f_y^0) defines the shift of the Gaussian envelope with respect to the origin in the Fourier domain. This defines the two components $G_{odd} = \Im(G_{\sigma,f_x^0,f_y^0})$ and $G_{even} = \Re(G_{\sigma,f_x^0,f_y^0})$ to construct the filters as described in Section 2.2.1 (compare with Daugman, 1985; Marčelja, 1980).

2.2.3. Mono- and Biphasic Temporal Filters

The second component required in the spatio-temporal filter generation process illustrated in **Figure 4** is the definition of mono- and bi-phasic temporal filters, T_{mono} and T_{bi}. To fit the experimental data of De Valois et al. (2000), we define (c.f. **Figure 6**)

$$T_{mono}(t) = G_{\sigma_{mono},\mu_{mono}}(t), \quad (22)$$

$$T_{bi}(t) = -s_1 \cdot G_{\sigma_{bi1},\mu_{bi1}}(t) + s_2 \cdot G_{\sigma_{bi2},\mu_{bi2}}(t), \quad (23)$$

with the unnormalized Gaussian function

$$G_{\sigma,\mu}(t) = \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right). \quad (24)$$

When the experimental findings are incorporated, it is only necessary to choose a value for μ_{bi1}. All other parameters can be inferred according to the experimental data from De Valois et al. (2000):

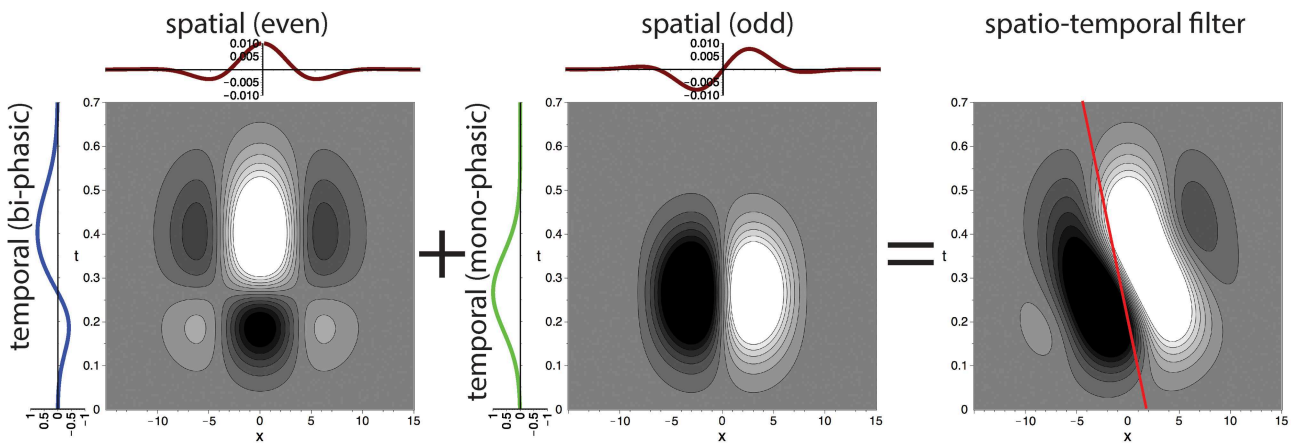

FIGURE 4 | Superposition of spatio-temporally separable filters creates motion-direction sensitive filter. The proposed spatio-temporal filter (right) is constructed by using the results of the singular value decomposition of the receptive field of motion directional cells by De Valois et al. (2000) (c.f. DeAngelis et al., 1995, their Figure 3). Two separable filters are superposed to create the final motion direction selective spatio-temporal filter. Each of the two filters is separable into a pair of a bi-phasic temporal and an even spatial or a mono-phasic temporal and an odd spatial filter, respectively (illustrated by line profile plots to the left and at the top). The red line indicates the preferred speed selectivity identified by a Fourier analysis of the filter function (c.f. Section 2.2.4).

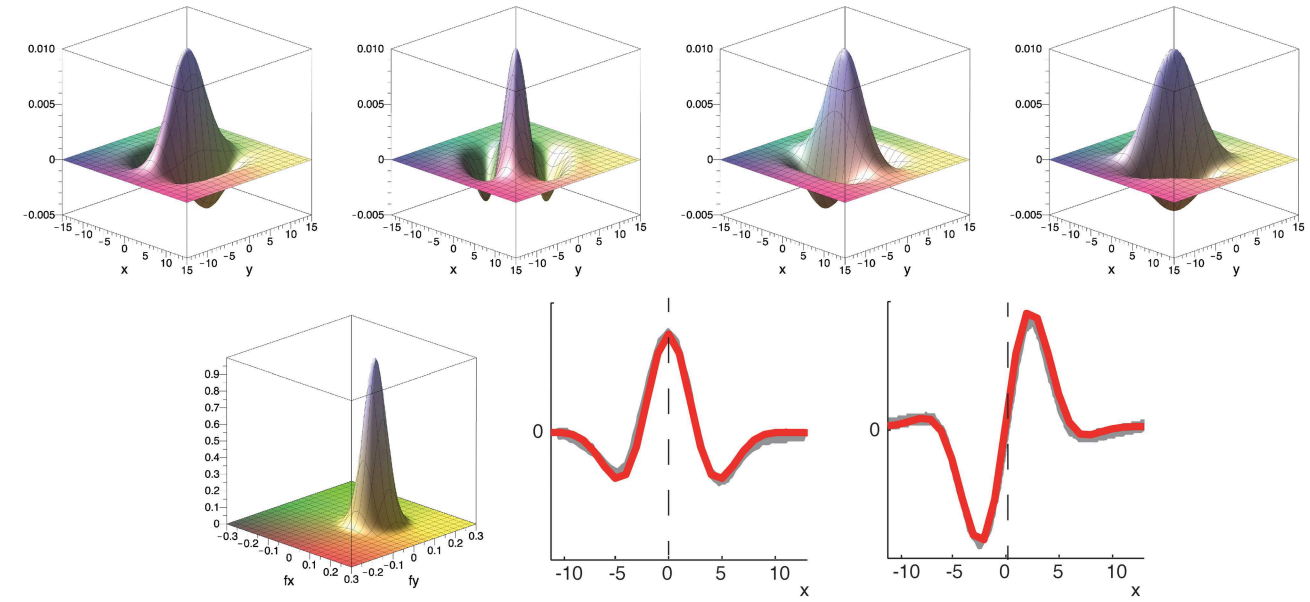

FIGURE 5 | Top row: Example of filter bank of four oriented Gabor filters for $\Phi = [0°, 45°, 90°, 135°]$ (to resemble the spatial receptive field weights; only even component is shown). **Bottom row:** Corresponding filter in Fourier domain and illustration of spatial fitting of analytical (red) and experimental data (gray) from De Valois et al. (2000). Parameters in all plots have been set to $\sigma = 25$ and $f_0 = 0.08$.

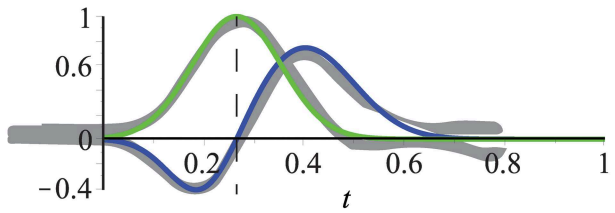

FIGURE 6 | Temporal filters fit experimental data (gray) from De Valois et al. (2000). Kernels consist of one or two Gaussians which define a mono- and bi-phasic temporal filter, respectively. The mean of the mono-phasic kernel has been set to $\mu_{bi1} = 0.2$ with the remaining parameters fitted to the experimental data (see text for details). Dashed line highlights that the peak of the mono-phasic kernel (green) is located at the zero-crossing of the bi-phasic kernel (blue).

