Distance Surface for Event-Based Optical Flow

Mohammed Almatrafi, *Member, IEEE*, Raymond Baldwin,
Kiyoharu Aizawa, *Fellow, IEEE*, and Keigo Hirakawa, *Senior Member, IEEE*

Abstract—We propose DistSurf-OF, a novel optical flow method for neuromorphic cameras. Neuromorphic cameras (or event detection cameras) are an emerging sensor modality that makes use of dynamic vision sensors (DVS) to report asynchronously the log-intensity changes (called "events") exceeding a predefined threshold at each pixel. In absence of the intensity value at each pixel location, we introduce a notion of "distance surface"—the distance transform computed from the detected events—as a proxy for object texture. The distance surface is then used as an input to the intensity-based optical flow methods to recover the two dimensional pixel motion. Real sensor experiments verify that the proposed DistSurf-OF accurately estimates the angle and speed of each events.

Index Terms—Motion estimation, optical flow, dynamic vision sensor, neuromorphic camera

1 Introduction

Optical flow refers to the task of estimating the apparent motion in a visual scene. It has been a major topic of research in computer vision for the past few decades due to the significant role it plays in various machine vision applications, including navigation [1], [2], [3], [4], segmentation [5], [6], image registration [7], tracking [8], [9], [10], and motion analysis [11]. While remarkable progress have been made since the original concepts were introduced by Horn-Schunck [12] and Lucas-Kanade [13], optical flow in the presence of fast motion and occlusions remains a major challenge today [14], [15], [16].

In recent years, neuromorphic cameras have gained popularity in applications that require cameras to handle high dynamic range and fast scene motion scenes. Unlike the conventional active pixel sensor (APS) that records an image intensity at a (slow) synchronous frame-rate, DVS in neuromorphic cameras asynchronously reports spikes called "events" when the log-brightness change exceeds a fixed threshold. Since these events only occur at object edges, they are very sparse. See Fig. 1a. DVS represents a significant reduction in memory storage and computational cost, increase in temporal resolution (+800kHz), higher dynamic range (+120dB), and lower latency (in the order of microseconds). Thus, neuromorphic cameras have the potential to improve the performance of optical flow methods, which are currently limited by the slow frame rate of the conventional camera's hardware. The main challenges to working with neuromorphic cameras, however, is the lack of the notion of pixel intensity, which renders conventional image processing and computer vision tools useless.

In this work, we propose *DistSurf-OF*, a novel DVS-based optical flow method that is robust to complex pixel motion vectors and scenes. We achieve this by introducing a novel notion of "distance surface," designed to corroborate pixel velocity from multiple edge pixels of varying edge orientations. We interpret the distance surface as a proxy for pixel intensity values in conventional cameras and treat its spatial derivatives as the "object textures" of non-edge pixels. This disambiguates the pixel motion and its temporal derivative as the encoding of the texture changes over time. See Figs. 1b, 1c, and 1d. The computed distance surface derivatives are then used as an input to the standard optical flow methods to recover the two dimensional pixel motion.

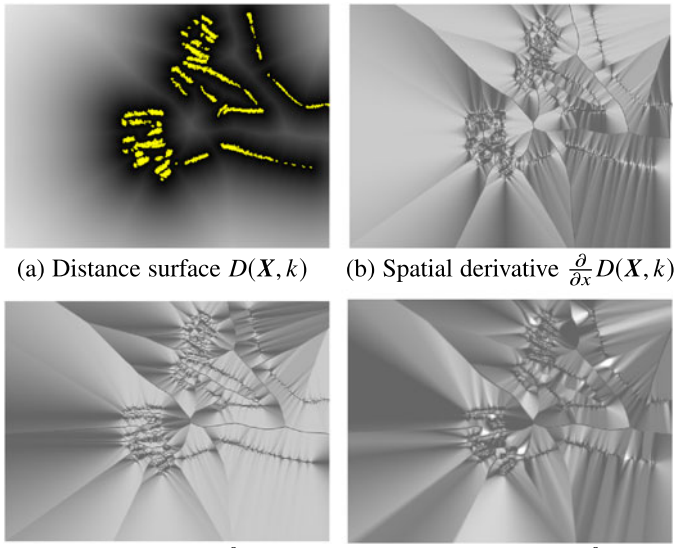
The main contributions of this paper are as follows:

- *Distance Surface:* We assign an intensity value to each pixel based on the proximity to the spatially closest detected event pixel. These values represent the object shapes by the relative positions of their edges, which satisfy the optical flow equations.
- *DistSurf-OF:* We recover the pixel motion field from the spatial-temporal gradients of the distance surface. The computed motion field draws on multiple events corresponding to multiple edge orientations, improving the robustness to motion and scene complexity.
- *Noise Robustness Study:* We employ event denoising to improve optical flow performance. We also analyze its positive impact on existing DVS-optical flow methods.
- *DVSMOTION20:* We present a new optical flow dataset of complex scenes and camera motion. Ground truth motion can be inferred from the rotational camera motion measured by the inertial measurement unit.

The remainder of this paper is organized as follows. In Section 2, we briefly review the requisite background

- *M. Almatrafi is with the Department of Electronic and Communication Engineering, Umm Al-Qura University, Al-Lith 28434, Saudi Arabia. E-mail: mmmatrafi@uqu.edu.sa.*
- *R. Baldwin is with the Department of Electrical and Computer Engineering, University of Dayton, Dayton, Ohio 45469. E-mail: baldwinr2@udayton.edu.*
- *K. Aizawa is with the University of Tokyo, Bunkyo City, Tokyo 113-8656, Japan. E-mail: aizawa@hal.t.u-tokyo.ac.jp.*
- *K. Hirakawa is with the Department of Electrical and Computer Engineering, University of Dayton, Dayton, Ohio 45469. E-mail: khirakawa1@udayton.edu.*

Manuscript received 19 Apr. 2020; revised 27 Mar. 2020; accepted 4 Apr. 2020. Date of publication 13 Apr. 2020; date of current version 3 June 2020. (Corresponding author: M. Almatrafi.)
Recommended for acceptance by A. Chakrabarti.
Digital Object Identifier no. 10.1109/TPAMI.2020.2986748



(a) Distance surface $D(\boldsymbol{X}, k)$ (b) Spatial derivative $\frac{\partial}{\partial x} D(\boldsymbol{X}, k)$

(c) Spatial derivative $\frac{\partial}{\partial y} D(\boldsymbol{X}, k)$ (d) Temporal derivative $\frac{\partial}{\partial t} D(\boldsymbol{X}, k)$

Fig. 1. Spatial/temporal derivatives of the proposed distance surface on *hand* sequence. Yellow pixels denote denoised events.

materials. In Section 3, we then propose the proposed distance surface-based optical flow method for DVS cameras. Section 4 outlines methods to increase algorithm robustness. We describe DVSMOTION20 and present the real-data experimental results in Section 5 before making the concluding remarks in Section 6.

2 Background and Related Work

2.1 Frame-Based Optical Flow

Let $I : \mathbb{Z}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ denote an intensity video, where $I(\boldsymbol{X}, t)$ is the pixel radiance at pixel $\boldsymbol{X} = (x, y)^T \in \mathbb{Z}^2$ of a video frame at time $t \in \mathbb{R}$. Known as the "brightness constancy assumption," optical flow is derived from the hypothesis that the the pixel intensities of translated objects remain constant over time [12]:

$$I(\boldsymbol{X} + \Delta\boldsymbol{X}, t + \Delta t) = I(\boldsymbol{X}, t), \tag{1}$$

where $\Delta\boldsymbol{X} = (\Delta x, \Delta y)$ and Δt denote spatial and temporal translations, respectively. Assuming that such translations are small, (1) is expanded via first-order Taylor series to give rise to the well-known "optical flow equation":

$$\nabla I(\boldsymbol{X}, t)\boldsymbol{V}(\boldsymbol{X}, t) + I_t(\boldsymbol{X}, t) \approx 0, \tag{2}$$

where $\nabla I(\boldsymbol{X}, t) = (\frac{\partial}{\partial x} I(\boldsymbol{X}, t), \frac{\partial}{\partial y} I(\boldsymbol{X}, t))$ and $I_t(\boldsymbol{X}, t) = \frac{\partial}{\partial t} I(\boldsymbol{X}, t)$ are the spatial gradient and temporal derivative, respectively. The goal of the optical flow task is to estimate the two dimensional pixel motion field $\boldsymbol{V}(\boldsymbol{X}, t) = (\frac{\Delta x}{\Delta t} \frac{\Delta y}{\Delta t})^T$.

