Event-based Shape from Polarization

Manasi Muglikar [1]   Leonard Bauersfeld [1]   Diederik Paul Moeys [2]   Davide Scaramuzza [1]

[1]Robotics and Perception Group, University of Zurich, Switzerland
[2]Advanced Sensors and Modelling Group, SONY R&D Center Europe, SL1

Abstract

State-of-the-art solutions for Shape-from-Polarization (SfP) suffer from a speed-resolution tradeoff: they either sacrifice the number of polarization angles measured or necessitate lengthy acquisition times due to framerate constraints, thus compromising either accuracy or latency. We tackle this tradeoff using event cameras. Event cameras operate at microseconds resolution with negligible motion blur, and output a continuous stream of events that precisely measures how light changes over time asynchronously. We propose a setup that consists of a linear polarizer rotating at high speeds in front of an event camera. Our method uses the continuous event stream caused by the rotation to reconstruct relative intensities at multiple polarizer angles. Experiments demonstrate that our method outperforms physics-based baselines using frames, reducing the MAE by 25% in synthetic and real-world datasets. In the real world, we observe, however, that the challenging conditions (i.e., when few events are generated) harm the performance of physics-based solutions. To overcome this, we propose a learning-based approach that learns to estimate surface normals even at low event-rates, improving the physics-based approach by 52% on the real world dataset. The proposed system achieves an acquisition speed equivalent to 50 fps (>twice the framerate of the commercial polarization sensor) while retaining the spatial resolution of 1MP. Our evaluation is based on the first large-scale dataset for event-based SfP.

Code, dataset and video are available under:
https://rpg.ifi.uzh.ch/esfp.html
https://youtu.be/sF3Ue2Zkpec

1. Introduction

Polarization cues have been used in many applications across computer vision, including image dehazing [41], panorama stitching and mosaicing [42], reflection removal [21], image segmentation [25], optical flow gyroscope, [47] and material classification [5]. Among these, Shape-from-Polarization (SfP) methods exploit changes in polariza-
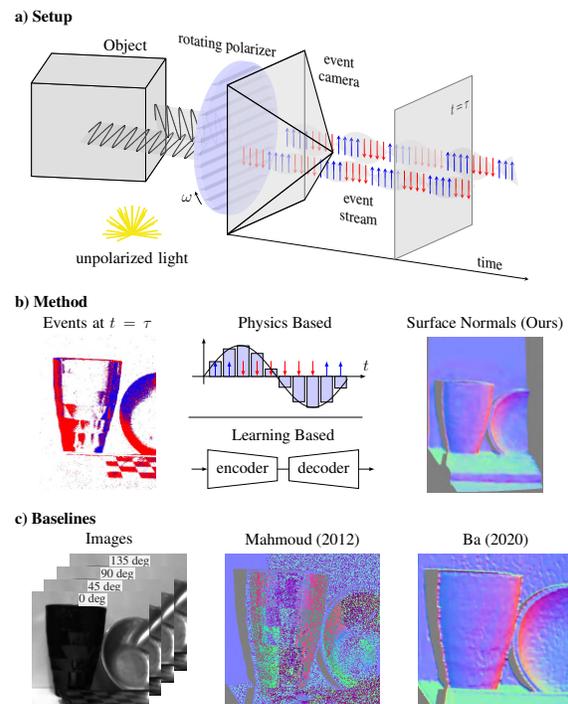
Figure 1. Surface normal estimation using event-based SfP. (a) Rotating a polarizer in front of an event camera creates sinosoidal changes in intensities, triggering events. (b) The proposed event-based method uses the continuous event stream to reconstruct relative intensities at multiple polarizer angles which is used to estimate surface normals using physics-based and learning-based method. (c) Our approach outperforms image-based baselines [24, 51].

tion information to infer geometric properties of an object [2, 18, 22, 49, 51]. It uses variations in radiance under different polarizer angles to estimate the 3D surface of a given object. In particular, when unpolarized light is reflected from a surface, it becomes partially polarized depending on the geometry and material of the surface. Surface normals, and thus 3D shape, can then be estimated by orienting a polarizing filter in front of a camera sensor and studying the relationship between the polarizer angle and the magnitude of light transmission. SfP has a number of advantages over both active and passive depth sensing methods. Unlike ac-

(a) Division of focal plane

(b) Division of time
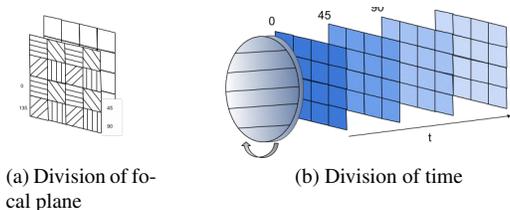
Figure 2. Illustration of SfP methods.

Dataset	Modality	(Resolution)	Size
Polar3D [18]	6 Images(DoT)	18 MP	3
DeepSfP [51]	4 Images(DoFP)	1224×1024	236
SPW [22]	4 Images(DoFP)	1224×1024	522
ESfP- Synthetic (Ours)	Events (DoT) + 12 Images(DoT)	512×512	104
ESfP- Real (Ours)	Events (DoT) + 4 Images (DoFP)	1280×720	90

Table 1. Summary of publicly available datasets for SfP.

tive depth sensors that use structured light (SL) [7, 48] or time-of-flight (ToF), SfP is not limited by material type and can be applied to non-Lambertian surfaces like transparent glass and reflective, metallic surfaces.

