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

Guillermo Gallego [†]   Mathias Gehrig [†]   Davide Scaramuzza [†]
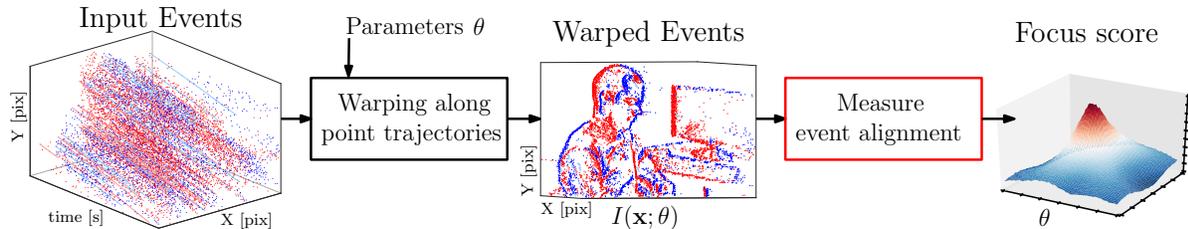
Figure 1: *Motion Compensation Framework.* Events in a space-time window are warped according to point trajectories described by motion parameters $\boldsymbol{\theta}$, resulting in an image of warped events (IWE) $I(\mathbf{x}; \boldsymbol{\theta})$. Then, a *focus loss function* of I measures how 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 bright-ness *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] possess outstanding properties compared to traditional cameras: very high dynamic range ($140\,\mathrm{dB}$ vs. $60\,\mathrm{dB}$), 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

1

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):

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

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

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

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:

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

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, $\delta(\mathbf{x} - \boldsymbol{\mu}) \approx \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \epsilon^2 \texttt{Id})$, with typically $\epsilon = 1$ pixel.

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

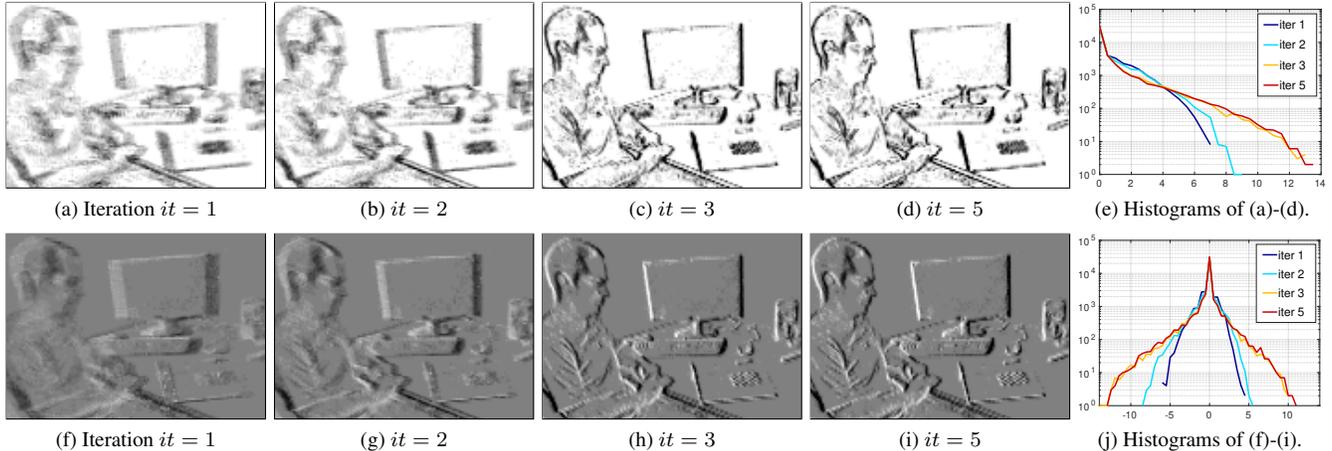|(f) Iteration $it = 1$ | (g) $it = 2$ | (h) $it = 3$ | (i) $it = 5$ | (j) Histograms of (f)-(i). |

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 θ 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:

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

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

3.3. 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

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

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, $\text{Hess}(I)$. We use the (squared) norm of the Laplacian,

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

or the (squared) Frobenius norm of the Hessian,

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

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):

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

(a) IWE (3) $I(\mathbf{x};\boldsymbol{\theta})$ with $b_k = 1$.   (b) Support map $F(I(\mathbf{x};\boldsymbol{\theta})) - F(0)$.
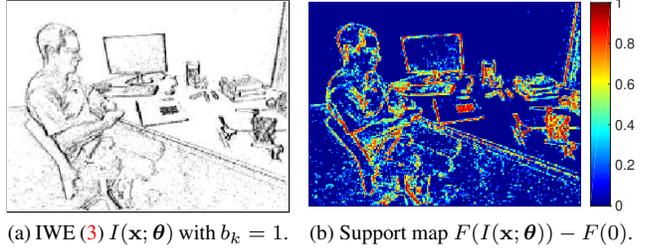
Figure 3: Support map of an IWE. Red regions contribute more to the support than blue regions. The integral of the support map in Fig. 3b gives the support (8) of the IWE.

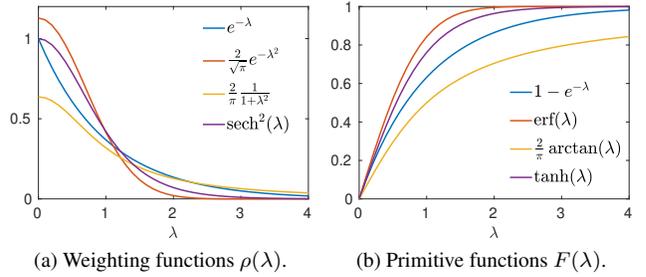(a) Weighting functions $\rho(\lambda)$.   (b) Primitive functions $F(\lambda)$.

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

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)

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

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,

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

and the mean absolute value,

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

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:

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

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,

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

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

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

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

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

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:

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

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

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

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:

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

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 $\approx \pm 1000\,°/\mathrm{s}$, that is recorded using a motion-capture system with submillimeter precision at $200\,\mathrm{Hz}$. 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.

6

Focus Loss Function	Boxes (RMS) $\omega_{max} \approx \pm 670\,°/s$		Poster (RMS) $\omega_{max} \approx \pm 1000\,°/s$		Time
	w/o	w/	w/o	w/	[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 $°/s$) 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 (ω_{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 ($\approx 2\,$ms). The MS and the MAV are also fast.

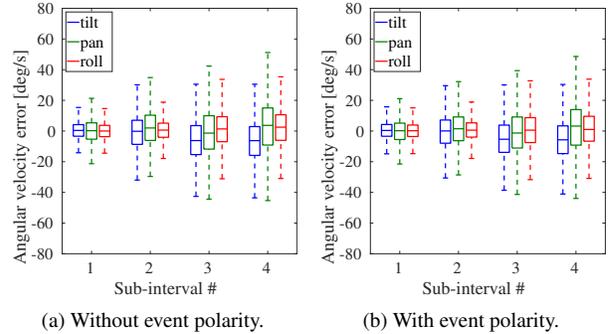(a) Without event polarity. (b) With event polarity.

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: $\approx 18\,°/s$ RMS (see Table 2), i.e., $2.7\,\%$ with respect to $|\omega_{max}| = 670\,°/s$.

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 $\approx 1.1\,$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 $\boldsymbol{\theta} \equiv \mathbf{v}$: $\mathbf{x}'_k = \mathbf{x}_k - (t_k - t_{ref})\mathbf{v}$. 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.
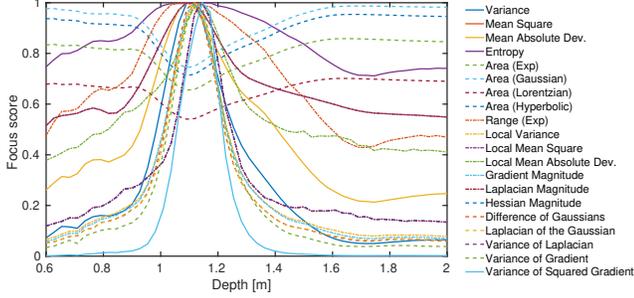
Figure 6: Depth estimation for a patch. Plots of several focus losses as a function of depth. Each focus loss curve is normalized (e.g., by its maximum), to stay in $[0, 1]$.

(a) Patch and three candidate flow vectors

(b) Variance (4).

(c) Gradient Magnitude (5).

(d) Laplacian Magnitude (6).

(e) Variance of Gradient Magnitude.
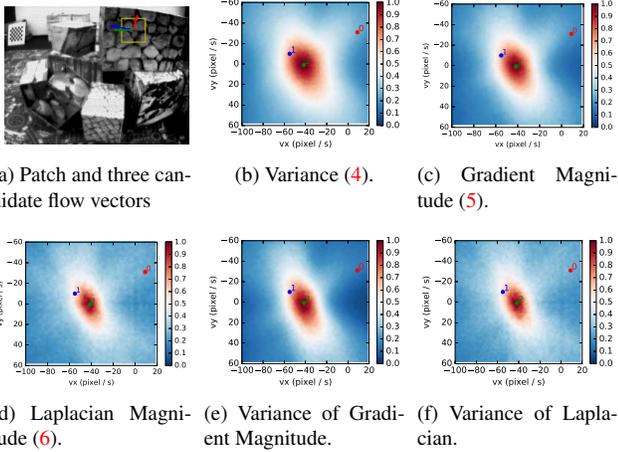
(f) Variance of Laplacian.

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
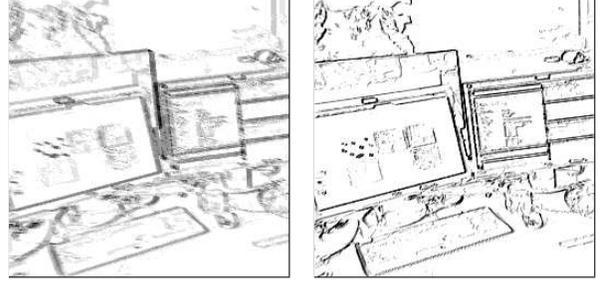

Figure 8: IWE before (Left) and after (Right) motion compensation by a neural network similar to [39], trained using the gradient magnitude loss (5).

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.

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

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

A. Notation

A.1. L^p norm of a vector-valued function

We define the L^p norm of a vector-valued function $\mathbf{f}$: $\Omega \subset \mathbb{R}^d \rightarrow \mathbb{R}^n$ with components $\{f_i\}_{i=1}^n$ (i.e., $\mathbf{f} = (f_1, \ldots, f_n)^\top$) by

$$\|\mathbf{f}\|_{L^p(\Omega)} \doteq \left(\int_\Omega \|\mathbf{f}(\mathbf{x})\|^p d\mathbf{x} \right)^{1/p}, \qquad (19)$$

where $\|\mathbf{f}(\mathbf{x})\| = (\sum_{i=1}^n |f_i(\mathbf{x})|^p)^{1/p}$ is the p-norm in $\mathbb{R}^n$. With this convention,

$$\|\mathbf{f}\|_{L^p(\Omega)}^p = \sum_{i=1}^n \int_\Omega |f_i(\mathbf{x})|^p d\mathbf{x} = \sum_{i=1}^n \|f_i\|_{L^p(\Omega)}^p, \quad (20)$$

that is, the p-th power of the norm of $\mathbf{f}$ is the sum of the p-th power of the norms of its components.

The two most common cases are $p = 1$ and $p = 2$, which, for the gradient of an image, $\nabla I = (I_x, I_y)^\top$, yield simple expressions:

$$\|\nabla I\|_{L^1(\Omega)} = \int_\Omega \left(|I_x(\mathbf{x})| + |I_y(\mathbf{x})| \right) d\mathbf{x} \qquad (21)$$

and

$$\|\nabla I\|_{L^2(\Omega)}^2 = \int_\Omega \left(I_x^2(\mathbf{x}) + I_y^2(\mathbf{x}) \right) d\mathbf{x}. \qquad (22)$$

A.2. Hessian Matrix

The Hessian matrix of a function, such as the IWE (used in (7)), is denoted by

$$\mathrm{Hess}(I) = \begin{pmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{pmatrix}, \qquad (23)$$

where the subscripts indicate derivatives.

The trace of the Hessian matrix is the Laplacian, which is used to define loss function (6).

B. Area of the Image of Warped Events

To measure the "thickness" of the edges of the IWE (e.g., Fig. 3a), one could count the number of pixels with count of warped events above a threshold, e.g., one event. However, this is brittle since it depends on this arbitrary threshold. We propose to define the above-mentioned edge thickness or "area" of an edge-like image like the IWE (3) in a more sensible way as a weighted sum of the interior of the level sets of the image, as we show next.