The pixel motion field cannot be estimated directly from the under-determined system of equations in (2) because the component of $\boldsymbol{V}$ parallel to the edges lives in the null-space of $\nabla I(\boldsymbol{X}, t)$. We overcome this issue—commonly referred to as the "aperture problem"—by imposing additional constraints. An example of such constraint is the flow magnitude minimization. However, its solution is a motion perpendicular in direction to the edge orientation—a phenomenon often referred to as the "normal flow"—which does not necessarily represent the actual two dimensional motion of the object in general. Overcoming the normal flow problem requires diversifying the gradient orientation $\nabla I(\boldsymbol{X}, t)$ by incorporating multiple pixels. The "local spatial consistency" constraint proposed by Lucas-Kanade helps overcome noise and variations by requiring $\boldsymbol{V}(\boldsymbol{X}, t)$ to be the same within a spatial neighborhood [13]. Similarly, Horn-Schunck (HS) introduced "global spatial constancy" criteria promoting smoothness of $\boldsymbol{V}(\boldsymbol{X}, t)$ globally by adding a quadratic penalty to (2) as a regularization term [12]. More recently developed methods improve upon these classical optical flow methods to yield state-of-the-art performance [17] by modifying penalty terms [14], [16], [18], [19] and leveraging phase [20], [21], [22] and block [23], [24] correlations.

Conventional optical flow methods applied to an intensity video sequence with a relatively slow frame rate (e.g., 30 or 60 frames per second) often leads to unreliable optical flow. Instability stems from large translations $\Delta\boldsymbol{X}$ and Δt in the presence of fast motion [16], [25], which invalidate the first order Taylor series approximation in (2). Similarly, variations in illumination conditions and occlusion cause changes over time in the intensity values of spatially translated features, negating the brightness constancy assumption in (1). Current solutions to deal with these challenges include spatial pyramids [25], [26], [27], [28], local layering [29], and robust penalty functions such as generalized Charbonnier [17], Charbonnier [18], and Lorentzian [28]. Additionally, the reliability of optical flow can be also enhanced by outlier removals [30], texture decomposition [30], [31], and smoothing filters [14], [18] to preserve the brightness assumption.

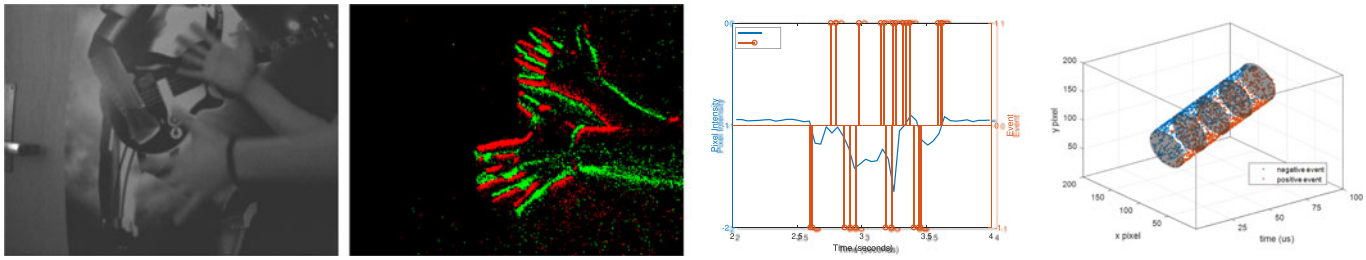
2.2 Dynamic Vision Sensor (DVS)

Instead of recording an image intensity at a synchronous frame-rate, neuromorphic cameras record asynchronous positive and negative spikes called events. These events are generated when the log-brightness change exceeds a fixed threshold ℓ:

$$\begin{aligned} t_{i+1}(\boldsymbol{X}) &= \arg\min_t \left\{ t > t_i \middle| \left| \log\left(\frac{I(\boldsymbol{X}, t)}{I(\boldsymbol{X}, t_i)}\right) \right| > \ell \right\} \\ p_{i+1}(\boldsymbol{X}) &= \text{sign}(I(\boldsymbol{X}, t_{i+1}) - I(\boldsymbol{X}, t_i)), \end{aligned} \tag{3}$$

where $t_{i+1}(\boldsymbol{X})$ denotes the time of event occurrence, whose accuracy is in the order of microseconds; and $p_{i+1}(\boldsymbol{X})$ is the polarity, indicating whether the intensity change is darker (-1) or brighter (+1). The sparsity of the events reduces throughput and memory storage considerably, enabling high temporal resolution and low latency. It has been widely speculated that such characteristics can help overcome the limitations of conventional frame-based optical flow methods. Specifically, the microseconds resolution implies $\Delta\boldsymbol{X}$ and Δt are small, better preserving the validity of the Taylor series expansions in (2) and minimizing the risks of occlusion, even in the presence of fast motion.

Optical flow for DVS is a task of determining the velocities of the pixels that generated the observed events. Because of the fact that DVS outputs lack the notion of pixel intensity (and brightness constancy assumption in (1) is largely invalid), optical flow requires an entirely new



(a) APS Data (b) DVS Data (c) APS vs DVS as function of time (d) Spatial-temporal data

Fig. 2. Example outputs from DAViS sensor, which has APS and DVS readout circuits sharing the same photodiode. (a) APS's synchronous outputs record pixel intensities. (b) DVS's asynchronous events report positive (green) and negative (red) changes in a pixel intensity. (c) Superposition of DVS and APS outputs of one particular pixel. (d) An example of a dark disk moving quickly across the scene. APS synchronously records the disk every 25 ms (at 40 Hz). DVS operates at 1 MHz, tracking the disks movement in between APS frames.

approach. Prior efforts were aimed at creating a proxy to image intensity (via accumulation of events over a temporal window) [32], the spatial gradient images (via central difference) [33], and the temporal derivative (via second order backward difference) [34]. Early DVS-specific optical flow approaches include local plane fitting that infers the pixel motion by fitting spatial-temporal manifold to the events based on their time stamps [33], [34], [35], edge orientations estimation [36], [37], and block matching [38], [39]. Recent event-based techniques simultaneously estimate optical flow along with other machine vision tasks, such as intensity estimation [7], depth along with motion reconstruction [2], contrast maximization [2], [6], and segmentation [5], [6]. Appealing to recent successes of machine learning approaches to optical flow [40], [41], [42], learning-based optical flow methods for neuromorphic cameras have recently been proposed [4], [43], [44], [45]. Finally, the method in [46] exploits the high spatial fidelity of APS and temporal fidelity of DVS to compute spatial temporal derivatives for hybrid hardware called DAViS [47]. Benchmarking datasets for DVS-based optical flow and real-time performance evaluation platform have helped accelerate the progress of research in optical flow for neuromorphic cameras [33], [43], [46].

3 Proposed: DistSurf Optical Flow

We propose a new DVS-based optical flow method designed to estimate the apparent motion in the scene at the edge pixels. See Fig. 3 for an overview. The underlying assumption is that the pixel spatial velocities of a rigid- or semirigid-body object are slowly varying. As such, the pixel velocity of a pixel internal to a semiregid-body object can be inferred from the edge pixels surrounding it. We propose a novel notion of "distance surface" as a way to leverage multiple edges of a semirigid-bodied object and as a proxy for object textures—incorporating multiple edges avoids the pitfalls of normal flow that many existing DVS-based optical flow algorithms suffer from.

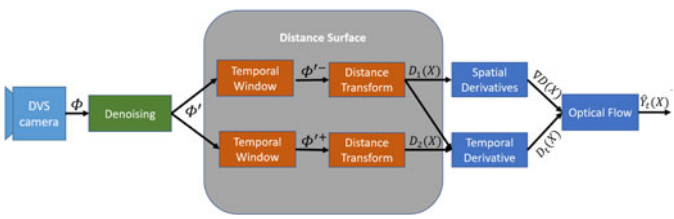


Fig. 3. System diagram for DistSurf-OF.