- The bi-phasic scaling factors s_1 and s_2 are adapted to the minimum and maximum values of the experimental data relative to the maximum value of the monophasic kernel (which is one), i.e., $s_1 = 1/2$ and $s_2 = 3/4$.
- A good fit with the experimental data reported in De Valois et al. (2000) is achieved by setting the relation between the mean values to $\mu_{bi2} = 2\mu_{bi1}$.
- The standard deviations σ_{mono} and σ_{bi1} are chosen such that the Gaussians are almost zero for $t = 0$, i.e., $\sigma_{mono} = \mu_{mono}/3$, $\sigma_{bi1} = \mu_{bi1}/3$ (3σ–rule; 99.7% of the values lie within three standard deviations of the mean in a normal distribution).
- The standard deviation of the second Gaussian of the bi-phasic kernel is about 3/2 of that of the first, i.e., $\sigma_{bi2} = \frac{3}{2} \cdot \sigma_{bi1} = \frac{1}{2} \cdot \mu_{bi1}$.
- The mean of the mono-phasic kernel μ_{mono} is given by the zero-crossing of the biphasic kernel, i.e., $\mu_{mono} = \frac{1}{5} \cdot \left(1 + \mu_{bi1} \cdot \sqrt{36 + 10 \cdot \ln(s_1/s_2)}\right)$.

Figure 6 illustrates that these settings result in a good fit of the temporal filters with the experimental data reported in De Valois et al. (2000). We will now construct the full spatio-temporal selective filters as outlined in **Figure 4**.

2.2.4. Combined Spatio-Temporal Filter

The full spatio-temporal filter F is defined according to the scheme of **Figure 4**, i.e., by the sum of two products consisting of the odd-spatial $G_{odd} = \Im(G_{\sigma,f_x^0,f_y^0})$, the monophasic temporal T_{mono}, the even-spatial $G_{even} = \Re(G_{\sigma,f_x^0,f_y^0})$, and the biphasic temporal filter T_{bi} (c.f. **Figure 7**):

$$F(x, y, t) = \Im(G_{\sigma,f_x^0,f_y^0}(x, y)) \cdot T_{mono}(t) + \Re(G_{\sigma,f_x^0,f_y^0}(x, y)) \cdot T_{bi}(t)\,. \tag{25}$$

The preferred speed of the filter can be determined by an analysis of the Fourier transform $\hat{F}(f_x, f_y, f_t)$ of the filter function $F(x, y, t)$. From the location $(f_t^{max}, f_x^{max}, f_y^{max})$ where $\hat{F}$ is maximal we can infer the filter's preferred normal velocity, i.e., the velocity parallel to the gradient of the luminance edge (n in **Figure 3**) with maximal filter response, using the following two relations:

- The motion constraint equation in the frequency domain: $f_x u + f_y v + f_t = 0$, i.e., $f \cdot u = -f_t$.

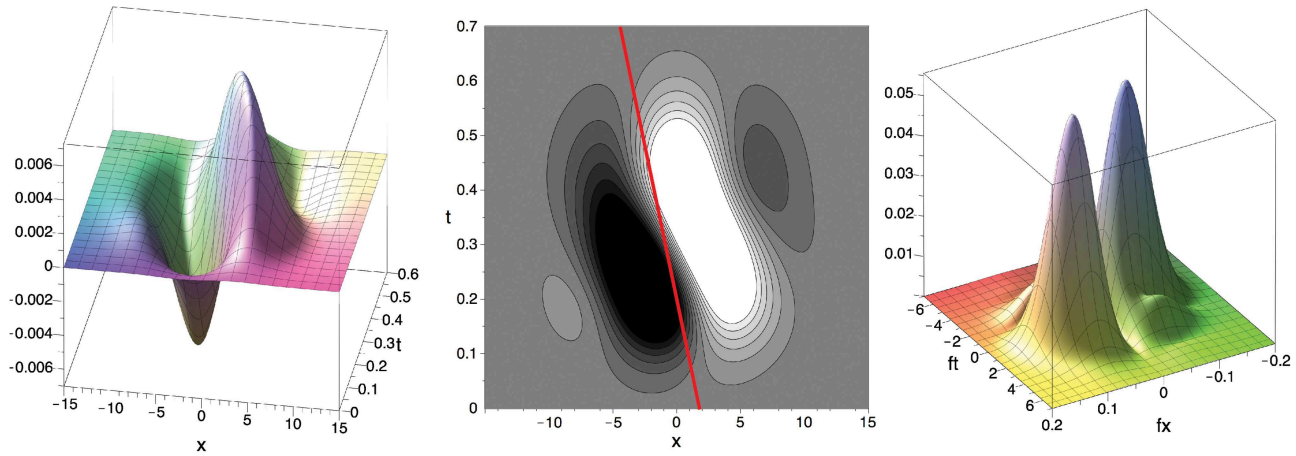

FIGURE 7 | Filter F in x–t–space (left, center; c.f. DeAngelis et al. (1995), Their Figure 3F) and in Fourier domain (right; absolute value) for $\sigma = 25$, $f_x^0 = f_y^0 \approx 0.057$ ($f_0 = \sqrt{(f_x^0)^2 + (f_y^0)^2} = 0.08$) and $\mu_{bi1} = 0.2$ (as in Figure 4). Red line illustrates the velocity component in the x–t-domain ($\approx 8.61 pixel/s$) inferred from the values maximizing the absolute value of the Fourier transform $|\hat{F}|$ of the filter F ($f_t^{max} \approx 0.965, f_x^{max} \approx f_y^{max} \approx 0.057$). See text for details.

- $(u_\perp, v_\perp)$ is orthogonal to the luminance edge, i.e., parallel to (f_x^{max}, f_y^{max}). Thus, the scalar product of $f^{max} = (f_x^{max}, f_y^{max})$ and $u_\perp = (u_\perp, v_\perp)$ is equal to $f^{max} \cdot u_\perp = \|f\| \cdot \|u\|$.

Combining both equations, we obtain $-f_t^{max} = \|f^{max}\| \cdot \|u_\perp\|$, i.e., the speed $s = \|u_\perp\|$ is given as $s = \|u_\perp\| = -f_t^{max}/\|f^{max}\|$. The velocity can now be obtained by scaling the normalized gradient direction $f^{max}/\|f^{max}\|$ with $\|u_\perp\| = \sqrt{u_\perp^2 + v_\perp^2}$ gaining

$$u_\perp = -\frac{f_t}{(f_x^{max})^2 + (f_y^{max})^2} \cdot f_x \,, \tag{26}$$

$$v_\perp = -\frac{f_t}{(f_x^{max})^2 + (f_y^{max})^2} \cdot f_y \,, \tag{27}$$

in $pixel/s$. For the parameter values that fit the experimental data from De Valois et al. (2000), i.e., $\sigma = 25, f_x^0 = f_y^0 \approx 0.057$ ($f_0 = \sqrt{(f_x^0)^2 + (f_y^0)^2} = 0.08$) and $\mu_{bi1} = 0.2$, we numerically determined the values as $f_t^{max} = 0.974, f_x^{max} = f_y^{max} = 0.057$ which maximize $|\hat{F}|$. Thus, the fitted spatio-temporal selective filter F is maximally selective for the velocity $(u_\perp, v_\perp) = (8.61, 8.61) pixel/s$, i.e., a speed of $12.2 pixel/s$.