Despite these advantages, however, estimating high-quality surface normals from polarization images is still an open challenge. *Division of Focal Plane (DoFP)* methods [22, 32, 33, 51] trade-off spatial resolution for latency and allow for the capture of four polarizations in the same image. This is achieved through a complex manufacturing process that requires precisely placing a micro-array of four polarization filters on the image sensor [32, 33], as shown in Fig. 2a Despite the reduced latency, this system constrains the maximum number of polarization angles that can be captured, potentially impacting the accuracy of the estimates as we show in our results.Additionally, the spatial resolution of the sensor is also reduced, requiring further mosaicing-based algorithms for high-resolution reconstruction [50]. On the other hand, *Division of Time (DoT)* methods [2, 18, 49] provide full-resolution images and are not limited in the number of polarization angles they can capture thanks to a rotating polarizing filter put in front of the image sensor. The frame rate of the sensing camera, however, effectively limits the rate at which the filter can rotate, increasing the acquisition time significantly (acquisition time $= N/f$, where N is the number of polarizer angles and f is the framerate of the camera). For this reason, commercial solutions, such as the Lucid Polarisens [33], favor DoFP, despite the lower resolution of both polarization angles and image pixels. To overcome this shortcoming, recently, significant progress has been made with data-driven priors [22, 51]. However, these solutions still fall short in terms of computational complexity when compared to DoT methods. A solution able to bridge the accuracy of DoT with the speed of DoFP is thus still lacking in the field.

In this paper, we tackle the speed-resolution trade-off using event cameras. Event cameras are efficient high-speed vision sensors that asynchronously measure changes in brightness intensity with microsecond resolution. We exploit these characteristics to design a DoT approach able to operate at high acquisition speeds (up to $5,000$ fps vs. 22 fps of standard frame-based devices) and full-resolution (1280×720) s shown in Fig. 1. Thanks to the working principles of event-cameras, our sensing device provides a continuous stream of information for estimating the sur-

face normal as compared to the discrete intensities captured at fixed polarization angles of traditional approaches. We present two algorithms to estimate surface normals from events, one using geometry and the other based on a learning-based approach. Our geometry-based method takes advantage of the continuous event stream to reconstruct relative intensities at multiple polarizer angles, which are then used to estimate the surface normal using traditional methods. Since events provide a temporally rich information, this results in better reconstruction of intermediate intensities. This leads to an improvement of upto 25% in surface normal estimation, both on the synthetic dataset and on the real-world dataset.On the real dataset, however, the non-idealities of the event camera introduce a lower fill-rate (percentage of pixels triggering events) of 3.6% in average (refer Section 3.1). To overcome this, we propose a deep learning framework which uses a simple U-Net network to predict the dense surface normals from events. Our data-driven approach improves the accuracy over the geometry-based method by 52%. Our contributions can be summarized as follows:

- A novel approach for shape-from-polarization using an event camera. Our approach utilizes the rich temporal information of events to reconstruct event intensities at multiple polarization angles. These event intensities are then used to estimate the surface normal. Our method outperforms previous state-of-the-art physics-based approaches using images by 25% in terms of accuracy.

- A learning-based framework which predicts surface normals using events to solve the issue of low fill-rate common in the real-world. This framework improves the estimation over physics-based approach by 52% in terms of angular error.

- Lastly, we present the *first* large scale dataset containing over 90 challenging scenes for SfP with events and images. Our dataset consists of events captured by rotating a polarizer in front of an event camera, as well as images captured using the Lucid Polarisens [33].

2. Related work

Frame-based SfP methods Shape-from-polarization estimates the normal of each point on an object's surface through Fresnel equations [8] by measuring the azimuthal and zenithal angles at each pixel. However, since a linear polarizer can only distinguish polarized light modulus

2π, shape-from-polarization is often regarded as an underdetermined problem and additional constraints are typically required to solve ambiguities in the measurements.

Prior SfP systems that addressed the problem with traditional cameras are thoroughly discussed in [11]. Early methods relied on assumptions about object surface and environment lighting to constrain the problem, such as pure specular [37] or pure diffusion reflections [1] and surface convexity [4, 15], but they typically fail when the theoretical model is violated. In order to ensure uniqueness in the solution, several methods exploit depth and geometric information to guide the surface reconstruction operation. A line of research [3, 26–28] measures polarization information from multiple view points, while other authors exploit coarse depth maps from low-cost depth sensors, such as Kinect [19] or RGBD cameras [53], to obtain additional geometric cues. In single-view, spectral information measured by multi-band cameras can also be used to solve the π-ambiguity, as well as estimating refractive properties of the material useful for 3D reconstruction [15, 15, 44]. Another class of methods considers photometric information to disambiguate SfP normal estimates by combining photometric stereo [1, 10, 31] and photometric constraints [24] to impose shading constraints from multiple light directions.

More recently, [22, 51] presented a data-driven approach to solve the ambiguity, resulting in state-of-the-art performance. Both these approaches, however, use a Polarisens sensor [33] which compromises spatial resolution for faster acquisition time, thus limiting the potential of the learning-based approach. We propose to address the shortcoming arsing form the spatial-temporal resolution trade-off using an event camera and take advantage of deep learning methods to tackle non-idealities of the event camera.

Event-based SfP methods Since event cameras are novel sensors, there are very few papers that combine polarization information with events. Recently, [14] combined the DoFP approach with an event camera by manufacturing a micro array of polarizers and placing it on top of the event camera sensor. This requires a very precise manufacturing process in which the polarizers are manually aligned with the pixels by inducing polarized motion for the pixels to see. Due to the low yield resulting from this process and to low-spatial resolution issues analogous to standard cameras, such as mosaicing, commercial solution are currently not available. Moreover, this approach also requires relative motion between the camera and the scene to capture event information, which further limits its application. In contrast, our approach, making use of a rotating polarizing filter, exploits the full spatial resolution of the event camera and does not require camera motion. Since event cameras have a high temporal resolution, this approach enables us to rotate the filter at very high speeds (upto 1500 RPM), thus enabling high-speed and high spatial resolution SfP.

SfP Datasets Prior SfP datasets are summarized in Table 1. The datasets were collected using standard cameras with either *DoT* [18] or *DoFP* [22, 51]. Kadambi *et al.* [18] presented a small dataset which was collected using Canon Rebel T3i DSLR camera and a linear polarizer and images were collected at 6 discrete polarizer angles. The advent of the Lucid Polarsens [33] camera enabled the generation of large scale dataset proposed by Ba *et al.* [51] and Lei *et al.* [22]. These datasets capture only 4 polarization angles as is the case with *DoFP*-based methods, and do not contain events. To push the limits of event-based SfP, in this paper, we propose two datasets, synthetic and real-world, which contain events, images and accurate groundtruth in challenging scenarios.