3.1 Distance Surface

We are interested in recovering the two dimensional motion field $V(X,t)$ from events detected immediately before and immediately after the time $t \in \mathbb{R}$. Define $\Phi \subset \mathbb{Z}^2$ as a set of pixel indices corresponding to detected events that occurred within a temporal window $[t-\Delta t, t)$, as follows:

$$\Phi = \{X \in \mathbb{Z}^2 | \exists i \ni t - \Delta t \leq t_i(X) < t\}, \tag{4}$$

where $t_i(X)$ denotes the time-stamps of the events as described in Section 2. Let $d : \mathbb{Z}^2 \times \mathbb{Z}^2 \to \mathbb{R}$ be a distance measure function, with $d(X,Y)$ representing the spatial distance between pixels $X = (X_1, X_2)^T \in \mathbb{Z}^2$ and $Y = (Y_1, Y_2)^T \in \mathbb{Z}^2$. In this work, we consider the L^2 norm taking the form:

$$d(X,Y) = \sqrt{(X-Y)^T(X-Y)}. \tag{5}$$

Then, a distance transform converts the sets $\Phi \in \mathbb{Z}$ to gray-level image $D : \mathbb{Z}^2 \to \mathbb{R}$ by:

$$D(X) = \min_{Y \in \Phi} d(X,Y) = d\Big(X, \widehat{Y}(X)\Big). \tag{6}$$

with corresponding indexes $\widehat{Y} : \mathbb{Z}^2 \to \mathbb{Z}^2$ as edge pixels deemed "closest" to X:

$$\widehat{Y}(X) = \arg\min_{Y \in \Phi} d(X,Y). \tag{7}$$

An example of distance transform is illustrated in Fig. 4. It is clear that $D(X) = 0$ when $X \in \Phi$. At a non-edge pixel $X \notin \Phi$, the pixel intensity $D(X) > 0$ represents the distance to the nearest edge pixel $\widehat{Y}(X) \in \Phi$.

We propose to use the gray-level image $D(X)$ computed from DVS as a proxy for intensity value in APS images—a notion we hereafter refer to as "distance surface." That is, we treat the distance surface $D(X)$ as object textures on pixels that are internal to an object, away from edges. While a full justification for using distance surface in the context of optical flow is provided in Section 3.2, we can already see in Fig. 4 some of the reasons that the notion of distance surface is well suited for the optical flow task—although $D(X)$ is a smooth function of X, the spatial gradient of $D(X)$ is influenced by the selection of different edge pixels $\widehat{Y}(X) \in \Phi$ closest to pixel X internal to the rigid-bodied object. As such, an optical flow computed from the distance surface implicitly incorporates multiple edge pixels (of multiple edge orientations) to establish an object motion, so as to

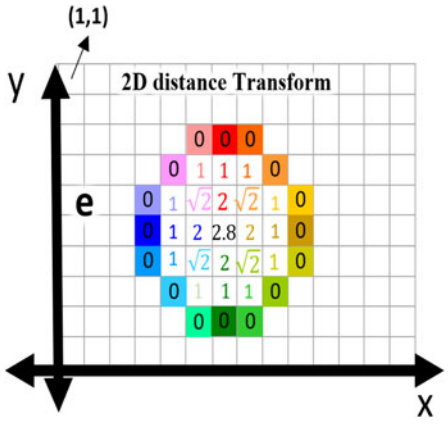


Fig. 4. An example of a distance transform. Solid colored pixels indicate event/edge pixels. Numbers indicate the minimum distance between a pixel and an edge. The color of the numbers indicate the closest edge pixel.

overcome the normal flow problem that hampers many optical flow algorithms.

Our work in distance surface can be interpreted as an alternative to the "time surface" [48]—time surface, too, can be interpreted as a type of a distance transform (in the temporal domain):

$$T(\boldsymbol{X}) = \min_{t_i(\boldsymbol{X}) < t} \|t - t_i(\boldsymbol{X})\|. \tag{8}$$

Using time surface as a proxy for pixel intensity values in APS images has been proven useful in pattern recognition applications [49], [50]. However, the brightness constancy assumption is invalid for $T(\boldsymbol{X}, k)$, making it a poor choice for optical flow.

Finally, note that the definition of distance transform in (6) makes no distinction of the polarity of the DVS generated events, as our goal is to recover the edges of the rigid-bodied objects surrounding the internal pixels X—an intuition confirmed via empirical experiments. Since the polarity of the events occurring at the object boundary is an encoding of the contrast between the foreground and the background, coupling the pixel motion estimation to the background texture properties only destabilized our optical flow method.

3.2 Distance Surface Optical Flow Equation

The distance surface in (6) is very well matched for the DVS optical flow task. To understand why this is the case, we begin by first proving that the distance surface satisfies the optical flow equation in (2). Taking a spatial gradient of the distance surface, we have the following relation:

$$\begin{aligned}\nabla D(\boldsymbol{X}) &= \frac{\left(\boldsymbol{X} - \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)^T}{d\left(\boldsymbol{X}, \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)} - \frac{\left(\boldsymbol{X} - \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)^T J_{\widehat{\boldsymbol{Y}}}(\boldsymbol{X})}{d\left(\boldsymbol{X}, \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)} \\ &= \frac{\left(\boldsymbol{X} - \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)^T}{d\left(\boldsymbol{X}, \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)},\end{aligned} \tag{9}$$

where above, we used the fact that the vector $\boldsymbol{X} - \widehat{\boldsymbol{Y}}(\boldsymbol{X})$ is in the nullspace of the Jacobian matrix $J_{\widehat{\boldsymbol{Y}}} \in \mathbb{R}^{2\times 2}$ in the last step (see Fig. 6). Similarly, consider the temporal derivative of the distance surface:

$$D_t(\boldsymbol{X}) := \frac{\partial}{\partial t} D(\boldsymbol{X}) = -\frac{\left(\boldsymbol{X} - \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)^T}{d\left(\boldsymbol{X}, \widehat{\boldsymbol{Y}}(\boldsymbol{X})\right)} \widehat{\boldsymbol{Y}}_t(\boldsymbol{X}), \tag{10}$$

where $\widehat{\boldsymbol{Y}}_t(\boldsymbol{X}) := \frac{\partial}{\partial t}\widehat{\boldsymbol{Y}}(\boldsymbol{X})$ is the event pixel velocity we are after. Thus, we arrive at a new optical flow equation for distance surface:

$$\nabla D(\boldsymbol{X})\widehat{\boldsymbol{Y}}_t(\boldsymbol{X}) + D_t(\boldsymbol{X}) = 0. \tag{11}$$

Note that unlike the classical optical flow equation in (2), distance surface optical flow equation in (11) is exact (i.e., derived without Taylor series approximation), meaning it is robust to fast motions.

Hence, we propose to recover the event pixel velocity $\widehat{\boldsymbol{Y}}_t(\boldsymbol{X})$ by leveraging distance surface optical flow equation in (11). That is, we solve for $\widehat{\boldsymbol{Y}}_t(\boldsymbol{X})$ using classical optical flow methods. Spatial derivative $\nabla D(\boldsymbol{X})$ may be computed directly from the distance transform $D(\boldsymbol{X})$ using a spatial derivative convolution filter (the same way $\nabla I(\boldsymbol{X})$ in (2) is computed in conventional optical flow methods):

$$\nabla D(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial}{\partial X_1} D(\boldsymbol{X}) \\ \frac{\partial}{\partial X_2} D(\boldsymbol{X}) \end{bmatrix} = \begin{bmatrix} H_X(\boldsymbol{X}) * D(\boldsymbol{X}) \\ H_Y(\boldsymbol{X}) * D(\boldsymbol{X}) \end{bmatrix}. \tag{12}$$

for some horizontal and vertical spatial derivative filters $H_X, H_Y : \mathbb{Z}^2 \to \mathbb{R}$. For temporal derivative, we consider difference of two distance surfaces:

$$D_t(\boldsymbol{X}) = \frac{1}{\Delta t}\left(\left(\min_{\boldsymbol{Y}\in\Phi^+} d(\boldsymbol{X}, \boldsymbol{Y})\right) - \left(\min_{\boldsymbol{Y}\in\Phi^-} d(\boldsymbol{X}, \boldsymbol{Y})\right)\right), \tag{13}$$

where the sets $\Phi^-, \Phi^+ \subset \mathbb{Z}^2$ correspond to a narrow temporal window immediately before and after $t \in \mathbb{R}$:

$$\begin{aligned}\Phi^- &= \{\boldsymbol{X} \in \mathbb{Z}^2 | \exists i \ni t - \Delta t \le t_i(\boldsymbol{X}) < t\} \\ \Phi^+ &= \{\boldsymbol{X} \in \mathbb{Z}^2 | \exists i \ni t \le t_i(\boldsymbol{X}) < t + \Delta t\}.\end{aligned} \tag{14}$$

Spatial and temporal gradients $\nabla D(\boldsymbol{X})$ and $D_t(\boldsymbol{X})$ may be used subsequently as the input to *any* conventional optical flow method to yield an estimation of pixel velocity $\widehat{\boldsymbol{Y}}_t(\boldsymbol{X})$ at every pixel.