2.2.5. Response Normalization

The spatio-temporal filter mechanism is combined with a stage of down-modulating lateral divisive inhibition. Such response normalization was shown to have a multitude of favorable properties such as the decrease in response gain and latency observed at high contrasts, the effects of masking by stimuli that fail to elicit responses of the target cell when presented alone, the capability to process a high dynamic range of response activations (Heeger, 1992; Carandini et al., 1997; Koch, 1999; Sceniak et al., 1999; Frégnac et al., 2003; Tsui et al., 2010), and the ability to resolve ambiguous motion estimates at, for example, straight contours without knowledge about the edges of the contour (aperture problem Wallach, 1935; Nakayama and Silverman, 1988; Wuerger et al., 1996). To account for such nonlinearities we add a stage of divisive normalization to test whether it is also suited to enhance flow estimated from the output of DVSs. Based on our previous modeling (e.g., Raudies et al., 2011; Brosch and Neumann, 2012, 2014a), we employ a dynamic neuron model of membrane potentials p and a mean firing rate generated by the monotonically increasing function $\Psi(p)$. The full dynamic equation reads

$$\dot{p}_i = -\alpha_p p_i + (\beta - p_i) \cdot I_i - p_i \cdot \Psi_q(q_i) \,, \tag{28}$$

$$\dot{q}_i = -\alpha_q q_i + \sum_{j \in \mathcal{N}_i} c_j \Psi_I(I_j) \,, \tag{29}$$

with I_i denoting the input and c_j denote the spatio-temporal weighting coefficients of the local neighborhood $\mathcal{N}_i$ of neuron i in the space-time-feature domain (see Brosch and Neumann, 2014a for more details of an even more generalized circuit model). At equilibrium, the following state equations can be derived

$$q_i^\infty = \frac{1}{\alpha_q} \sum_{j \in \mathcal{N}} c_j \Psi_I(I_j) \,, \tag{30}$$

$$p_i^\infty = \frac{\beta I_i}{\alpha_p + I_i + \Psi_q\left(\frac{1}{\alpha_q} \sum_{j \in \mathcal{N}_i} c_j \Psi_I(I_j)\right)} \,. \tag{31}$$

Another favorable property of divisive normalization has been the observation that it can approximate a process dubbed *radial Gaussianization* which minimizes the statistical dependency of coefficients in image coding (Lyu and Simoncelli, 2008b, 2009a):

$$(p_{norm})_i = \frac{I_i}{\left(b + \sum_j c_j I_j^2\right)^{1/2}} \,, \tag{32}$$

where b is a scalar scaling coefficient and c_j denote the weighting coefficients for the activations in the surrounding neighborhood in the space-feature domain [as in Equation (29)]. When the coefficients are learned from a test set (Lyu and Simoncelli, 2009a), it was shown to approximate optimal minimization of statistical dependency, i.e., radial Gaussianization. Here, we test whether this is also true for Gaussian weights (in accordance with neurophyiological findings Bonin et al., 2005) and a slightly different but biologically inspired normalization scheme as outlined in Equation (31). Therefore, the normalization scheme adopted here can only lead to an approximate decorrelation of input encoding. We will, therefore, demonstrate experimentally the impact of the divisive normalization of the spatio-temporal input filtering.

3. Results

In addition to the main part describing the theoretical investigations outlined in the previous sections, we conducted a series of experiments to validate the modeling approach and its theoretical properties. The parameters of the spatio-temporal filters were chosen such that they fit the experimental data as reported in De Valois et al. (2000) (up to scaling), namely $\mu_{bi1} = 0.2$ for the temporal filter components, and $\sigma = 25$, $f_0 = 0.08$ for the spatial filter components. The parameters of the normalization mechanism in Equation (31) were set to $\beta = 1$, $\alpha_p = 0.1$, $\alpha_q = 0.002$, c_j resemble the coefficients of a Gaussian kernel with $\sigma = 3.6$, and $\Psi_I(x) = \Psi_q(x) = \max(0, x)$ denotes a rectifying transfer function.

First, we probed the model using simple and more complex stimuli with translatory and rotational motion to demonstrate the detection performance and noise characteristics of the initial (linear and non-linear filtering of the input). Second, we studied the impact of the normalization stage on the initial filter responses. Third, the model was probed by stimuli with transparent overlaid motion patterns to test the segregation into multiple motion directions at a single spatial location (see e.g., Braddick et al., 2002; Edwards and Nishida, 1999; Treue et al., 2000).

3.1. Detection of Translatory and Rotational Movements

At each location the filter creates a population code of length N with each entry corresponding to the response of a spatio-temporal filter with motion direction selectivity θ_k. For visualization purposes (**Figure 8**), the velocity components u_p and v_p are inferred from the initial responses $I_{p;k}, k \in \{1, \ldots, N\}$ at each location p by summing them up according to

$$\begin{pmatrix} u_p \\ v_p \end{pmatrix} = \sum_{k=1}^{N} I_k \cdot \begin{pmatrix} \cos(2\pi(k-1)/N) \\ -\sin(2\pi(k-1)/N) \end{pmatrix}, \tag{33}$$

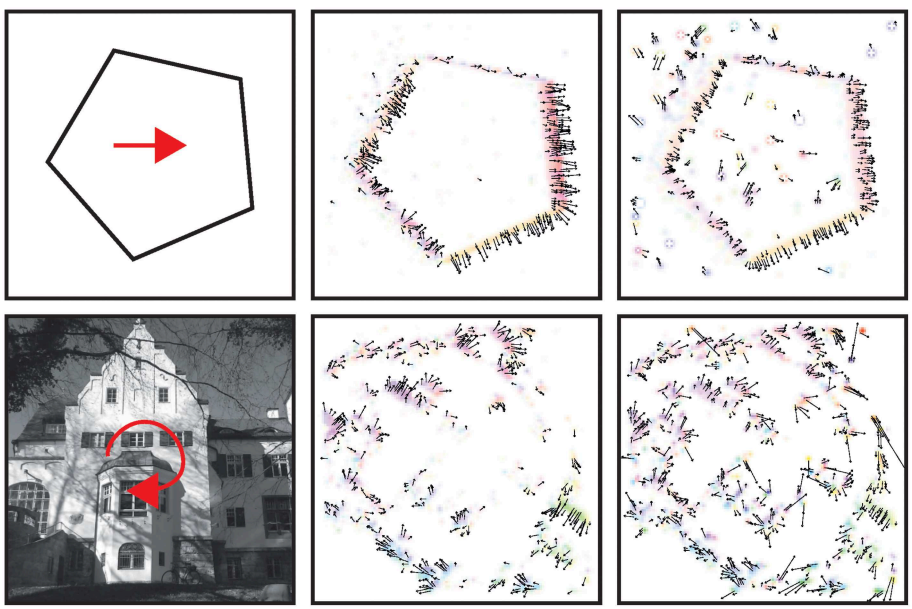

FIGURE 8 | Responses to input stimuli with translatory and rotational motion. From left to right: Test stimulus and vector field of initial motion estimation using the filter mechanism in Equation (25) and after normalization (red arrows are not part of stimulus; only two representative stimuli are shown due to space constraints). **First row:** Translatory motion stimulus illustrates that a majority of the responses point into the normal flow-direction, i.e., orthogonal to the stimulus boundaries. **Last row:** A rotational stimulus has been employed to validate that the filter also works for different speeds (slow motion close to the center and fast motion at the more distant regions). See Section 3.2 for details about the normalization mechanism. A comparison of initial and normalized flow estimation demonstrates that responses within line segments are reduced while responses at corners or noise are enhanced (that could be compensated by feedback from higher stages Brosch and Neumann, 2014b).

effectively implementing a local vector addition of component estimates. The tests utilize stimuli of translatory and rotational motion. The visualized results (**Figure 8**) demonstrate that the filter based approach robustly computes estimates of contour motion, i.e., locations of apparently moving contrasts and object boundaries (Barranco et al., 2014).

