Focus Is All You Need: Loss Functions For Event-based Vision

Guillermo Gallego [†] Mathias Gehrig [†] Davide Scaramuzza [†]

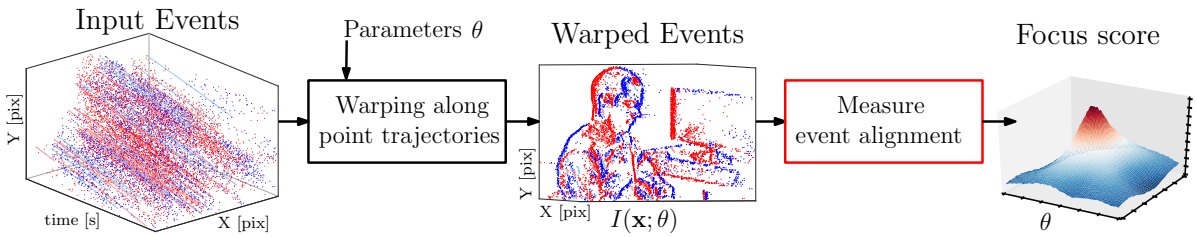


Figure 1: *Motion Compensation Framework.* Events in a space-time window are warped according to point trajectories described by motion parameters $\boldsymbol{\theta}$, resulting in an image of warped events (IWE) $I(\mathbf{x};\boldsymbol{\theta})$. Then, a *focus loss function* of I measures how well events are aligned along the point trajectories. This work proposes multiple focus loss functions for event alignment (last block in the figure and Table 1) for tasks such as rotational motion, depth and optical flow estimation.

Abstract

Event cameras are novel vision sensors that output pixel-level brightness changes ("events") instead of traditional video frames. These asynchronous sensors offer several advantages over traditional cameras, such as, high temporal resolution, very high dynamic range, and no motion blur. To unlock the potential of such sensors, motion compensation methods have been recently proposed. We present a collection and taxonomy of twenty two objective functions to analyze event alignment in motion compensation approaches (Fig. 1). We call them focus loss functions *since they have strong connections with functions used in traditional shape-from-focus applications. The proposed loss functions allow bringing mature computer vision tools to the realm of event cameras. We compare the accuracy and runtime performance of all loss functions on a publicly available dataset, and conclude that the variance, the gradient and the Laplacian magnitudes are among the best loss functions. The applicability of the loss functions is shown on multiple tasks: rotational motion, depth and optical flow estimation. The proposed focus loss functions allow to unlock the outstanding properties of event cameras.*

1. Introduction

Event cameras are bio-inspired sensors that work radically different from traditional cameras. Instead of capturing brightness images at a fixed rate, they measure brightness *changes* asynchronously. This results in a stream of *events*, which encode the time, location and polarity (sign) of the brightness changes. Event cameras, such as the Dynamic Vision Sensor (DVS) [1] posses outstanding properties compared to traditional cameras: very high dynamic range (140 dB vs. 60 dB), high temporal resolution (in the order of µs), and do not suffer from motion blur. Hence, event cameras have a large potential to tackle challenging scenarios for standard cameras (such as high speed and high dynamic range) in tracking [2–9], depth estimation [10–19], Simultaneous Localization and Mapping [20–27], and recognition [28–32], among other applications. However, novel methods are required to process the unconventional output of these sensors in order to unlock their potential.

Motion compensation approaches [15, 19, 33–42] have been recently introduced for processing the visual information acquired by event cameras. They have proven successful for the estimation of motion (optical flow) [34–36, 38, 39], camera motion [33, 35], depth (3D reconstruction) [15, 19, 38, 39] as well as segmentation [36, 40, 41]. The main idea of such methods consists of searching for point trajectories on the image plane that maximize *event alignment* [33, 35] (Fig. 1, right), which is measured using some loss function of the events warped according to such trajectories. The best trajectories produce sharp, motion compensated images that reveal the brightness patterns causing the events (Fig. 1, middle).

In this work, we build upon the motion compensation framework [35] and extend it to include twenty more loss functions for applications such as ego-motion, depth and

[†] Dept. Informatics, Univ. of Zurich and Dept. Neuroinformatics, Univ. of Zurich and ETH Zurich

Focus Loss Function	Type	Spatial?	Goal
Variance (4) [33, 35]	Statistical	No	max
Mean Square (9) [33, 36]	Statistical	No	max
Mean Absolute Deviation (10)	Statistical	No	max
Mean Absolute Value (11)	Statistical	No	max
Entropy (12)	Statistical	No	max
Image Area (8)	Statistical	No	min
Image Range (13)	Statistical	No	max
Local Variance (14)	Statistical	Yes	max
Local Mean Square	Statistical	Yes	max
Local Mean Absolute Dev.	Statistical	Yes	max
Local Mean Absolute Value	Statistical	Yes	max
Moran's Index (17)	Statistical	Yes	min
Geary's Contiguity Ratio (18)	Statistical	Yes	max
Gradient Magnitude (5)	Derivative	Yes	max
Laplacian Magnitude (6)	Derivative	Yes	max
Hessian Magnitude (7)	Derivative	Yes	max
Difference of Gaussians	Derivative	Yes	max
Laplacian of the Gaussian	Derivative	Yes	max
Variance of Laplacian	Stat. & Deriv.	Yes	max
Variance of Gradient	Stat. & Deriv.	Yes	max
Variance of Squared Gradient	Stat. & Deriv.	Yes	max
Mean Timestamp on Pixel [37]	Statistical	No	min

Table 1: List of objective functions considered.

optical flow estimation. We ask the question: *What are good metrics of event alignment?* In answering, we noticed strong connections between the proposed metrics (Table 1) and those used for shape-from-focus and autofocus in conventional, frame-based cameras [43, 44], and so, we called the event alignment metrics "*focus loss functions*". The extended framework allows mature computer vision tools to be used on event data while taking into account all the information of the events (asynchronous timestamps and polarity). Additionally, it sheds light on the event-alignment goal of functions used in existing motion-compensation works and provides a taxonomy of loss functions for event data.

Contributions. In summary, our contributions are:

1. The introduction and comparison of twenty two focus loss functions for event-based processing, many of which are developed from basic principles, such as the "area" of the image of warped events.
2. Connecting the topics of shape-from-focus, autofocus and event-processing by the similar set of functions used, thus allowing to bring mature analysis tools from the former topics into the realm of event cameras.
3. A thorough evaluation on a recent dataset [45], comparing the accuracy and computational effort of the proposed focus loss functions, and showing how they can be used for depth and optical flow estimation.

The rest of the paper is organized as follows. Section 2 reviews the working principle of event cameras. Section 3 summarizes the motion compensation method and extends it with the proposed focus loss functions. Experiments are carried out in Section 4 comparing the loss functions, and conclusions are drawn in Section 5.