In our implementation, we used the classical Horn Schunck approach [12] aimed at minimizing the global energy functional $\boldsymbol{E}_{HS} : \ell^2(\mathbb{Z}^2) \to \mathbb{R}$ given by

$$\boldsymbol{E}_{HS}\left(\widehat{\boldsymbol{Y}}_t(\boldsymbol{X})\right) = \sum_{\boldsymbol{X}\in\mathbb{Z}^2} \rho\left(\nabla D(\boldsymbol{X})\widehat{\boldsymbol{Y}}_t(\boldsymbol{X}) + D_t(\boldsymbol{X})\right) + \lambda\, \rho\left(\nabla\widehat{\boldsymbol{Y}}_t(\boldsymbol{X})\right). \tag{15}$$

Here, λ is the regularization parameter (set to 0.1 in this work) and $\rho(x) = \log\left(1 + \frac{\|x\|^2}{2\sigma^2}\right)$ denotes Lorentzian robust penalty function [28] as implemented in [16]. The regularization term in effect imposes spatial consistency on the estimated event pixel velocities based on neighboring events that comprise an object boundary shape. By relying on multiple events within the contour of the object, we diversify the edge orientations and reduce the risk of normal flow motion artifacts, increase robustness to noise, etc. Readers

Fig. 5. An illustration of distance transform (shown using L^1 distance measure function rather than L^2 used in our algorithm because it is easier to follow). Black pixels denote detected events. Distance transform is robust to "missing" events and "jagged edges"—as evidenced by orange- and green-shaded pixels, the affected pixels are negligibly different from the desired distance transform values. By contrast, false positive events deteriorate a large number of the pixels (shaded in blue) by a large margin.

are reminded that while the method in (15) yields a state-of-the-art result, the proposed DistSurf-OF framework is agnostic to the choice of intensity-based optical flow method it is paired with in general.

Recalling the mapping in (7), suppose there are multiple non-edge pixels $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N\}$ that map back to the same edge pixel:

$$\boldsymbol{Y} = \widehat{\boldsymbol{Y}}(\boldsymbol{X}_1) = \widehat{\boldsymbol{Y}}(\boldsymbol{X}_2) = \ldots = \widehat{\boldsymbol{Y}}(\boldsymbol{X}_N). \tag{16}$$

Then technically, $\left\{\widehat{\boldsymbol{Y}}_t(\boldsymbol{X}_1), \widehat{\boldsymbol{Y}}_t(\boldsymbol{X}_2), \ldots, \widehat{\boldsymbol{Y}}_t(\boldsymbol{X}_N)\right\}$ are all valid estimates of the event pixel velocity $\boldsymbol{V}(\boldsymbol{Y}, t)$. In our work, we chose a simplistic and computationally efficient approach. Since $\widehat{\boldsymbol{Y}}(\boldsymbol{Y}) = \boldsymbol{Y}$ (i.e., an event pixel $\boldsymbol{Y} \in \Phi$ is closest to itself), we assigned the estimated distance surface velocity to the final event pixel velocity $\boldsymbol{V}(\boldsymbol{Y}, t)$ as follows:

$$\boldsymbol{V}(\boldsymbol{Y}, t) = \widehat{\boldsymbol{Y}}_t(\boldsymbol{Y}). \tag{17}$$

However, a more sophisticated approach combining $\left\{\widehat{\boldsymbol{Y}}_t(\boldsymbol{X}_1), \widehat{\boldsymbol{Y}}_t(\boldsymbol{X}_2), \ldots, \widehat{\boldsymbol{Y}}_t(\boldsymbol{X}_N)\right\}$ to yield the final estimate $\boldsymbol{V}(\boldsymbol{Y}, t)$ may increase robustness to perturbations. This is left as a future topic of investigation.

4 Robustness Analysis and Denoising

Let Ψ in (4) denote the "ideal" event detection corresponding to the image edges. DVS suffers from considerably high noise (random events) along with the signal due to multiple factors such as electronic noise and sensor heat [51]. As such, the set of actual observed events Φ is a perturbed version of the ideal set of events Ψ in the following sense:

$$\Phi = \{\Psi \cap \bar{\Lambda}\} \cup \Omega, \tag{18}$$

where $\Omega \subset \mathbb{Z}^2$ is a set of random DVS activations (i.e., false positives); $\Lambda \subset \Psi$ is a set of "holes" or missing events (i.e., false negatives) randomly excluded from Ψ; and $\bar{\Lambda}$ denotes the complimentary set to Λ. *What is the practical impact of computing the distance transform using Φ instead of Ψ?*

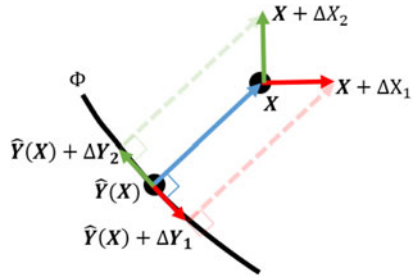


Fig. 6. Event $\widehat{\boldsymbol{Y}}(\boldsymbol{X}) \in \Phi$ is the point on the edge manifold Φ closest to pixel location $\boldsymbol{X}$. At $\widehat{\boldsymbol{Y}}(\boldsymbol{X})$, the gradient vectors (red and green arrows) are orthogonal to $\boldsymbol{X} - \widehat{\boldsymbol{Y}}(\boldsymbol{X})$ (blue arrow). Thus, $\boldsymbol{X} - \widehat{\boldsymbol{Y}}(\boldsymbol{X})$ lives in the nullspace of the Jacobian matrix in (9) (whose column vectors are the gradient vectors of $\widehat{\boldsymbol{Y}}(\boldsymbol{X})$).

The distance transform in (6) is robust to "holes" in the edge pixels of the binary frame Ψ. To understand why this is the case, let $\boldsymbol{Y} \in \Psi$ and $\boldsymbol{Z} \in \Psi$ be the edge pixels where $d(\boldsymbol{X}, \boldsymbol{Y}) < d(\boldsymbol{X}, \boldsymbol{Z})$—that is, $\boldsymbol{Y}$ is closer to $\boldsymbol{X}$ than $\boldsymbol{Z}$. Then by the triangle inequality of distance measure functions, we have the following relation:

$$d(\boldsymbol{X}, \boldsymbol{Z}) \leq d(\boldsymbol{X}, \boldsymbol{Y}) + d(\boldsymbol{Y}, \boldsymbol{Z}). \tag{19}$$

Suppose further that $\boldsymbol{Y} \in \Lambda$ is missing (i.e., $\boldsymbol{Y} \notin \Phi$). Then the penalty for replacing $d(\boldsymbol{X}, \boldsymbol{Y})$ by $d(\boldsymbol{X}, \boldsymbol{Z})$ is

$$0 < d(\boldsymbol{X}, \boldsymbol{Z}) - d(\boldsymbol{X}, \boldsymbol{Y}) \leq d(\boldsymbol{Y}, \boldsymbol{Z}). \tag{20}$$

Owing to the fact that edge pixels occur in clusters (i.e., $d(\boldsymbol{Y}, \boldsymbol{Z})$ is small), we conclude $d(\boldsymbol{X}, \boldsymbol{Y}) \approx d(\boldsymbol{X}, \boldsymbol{Z})$. Thus, the distance transform $D(\boldsymbol{X}, k)$ is largely invariant to random exclusions of events Λ in Ψ:

$$\min_{\boldsymbol{Y} \in \Psi} d(\boldsymbol{X}, \boldsymbol{Y}) \approx \min_{\boldsymbol{Y} \in \Psi \cap \bar{\Lambda}} d(\boldsymbol{X}, \boldsymbol{Y}). \tag{21}$$

Same analysis applies to jagged edges, where detected events at the edges are displaced by one or two pixels. See Fig. 5 for examples.

On the other hand, the distance surface is vulnerable to randomly activated events (i.e., false events) in Ω. Let us rewrite the distance transform as follows:

$$\min_{\boldsymbol{Y} \in \Psi \cup \Omega} d(\boldsymbol{X}, \boldsymbol{Y}) = \min\left\{\min_{\boldsymbol{Y} \in \Psi} d(\boldsymbol{X}, \boldsymbol{Y}),\ \min_{\boldsymbol{Z} \in \Omega} d(\boldsymbol{X}, \boldsymbol{Z})\right\}. \tag{22}$$

In other words, random distance transform $\min_{\boldsymbol{Z} \in \Omega} d(\boldsymbol{X}, \boldsymbol{Z})$ is a source of significant degradation to the desired distance surface $\min_{\boldsymbol{Y} \in \Psi} d(\boldsymbol{X}, \boldsymbol{Y})$. As such, the severity of degradation increases with the distance $d(\boldsymbol{X}, \boldsymbol{Y})$. See Fig. 5. Therefore, the proposed use of distance surface would benefit from denoising randomly activated events.