3.2. Response Normalization

A well known problem to motion detection is the estimation of ambiguous motion at e.g., straight contours (aperture problem). Locally only the normal flow direction can be measured which might not coincide with the true direction because the motion component parallel to a contrast edge is unknown (**Figure 9**, left). As suggested in Tsui et al. (2010), normalization can help to suppress responses at ambiguous parts of a contour (i.e., the inner parts of an extended contrast or line) and to enhance responses at line ends or sharp corners (c.f. **Figure 9B**). **Figure 9** shows motion histograms of the tilted bar in **Figure 8** (top) as a result of the initial filtering in the model (left) and with normalization (right). These results indicate that normalization significantly improves the histograms to better represent the true motion direction (**Figure 9A**; blue lines).

In Section 2.2.5, we point out that divisive normalization can effectively approximate radial Gaussianization, i.e., a reduction of the dependency between components within a population code. Here, we empirically validate that the divisive normalization described in Equation (31) indeed reduces the dependency within the population of motion selective cells. We quantify the statistical dependency of the multivariate representation by using multi-information (MI) (Studený and Vejnarová, 1998), which is defined as the Kullback-Leibler divergence (Cover and Thomas, 2006; Lyu and Simoncelli, 2009a) between the joint distribution $p(x_1, x_2, \ldots, x_d)$ and the product of its marginals

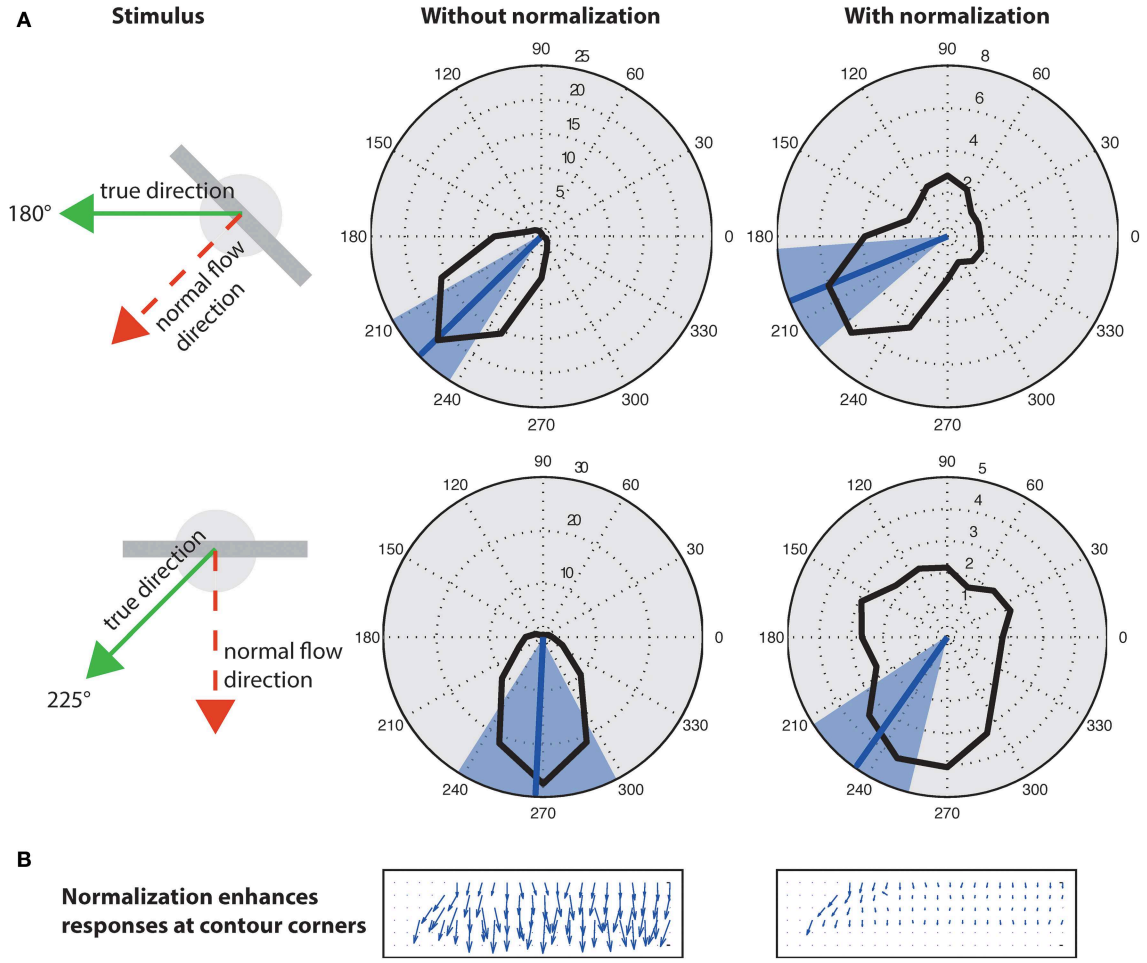

FIGURE 9 | Effect of normalization on initial flow estimation of two oblique bars moved with 45° difference to their orientation. (A), Left: Initial estimates elicit maximum response for normal flow direction (blue line; blue areas indicates standard deviation). Right: Surround inhibition enhances responses at the corners, effectively biasing the estimate toward the true motion direction. The circles on the elongated bar stimulus show the size of kernel weighting functions. **(B)** Normalization suppresses responses "within" the line and enhances responses at its endings (line ends, corners) that encode the true motion direction (c.f. Guo et al., 2006; Tsui et al., 2010).

$$MI(I) = D_{KL}\left(p(I)\middle\|\prod_k p(I_k)\right) = \left[\sum_{k=1}^{d} H(I_k)\right] - H(I) \quad (34)$$

$$= \sum_{x_i\in\mathcal{I}_1}\sum_{x_i\in\mathcal{I}_2}\cdots\sum_{x_i\in\mathcal{I}_d} p(x_1, x_2, \ldots, x_d) \cdot \log\frac{p(x_1, x_2, \ldots, x_d)}{p(x_1)p(x_2)\cdots p(x_d)}, \quad (35)$$

where $H(I) = \int p(I)\log(p(I))\,dI$ is the differential entropy of the representation I, and $H(I_k)$ denotes the differential entropy of the kth component of I (Lyu and Simoncelli, 2009a). To calculate the required probability estimates, we employ binary variables indicating motion for $d = 4$ movement directions. As theoretically predicted by the connection to radial Gaussianization, the MI for the stimulus shown in **Figure 9** is reduced from $MI(I) = 0.042$ (0.090 for the second example) before normalization to $MI(I_{norm}) = 0.028$ (0.027 for the second example) after the normalization stage. Thus, divisive normalization employed here does not entirely decorrelate the movement representation (which would imply $MI(I_{norm}) = 0$) but significantly reduces it.

3.3. Spatio-Temporal Filtering and Transparent Motion

Unlike the motion of opaque surfaces transparent motion is perceived when multiple motions are presented in the same part of visual space. Few computational model mechanisms have been proposed in the literature that allow to segregate multiple motions (see e.g., Raudies and Neumann, 2010; Raudies et al., 2011 which include recent overviews). All such model approaches are based on frame-based inputs. For that reason, we investigate how transparent motion induced by random dot patterns moving in different directions is represented in event-clouds originating from DVSs. In general, filter-based mechanisms are able to encode estimated motions for multiple directions at a single location. In contrast, it is not possible to fit a plane at positions where two (or multiple) event clouds generated by, for example, two crossing pedestrians intersect without applying additional knowledge. The filter mechanisms proposed in this work naturally encode motion directions within the uncertainty of the integration fields (c.f. **Figures 10A,B**). In order to build such a filter bank, the frequency space in **Figure 10B** needs to be sampled properly in accordance with the theoretical analysis outlined in Section 2 (c.f. **Table 1**).