3. Event-based Shape from Polarization

Event-cameras are novel, bio-inspired sensors that asynchronously measure *changes* (i.e., temporal contrast) in illumination at every pixel, at the time they occur [12, 23, 35, 45]. In particular, an event camera generates an event $e_k = (\mathbf{x}_k, t_k, p_k)$ at time t_k when the difference of logarithmic brightness at the same pixel $\mathbf{x}_k = (x_k, y_k)^\top$ reaches a predefined threshold C:

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

where $p_k \in \{-1, +1\}$ is the sign (or polarity) of the brightness change, and Δt_k is the time since the last event at the pixel $\mathbf{x}_k$. The result is a sparse sequence of events which are asynchronously triggered by illumination changes.

We consider the task of surface normal estimation using a polarizer and an event camera, as illustrated in Fig. 1. The polarizer, rotating in front of the event camera at speed ω using a motor, changes the illumination of the incoming light. Given the orientation ϕ_{pol} of the polarizer, we can express the light intensity $I(\phi_{\text{pol}})$ passing through it as:

$$I(\phi_{\text{pol}}) = I_{\text{un}}(1 + \rho \cdot \cos(2(\phi_{\text{pol}} - \phi)), \qquad (2)$$

where I_{un} is the unpolarized intensity, ρ is the degree of polarization and ϕ is the angle of polarization of the surface. When rotating the polarizer, the sinusoidal intensity changes trigger events in the event camera. For instance, when observing two surfaces which are perpendicular to each other with our setup, the corresponding events have opposite polarity, as shown in Fig. 1. This is because the underlying sinusoids are phase shifted by $90°$.

We take advantage of the close relationship between surface orientation and event generation to develop a principled procedure for estimating surface normals. By continuously capturing how light is affected by the polarizer as it rotates, events offer increased information than what is provided by regular frame-based devices.

Based on these considerations, we start in Section 3.1 by describing how events can directly be used to estimate

the normals in a physics-based fashion. This approach, however, assumes events are generated following an ideal model, which might limit its performance when edge-cases are met in the real world due to nonidealities. Leveraging these findings, we also propose a deep learning method that improves predictions in Section 3.2.

3.1. Physics-based ESfP

Over one rotation of the polarizer, the intensity at each pixel $\mathbf{x}_p$ follows a sine wave, as depicted in Fig. 1. By combining this behavior with the event generation model in Eq. (1), we can estimate the intensity $I_e(\mathbf{x}_p)$ at any time τ as:

$$L(\tau) = L(\tau - \delta t) + p_t C = L(0) + \sum_{t=0}^{\tau} p_t C \quad (3)$$

$$I_e(\tau) = \exp\left(L(0)\right) \cdot \exp\left(\sum_{t=0}^{\tau} p_t C\right), \quad (4)$$

where we dropped the pixel location x_p from the logarithmic brightness for readability.

Since events only capture relative intensity changes and not absolute ones, there is no way to recover the constant $e^{L(0)}$ without direct intensity measurements. However, we show below that for the computation of the surface normal, the knowledge of absolute intensity is not required, and we can thus avoid estimating $e^{L(0)}$.

Since $\phi_{pol} = \omega \cdot t$, we can use the above formulation to reconstruct "event-intensities" (I_e) *at multiple angles ϕ_{pol}*, rather than time instants. These "event-intensities" are then used to estimate the surface normal using traditional algorithms. Below, we show the estimation of ρ and ϕ using the event intensities computed at 4 polarizer angles as an example, where we change the notation from from $I_e(\cdot)$ to $I_e[\cdot]$ to highlight the change in variable from time to angles. Notice that, in practice, we estimate event intensities at 12 polarizer angles. Nonetheless, in both cases, we can demonstrate that estimating surface normal does not require the knowledge of absolute intensity. The SfP equations [18, 22, 51] can be substituted with our event intensities as follows:

$$\phi = 0.5 \cdot \arctan \frac{I_e[\pi/4] - I_e[3\pi/4]}{I_e[0] - I_e[\pi/2]} \quad (5)$$

$$= 0.5 \cdot \arctan \frac{e^{\sum_{t=0}^{\pi/4\omega} p_t C - \sum_{t=0}^{3\pi/4\omega} p_t C}}{-\sum_{t=0}^{\pi/2\omega} p_t C} \quad (6)$$

Similarly, we can estimate ρ as,

$$\rho = 0.5 \cdot \arctan \frac{\sqrt{(I_e[\pi/2]^2 + (I_e[\pi/4]^2 - I_e[3\pi/4]^2)}}{I_e[\pi/2] + I_e[\pi/4] + I_e[3\pi/4]}, \quad (7)$$

$$= 0.5 \cdot \arctan \frac{\sqrt{e^{2\sum_{t=0}^{\pi/2\omega} p_t C} + \left(e^{\sum_{t=0}^{\pi/4\omega} p_t C} - e^{\sum_{t=0}^{3\pi/4\omega} p_t C}\right)^2}}{e^{\sum_{t=0}^{\pi/2\omega} p_t C} + e^{\sum_{t=0}^{\pi/4\omega} p_t C} + e^{\sum_{t=0}^{3\pi/4\omega} p_t C}}$$

Using the above ρ and ϕ, the azimuth (α) and zenith angle (θ) of the surface normal can be estimated. We refer the reader to the supplementary materials for a complete derivation, which takes into account all the 12 angles we consider in our approach and the estimation of surface normal from ρ and θ. Since we reconstruct event intensities, any traditional algorithm developed for frame-based approaches can be utilized along with our approach. This opens the door to build upon the past research and develop better phyiscs-based algorithms from events.

