Indoor Lighting Estimation using an Event Camera

Zehao Chen[1,2#] Qian Zheng[3#]   Peisong Niu[1]   Huajin Tang[1]   Gang Pan[1,2*]

[1] College of Computer Science and Technology, Zhejiang University, Hangzhou, China   [2] Zhejiang Lab, Hangzhou, China
[3] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

zehao@zju.edu.cn, zhengqian@ntu.edu.sg, {psacfc, htang, gpan}@zju.edu.cn

Abstract

Image-based methods for indoor lighting estimation suffer from the problem of intensity-distance ambiguity. This paper introduces a novel setup to help alleviate the ambiguity based on the event camera. We further demonstrate that estimating the distance of a light source becomes a well-posed problem under this setup, based on which an optimization-based method and a learning-based method are proposed. Our experimental results validate that our approaches not only achieve superior performance for indoor lighting estimation (especially for the close light) but also significantly alleviate the intensity-distance ambiguity.

1. Introduction

Obtaining lighting information is a classic problem in computer vision and graphics. It contributes to solving a variety of vision tasks, such as photometric stereo [29, 78], virtual object compositing [31], and scene understanding [55]. A field of researches study to estimate or calibrate lighting by taking a single image of an illuminated object (*e.g.*, [9, 34]) or scene (*e.g.*, [59, 35]) as the input.

Early work assumes the light source to be distant and focuses on the estimation of light direction (*e.g.*, [15]). This assumption is often violated for indoor scenes due to the localized light sources. Recent advances tackle this problem by either estimating a spatially-varying lighting at different scene points (*e.g.*, [16]) or predicting the light source positions in 3D space (*e.g.*, [14]). However, as these approaches take a single image as the input, this inherently ill-posed problem is likely to suffer from the problem of *intensity-distance ambiguity*. Because the light distance inferred from a given intensity (recorded by an image) is not guaranteed to be unique if the light source intensity changes [36].

Recently, the event camera (*e.g.*, [37]) has attracted the attention of many academics due to its advantages of high
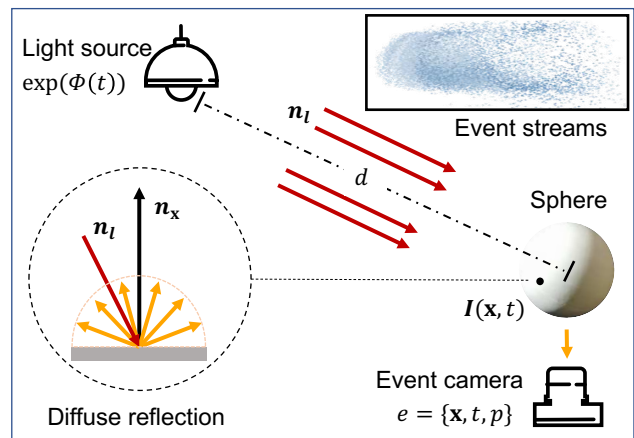

Figure 1. The setup of our method: using an event camera to capture the intensity changes on a purely diffuse sphere, which is placed in a dark room, for the split second of turning light on.

temporal resolution, high dynamic range, and sensitivity to small intensity changes. And it has been used to solve various vision tasks such as 3D reconstruction [64], optical flow estimation [13], and segmentation [62].

In this paper, we leverage the event camera to alleviate the intensity-distance ambiguity for indoor lighting estimation. Our basic idea is to use more information to better constrain the estimation of light source parameters, because the event camera can capture signals in a split second. To be specific, we introduce a novel setup as shown in Figure 1. With such a setup, we observe that the intensity-distance ambiguity can hardly be found for the event streams (Figure 2). We detail the analysis about the ambiguity in Section 3.2, which is based on our analytic formulation of event streams (Section 3.1). We further show that estimating the distance of light source becomes a well-posed problem with the input of event streams, based on which an optimization-based method is proposed (Section 4.1). We also propose a learning-based method for robust lighting estimation (Section 4.2). To evaluate our methods, we collect two testing datasets with paired data captured by a traditional camera and an event camera (Section 4.2). Experimental results

---

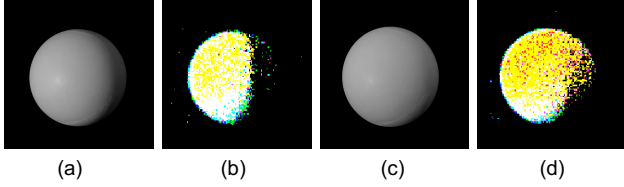[#]Equal contribution. [*]Corresponding author.

Figure 2. (a)&(b) (or (c)&(d)) are paired data, captured with the same light-emitting diode, *i.e.*, LED, 3w (or 5w), and the same distance of light source 119 *cm* (or 213 *cm*). (a)&(c) are RGB images captured by a digital camera while (b)&(d) are event streams captured by an event camera (all event streams in this paper are visualized based on the method in [81]). Even though (a)&(c) have a similar appearance, (b)&(d) display discriminative features.

demonstrate that our methods not only achieve superior performance for lighting estimation, but also significantly alleviate the intensity-distance ambiguity. In addition, we provide a byproduct application to classify the types of the light source. Our contributions are summarized as:

1. We introduce a novel setup, based on which we show that the intensity-distance ambiguity can be alleviated for the problem of indoor lighting estimation.

2. We show that estimating lighting distance becomes a well-posed problem with our setup, based on which we propose an optimization-based method and a learning-based method.

3. We demonstrate that our methods not only achieve superior performance but also significantly alleviate the intensity-distance ambiguity for the problem of indoor lighting estimation. Our byproduct of lamp classification also verifies the effectiveness of our methods.

2. Related Work

2.1. Lighting Estimation

Existing lighting estimation methods can be roughly categorized according to whether the assumption of distant light is taken. Such an assumption is often applied to the outdoor scenes [26, 27, 25, 76, 24, 45, 33], or found from methods that rely on an object, *e.g.*, light probes [9, 10, 6, 34], faces [75, 5], cars [20], bags of chips [50], glass [77], or illuminated objects [6, 42, 19]. The assumption of distant light often violates for indoor scenes. Except for a few early works [15], recent advances address the problem of spatially-varying lighting estimation. These methods either recover a spherical panorama for each point of the given scene [16, 59, 35, 61, 65] or directly predict the position of light sources in 3D space [14].

In this paper, we also consider the problem of parametric indoor lighting estimation. Different from existing methods, we also predict the intensity changing function of the light source when turning it on.

2.2. Event-based Camera

The event camera (*e.g.*, DVS [37]) is a bio-inspired sensor that imitates the neural structure of the eye. Different from traditional cameras that directly collect the radiant intensity, event cameras capture small changes of intensity with high temporal resolution (in microseconds) with significantly lower power consumption. These advantages are imposed to solve versatile vision tasks such as tracking [18, 43, 52, 74, 69], optical flow estimation [13, 11, 62, 17, 49, 2, 47, 1], image restoration or enhancement [54, 66, 53, 48, 38, 30, 8, 67, 70, 22, 63], depth estimation [81, 3, 80, 79, 64], SLAM [12, 44, 39], segmentation and recognition [62, 46, 4, 41, 68, 58, 56, 7, 28, 40].