Denoising used in this work is a modified version of the filtering proposed in [52]. In this work, event $t_i(\boldsymbol{X})$ is classified into one of the following three categories: Background Activity (BA), Inceptive Event (IE), or Trailing Event (TE). They are defined as:

$$\begin{aligned} \text{BA}: &\{t_i(\boldsymbol{X}) - t_{i-1}(\boldsymbol{X}) > \tau\} \cap \{t_{i+1}(\boldsymbol{X}) - t_i(\boldsymbol{X}) > \tau\} \\ \text{IE}: &\{t_i(\boldsymbol{X}) - t_{i-1}(\boldsymbol{X}) > \tau\} \cap \{t_{i+1}(\boldsymbol{X}) - t_i(\boldsymbol{X}) < \tau\} \\ \text{TE}: &\{t_i(\boldsymbol{X}) - t_{i-1}(\boldsymbol{X}) < \tau\}, \end{aligned} \tag{23}$$

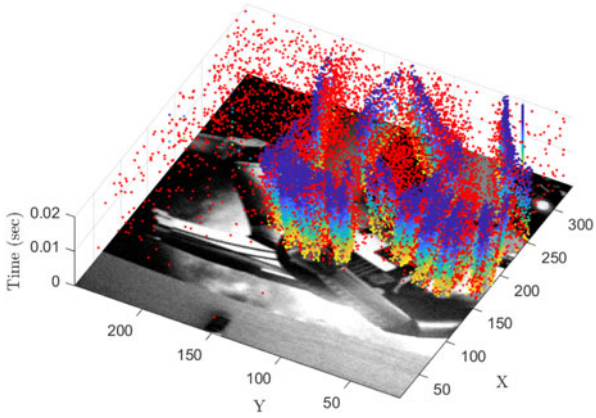


Fig. 7. Example of denoising. Red points indicate BA (or randomly activated events) that we filter out. Our optical flow method is being applied to IE and TE events that comprise remainder of points.

where τ is a threshold value. Intuitively, a single log-intensity change often trigger multiple events of the same polarity in rapid temporal succession. IE corresponds to the first of these events, indicating an arrival of an edge. IE is followed by the TE, which is proportional in number to the magnitude of the log-intensity change that occurred with the inceptive event. Remaining events are called BA, and they are attributed to noise or random activation events.

In recognition tasks, it was demonstrated empirically that IE was shown to be most useful for describing object shapes [52]. In our work, however, we are concerned about the negative impact of random activations (BA) on the distance surfaces. Thus we exclude BA from Φ in (4), as follows:

$$\Phi' = \left\{ \boldsymbol{X} \in \mathbb{Z}^2 \middle| \begin{array}{c} \exists i \ni \{t - \Delta t \leq t_i(\boldsymbol{X}) < t\} \cap \\ \{\{t_i(\boldsymbol{X}) - t_{i-1}(\boldsymbol{X}) < \tau\} \cup \{t_{i+1}(\boldsymbol{X}) - t_i(\boldsymbol{X}) < \tau\}\} \end{array} \right\}. \tag{24}$$

The effectiveness of BA exclusion is evident in Fig. 7.

In practice, denoising by BA exclusion improves the accuracy of all DVS-based optical flow methods, not limited to DistSurf-OF. The results in Section 5 are shown with and without the same denoising method applied to all optical flow methods.

5 Experimental Result

5.1 DVSMOTION20 Dataset

The existing benchmarking datasets have played a critical role in the progress of research in optical flow for neuromorphic cameras [33], [43], [46]. However, they are not without shortcomings. Motion and the scene contents in some sequences of [33] are overly short, simplistic, and unnaturally favor normal flow (where the motion is perpendicular to the edge orientation). The dataset in [43] update ground truth motion at 20 Hz sampling rate—slow considering that DVS is accurate to microseconds. The spatial resolution of sequences in [46] is small.

Therefore, we collected a new dataset—called DVSMOTION20—using IniVation DAViS346 camera in attempt to further enhance the progress of DVS-based optical flow methods. The DAViS346 camera has a 346 $\times$ 260 spatial resolution and outputs frames (APS) up to 60 frames per seconds, events in microsecond resolution, and a 6-axis IMU

Fig. 8. Camera setup for DVSMOTION20 collection. Gimbal limits camera motion while centering the focal point at the origin.

data at around 1 kHz sampling rate. We used a standard checkerboard calibration target to recover the intrinsic parameters of the camera. We infer ground truth pixel velocity stemming from camera motion using the inertial measurement unit (IMU), similar to prior benchmarkings in [33], [46]. Specifically, we placed the camera on a gimbal as shown in Fig. 8, restricting the movement to yaw, pitch, and roll rotations (i.e., no translations). This restriction to angular rotational motion ensures that pixel velocity can be recovered entirely from gyroscope data.

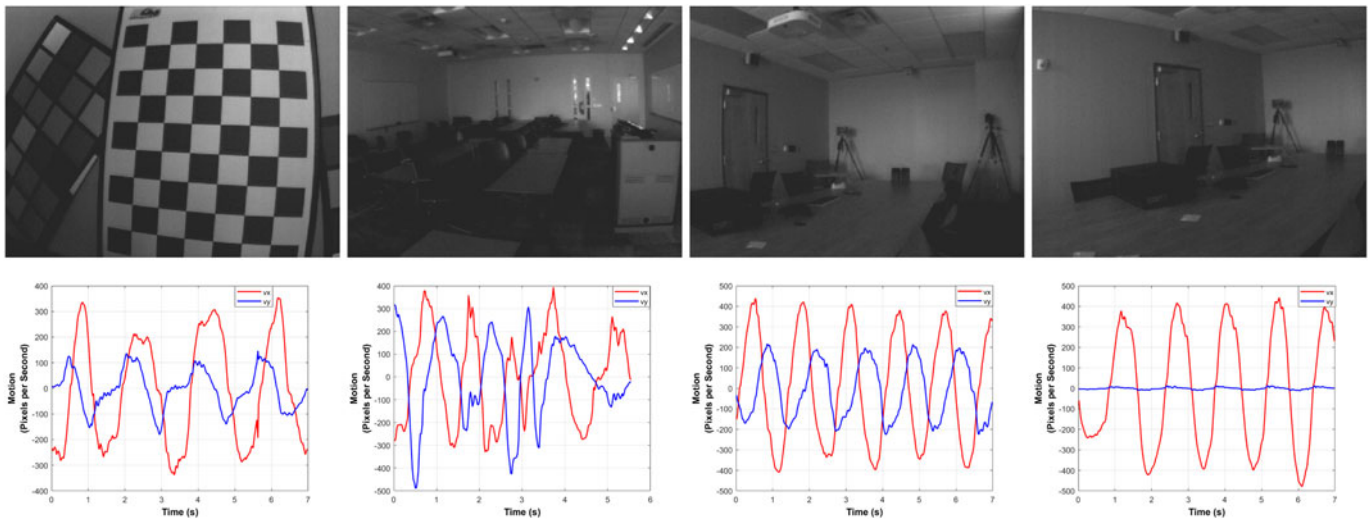
DVSMOTION20 dataset contains four real indoor sequences (*checkerboard*, *classroom*, *conference room*, and *conference room translation*). Each scene was captured for around 13-16 seconds with the first three seconds containing no motion for IMU calibration; 7-8 seconds of DVS data following the IMU calibration is used in the performance evaluation in Section 5.3. Each recorded data files are about 500 MB in size. See Fig. 9 for example frame content and the IMU trajectories. Although the sequences were restricted to camera motion with a stationary scene, all except for *conference room translation* contain fast motion with complex random camera movements—they should be challenging to optical flow methods that tend to yield normal flow. In *conference room translation* the motion was purely horizontal (yaw rotation). It helps verify the hypothesis that most optical flow methods generally work better when the pixel motion is simple.

We provide additional sequences in DVSMOTION20 (called *hands* and *cars*) containing multiple object motions (i.e., not camera motion). Object motion is spatially local by nature, but with strong motion boundaries. In our sequences, cars and hands moving in the opposite direction intersect, resulting in motion occlusion. See Fig. 11. Visual inspection of the estimated motion vectors are more than adequate for showing optical flow failures in presence of occlusion or large object motion (which happens frequently), even if the ground truth motion vectors are not available (because object motion cannot be inferred from IMU).