2. Event-based Camera Working Principle

Event-based cameras, such as the DVS [1], have independent pixels that output "events" in response to *brightness changes*. Specifically, if $L(\mathbf{x}, t) \doteq \log I(\mathbf{x}, t)$ is the logarithmic brightness at pixel $\mathbf{x} \doteq (x, y)^\top$ on the image plane, the DVS generates an event $e_k \doteq (\mathbf{x}_k, t_k, p_k)$ if the change in logarithmic brightness at pixel $\mathbf{x}_k$ reaches a threshold C (e.g., 10-15% relative change):

$$\Delta L \doteq L(\mathbf{x}_k, t_k) - L(\mathbf{x}_k, t_k - \Delta t_k) = p_k\, C, \qquad (1)$$

where t_k is the timestamp of the event, Δt_k is the time since the previous event at the same pixel $\mathbf{x}_k$ and $p_k \in \{+1, -1\}$ is the event polarity (i.e., sign of the brightness change).

Therefore, each pixel has its own sampling rate (which depends on the visual input) and outputs data proportionally to the amount of motion in the scene. An event camera does not produce images at a constant rate, but rather a stream of *asynchronous*, sparse events in space-time (Fig. 1, left).

3. Methodology

3.1. Motion Compensation Framework

In short, the method in [35] seeks to find the point-trajectories on the image plane that maximize the alignment of corresponding events (i.e., those triggered by the same scene edge). Event alignment is measured by the *strength of the edges* of an image of warped events (IWE), which is obtained by aggregating events along candidate point trajectories (Fig. 1). In particular, [35] proposes to measure edge strength (which is directly related to *image contrast* [46]) using the variance of the IWE.

More specifically, the events in a set $\mathcal{E} = \{e_k\}_{k=1}^{N_e}$ are geometrically transformed

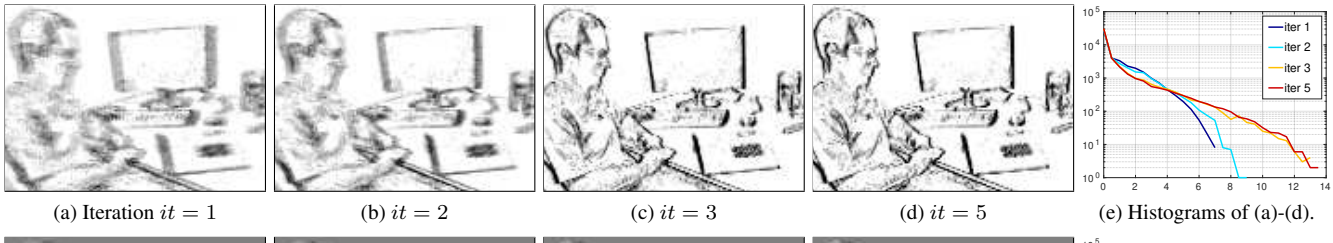
$$e_k \doteq (\mathbf{x}_k, t_k, p_k) \;\mapsto\; e'_k \doteq (\mathbf{x}'_k, t_{\text{ref}}, p_k) \qquad (2)$$

according to a point-trajectory model $\mathbf{W}$, resulting in a set of warped events $\mathcal{E}' = \{e'_k\}_{k=1}^{N_e}$ "flattened" at a reference time t_{ref}. The warp $\mathbf{x}'_k = \mathbf{W}(\mathbf{x}_k, t_k; \boldsymbol{\theta})$ transports each event along the point trajectory that passes through it (Fig. 1, left), until t_{ref} is reached, thus, taking into account the space-time coordinates of the event. The vector $\boldsymbol{\theta}$ parametrizes the point trajectories, and hence contains the motion or scene parameters.

The image (or histogram) of warped events (IWE) is given by accumulating events along the point trajectories:

$$I(\mathbf{x}; \boldsymbol{\theta}) \doteq \sum_{k=1}^{N_e} b_k\, \delta(\mathbf{x} - \mathbf{x}'_k(\boldsymbol{\theta})), \qquad (3)$$

where each pixel $\mathbf{x}$ sums the values b_k of the warped events $\mathbf{x}'_k$ that fall within it ($b_k = p_k$ if polarity is used or $b_k = 1$ if polarity is not used; see Fig. 2). In practice, the Dirac delta δ is replaced by a smooth approximation, such as a Gaussian, $\delta(\mathbf{x} - \boldsymbol{\mu}) \approx \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \epsilon^2 \mathtt{Id})$, with typically $\epsilon = 1$ pixel.



(a) Iteration $it = 1$ (b) $it = 2$ (c) $it = 3$ (d) $it = 5$ (e) Histograms of (a)-(d).

(f) Iteration $it = 1$ (g) $it = 2$ (h) $it = 3$ (i) $it = 5$ (j) Histograms of (f)-(i).

Figure 2: Evolution of the image of warped events (IWE, (3)) as the focus loss is optimized, showing how the IWE sharpens as the motion parameters $\boldsymbol{\theta}$ are estimated. Motion blur (left) due to event misalignment (in the example, dominantly in horizontal direction) decreases as warped events become better aligned (right). Top row: without polarity ($b_k = 1$ in (3)); Bottom row: with polarity ($b_k = p_k$ in (3)). The last column shows the histograms of the images; the peak at zero corresponds to the pixels with no events (white in the top row, gray in the bottom row).

The contrast [35] of the IWE (3) is given by its variance:

$$\mathrm{Var}\big(I(\mathbf{x};\boldsymbol{\theta})\big) \doteq \frac{1}{|\Omega|}\int_{\Omega}(I(\mathbf{x};\boldsymbol{\theta}) - \mu_I)^2 d\mathbf{x} \qquad (4)$$

with mean $\mu_I \doteq \frac{1}{|\Omega|}\int_{\Omega} I(\mathbf{x};\boldsymbol{\theta})d\mathbf{x}$. Discretizing into pixels, it becomes $\mathrm{Var}(I) = \frac{1}{N_p}\sum_{i,j}(h_{ij} - \mu_I)^2$, where N_p is the number of pixels of $I = (h_{ij})$ and $\mu_I = \frac{1}{N_p}\sum_{i,j} h_{i,j}$.

3.2. Our Proposal: Focus Loss Functions

In (3) a group of events $\mathcal{E}$ has been effectively converted, using the spatio-temporal transformation, into an image representation (see Fig. 2), aggregating the little information from individual events into a larger, more descriptive piece of information. Here, we propose to exploit the advantages that such an image representation offers, namely bringing in tools from image processing (histograms, convolutions, Fourier transforms, etc.) to analyze event alignment.

We study different image-based event alignment metrics (i.e., loss functions), such as (4); that is, we study different objective functions that can be used in the second block of the diagram of Fig. 1 (highlighted in red). To the best of our knowledge we are the first to address the following three related topics: (*i*) establishing connections between event alignment metrics and so-called "focus measures" in shape-from-focus (SFF) and autofocus (AF) in conventional frame-based imaging [43, 44], (*ii*) comparing the performance of the different metrics on event-based vision problems, and (*iii*) showing that it is possible to design new focus metrics tailored to edge-like images like the IWE.

Table 1 lists the multiple focus loss functions studied, the majority of which are newly proposed, and categorizes them according to their nature. The functions are classified according to whether they are based on statistical or derivative operators (or their combination), according to whether they depend on the spatial arrangement of the pixel values or not, and according to whether they are maximized or minimized.

The next sections present the focus loss functions using two image characteristics related to edge strength: *sharpness* (Section 3.3) and *dispersion* (Section 3.4). But first, let us discuss the variance loss, since it motivates several focus losses.