Most the existing works produce the event signal based on the relative motion between the camera and the scene. In contrast, this paper focuses on the signal brought by the intensity changes of a light source, in a static scene without relative motion. To the best of our knowledge, this is the first attempt to use event cameras for lighting estimation.

3. Modeling

This section models our setup shown in Figure 1. The scene is considered to be static during the split second. Due to the short period, the fact that both the camera and the sphere stay static, and the time window of the event camera is very short (less than 0.1 second), we consider our setup to be *single shot*.

3.1. Analytic Formulation of Event Streams

Thanks to the static scene, we build the analytic formulation of event streams based on the radiant intensity.

Analytic formulation of radiant intensity. As the distance of the light source (> 70 *cm*) is much larger than the size of the sphere (8 *cm*), we assume that the direction and intensity of incoming light for all points on the sphere surface are the *same*, similar to other topics (*e.g.*, photometric stereo [57]). We also assume the illuminance of emitting light obeys the inverse square law [36]. Then the incoming light on the sphere surface S at time t can be formulated as

$$L(t) = \frac{\exp(\Phi(t))}{4\pi d^2}, \tag{1}$$

where d is the distance between the light source and sphere. We consider the intensity changing function as a power-on step function [32], and represent it as $\exp(\Phi(t))$. As we use a purely diffuse sphere, the radiant intensity of point $\mathbf{x}$ at time t can be formulated according to the Lambertian reflectance model

$$I(\mathbf{x}, t) = \rho L(t) \max(\langle \mathbf{n}_l, \mathbf{n}_\mathbf{x} \rangle, 0), \quad t \in [t_0, t_n], \quad \mathbf{x} \in S, \tag{2}$$

where ρ is the albedo which is a constant, $\mathbf{n}_l$ is the direction of the light source, $\mathbf{n}_\mathbf{x}$ is the surface normal of point $\mathbf{x}$, the

action of turning light source on starts at time t_0, and $\Phi(t)$ is stable at time t_n.

Radiant intensity from event streams. On the other hand, the event camera captures the intensity changes for each point $\mathbf{x} \in S$ for the split second of turning light on, and produces a stream of asynchronous sparse events $e \doteq (\mathbf{x}, t, p)$ [37]. As the intensity changes are monotonically non-decreasing for this split second, we have polarity $p = 1$ and such the event streams can be described as[1]

$$e(\mathbf{x}, t) \doteq \begin{cases} 1, \text{if } \Delta I = C, \\ \text{none}, \text{otherwise}. \end{cases} \quad t \in [t_0, t_n], \; \mathbf{x} \in S, \quad (3)$$

where

$$\Delta I = \lg I(\mathbf{x}, t) - \lg I(\mathbf{x}, t - \Delta t), \quad (4)$$

$I(\mathbf{x}, t)$ is the intensity captured by the camera for a point $\mathbf{x}$ at time t, C is the contrast threshold and is obtained from the camera configuration (0.812 in this paper), Δt is the time since the last event is triggered at position $\mathbf{x}$. Due to the darkroom, all initial intensity captured by the camera can be regarded as the same value I_0, i.e., $\forall \mathbf{x} \in S, I(\mathbf{x}, t_0) = I_0$. We set $I_0 = 1$ to simplify the computation. As the position of point $\mathbf{x}$ regarding the event camera is unchanged during this split second, the radiant intensity of $\mathbf{x}$ at time t can be calculated based on the summation of events

$$I_e(\mathbf{x}, t) = I_0 \exp(C \int_{t_0}^{t} e(\mathbf{x}, t) \mathrm{d}t), \quad t \in [t_0, t_n], \quad \mathbf{x} \in S. \quad (5)$$

The equality of $I_e(\mathbf{x}, t)$ and $I(\mathbf{x}, t)$ builds the analytic formulation of $e(\mathbf{x}, t)$. This formulation describes the relation between $e(\mathbf{x}, t)$ and the parameters of light source, i.e., distance d, direction $\mathbf{n}_l$, intensity changing function $\Phi(t)$.

3.2. Ambiguity Analysis

According to Equation (1), the intensity of incoming light is determined by $\Phi(t)$ and d. We consider an object or a scene is illuminated by two different light sources with intensity functions of $\Phi_1(t)$ and $\Phi_2(t)$, respectively.

For image-based lighting estimation methods, intensity-distance ambiguity exists if $\exp(\Phi_1(t_n)) = \alpha \exp(\Phi_2(t_n))$ holds, where α is a positive constant. For our methods that take event streams as the inputs, intensity-distance ambiguity exists only if the following equation holds

$$\exp(\Phi_1(t)) = \alpha \exp(\Phi_2(t)), \qquad \forall t \in [t_0, t_n]. \quad (6)$$

As $\Phi(t)$ is determined by a sophisticated physical process including inherent factors such as current, resistance, and
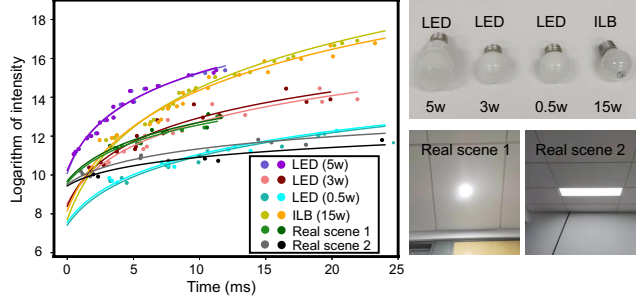

Figure 3. For each $t_t \in [t_0, t_n]$ where events are triggered, $\Phi(t_t)$ can be calculated based on the analytic formulation of $e(\mathbf{x}, t)$, with known d and $\mathbf{n}_l$. We plot points $\{t_t, \Phi(t_t)\}$ ($\Phi(t_t)$ is averaged over different $\mathbf{x}$ at time t_t) according to eight event streams (represent by different colors). For each set of points, we fit it based on Equation (7). The data are from controlled dataset and wild dataset.

the arrangement&number of LEDs, we argue that any two $\Phi(t)$ can hardly satisfy the constraint in Equation (6).

Obviously, satisfying the constraint in Equation (6) is much more difficult than that for images. Therefore, our setup is expected to alleviate the intensity-distance ambiguity for lighting estimation. The experimental validation about the ambiguity can be found in Section 5.2.

3.3. Approximating $\Phi(t)$

Although $\Phi(t)$ is determined by a sophisticated physical process, it is determined by finite inherent factors. Inspired by the shapes of curves in Figure 3, we use an empirical formulation to approximate $\Phi(t)$[2]

$$\Phi(t) = a \lg(t + c) + b, \quad (7)$$

where a and b are used to formulate the inherent factors of the light source. $c \in [0, 2000]$ is used to model the time delay from the moment of turning light on to t_0 to tolerate the fluctuation caused by unexpected noises. We observe that approximation in Equation (7) works well for the estimation of light source distance [3]. The experimental validation can be found in Section 5.3.