5.2 Setup and Comparisons

We compared DistSurf-OF to several state-of-the-art DVS-based optical flow algorithms including LK-DVS [32], LP [35], and EV-FlowNet [43]. In particular, the learning-based method in [43] (trained originally using MVSEC) is fine tuned with sequences in DVSMOTION20. Since training and testing data are identical, results we show in this paper represents its *best case scenario* performance.

We also compared to the state-of-the-art DAVIS optical flow method DAVIS-OF [46]. Because this method makes



(a) Checkerboard (b) Classroom (c) Conference Room (d) Conference Room Translation

Fig. 9. (Top row) Example APS frames extracted from DVSMOTION20 dataset. (Bottom row) IMU measurements recording the trajectory of the camera motion.

use of DVS (events, high temporal fidelity) as well as APS (intensity, high spatial fidelity) data, we expect it to perform better than the DVS-only methods. As evidenced in Table 1, however, the performance of proposed DistSurf-OF comes surprisingly close to this DAVIS method in some sequences.

To ensure a fair evaluation, we applied the same Δt =5 ms to all methods using the temporal window in the evaluation. That is, we output a motion field every 5 ms. We also found empirically that the denoising technique in (24) improved the performance of *all* DVS-based optical flow methods (not just ours). Thus, the results shown in Section 5.3 are shown with and without the same denoising for all optical flow methods.

The proposed DistSurf-OF method was implemented and run on MATLAB 2019b operating on Lenovo ThinkStation P520C. In our implementation, we fixed the temporal window Δt used for edge map in (4) and the predefined threshold parameter τ for denoising in (24) to 5ms. The spatial gradient filter used in (12) was the 4-point central difference (with mask coefficients $(\frac{1}{12}(-1, 8, 0, -8, 1))$ [14]. With these configurations, the execution time to yield a motion field for the 346×260 spatial resolution sensor was around 0.737s. However, the computation of operations specific to DistSurf-OF (event denoising, temporal windowing, distance transform, and spatial/temporal derivatives) take only 35 ms, while the (intensity-based) optical flow method in [16] takes 0.702s. Therefore, users have the freedom to choose different intensity-based optical flow methods to pair with DistSurf based on the accuracy and speed requirements of the application. The computation of DistSurf-OF itself can be sped up further by GPU-based parallel coding, as well as fast distance transform implementations (e.g., [55]) that reduce complexity from $\mathcal{O}(n^2)$ to $\mathcal{O}(n \log n)$ or even $\mathcal{O}(n)$. The code and the DVSMOTION20 dataset (with the accompanying ground truth motion field) are made available at issl.udayton.edu.

5.3 Results and Discussion

The performances of DistSurf-OF and the state-of-the-art DVS/DAVIS optical flow methods on sequences shown in Fig. 10 are reported in Table 1. For error statistics, we show average angular error (AAE) and the relative average endpoint error (RAEE), referring to the pixel motion magnitude and angle errors, respectively [33] and [46]. The stability of each method can be inferred from the standard deviations of the error statistics.

Among the DVS-based optical flow methods, DistSurf-OF has a clear advantage. Average angular error is consistently below 7° in all sequences, while the motion magnitude as evaluated by AEE is also the smallest. Contrast this to the other state-of-the-art DVS-based optical flow methods in [32], [35], [36], whose angular error exceeds 10° in all but simplistic motion sequence (*translation conference*). Method in [35] failed in three of the four sequences. One can also gauge the effectiveness of the denoising on the state-of-the-art methods by comparing the "DVS" and "DVS+denoising" data types in Table 1. In all but the *checkerboard* sequence, optical flow applied to IE+TE in (24) was more accurate than same methods applied to IE+TE+BA.

The performance of DistSurf-OF is closer to that of the state-of-the-art DAViS-OF method in [46]. Owing to the fact that the latter method leverages both DVS and APS data, it is indeed expected to perform better than DVS-only methods. With an angular error around 5°, there are about 2° only separating the performance of DistSurf-OF (DVS only) and DAViS-OF (DVS+APS).

The trends in Table 1 can be visually confirmed by the results in Fig. 10. DistSurf-OF in Fig. 10e yields stable and satisfactory results in terms of motion orientation and magnitude, closely resembling the ground truth motion field in Fig. 10a and DAViS-OF in Fig. 10b. In particular, the motion detected by the proposed method in *checkerboard* sequence is in the correct orientation, not normal to the edge direction. Contrast this to EV-FlowNet in Fig. 10d, whose estimated motion direction is predominantly horizontal (perpendicular to the vertical edges). LK-DVS in Fig. 10c is more accurate than EV-FlowNet in terms of motion orientation, but lacks spatial consistency.

The other sequences are rich with diverse edge orientations and edge length, making it possible to assess the

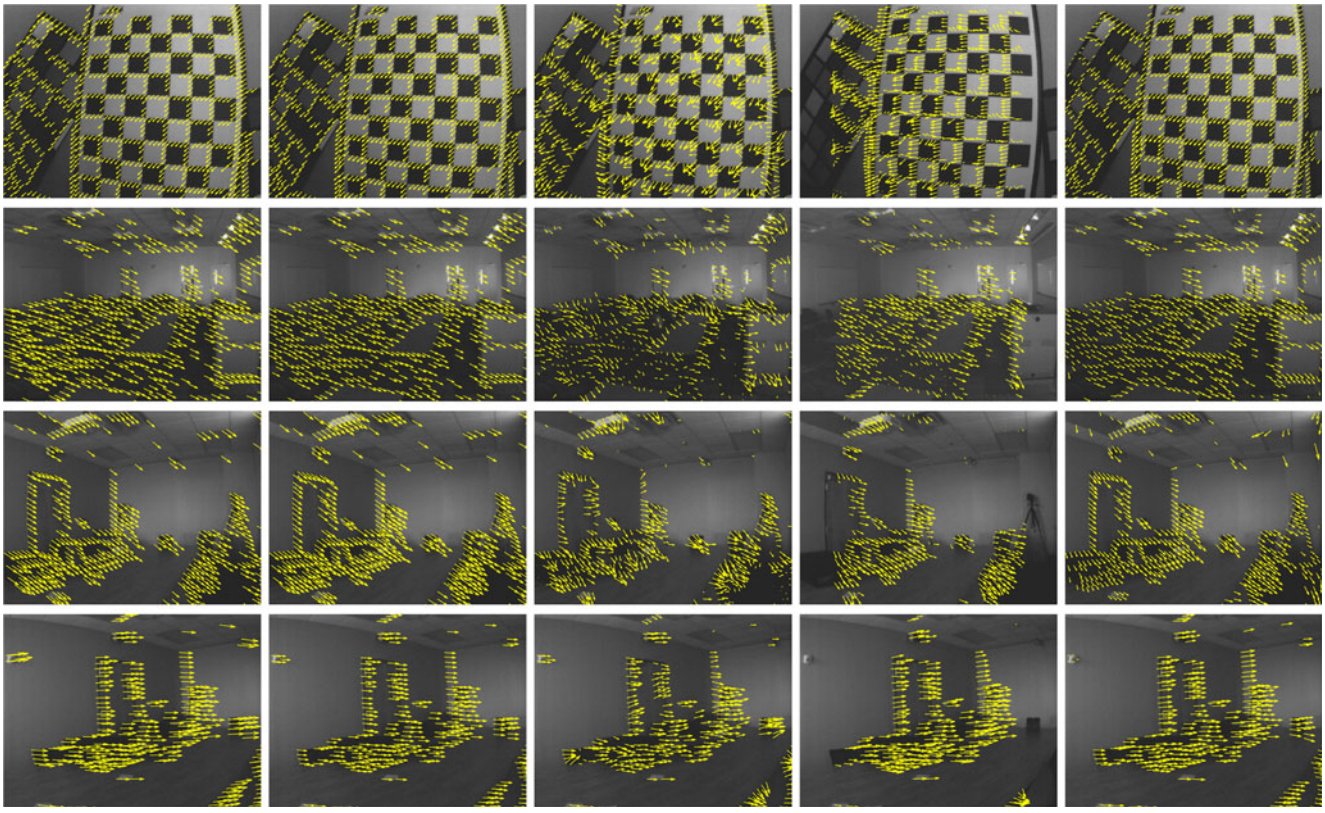


Fig. 10. Optical flow results on DVSMOTION20 sequences (shown with denoising). Arrow orientation and magnitude indicate the estimated pixel motion orientation and speed of the observed events. Ground truth is obtained from IMU. See Section 5.1. Motion is overlaid on APS for visualization.

optical flow method's robustness to real-world variations. DistSurf-OF handled them well, save for the events occurring very close to the boundaries of the sensor. The quality of the estimated motion field is comparable to that of the DAVIS-OF. EV-Flownet performed better in the *checkerboard* sequence, although the proposed method was still better when comparing the orientations of the pixel motions event-for-event to the ground truth motion. LK-DVS is unable to resolve the spatial inconsistency problem.