B.1. Definition of the Area of an Image

Using a Gaussian function (kernel) as a smooth approximation to the Dirac delta, $\delta(\mathbf{x} - \boldsymbol{\mu}) \approx \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \sigma^2 \mathtt{Id})$, the image of warped events (3) has, strictly speaking, an unbounded support (area of pixels with non-zero value). To have a meaningful support measure, we instead count the number of pixels with value greater than[1] λ,

$$\mathrm{supp}(I; \lambda) \doteq \int_\Omega H(I(\mathbf{x}) > \lambda) \, d\mathbf{x}, \qquad (24)$$

where $H(\cdot)$ is the Heaviside function. Fig. 9 shows several examples of it. This figure also illustrates the principle of area minimization, for a 1-D signal (3) with just two warped events. As observed, the area or thickness of I is minimized if the events are warped to the same location ($\Delta \mathbf{x}' = \mathbf{0} \iff \mathbf{x}_i' = \mathbf{x}_j'$), which is the desired event alignment condition of corresponding events.

To have a support metric that does not depend on the particular value of the threshold λ used (for fixed kernel width σ), we sum (24) over all threshold values,

$$\mathrm{supp}(I) \doteq \int_0^\infty \rho(\lambda) \, \mathrm{supp}(I; \lambda) \, d\lambda, \qquad (25)$$

using a decreasing weighting function ρ, such as $e^{-\lambda}$, thus emphasizing the areas corresponding to $\lambda \approx 0$ over those associated to $\lambda \gg 0$. In this way, an algorithm minimizing (25) will focus its attention on decreasing the area contribution of small thresholds, which are more important since the areas of larger thresholds are smaller due to the λ-support sets $\{\mathbf{x} \in \mathbb{R}^2 \mid I(\mathbf{x}) > \lambda\}$ forming a family of nested subsets.

---

[1]We assume that $I(\mathbf{x}) \geq 0$ either because event polarity is not used ($b_k = 1$ in (3)) or because the support here defined is applied to images of positive and negative events separately, and the results are added.
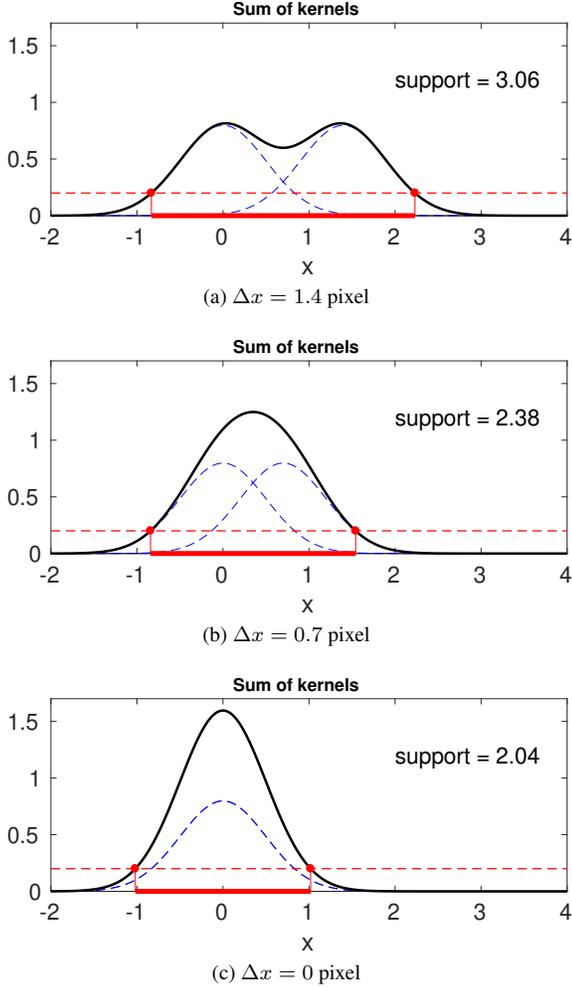
(a) $\Delta x = 1.4$ pixel

(b) $\Delta x = 0.7$ pixel

(c) $\Delta x = 0$ pixel

Figure 9: Illustration of the "area" or support of a 1-D image (3) with two events. The sum of the Gaussian kernels centered on each warped event (in blue) produces $I(x)$ (in black), whose support (i.e., the set $\{x \in \mathbb{R} \,|\, I(x) > \lambda = 0.2\}$) is displayed in solid red. The more aligned the events (smaller $\Delta x = x'_2 - x'_1$), the smaller the support of $I(x)$.

Notice that it is not possible to use $\rho = \text{const}$ since this leads to $\text{supp}(I) = N_e$, which does not depend on the motion parameters $\boldsymbol{\theta}$ we wish to optimize for. Using weighting functions with unit area (i.e., $\int_0^\infty \rho(\lambda)d\lambda = 1$) allows us to interpret (25) as a convex combination of supports (24), thus setting the correct scale so that (25) has the same units as (24).
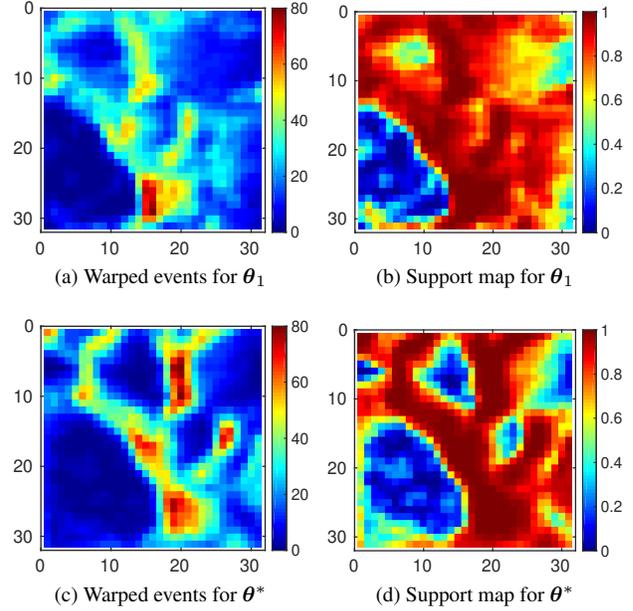

(a) Warped events for $\boldsymbol{\theta}_1$

(b) Support map for $\boldsymbol{\theta}_1$

(c) Warped events for $\boldsymbol{\theta}^*$

(d) Support map for $\boldsymbol{\theta}^*$

Figure 10: Illustration of the area (26) of a patch/image of warped events. Warped events $I(\mathbf{x}; \boldsymbol{\theta})$ (left column) and corresponding per-pixel support maps $F(I(\mathbf{x}; \boldsymbol{\theta})/10) - F(0)$ in (26) (right column), for two different motion parameters $\boldsymbol{\theta}$ (top: $\boldsymbol{\theta}_1 \equiv$ suboptimal; bottom: $\boldsymbol{\theta}_2 \equiv \boldsymbol{\theta}^*$ optimal). Support scores (26): $\text{supp}(I) = 0.72\,|\Omega|$ (top) and $\text{supp}(I) = 0.65\,|\Omega|$ (bottom), with $|\Omega| = 31^2 = 961$ pixels. The bottom patch has a smaller area (i.e., thinner edges) than the top patch, thus showing a better event alignment.

B.2. Simplification of the Area of an Image

Substituting (24) in (25) and swapping the order of integration gives

$$
\begin{aligned}
\text{supp}(I) &= \int_\Omega \int_0^\infty \rho(\lambda)\, H(I(\mathbf{x}) > \lambda)\, d\lambda\, d\mathbf{x} \\
&= \int_\Omega \int_0^{I(\mathbf{x})} \rho(\lambda)\, d\lambda\, d\mathbf{x} \\
&= \int_\Omega [F(\lambda)]_0^{I(\mathbf{x})}\, d\mathbf{x} \\
&= \int_\Omega \big(F(I(\mathbf{x})) - F(0)\big)\, d\mathbf{x}, \quad (26)
\end{aligned}
$$

where $F(\lambda) \doteq \int \rho(\lambda)d\lambda$ is a primitive of ρ, and $F(0)$ is constant. This is an advantageous expression compared to (25), since it states that $\text{supp}(I)$ can be computed using the values of $I(\mathbf{x})$ directly, without having to compute (24) for every threshold λ and then sum up the results. By using a continuous image formulation, we have analytically integrated the partial sums (24).

Fig. 10 illustrates (26). It shows the warped events

$I(\mathbf{x}; \boldsymbol{\theta})$ on a 31×31 image patch for two different parameters $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ (depth values, in this example [35]). It also shows the corresponding integrands of (26), or "per-pixel support maps" $F(I(\mathbf{x}; \boldsymbol{\theta})/\lambda_0) - F(0) \equiv 1 - \exp(-(I(\mathbf{x}; \boldsymbol{\theta})/\lambda_0))$, with $\lambda_0 = 10$ warped events. Pixels with $I(\mathbf{x}) \gtrsim \lambda_0$ events contribute more to the support (26) than pixels with $I(\mathbf{x}) \lesssim \lambda_0$ events, as shown in the support maps (right column of Fig. 10), which are color-coded from blue (low contribution) to red (high contribution).

Basically, the red regions of the support maps approximately indicate the area of the IWE, whereas the blue regions indicate the pixels where few warped events accumulate and therefore do not effectively contribute to the area of the IWE. Clearly, the bottom patch has a smaller area (i.e., thinner edges) than the top patch, as indicated by the smaller area of the red regions. The image area (26) is used to define the focus loss function (8).

C. Loss Function: Image Entropy

As anticipated in (12), event alignment may be achieved by maximizing the entropy of the IWE, where

$$H\left(p(z)\right) \doteq - \int_{-\infty}^{\infty} p(z) \log p(z) dz \qquad (27)$$

is Shannon's (differential) entropy for a continuous random variable whose density function (PDF) is $p(z)$.

The PDF of an image is approximated by its histogram, normalized to have unitary area. In a continuous formulation, this is written as (see [56])

$$p_I(z) \doteq \frac{1}{|\Omega|} \int_{\Omega} \delta\left(z - I(\mathbf{x}; \boldsymbol{\theta})\right) d\mathbf{x}, \qquad (28)$$

using the Dirac delta. This equation intuitively says that $p_I(z)$ is computed as a ratio of areas: the "number of pixels" of the IWE with value z, divided by the total "number of pixels", $|\Omega| = \int_{\Omega} d\mathbf{x} = N_p$.

Substituting (28) into (27), the entropy of the IWE becomes

$$
\begin{aligned}
H\left(p_I(z)\right) &= - \int_{-\infty}^{\infty} p_I(z) \log p_I(z) dz \\
&= - \int_{-\infty}^{\infty} \frac{1}{|\Omega|} \int_{\Omega} \delta(z - I(\mathbf{x})) d\mathbf{x} \log p_I(z) dz \\
&= - \frac{1}{|\Omega|} \int_{\Omega} \left(\int_{-\infty}^{\infty} \delta(z - I(\mathbf{x})) \log p_I(z) dz \right) d\mathbf{x} \\
&= - \frac{1}{|\Omega|} \int_{\Omega} \log p_I\left(I(\mathbf{x})\right) d\mathbf{x}. \qquad (29)
\end{aligned}
$$

Observe that the entropy is maximized by favoring large values of $\log(1/p_I(I(\mathbf{x})))$ over smaller ones. Since $\log$ is concave, it means that large values of $1/p_I(I(\mathbf{x}))$ are favored, i.e., small values of $p_I(I(\mathbf{x}))$ are favored. For a PDF

that is concentrated around $I = 0$ (large $p_I(0)$, as shown on the last column of Fig. 2), favoring small density values implies that they must be achieved away from $I = 0$, i.e., for large $|I|$ values (which are caused by the aggregation of aligned events). Thus, maximizing the entropy increases the range of $I(\mathbf{x})$, producing a higher contrast image.

D. Loss Function: Image Range

We measure the image range by means of the support of its PDF (28),

$$\mathrm{supp}(p_I) \doteq \int_0^{\infty} \rho(\lambda) \, \mathrm{supp}(p_I(z); \lambda) \, d\lambda, \qquad (30)$$

where the weight function $\rho(\lambda) \geq 0$ emphasizes the contributions of small $|\lambda|$ over those of large $|\lambda|$, according to the typical shape of the PDF of the event image (concentrated around $\lambda = 0$).

Mimicking the steps in Section B, Eq. (30) can be rewritten as

$$
\begin{aligned}
\mathrm{supp}(p_I) &= \int_{-\infty}^{\infty} \int_0^{\infty} \rho(\lambda) \, H(p_I(z) > \lambda) \, d\lambda \, dz \\
&= \int_{-\infty}^{\infty} \int_0^{p_I(z)} \rho(\lambda) \, d\lambda \, dz \\
&= \int_{-\infty}^{\infty} [F(\lambda)]_0^{p_I(z)} \, dz \\
&= \int_{-\infty}^{\infty} \left(F(p_I(z)) - F(0) \right) dz, \qquad (31)
\end{aligned}
$$

where $F(\lambda)$ is a primitive of $\rho(\lambda)$, and $F(0)$ is constant.

The motion parameters are found by maximizing (31), i.e., (13). The same weighting functions and primitives as for the image area (Section 3.3) may be used (with even symmetry if event polarity is used in the IWE (3)). This approach is inspired by the maximization of the entropy of the PDF of the image of warped events, as explained in Section C.

E. Loss Function: Spatial Autocorrelation

E.1. Moran's *I* Index

Moran's *I* index (17) (or "serial correlation coefficient" [47]) is a measure of spatial autocorrelation, i.e., it measures how similar is one object with respect to its neighbors. It is a concept that applies to variables whose values are known in unstructured grids (spatial units), in general (see Fig. 11).