To test the encoding of motion transparency, we probed the model by using simulated event-based sensor outputs of two superimposed random-dot patterns moving in orthogonal directions with the same speed. The spatio-temporal event-cloud generated by the moving dots is rather noisy and the component motions appear rather indistinguishable by eye. **Figure 10C** shows such events for individual dots and integrated over a small temporal window (directions are indicated by the blue and red arrows for illustrative purposes). As can be seen in **Figure 10D** the filter response clearly encodes both movement directions which could not be achieved by a plane-fitting approach without incorporating knowledge about the number of movement directions.

4. Discussion

This paper investigates mechanisms for motion estimation given event-based input generation and representation. The proposed mechanism has been motivated from the perspective of sampling the plenoptic function such that specific temporal changes in the optic array are registered by the sensory device. The temporal sampling is based on significant changes in the (log) luminance distribution at individual sensory elements (pixels). These operate at a very low latency by generating events whenever the local luminance function has undergone a super-threshold increment or decrement. This is fundamentally different from common frame-based approaches of image acquisition where a full image is recorded at fixed intervals leading to a largely redundant

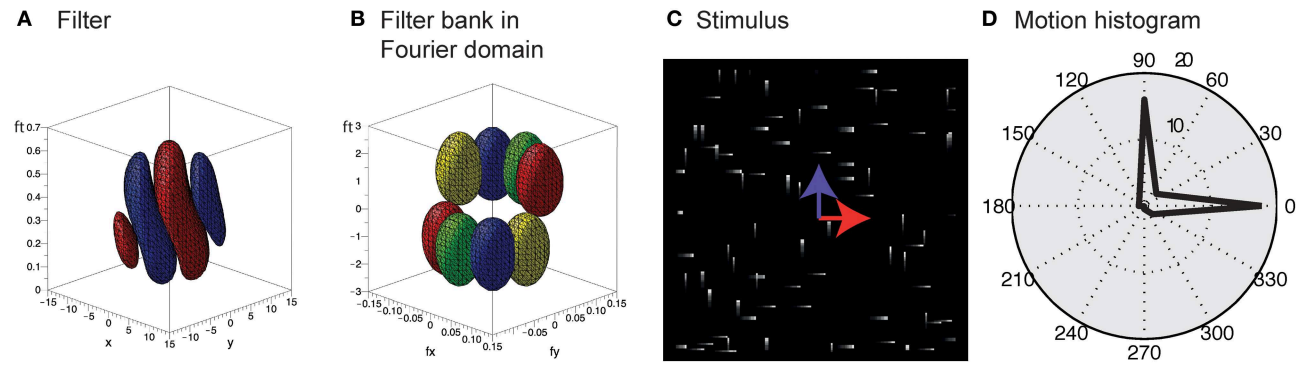

FIGURE 10 | Encoding of motion transparency in the proposed model. (A) Illustration of a single spatio-temporal filter (surfaces indicate that $F = \pm 0.0005$ for red/blue). Note that this filter resembles measurements of, for example, the cat's striate cortex (e.g., DeAngelis et al., 1995, Their Figure 2B). **(B)** Illustration of the preferred frequencies (surfaces indicate that $|\hat{F}| = 0.1$) of four filters of a filter bank in the Fourier domain [red pair of ellipsoids corresponds to Fourier spectrum of the filter shown in **(A)**]. Note that the combined ellipsoids sample the frequency space, with each pair responding to a certain speed and motion direction. **(C)** Stimulus consisting of a random dot pattern of dots moving to the right or to the top with equal speeds. **(D)** Motion histogram of filter responses. While it is not possible to fit a plane to the resulting event cloud, the proposed filter based approach encodes both movement directions.

TABLE 1 | Effect of different settings of parameters f_0 and μ_{bi} on the speed selectivity (in *pixel*/*s*), i.e., f_t^0/f_x^0, with f_x^0 and f_t^0 maximizing $|\hat{F}|$ for $\sigma = 25$.

$f_0 \backslash \mu_{bi}$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
0.04	100.65	50.33	33.55	25.16	20.13	16.78	14.38	12.58	11.18
0.05	78.76	39.38	26.26	19.69	15.75	13.13	11.25	9.85	8.75
0.06	65.10	32.55	21.70	16.28	13.02	10.85	9.30	8.14	7.23
0.07	55.67	27.84	18.56	13.92	11.13	9.28	7.95	6.96	6.18
0.08	48.69	24.34	16.23	12.17	9.74	8.11	6.96	6.09	5.41
0.09	43.28	21.64	14.42	10.82	8.66	7.21	6.18	5.41	4.81
0.10	38.95	19.48	12.98	9.74	7.79	6.49	5.56	4.87	4.33
0.11	35.41	17.70	11.80	8.85	7.08	5.90	5.06	4.43	3.93
0.12	32.46	16.23	10.82	8.11	6.49	5.41	4.64	4.06	3.61

The different parameter configurations allow to realize filter banks selective for a wide range of different speeds. If, for example, motion with a speed of about 10*pixels*/*s needs to be detected, the table reveals that the parameters can be chosen as* $(f_0, \mu_{bi}) = (0.10, 0.20)$ *or as* $(0.05, 0.40)$*, for example. Further adaptation of the standard deviations of the spatial and temporal kernels according to our theoretical results allows to realize optimal sampling of the Fourier-domain. For large enough* σ *(as for this table), the speed-selectivity does hardly depend on the parameter* σ*. For small* σ*, however, we noticed a strong impact which needs to be considered in the creation of a properly tuned filter bank.*

signal representations. Our focus is on motion computation and the proposed approach is different from previous approaches in several respects. In a nutshell, our paper makes three main contributions:

- We first investigate fundamental aspects of the local structure of lightfields for stationary observers and local contrast motion of the spatio-temporal luminance function. In particular, we emphasize the structure of local contrast information in the space-time domain and their encoding by events to build up an address-event representation (AER).
- Based on these results we derive several constraints on the kind of information that can be extracted from event-based sensory acquisition using the AER principle. This allows us to challenge several previous approaches and to develop a unified formulation in a common framework of event-based motion detection.
- We have shown that response normalization as part of a canonical microcircuit for motion detection is also applicable for event-based flow for which it reduces motion ambiguity and contributes to making the localized measures of filtering statistically more independent.

These different findings will be discussed in more detail in the following sections.

4.1. Previous Related Computational Models

So far, only relatively few investigations have been published that report on how classical approaches developed in computer vision can be adapted to event-based sensory input and how the quality of the results changes depending on the new data representation framework. Examples are Benosman et al. (2012, 2014) for optical flow computation and (Rogister et al., 2012; Piatkowska et al., 2013; Camuñas Mesa et al., 2014) for stereo vision. Furthermore, other authors show future applications of this new sensor technology that have the potential to provide fast, robust and highly efficient sensory processing in various domains and challenging scenarios (e.g., Fu et al., 2008; Drazen et al., 2011). Even further, most recent work has elucidated how fast event-based sensing technology can be utilized to improve the performance of computer vision motion estimation approaches and how frame-based imagery may help stabilizing the raw event-based motion processing (Barranco et al., 2014).