3.4. Parametric Lighting

We use the parametric lighting-to-environment map projection function described in GH19 [14] to convert parametric light into an environment map. Unlike GH19 [14], our method cannot estimate color, since the event camera can only capture intensity changes. In addition, since the

---

[1]We replace the condition $\Delta I \geq C$ by $\Delta I = C$. This replacement has little impact on our following analysis. It only brings a consistent transformation to function $\Phi(t)$, while the accurate $\Phi(t)$ is not necessarily required in our solution or analysis.

[2]The curve describes the intensity changes caused by transient current spikes in a rise time [72] and a step function can always fit well.

[3]The function in Equation (7) can be revised with more parameters (e.g., higher-order polynomials) for more complicated cases only if the numbers of the unknown is smaller than the number of constraints used in our paper (e.g., 7).

distance of the light source (> 70 *cm*) is much larger than its radius (~ 10 *cm*), we replaced the angular size [73] in GH19 [14] by s/d, where s is a fixed value (10 *cm*) indicating the radius of the point light source and d is the distance of the light source. The projection function can be written as:

$$f(\mathbf{n}_l, s, d, \mathbf{n_x}) = \exp(\frac{\mathbf{n}_l \cdot \mathbf{n_x} - 1}{\frac{1}{4\pi}\frac{s}{d}}) \qquad (8)$$

4. Proposed Method

Inspired by the analytic formulation of event streams, our indoor lighting estimation aims to estimate the light source parameters of distance d, direction $\mathbf{n}_l$, and intensity changing function $\Phi(t)$ by taking event streams as the input.

4.1. Optimization-based Method

Based on Equation (2) and Equation (5), our indoor lighting estimation is to solve an over-determined system: $\forall t \in [t_0, t_n], \forall \mathbf{x} \in S, \|I_e(\mathbf{x}, t) - I(\mathbf{x}, t)\| = 0$, which can be achieved by the following minimization[4]

$$\min_{d, \mathbf{n}_l, a, b, c} \int_S \int_{t_0}^{t_n} \|I_e(\mathbf{x}, t) - I(\mathbf{x}, t)\| \mathrm{d}t \mathrm{d}\mathbf{x}. \qquad (9)$$

To solve Equation (9), we first optimize the light direction $\mathbf{n}_l$ by minimizing the difference between the analytic illuminated region and that captured by the event camera:

$$\min_{\mathbf{n}_l} \int_S \|B(\int_{t_0}^{t_n} I_e(\mathbf{x}, t)\mathrm{d}t) - B(\int_{t_0}^{t_n} I(\mathbf{x}, t_n)\mathrm{d}t)\| \mathrm{d}\mathbf{x}, \qquad (10)$$

where function $B(I)$ produces the binarization result of I,

$$B(I) = \begin{cases} 1, & \text{if } I > I_0 \\ 0, & \text{otherwise}, \end{cases} \qquad (11)$$

and we have $B(\int_{t_0}^{t_n} I(\mathbf{x}, t_n)\mathrm{d}t) = B(\max(\langle \mathbf{n}_l, \mathbf{n_x} \rangle, 0))$ This is achieved by the stochastic gradient descent technique[60] with $\mathbf{n}_l^0$ initialized as $(1, 0, 0)$. With the estimated $\mathbf{n}_l$, we alternatively optimize d and $\{a, b, c\}$ with an iteration scheme, and they are initialized as $\{5, 0.01, 100\}$, which is also solved by the stochastic gradient descent technique[60].[5]

4.2. Learning-based Method

Lighting estimation by the optimization-based method may produce unreliable results due to the unexpected random noise in event streams and the fact that $I(\mathbf{x}, t)$ is not completely equivalent to $I_e(\mathbf{x}, t)$. Besides, the problem of

---
[4]We integrate ρ to $\Phi(t)$.

[5]More details about our algorithm can be found in the supplementary material.
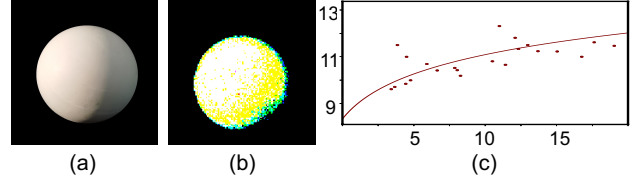

Figure 4. Example of an unreliable prediction by our optimization-based method (ground truth *vs.* prediction: 182 *cm vs.* 270 *cm*). (a) The RGB image. (b) Event streams. (c) A set of points $\{t_t, \Phi(t_t)\}$ obtained as the same way in Figure 3.

the local minimum leads to less accurate estimation (Section 5.3). Figure 4 shows an example of such an unreliable prediction, *i.e.*, $\Phi(t)$ cannot be well fitted due to the outliers from unexpected noise. To this end, we propose a learning-based method in this section.

Our learning-based method focuses on the estimation of d and $\{a, b, c\}$ and leaves the prediction of $\mathbf{n}_l$ for the optimization-based method. Because the optimization in Equation (10) addresses the influence of noise by integrating signals with respect to time t and is free from the problem of local minimum due to the binarization. The overview of our framework is shown in Figure 5.

Streams pre-processing. Similar to several previous methods, we pre-process the event stream so that it can be used as an input of a neural network. We adopt a similar data pre-processing strategy as that in [81]. The temporal dimension (*i.e.*, 24 *ms*) is discretized into 24 bins. Temporal information is preserved by the sequential order of these bins. For the temporal information in each bin, we encode the event streams according to the relative timestamp of the first event for each point $\mathbf{x}$. We also align the input to make their elevation angles uniform based on the estimated $\mathbf{n}_l$ (*i.e.*, $45°$). Experimental results show that such a pre-processing strategy is sufficient to obtain discriminative cues for lighting estimation.

Network architecture and loss functions. The input passed through the headless ResNet-50 architecture [23] and produce a 2048-dimensional latent vector. This vector is then fed to three fully-connected layers with 1024 units, 500 units and 100 units respectively to regress d and $\{a, b, c\}$ separately. We use L_1 loss [21] to perform supervised training and the loss function is

$$\mathcal{L} = L_1(d_{\text{gt}}, d_{\text{pre}}) + w_1 L_1(a_{\text{gt}}, a_{\text{pre}}) + \qquad (12)$$
$$w_2 L_1(b_{\text{gt}}, b_{\text{pre}}) + w_3 L_1(c_{\text{gt}}, c_{\text{pre}}) \qquad (13)$$