If the variable of interest z consists of the intensity values of an image, $z_i = I(\mathbf{x}_i)$, which is defined on a regular (pixel) lattice $\{\mathbf{x}_i\}$, and the weights w_{ij} are shift-invariant (they do not depend on the particular location of pixels i and j, only on their relative spatial arrangement) and symmetric
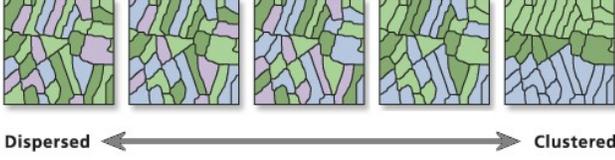
11

Figure 11: Illustration of spatial autocorrelation by Moran's index. A negative index indicates dispersion, whereas a positive index indicates clustering. The IWE differs from the figure above in the sense that its pixel values vary continuously with respect to the warping parameters, i.e., they are not fixed values that move around as in the figure. Image courtesy of ArcGIS.com https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/spatial-autocorrelation.htm

$w_{ij} = w_{ji}$, then it is possible to write Moran's I index using a convolution. In the formalism of continuous images $z(\mathbf{x})$ over a domain Ω, Moran's I index becomes

$$\mathrm{Moran}(z) \doteq \frac{1}{|\Omega|} \int_\Omega z^s(\mathbf{x}) \left(z^s(\mathbf{x}) * \tilde{w}(\mathbf{x}) \right) d\mathbf{x}, \quad (32)$$

where the standardized image $z^s(\mathbf{x}) \doteq (z(\mathbf{x}) - \bar{z})/\sigma_z$ is obtained by normalizing z with its mean $\bar{z}$ and variance σ_z^2 over Ω. The weights $\tilde{w}(\mathbf{x})$ should produce, in the convolution $z^s(\mathbf{x}) * \tilde{w}(\mathbf{x})$, a sum of the neighboring values of $z^s(\mathbf{x})$ (excluding the central value at $\mathbf{x}$). Thus, it is natural to consider the weights from a Gaussian kernel $G_\sigma(\mathbf{x})$ with a zero at the origin:

$$\tilde{w}(\mathbf{x}) = \frac{G_\sigma(\mathbf{x}) - G_\sigma(\mathbf{0})\delta(\mathbf{x})}{1 - G_\sigma(\mathbf{0})}. \quad (33)$$

The integrand of (32) is the local Moran's I index, and it is the element-wise product of the standardized variable z^s with a low-pass filtered version of itself. It is positive if both z^s and neighboring values $z^s(\mathbf{x}) * \tilde{w}(\mathbf{x})$ are higher or lower than the mean; and it is negative if the value and neighboring values are on opposite sides of the mean (one higher, the other lower). Increasing event alignment corresponds to favoring negative local Moran indices (dissimilar IWE pixels next to each other), and therefore, a negative (global) Moran's I index.

E.2. Geary's Contiguity Ratio

Geary's contiguity ratio is a generalization of Von Neumann's ratio [50] of the mean square successive difference (numerator of (18)) to the variance (denominator of (18)). Geary's contiguity ratio is non-negative, and its mean is 1 for random images. Values of C significantly lower than 1 demonstrate positive spatial autocorrelation (the variable of interest is regarded as *contiguous*), while values significantly higher than 1 illustrate negative autocorrelation.

In the formalism of continuous images, Geary's contiguity ratio can be written as

$$C(z) = \frac{1}{2} \frac{1}{|\Omega|} \int_\Omega c(\mathbf{x}) d\mathbf{x}, \quad (34)$$

with local score efficiently computed using convolutions:

$$c(\mathbf{x}) \doteq (z^s(\mathbf{x}))^2 + (z^s(\mathbf{x}))^2 * \tilde{w}(\mathbf{x}) - 2z^s(\mathbf{x}) \left(z^s * \tilde{w} \right)(\mathbf{x}). \quad (35)$$

Notice that the last term in (35) also appears in Moran's I index (32). Thus, Geary's C is inversely related to Moran's I, but they are not identical.

Homogeneous regions of an image $z(\mathbf{x})$ have a positive spatial autocorrelation, indicated by $c(\mathbf{x}) < 1$. The regions with large values of (35), i.e., negative spatial autocorrelation, are those corresponding to the edges of the objects (dissimilar intensity values on each side of the edge). Thus, Geary's local statistic (35) acts as an edge detector (see Figs. 13 or 15).

F. Loss Function: Aggregation of Local Statistics

Similarly to (14), aggregating other local statistics of the IWE also yield focus measures. For example, the aggregation of the local mean absolute deviation (ALMAD),

$$\mathrm{ALMAD}(I) \doteq \int_\Omega \mathrm{MAD}(\mathbf{x}; I) d\mathbf{x}, \quad (36)$$

with

$$\mathrm{MAD}(\mathbf{x}; I) \doteq \frac{1}{|B(\mathbf{x})|} \int_{B(\mathbf{x})} |I(\mathbf{u}; \boldsymbol{\theta}) - \mu(\mathbf{x}; I)| d\mathbf{u}, \quad (37)$$

is closely related to (14) since both aggregate local measures that are edge-detectors of I ((15) uses the the L^2 norm, whereas (37) uses the L^1 norm). Using a weighted neighborhood (e.g., Gaussian kernel G_σ), (37) can be efficiently approximated by the formula with two convolutions:

$$\mathrm{MAD}(\mathbf{x}; I) \approx \left| I(\mathbf{x}) - \left(I(\mathbf{x}) * G_\sigma(\mathbf{x}) \right) \right| * G_\sigma(\mathbf{x}), \quad (38)$$

where the inner convolution approximates the local mean, $\mu(\mathbf{x}; I) \approx I(\mathbf{x}) * G_\sigma(\mathbf{x})$, and the outer convolution averages the magnitude of the local, centered IWE (integrand of (37)) over the neighborhood around $\mathbf{x}$.

Omitting the local mean in (15) and (37) leads to local versions of the MS and the MAV, respectively. These operators, however, are not edge detectors; nevertheless, they also work as focus loss functions since the images on which they are applied, the IWEs, are edge-like images (the events are brightness changes, i.e., they are related to the temporal derivative of the brightness signal). Similarly to the (global) MAV, the local MAV does not provide enough information to estimate the warp parameters $\boldsymbol{\theta}$ if polarity is not used (as indicated in Table 2).

G. Plots of the Local Loss Maps

Most of the loss functions considered can be written as integrals over the IWE domain. Figs. 12, 13, 14 and 15 visualize the integrands (i.e., "local loss") of most focus loss functions, for two scenes: dynamic and boxes from the dataset [45]. Images are given in pairs (with a common caption below the images): local loss before optimization (without motion compensation, on the left), and after optimization of the corresponding focus loss function (motion-compensated, on the right). Each image pair shares the same color scale for proper visualization of how the local loss changes before and after optimization. For reference, since the local loss are transformations of the IWE, Figs. 12 and 14 also provide, on the top right, the IWE before and after optimization with one of the loss functions (the variance).

The local loss of area-based loss functions is the support map, as in Figs. 3b and 10. The focus loss given by the IWE range (13) is not expressed as an integral over the image domain, therefore, no image integrand is visualized in the above-mentioned figures. MAV and local MAV are not displayed either since they cannot be optimized with respect to the parameters.

Notice that all local loss maps are represented using the same color scheme, from blue (small) to large (yellow). For objective functions formulated as maximization problems (variance, gradient magnitude, etc.), visually good maps are those that are almost "blue" for IWEs with bad event-alignment parameters, and that clearly show "yellow" regions where events align (due to good parameters $\boldsymbol{\theta}$). For objective functions formulated as minimization problems (e.g., area-based, loss based on the mean timestamp per pixel), the situation is the opposite: good local loss maps become less yellowish and more blueish as event alignment improves due to good parameters $\boldsymbol{\theta}$.

H. Additional Experiments on Accuracy Evaluation

Tables 3 and 4 provide further quantitative evaluations of the considered focus loss functions. Results correspond to the boxes and poster sequences in [45], undergoing a rotational motion with velocities close to $1000\,°/\text{s}$. Looking at the RMS errors, these are small compared to the excursion of the signal. The RMS columns of these tables are summarized in Table 2.

I. Additional Plots of Focus Loss Functions in Optical Flow Space

Figs. 16 and 17 present experiments with events in a small space-time window (31×31 pixels and $\Delta t = 200\,\text{ms}$), yielding approximately 2000 events, from a sequence of the dataset [45]. The goal of these figures is to visualize the "shape" or "signature" of the focus loss functions (as heat maps, pseudo-colored from blue to red).

The top-left image shows the patch on the intensity frame (not used) corresponding to the space-time window of events, highlighted in yellow, and three candidate flow vectors $\{\boldsymbol{\theta}_i\}_{i=0}^2$ (marked with red, blue and green arrows, respectively). The ground truth flow is close to $\boldsymbol{\theta}_2 = (-40, 0)\text{pixel/s}$. The top row also shows the warped events (IWE patch) using the three flow vectors, without polarity in the IWE (Fig. 16) or with polarity (Fig. 17). The remaining rows show the focus loss functions in optical flow space, with $\pm 60\,\text{pixel/second}$ around $\boldsymbol{\theta}_2$. Some focus functions are designed to be maximized (and therefore should present a local maximum at $\boldsymbol{\theta}_2$), while others are designed to be minimized (and should present a local minimum at $\boldsymbol{\theta}_2$).

Without Polarity. Rows 2 and 3 of the figures show the variance, MS, MAD, MAV and their aggregation of local versions. They are all visualized in the range $[0, 1]$, by dividing by the maximum value of the focus function. Without using polarity (Fig. 16), the variance presents a nice peak at the correct optical flow $\boldsymbol{\theta}_2$, the peaks of the MS and MAD functions are not as pronounced, and the MAV function does not have the maximum at the ground truth location (we explained that, without polarity, the MAV cannot be used to estimate $\boldsymbol{\theta}$). The local versions (third row of Fig. 16) are slightly narrower than the global versions. The fourth row presents the four area-based focus losses (Section B), whose goal is to be minimized, and indeed, they present a local minimum at $\boldsymbol{\theta}_2$. There are not big differences in these four area-based losses. The fifth row shows more statistics-based losses. The range and Geary's C show a local maximum at the correct flow. Moran's index shows a local minimum, as expected, at the correct flow. The entropy, without using polarity, does not have a local maximum at the correct flow. Instead, using polarity (Fig. 17), it does have a local maximum at $\boldsymbol{\theta}_2$. The last two rows of Figs. 16 and 17 show focus loss functions based on the IWE derivatives and their variances (composite losses). They all present a clear peak at the correct depth (as the case of the variance and local variance); some of them are more narrow than others (all are visualized in the range $[0, 1]$, for ease of comparison). The gradient magnitude (based on Sobel operator), the DoG magnitude and the LoG magnitude seem to be the smoothest of these two rows.

With Polarity. Fig. 17 shows the results on the same experiment as Fig. 16, but using event polarity in the IWE. In a scene with approximately equal number of dark-to-bright and bright-to-dark transitions, the number of positive and negative events is approximately balanced, and so, the mean of the IWE is approximately zero. Thus, in this