Loss Function: Image Variance (Contrast). An event alignment metric used in [33, 35] is the variance of the IWE (4), known as the RMS *contrast* in image processing [46]. The variance is a statistical measure of *dispersion*, in this case, of the pixel values of the IWE, regardless of their spatial arrangement. Thus, event alignment (i.e., edge strength of the IWE) is here assessed using statistical principles (Table 1). The motion parameters $\boldsymbol{\theta}$ that best fit the events are obtained by maximizing the variance (4). In the example of Fig. 2, it is clear that as event alignment increases, so does the visual contrast of the IWE.

Fourier Interpretation: Using (*i*) the formula relating the variance of a signal to its mean square (MS) and mean, $\mathrm{Var}(I) = \mathrm{MS} - \mu_I^2$, and (*ii*) the interpretations of MS and squared mean as the total "energy" and DC component of the signal, respectively, yields that the variance represents the AC component of the signal, i.e., the energy of the oscillating, high frequency, content of the IWE. Hence, the IWE comes into focus (Fig. 2) by increasing its AC component; the DC component does not change significantly during the process. This interpretation is further developed next.

3.3. Image Sharpness

In the Fourier domain, sharp images (i.e., those with strong edges, e.g., Fig. 2, right) have a significant amount of energy concentrated at high frequencies. The high-frequency content of an image can be assessed by measuring the magnitude of its derivative since derivative operators act as band-pass or high-pass filters. The magnitude is given by any norm; however, the L^2 norm is preferred since it admits an inner product interpretation and simple derivatives. The following losses are built upon this idea, using first and second derivatives, respectively. Similar losses, based on the DCT or wavelet transforms [44] instead of the Fourier transform are also possible. The main idea remains the same: measure the high frequency content of the IWE (i.e., edge strength) and find the motion parameters that maximize it, therefore maximizing event alignment.

Loss Function: Magnitude of Image Gradient. Event alignment is achieved by seeking the parameters $\boldsymbol{\theta}$ of the point-trajectories that maximize

$$\|\nabla I\|^2_{L^2(\Omega)} \doteq \int_\Omega \|\nabla I(\mathbf{x})\|^2 d\mathbf{x}, \tag{5}$$

where $\nabla I = (I_x, I_y)^\top$ is the gradient of the IWE I (subscripts indicate derivative: $I_x \equiv \partial I/\partial x$), and its magnitude is measured by an L^p norm (suppl. material), e.g., the (squared) L^2 norm: $\|\nabla I\|^2_{L^2(\Omega)} = \int_\Omega (I_x^2(\mathbf{x}) + I_y^2(\mathbf{x}))d\mathbf{x}$.

Loss Function: Magnitude of Image Hessian. For these loss functions, event alignment is attained by maximizing the magnitude of the second derivatives (i.e., Hessian) of the IWE, $\mathrm{Hess}(I)$. We use the (squared) norm of the Laplacian,

$$\|\Delta I\|^2_{L^2(\Omega)} \doteq \|I_{xx} + I_{yy}\|^2_{L^2(\Omega)}, \tag{6}$$

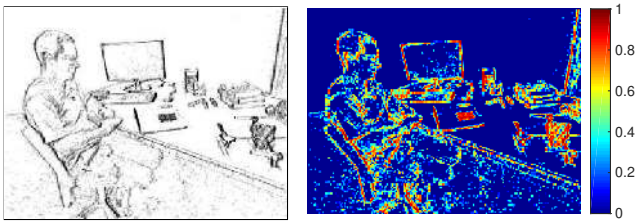
or the (squared) Frobenius norm of the Hessian,

$$\|\,\mathrm{Hess}(I)\|^2_{L^2(\Omega)} \doteq \|I_{xx}\|^2_{L^2(\Omega)} + \|I_{yy}\|^2_{L^2(\Omega)} + 2\|I_{xy}\|^2_{L^2(\Omega)}. \tag{7}$$

Loss Functions: DoG and LoG. The magnitude of the output of established band-pass filters, such as the Difference of Gaussians (DoG) and the Laplacian of the Gaussian (LoG), can also be used to assess the sharpness of the IWE.

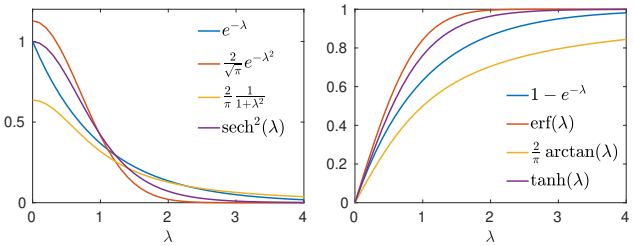
Loss Function: Image Area. Intuitively, sharp, motion-compensated IWEs, have *thinner edges* than uncompensated ones (Fig. 2). We now provide a definition of the "thickness" or "support" of edge-like images (see supplementary material) and use it for event alignment. We propose to minimize the support (i.e., area) of the IWE (3):

$$\mathrm{supp}(I(\mathbf{x};\boldsymbol{\theta})) = \int_\Omega \left(F(I(\mathbf{x};\boldsymbol{\theta})) - F(0)\right) d\mathbf{x}, \tag{8}$$



(a) IWE (3) $I(\mathbf{x};\boldsymbol{\theta})$ with $b_k = 1$. (b) Support map $F(I(\mathbf{x};\boldsymbol{\theta})) - F(0)$.

Figure 3: Support map of an IWE. Red regions contribute more to the support than blue regions. The integral of the support map in Fig. 3b gives the support (8) of the IWE.



(a) Weighting functions $\rho(\lambda)$. (b) Primitive functions $F(\lambda)$.

Figure 4: Weighting functions $\rho(\lambda)$ and their primitives $F(\lambda)$, used to define the support (8) of the IWE.

where $F(\lambda) \doteq \int \rho(\lambda) d\lambda$ is the primitive of a decreasing weighting function $\rho(\lambda) \geq 0$, for $\lambda \geq 0$. Fig. 3 shows an IWE and its corresponding support map, pseudo-colored as a heat map: red regions (large IWE values) contribute more to the support (8) than blue regions (small IWE values).

Our definition of the support of the IWE is very flexible, since it allows for different choices of the weighting function (Fig. 4). Specifically, we consider four choices:

1. Exponential: $\rho(\lambda) = e^{-\lambda}$, $F(\lambda) = 1 - e^{-\lambda}$.
2. Gaussian: $\rho(\lambda) = \frac{2}{\sqrt{\pi}}e^{-\lambda^2}$, $F(\lambda) = \mathrm{erf}(\lambda)$.
3. Lorentzian: $\rho(\lambda) = \frac{2}{(1+\lambda^2)\pi}$, $F(\lambda) = \frac{2}{\pi}\arctan(\lambda)$.
4. Hyperbolic: $\rho(\lambda) = \mathrm{sech}^2(\lambda)$, $F(\lambda) = \tanh(\lambda)$.

3.4. Image Dispersion (Statistics)

Besides the variance (4), other ways to measure dispersion are possible. These may be categorized according to their spatial character: global (operating on the IWE pixels regardless of their arrangement) or local. Among the global metrics for dispersion we consider the mean square (MS), the mean absolute deviation (MAD), the mean absolute value (MAV), the entropy and the range. In the group of local dispersion losses we have: local versions of the above-mentioned global losses (variance, MS, MAD, MAV, etc.) and metrics of spatial autocorrelation, such as Moran's I index [47] and Geary's Contiguity Ratio [48].