where subscript 'gt' and 'pre' represent the ground truth and prediction of a variable. We empirically set the weight $w_1 = 70, w_2 = 40, w_3 = 0.001$ in our experiment.
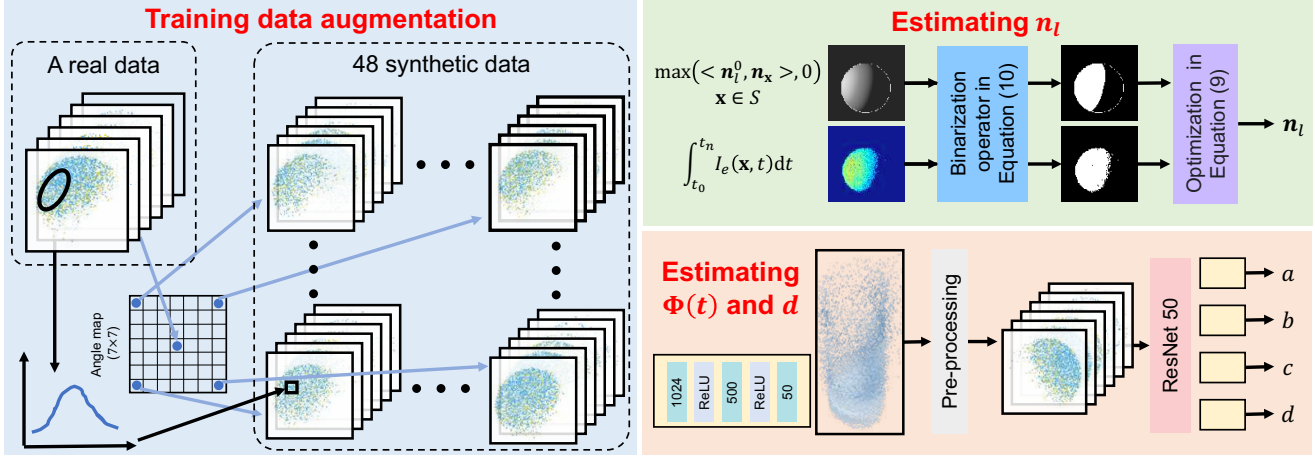
Figure 5. The framework of our learning-based method. Training data augmentation: each real data will generate 48 synthetic data based on the angle map. Estimating $\mathbf{n}_l$: we estimate the direction of the light source by minimizing the difference between the analytic illuminated area and that captured by the event camera. Estimating $\Phi(t)$ and d: the event streams are discretized into frames and then fed to our network.
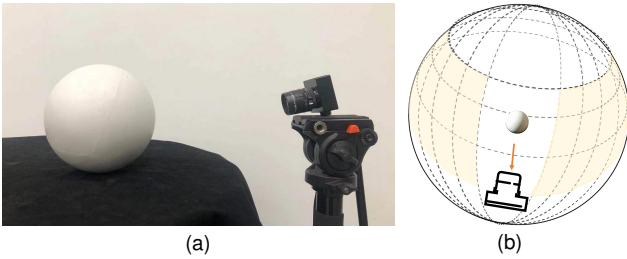

Figure 6. (a) Illustration of our data collection. (b) Schematic diagram of data collection. We randomly collect the same number of samples in each colored area, where the elevation angle ranges from $0°$ to $60°$, and azimuth angle ranges from $0°$ to $75°$ and from $105°$ to $180°$. Note that the variation of elevation angle is less considered as it is aligned in our pre-processing procedure.

Training details. We implement the network using Pytorch [51] and our entire network contains 3.6M parameters. The parameters of ResNet-50 are initialized by their pre-trained model. The network is trained for 100 epochs with an early stopping mechanism. The Adam optimizer is used with a learning rate of 0.001 and a batch size of 64.

Data collection. We collect two types of data for training and testing: controlled dataset and wild dataset. The controlled dataset is to investigate the effectiveness of our method that alleviates the intensity-distance ambiguity. Specifically, we collect 400 real data with different d ranges from 70 cm to 300 cm and different $\mathbf{n}_l$ evenly sampled from the orange surface as shown in Figure 6 [6]. To simulate dif-

ferent types of lighting, we use LED lamps with different powers (*i.e.*, 0.5w, 3w, 5w) and an ILB (15w) to capture these data (100 for each). Their corresponding RGB images (shown in Figure 7(a)) on the illuminated sphere are also captured for comparison with image-based methods. The wild dataset is to investigate the robustness and practicability of our methods. Specifically, we collect 363 real data from 19 different real scenes. Most of the wild dataset are captured with a single light source, and only a few of them are captured with multiple light sources where lights are close to each other. Due to the close distance between multiple light sources, we also consider them to be single light sources. The mean of light distances in wild dataset is 198cm. Each data includes event data, an RGB image, the position of the light source, and the environment map collected at the same position with a light probe (shown in Figure 7(b)) [7].

Data processing. We identify time $t_0 = 0$ as the moment when 30 events appeared within 1 ms. To balance the efficiency and efficient, we set the split second of turning the light on to be 24 ms, *i.e.*, $t_n = 24$. We find it is sufficient to extract discriminative features for lighting estimation.

Training and testing data. We randomly select 160 data from controlled dataset and augment $7680 = 48 \times 160$ data for training. Note that these 160 data are *not* included in our training dataset. The augmentation is achieved by using another 48 different $\mathbf{n}_l$ to synthesize 48 new data based on

---

[6]We suppress the influence of inter-reflection by putting a black cloth between the sphere and the upholder and keeping the distance between the

sphere and other objects as far as possible.

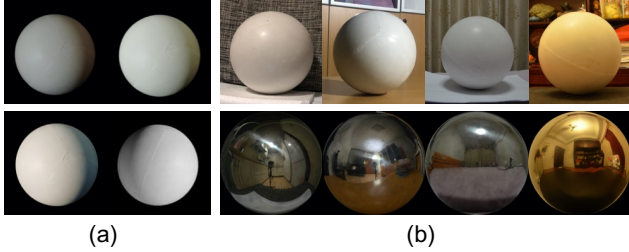[7]We transform light probes as panoramas through HDRshop. HDRshop: https://vgl.ict.usc.edu/HDRShop/.

Figure 7. (a) RGB data in controlled dataset (upper left: LED lamps (3w), upper right: LED lamps (5w), bottom left: ILB lamp (15w), bottom right: LED lamps (0.5w)) (b) RGB data in wild dataset (upper), their corresponding light probes (bottom).
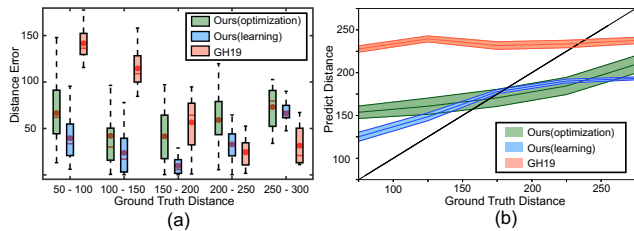

Figure 8. (a).The error distribution is displayed using the box-and-whisker plot. Red lines indicate mean values, top and bottom bounds of boxes indicate the first and third quartile values, and the top and bottom ends of the vertical black lines indicate the minimum and maximum values. (b).The distance distribution. The middle line of each curve indicates the mean value. The difference between sidelines and middle lines for each curve reflects the standard deviation values.

each of 160 data. We build the distribution of events, and let the distribution of synthetic data be the same as that of real data, regarding each point with the same $< \mathbf{n}_l, \mathbf{n_x} >$. We augmented 48 data for each data by increasing the original $\mathbf{n}_l$ by 4°, 5° and 6°, which brings $(3 * 2 + 1)^2 - 1 = 48$ surrounding directions as shown in Figure 5 (angle map). We use the remaining 240 data from controlled dataset and all data in wild dataset for testing.