Focus Loss Function	Without polarity						With polarity					
	ω_x	ω_y	ω_z	μ	σ	RMS	ω_x	ω_y	ω_z	μ	σ	RMS
Variance (4) [33,35]	15.69	19.53	20.34	-0.31	18.42	18.52	16.03	19.67	21.12	-0.64	18.78	18.94
Mean Square (9) [33,36]	16.05	20.03	23.70	-0.49	19.83	19.93	16.04	19.67	21.36	-0.57	18.86	19.02
Mean Absolute Deviation (10)	15.40	**18.99**	23.99	-0.49	19.40	19.46	15.93	19.15	23.67	-0.57	19.40	19.58
Mean Absolute Value (11)	-	-	-	-	-	-	15.95	19.16	24.22	-0.61	19.60	19.77
Entropy (12)	19.29	21.73	44.48	0.03	28.14	28.50	18.03	20.50	41.10	-0.20	26.19	26.54
Area (8) (Exp)	19.70	21.57	53.23	0.10	31.41	31.50	15.89	19.12	23.60	-0.27	19.39	19.54
Area (8) (Gaussian)	18.12	20.20	39.22	0.10	25.78	25.85	15.67	**18.77**	22.11	-0.33	18.69	18.85
Area (8) (Lorentzian)	20.56	21.69	55.04	0.42	32.30	32.43	15.48	19.36	28.11	-0.02	20.85	20.98
Area (8) (Hyperbolic)	18.71	20.67	47.99	0.07	29.05	29.13	15.69	18.90	22.86	-0.27	18.99	19.15
Range (13) (Exp)	18.68	22.72	44.59	0.39	28.16	28.66	17.29	19.38	49.50	-0.48	28.23	28.72
Local Variance (14)	16.17	19.55	18.90	-0.41	17.94	18.21	16.04	19.55	19.61	-0.21	18.15	18.40
Local Mean Square	18.37	21.42	34.65	0.59	24.31	24.81	16.68	20.21	22.67	-0.05	19.53	19.86
Local Mean Absolute Deviation	16.36	19.39	28.35	-0.35	21.10	21.37	15.78	19.06	21.37	-0.20	18.47	18.74
Local Mean Absolute Value	-	-	-	-	-	-	18.13	20.17	34.00	-0.42	23.61	24.10
Moran's Index (17)	17.87	20.74	34.24	-0.99	23.87	24.28	16.97	20.01	33.31	-0.08	23.18	23.43
Geary's Contiguity Ratio (18)	17.48	20.21	33.93	-0.58	23.50	23.87	16.00	19.42	23.07	0.04	19.25	19.50
Gradient Magnitude (5)	16.12	19.53	**17.84**	-0.71	**17.58**	**17.83**	15.90	19.46	18.93	-0.79	17.91	18.10
Laplacian Magnitude (6)	15.63	20.92	18.42	-0.01	18.09	18.32	14.52	19.93	**18.28**	**0.03**	**17.36**	**17.58**
Hessian Magnitude (7)	16.26	19.69	19.29	-0.22	18.14	18.41	15.89	19.43	18.46	-0.30	17.70	17.93
Difference of Gaussians (DoG)	14.79	20.22	27.54	0.66	20.50	20.85	14.99	20.52	22.25	0.43	18.92	19.25
Laplacian of the Gaussian (LoG)	**14.64**	20.45	26.00	0.50	20.09	20.36	**14.42**	20.09	18.80	0.47	17.49	17.77
Variance of Laplacian	16.30	19.70	18.78	-0.45	18.02	18.26	15.95	19.45	18.71	-0.32	17.81	18.01
Variance of Gradient	16.26	19.71	20.11	**-0.00**	18.49	18.69	16.10	19.75	21.39	0.31	18.84	19.08
Variance of Squared Gradient	16.50	20.02	19.64	-0.10	18.46	18.72	16.22	20.08	20.56	0.06	18.72	18.95
Mean Timestamp on Pixel [37]	42.48	39.40	166.81	0.54	82.87	82.89	-	-	-	-	-	-

Table 3: *Accuracy and Timing Comparison of Focus Loss Functions.* Angular velocity errors (in $\deg/\text{s}$) of the motion compensation method [35] (with or without polarity) with respect to motion-capture system. The six columns per case are the errors in each component of the angular velocity and their mean, standard deviation and RMS values. Processing $N_e = 30\,000$ events, warped onto an image of 240×180 pixels (DAVIS camera [52]). Sequence: `boxes_rotation` from the Event Camera Dataset [45]. Best value per column is in bold.

case, the MS is approximately equal to the variance, and the MAV approximates the MAD. This is noticeable in the second row of Fig. 17. A similar trend is observe in the local versions of the four above statistics, albeit the local variance and MAD present narrower peaks than the local MS and MAV, respectively. The area-based focus functions are computed by splitting the events according to polarity, computing the areas of the two resulting IWEs and adding their area values. The corresponding plots, in the fourth row of Fig. 17 are similar to those without polarity (Fig. 16, except for the vertical scale). The entropy and range improve if event polarity is used, basically because the PDF becomes double-sided and it allows us to distinguish positive and negative IWE edges/values. Moran's index and Geary's C ratio are not good focus losses if polarity is used, since they present brittle local minimum/maximum, respectively. The derivative-based losses present a clear peak at the correct flow, and slightly more pronounced than their counterparts in Fig. 16.

J. Additional Plots on Depth Estimation

Fig. 18 shows in more detail Fig. 6: depth estimation for a patch from a sequence of the dataset [45]. The sequence was recorded with a DAVIS camera [52], camera poses were recorded by a motion-capture system, and the camera was calibrated, so the only unknown is the scene structure (i.e., depth). Fig. 18b shows how the values of the focus functions vary with respect to the depth parameter $\boldsymbol{\theta} \equiv Z$, for the events corresponding to the patch highlighted in Fig. 18a. Remarkably, the focus curves have a smooth variation, with a clear extrema around the correct depth value. Fig. 18a shows the warped events (IWE) for a depth $Z = 1.11\,\text{m}$, close to the peaks of the focus curves. The IWE is pseudo-colored, from few event count (blue) to large event count (red). It shows that the events produce a sharp image at the patch location, whereas other parts of the image are "out of focus" since events do not align at that depth [15,35]. Clearly, some focus curves are narrower than others (Fig. 18b), showing better properties for determining the optimal depth location. We observe that the entropy and

Focus Loss Function	Without polarity						With polarity					
	ω_x	ω_y	ω_z	μ	σ	RMS	ω_x	ω_y	ω_z	μ	σ	RMS
Variance (4) [33, 35]	26.67	20.24	30.95	**0.45**	25.93	25.96	26.50	19.86	26.80	**-0.20**	24.36	24.39
Mean Square (9) [33, 36]	28.44	25.89	47.97	3.70	33.43	34.10	26.90	21.23	30.79	1.19	26.14	26.31
Mean Absolute Deviation (10)	27.25	22.16	42.66	1.25	30.43	30.70	27.11	21.73	40.03	1.71	29.40	29.62
Mean Absolute Value (11)	-	-	-	-	-	-	27.18	22.19	40.33	2.18	29.65	29.90
Entropy (12)	32.81	41.12	68.70	3.63	46.98	47.54	27.65	22.73	49.27	2.36	32.98	33.21
Area (8) (Exp)	33.12	22.50	73.75	-0.73	42.94	43.12	26.31	19.53	33.36	-0.42	26.36	26.40
Area (8) (Gaussian)	28.40	21.07	54.02	-0.59	34.36	34.50	26.15	19.20	30.70	-0.41	25.31	25.35
Area (8) (Lorentzian)	28.51	21.69	57.41	-0.92	35.71	35.86	26.30	19.63	33.77	-0.59	26.52	26.57
Area (8) (Hyperbolic)	27.39	20.65	50.78	-0.84	32.82	32.94	26.27	19.35	32.01	-0.41	25.84	25.88
Range (13) (Exp)	32.18	37.73	61.70	3.44	43.32	43.87	24.48	19.53	45.99	1.56	29.86	30.00
Local Variance (14)	26.95	20.49	28.87	1.11	25.34	25.44	26.48	20.32	25.66	0.88	24.06	24.15
Local Mean Square	28.47	25.58	47.79	3.79	33.28	33.95	26.74	21.25	31.41	1.26	26.30	26.47
Local Mean Absolute Deviation	47.61	53.19	84.85	4.55	61.67	61.89	26.42	20.03	29.41	0.95	25.22	25.29
Local Mean Absolute Value	-	-	-	-	-	-	27.19	22.34	41.57	2.22	30.10	30.37
Moran's Index (17)	28.32	21.57	47.31	1.39	32.32	32.40	27.70	20.99	44.18	1.32	30.90	30.96
Geary's Contiguity Ratio (18)	27.63	20.78	31.44	0.89	26.56	26.61	26.52	20.25	28.90	1.19	25.15	25.23
Gradient Magnitude (5)	26.61	20.50	**24.69**	0.61	**23.85**	**23.93**	26.34	20.01	24.40	**-0.20**	23.54	23.58
Laplacian Magnitude (6)	27.24	20.74	26.76	0.64	24.85	24.91	26.29	20.25	24.46	0.61	23.60	23.67
Hessian Magnitude (7)	27.20	22.40	26.81	0.95	25.39	25.47	26.36	20.24	24.61	0.80	23.66	23.74
Difference of Gaussians (DoG)	**24.51**	18.01	30.97	1.45	24.41	24.50	24.29	17.85	24.30	0.75	**22.10**	**22.15**
Laplacian of the Gaussian (LoG)	24.66	**17.96**	32.85	1.27	25.10	25.15	**24.20**	**17.78**	30.04	1.30	23.94	24.01
Variance of Laplacian	26.85	27.84	25.10	0.77	26.52	26.59	26.26	20.23	**24.38**	0.53	23.56	23.62
Variance of Gradient	26.85	27.84	25.10	0.77	26.52	26.60	26.63	20.29	25.44	0.96	24.12	24.22
Variance of Squared Gradient	26.80	21.14	30.36	1.29	25.97	26.10	26.72	20.94	25.63	1.13	24.32	24.43
Mean Timestamp on Pixel [37]	64.94	87.19	211.4	0.94	121.1	121.2	-	-	-	-	-	-

Table 4: *Accuracy Comparison of Focus Loss Functions* on the `poster_rotation` sequence from dataset [45]. Angular velocity errors (in $\deg/s$) of the motion compensation method [35] (with or without polarity) with respect to motion-capture system. The six columns per case are the errors in each component of the angular velocity and their mean, standard deviation and RMS values. Processing $N_e = 30\,000$ events, warped onto an image of 240×180 pixels (DAVIS camera [52]). On each column, the best value is highlighted in bold.

the range curves have wide peaks, and therefore do not determine depth very precisely.

Semi-dense 3D Reconstruction. Fig. 19 shows depth estimation for every pixel of a reference view along the trajectory of the event camera. For every pixel, we compute focus curves, as those in Fig. 18b, and select the depth at the peak. To capture fine spatial details, the focus functions are computed on patches of 3×3 pixels in the reference view, weighted by a Gaussian kernel to emphasize the contribution of the center pixel. We also record the value of the focus function at the peak for every pixel of the reference view. These values are displayed as a "focus confidence map" in Fig. 19. For better visualization, the focus values are represented in negative form, from bright (low focus value) to dark (high focus value). The confidence map is used to select the pixels in the reference view with largest focus, i.e., the pixels for which depth is most reliably estimated. The above selection yields a semi-dense depth map, which is displayed color coded, overlaid on the intensity frame from the DAVIS camera [52] at the reference view.

We used adaptive thresholding [46, p.780] on the focus confidence map, and a median filter to remove spike noise from the depth map. As it is seen, depth is most reliably estimated at strong brightness edges of the scene. Finally, the depth map is also visualized as a point cloud, color-coded according to depth (Fig. 19).

The figure compares some representative focus functions. In general, we obtain good depth 3D reconstructions with the methods tested. Some methods produce slightly noisier 3D reconstructions than others, and some recover more edges than others. This is due to both the shape of the focus confidence maps and the adaptive thresholding parameters. We observe that focus functions as simple as the local mean square (MS) or the local mean absolute deviation (MAD) produce good results. These semi-dense 3D reconstruction methods may be used as the mapping module of an event-based visual odometry system, such as [23], to enable camera pose estimation from the 3D reconstructed scene.
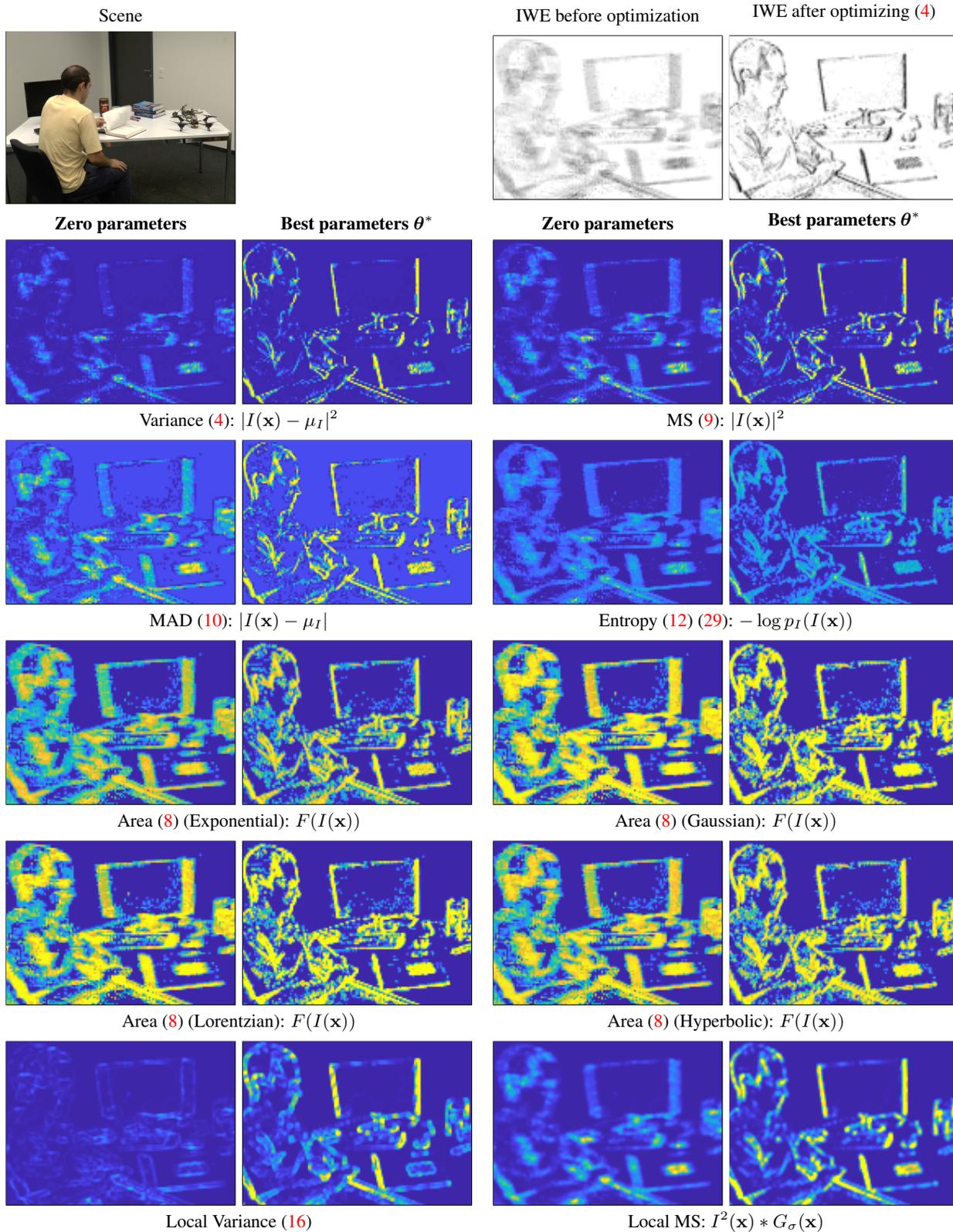
Scene

