Event-ID: Intrinsic Decomposition Using an Event Camera

Anonymous Authors

ABSTRACT

Reconstructing 3D scenes from multi-view images is challenging, especially in adverse conditions. We propose a novel event-based intrinsic decomposition framework that leverages events and images for stable decomposition under extreme scenarios. Our method is based on two observations: event cameras maintain good imaging quality, and events from different viewpoints exhibit similarity in diffuse regions while varying in specular regions. We establish an event-based reflectance model and introduce an event-based warping method to extract specular clues. Our two-part framework constructs a radiance field and decomposes the scene into normal, material, and lighting. Experimental results demonstrate superior performance compared to state-of-the-art methods. Our contributions include an event-based reflectance model, event warping-based consistency learning, and a framework for event-based intrinsic decomposition.

CCS CONCEPTS

• **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

KEYWORDS

Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

1 INTRODUCTION

Intrinsic decomposition from multi-view images is a fundamental task, that enables a range of downstream applications including view synthesis [5, 7, 62, 71], relighting [2, 5, 7, 21, 24, 39, 62, 73], object insertion [4, 19, 73], and digital heritage preservation [58]. Yet, present methodologies [5, 6, 32, 41, 51, 62, 75, 78, 83, 84] presuppose that the input images are sharp and well-exposed.

During the execution of downstream tasks, practitioners encounter various and demanding environments. These may involve scenarios such as reconstructing murals [64] in low-light caves, capturing scenes affected by severe camera shake-induced blur, or dealing with a high dynamic range that leads to over-exposed images. Under these conditions, the compromised image quality impedes the effectiveness of conventional methods, which struggle to discern precise cues for intrinsic decomposition, such as detailed gradient information and color distribution.

ACM MM, 2024, Melbourne, Australia

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

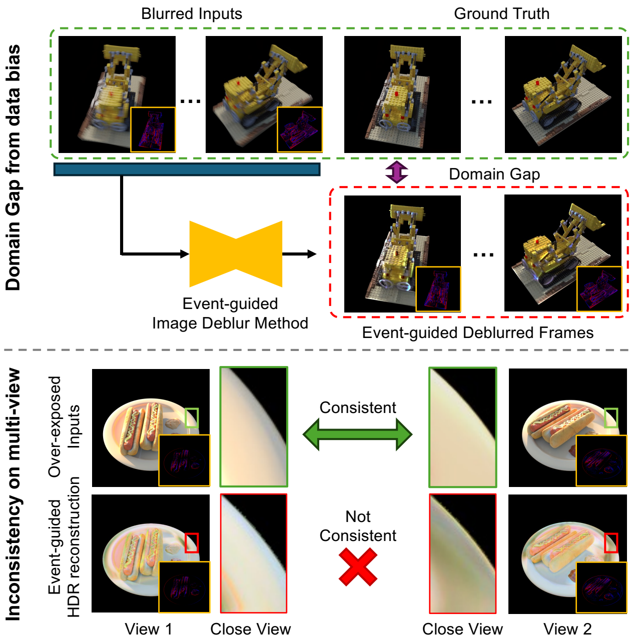

Figure 1: Demonstration of two primary challenges posed by event-guided techniques. (Top) The domain gap between predictions from event-guided image deblurring and ground truth. While existing methods [69] enhance clarity in blurred inputs, they often introduce significant color shifts that can impact intrinsic decomposition results. (Bottom) Multi-view inconsistency arises from frame-by-frame event-guided HDR reconstruction [74]. The method neglects 3D consistency during HDR recovery, focusing solely on time-related event information. This lack of 3D consistency increase the ill-posedness of intrinsic decomposition and might severely compromise performance.

Event cameras [16], with their high dynamic range and high temporal resolution capabilities, outperform traditional cameras in maintaining quality imaging under adverse conditions. The growing ubiquity of hybrid camera technology [10, 68], which integrates both events and images, paves the way for leveraging event information to substantially improve image quality. However, using event-guided methods [33, 35, 74] to enhance the input images does not yield satisfactory results (shown in Fig. 1). Firstly, most of these methods are data-driven, and when the test data has a domain gap with the training data, it can easily affect the results. Secondly, these methods typically process images on a frame-by-frame basis, which results in a loss of 3D consistency across multi-views [77].