We here focus on the detection of flow from spatio-temporal motion on the basis of event-based sensor input. We utilize the dynamic-vision sensor (DVS) that emulates the major processing cascade of the retina from sensors to ganglion cells (Lichtsteiner et al., 2008; Liu and Delbruck, 2010). Based on the formulation of a local spatio-temporal surface patch at a significant luminance transition that moves along either direction, we have first categorized event-based flow estimation models. This allows us to provide a more systematic overview and to identify rather principled approaches. Based in these prerequisites, we have shown that gradient-based methods like (Benosman et al., 2012) are generally not stable in terms of their input feature estimation. The main reason is rooted in the potentially very small number of events generated at a single location (c.f. **Figure 1**). Based on these investigations we have further shown that the numerical approximation of the gradients, like in Benosman et al. (2012), has methodological deficiencies that may lead to inconclusive motion estimates. On formal grounds, we have demonstrated that a gradient-based motion detection and integration scheme, using the scheme of Lucas and Kanade (1981), can be utilized to numerically estimate second-order spatio-temporal derivatives on a function that represents the temporal derivative of the luminance distribution. This requires to employ proper numerical difference schemes which also demonstrates the disadvantage of increased noise sensitivity (Section 2.1.3).

In contrast, methods exploiting the local structure of the cloud of events are more robust in general. Here, we compared different approaches. First, we reviewed methods fitting an oriented plane to the event cloud. We derived equations which demonstrate that the orientation parameters of the plane directly encode the velocity [see Equation (18)]. The benefit of such an approach against the above-mentioned numerical derivative scheme is that it works even in the case of only a few generated events. Of course, the goodness of fit depends on the size of the spatio-temporal neighborhood. However, if we consider a neighborhood that is too small then the plane fit may eventually become arbitrary and

thus instable. If the neighborhood is too large then the chances increase that the event cloud contains structure that is not well approximated by a local plane. This also applies to the case of multiple motions, such as in the case of, e.g., occlusions due to opposite motions, limp motion in articulations, or in case of transparent motion stimuli.

Based on these insights we suggest a novel filter that samples the event-cloud along different spatio-temporal orientations. Its construction "reverses" the singular-value decomposition conducted of V1 receptive fields to construct direction-selective cells with spatio-temporally inseparable receptive fields (De Valois and Cottaris, 1998; De Valois et al., 2000). The conducted theoretical analysis allows to realize a spatio-temporally selective filter bank. Our investigation is similar to Escobar et al. (2009) who seek to specify the spatio-temporal selectivity. In contrast, our mechanism is directly derived from physiological findings. Perhaps the most similar scheme in comparison to our model is the one proposed by Adelson and Bergen (1985) which also suggests to derive spatio-temporally selective kernels by superposing different receptive fields. In their work, a spatial quadrature pair and two bi-phasic temporal kernels (in contrast to the mono- and bi-phasic kernels employed in our work) are combined (Adelson and Bergen, 1985) (compare also the review Emerson et al., 1992 and Borst and Egelhaaf, 1993). This scheme was motivated to resemble the spatio-temporal correlation scheme for motion detection (Hassenstein and Reichardt, 1956; Reichardt, 1957). In contrast to their approach, we rely upon the superposition of space-time separable filters with *out-of-phase* temporal modulation filter-responses. In addition to the main analysis, our test applications of the model implementation successfully demonstrate the functionality of such initial filtering for motion detection from spatio-temporal event clouds.

Compared to plane-fitting models (as suggested by, e.g., Benosman et al., 2014) we have shown that our model has the advantage that it can encode multiple motion directions at a single location, such as, e.g., (semi-) transparent motion (**Figure 10**; compare, e.g., Snowden et al., 1991; Treue et al., 2000; see e.g., Raudies and Neumann, 2010; Raudies et al., 2011 for a detailed discussion of motion transparency computation).

4.2. Non-Linear Response Normalization by Divisive Inhibition

In order to account for non-linearities in the response properties of cortical cells (Carandini et al., 1997) several models have been proposed to arrive at a neural circuit to define canonical computational mechanism (e.g., Kouh and Poggio, 2008; Carandini and Heeger, 2012). These and other models employ a mechanism of divisive inhibition of the surround activity (also used here) that has been suggested to explain findings ranging from gain control (Ayaz and Chance, 2009; Louie et al., 2011) over attention effects (Reynolds and Heeger, 2009; Lee and Maunsell, 2009; Montijn et al., 2012) to normalization in multi-sensory integration (Ohshiro et al., 2011). Tsui et al. (2010) have demonstrated that cells in the motion-sensitive area MT can properly respond to motion directions even for tilted bars although the normal flow directions signaled by component sensitive V1 cells should bias the motion selectivity in a direction orthogonal to the tilt direction. These authors suggest a divisive normalization that operates upon the static filters of oriented contrast filtering *before* the separate temporal filter. Such a scheme is rather implausible mechanistically. We therefore developed a scheme that employs the pool normalization after the stage of spatio-temporal event-input filtering (c.f. Brosch and Neumann, 2014b). The simulation results using oriented bar stimuli further confirm findings of Guo et al. (2006) in which enhanced responses were shown at the bar ends while the responses along the extended boundary of the bar are significantly reduced [consistent with earlier investigations Bolz and Gilbert, 1986]. While Escobar et al. (2008) showed that such a reduction of uncertainty can be achieved by using subtractive surround inhibition the proposal by Bayerl and Neumann (2004) suggests that feedback can reduce such redundant aperture responses. Taken together, the proposed model not only demonstrates that response normalization of initial motion detection successfully operates for event-based representations but suggests a reasonably simple account for the recent experimental observations (Tsui et al., 2010) using lateral interactions.

Based on statistical investigations, a decorrelation of the responses of a group of cells into rather independent components has been suggested in Lyu and Simoncelli (2008a, 2009a), dubbed *radial Gaussianization* to account for the broadening of the tuning curves. Since we showed certain similarities but also deviations from the model proposed here, we employed an information theoretic measure which confirms that the normalization scheme decorrelates input representations by decreasing the multi-information even without special parameter learning from a test set (Studený and Vejnarová, 1998; Lyu and Simoncelli, 2009a,b). This might be beneficial in light of coding principles (to support a sparse coding mechanism, Olshausen and Field, 2004) and to better deal with the variability of the overall motion stimulus configuration. For example, most model mechanisms have been employed by assuming (implicitly or explicitly) that the motion can be approximated locally by translatory motion. However, for cases of rotations, the intersection-of-constraints mechanism (Adelson and Movshon, 1982) fails as there is no common point of intersection from local estimates (Caplovitz et al., 2006). We suggest that such a stage of normalization in real-world motions reduces the response to ambiguous parts of a stimulus, like the center of an extended contrast. At the same time due to the reduced mutual dependency of individual responses in a population the rotation components can be combined into a more global configuration more easily. This is exemplified by demonstrating the effective pushing of the motion response histogram toward the true motion direction (**Figure 9**) similar to Tsui et al. (2010) (see Pack and Born, 2001 for a discussion of an account to solve the aperture problem in area MT).

4.3. Summary

Motion estimation from the output of an asynchronous event-based vision sensor requires adapted methods. Here, we conducted for the first time a theoretical investigation that systematically categorizes event-based flow estimation models with respect to their underlying methods, namely gradient-based

methods and algorithms exploiting the locally approximated plane-like structure of the cloud of events. In addition to analyzing existing gradient-based methods inconsistently mixing first and second order derivatives we proposed a novel consistent gradient-based algorithm. Even further, we showed that gradient-based methods in general suffer from strong noise originating from the limited number of events occurring at a single location. Methods exploiting the local plane-like shape of the event-cloud, on the other hand, were shown to be suitable for motion originating from a single object. In addition, we derived an explicit formula to derive the velocity from the parameters of the plane. For filter-based approaches, we proposed and analyzed a novel biologically inspired algorithm and demonstrated that it can also deal with motion transparency, i.e., it can represent different motion directions at a single location. Finally, we analyzed the impact of a stage of response normalization. We demonstrated that it is applicable to flow originating from event-based vision sensors, that it reduces motion ambiguity, and that it improves statistical independence of motion responses. All the theoretical findings were underpinned by simulation results which confirm that the model robustly estimates flow from event-based vision sensors.