IWE before optimization

IWE after optimizing (4)

Zero parameters

Best parameters θ^*

Zero parameters

Best parameters θ^*

Variance (4): $|I(\mathbf{x}) - \mu_I|^2$

MS (9): $|I(\mathbf{x})|^2$

MAD (10): $|I(\mathbf{x}) - \mu_I|$

Entropy (12) (29): $-\log p_I(I(\mathbf{x}))$

Area (8) (Exponential): $F(I(\mathbf{x}))$

Area (8) (Gaussian): $F(I(\mathbf{x}))$

Area (8) (Lorentzian): $F(I(\mathbf{x}))$

Area (8) (Hyperbolic): $F(I(\mathbf{x}))$

Local Variance (16)

Local MS: $I^2(\mathbf{x}) * G_\sigma(\mathbf{x})$

Figure 12: Visualization of the local loss (i.e., integrands of the Focus Loss Functions), pseudo-colored from blue (small) to yellow (large). Same scene as in the top row of Fig. 2 (i.e., without using polarity). Images are given in pairs: local loss before optimization (no motion compensation, Left), and after optimization of the corresponding focus loss function (motion-compensated, Right). The local loss of area-based loss functions is the "support map", as in Figs. 3b and 10.
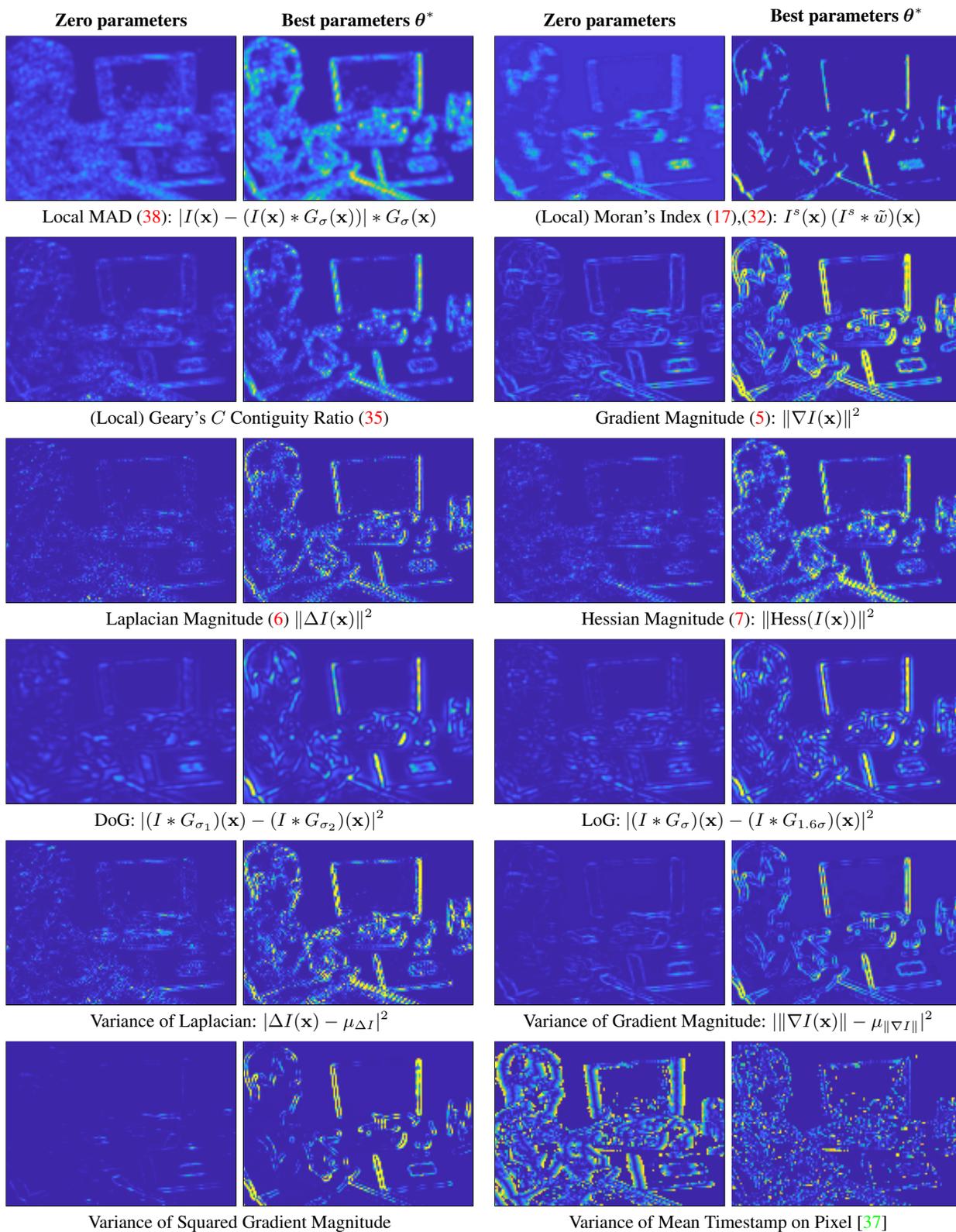
16

Zero parameters — **Best parameters θ^*** — **Zero parameters** — **Best parameters θ^***

Local MAD (38): $|I(\mathbf{x}) - (I(\mathbf{x}) * G_\sigma(\mathbf{x}))| * G_\sigma(\mathbf{x})$

(Local) Moran's Index (17),(32): $I^s(\mathbf{x})\,(I^s * \tilde{w})(\mathbf{x})$

(Local) Geary's C Contiguity Ratio (35)

Gradient Magnitude (5): $\|\nabla I(\mathbf{x})\|^2$

Laplacian Magnitude (6) $\|\Delta I(\mathbf{x})\|^2$

Hessian Magnitude (7): $\|\mathrm{Hess}(I(\mathbf{x}))\|^2$

DoG: $|(I * G_{\sigma_1})(\mathbf{x}) - (I * G_{\sigma_2})(\mathbf{x})|^2$

LoG: $|(I * G_\sigma)(\mathbf{x}) - (I * G_{1.6\sigma})(\mathbf{x})|^2$

Variance of Laplacian: $|\Delta I(\mathbf{x}) - \mu_{\Delta I}|^2$

Variance of Gradient Magnitude: $|\|\nabla I(\mathbf{x})\| - \mu_{\|\nabla I\|}|^2$

Variance of Squared Gradient Magnitude

Variance of Mean Timestamp on Pixel [37]

Figure 13: Visualization of local scores (i.e., integrands) of the Focus Loss Functions (continuation).

17

Scene

IWE before optimization   IWE after optimizing (4)

| Zero parameters | Best parameters θ^* | Zero parameters | Best parameters θ^* |

Variance (4): $|I(\mathbf{x}) - \mu_I|^2$ MS (9): $|I(\mathbf{x})|^2$

MAD (10): $|I(\mathbf{x}) - \mu_I|$ Entropy (12) (29): $-\log p_I(I(\mathbf{x}))$

Area (8) (Exponential): $F(I(\mathbf{x}))$ Area (8) (Gaussian): $F(I(\mathbf{x}))$

Area (8) (Lorentzian): $F(I(\mathbf{x}))$ Area (8) (Hyperbolic): $F(I(\mathbf{x}))$

Local Variance (16)   Local MS: $I^2(\mathbf{x}) * G_\sigma(\mathbf{x})$
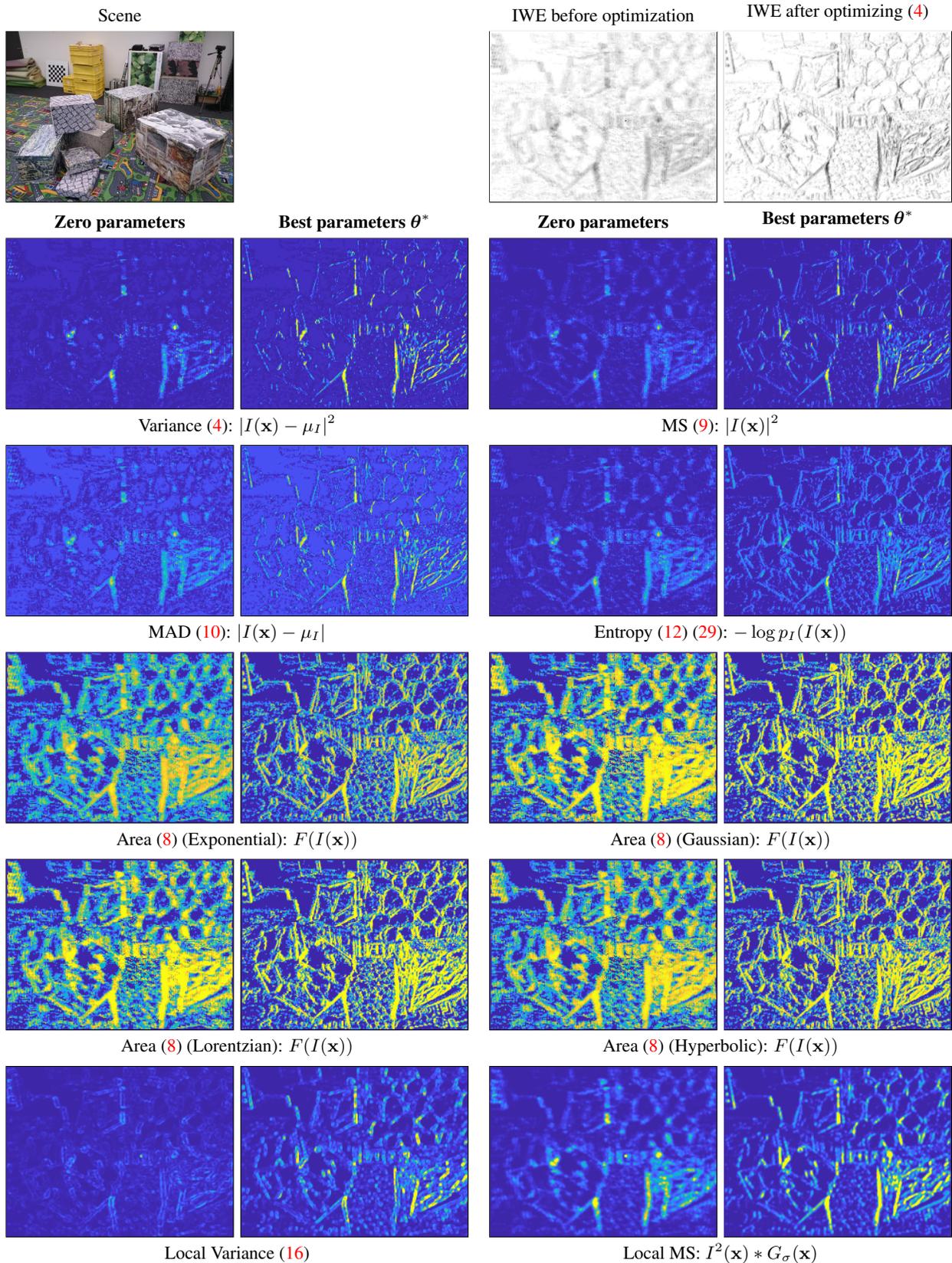
Figure 14: Visualization of the local loss (i.e., integrands of the Focus Loss Functions). Scene `boxes`, IWE not without using polarity. Same notation as Fig. 12.