5. Experiments

We perform the comparison with a parametric method GH19 [14] and an object image method WP18 [71]. GH19 [14] predicts the 3D positions of three light sources in an indoor scene, the one which is closest to the ground truth is used for our comparison. WP18 [71] provides three pre-train models of different materials, the best one is used for our comparison. We define the direction error as the angle between the direction of prediction and that of ground truth. Similar to [14], we use metric RMSE to evaluate the quantitative performance of estimated light distance.

Table 1. Quantitative performance comparison in terms of RMSE for distance and direction estimation in controlled dataset (mean ± std).

Method	Distance	Direction
Ours (optimization)	51.70±62.59	13.4° ±5.26°
Ours (learning)	**27.05±35.78**	**13.4°±5.26°**
GH19 [14]	75.63±89.99	38.44° ±7.77°

Table 2. Quantitative performance comparison in terms of RMSE for distance and direction estimation in wild dataset (mean ± std).

Method	Distance	Direction
Ours (optimization)	53.0±67.9	28.8° ±34.3°
Ours (learning)	**34.0±46.5**	**28.8°±34.3°**
GH19 [14]	112.4±126.8	43.0° ±48.6°

5.1. Overall Performance

Controlled dataset. Table 1 and Figure 8 display the quantitative performance in controlled dataset. Since WP18 [71] cannot predict the lighting distance, we only compare with GH19 [14]. As can be observed, both our optimization-based method and learning-based method achieve much better overall performance regarding mean and standard deviation as compared with GH19 [14]. The performance advantage of our methods benefits from more constraints from event streams than those from an image for lighting estimation. It can also be found that our learning-based method outperforms the optimization-based method, due to the unreliable prediction of our optimization-based method as introduced in Section 4.2. Although our methods achieve superior performance advantage over GH19 [14] for a close distance (*i.e.*, $< 200cm$), it fails to outperform GH19 [14] for a large distance (*i.e.*, $> 200cm$). This is for two reasons: 1) A larger distance indicates a smaller intensity of the incoming light on the sphere, leading to a smaller number of events triggered. This makes the random noise more dominant. The unreliable input of event streams degrades the performance of our methods. 2) GH19 [14] tends to produce results between 200 *cm* and 300 *cm* for most images as shown in Figure 8 (b).

Wild dataset. Table 2 displays the quantitative performance in wild dataset. As can be observed, due to the more complex scenarios (*i.e.*, different light source shapes) in wild dataset, the results are not as good as controlled dataset, Even so, our method achieves better results as compared with GH19 [14], which indicates its good generalization to real scenarios. We further perform visual quality study for outputs of our methods, GH19 [14], and WP18 [71]. Figure 9 shows the environment maps results of real indoor scenes. As can be observed, our method produces a much reliable estimation as compared with that from GH19 [14] and WP18 [71]. We also investigate the visual quality performance by inserting objects into indoor
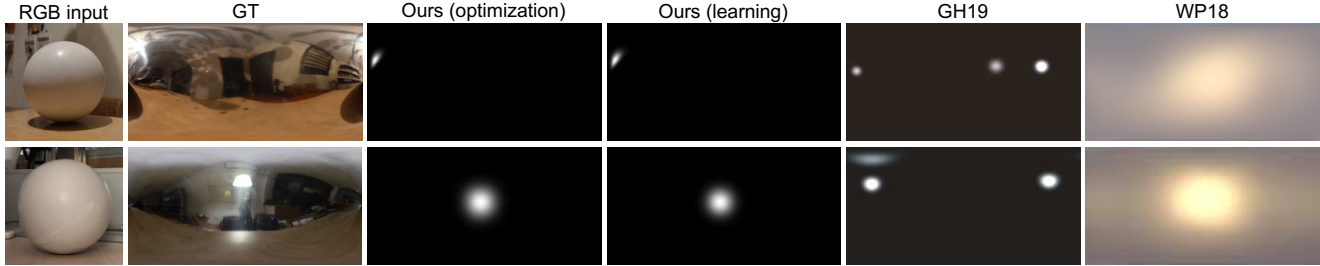
Figure 9. Visual comparison of environment maps for data in wild datasets. From left to right: the RGB input, the ground truth, results from our optimization-based method, our learning-based method, GH19 [14], and WP18 [71].
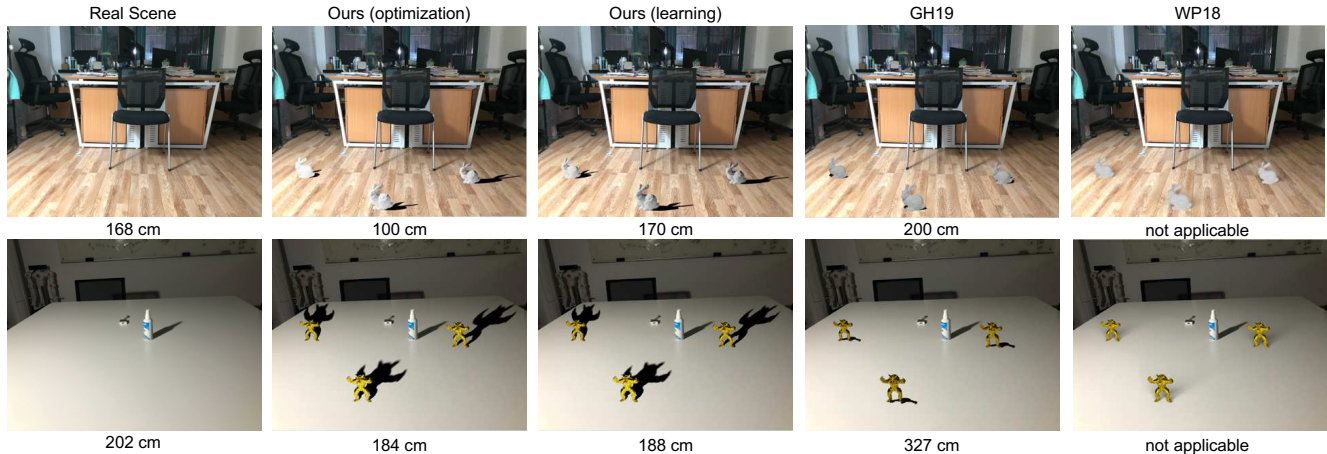

Figure 10. Visual comparison of object insertion results for data in wild datasets. From left to right: the scene image, results from our optimization-based method, our learning-based method, GH19 [14], and WP18 [71]. Numbers below each image indicates the ground truth or predicted result of light source distance. The input data of our method is captured by putting the diffuse sphere on the desk (in the middle of each scene).

scenes in the real world. Kindly note that the natural lamps in these scenes (*i.e.*, ceiling lamp and table lamp) are quite different from those lights in the training data. As can be observed from Figure 10, the inserted objects are more realistic based on results from our methods than those from GH19 [14] and WP18 [71]. For example, the cast shadows rendered with our results are more consistent with those in the scene, while those rendered with results from GH19 [14] are very concentrated, caused by their incorrectly predicted direction and distance of the light source. The virtual objects at several positions in WP18 [71] are same since their methods assume that the light source is distant. The performance advantage of our optimization-based method and learning-based method validates the good generalization of our methods to indoor scenes in the real world.