Loss Function: Mean Square. The Mean Square (MS)

$$\mathrm{MS}(I) = \|I(\mathbf{x};\boldsymbol{\theta})\|^2_{L^2(\Omega)} \,/\, |\Omega|, \tag{9}$$

also measures the dispersion of the IWE (with respect to zero). As anticipated in the Fourier interpretation of the variance, the MS is the total energy of the image, which comprises an oscillation (i.e., dispersion) part and a constant part (DC component or squared mean).

Loss Function: Mean Absolute Deviation (MAD) and Mean Absolute Value (MAV). The analogous of the variance and the MS using the L^1 norm are the MAD,

$$\mathrm{MAD}(I) = \|I(\mathbf{x};\boldsymbol{\theta}) - \mu_I\|_{L^1(\Omega)} \,/\, |\Omega|, \tag{10}$$

and the mean absolute value,

$$\mathrm{MAV}(I) = \|I(\mathbf{x};\boldsymbol{\theta})\|_{L^1(\Omega)} \,/\, |\Omega|, \tag{11}$$

respectively. The MAV provides a valid loss function to estimate $\boldsymbol{\theta}$ if polarity is used. However, if polarity is not used, the MAV coincides with the IWE mean, μ_I (which counts the warped events in Ω), and since this value is typically constant, the MAV does not provide enough information to estimate $\boldsymbol{\theta}$ in this case (as will be noted in Table 2).

Loss Function: Image Entropy. Information entropy measures the uncertainty or spread (i.e., dispersion) of a distribution, or, equivalently, the reciprocal of its concentration. This approach consists of maximizing Shannon's entropy of the random variable given by the IWE pixels:

$$H\left(p_I(z)\right) = -\int_{-\infty}^{\infty} p_I(z) \log p_I(z) dz \tag{12}$$

The PDF of the IWE is approximated by its histogram (normalized to unit area). Comparing the PDFs of the IWEs before and after motion compensation (last column of Fig. 2), we observe two effects of event alignment: (i) the peak of the distribution at $I = 0$ increases since the image regions with almost no events grow (corresponding to homogeneous regions of the brightness signal), and (ii) the distribution spreads out away from zero: larger values of $|I(\mathbf{x})|$ are achieved. Hence, the PDF of the motion-compensated IWE is more concentrated around zero and more spread out away from zero than the uncompensated one. Concentrating the PDF means decreasing the entropy, whereas spreading it out means increasing the entropy. To obtain a high contrast IWE, with sharper edges, the second approach must dominate. Hence, parameters $\boldsymbol{\theta}$ are obtained by maximizing the entropy of the IWE (12).

Entropy can also be interpreted as a measure of diversity in information content, and since (i) sharp images contain more information than blurred ones [49], and (ii) our goal is to have sharp images for better event alignment, thus our goal is to maximize the information content of the IWE, which is done by maximizing its entropy.

Loss Function: Image Range. As is well known, contrast is a measure of the oscillation (i.e., dispersion) of a signal with respect to its background (e.g., Michelson contrast), and the range of a signal, $\mathrm{range}(I) = I_{\max} - I_{\min}$, measures its maximum oscillation. Hence, maximizing the range of the IWE provides an alternative way to achieve event alignment. However, the min and max statistics of an image are brittle, since they can drastically change by modifying two pixels. Using the tools that lead to (8) (suppl. material), we propose to measure image range more sensibly by means of the support of the image PDF,

$$\mathrm{supp}(p_I) = \int_{-\infty}^{\infty} \left(F(p_I(z)) - F(0)\right) dz, \tag{13}$$

where $F(\lambda)$ is a primitive of the weight function $\rho(\lambda) \geq 0$ (as in Fig. 4). Fig. 2 illustrates how the range of the IWE is related to its sharpness: as the IWE comes into focus, the image range (support of the histogram) increases.

Next, we present focus loss functions based on local versions of the above global statistics.

Loss Function: Local Variance, MS, MAD and MAV. Mimicking (5), which aggregates local measures of image sharpness to produce a global score, we may aggregate the local variance, MS, MAD or MAV of the IWE to produce a global dispersion score that we seek to maximize. For example, the aggregated local variance (ALV) of the IWE is

$$\mathrm{ALV}(I) \doteq \int_{\Omega} \mathrm{Var}(\mathbf{x}; I) d\mathbf{x}, \tag{14}$$

where the local variance in a neighborhood $B(\mathbf{x}) \subset \Omega$ centered about the point $\mathbf{x}$ is given by

$$\mathrm{Var}(\mathbf{x}; I) \doteq \frac{1}{|B(\mathbf{x})|} \int_{B(\mathbf{x})} \big(I(\mathbf{u};\boldsymbol{\theta}) - \mu(\mathbf{x}; I)\big)^2 d\mathbf{u}, \tag{15}$$

with local mean $\mu(\mathbf{x}; I) \doteq \int_{B(\mathbf{x})} I(\mathbf{v};\boldsymbol{\theta})\, d\mathbf{v} / |B(\mathbf{x})|$ and $|B(\mathbf{x})| = \int_{B(\mathbf{x})} d\mathbf{u}$. The local variance of an image (15) is an edge detector similar to the magnitude of the image gradient, $\|\nabla I\|$. It may be estimated using a weighted neighborhood $B(\mathbf{x})$ by means of convolutions with a smoothing kernel $G_\sigma(\mathbf{x})$, such as a Gaussian:

$$\mathrm{Var}(\mathbf{x}; I) \approx \left(I^2(\mathbf{x}) * G_\sigma(\mathbf{x})\right) - \left(I(\mathbf{x}) * G_\sigma(\mathbf{x})\right)^2. \tag{16}$$

Based on the above example, local versions of the MS, MAD and MAV can be derived (see suppl. material).

Loss Function: Spatial Autocorrelation. Spatial autocorrelation of the IWE can also be used to assess event alignment. We present two focus loss functions based on spatial autocorrelation: Moran's I and Geary's C indices.

Moran's I *index* is a number between -1 and 1 that evaluates whether the variable under study (i.e., the pixels of the IWE) is clustered, dispersed, or random. Letting $z_i = I(\mathbf{x}_i)$ be the value of the i-th pixel of the IWE, Moran's index is

$$\mathrm{Moran}(I) \doteq \frac{\sum_{i,j} w_{ij}(z_i - \bar{z})(z_j - \bar{z})/W}{\sum_i (z_i - \bar{z})^2 / N_p}, \tag{17}$$

where $\bar{z} \equiv \mu_I$ is the mean of I, N_p is the number of pixels of I, (w_{ij}) is a matrix of spatial weights with zeros on the diagonal (i.e., $w_{ii} = 0$) and $W = \sum_{i,j} w_{ij}$. The weights w_{ij} encode the spatial relation between pixels z_i and z_j: we use weights that decrease with the distance between pixels, e.g., $w_{ij} \propto \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_M^2)$, with $\sigma_M \approx 1$ pixel.