Event non-idealities An event camera, like any physical sensor, is affected by noise caused by nonidealities of its individual electric circuits and the manufacturing process. The most important parameter that affects the noise generation model is the contrast sensitivity threshold C. If C is too high, this results in less number events triggered, thus increasing the signal to noise ratio; however, if it C is set too low, a higher number of noisy events are generated. In our application, it is favourable to have a low contrast threshold, in the order of 1-5% as the polarization signal is often weak (especially in low light conditions). Commercial available event sensors [36] however, are manufactured for a general purpose use of event cameras, and therefore have a very high threshold, typically ranging between 30% and 50%[1]. This directly affects the number of events that are triggered (or the fill-rate), resulting in a sparse estimation. Since the polarization signal is especially weak for front-parallel surfaces in real world scenes, we find that these surfaces rarely trigger any event. To address this sensor issue, we propose to tackle the problem using a data-driven approach.

3.2. Learning-based ESfP

We now describe our approach to solve the problem of shape-from-polarization using deep learning. As described above, events can be used to reconstruct event intensities at multiple polarization angles. Inspired by the frame-based solution of Ba *et al.* [22, 51], we propose to tackle the problem of estimating the surface normals using a U-Net [40] architecture. The network predicts the surface normal $\mathbf{N}$ as a 3 channel tensor $\mathbf{N} = (\sin\theta\cos\alpha, \sin\theta\sin\alpha, \cos\theta)$, where θ and α are the zenith and azimuth angle respectively.

To adopt this network to event data, we need to convert the sparse event stream into a frame-like event representation that is compatible with the CNN layers. A straightforward approach, which demonstrated remarkable performance on several tasks [13, 17, 30, 39], is to use the voxel grid [52] representation as input to the encoder network. This representation splits the temporal domain into bins, and aggregates events in each pixel as the sum of their polarity. However, doing so, empirically results in a higher mean

---

[1]The contrast threshold can be quantified as the percentage of intensity change with respect to the previous intensity value.

absolute error, most likely due to events with opposite polarity in the same bin nullifying each other, and thus resulting in an empty feature. Please refer the supplementary material for more details. We therefore design a novel event representation specifically suited for the SfP task. In Equation. 3, we observe that the accumulated sum of event polarities is directly related to the image intensity at a given time instant. Based on this consideration, we propose the *Cumulative Voxel Grid Representation* (CVGR), a variation of the voxel grid [52] that preserves polarity information even across long time intervals. Similar to previous works on learning with events [30, 39, 46], we first build a voxel grid [52] by partitioning the time domain into $b = 0, ..., B - 1$ equally-sized temporal windows, each aggregating the sequence of events $\mathcal{E}_b = \left\{ e_k \mid t_k \in \left[\frac{b \cdot T}{B}, \frac{(b+1) \cdot T}{B} \right] \right\}$. We then compute the cumulative sum over the bins and multiply this quantity with the contrast threshold.

$$E(x, y, b) = \sum_{i=0}^{i=b} C \cdot V(x, y, i) = \sum_{i=0}^{i=b} C \Big(\sum_{\substack{e_k \in \mathcal{E}_i: \\ x_k = x, y_k = y}} p_k \Big). \quad (8)$$

This final representation E is used as an input to the network. In our experiments, we use a voxel grid with 8 bins, thus our CVGR has the dimension $H \times W \times 8$.

This novel event-representations is then fed to a 8-level U-Net encoder-decoder architecture which densely regresses the normals. We apply cosine similarity loss function on the unit normalized predictions for training. We show this simple architecture is enough to recover high-quality normal vectors, even in locations where no event has been triggered, thus directly addressing the weakness of purely model-based solutions.

4. Datasets

To the best of our knowledge, there exists no dataset on which the proposed approach can be evaluated. We, therefore, first evaluate our approach on a synthetic dataset, rendered using an accurate polarimetric renderer [16] to perform controlled experiments. To evaluate our approach in the real world, we also build a prototype of the proposed setup using an event camera and a polarizing filter. We provide more details in the next sections.

4.1. ESfP-Synthetic Dataset

We use the Mitsuba renderer [16] to render a scene consisting of a textured mesh illuminated with a point light source. A polarizer lens was rotated in front of the camera between 0 and 180 degrees with 15 degrees intervals, resulting in 12 rendered images. Images thus collected were then used to simulate events using ESIM [38] with a contrast sensitivity threshold of 5%. The dataset consists of 89 training and 15 test sequences, each containing the images captured with the polarizer, together with the events generated from these, as well as the groundtruth surface normal

Figure 3. Hardware setup of event camera coupled with a rotating linear polarizer.

provided by the renderer. Meshes used from data generation were obtained from the *Google Scanned Objects* dataset [9] and the textures derived from [6]. More details are provided in the supplementary material. To increase data variability, the position of the light source was randomized and the mesh rotated for each scene.

4.2. ESfP-Real Dataset

To evaluate our approach in the real world, we built a prototype using a Prophesee Gen 4 event camera [12] and a *Breakthrough Photography X4 CPL* [34] linear polarizer with a quarter-wave plate, commonly referred to as a circular polarizer (CPL). The prototype is shown in Fig. 3. The polarizer can rotate at speeds of up to 1500 rpm using a brushless DC motor. The groundtruth is generated using Event-based Structured Light (ESL) [29], which consists of a laser point projector combined with an event camera. It provides accurate groundtruth thanks to laser scanner, while also having a small acquisition time (16 ms). Another advantage of using ESL is that the groundtruth is always in the frame of the event camera and therefore no additional alignment is required.

We use this setup to collect the *first* large scale dataset consisting of several objects with different textures and shapes, and featuring multiple illumination and scene depths, for a total of 90 scenes. Additionally, we use a Lucid Polarisens camera [33] to collect polarization images of the same scene. The images are obtained at 4 polarization angles $\{0, 45, 90, 135\}$ with a resolution of 1224×1220. We rectify the images and aligned them with the event camera to enable fair comparison with available image-based only methods. The dataset was recorded with the rotation speeds of 150RPM with illumination of 200lux.