To achieve this, we exploit the two primary attributes of events: 1) Event camera still maintains good imaging quality in these extreme scenarios[1]; 2) Events generated from different viewpoints

[1] Event camera is capable of capturing color information [57].

are relatively similar in diffuse regions, while in specular regions, they vary with changes in viewpoint. Based on the first point, we established an event-based reflectance model that captures the relationship between geometry, materials, lighting, and events in a 3D scene. Besides, with the second point, we introduce an event-based warping method that extracts specular-related clues from events to mitigate the impact of specular reflection on the intrinsic decomposition task.

To the end, based on the proposed reflectance model and specular-related clues from events, we propose an **Event**-based **I**ntrinsic **D**ecomposition method (**Event-ID**) taking the inputs of both events and images. In specific, we employ four separate Multilayer Perceptrons (MLPs) to represent the geometry, the surface normals, the material properties of the surface, and the scene lighting. Besides, we additionally modeling the specular variation on the changing of events to enable the specular-aware method, for recovering the challenge geometry and base color information on over-saturated specular region. Our method achieves state-of-the-art results under low-light conditions, overexposure, and when the input image is blurry. Compared to other methods, our approach can perform stable intrinsic decomposition even under extreme conditions. The main contributions of this paper can be summarized as follows:

(1) We build an event-based reflectance model that establishes the relationship between the geometry, materials, lighting, and events within a 3d scene.
(2) Observing the multi-view consistency of events, we extract specular-related clues from events and apply these clues to assist in intrinsic decomposition.
(3) Leveraging the event-based reflectance model and specular-related clues derived from events, we propose a framework for intrinsic decomposition using events, which enables relighting under extreme conditions.

2 RELATED WORKS

Intrinsic decomposition from multi-view images. There are several works [4, 26, 37, 43, 52, 59, 80] on extracting scene geometry, materials, and lighting from multi-view images. Here, we focus on introducing the efforts related to neural representation [5, 6, 32, 41, 51, 62, 75, 78, 83, 84]. NeRFactor [83] recovers an object's shape and reflectance from multi-view images under a single unknown lighting condition. TensoIR [32] model secondary shading effects (like shadows and indirect lighting). PhySG [81] and NeILF++ [78] obtain surface normal by optimizing the Signed Distance Field (SDF). Building upon the foundation of NeILF [75], NeILF++ [78] incorporates inter-object reflections, yielding more accurate geometric information. GS-IR [41] employs 3D Gaussian Splatting for the estimation of scene geometry, surface material, and environmental illumination.

These methods typically require input images to be clear and well-exposed, resulting in subpar performance in scenarios where the image quality is compromised. Our approach capitalizes on the high temporal resolution and high dynamic range of events, ensuring stable performance even in challenging lighting conditions.

Warping-based Consistency Learning. Warping-based consistency learning is commonly employed in multi-view stereo [65, 70, 72, 79], as well as in neural implicit surface learning [13, 15, 20], for the purpose of 3D reconstruction. It harnesses the power of inter-image correspondences through differentiable warping operations to enhance the reconstruction process. NeuralWarp [13] applies to warp to points sampled along a ray towards source images for RGB value extraction. Geo-Neus [15] warps grayscale patches around the anticipated surface points to adjacent images, ensuring multi-view geometric consistency. Ref-NeuS [20] proposes leveraging an anomaly detection framework to compute an explicit reflection score, utilizing multi-view contextual information to accurately identify and localize reflective surfaces within the scene. When the imaging quality of the image degrades, the performance of these methods tends to diminish.

Current warping methods based on events are predominantly used for optical flow estimation [17, 18, 61, 85], motion estimation [1, 11, 34, 85], and event alignment [18, 23, 63]. These methods focus on 2D warping and alignment, without addressing the alignment of 3D scenes. Moreover, these warping techniques are grounded in contrast maximization approaches [18, 63], which all rely on the assumption of photometric consistency and fail to consider the specularities present within the events. In contrast, our method exploits the inconsistency of events between diffuse and specular regions to extract specular-related information from 3D scenes. Compared to image-based methods, my approach remains robust and yields stable results even in scenarios with poor image quality.