Local MAD (38): $|I(\mathbf{x}) - (I(\mathbf{x}) * G_\sigma(\mathbf{x}))| * G_\sigma(\mathbf{x})$

(Local) Moran's Index (17), (32): $I^s(\mathbf{x})\,(I^s * \tilde{w})(\mathbf{x})$

(Local) Geary's C Contiguity Ratio (35)

Gradient Magnitude (5): $\|\nabla I(\mathbf{x})\|^2$

Laplacian Magnitude (6) $\|\Delta I(\mathbf{x})\|^2$

Hessian Magnitude (7): $\|\mathrm{Hess}(I(\mathbf{x}))\|^2$

DoG: $|(I * G_{\sigma_1})(\mathbf{x}) - (I * G_{\sigma_2})(\mathbf{x})|^2$

LoG: $|(I * G_\sigma)(\mathbf{x}) - (I * G_{1.6\sigma})(\mathbf{x})|^2$

Variance of Laplacian: $|\Delta I(\mathbf{x}) - \mu_{\Delta I}|^2$

Variance of Gradient Magnitude: $|\|\nabla I(\mathbf{x})\| - \mu_{\|\nabla I\|}|^2$

Variance of Squared Gradient Magnitude

Variance of Mean Timestamp on Pixel [37]
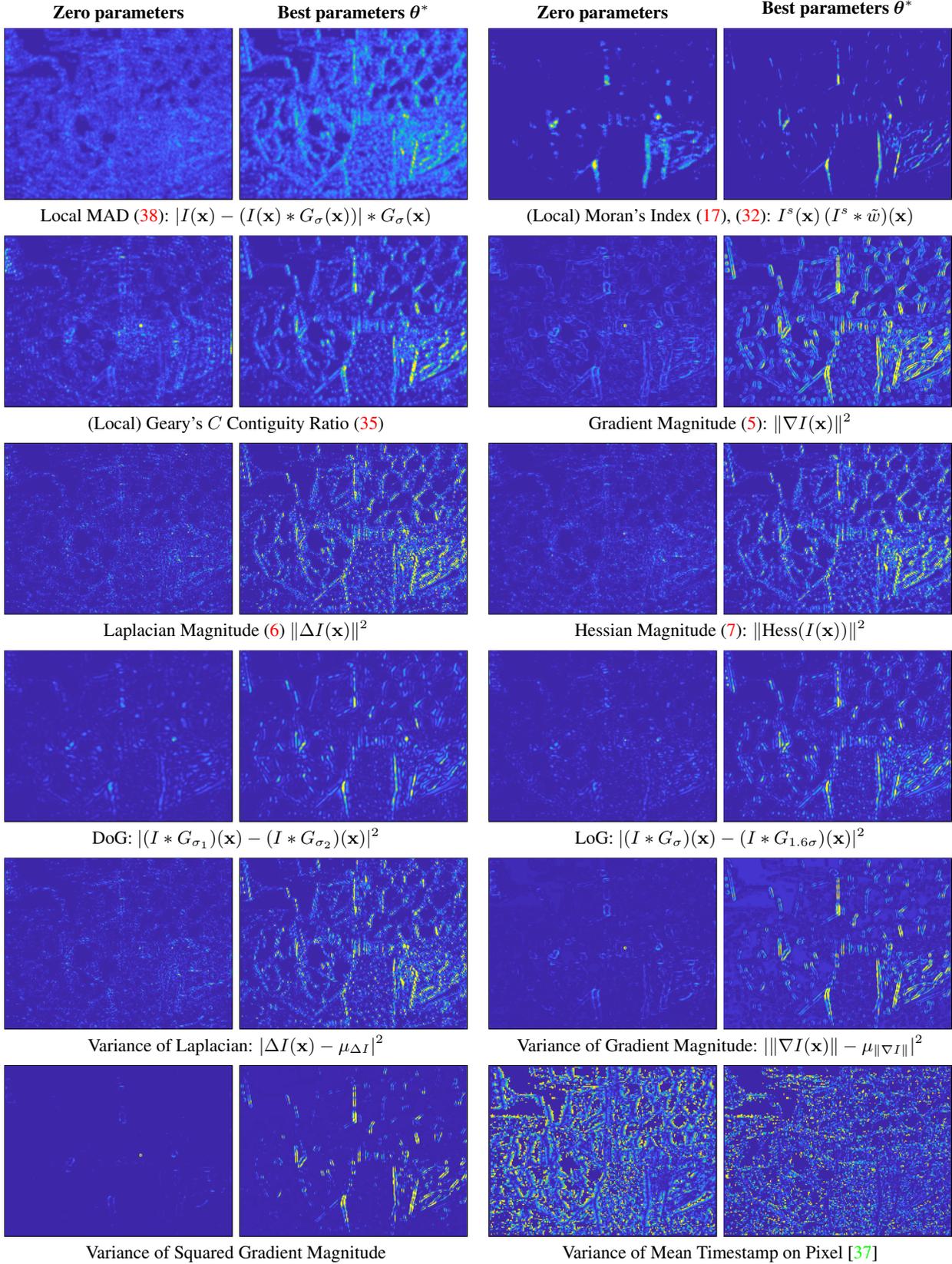
Figure 15: Visualization of local scores (i.e., integrands) of the Focus Loss Functions (continuation).
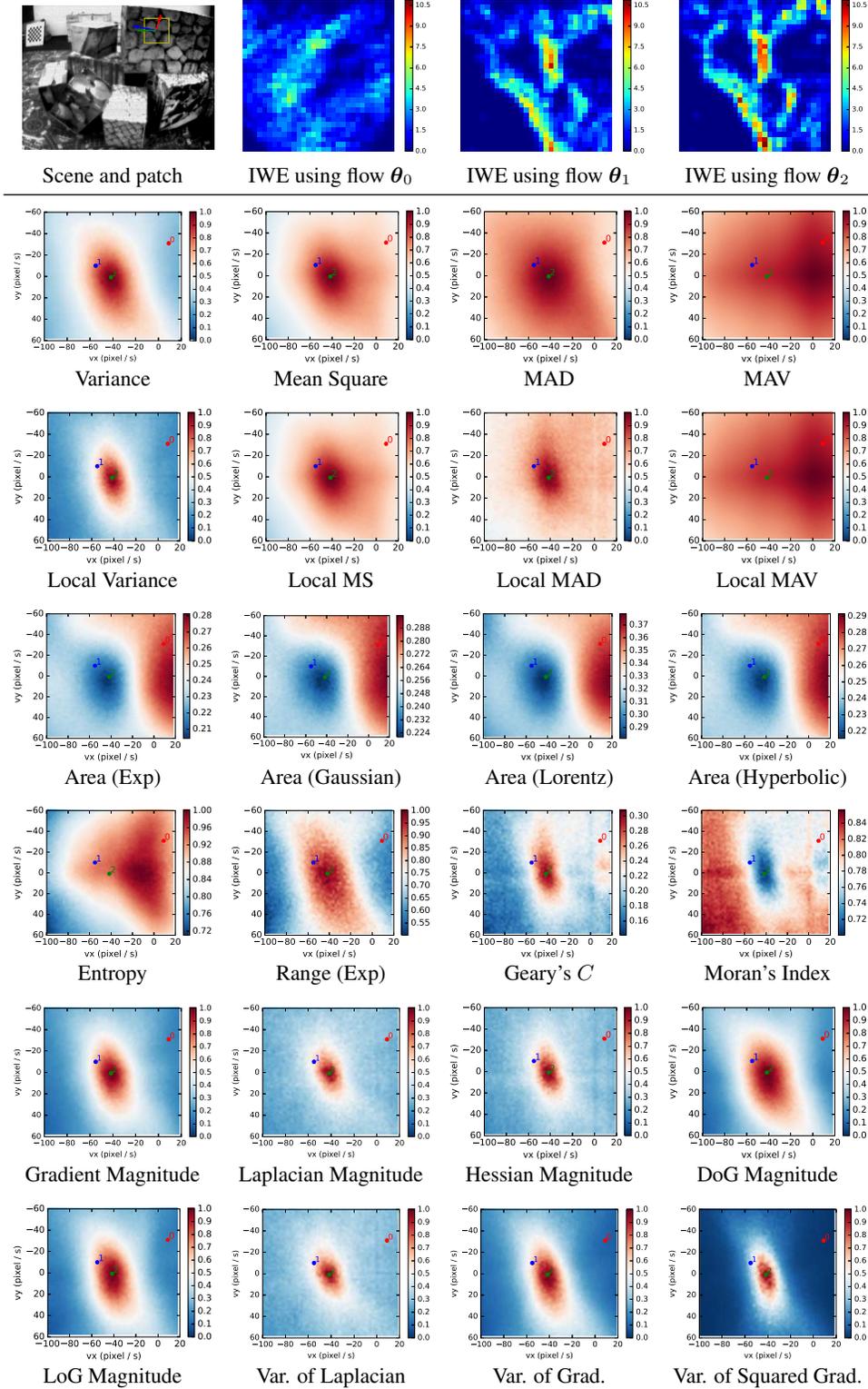
Figure 16: Visualization of the Focus Loss Functions (as heat maps, pseudo-colored from blue to red). The top-left image shows a selected patch (highlighted in yellow) and three candidate flow vectors $\{\boldsymbol{\theta}_i\}_{i=0}^2$. The ground truth flow is close to $\boldsymbol{\theta}_2 = (-40, 0)$pixel/s. The top row also shows the warped events (IWE patch) using the three flow vectors, without using polarity in the IWE. The remaining rows show the focus loss functions in optical flow (i.e., image velocity) space.
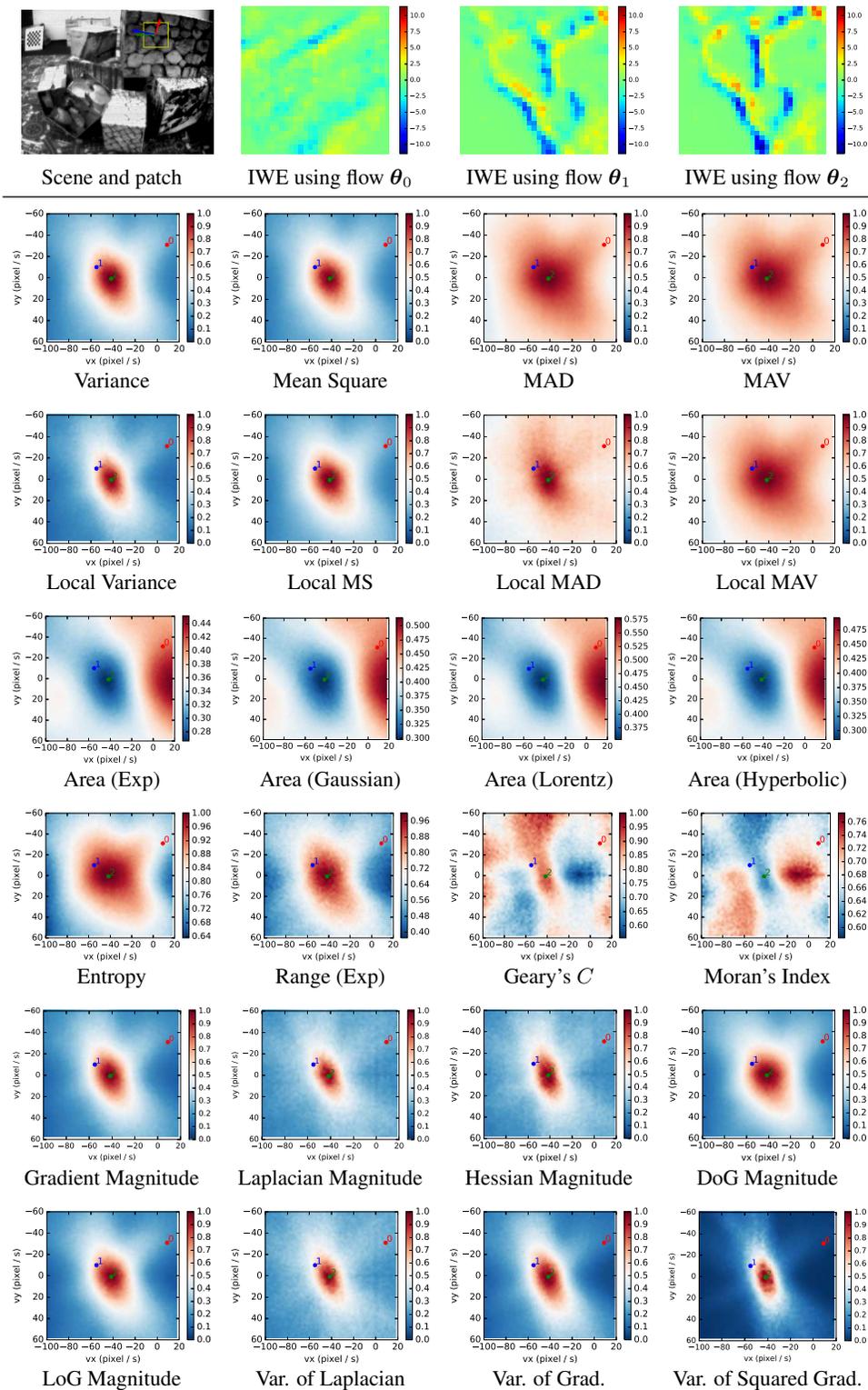
Figure 17: Visualization of the Focus Loss Functions (as heat maps, pseudo-colored from blue to red). Same notation as in Fig. 16, but using polarity in the IWE.