5. Experiments

This section evaluates the performance of our event-based SfP system for surface normal estimation. We begin by introducing the baselines and performance metrics which will be used for evaluation on our dataset. Then, we perform experiments on the ESfP-Synthetic dataset to quantify the accuracy of our proposed physics-based and

Image	Events	[24]	[43]	Ours	GT
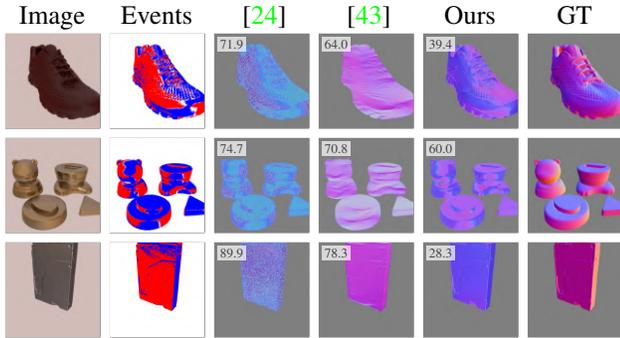
Figure 4. Comparison of physics-based methods on ESfP-Synthetic. We report the angular error on the top-left corner of each picture.

Method	Input	Task	Angular Error ↓ Mean	Accuracy ↑ AE<11.25	AE<22.5	AE<30
Mahmoud *et al.* [24]	4 I	Physics-based	80.923	0.034	0.065	0.085
Smith *et al.* [43]	4 I	Physics-based	67.684	0.010	0.047	0.106
Smith *et al.* [43]	12 I	Physics-based	62.476	0.007	0.043	0.097
Ours (P)	E	Physics-based	58.196	0.007	0.046	0.095
Ba *et al.* [51]	4 I	Learning-based	**24.509**	**0.357**	**0.623**	**0.718**
Ours (L)	E	Learning-based	27.953	0.263	0.527	0.655

Table 2. State-of-the-art comparison on ESfP-Synthetic in terms of accuracy and angular error. The *Input* column report whether the method uses events (E) or images (I), in which case the number of frames is also reported. We underline best results among the physics-based methods, and use bold text for learning-based ones.

learning-based methods. Finally, we evaluate these approach on our ESfP-Real dataset and study the proposed approach on real scenes under challenging conditions.

Implementation details We implement our physics-based approach in Python. As noted in Section 3.1, we reconstruct event intensities at 12 polarizer angles. Our physics-based approach does not require any hyperparameter tuning. The proposed learning-based approach was implemented in PyTorch. We train our network for 1000 epochs with a batch size of 4 on NVIDIA Tesla V100. We used a learning rate of $1e-4$ and Adam [20] as optimizer.

Evaluation Metrics We use 4 metrics to quantify the predicted surface normals, namely *Mean Angular Error* (MAE), $\%$ *Angular Error under* $11.25°$ (AE<11.25), $\%$ *Angular Error under* $22.5°$ (AE<22.5) and $\%$ *Angular Error under* $30°$ (AE<30). The first is a widely used metric for computing the angular error of the predicted surface normal (lower is better) [22, 51], while the last three, which we also refer to as angular accuracy, measure the percentage of pixels under $11.25°$, $22.5°$ and $30°$ of angular error (higher is better). We introduce another metric called *fill-rate* to compare events and images. The fill-rate measures the percentage of pixels that are triggered by SfP, i.e., by the polarizing filter's rotation. For image-based approaches, this always equals to 1, since images capture intensity for every pixel. However, for events this becomes an important parameter, since the number of triggered pixels depends on the contrast sensitivity of the camera as explained in Section 3.1.
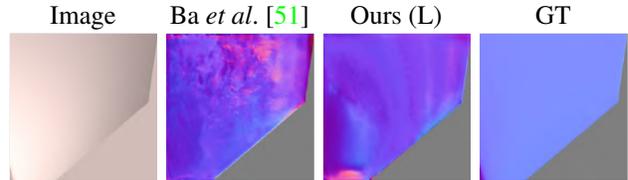
Figure 5. Comparison of our learning-based method against image-counterpart on ESfP-Synthetic scene with low-contrast.

Baselines We evaluate the proposed methods against frame-based SfP solutions. As we propose to tackle the problem both in a model-based and a data-driven manner, we include state-of-the-art methods from both categories as baselines in our evaluation. Smith *et al.* [43] and Mahmoud *et al.* [24] recover the surface normal using physics-based SfP methods[2], whereas Ba *et al.* [51] uses a learning-based SfP method. We re-implemented the last approach and retrained it on our datasets using only the images as input.

Note, while Lei *et al.* [22] improved upon the works of Ba *et al.*, by introducing view-encoding as an input to the network to generalize better to natural scenes, the goal of this paper is to show the performance of our learning-based framework on object-level scenes used in Ba *et al.*. We therefore, do not compare our approach against Lei *et al.*.

The event-based SfP solution proposed in [14] is not commercially available and requires specific manufacturing in order to be replicated. Moreover, this sensor requires relative motion between the scene and the camera and would not result in any surface normal estimation in our experiments. We, therefore, do not include comparison to this method in our experiments. In the following sections, we will first evaluate our methods in a simulated setting where precise ground truth allows for an accurate evaluation, and then proceed by studying their performance on the real world dataset we collected.

5.1. Results on ESfP-Synthetic

Table 2 shows the performance of our event-based approaches (both geometry-based and learning-based SfP) on the synthetic dataset. As it can be observed, frames-based solutions leveraging a physics-based approach are limited by the number of polarization images (and thus angles) which can be used for inference. When the same approach is used, i.e., in the case of Smith *et al.* [43], simply increasing the number of images from 4 to 12 decreases the angular error by nearly 8%. This gain in performance, however, comes at the cost of either spatial resolution or acquisition speed. In the case of a DoT approach, for instance, because the acquisition time is linearly proportional to the number of frames acquired, increasing the angles from 4 to 12 triples the acquisition time at a given frame-rate.