Events for geometry, lighting, and material estimation. Event cameras, characterized by their high temporal resolution and high dynamic range, have been primarily utilized in previous research for depth estimation within scenes [18, 27, 48, 53, 85]. Additionally, there are efforts focused on reconstructing the three-dimensional shape of objects [3]. [22, 49] takes advantage of the high temporal resolution of event cameras by combining polarization information with events to estimate shape. Within the field of lighting estimation using event cameras, [12] exploits the high temporal resolution characteristic of events to capture the brightness changes of objects at the instant the light is switched on, in order to estimate the distance of the light source. To date, there has not been work on utilizing event cameras to estimate material properties.

To the best of our knowledge, we are the first to use event cameras for the estimation of geometry, materials, and lighting simultaneously.

Events for neural radiance fields. Recent research [31, 36, 42, 44, 54, 57] has sought to integrate events with Neural Radiance Fields (NeRF) to improve neural radiance fields reconstruction. Given that event cameras only capture changes in intensity rather than absolute values, methods [31, 42, 57] that exclusively depend on events as inputs require the manual assignment of a background intensity value. This approach could potentially lead to inaccuracies in the intensity of the reconstructed scene. In response to this issue, other techniques [36, 44, 54] employ both images and events as inputs.

Although existing research has primarily focused on extracting radiance and geometric details from 3D scenes, they largely overlook the reconstruction of material properties and lighting conditions. Our method stands out by reconstructing the geometry from events while concurrently capturing detailed material attributes and lighting nuances.

3 MODELING

In this section, we present our event-based intrinsic decomposition framework. Section 3.1 describes the events-based reflectance model, which captures the relationship between events and the geometry, materials, and lighting in a 3D scene. In Section 3.2, we explain how to extract specular-related information from events and how this information can be utilized in the intrinsic decomposition task. Finally, Section 3.3 provides a detailed description of our event-based intrinsic decomposition framework.

3.1 Events-based Reflectance Model

Event generation model. An event $e = (\mathbf{u}, t, p)$ at pixel position $\mathbf{u} = (u, v)$ and time t with polarity $p \in \{-1, +1\}$ is generated when the logarithmic change of brightness I since the last event at the pixel $\mathbf{x}$ and time $t - \Delta\mathrm{t}$ exceeds a threshold Θ ($\Theta > 0$) and the event e can be represented [55] as:

$$e(\mathbf{u}, t) = \left\lfloor \frac{\ln(\mathrm{I}(\mathbf{u}, \mathrm{t})) - \ln(\mathrm{I}(\mathbf{u}, \mathrm{t} - \Delta\mathrm{t}))}{\Theta} \right\rfloor. \tag{1}$$

Event-based reflectance model. From Equation (1), we can find that an event represents a change in brightness caused by the movement of the viewpoint in world space[2] within time Δt. As shown in Figure 2, We assume that at time t, an event camera captures the brightness of point $\mathbf{x}_t$ in the 3D scene at pixel $\mathbf{u}$, and at time $t - \Delta t$, the brightness at pixel $\mathbf{u}$ (image space) corresponds to the brightness of point $\mathbf{x}_{t-\Delta t}$ in the 3D scene (world space). The event-based reflectance model can be described as:

$$e(\mathbf{u}, t) = \left\lfloor \frac{\ln(\mathrm{L}(\omega_{\mathrm{o}}^{\mathrm{t}}, \mathbf{x}_{\mathrm{t}})) - \ln(\mathrm{L}(\omega_{\mathrm{o}}^{\mathrm{t}-\Delta\mathrm{t}}, \mathbf{x}_{\mathrm{t}-\Delta\mathrm{t}}))}{\Theta} \right\rfloor, \tag{2}$$

$$L(\omega_o^t, \mathbf{x}) = \int_{\Omega} f(\omega_o^t, \omega_i, \mathbf{x}) L_i(\omega_i, \mathbf{x})(\omega_i \cdot \mathbf{n}) d\omega_i, \tag{3}$$

where ω_o^t is the viewing direction of the outgoing light at time t, $\mathbf{n}$ is the surface normal, L_i is the incident light from direction ω_i, and f is the Bidirectional Reflectance Distribution Function [46] (BRDF) properties of the surface point $\mathbf{x}$.