21
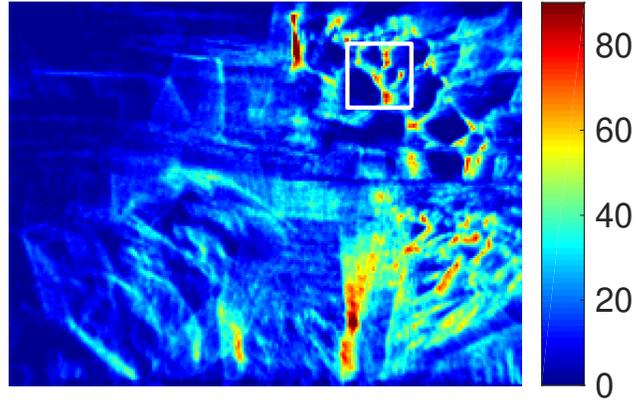
(a) Image of Warped Events IWE (i.e., slice of the Disparity Space Image (DSI) [15]) corresponding to depth $\theta \equiv Z = 1.11\,\mathrm{m}$. Color scale represents the event count, from blue (few events) to red (many events).
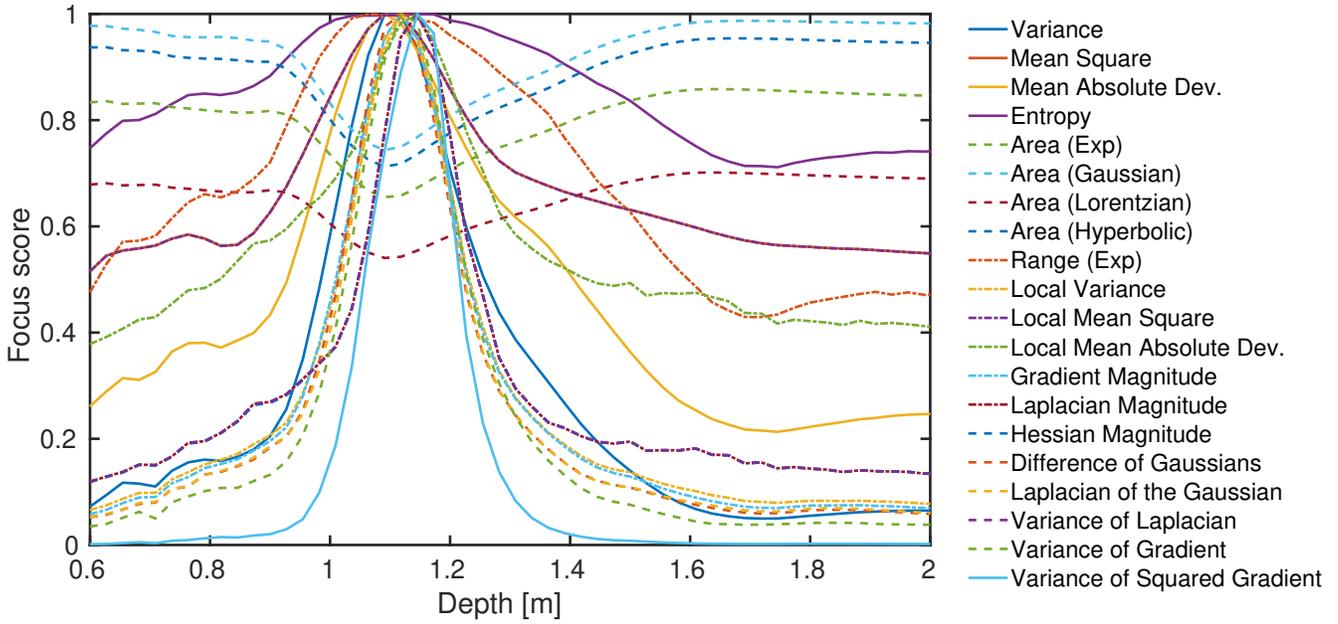
(b) Focus functions vs depth for the events corresponding to the patch highlighted in Fig. 18a.

Figure 18: *Depth Estimation*. Focus functions in Fig. 18b are shown normalized to the range $[0, 1]$ for easier visualization.
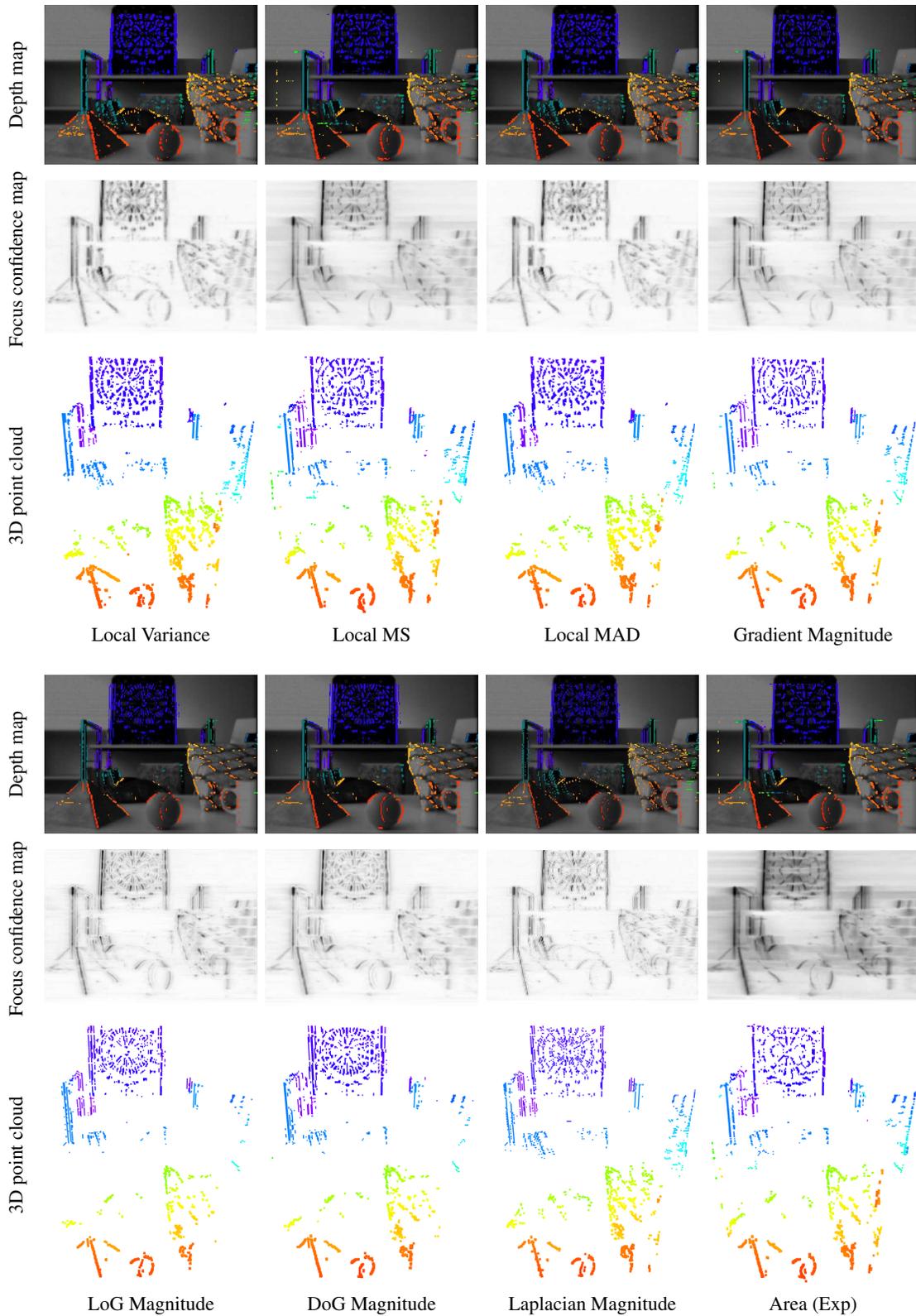
Figure 19: *3-D reconstruction* of a scene using several focus loss functions. Semi-dense depth maps (overlaid on grayscale frames) and point clouds are pseudo-colored according to depth, from red (close) to blue (far), in the range $0.45\,\mathrm{m}$ to $2.4\,\mathrm{m}$. Sequence slider_depth from [45], with $N_e = 1\,000\,000$ events over a time span of $2.93\,\mathrm{s}$ and a camera baseline of $85\,\mathrm{cm}$.

K. Analytical Derivatives of Focus Loss Functions

In this section, we provide the analytical derivatives of some of the focus loss functions used. A significant advantage of the proposed focus loss functions is that they are defined in terms of well-known analytical operations, and so, we can use powerful tools from Calculus to compute and simplify their derivatives. This becomes useful in the optimization framework of Fig. 1, both, for speed-up and increased accuracy over numerical derivatives.

Let the vector of parameters defining the space-time warping of events be $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_M)^\top \in \mathbb{R}^M$.

Derivative of the IWE. The derivative of the IWE (3) with respect to the warp parameters $\boldsymbol{\theta}$ is, replacing the Dirac delta with an approximation, such as a Gaussian, $\delta(\mathbf{x}) \approx \mathcal{N}(\mathbf{x}; \mathbf{0}, \epsilon^2 \mathtt{Id})$, and using the chain rule,

$$\frac{\partial I}{\partial \boldsymbol{\theta}} = -\sum_{k=1}^{N_e} b_k \, \nabla \mathcal{N}(\mathbf{x} - \mathbf{x}'_k(\boldsymbol{\theta}); \mathbf{0}, \epsilon^2 \mathtt{Id}) \, \frac{\partial \mathbf{x}'_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \tag{39}$$

where $\nabla \mathcal{N}$ is the gradient of the 2D Gaussian PDF, and the derivative of the warp is a purely geometric term: $\frac{\partial \mathbf{x}'_k(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{W}'(\mathbf{x}'_k, t_k; \boldsymbol{\theta})$.

Loss Function: Mean Square (MS). The derivative of the MS, (9), is, by the chain rule,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{MS}(I) = \frac{1}{|\Omega|} \int_\Omega 2I(\mathbf{x}) \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} d\mathbf{x}. \tag{40}$$

Loss Function: Variance. The derivative of the variance of the IWE (4) is given in [33]. Letting

$$I^c(\mathbf{x}) \doteq I(\mathbf{x}) - \mu(I) \tag{41}$$

be the centered IWE, the derivative of the variance is, by the chain rule,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{Var}(I) = \frac{1}{|\Omega|} \int_\Omega 2I^c(\mathbf{x}) \frac{\partial I^c(\mathbf{x})}{\partial \boldsymbol{\theta}} d\mathbf{x}, \tag{42}$$

(formally, the same formula as (40), but with the centered IWE playing the role of the IWE in (40)) with

$$\frac{\partial I^c}{\partial \boldsymbol{\theta}} = \frac{\partial I}{\partial \boldsymbol{\theta}} - \mu\left(\frac{\partial I}{\partial \boldsymbol{\theta}}\right) \tag{43}$$

since the mean $\mu(\cdot)$ and the derivative are linear operators, and therefore, commute. The previous result (39) may be substituted in (43).

Loss Function: Mean Absolute Value (MAV). Derivative of the MAV is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{MAV}(I) = \frac{1}{|\Omega|} \int_\Omega \mathrm{sign}(I(\mathbf{x})) \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} d\mathbf{x}. \tag{44}$$

From a numerical point of view, it is sensible to replace $\mathrm{sign}(x)$ by a smooth approximation, e.g., $\mathrm{sign}(x) \approx \tanh(kx)$, with parameter $k \gg 1$ controlling the width of the transition around $x = 0$ (see [53]).

Loss Function: Mean Absolute Deviation (MAD). The derivative of the MAD, (10), is, by the chain rule and using (43),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{MAD}(I) = \frac{1}{|\Omega|} \int_\Omega \mathrm{sign}(I^c(\mathbf{x})) \frac{\partial I^c(\mathbf{x})}{\partial \boldsymbol{\theta}} d\mathbf{x}. \tag{45}$$

Loss Function: Entropy. The derivative of entropy (12) is, stemming from (29),

$$\frac{\partial}{\partial \boldsymbol{\theta}} H\left(p_I(z)\right) = \int_\Omega -\frac{1}{|\Omega|} \frac{p_I'(I(\mathbf{x}))}{p_I(I(\mathbf{x}))} \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \, d\mathbf{x},$$ (46)

where $p_I'(z) = dp_I/dz$. This formula can be obtained by differentiating (29) and applying the chain rule,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log p_I\left(I(\mathbf{x})\right) = \frac{p_I'\left(I(\mathbf{x})\right)}{p_I\left(I(\mathbf{x})\right)} \frac{\partial}{\partial \boldsymbol{\theta}} I(\mathbf{x}).$$ (47)

Implementation Details: We approximate the density function $p_I(z)$ by a smooth histogram of the IWE (3). First, a high-resolution histogram (e.g., 200 bins) is computed and normalized to unit area (like a PDF), and then it is smoothed by a Gaussian filter, e.g., of standard deviation $\sigma = 5$ bins. The smoothed PDF is the convolution

$$p_I^\sigma(z) \doteq p_I(z) * \mathcal{N}(z; 0, \sigma^2),$$ (48)

and it is straightforward to show, using the convolution properties and interchanging the order of integration, that it is equivalent to the function obtained by replacing the Dirac delta in (28) with a smooth approximation (such as the Gaussian: $\delta \leftarrow \mathcal{N}$):

$$p_I^\sigma(z) = \frac{1}{|\Omega|} \int_\Omega \mathcal{N}(z - I(\mathbf{x}); 0, \sigma^2) d\mathbf{x}.$$ (49)

Smoothing (i.e., filtering) mitigates the noise due to bin discretization and improves robustness (size of the basin of attraction) of the entropy-based focus function (12). Thus, the term in the integrand of (46) is approximated by

$$\frac{p_I'(z)}{p_I(z)} \approx \frac{(p_I^\sigma)'(z)}{p_I^\sigma(z)},$$ (50)

where the derivative $(p_I^\sigma)'$ is computed using central, finite-differences on the samples of $p_I^\sigma(z)$. Linear interpolation is used to interpolate the samples of p_I^σ and $(p_I^\sigma)'$.