On the other hand, our approach is not limited by

---

[2]https://github.com/waps101/depth-from-polarisation
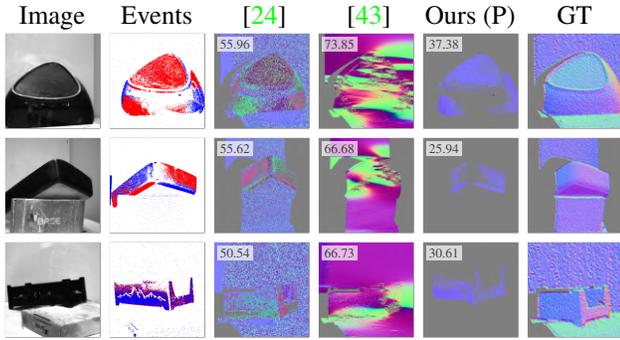
Figure 6. Comparison of physics-based methods using images [24] and [43] on the ESfP-Real. Our physics-based method results in sparse estimation as explained in Section 3.1

Method	Input	Angular Error ↓	Accuracy ↑		
		Mean	AE<11.25	AE<22.5	AE<30
Smith *et al.* [43]	I	72.525	0.009	0.034	0.058
Mahmoud *et al.* [24]	I	56.278	0.032	0.091	0.163
Ours (P)	E	<u>38.786</u>	<u>0.087</u>	<u>0.22</u>	<u>0.452</u>
Ba *et al.* [51]	I	**26.157**	**0.103**	**0.467**	**0.710**
Ours (L)	E	26.677	0.105	0.452	0.691

Table 3. State-of-the-art comparison on ESfP-Real in terms of angular error and accuracy. We use underlined text to mark the best results within the physics-based category, and use bold text for best results among learning-based ones.

this tradeoff thanks to the high-temporal resolution of the event camera. Exploiting the continuous stream of events acquired when observing the light being polarized, our geometry-based approach outperforms other physics-based methods using 4 images by 14% in angular error. In Fig. 4, we show qualitative results of our approach against the baselines. While our learning-based approach boosts the performance compared to physics-based methods significantly, it falls short of the image-based counterpart. The reason is because the learning-based method tries to hallucinate the surface normal prediction in the absence of events, whereas the image counterpart still has a dense spatial information to guide the network to interpolate the surface normals. In future, it would be worth investigating a combination of events and images to improve the performance.

5.2. Results on ESfP-Real

We also compare these methods on the real dataset in Table 3 and report qualitative results in Fig. 6. We observe similar conclusions as with the synthetic dataset. The proposed geometry-based approach outperforms previous physics-based solutions by 25.8% in angular error.As mentioned in Section 3.1, a real event camera introduces several non-idealities, one of them being the contrast sensitivity. With a typical event camera the contrast sensitivity cannot be usually set lower than 30%, resulting in the inability to capture small intensity changes. An example of this phenomenon is given in Fig. 7, where the presence
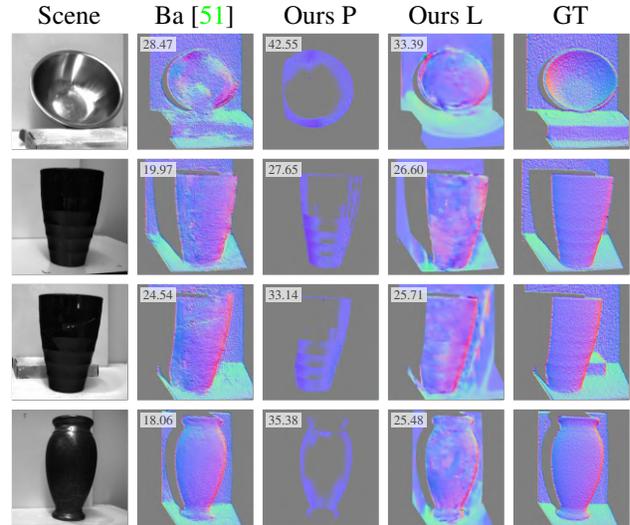
Figure 7. Comparison of best image-baseline (Ba *et al.*) against ours event-based methods on the ESfP-Real dataset

of fronto-parallel geometries causes a polarization signal and therefore a small fill-rate metric. On average, for this dataset, the fill-rate is around 3.6%. This, however, is exactly the condition where the proposed learning based approach excel. By leveraging learned data-priors, it can recover accurate normals even in the presence of very few events. The improvements in both angular error and accuracy, show the proposed method outperforms all baselines, despite falling short by a small margin to Ba *et al.* [51]. We emphasize, however, that these results were obtained under ideal lighting conditions, where data-priors of learning-based approaches are enough to compensate for the limited polarization information that standard sensors can provide. We now show the performance of our methods when these perfect conditions are not satisfied against baselines.

Effect of illumination Fig. 8 highlights the high dynamic range (HDR) advantage of using an event camera for SfP under different illumination conditions. We increased the illumination from 528 Lux to 1442 Lux and evaluated the performance our proposed approach and compared it to the image-baselines. The corresponding images and events are shown. As it can be noticed, the performance of our methods remains consistent across different illumination conditions. This is case for both the model-based and learning-based variants, highlighting that events provide more information than images, and are thus more suited for these challenging conditions.