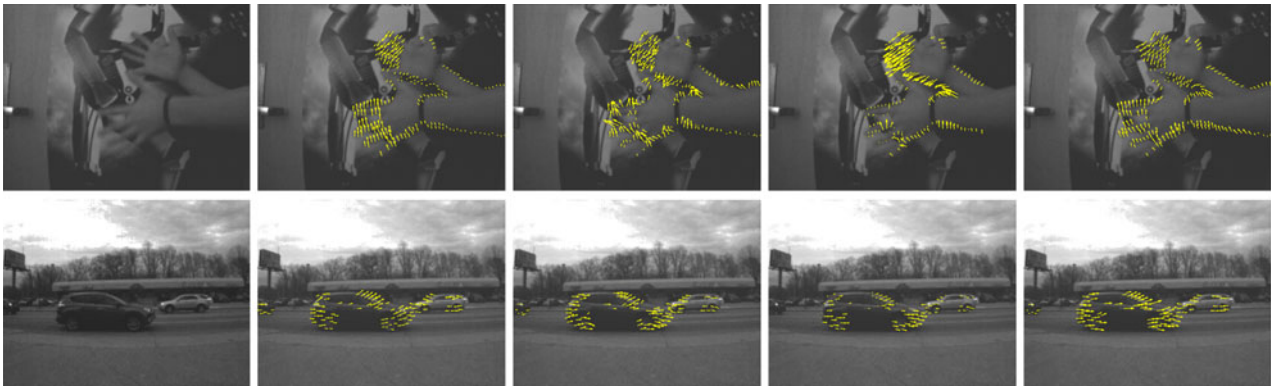
Finally, Fig. 11 shows more challenging sequences containing multiple objects moving in opposite directions. In the *hand* sequence where two hands cross in front of a textured background, DAVIS-OF and DistSurf-OF are able to track individual fingers and their corresponding directions accurately, and the motion boundaries where the hands and the arms cross each other is well defined. Despite the lack of spatial consistency, LK-DVS largely detects the orientation of the finger movement. The high concentration of events in the fingers seem to confuse the pixel motion estimation in EV-FlowNet, with inconsistent velocity magnitudes and orientations.

In the *car* sequence, the motion estimated by all methods on the windshield of the foreground car seem to point towards the sky erroneously—meaning all optical flow methods yielded normal flow. In the remainder of the foreground car and the background, DistSurf-OF's motions are horizontally oriented with consistent motion magnitudes. EV-FlowNet estimates of the motion orientation in the remainder of foreground car are more stable, but the motion magnitudes vary considerably; and the estimated background car motion is inconsistent with the context. The quality of LK-DVS is comparable to DistSurf-OF.

TABLE 1
Average Angular Error (AAE) and Relative Average end-Point Error (RAEE) With Standard Deviations

Data Type	Method+Enhancements	Checkerboard		Classroom		Conference Room		Conference Room Translation	
		RAEE [%]	AAE [°]	RAEE [%]	AAE [°]	RAEE [%]	AAE [°]	RAEE [%]	AAE [°]
DVS	LK-DVS [32] + [54]	39.30 ± 21.4	13.19 ± 14.40	53.42 ± 23.13	20.45 ± 20.68	47.39 ± 25.43	21.49 ± 26.41	48.21 ± 25.87	21.19 ± 28.88
	LP [35] + [54], [55]	121.66 ± 35.35	91.00 ± 31.03	110.76 ± 40.56	92.61 ± 50.24	117.45 ± 42.68	98.10 ± 50.40	117.10 ± 143.71	81.79 ± 47.76
	EV-FlowNet [43]	36.12 ± 23.67	15.86 ± 16.61	53.42 ± 45.92	18.77 ± 14.33	41.24 ± 31.49	18.37 ± 18.29	28.37 ± 25.00	10.63 ± 14.56
DVS + denoising	LK-DVS [32] + [54]	42.32 ± 20.53	12.06 ± 11.15	32.64 ± 17.29	11.66 ± 11.48	30.81 ± 16.91	11.87 ± 11.37	31.48 ± 16.81	10.97 ± 11.46
	LP [35] + [54], [55]	104.22 ± 17.06	72.07 ± 18.23	103.09 ± 31.79	82.96 ± 42.53	102.08 ± 28.67	76.35 ± 40.82	112.25 ± 111.23	78.23 ± 40.20
	EV-FlowNet [43]	38.56 ± 24.03	16.97 ± 16.17	44.62 ± 24.18	15.14 ± 16.77	38.52 ± 22.21	16.27 ± 15.92	23.62 ± 20.06	8.04 ± 11.82
	DistSurf-OF (proposed)	**20.54** ± 10.36	**5.89** ± 5.11	**25.62** ± 13.27	**6.84** ± 6.26	**18.71** ± 12.38	**5.67** ± 5.75	**18.71** ± 12.95	**5.42** ± 5.38
DAViS	DAViS-OF [46]	17.56 ± 8.18	3.80 ± 2.64	20.00 ± 6.60	5.01 ± 2.69	19.77 ± 6.32	5.48 ± 2.71	23.76 ± 5.74	3.77 ± 1.97

***Bold** font indicates the best DVS-based optical flow performance.*



(a) Frame (b) DAViS-OF [46] (c) LK-DVS [32] (d) EV-FlowNet [53] (e) DistSurf-OF

Fig. 11. Optical flow results on DVSMOTION20 sequences (shown with denoising) containing object motions and motion occlusions. (top row) Hand and (bottom row) car sequences.

6 Conclusion

We proposed the notion of *distance surface* for performing optical flow tasks in neuromorphic cameras. We proposed to use the distance transforms computed from the events generated from DVS as a proxy for object textures. We rigorously proved that distance surface satisfy optical flow equations, and the event pixel motion recovered by DistSurf-OF are highly accurate. We verified the effectiveness of our method using DVSMOTION20 dataset. For future work, we plan to investigate whether DistSurf can be combined with APS (similar in style to [46]) to further improve the optical flow accuracy.

Acknowledgments

This work was supported in part by Ford University Research Program and the Japan National Institute of Information and Communications Technology.

References

[1] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," *IEEE Trans. Robotics*, vol. 34, no. 6, pp. 1425–1440, Dec. 2018.

[2] G. Gallego, H. Rebecq, and D. Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3867–3876.

[3] G. Gallego, M. Gehrig, and D. Scaramuzza, "Focus is all you need: Loss functions for event-based vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12 280–12 289.

[4] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 989–997.

[5] T. Stoffregen and L. Kleeman, "Simultaneous optical flow and segmentation (sofas) using dynamic vision sensor," in *Proc. Australas. Conf. Robot. Autom.*, 2017, pp. 52–61.

[6] T. Stoffregen, G. Gallego, T. Drummond, L. Kleeman, and D. Scaramuzza, "Event-based motion segmentation by motion compensation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 7244–7253.

[7] P. Bardow, A. J. Davison, and S. Leutenegger, "Simultaneous optical flow and intensity estimation from an event camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 884–892.

[8] H. Kim, A. Handa, R. Benosman, S. Ieng, and A. Davison, "Simultaneous mosaicing and tracking with an event camera," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 1–12.

[9] X. Clady, J.-M. Maro, S. Barré, and R. B. Benosman, "A motion-based feature for event-based pattern recognition," *Front. Neurosci.*, vol. 10, 2017, Art. no. 594.

[10] D. Gehrig, H. Rebecq, G. Gallego, and D. Scaramuzza, "Asynchronous, photometric feature tracking using events and frames," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 750–765.

[11] H. G. Chen *et al.*, "Fast retinomorphic event-driven representations for video gameplay and action recognition," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 276–290, Oct. 2019.

[12] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.

[13] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artif. Intell.* -Vol. 2, 1981, pp. 674–679.

[14] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, 1994.

[15] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.

[16] D. Sun, S. Roth, and M. J. Black, "A quantitative analysis of current practices in optical flow estimation and the principles behind them," *Int. J. Comput. Vis.*, vol. 106, no. 2, pp. 115–137, 2014.

[17] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2432–2439.

[18] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 211–231, 2005.

[19] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.

[20] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 77–104, 1990.

[21] D. J. Fleet and A. D. Jepson, "Stability of phase information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 12, pp. 1253–1268, Dec. 1993.

[22] T. Gautama and M. M. Van Hulle, "A phase-based approach to the estimation of the optical flow field using spatial filtering," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1127–1136, Sep. 2002.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[24] T. Kroeger, R. Timofte, D. Dai, and L. Van Gool, "Fast optical flow using dense inverse search," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 471–488.