A positive Moran's index value indicates tendency toward clustering pixel values while a negative Moran's index value indicates tendency toward dispersion (dissimilar values are next to each other). Edges correspond to dissimilar values close to each other, and so, we seek to minimize Moran's index of the IWE.

Geary's Contiguity Ratio is a generalization of Von Neumann's ratio [50] of the mean square successive difference to the variance:

$$C(I) \doteq \frac{1}{2} \frac{\sum_{i,j} w_{ij}(z_i - z_j)^2/W}{\sum_i (z_i - \bar{z})^2/(N_p - 1)}. \tag{18}$$

It is non-negative. Values around 1 indicate lack of spatial autocorrelation; values near zero are positively correlated, and values larger than 1 are negatively correlated. Moran's I and Geary's C are inversely related, thus event alignment is achieved by maximizing Geary's C of the IWE.

3.5. Dispersion of Image Sharpness Values

We have also combined statistics-based loss functions with derivative-based ones, yielding more loss functions (Table 1), such as the variance of the Laplacian of the IWE $\mathrm{Var}(\Delta I(\mathbf{x};\boldsymbol{\theta}))$, the variance of the magnitude of the IWE gradient $\mathrm{Var}(\|\nabla I(\mathbf{x};\boldsymbol{\theta})\|)$, and the variance of the squared magnitude of the IWE gradient $\mathrm{Var}(\|\nabla I(\mathbf{x};\boldsymbol{\theta})\|^2) \equiv \mathrm{Var}(I_x^2 + I_y^2)$. They apply statistical principles to local focus metrics based on neighborhood operations (convolutions).

3.6. Discussion of the Focus Loss Functions

Connection with Shape-from-Focus and Autofocus: Several of the proposed loss functions have been proven successful in shape-from-focus (SFF) and autofocus (AF) with conventional cameras [43, 44, 51], showing that there is a strong connection between these topics and event-based motion estimation. The principle to solve the frame-based and event-based problems is the same: maximize a focus score of the considered image or histogram. In the case of conventional cameras, SFF and AF maximize edge strength at each pixel of a focal stack of images (in order to infer depth). In the case of event cameras, the IWE (an image representation of the events) plays the role of the images in the focal stack: varying the parameters $\boldsymbol{\theta}$ produces a different "slice of the focal stack", which may be used not only to estimate depth, but also to estimate other types of parameters $\boldsymbol{\theta}$, such as optical flow, camera velocities, etc.

Spatial Dependency: Focus loss functions based on derivatives or on local statistics imply neighborhood operations (e.g., convolutions with nearby pixels of the IWE), thus they depend on the spatial arrangement of the IWE pixels (Table 1). Instead, global statistics (e.g., variance, entropy, etc.) do not directly depend on such spatial arrangement[1]. The image area loss functions (8) are integrals of point-wise functions of the IWE, and so, they do not depend on the spatial arrangement of the IWE pixels. The PDF of the IWE also does not have spatial dependency, nor do related losses, such as entropy (12) or range (13). Composite focus losses (Section 3.5), however, have spatial dependency since they are computed from image derivatives.

Fourier Interpretation: Focus losses based on derivatives admit an intuitive interpretation in the Fourier domain: they measure the energy content in the high frequencies (i.e., edges) of the IWE (Section 3.3). Some of the statistical focus losses also admit a frequency interpretation. For example, image variance quantifies the energy of the AC portion (i.e., oscillation) of the IWE, and the MS measures the energy of both, the AC and DC components. Other focus functions, such as entropy, do not admit such a straightforward Fourier interpretation related to edge strength.

4. Experiments

In this section we compare the performance of the proposed focus loss functions, which depends on multiple factors, such as, the task for which it is used, the number of events processed, the data used, and, certainly, implementation. These factors produce a test space with an intractable combinatorial size, and so, we choose the task for best assessing accuracy on real data and then provide qualitative results on other tasks: depth and optical flow estimation.

4.1. Accuracy and Timing Evaluation

Rotational camera motion estimation provides a good scenario for assessing accuracy since camera motion can be reliably obtained with precise motion capture systems. The acquisition of accurate per-event optical flow or depth is less reliable, since it depends on additional depth sensors, such as RGB-D or LiDAR, which are prone to noise.

Table 2 compares the accuracy and timing of the focus loss functions on the above scenario, using data from [45]. The data consists of sequences of one minute duration, with increasing camera motion, reaching up to $\approx \pm 1000\,^\circ/\mathrm{s}$, that is recorded using a motion-capture system with sub-millimeter precision at 200 Hz. Each entry in Table 2 consists of an experiment with more than 160 million events, processed in groups of $N_e = 30000$ events. The focus loss functions are optimized using non-linear conjugate gradient, initialized by the estimated angular velocity $\boldsymbol{\theta}$ for the previous group of events. As the table reports, the angular velocity errors (difference between the estimated angu-

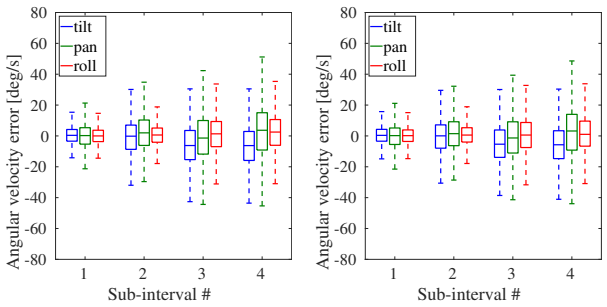
[1] The IWE consists of warped events, which depend on the location of the events; however, by "directly" we mean that the focus loss function, by itself, does not depend on the spatial arrangement of the IWE values.

Focus Loss Function	Boxes (RMS) $\omega_{\max} \approx \pm 670\,^\circ/\mathrm{s}$ w/o	Boxes (RMS) $\omega_{\max} \approx \pm 670\,^\circ/\mathrm{s}$ w/	Poster (RMS) $\omega_{\max} \approx \pm 1000\,^\circ/\mathrm{s}$ w/o	Poster (RMS) $\omega_{\max} \approx \pm 1000\,^\circ/\mathrm{s}$ w/	Time [µs]
Variance (4) [33, 35]	18.52	18.94	25.96	24.39	**16.90**
Mean Square (9) [33, 36]	19.93	19.02	34.10	26.31	25.11
Mean Absolute Deviation (10)	19.46	19.58	30.70	29.62	78.11
Mean Absolute Value	-	19.77	-	29.90	23.89
Entropy (12)	28.50	26.54	47.54	33.21	271.85
Area (8) (Exp)	31.50	19.54	43.12	26.40	160.56
Area (8) (Gaussian)	25.85	18.85	34.50	25.35	1098.64
Area (8) (Lorentzian)	32.43	20.98	35.86	26.57	777.98
Area (8) (Hyperbolic)	29.13	19.15	32.94	25.88	1438.15
Range (13) (Exp)	28.66	28.72	65.33	32.23	263.11
Local Variance (14)	18.21	18.40	25.44	24.15	78.48
Local Mean Square	24.81	19.86	33.95	26.47	137.20
Local Mean Absolute Dev.	21.37	18.74	61.89	25.29	177.15
Local Mean Absolute Value	-	24.10	-	30.37	243.58
Moran's Index (17)	24.28	23.43	32.40	30.96	116.39
Geary's Contiguity Ratio (18)	23.87	19.50	26.61	25.23	181.73
Gradient Magnitude (5)	**17.83**	18.10	**23.93**	23.58	128.46
Laplacian Magnitude (6)	18.32	**17.58**	24.91	23.67	293.80
Hessian Magnitude (7)	18.41	17.93	25.47	23.74	569.55
Difference of Gauss. (DoG)	20.85	19.25	24.50	**22.15**	189.90
Laplacian of the Gauss. (LoG)	20.36	17.77	25.15	24.01	127.65
Variance of Laplacian	18.26	18.01	26.59	23.62	327.60
Variance of Gradient	18.69	19.08	26.60	24.22	872.03
Variance of Squared Gradient	18.72	18.95	26.10	24.43	653.62
Mean Time on Pixel [37]	82.89	-	121.20	-	24.43