5.2. Validation for Ambiguity Alleviation

To further investigate the effectiveness of our method that alleviates the intensity-distance ambiguity, we collect another testing dataset. To be specific, we collect 20 paired data from controlled dataset, each pair contains two images as well as their corresponding event streams. We use dif-

ferent lamps and carefully adjust the light source distance to ensure the images in each pair have a similar appearance while quite different light distances d.

As can be observed from Table 3, our methods achieve much smaller RMSE for all these paired data while producing more discriminative results for two data from each pair. Figure 11 visualizes the distributions of predicted results. It can be found that our methods successfully distinguish close (green dots) and far (orange dots) light source distance and produce more accurate predictions. GH19 [14] fails to separate data captured based on close (yellow dots) and far lights (blue dots). These results clearly show that our method successfully alleviates the ambiguity between light source intensity and light source distance.

5.3. Validation for Approximation in Equation (7)

Since the controlled dataset contains the same number of different types of light sources, in this section we validate the approximation in Equation (7) with controlled dataset. As there is no ground truth of $\Phi(t)$ or $\{a, b\}$, we regard $\{a, b\}$ fitted with known d and $\mathbf{n}_l$ as the 'ground truth'. As can be observed from Figure 12 (left), $\{a, b\}$ are well sepa-
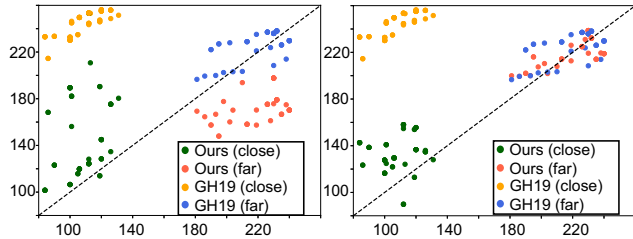
Figure 11. The x-axis represents the ground truth, and the y-axis represents the predicted distance. We compare our optimization-based method (left) and our learning-based method (right) with GH19 [14]. Different colors of dots represent data captured either with a close light source (green&yellow) or a far one (blue&orange).
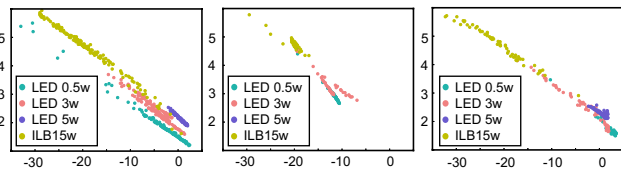

Figure 12. The distribution of $\{a, b\}$ for our 240 testing data from controlled dataset. x-axis represents variable a while y-axis represents variable b. Left: $\{a, b\}$ fitted by optimizing Equation (9) with known d and $\mathbf{n}_l$, and are considered as the 'ground truth'. Middle: $\{a, b\}$ fitted by optimizing Equation (9) with unknown d and $\mathbf{n}_l$. Right: $\{a, b\}$ calculated by our learning-based method. Different colors represent data captured by different lamps.

Table 3. Quantitative performance comparison for distance estimation (mean $\pm$ std). The average difference reflects the difference between two predicted distances for each paired data whose RGB images contains intensity-distance ambiguity.

Method	Average error	Average difference
Ours (optimization)	40.50±47.20	38.09±44.90
Ours (learning)	**28.87±35.56**	**76.37±77.10**
GH19 [14]	75.50±96.55	10.90±11.88

rated regarding different lamps. This observation verifies the practicability of our approximation to $\Phi(t)$ in Equation (7). Note that the fluctuation of $\{a, b\}$ for each lamp is due to different initial currents at the moment of turning light on, caused by the fact that our light sources are powered by the alternating current.

We also illustrate the distributions of $\{a, b\}$ calculated by our methods. As shown in Figure 12 (middle), although the distribution of $\{a, b\}$ by our optimization-based method is less accurate in absolute spatial space, it also contains separated clusters regarding different lamps. Considering both a and b have positive correlation with $\Phi(t)$, the accurate estimation of d by our optimization-based method (Table 1), and also the well separated clusters in Figure 12 (middle), we consider our optimization-based method is troubled by the problem of local minimum. Our learning-based method produces a much more accurate distribution of $\{a, b\}$ and which is considered to be free from the local minimum

Table 4. Light source classification results of our 240 testing data from controlled dataset.

Lighting type	LED(0.5w)	LED(3w)	LED(5w)	ILB (15w)
Accuracy(%)	15.0	41.7	100.0	90.0

problem, as shown in Figure 12 (right).

5.4. A Byproduct for Lamp Classification

We investigate a byproduct application of lamp classification in this section. The results from our optimization-based are not reported due to the local minimum problem. A simple strategy of lamp classification can be achieved by comparing an estimated $\{a, b\}$ with all $\{a, b\}$ in Figure 12 (left). Table 4 shows the classification accuracy using our learning-based method. As can be observed, our method achieves a very high accuracy rate for the ILB lamp (15w) and the LED lamp (5w) as the strong lighting produces more reliable event streams for our estimation.

6. Conclusion

In this paper, we leverage the event camera to alleviate the intensity-distance ambiguity for parametric indoor lighting estimation based on its advantages of high temporal resolution and high dynamic range. To the best of our knowledge, we are the first to estimate lighting using event cameras. We introduce a novel step, *i.e.*, using an event camera to capture the intensity changes on a purely diffuse sphere, which is placed in a dark room, for the split second of turning the light on. We build the analytic formulation of event streams through the radiant intensity. We then propose an optimization-based method and a learning-based method for lighting estimation. The comparison with the state-of-the-art method demonstrates that our methods not only significantly alleviate the intensity-distance ambiguity but also achieve superior performance for lighting estimation.

Limitations. Although our methods alleviate the intensity-distance ambiguity for the indoor lighting estimation, they require a controlled environment such as the dark scene, single light source, and the action of turning light on. And our methods cannot be applied to natural lighting. The diffuse sphere also limits our methods to predict high-frequency lighting.