Loss Function: Image Area. The derivative of the area (8) is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \operatorname{supp}(I) = \int_\Omega \rho(I(\mathbf{x})) \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} d\mathbf{x},$$ (51)

with weighting functions $\rho(\lambda)$. See Section 3.3 for different choices (exponential, Gaussian, Lorentzian and Hyperbolic).

Loss Function: Image Range. The derivative of the support of the PDF of the IWE (13) is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \operatorname{supp}(p_I) = \frac{1}{|\Omega|} \int_\Omega \rho'\left(p_I(I(\mathbf{x}))\right) \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} d\mathbf{x},$$ (52)

where $\rho'(\lambda) = d\rho/d\lambda$ is the derivative of the weighting function. This can be shown using the chain rule and the derivative of the PDF of the IWE (28), expressed in terms of the derivative of the Dirac delta.

Result: Derivative of a Convolution. The derivative of a convolution of the IWE with a kernel $K(\mathbf{x})$ is computed component-wise:

$$\frac{\partial}{\partial \boldsymbol{\theta}} \underbrace{\underbrace{(I * K)(\mathbf{x})}_{1\times 1}}_{1\times M} = \left(\ldots, \frac{\partial}{\partial \theta_i}(I * K)(\mathbf{x}), \ldots\right) = \left(\ldots, \frac{\partial I(\mathbf{x})}{\partial \theta_i} * K(\mathbf{x}), \ldots\right) \doteq \underbrace{\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}}}_{1\times M} * \underbrace{K(\mathbf{x})}_{1\times 1}.$$ (53)

Loss Function: Local MS. The derivative of the aggregated local MS of the IWE is, by the chain rule and (53),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_\Omega I^2(\mathbf{x}) * G_\sigma(\mathbf{x}) \, d\mathbf{x} = \int_\Omega \left(2I(\mathbf{x}) \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}}\right) * G_\sigma(\mathbf{x}) \, d\mathbf{x}.$$ (54)

Loss Function: Local Variance. The derivative of the aggregated local variance of the IWE is, by the chain rule on (14),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_\Omega \mathrm{Var}(\mathbf{x}; I)\, d\mathbf{x} = \int_\Omega \frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{Var}(\mathbf{x}; I)\, d\mathbf{x}, \tag{55}$$

where, using (16) and result (53), the integrand becomes

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathrm{Var}(\mathbf{x}; I) \overset{(16)}{\approx} \left(2I(\mathbf{x}) \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) * G_\sigma(\mathbf{x}) - 2(I(\mathbf{x}) * G_\sigma(\mathbf{x})) \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} * G_\sigma(\mathbf{x}) \right). \tag{56}$$

Loss Function: Local MAV. The derivative of the aggregated local MAV of the IWE is, using the chain rule and the compact notation (53),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_\Omega |I(\mathbf{x})| * G_\sigma(\mathbf{x})\, d\mathbf{x} = \int_\Omega \left(\mathrm{sign}(I(\mathbf{x})) \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) * G_\sigma(\mathbf{x})\, d\mathbf{x}. \tag{57}$$

Loss Function: Local MAD. Letting $I^c(\mathbf{x}) \doteq I(\mathbf{x}) - \mu(I)(\mathbf{x})$ be the locally-centered IWE, with local mean $\mu(I)(\mathbf{x}) \doteq I(\mathbf{x}) * G_\sigma(\mathbf{x})$, the derivative of the aggregated local MAD of the IWE is, using the chain rule and the compact notation (53),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_\Omega |I^c(\mathbf{x})| * G_\sigma(\mathbf{x})\, d\mathbf{x} = \int_\Omega \left(\mathrm{sign}(I^c(\mathbf{x})) \frac{\partial I^c(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) * G_\sigma(\mathbf{x})\, d\mathbf{x}, \tag{58}$$

with derivative of the locally-centered IWE

$$\frac{\partial I^c(\mathbf{x})}{\partial \boldsymbol{\theta}} = \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} - \mu\left(\frac{\partial I}{\partial \boldsymbol{\theta}} \right)(\mathbf{x}) \tag{59}$$

since the local mean and the derivative are linear operators, and hence, commute.

Loss Function: Gradient Magnitude. The derivative of the squared magnitude of the gradient of the IWE (5) is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \|\nabla I\|_{L^2(\Omega)}^2 = \int_\Omega 2(\nabla I(\mathbf{x}))^\top \frac{\partial \nabla I(\mathbf{x})}{\partial \boldsymbol{\theta}} d\mathbf{x}, \tag{60}$$

with $2 \times M$ matrix (assuming equality of mixed derivatives by Schwarz's theorem),

$$\frac{\partial \nabla I(\mathbf{x})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\theta}} I_x \\ \frac{\partial}{\partial \boldsymbol{\theta}} I_y \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x} \frac{\partial I}{\partial \boldsymbol{\theta}} \\ \frac{\partial}{\partial y} \frac{\partial I}{\partial \boldsymbol{\theta}} \end{pmatrix}. \tag{61}$$

Hence, (60) becomes

$$(60) = \int_\Omega 2 \left(I_x(\mathbf{x}) \frac{\partial}{\partial x} \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) + I_x(\mathbf{x}) \frac{\partial}{\partial y} \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) \right) d\mathbf{x}. \tag{62}$$

Loss Function: Laplacian Magnitude. Following similar steps as for the gradient magnitude, the derivative of the magnitude of the Laplacian of the IWE (6) is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \|\Delta I\|_{L^2(\Omega)}^2 = \int_\Omega 2\Delta I(\mathbf{x}) \Delta \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) d\mathbf{x}. \tag{63}$$

where we defined

$$\Delta \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) \doteq \frac{\partial^2}{\partial x^2} \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right) + \frac{\partial^2}{\partial y^2} \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right). \tag{64}$$

Loss Function: Hessian Magnitude. The derivative of Hessian magnitude (7) is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \| \operatorname{Hess}(I) \|_{L^2(\Omega)}^2 = \sum_{i,j} \frac{\partial}{\partial \boldsymbol{\theta}} \| I_{x_i x_j} \|_{L^2(\Omega)}^2, \tag{65}$$

where x_i, x_j are variables x, y of the image plane. Using once more Schwarz's theorem to swap the differentiation order, each of the four terms in the sum (65) is

$$\frac{\partial}{\partial \boldsymbol{\theta}} \| I_{x_i x_j} \|_{L^2(\Omega)}^2 = \int_{\Omega} 2 I_{x_i x_j}(\mathbf{x}) \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} \right)_{x_i x_j} d\mathbf{x}, \tag{66}$$

where subscripts x_i, x_j indicate differentiation with respect to those variables.

Loss Function: Difference of Gaussians (DoG). The derivative of the squared difference of Gaussians applied to the IWE can be written in a compact way, using (53) on the $\operatorname{DoG}(\mathbf{x}) \doteq (G_{\sigma_1} - G_{\sigma_2})(\mathbf{x})$ filter, with $\sigma_1 > \sigma_2$, as

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\Omega} (I(\mathbf{x}) * \operatorname{DoG}(\mathbf{x}))^2 \, d\mathbf{x} = \int_{\Omega} 2 \left(I(\mathbf{x}) * \operatorname{DoG}(\mathbf{x}) \right) \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} * \operatorname{DoG}(\mathbf{x}) \right) d\mathbf{x}, \tag{67}$$

In expanded form,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\Omega} (I(\mathbf{x}) * \operatorname{DoG}(\mathbf{x}))^2 \, d\mathbf{x} = \int_{\Omega} 2 \left((I * G_{\sigma_1})(\mathbf{x}) - (I * G_{\sigma_2})(\mathbf{x}) \right) \left(\frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} * G_{\sigma_1}(\mathbf{x}) - \frac{\partial I(\mathbf{x})}{\partial \boldsymbol{\theta}} * G_{\sigma_2}(\mathbf{x}) \right) d\mathbf{x} \tag{68}$$

In the experiments, we used $\sigma_1 = 1$ pixel and $\sigma_2 = 3\sigma_1$.

Loss Function: Laplacian of the Gaussian (LoG). The derivative of the squared difference of the Laplacian of the Gaussian of the IWE can be computed from the formula for the DoG, using the fact that the DoG approximates the LoG if $\sigma_2 \approx 1.6\sigma_1$.

Loss Function: Variance of the Laplacian. Derivative of the variance of the Laplacian:

$$\operatorname{Var}(\Delta I) \doteq \frac{1}{|\Omega|} \int_{\Omega} (\Delta I(\mathbf{x}) - \mu_{\Delta I})^2 d\mathbf{x}, \tag{69}$$

where the mean is $\mu_{\Delta I} \doteq \frac{1}{|\Omega|} \int_{\Omega} \Delta I(\mathbf{x}) d\mathbf{x}$. The derivative of (69) with respect to the parameters $\boldsymbol{\theta}$ is, using (64),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \operatorname{Var}(\Delta I) = \frac{1}{|\Omega|} \int_{\Omega} 2(\Delta I - \mu_{\Delta I}) \left(\Delta \left(\frac{\partial I}{\partial \boldsymbol{\theta}} \right) - \mu \left(\Delta \left(\frac{\partial I}{\partial \boldsymbol{\theta}} \right) \right) \right) d\mathbf{x}. \tag{70}$$

Loss Function: Variance of the Squared Gradient Magnitude. The derivative of the variance of the squared gradient magnitude

$$\operatorname{Var}(\|\nabla I\|^2) \doteq \frac{1}{|\Omega|} \int_{\Omega} (\|\nabla I(\mathbf{x})\|^2 - \mu_{\|\nabla I\|^2})^2 d\mathbf{x}, \tag{71}$$

with mean $\mu_{\|\nabla I\|^2} \doteq \frac{1}{|\Omega|} \int_{\Omega} \|\nabla I(\mathbf{x})\|^2 d\mathbf{x}$, is, using (61),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \operatorname{Var}(\|\nabla I\|^2) = \frac{1}{|\Omega|} \int_{\Omega} 2(\|\nabla I\|^2 - \mu_{\|\nabla I\|^2}) \left(2(\nabla I)^\top \frac{\partial \nabla I}{\partial \boldsymbol{\theta}} - \mu \left(2(\nabla I)^\top \frac{\partial \nabla I}{\partial \boldsymbol{\theta}} \right) \right) d\mathbf{x} \tag{72}$$

Loss Function: Variance of the Gradient Magnitude. Derivative of the variance of the gradient magnitude

$$\operatorname{Var}(\|\nabla I\|) \doteq \frac{1}{|\Omega|} \int_{\Omega} (\|\nabla I(\mathbf{x})\| - \mu_{\|\nabla I\|})^2 d\mathbf{x}, \tag{73}$$

with mean $\mu_{\|\nabla I\|} \doteq \frac{1}{|\Omega|} \int_{\Omega} \|\nabla I(\mathbf{x})\| d\mathbf{x}$, is, also using (61),

$$\frac{\partial}{\partial \boldsymbol{\theta}} \operatorname{Var}(\|\nabla I\|) = \frac{1}{|\Omega|} \int_{\Omega} 2(\|\nabla I\| - \mu_{\|\nabla I\|}) \left(\frac{(\nabla I)^\top}{\|\nabla I\|} \frac{\partial \nabla I}{\partial \boldsymbol{\theta}} - \mu \left(\frac{(\nabla I)^\top}{\|\nabla I\|} \frac{\partial \nabla I}{\partial \boldsymbol{\theta}} \right) \right) d\mathbf{x}. \tag{74}$$

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, 14, 22

[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, 15

[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 Galluppi, 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, 14, 15, 24

[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, 11, 14, 15

[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, 14, 15

[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, 14, 15, 17, 19

[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," in *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019. 1, 8

[40] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza, "Event-based motion segmentation by motion compensation," in *Int. Conf. Comput. Vis. (ICCV)*, 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," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, 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, 13, 14, 15, 23

[46] Rafael C. Gonzalez and Richard Eugene Woods, *Digital Image Processing*. Pearson Education, 2009. 2, 3, 15

[47] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, p. 17, June 1950. 4, 11

[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, 12

[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, 14, 15

[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, 24

[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

[56] Guillermo Gallego, Anthony Yezzi, Francesco Fedele, and Alvise Benetazzo, "Variational stereo imaging of oceanic waves with statistical constraints," *IEEE Trans. Image Process.*, vol. 22, pp. 4211–4223, Nov. 2013. 11