[25] L. Bao, Q. Yang, and H. Jin, "Fast edge-preserving patchmatch for large displacement optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3534–3541.

[26] E. Memin and P. Perez, "A multigrid approach for hierarchical motion estimation," in *Proc. 6th Int. Conf. Comput. Vis.*., 1998, pp. 933–938.

[27] E. Mémin and P. Pérez, "Hierarchical estimation and segmentation of dense motion fields," *Int. J. Comput. Vis.*, vol. 46, no. 2, pp. 129–155, 2002.

[28] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Comput. Vis. Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

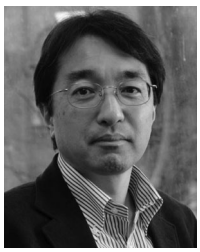
[29] D. Sun, C. Liu, and H. Pfister, "Local layering for joint motion estimation and occlusion detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1098–1105.

[30] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for TV-L 1 optical flow," in *Statistical and Geometrical Approaches to Visual mzotion Analysis*. Berlin, Germany: Springer, 2009, pp. 23–45.

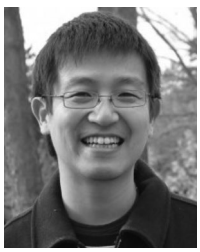
[31] A. Wedel, T. Pock, J. Braun, U. Franke, and D. Cremers, "Duality TV-L1 flow with fundamental matrix prior," in *Proc. IEEE 23rd Int. Conf. Image Vis. Comput. New Zealand*, 2008, pp. 1–6.
[32] R. Benosman, S.-H. Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan, "Asynchronous frameless event-based optical flow," *Neural Netw.*, vol. 27, pp. 32–37, 2012.
[33] B. Rueckauer and T. Delbruck, "Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor," *Fron. Neurosci.*, vol. 10, 2016, Art. no. 176.
[34] T. Brosch, S. Tschechne, and H. Neumann, "On event-based optical flow detection," *Front. Neurosci.*, vol. 9, 2015, Art. no. 137.
[35] R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, "Event-based visual flow," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 407–417, Feb. 2014.
[36] T. Delbruck, "Frame-free dynamic digital vision," in *Proc. Int. Symp. Secure-Life Electronics Adv. Electronics Quality Life Soc.*, 2008, pp. 21–26.
[37] J. H. Lee, P. K. Park, C.-W. Shin, H. Ryu, B. C. Kang, and T. Delbruck, "Touchless hand gesture ui with instantaneous responses," in *Proc. 19th IEEE Int. Conf. Image Process.*, 2012, pp. 1957–1960.
[38] M. Liu and T. Delbruck, "Block-matching optical flow for dynamic vision sensors: Algorithm and FPGA implementation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2017, pp. 1–4.
[39] M. Liu and T. Delbruck, "Abmof: A novel optical flow algorithm for dynamic vision sensors," 2018. *arXiv:1805.03988*
[40] J. Y. Jason, A. W. Harley, and K. G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 3–10.
[41] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2462–2470.
[42] S. Meister, J. Hur, and S. Roth, "Unflow: Unsupervised learning of optical flow with a bidirectional census loss," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 7251–7259.
[43] A. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Ev-flownet: Self-supervised optical flow estimation for event-based cameras," in *Proc. Robotics: Sci. Syst.*, 2018, doi: 10.15607/RSS.2018.XIV.062.
[44] C. Ye, A. Mitrokhin, C. Parameshwara, C. Fermüller, J. A. Yorke, and Y. Aloimonos, "Unsupervised learning of dense optical flow and depth from sparse event data," 2018. *arXiv:1809.08625*
[45] F. Paredes-Valles, K. Y. W. Scheper, and G. C. H. E. De Croon, "Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2019.2903179.
[46] M. M. Almatrafi and K. Hirakawa, "Davis camera optical flow," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 396–407, Oct. 2019.
[47] C. Brandli, R. Berner, M. Yang, S.-C. Liu, and T. Delbruck, "A 240 $\times$ 180 130 db 3 μs latency global shutter spatiotemporal vision sensor," *IEEE J. Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, Oct. 2014.
[48] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "Hots: A hierarchy of event-based time-surfaces for pattern recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1346–1359, Jul. 2017.
[49] I. Alzugaray and M. Chli, "Asynchronous corner detection and tracking for event cameras in real time," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3177–3184, Oct. 2018.
[50] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, "Hats: Histograms of averaged time surfaces for robust event-based object classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1731–1740.
[51] V. Padala, A. Basu, and G. Orchard, "A noise filtering algorithm for event-based asynchronous change detection image sensors on truenorth and its implementation on truenorth," *Front. Neurosci.*, vol. 12, 2018, Art. no. 118.
[52] R. W. Baldwin, M. Almatrafi, J. R. Kaufman, V. Asari, and K. Hirakawa, "Inceptive event time-surfaces for object classification using neuromorphic cameras," in *Proc. Int. Conf. Image Anal. Recognit.*, 2019, pp. 395–403.
[53] B. Rueckauer and T. Delbruck, "Evaluation of event-based algorithms for optical flow with ground-truth from inertial measurement sensor," 2016. [Online]. Available: http://sensors.ini.uzh.ch/news_page/ds-flow-2016.html
[54] F. Barranco, C. Fermüller, and Y. Aloimonos, "Contour motion estimation for asynchronous event-driven cameras," *Proc. IEEE*, vol. 102, no. 10, pp. 1537–1556, Oct. 2014.
[55] C. Karam, K. Sugimoto, and K. Hirakawa, "Fast convolutional distance transform," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 853–857, Jun. 2019.

Mohammed Almatrafi (Member, IEEE) received the BS (Hons.) degree in electrical engineering from Umm Al-Qura University, Makkah, Saudi Arabia, in 2011, the MS and PhD degrees in electrical and computer engineering from the University of Dayton, Dayton, OH, USA, in 2015 and 2019, respectively. He is currently an assistant professor and the vice dean of the College of Engineering for Postgraduate Studies and Scientific Research at Umm Al-Qura University, Al-Lith, Saudi Arabia. His research is focused on image processing, neuromorphic cameras, and computer vision.

Wes Baldwin received the BS degree in computer engineering from Kettering University, in 2002, and the MS degree in electrical and computer engineering, in 2005 from the University of Illinois at Chicago. He is currently working towards the PhD degree in electrical and computer engineering at the University of Dayton. His research is focused on machine learning, image processing, and neuromorphic cameras.



Kiyoharu Aizawa (Fellow, IEEE) received the BE, ME, and Dr. Eng. degrees in electrical engineering from the University of Tokyo, in 1983, 1985, and 1988, respectively. He is currently a professor with the Department of Information and Communication Engineering of the University of Tokyo. He was a visiting assistant professor at the University of Illinois from 1990 to 1992. He received the 1987 Young Engineer Award and the 1990, 1998 Best Paper Awards, the 1991 Achievement Award, 1999 Electronics Society Award from IEICE Japan, and the 1998 Fujio Frontier Award, the 2002 and 2009 Best Paper Award, and 2013 Achievement Award from ITE Japan. He received the IBM Japan Science Prize in 2002. He is on Editorial Boards of *IEEE MultiMedia*, *ACM TOMM*, *APSIPA Transactions on Signal and Information Processing*, and *International Journal of Multimedia Information Retrieval*. He served as the editor in chief of *Journal of ITE Japan*, an associate editor of *IEEE Trans. Image Processing*, *IEEE Trans. CSVT* and *IEEE Trans. Multimedia*. He was a president of ITE and ISS society of IEICE, 2019 and 2018, respectively. He has served a number of international and domestic conferences; he was a general co-chair of ACM Multimedia 2012 and ACM ICMR2018. He is a fellow of IEICE, ITE and a council member of Science Council of Japan. His research interest is in multimedia applications, image processing, and computer vision



Keigo Hirakawa (Senior Member, IEEE) received the BS degree (Hons.) in electrical engineering from Princeton University, Princeton, NJ, USA, in 2000, the MS and PhD degrees in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2003 and 2005, respectively, and the MM degree (Hons.) in jazz performance studies from the New England Conservatory of Music, Boston, MA, USA, in 2006. He was a research associate with Harvard University, Cambridge, MA, USA from 2006 to 2009. He is currently an associate professor with the University of Dayton, Dayton, OH, USA. He is currently the head of the Intelligent Signal Systems Laboratory, University of Dayton, where the group focuses on statistical signal processing, color image processing, and computer vision.