Acknowledgments

References

[1] Himanshu Akolkar, Sio Hoi Ieng, and Ryad Benosman. Real-time high speed motion prediction using fast aperture-robust event-driven visual flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2

[2] Mohammed Mutlaq Almatrafi, Raymond Baldwin, Kiyoharu Aizawa, and Keigo Hirakawa. Distance surface for event-based optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2

[3] Alexander Andreopoulos, Hirak J Kashyap, Tapan K Nayak, Arnon Amir, and Myron D Flickner. A low power, high throughput, fully event-based stereo system. In *Proc. of Computer Vision and Pattern Recognition*, 2018. 2

[4] Yin Bi, Aaron Chadha, Alhabib Abbas, Eirina Bourtsoulatze, and Yiannis Andreopoulos. Graph-based object classification for neuromorphic vision sensing. In *Proc. of Internatoinal Conference on Computer Vision*, 2019. 2

[5] Dan A Calian, Jean-François Lalonde, Paulo Gotardo, Tomas Simon, Iain Matthews, and Kenny Mitchell. From faces to outdoor light probes. In *Computer Graphics Forum*. Wiley Online Library, 2018. 2

[6] Dan A Calian, Kenny Mitchell, Derek Nowrouzezahrai, and Jan Kautz. The shading probe: Fast appearance acquisition for mobile ar. In *Proc. of ACM SIGGRAPH Asia*. 2013. 2

[7] Marco Cannici, Marco Ciccone, Andrea Romanoni, and Matteo Matteucci. A differentiable recurrent surface for asynchronous event-based data. In *Proc. of European Conference on Computer Vision*, 2020. 2

[8] Jonghyun Choi, Kuk-Jin Yoon, et al. Learning to super resolve intensity images from events. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[9] Paul Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. of ACM SIGGRAPH*. 2008. 1, 2

[10] Paul Debevec, Paul Graham, Jay Busch, and Mark Bolas. A single-shot light probe. In *Proc. of ACM SIGGRAPH*. 2012. 2

[11] Guillermo Gallego, Mathias Gehrig, and Davide Scaramuzza. Focus is all you need: Loss functions for event-based vision. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[12] Guillermo Gallego, Jon EA Lund, Elias Mueggler, Henri Rebecq, Tobi Delbruck, and Davide Scaramuzza. Event-based, 6-dof camera tracking from photometric depth maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 2

[13] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proc. of Computer Vision and Pattern Recognition*, 2018. 1, 2

[14] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *Proc. of Internatoinal Conference on Computer Vision*, 2019. 1, 2, 3, 4, 6, 7, 8

[15] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. Learning to predict indoor illumination from a single image. *arXiv preprint arXiv:1704.00090*, 2017. 1, 2

[16] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. Fast spatially-varying indoor lighting estimation. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 1, 2

[17] Daniel Gehrig, Antonio Loquercio, Konstantinos G Derpanis, and Davide Scaramuzza. End-to-end learning of representations for asynchronous event-based data. In *Proc. of Internatoinal Conference on Computer Vision*, 2019. 2

[18] Daniel Gehrig, Henri Rebecq, Guillermo Gallego, and Davide Scaramuzza. Asynchronous, photometric feature tracking using events and frames. In *Proc. of European Conference on Computer Vision*, 2018. 2

[19] Stamatios Georgoulis, Konstantinos Rematas, Tobias Ritschel, Mario Fritz, Tinne Tuytelaars, and Luc Van Gool. What is around the camera? In *Proc. of Internatoinal Conference on Computer Vision*, 2017. 2

[20] Stamatios Georgoulis, Konstantinos Rematas, Tobias Ritschel, Efstratios Gavves, Mario Fritz, Luc Van Gool, and Tinne Tuytelaars. Reflectance and natural illumination from single-material specular objects using deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 2

[21] Ross Girshick. Fast r-cnn. In *Proc. of Internatoinal Conference on Computer Vision*, 2015. 4

[22] Jin Han, Chu Zhou, Peiqi Duan, Yehui Tang, Chang Xu, Chao Xu, Tiejun Huang, and Boxin Shi. Neuromorphic camera guided high dynamic range imaging. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. of Computer Vision and Pattern Recognition*, 2016. 4

[24] Yannick Hold-Geoffroy, Akshaya Athawale, and Jean-François Lalonde. Deep sky modeling for single image outdoor lighting estimation. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[25] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Sunil Hadap, Emiliano Gambaretto, and Jean-François Lalonde. Deep outdoor illumination estimation. In *Proc. of Computer Vision and Pattern Recognition*, 2017. 2

[26] Lukas Hosek and Alexander Wilkie. An analytic model for full spectral sky-dome radiance. *ACM Transactions on Graphics (TOG)*, 2012. 2

[27] Lukáš HošekHošek and Alexander Wilkie. Adding a solar-radiance function to the hošek-wilkie skylight model. *IEEE computer graphics and applications*, 2013. 2

[28] Yuhuang Hu, Tobi Delbruck, and Shih-Chii Liu. Learning to exploit multiple vision modalities by using grafted networks. In *Proc. of European Conference on Computer Vision*, 2020. 2

[29] Katsushi Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1981. 1

[30] Zhe Jiang, Yu Zhang, Dongqing Zou, Jimmy Ren, Jiancheng Lv, and Yebin Liu. Learning event-based motion deblurring. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[31] Kevin Karsch, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, Hailin Jin, Rafael Fonte, Michael Sittig, and David Forsyth. Automatic scene inference for 3d object compositing. *ACM Transactions on Graphics (TOG)*, 2014. 1

[32] Hyun-Ho Kim, Sang-Hyun Choi, Sang-Hyun Shin, Young-Ki Lee, Seok-Moon Choi, and Sung Yi. Thermal transient characteristics of die attach in high power led pkg. *Micro-electronics Reliability*, 2008. 2

[33] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Estimating natural illumination from a single outdoor image. In *Proc. of Internatoinal Conference on Computer Vision*. IEEE, 2009. 2

[34] Chloe LeGendre, Wan-Chun Ma, Graham Fyffe, John Flynn, Laurent Charbonnel, Jay Busch, and Paul Debevec. Deeplight: Learning illumination for unconstrained mobile mixed reality. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 1, 2

[35] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 1, 2

[36] Miao Liao, Liang Wang, Ruigang Yang, and Minglun Gong. Light fall-off stereo. In *Proc. of Computer Vision and Pattern Recognition*. IEEE, 2007. 1, 2

[37] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. A 128 128 120 db 15 s latency asynchronous temporal contrast vision sensor. *IEEE JOURNAL OF SOLID-STATE CIRCUITS*, 2008. 1, 2, 3

[38] Songnan Lin, Jiawei Zhang, Jinshan Pan, Zhe Jiang, Dongqing Zou, Yongtian Wang, Jing Chen, and Jimmy Ren. Learning event-driven video deblurring and interpolation. In *Proc. of European Conference on Computer Vision*, 2020. 2

[39] Daqi Liu, Alvaro Parra, and Tat-Jun Chin. Globally optimal contrast maximisation for event-based motion estimation. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[40] Qianhui Liu, Gang Pan, Haibo Ruan, Dong Xing, Qi Xu, and Huajin Tang. Unsupervised aer object recognition based on multiscale spatio-temporal features and spiking neurons. *IEEE transactions on neural networks and learning systems*, 2020. 2

[41] Qianhui Liu, Haibo Ruan, Dong Xing, Huajin Tang, and Gang Pan. Effective aer object classification using segmented probability-maximization learning in spiking neural networks. In *AAAI*, 2020. 2