Effect of speed In this section, we assess the performance of our method under different rotation speeds of the motor. In Table 4, we find the performance of both our methods improves slightly with an increase in the speed. This can be explained as at higher rotation speeds, the rate of intensity change is higher, which makes non-linearities and non-

Speed (RPM)	Task	Angular Error ↓	Accuracy ↑		
		Mean	AE<11.25	AE<22.5	AE<30
57.5	Physics-based	44.29	0.01026	0.1294	0.1998
171.25	Physics-based	44.13	0.01028	0.1297	0.2000
308.75	Physics-based	44.02	0.01043	0.1302	0.2008
57.5	Learning-based	29.47	0.147	0.441	0.613
171.25	Learning-based	**27.10**	**0.177**	**0.498**	**0.667**
308.75	Learning-based	27.47	0.172	0.491	0.656

Table 4. Ablation study on the effect of using different angular velocities for the rotation of the polarizing lens on the proposed ESfP-Real dataset.
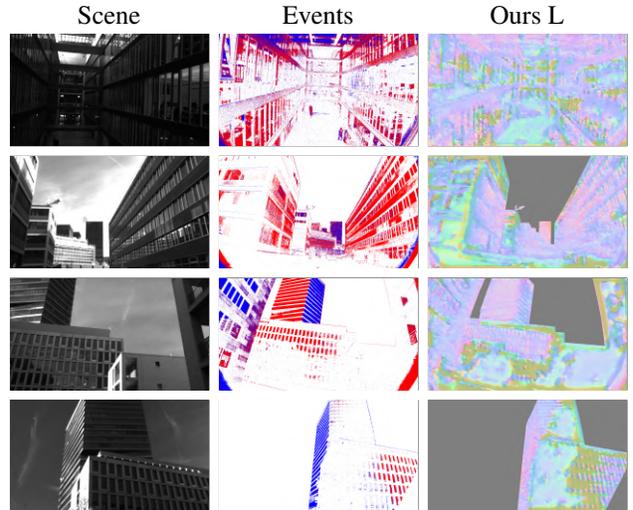
Figure 9. Performance of our method on outdoor scenes

Figure 8. Varying the ambient illumination has a drastic effect on the image-based baselines (bar plot: green colors), whereas both our methods (bar plot: blue colors), perform consistently.

idealities in the event-camera's less prominent, resulting in a better quality of events. Please refer to the suppl. material.

5.3. Generalization to outdoor scenes

In Fig. 9, we show that our method generalizes to outdoor scenes as well, despite being trained on single-object and near-depth scenes. Fine tuning on this dataset is not possible due to the unavailability of any groundtruth. In the future, the proposed work would benefit by adopting a view-encoding proposed in [22] here.

6. Conclusions

We introduce a novel method for surface normal estimation using a rotating linear polarizer and an event camera. To the best of our knowledge, we are the first to propose a principled way of estimating the surface normals from polarization events using a physics-based solution. Our method takes advantage of the rich temporal event information to reconstruct relative event intensities at multiple polarizer angles. We also propose a learning-based method to overcome the non-idealities of the physical sensor which improves the performance by 52% in terms of MAE. Despite reduced performance in the presence of event camera non-idealities, we believe our derivations could pave the way to future hybrid solutions to increased robustness and interpretability.

7. Acknowledgement

References

[1] Gary A Atkinson. Polarisation photometric stereo. *Computer Vision and Image Understanding*, 160:158–167, 2017. 3

[2] Gary A. Atkinson and Jürgen D. Ernst. High-sensitivity analysis of polarization by surface reflection. *Machine Vision and Applications*, 29(7):1171–1189, Aug. 2018. 1, 2

[3] Gary A Atkinson and Edwin R Hancock. Multi-view surface reconstruction using polarization. In *ICCV*, volume 2, page 3, 2005. 3

[4] Gary A Atkinson and Edwin R Hancock. Recovery of surface orientation from diffuse polarization. *IEEE transactions on image processing*, 15(6):1653–1664, 2006. 3

[5] Seung-Hwan Baek, Daniel S Jeon, Xin Tong, and Min H Kim. Simultaneous acquisition of polarimetric svbrdf and normals. *ACM Trans. Graph.*, 37(6):268–1, 2018. 1

[6] Seung-Hwan Baek, Tizian Zeltner, Hyun Jin Ku, Inseung Hwang, Xin Tong, Wenzel Jakob, and Min H. Kim. Image-based acquisition and modeling of polarimetric reflectance. *ACM Transactions on Graphics (Proc. SIGGRAPH 2020)*, 2020. 5

[7] Intel RealSense Depth cameras. https://www.intelrealsense.com/coded-light/. 2

[8] Edward Collett. *Field Guide to Polarization*. SPIE, Sept. 2005. 2

[9] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B. McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items, 2022. 5

[10] Ondrej Drbohlav and Radim Sara. Unambiguous determination of shape from photometric stereo with unknown light sources. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 1, pages 581–586. IEEE, 2001. 3

[11] Jean-Denis Durou, Maurizio Falcone, Yvain Quéau, and Silvia Tozza. *Advances in photometric 3d-reconstruction*. Springer, 2020. 3

[12] Thomas Finateu, Atsumi Niwa, Daniel Matolin, Koya Tsuchimoto, Andrea Mascheroni, Etienne Reynaud, Pooria Mostafalu, Frederick Brady, Ludovic Chotard, Florian LeGoff, Hirotsugu Takahashi, Hayato Wakabayashi, Yusuke Oike, and Christoph Posch. A 1280x720 back-illuminated stacked temporal contrast event-based vision sensor with $4.86\mu m$ pixels, 1.066geps readout, programmable event-rate controller and compressive data-formatting pipeline. In *IEEE Intl. Solid-State Circuits Conf. (ISSCC)*, 2020. 3, 5

[13] Mathias Gehrig, Mario Millhäusler, Daniel Gehrig, and Davide Scaramuzza. E-raft: Dense optical flow from event cameras. In *International Conference on 3D Vision (3DV)*, 2021. 4

[14] Viktor Gruev, Germain Haessig, Damien Joubert, Justin Haque, Yingkai Chen, Moritz Milde, and Tobi Delbruck. Division of focal plane asynchronous polarization imager. In *Polarization: Measurement, Analysis, and Remote Sensing XV*, volume 12112, page 1211208. SPIE, 2022. 3, 6

[15] Cong Phuoc Huynh, Antonio Robles-Kelly, and Edwin R Hancock. Shape and refractive index from single-view spectro-polarimetric images. *International journal of computer vision*, 101(1):64–94, 2013. 3

[16] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer. 2022. https://mitsuba-renderer.org. 5

[17] Daniel Gehrig Javier Hidalgo-Carrio and Davide Scaramuzza. Learning monocular dense depth from events. *IEEE International Conference on 3D Vision.(3DV)*, 2020. 4