Table 2: *Accuracy and Timing Comparison of Focus Loss Functions*. RMS angular velocity errors (in $^\circ/\mathrm{s}$) of the motion compensation method [35] (with (w/) or without (w/o) polarity) with respect to motion-capture system. Processing $N_e = 30\,000$ events, warped onto an image of 240×180 pixels (DAVIS camera [52]), takes 2040.24 µs, the runtime of the focus loss functions is given in the last column. Sequences `boxes` and `poster` from dataset [45]. The best value per column is highlighted in bold.

lar velocity $\boldsymbol{\theta}$ that optimizes the focus loss function and the velocity provided by the motion-capture system) vary across loss functions. These columns summarize the RMS errors (with and without using polarity), which are reported in more detail in the supplementary material. Among the most accurate focus loss functions are the ones based on the derivatives of the IWE, high-pass filters and the ones based on the variance (global or local), with less than 2.6 % error. The entropy and support-based losses (area and range) are not as accurate, yet the errors are small compared to the angular velocity excursions ($\omega_{\max}$) in the sequences (less than 7 % error). In general, Table 2 shows that using polarity is beneficial; the results are slightly more accurate using polarity than not using it. The error boxplots of the gradient magnitude (5), one of the most accurate focus loss functions on Table 2, are given in Fig. 5.

Computational Cost. The last column of Table 2 compares the losses in terms of computational effort, on a single-core 2.7 GHz CPU. The variance is the fastest, with 17 µs (possibly due to efficient code in OpenCV), and its runtime is negligible compared to the time required to warp the events (≈2 ms). The MS and the MAV are also fast.



(a) Without event polarity. (b) With event polarity.

Figure 5: Box plots of angular velocity error (estimated vs. ground truth), per axis, for four subintervals of 15 s, using the gradient magnitude (5) as focus loss function. The method produces small errors: ≈18 $^\circ/\mathrm{s}$ RMS (see Table 2), i.e., 2.7 % with respect to $|\omega_{\max}| = 670\,^\circ/\mathrm{s}$.

Other focus functions take longer, due to effort spent in spatial operations (convolutions), and/or transcendental functions (e.g., exp). The variance is a notable trade-off between accuracy and speed[2]. Computing the IWE has $O(N_e)$ complexity, but it is, to some extent, parallelizable. Loss functions act on the IWE (with N_p pixels), thus they have $O(N_p)$ complexity.

4.2. Depth Estimation

We also qualitatively compared the loss functions in the context of depth estimation. Here, the IWE is a slice of the disparity space image (DSI) computed by back-projecting events into a volume [15]. The DSI plays the role of the focal stack in SFF and AF (Section 3.6). Depth at a pixel of the reference view ($\boldsymbol{\theta} \equiv Z$) is estimated by selecting the DSI slice with extremal focus along the optical ray of the pixel. Fig. 6 shows the focus loss functions along an optical ray for the events in a small space-time window around the pixel. The values of the focus losses come to an extremal at a common focal point (at depth ≈1.1 m). The plots show that, in general, the focus functions monotonically decrease/increase as the distance from the focal plane increases, forming the basin of attraction of the local extrema. Composite losses (Section 3.5) have a narrow peak, whereas others, such as the area losses have a wider peak.

4.3. Optical Flow Estimation

The profile of the loss function can also be visualized for 2D problems such as patch-based optical flow [35]. Events in a small space-time window (e.g., 15×15 pixels) are warped according to a feature flow vector $\boldsymbol{\theta} \equiv \mathbf{v}$: $\mathbf{x}'_k = \mathbf{x}_k - (t_k - t_{\text{ref}})\mathbf{v}$. Fig. 7 shows the profiles of several

[2]Implementation plays a major role in runtime; thus the above figures are illustrative. We built on top of the C++ Standard Library and OpenCV, but did not optimize the code or used function approximation [53, 54]. [37,40] suggest considerable speed-up factors if warping is done on a GPU.

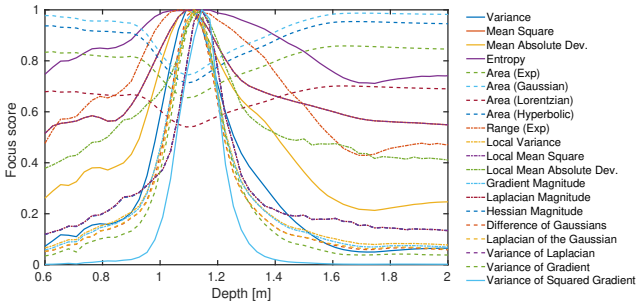
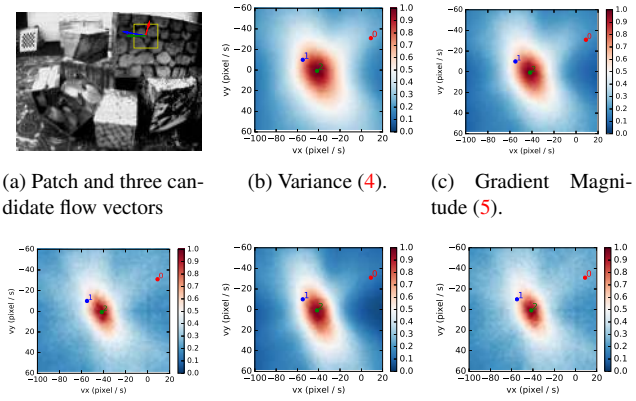


Figure 6: Depth estimation for a patch. Plots of several focus losses as a function of depth. Each focus loss curve is normalized (e.g., by its maximum), to stay in $[0, 1]$.



(a) Patch and three candidate flow vectors

(b) Variance (4).

(c) Gradient Magnitude (5).

(d) Laplacian Magnitude (6).

(e) Variance of Gradient Magnitude.

(f) Variance of Laplacian.

Figure 7: Optical flow of a patch (feature). Plots of several focus losses as a function of the two optical flow parameters of the feature. Each focus loss surface is normalized by its maximum. At the correct optical flow vector, some surfaces show a narrow peak, whereas others present a broader peak.

competitive focus losses. They all have a clear extremal at the location of the visually correct ground truth flow (point 0, green arrow in Fig. 7a). The Laplacian magnitude and its variance show the narrowest peaks. Plots of more loss functions are provided in the supplementary material.