Author Contributions

Designing the models/experiments: TB, ST, and HN. Mathematical and theoretical analysis: TB and HN. Spatio-temporal filter-analysis: TB. Experimental investigations: ST. Manuscript preparation: TB, ST, and HN.

Acknowledgments

The work has been supported by the Transregional Collaborative Research Center "A Companion Technology for Cognitive Technical Systems" (SFB/TR-62) funded by the German Research Foundation (DFG). We thank the reviewers for their careful reading and constructive criticism that helped to improve the manuscript.

References

Adelson, E. H., and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* 2, 284–299. doi: 10.1364/JOSAA.2.000284

Adelson, E. H., and Bergen, J. R. (1991). "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, eds M. S. Landy and J. A. Movshon (Cambridge, MA: MIT Press), 3–20.

Adelson, E. H., and Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature* 300, 523–525. doi: 10.1038/300523a0

Ayaz, A., and Chance, F. S. (2009). Gain modulation of neuronal responses by subtractive and divisive mechanisms of inhibition. *J. Neurophysiol.* 101, 958–968. doi: 10.1152/jn.90547.2008

Barranco, F., Fermüller, C., and Aloimonos, Y. (2014). Contour motion estimation for asynchronous event–driven cameras. *Proc. IEEE* 102, 1537–1556. doi: 10.1109/JPROC.2014.2347207

Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *Int. J. Comput. Vis.* 12, 43–77. doi: 10.1007/BF01420984

Bayerl, P., and Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Comput.* 16, 2041–2066. doi: 10.1162/0899766041732404

Benosman, R., Clercq, C., Lagorce, X., Ieng, S.-H., and Bartolozzi, C. (2014). Event–based visual flow. *Neural Netw. Learn. Syst.* 25, 407–417. doi: 10.1109/TNNLS.2013.2273537

Benosman, R., Ieng, S.-H., Clercq, C., Bartolozzi, C., and Srinivasan, M. (2012). Asynchronous frameless event–based optical flow. *Neural Netw.* 27, 32–37. doi: 10.1016/j.neunet.2011.11.001

Bolz, J., and Gilbert, C. D. (1986), Generation of end–Inhibition in the visual cortex via interlaminar connections. *Nature* 320, 362–365. doi: 10.1038/320362a0

Bonin, V., Mante, V., and Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. *J. Neurosci.* 25, 10844–10856. doi: 10.1523/JNEUROSCI.3562-05.2005

Borst, A., and Egelhaaf, M. (1993). "Detecting visual motion: theory and models," in *Visual Motion and its Role in the Stabilization of Gaze*, eds F. A. Miles and J. Wallman (Amsterdam: Elsevier), 3–27.

Braddick, O. J., Wishart, K. A., and Curran, W. (2002). Directional performance in motion transparency. *Vis. Res.* 42, 1237–1248. doi: 10.1016/S0042-6989(02)00018-4

Brosch, T., and Neumann, H. (2012). "The brain's sequential parallelism: perceptual decision-making and early sensory responses," in *ICONIP (Part II), Volume 7664 of LNCS* (Berlin; Heidelberg), 41–50.

Brosch, T., and Neumann, H. (2014a). Computing with a canonical neural circuits model with pool normalization and modulating feedback. *Neural Comput.* 26, 2735–2789. doi: 10.1162/NECO_a_00675

Brosch, T., and Neumann, H. (2014b). Interaction of feedforward and feedback streams in visual cortex in a firing–rate model of columnar computations. *Neural Netw.* 54, 11–16. doi: 10.1016/j.neunet.2014.02.005

Camuñas Mesa, L. A., Serrano-Gotarredona, T., Ieng, S. H., Benosman, R. B., and Linares-Barranco, B. (2014). On the use of orientation filters for 3D reconstruction in event–driven stereo vision. *Front. Neurosci.* 8:48. doi: 10.3389/fnins.2014.00048

Caplovitz, G. P., Hsieh, P.-J., and Tse, P. U. (2006). Mechanisms underlying the perceived angular velocity of a rigidly rotating Object. *Vis. Res.* 46, 2877–2893. doi: 10.1016/j.visres.2006.02.026

Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136

Carandini, M., Heeger, D. J., and Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17, 8621–8644.

Cover, T. M., and Thomas, J. A. (2006). *Elements of Information Theory, 2nd Edn.* (Hoboken, NJ: Wiley & Sons).

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two–dimensional visual cortical filters. *J. Opt. Soc. Am.* 2, 1160–1169. doi: 10.1364/JOSAA.2.001160

De Valois, R. L., and Cottaris, N. P. (1998). Inputs to directionally selective simple cells in macaque striate cortex. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14488–14493. doi: 10.1073/pnas.95.24.14488

De Valois, R. L., Cottaris, N. P., Mahon, L. E., Elfar, S. D., and Wilson, J. A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vis. Res.* 40, 3685–3702. doi: 10.1016/S0042-6989(00)00210-8

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1995). Receptive–field dynamics in the central visual pathways. *Trends Neurosci.* 18, 451–458. doi: 10.1016/0166-2236(95)94496-R

Delbrück, T., and Liu, S.-C. (2004). A silicon early visual system as a model animal. *Vis. Res.* 44, 2083–2089. doi: 10.1016/j.visres.2004.03.021

Drazen, D., Lichtsteiner, P., Häfliger, P., Delbrück, T., and Jensen, A. (2011). Toward real–time particle tracking using an event–based dynamic vision sensor. *Exp. Fluids* 51, 1465–1469. doi: 10.1007/s00348-011-1207-y

Edwards, M., and Nishida, S. (1999). Global–motion detection with transparent–motion signals. *Vis. Res.* 39, 2239–2249. doi: 10.1016/S0042-6989(98)00325-3

Emerson, R. C., Bergen, J. R., and Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vis. Res.* 32, 203–218. doi: 10.1016/0042-6989(92)90130-B

Escobar, M.-J., Masson, G. S., and Kornprobst, P. (2008). "A Simple mechanism to reproduce the neural solution of The Aperture Problem in Monkey Area MT," in *Proceedings of the Second French Conference on Computational Neuroscience* (Marseille).

Escobar, M.-J., Masson, G. S., Vieville, T., and Kornprobst, P. (2009). Action recognition using a bio–inspired feedforward spiking network. *Int. J. Comput. Vis.* 82, 284–301. doi: 10.1007/s11263-008-0201-1

Fermüller, C., and Aloimonos, Y. (1995). Qualitative egomotion. *IJCV* 15, 7–29. doi: 10.1007/BF01450848

Frégnac, Y., Monier, C., Chavane, F., Baudot, P., and Graham, L. (2003). Shunting inhibition a silent step in visual cortical computation. *J. Physiol.* 97, 441–451. doi: 10.1016/j.jphysparis.2004.02.004

Fu, Z., Delbruck, T., Lichtsteiner, P., and Culurciello, E. (2008). An address–event fall detector for assisted living applications. *Biomed. Circ. Syst.* 2, 88–96. doi: 10.1109/TBCAS.2008.924448

Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo* 11, 227–235. doi: 10.2307/1574154

Gibson, J. J. (1986). *The Ecological Approach to the Visual Perception of Pictures*, Hillsdale, NJ: Psychology Press.