[18] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3d: High-quality depth sensing with polarization cues. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3370–3378, 2015. 1, 2, 3, 4

[19] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Depth sensing using geometrically constrained polarization normals. *International Journal of Computer Vision*, 125(1):34–51, 2017. 3

[20] Diederik P. Kingma and Jimmy L. Ba. Adam: A method for stochastic optimization. *Int. Conf. Learn. Representations (ICLR)*, 2015. 6

[21] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[22] Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun, and Qifeng Chen. Shape from polarization for complex scenes in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12632–12641, June 2022. 1, 2, 3, 4, 6, 8

[23] 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*, 43(2):566–576, 2008. 3

[24] Ali H Mahmoud, Moumen T El-Melegy, and Aly A Farag. Direct method for shape recovery from polarization and shading. In *2012 19th IEEE International Conference on Image Processing*, pages 1769–1772. IEEE, 2012. 1, 3, 6, 7, 8

[25] Haiyang Mei, Bo Dong, Wen Dong, Jiaxi Yang, Seung-Hwan Baek, Felix Heide, Pieter Peers, Xiaopeng Wei, and Xin Yang. Glass segmentation using intensity and spectral polarization cues. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022. 1

[26] Miyazaki, Tan, Hara, and Ikeuchi. Polarization-based inverse rendering from a single view. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 982–987 vol.2, 2003. 3

[27] Daisuke Miyazaki, Masataka Kagesawa, and Katsushi Ikeuchi. Transparent surface modeling from a pair of polarization images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):73–82, 2004. 3

[28] Daisuke Miyazaki, Takuya Shigetomi, Masashi Baba, Ryo Furukawa, Shinsaku Hiura, and Naoki Asada. Surface normal estimation of black specular objects from multiview polarization images. *Optical Engineering*, 56(4):041303, 2016. 3

[29] Manasi Muglikar, Guillermo Gallego, and Davide Scaramuzza. ESL: Event-based structure light. In *IEEE International Conference on 3D Vision.(3DV)*, December 2021. 5

[30] Manasi Muglikar, Diederik Paul Moeys, and Davide Scaramuzza. Event guided depth sensing. In *2021 International Conference on 3D Vision (3DV)*. IEEE, Dec. 2021. 4, 5

[31] Trung Ngo Thanh, Hajime Nagahara, and Rin-ichiro Taniguchi. Shape and light directions from shading and polarization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2310–2318, 2015. 3

[32] PolarM polarization camera (2017). http://www.4dtechnology.com/products/polarimeters/polarcam/. 2

[33] Lucid Vision Phoenix polarization camera (2018). https://thinklucid.com/product/phoenix-5-0-mp-polarized-model/. 2, 3, 5

[34] Breakthrough Photography X4 Polarizer. https://breakthrough.photography/products/x4-circular-polarizer. 5

[35] Christoph Posch, Daniel Matolin, and Rainer Wohlgenannt. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE J. Solid-State Circuits*, 46(1):259–275, Jan. 2011. 3

[36] Prophesee Evaluation Kits. https://www.prophesee.ai/event-based-evk/, 2020. 4

[37] Stefan Rahmann and Nikos Canterakis. Reconstruction of specular surfaces using polarization imaging. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001. 3

[38] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. ESIM: an open event camera simulator. In *Conf. on Robotics Learning (CoRL)*, 2018. 5

[39] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019. 4, 5

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. 4

[41] Y.Y. Schechner, S.G. Narasimhan, and S.K. Nayar. Instant dehazing of images using polarization. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, 2001. 1

[42] Y.Y. Schechner and S.K. Nayar. Generalized mosaicing: polarization panorama. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):631–636, 2005. 1

[43] William A. P. Smith, Ravi Ramamoorthi, and Silvia Tozza. Height-from-polarisation with unknown lighting or albedo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(12):2875–2888, 2019. 6, 7, 8

[44] Christophe Stolz, Mathias Ferraton, and Fabrice Meriaudeau. Shape from polarization: a method for solving zenithal angle ambiguity. *Optics letters*, 37(20):4218–4220, 2012. 3

[45] Yunjae Suh, Seungnam Choi, Masamichi Ito, Jeongseok Kim, Youngho Lee, Jongseok Seo, Heejae Jung, Dong-Hee Yeo, Seol Namgung, Jongwoo Bong, Jun seok Kim, Paul K. J. Park, Joonseok Kim, Hyunsurk Ryu, and Yongin Park. A 1280x960 Dynamic Vision Sensor with a 4.95-μm pixel pitch and motion artifact minimization. In *IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2020. 3

[46] Stepan Tulyakov, Alfredo Bochicchio, Daniel Gehrig, Stamatios Georgoulis, Yuanyou Li, and Davide Scaramuzza. Time lens++: Event-based frame interpolation with parametric non-linear flow and multi-scale fusion. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 5

[47] Masada Tzabari and Yoav Y. Schechner. Polarized optical-flow gyroscope. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*. Springer International Publishing, 2020. 1

[48] Microsoft Kinect v1. http://www.xbox.com/kinect. 2

[49] Lawrence B Wolff. Polarization vision: a new sensory approach to image understanding. volume 15, pages 81–93. Elsevier, 1997. 1, 2

[50] Rongyuan Wu, Yongqiang Zhao, Ning Li, and Seong G. Kong. Polarization image demosaicing using polarization channel difference prior. *Opt. Express*, 29(14):22066–22079, Jul 2021. 2

[51] Ba Yunhao, Gilbert Alex, Wang Franklin, Yang Jinfa, Chen Rui, Wang Yiqin, Yan Lei, Shi Boxin, and Kadambi Achuta. Deep shape from polarization. 2020. 1, 2, 3, 4, 6, 7, 8

[52] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based optical flow using motion compensation. In *Eur. Conf. Comput. Vis. Workshops (ECCVW)*, 2018. 4, 5

[53] Dizhong Zhu and William AP Smith. Depth from a polarisation+ rgb stereo pair. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7586–7595, 2019. 3