4.4. Unsupervised Learning of Optical Flow

The last row of Table 2 evaluates the accuracy and timing of a loss function inspired in the time-based IWE of [37]: the variance of the per-pixel average timestamp of warped events. This loss function has been used in [39] for unsupervised training of a neural network (NN) that predicts dense optical flow from events [55]. However, this loss function is considerably less accurate than all other functions (Table 2). This suggests that (*i*) loss functions based on timestamps of warped events are not as accurate as those based on event count (the event timestamp is already taken into account during warping), (*ii*) there is considerable room for

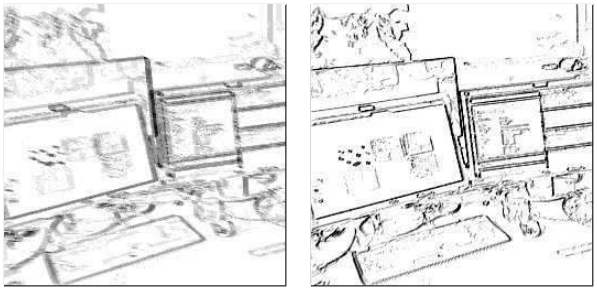


Figure 8: IWE before (Left) and after (Right) motion compensation by a neural network similar to [39], trained using the gradient magnitude loss (5).

improvement in unsupervised learning of optical flow if better loss function is used.

To probe the applicability of the loss functions to estimate dense optical flow, we trained a network inspired in [55] using (5) and a Charbonier prior on the flow derivative. The flow produced by the NN increases event alignment (Fig. 8). A deeper evaluation is left for future work.

5. Conclusion

We have extended motion compensation methods for event cameras with a library of twenty more loss functions that measure event alignment. Moreover, we have established a fundamental connection between the proposed loss functions and metrics commonly used in shape from focus with conventional cameras. This connection allows us to bring well-established analysis tools and concepts from image processing into the realm of event-based vision. The proposed functions act as focusing operators on the events, enabling us to estimate the point trajectories on the image plane followed by the objects causing the events. We have categorized the loss functions according to their capability to measure edge strength and dispersion, metrics of information content. Additionally, we have shown how to design new focus metrics tailored to edge-like images like the image of warped events. We have compared the performance of all focus metrics in terms of accuracy and time. Similarly to comparative studies in autofocus for digital photography applications [51], we conclude that the variance, the gradient magnitude and the Laplacian are among the best functions. Finally, we have shown the broad applicability of the functions to tackle essential problems in computer vision: ego-motion, depth, and optical flow estimation. We believe the proposed focus loss functions are key to taking advantage of the outstanding properties of event cameras, specially in unsupervised learning of structure and motion.

Acknowledgment

This work was supported by the Swiss National Center of Competence Research Robotics, the Swiss National Science Foundation and the SNSF-ERC Starting Grant.

References

[1] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck, "A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008. 1, 2

[2] Zhenjiang Ni, Sio-Hoï Ieng, Christoph Posch, Stéphane Régnier, and Ryad Benosman, "Visual tracking using neuromorphic asynchronous event-based cameras," *Neural Computation*, vol. 27, no. 4, pp. 925–953, 2015. 1

[3] Daniel Gehrig, Henri Rebecq, Guillermo Gallego, and Davide Scaramuzza, "Asynchronous, photometric feature tracking using events and frames," in *Eur. Conf. Comput. Vis. (ECCV)*, pp. 766–781, 2018. 1

[4] Elias Mueggler, Basil Huber, and Davide Scaramuzza, "Event-based, 6-DOF pose tracking for high-speed maneuvers," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, pp. 2761–2768, 2014. 1

[5] Andrea Censi and Davide Scaramuzza, "Low-latency event-based visual odometry," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, pp. 703–710, 2014. 1

[6] Guillermo Gallego, Jon E. A. Lund, Elias Mueggler, Henri Rebecq, Tobi Delbruck, and Davide Scaramuzza, "Event-based, 6-DOF camera tracking from photometric depth maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, pp. 2402–2412, Oct. 2018. 1

[7] David Tedaldi, Guillermo Gallego, Elias Mueggler, and Davide Scaramuzza, "Feature detection and tracking with the dynamic and active-pixel vision sensor (DAVIS)," in *Int. Conf. Event-Based Control, Comm. Signal Proc. (EBCCSP)*, 2016. 1

[8] Arren Glover and Chiara Bartolozzi, "Robust visual tracking with a freely-moving event camera," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, pp. 3769–3776, 2017. 1

[9] Ignacio Alzugaray and Margarita Chli, "Asynchronous corner detection and tracking for event cameras in real time," *IEEE Robot. Autom. Lett.*, vol. 3, pp. 3177–3184, Oct. 2018. 1

[10] Paul Rogister, Ryad Benosman, Sio-Hoi Ieng, Patrick Lichtsteiner, and Tobi Delbruck, "Asynchronous event-based binocular stereo matching," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 2, pp. 347–353, 2012. 1

[11] Stephan Schraml, Ahmed Nabil Belbachir, and Horst Bischof, "Event-driven stereo matching for real-time 3D panoramic vision," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 466–474, 2015. 1

[12] Nathan Matsuda, Oliver Cossairt, and Mohit Gupta, "MC3D: Motion contrast 3D scanning," in *IEEE Int. Conf. Comput. Photography (ICCP)*, pp. 1–10, 2015. 1

[13] Dongqing Zou, Ping Guo, Qiang Wang, Xiaotao Wang, Guangqi Shao, Feng Shi, Jia Li, and Paul-K. J. Park, "Context-aware event-driven stereo matching," in *IEEE Int. Conf. Image Process. (ICIP)*, pp. 1076–1080, 2016. 1

[14] Marc Osswald, Sio-Hoi Ieng, Ryad Benosman, and Giacomo Indiveri, "A spiking neural network model of 3D perception for event-based neuromorphic stereo vision systems," *Scientific Reports*, vol. 7, Jan. 2017. 1

[15] Henri Rebecq, Guillermo Gallego, Elias Mueggler, and Davide Scaramuzza, "EMVS: Event-based multi-view stereo—3D reconstruction with an event camera in real-time," *Int. J. Comput. Vis.*, vol. 126, pp. 1394–1414, Dec. 2018. 1, 7

[16] Zhen Xie, Shengyong Chen, and Garrick Orchard, "Event-based stereo depth estimation using belief propagation," *Front. Neurosci.*, vol. 11, Oct. 2017. 1

[17] Alexander Andreopoulos, Hirak J. Kashyap, Tapan K. Nayak, Arnon Amir, and Myron D. Flickner, "A low power, high throughput, fully event-based stereo system," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 7532–7542, 2018. 1

[18] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza, "Semi-dense 3D reconstruction with a stereo event camera," in *Eur. Conf. Comput. Vis. (ECCV)*, pp. 242–258, 2018. 1

[19] Alex Zihao Zhu, Yibo Chen, and Kostas Daniilidis, "Real-time time synchronized event-based stereo," in *Eur. Conf. Comput. Vis. (ECCV)*, pp. 438–452, 2018. 1