For traditional images, there is a one-to-one correspondence between points in image space and points in world space. This means a pixel value represents the result of a physically-based rendering [56] process at a specific location in the 3D scene, taking into account the geometry, material, and lighting at that position. However, for event cameras, since an event describes a change in brightness, a position in image space corresponds to two points in world space. That is, a single event value is related to the physically-based rendering results of two different locations in the 3D scene, influenced by their respective geometry, material, and lighting conditions.

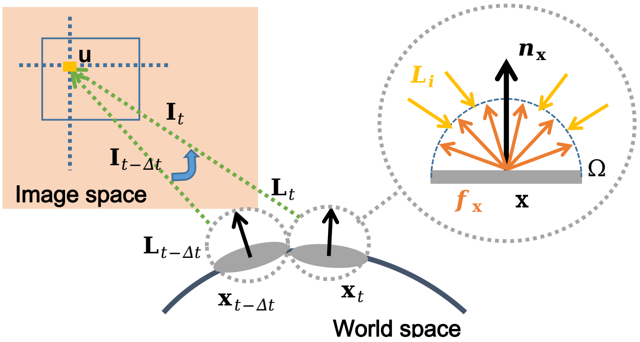
u
$\mathbf{I}_t$
$\mathbf{I}_{t-\Delta t}$
Image space
$\mathbf{L}_t$
$\mathbf{L}_{t-\Delta t}$
$\mathbf{x}_{t-\Delta t}$
$\mathbf{x}_t$
World space
L_i
$\boldsymbol{n}_{\mathbf{x}}$
Ω
$\boldsymbol{f}_{\mathbf{x}}$
$\mathbf{x}$

Figure 2: Illustration of events-based reflectance model. In image space, the generation of an event $e(\mathbf{u}, t)$ represents a change in brightness from position $\mathbf{x}_{t-\theta t}$ to position $\mathbf{x}_t$ in world space. The color at a position $\mathbf{x}$ in world space is determined by the surface normal $\mathbf{n}_x$, the BRDF properties f_x, and the incident light L_i at that position $\mathbf{x}$.

3.2 Event Clues for Intrinsic Decomposition

In 3D scenes, the diffuse component of objects exhibits the property of photometric consistency across different viewpoints, while the specular regions are view-dependent [29, 45, 86]. By leveraging this characteristic, it is possible to separate the diffuse and specular regions, which facilitates intrinsic decomposition [14, 38, 76].

Image-based methods often employ patch warping techniques to project image patches from different viewpoints onto the same viewpoint [13, 15, 20]. By computing the similarity between these warped patches, it is possible to determine whether a specular region exists at that particular viewpoint. However, when the quality of the images deteriorates (such as underexposure, overexposure, or blur), this warping approach [13, 15, 20] becomes ineffective. The degradation in image quality can lead to inaccuracies in the patch warping process, making it challenging to reliably compare the warped patches and identify specular regions. Consequently, the effectiveness of these image-based methods diminishes as the quality of the input images decreases (shown in Figure 3).

Compared to traditional cameras, event cameras offer unique advantages such as high dynamic range and high temporal resolution. These characteristics enable event cameras to maintain good image quality even in extreme lighting conditions and rapid motion scenarios where traditional cameras may struggle. Therefore, in extreme situations where traditional image-based methods may fail due to underexposure, overexposure, or motion blur, events provide a robust alternative for acquiring reliable specular clues (shown in Figure 3).

Event-based warping . As described in Section 3.1, for images, the brightness of each pixel corresponds one-to-one with a point in the world coordinates. However, an event represents the change in brightness of a pixel between time t and time $t - \Delta t$, which corresponds to two adjacent positions in the world space (shown in Figure 4). Due to the different imaging principles of images and events, warping methods designed for images cannot be directly applied to events. For event-based warping, when projecting

[2] In a static scene, this change in brightness is solely caused by the camera's motion.