[42] Stephen Lombardi and Ko Nishino. Reflectance and illumination recovery in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015. 2

[43] Jacques Manderscheid, Amos Sironi, Nicolas Bourdis, Davide Migliore, and Vincent Lepetit. Speed invariant time surface for learning to detect corner points with event-based cameras. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[44] Ana I Maqueda, Antonio Loquercio, Guillermo Gallego, Narciso García, and Davide Scaramuzza. Event-based vision meets deep learning on steering prediction for self-driving cars. In *Proc. of Computer Vision and Pattern Recognition*, 2018. 2

[45] Moustafa Meshry, Dan B Goldman, Sameh Khamis, Hugues Hoppe, Rohit Pandey, Noah Snavely, and Ricardo Martin-Brualla. Neural rerendering in the wild. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[46] Anton Mitrokhin, Zhiyuan Hua, Cornelia Fermuller, and Yiannis Aloimonos. Learning visual motion segmentation using event surfaces. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[47] Liyuan Pan, Miaomiao Liu, and Richard Hartley. Single image optical flow estimation with an event camera. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[48] Liyuan Pan, Cedric Scheerlinck, Xin Yu, Richard Hartley, Miaomiao Liu, and Yuchao Dai. Bringing a blurry frame alive at high frame-rate with an event camera. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[49] Federico Paredes-Vallés, Kirk Yannick Willehm Scheper, and Guido Cornelis Henricus Eugene De Croon. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[50] Jeong Joon Park, Aleksander Holynski, and Steven M Seitz. Seeing the world in a bag of chips. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[51] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 2019. 5

[52] Bharath Ramesh, Hong Yang, Garrick Michael Orchard, Ngoc Anh Le Thi, Shihao Zhang, and Cheng Xiang. Dart: distribution aware retinal transform for event-based cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[53] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. Events-to-video: Bringing modern computer vision to event cameras. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[54] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[55] Alexander G Schwing and Raquel Urtasun. Efficient exact inference for 3d indoor scene understanding. In *Proc. of European Conference on Computer Vision*, 2012. 1

[56] Yusuke Sekikawa, Kosuke Hara, and Hideo Saito. Eventnet: Asynchronous recursive event processing. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[57] B Shi, Z Mo, Z Wu, D Duan, SK Yeung, and P Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[58] Amos Sironi, Manuele Brambilla, Nicolas Bourdis, Xavier Lagorce, and Ryad Benosman. Hats: Histograms of averaged time surfaces for robust event-based object classification. In *Proc. of Computer Vision and Pattern Recognition*, 2018. 2

[59] Shuran Song and Thomas Funkhouser. Neural illumination: Lighting prediction for indoor environments. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 1, 2

[60] Suvrit Sra, Sebastian Nowozin, and Stephen J Wright. *Optimization for machine learning*. Mit Press, 2012. 4

[61] Pratul P Srinivasan, Ben Mildenhall, Matthew Tancik, Jonathan T Barron, Richard Tucker, and Noah Snavely. Lighthouse: Predicting lighting volumes for spatially-coherent illumination. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[62] Timo Stoffregen, Guillermo Gallego, Tom Drummond, Lindsay Kleeman, and Davide Scaramuzza. Event-based motion segmentation by motion compensation. In *Proc. of Internatoinal Conference on Computer Vision*, 2019. 1, 2

[63] Timo Stoffregen, Cedric Scheerlinck, Davide Scaramuzza, Tom Drummond, Nick Barnes, Lindsay Kleeman, and Robert Mahony. Reducing the sim-to-real gap for event cameras. In *Proc. Eur. Conf. Comput. Vis.*, 2020. 2

[64] Stepan Tulyakov, Francois Fleuret, Martin Kiefel, Peter Gehler, and Michael Hirsch. Learning an event sequence embedding for dense event-based deep stereo. In *Proc. of Internatoinal Conference on Computer Vision*, 2019. 1, 2

[65] Jonas Unger, Stefan Gustavson, and Anders Ynnerman. Spatially varying image based lighting by light probe sequences. *The Visual Computer*, 2007. 2

[66] Lin Wang, Yo-Sung Ho, Kuk-Jin Yoon, et al. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[67] Lin Wang, Tae-Kyun Kim, and Kuk-Jin Yoon. Eventsr: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[68] Yanxiang Wang, Bowen Du, Yiran Shen, Kai Wu, Guangrong Zhao, Jianguo Sun, and Hongkai Wen. Ev-gait: Event-based robust gait recognition using dynamic vision sensors. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[69] Yuanhao Wang, Ramzi Idoughi, and Wolfgang Heidrich. Stereo event-based particle tracking velocimetry for 3d fluid flow reconstruction. In *Proc. of European Conference on Computer Vision*, 2020. 2

[70] Zihao W Wang, Peiqi Duan, Oliver Cossairt, Aggelos Katsaggelos, Tiejun Huang, and Boxin Shi. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[71] Henrique Weber, Donald Prévost, and Jean-François Lalonde. Learning to estimate indoor lighting from 3d objects. In *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018. 6, 7

[72] Wikipedia contributors. Rise time — Wikipedia, the free encyclopedia, 2020. [Online; accessed 24-March-2021]. 3

[73] Wikipedia contributors. Angular diameter — Wikipedia, the free encyclopedia, 2021. [Online; accessed 25-March-2021]. 4

[74] Lan Xu, Weipeng Xu, Vladislav Golyanik, Marc Habermann, Lu Fang, and Christian Theobalt. Eventcap: Monocular 3d capture of high-speed human motions using an event camera. In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[75] Renjiao Yi, Chenyang Zhu, Ping Tan, and Stephen Lin. Faces as lighting probes via unsupervised deep highlight extraction. In *Proc. of European Conference on Computer Vision*, 2018. 2

[76] Jinsong Zhang, Kalyan Sunkavalli, Yannick Hold-Geoffroy, Sunil Hadap, Jonathan Eisenman, and Jean-François Lalonde. All-weather deep outdoor lighting estimation. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2

[77] Qian Zheng, Jinnan Chen, Zhan Lu, Boxin Shi, Xudong Jiang, Kim-Hui Yap, Ling-Yu Duan, and Alex C Kot. What does plate glass reveal about camera calibration? In *Proc. of Computer Vision and Pattern Recognition*, 2020. 2

[78] Qian Zheng, Ajay Kumar, Boxin Shi, and Gang Pan. Numerical reflectance compensation for non-lambertian photometric stereo. *IEEE Transactions on Image Processing*, 2019. 1

[79] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. Semi-dense 3d reconstruction with a stereo event camera. In *Proc. of European Conference on Computer Vision*, 2018. 2

[80] Alex Zihao Zhu, Yibo Chen, and Kostas Daniilidis. Realtime time synchronized event-based stereo. In *Proc. of European Conference on Computer Vision*, 2018. 2

[81] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based learning of optical flow, depth, and egomotion. In *Proc. of Computer Vision and Pattern Recognition*, 2019. 2, 4