[20] David Weikersdorfer, Raoul Hoffmann, and Jörg Conradt, "Simultaneous localization and mapping for event-based vision systems," in *Int. Conf. Comput. Vis. Syst. (ICVS)*, pp. 133–142, 2013. 1

[21] Hanme Kim, Ankur Handa, Ryad Benosman, Sio-Hoi Ieng, and Andrew J. Davison, "Simultaneous mosaicing and tracking with an event camera," in *British Mach. Vis. Conf. (BMVC)*, 2014. 1

[22] Hanme Kim, Stefan Leutenegger, and Andrew J. Davison, "Real-time 3D reconstruction and 6-DoF tracking with an event camera," in *Eur. Conf. Comput. Vis. (ECCV)*, pp. 349–364, 2016. 1

[23] Henri Rebecq, Timo Horstschäfer, Guillermo Gallego, and Davide Scaramuzza, "EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real-time," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 593–600, 2017. 1

[24] Christian Reinbacher, Gottfried Munda, and Thomas Pock, "Real-time panoramic tracking for event cameras," in *IEEE Int. Conf. Comput. Photography (ICCP)*, pp. 1–9, 2017. 1

[25] Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis, "Event-based visual inertial odometry," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 5816–5824, 2017. 1

[26] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza, "Ultimate SLAM? combining events, images, and IMU for robust visual SLAM in HDR and high speed scenarios," *IEEE Robot. Autom. Lett.*, vol. 3, pp. 994–1001, Apr. 2018. 1

[27] Elias Mueggler, Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," *IEEE Trans. Robot.*, vol. 34, pp. 1425–1440, Dec. 2018. 1

[28] Jun Haeng Lee, Tobi Delbruck, Michael Pfeiffer, Paul K.J. Park, Chang-Woo Shin, Hyunsurk Ryu, and Byung Chang Kang, "Real-time gesture interface based on event-driven processing from stereo silicon retinas," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2250–2263, 2014. 1

[29] Garrick Orchard, Cedric Meyer, Ralph Etienne-Cummings, Christoph Posch, Nitish Thakor, and Ryad Benosman,

"HFirst: A temporal approach to object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2028–2040, 2015. 1

[30] Xavier Lagorce, Garrick Orchard, Francesco Gallupi, Bertram E. Shi, and Ryad Benosman, "HOTS: A hierarchy of event-based time-surfaces for pattern recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, pp. 1346–1359, July 2017. 1

[31] Arnon Amir, Brian Taba, David Berg, Timothy Melano, Jeffrey McKinstry, Carmelo Di Nolfo, Tapan Nayak, Alexander Andreopoulos, Guillaume Garreau, Marcela Mendoza, Jeff Kusnitz, Michael Debole, Steve Esser, Tobi Delbruck, Myron Flickner, and Dharmendra Modha, "A low power, fully event-based gesture recognition system," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 7388–7397, 2017. 1

[32] Amos Sironi, Manuele Brambilla, Nicolas Bourdis, Xavier Lagorce, and Ryad Benosman, "HATS: Histograms of averaged time surfaces for robust event-based object classification," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 1731–1740, 2018. 1

[33] Guillermo Gallego and Davide Scaramuzza, "Accurate angular velocity estimation with an event camera," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 632–639, 2017. 1, 2, 3, 7

[34] Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis, "Event-based feature tracking with probabilistic data association," in *IEEE Int. Conf. Robot. Autom. (ICRA)*, pp. 4465–4470, 2017. 1

[35] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, pp. 3867–3876, 2018. 1, 2, 3, 7

[36] Timo Stoffregen and Lindsay Kleeman, "Simultaneous optical flow and segmentation (SOFAS) using Dynamic Vision Sensor," in *Australasian Conf. Robot. Autom. (ACRA)*, 2017. 1, 2, 7

[37] Anton Mitrokhin, Cornelia Fermuller, Chethan Parameshwara, and Yiannis Aloimonos, "Event-based moving object detection and tracking," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 2018. 1, 2, 7, 8

[38] Chengxi Ye, Anton Mitrokhin, Chethan Parameshwara, Cornelia Fermüller, James A. Yorke, and Yiannis Aloimonos, "Unsupervised learning of dense optical flow and depth from sparse event data," *arXiv preprint arXiv:1809.08625*, 2018. 1

[39] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," *arXiv preprint arXiv:1812.08156*, 2018. 1, 8

[40] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza, "Event-based motion segmentation by motion compensation," *arXiv preprint arXiv:1904.01293*, 2019. 1, 7

[41] Anton Mitrokhin, Chengxi Ye, Cornelia Fermuller, Yiannis Aloimonos, and Tobi Delbruck, "EV-IMO: Motion segmentation dataset and learning pipeline for event cameras," *arXiv preprint arXiv:1903.07520*, 2019. 1

[42] Timo Stoffregen and Lindsay Kleeman, "Event cameras, contrast maximization and reward functions: an analysis," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019. 1

[43] Parikshit Sakurikar and P. J. Narayanan, "Composite focus measure for high quality depth maps," in *Int. Conf. Comput. Vis. (ICCV)*, pp. 1623–1631, 2017. 2, 3, 6

[44] Said Pertuz, Domenec Puig, and Miguel Angel García, "Analysis of focus measure operators for shape-from-focus," *Pattern Recog.*, vol. 46, no. 5, pp. 1415–1432, 2013. 2, 3, 4, 6

[45] Elias Mueggler, Henri Rebecq, Guillermo Gallego, Tobi Delbruck, and Davide Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *Int. J. Robot. Research*, vol. 36, no. 2, pp. 142–149, 2017. 2, 6, 7

[46] Rafael C. Gonzalez and Richard Eugene Woods, *Digital Image Processing*. Pearson Education, 2009. 2, 3

[47] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, p. 17, June 1950. 4

[48] R. C. Geary, "The contiguity ratio and statistical mapping," *The Incorporated Statistician*, vol. 5, pp. 115–146, Nov. 1954. 4

[49] Firestone Lawrence, Cook Kitty, Culp Kevin, Talsania Neil, and Preston Kendall, "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991. 5

[50] John von Neumann, "Distribution of the ratio of the mean square successive difference to the variance," *Ann. Math. Stat.*, vol. 12, pp. 367–395, Dec. 1941. 6

[51] Hashim Mir, Peter Xu, and Peter van Beek, "An extensive empirical evaluation of focus measures for digital photography," in *Proc. SPIE*, vol. 9023, pp. 1–11, 2014. 6, 8

[52] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck, "A 240x180 130dB 3us latency global shutter spatiotemporal vision sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014. 7

[53] Guillermo Gallego, Daniel Berjón, and Narciso García, "Optimal polygonal L_1 linearization and fast interpolation of nonlinear systems," *IEEE Trans. Circuits Syst. I*, vol. 61, pp. 3225–3234, Nov. 2014. 7

[54] Daniel Berjón, Guillermo Gallego, Carlos Cuevas, Francisco Morán, and Narciso García, "Optimal piecewise linear function approximation for GPU-based applications," *IEEE Trans. Cybern.*, vol. 46, pp. 2584–2595, Nov. 2016. 7

[55] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis, "EV-FlowNet: Self-supervised optical flow estimation for event-based cameras," in *Robotics: Science and Systems (RSS)*, 2018. 8