Guo, K., Robertson, R., Nevado, A., Pulgarin, M., Mahmoodi, S., and Young, M. P. (2006). Primary visual cortex neurons that contribute to resolve the aperture problem. *Neuroscience* 138, 1397–1406. doi: 10.1016/j.neuroscience.2005.12.016

Hassenstein, B., and Reichardt, W. (1956). Systemtheoretische analyse der Zeit, Reihenfolgen und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Z. Naturforschg.* 11b, 513–524.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neurosci.* 9, 191–197. doi: 10.1017/S0952523800009640

Horn, B. K. P., and Schunck, B. G. (1981). Determining optical flow. *Artif. Intell.* 17, 185–203. doi: 10.1016/0004-3702(81)90024-2

Koch, C. (1999). *Biophysics of Computation: Information Processing in Single Neurons*, New York, NY: Oxford University Press.

Kouh, M., and Poggio, T. (2008). A canonical neural circuit for cortical nonlinear operations. *Neural Comput.* 20, 1427–1451. doi: 10.1162/neco.2008.02-07-466

Lee, J., and Maunsell, J. H. R. (2009). A normalization model of attentional modulation of single unit responses. *PLoS ONE* 4:e4651. doi: 10.1371/journal.pone.0004651

Lichtsteiner, P., Posch, C., and Delbruck, T. (2008). A 128×128 120 dB 15 μs Latency asynchronous temporal contrast vision sensor. *Solid–State Circ. IEEE J.* 43, 566–576. doi: 10.1109/JSSC.2007.914337

Liu, S.-C., and Delbruck, T. (2010). Neuromorphic sensory systems. *Curr. Opin. Neurobiol.* 20, 288–295. doi: 10.1016/j.conb.2010.03.007

Louie, K., Grattan, L. E., and Glimcher, P. W. (2011). Reward value–based gain control: divisive normalization in parietal cortex. *J. Neurosci.* 31, 10627–10639. doi: 10.1523/JNEUROSCI.1237-11.2011

Lucas, B. D., and Kanade, T. (1981). "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 1981 DARPA Image Understanding Workshop* (San Francisco, CA), 121–130.

Lyu, S., and Simoncelli, E. P. (2008a). *Nonlinear Extraction of 'Independent Components' of elliptically symmetric Densities using radial Gaussianization*, Technical report, Courant Institute of Mathematical Sciences, New York University.

Lyu, S., and Simoncelli, E. P. (2008b). "Nonlinear image representation using divisive normalization," in *IEEE Conference on Computer Vision and Pattern Recognition, 2008* (IEEE), 1–8.

Lyu, S., and Simoncelli, E. P. (2009a). Nonlinear extraction of independent components of natural images using radial Gaussianization. *Neural Comput.* 21, 1485–1519. doi: 10.1162/neco.2009.04-08-773

Lyu, S., and Simoncelli, E. P. (2009b), Reducing statistical dependencies in natural signals using radial Gaussianization. *Adv. Neural Inform. Process. Syst.* 21, 1–8.

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am.* 70, 1297–1300. doi: 10.1364/JOSA.70.001297

Mead, C. (1990). Neuromorphic electronic systems. *Proc. IEEE* 78, 1629–1636. doi: 10.1109/5.58356

Montijn, J. S., Klink, P. C., and van Wezel, R. J. A. (2012). Divisive normalization and neuronal oscillations in a single hierarchical framework of selective visual attention. *Font. Neural Circ.* 6, 1–17. doi: 10.3389/fncir.2012.00022

Nakayama, K., and Silverman, G. H. (1988). The aperture Problem–I. perception of nonrigidity and motion direction in translating sinusoidal lines. *Vis. Res.* 28, 739–746. doi: 10.1016/0042-6989(88)90052-1

Neumann, H., and Ottenberg, K. (1992). "Estimating Ramp–Edge Attributes from Scale–Space," in *Signal Processing VI: Theories and Applications, Vol. 1* (Elsevier), 603–606.

Ohshiro, T., Angelaki, D. E., and DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nat. Neurosci.* 14, 775–782. doi: 10.1038/nn.2815

Olshausen, B. A., and Field, D. J. (2004). Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* 14, 481–487. doi: 10.1016/j.conb.2004.07.007

Pack, C. C., and Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature* 409, 1040–1042. doi: 10.1038/35059085

Piatkowska, E., Belbachir, A. N., and Gelautz, M. (2013). "Asynchronous stereo vision for event-driven dynamic stereo sensor using an adaptive cooperative approach," in *International Conference on Computer Vision Workshops* (Sydney: IEEE), 45–50. doi: 10.1109/ICCVW.2013.13

Raudies, F., Mingolla, E., and Neumann, H. (2011). A model of motion transparency processing with local center–surround interactions and feedback. *Neural Comput.* 23, 2868–2914. doi: 10.1162/NECO_a_00193

Raudies, F., and Neumann, H. (2010). A model of neural mechanisms in monocular transparent motion perception. *J. Physiol. Paris* 104, 71–83. doi: 10.1016/j.jphysparis.2009.11.010

Reichardt, W. (1957). Autokorrelations–Auswertung als Funktionsprinzip des Zentralnervensystems (bei der optischen Bewegungswahrnehmung eines Insektes). *Z. Naturforschg.* 12b, 448–457.

Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002

Rogister, P., Benosman, R., Ieng, S.-H., Lichtsteiner, P., and Delbruck, T. (2012). Asynchronous event–based binocular stereo matching. *Neural Netw. Learn.* 23, 347–353. doi: 10.1109/TNNLS.2011.2180025

Sceniak, M. P., Ringach, D. L., Hawken, M. J., and Shapley, R. (1999). Contrast's effect on spatial summation by Macaque V1 Neurons. *Nat. Neurosci.* 2, 733–739. doi: 10.1038/11197

Snowden, R. J., Treue, S., Erickson, R. G., and Andersen, R. A. (1991). The response of area MT and VI neurons to transparent motion. *J. Neurosci.* 11, 2768–2785.

Snowden, R. J., and Verstraten, F. A. J. (1999). Motion transparency making models of motion perception transparent. *Trends Cogn. Sci.* 3, 369–377. doi: 10.1016/S1364-6613(99)01381-9

Studený, M., and Vejnarová, J. (1998). "The Multiinformation Function as a Tool for Measuring stochastic Dependence," in *Learning in Graphical Models (Kluwer Academic Publishers), Proceedings of the NATO Advanced Study Institute on Learning in graphical models,* 261–97. doi: 10.1007/978-94-011-5014-9_10

Treue, S., Hol, K., and Rauber, H.-J. (2000). Seeing multiple directions of motion–physiology and psychophysics. *Nat. Neurosci.* 3, 270–276. doi: 10.1038/72985

Tschechne, S., Brosch, T., Sailer, R., von Egloffstein, N., Abdul-Kreem, L. I., and Neumann, H. (2014). "On event-based motion detection and integration," in *8th International Conference on Bio–inspired Information and Communications Technologies* (ACM) (Boston), 298–305.

Tsui, J. M. G., Hunter, J. N., Born, R. T., and Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *J. Neurophysiol.* 103, 3123–3138. doi: 10.1152/jn.00654.2009

Wallach, H. (1935). Über visuell wahrgenommene Bewegungsrichtung. *Psychol. Forschung* 20, 325–380. doi: 10.1007/BF02409790

Wuerger, S., Shapley, R., and Rubin, N. (1996). "On the Visually Perceived Direction of Motion" by Hans Wallach: 60 Years Later. *Perception* 25, 1317–1367. doi: 10.1068/p251317

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.