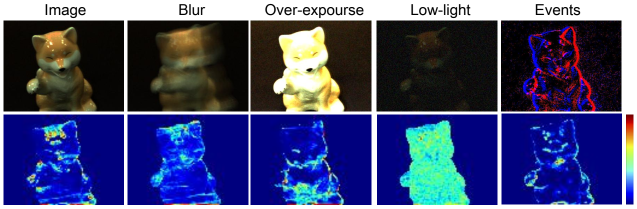

Figure 3: Illustration of specular reflection confidence in different scenes: (Top). Visualization of images / events from different scenes; (Bottom). The corresponding specular reflection confidence.

events from different viewpoints onto the same reference view, it is necessary to determine whether the two points in world space corresponding to the events from the source view are the same as the events in the reference view.

Because events are asynchronous and sparse, to reduce computational load, we slice the events along the temporal dimension, with events within the same slice sharing the same time. Since a single event carries too little information, to increase the effective information, we perform warping on pixel patches rather than on single pixel points.

Given a reference view event slice E_{ref} and a series of source view event slices $\{E^i_{\text{src}}\}$, our objective is to warp the source view event slices to the reference view. Shown in Figure 4, each event slice is associated with two positions in the world space, which we denote as $\mathbf{p}_{\{\text{A,B}\}}$. And these two positions have corresponding normal $\mathbf{n}_{\{\text{A,B}\}}$, rotation matrix $\mathbf{R}_{\{\text{A,B}\}}$, and translation vector $\mathbf{t}_{\{\text{A,B}\}}$. Like [15], the point x in the pixel patch q_{ref} of reference event slice E_{ref} is related to the corresponding point x' in the pixel patch q^i_{src} of the source event slice E^i_{src} via the plane-induced homography H [25]:

$$x = \mathbf{H}_{\text{A}}x', x = \mathbf{H}_{\text{B}}x', \tag{4}$$

$$\mathbf{H}_i = \mathbf{K}\left(\mathbf{R}^{\text{src}}_i(\mathbf{R}^{\text{ref}}_i)^T + \frac{\mathbf{R}^{\text{src}}_i((\mathbf{R}^{\text{src}}_i)^T t^{\text{src}}_i - (\mathbf{R}^{\text{ref}}_i)^T t^{\text{ref}}_i)\mathbf{n}^T_i}{\mathbf{n}^T_i\,\mathbf{p}_i}\right)\mathbf{K}^{-1}, \tag{5}$$

where $i \in \{A, B\}$, $\mathbf{K}$ donates the internal calibration matrix. We determine the validity of the source view event patch by calculating the distance between the center positions of the projections of two patches:

$$|\mathbf{H}_{\text{A}}x_{\text{center}} - \mathbf{H}_{\text{B}}x_{\text{center}}| < \frac{s}{2}, \tag{6}$$

where we set patch size s to 11. Figure 4 illustrates the correspondence of the same positions in world space as viewed from different perspectives.

Measuring the similarity between event patches. When measuring the similarity between image patches, previous works [15] convert RGB images to grayscale images and then use Normalized Cross Correlation (NCC) to assess the similarity between two grayscale image patches. However, since events are sparse, discrete values with polarity (+1, -1), and subject to noise, directly applying NCC to event patches may not be effective. To address this, we propose to accumulate events within a certain time interval and normalize the accumulated event patch. The specific steps are as follows:

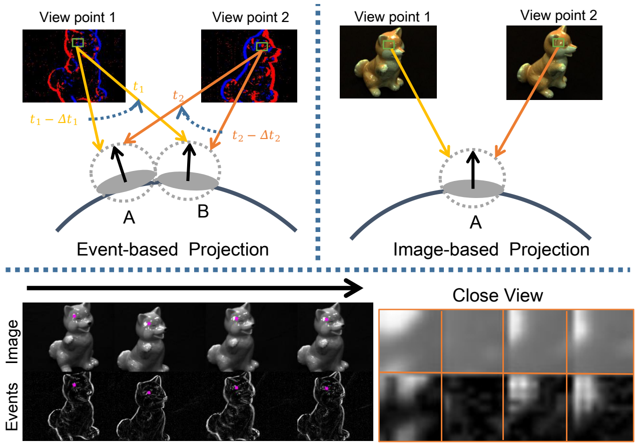

Figure 4: (Top). Illustration of the differences between event-based projection and image-based projection: For events, for two small adjacent areas A and B on the surface, the projection of A and B on the event camera is a small pixel patch over the time interval from $t - \Delta t$ to t. In contrast, for images, for a small area A on the surface, the projection of A on the image is a small pixel patch; (Bottom). The correspondence of the same world space location on event slices and images from different viewpoints.

$$E = \left|\frac{\sum_{\Delta t} e}{k}\right|, \tag{7}$$

where $\sum_{\Delta t} e$ represents the accumulated events within the specified time interval Δt, k is an empirically determined constant and we set k to 10. To mitigate the impact of noise in events on similarity measurement, we incorporate noise modeling when computing the NCC.

$$NCC(x, y) = \frac{\text{Cov}(x + \mathcal{N}(\mu, \sigma), y + \mathcal{N}(\mu, \sigma))}{\sqrt{\text{Var}(x + \mathcal{N}(\mu, \sigma)) \cdot \text{Var}(y + \mathcal{N}(\mu, \sigma))}}, \tag{8}$$

where $\mathcal{N}$ is a normal distribution with a mean μ of 0 and a standard deviation σ of 0.01.

Events clues. Based on the event-based warping and the metric for measuring similarity, we can obtain the similarity between the event patch of the reference view and a series of event patches from source views. Similar to [20], we believe that only the local region of the partial event patch contains specular reflections. Therefore, we use the variance of these similarities as the confidence measure for specular reflection,

$$M_s = \text{Var}(\{1 - NCC(E_{\text{ref}}, E^i_{\text{src}})\}), \tag{9}$$

The specular reflection confidence obtained from the events can serve as a clue for specular highlights, which helps mitigate the impact of specular reflections on intrinsic decomposition. Specifically, this confidence can be used as a soft mask, applied to the loss function to reduce the influence of specular regions. Compared to image-based approaches, event-based methods demonstrate stable performance even in extreme scenarios (shown in Figure 3).

3.3 Event-based Intrinsic Decomposition

The overview of our method is illustrated in Figure 5. We parameterize the 3D scene into four fields: the scene geometry is described by a signed distance field $\mathcal{G}$; surface normal is described by a normal field $\mathcal{N}$; material properties of the surface are described by a BRDF field $\mathcal{B}$; and the lighting of the scene is described by a light field $\mathcal{S}$. We optimize these four fields using both events and images.

Scene geometry. Following [78, 81], we represent the scene geometry as a Signed Distance Field (SDF). We obtain the position $\mathbf{x}$ of the object's surface points from this field, as well as the corresponding normal $\mathbf{n}_{\text{sdf}}$ and radiance $\mathbf{R}_{\text{vol}}$.

$$\mathcal{G} : \mathbf{x} \longrightarrow \{\mathbf{d}, \mathbf{n}_{\text{sdf}}, \mathbf{R}_{\text{vol}}\}, \tag{10}$$

where $\mathbf{d}$ represents the signed distance from the point $\mathbf{x}$ to the closest surface. Similar to [66], this field is optimized using a volume rendering technique. Due to the poor image quality in extreme scenarios, events tend to yield better imaging results compared to traditional images. Similar to [57], we utilize events as a supervisory signal to guide the optimization process, ensuring that the background color in the volume rendering closely matches the background color observed in the images. The loss $\mathcal{L}_{\text{vol}}$ can be written as,

$$\mathcal{L}_{\text{vol}} = ||M_{\text{color}}\Theta \sum_{\Delta t} e_{\mathbf{u}} - M_{\text{color}} E_{\mathbf{u}}||_2, \tag{11}$$

$$E_{\mathbf{u}} = \ln(\mathbf{R}_{\text{vol}}(\mathbf{u}, \text{t})) - \ln(\mathbf{R}_{\text{vol}}(\mathbf{u}, \text{t} - \Delta\text{t})), \tag{12}$$

where $\mathbf{R}_{\text{vol}}$ is the color output of volume rendering[3], M_{color} is described in [57], which is used to map the color information of the events onto the three RGB channels.

Surface normals of the scene. The surface normals within a 3D scene can be formulated as a normal field, which is encoded by a multi-layer perceptron (MLP). This MLP takes a point location $\mathbf{x}$ as input and outputs a surface normal $\mathbf{n}$.

$$\mathcal{N} : \mathbf{x} \longrightarrow \mathbf{n}, \tag{13}$$

We use the normal $\mathbf{n}_{\text{vol}}$ output by $\mathcal{G}$ as the supervisory signal and loss $\mathcal{L}_{\text{normal}}$ can be written as,

$$\mathcal{L}_{\text{normal}} = ||M_s \mathbf{n} - M_s \mathbf{n}_{\text{sdf}}||_2, \tag{14}$$

where M_s is described by Equation (9).

Material properties of the surface. Following [75, 78], the material properties of the surface can be formulated as a BRDF field, which is encoded by a MLP. This MLP takes a point location $\mathbf{x}$ as input and outputs a base color $\mathbf{b}$, a roughness r and a metallic m.

$$\mathcal{B} : \mathbf{x} \longrightarrow \{\mathbf{b}, r, m\}, \tag{15}$$

We use a simplified Disney principled BRDF [8] parameterization. The BRDF f in Equation (3) can be computed as,

$$f(\omega_o, \omega_i, \mathbf{n}) = f_d + f_s(\omega_o, \omega_i, \mathbf{n}), \tag{16}$$

$$f_d = \frac{1 - m}{\pi} \cdot \mathbf{b}, \tag{17}$$

$$f_s(\omega_o, \omega_i, \mathbf{n}) = \frac{D(h; r) \cdot F(\omega_o, h; \mathbf{b}, m) \cdot G(\omega_i, \omega_o, h; r)}{4 \cdot (\mathbf{n} \cdot \omega_i) \cdot (\mathbf{n} \cdot \omega_o)}, \tag{18}$$

where f_d is diffuse term, f_s is specular term, h is the half vector between the incident direction ω_i and the viewing direction ω_o. D, F, and G refer to the normal distribution function, the Fresnel term, and the geometry term respectively[4]. The loss $\mathcal{L}_{\text{base\_color}}$ can be written as,

$$\mathcal{L}_{\text{base\_color}} = ||M_s \frac{\mathbf{b}}{||\mathbf{b}||_2} - M_s \frac{\mathbf{R}_{\text{vol}}}{||\mathbf{R}_{\text{vol}}||_2}||_2, \tag{19}$$

Lighting of the scene. Like [75, 78], the incoming lights in the scene can be formulated as a neural incident light field, which is recorded by a MLP. The MLP takes a point location $\mathbf{x}$ and a direction ω_i as inputs, and returns an incident light $\mathbf{L}_i$,

$$\mathcal{S} : \{\mathbf{x}, \omega_i\} \longrightarrow \mathbf{L}_i, \tag{20}$$

Event-based surface rendering equation. Due to events being sparse and asynchronous, and containing noise, to improve efficiency and reduce the impact of noise, we employ the same slicing method as in section 3.2 for surface rendering. The surface rendering equation for events established in section 3.1 can be reformulated as follows,

$$E_{\mathbf{u}} = \ln(\mathbf{I}(\mathbf{u}, \text{t})) - \ln(\mathbf{I}(\mathbf{u}, \text{t} - \Delta\text{t})), \tag{21}$$

where $\mathbf{I}(\mathbf{u}, t)$ is described by Equation (3). And the $E_{\mathbf{u}}$ can be supervisd by events generated at $\mathbf{u}$, the loss $\mathcal{L}_{\text{surf}}$ is,

$$\mathcal{L}_{\text{surf}} = ||M_{\text{color}} E_{\mathbf{u}} - M_{\text{color}}\Theta \sum_{\Delta t} e_{\mathbf{u}}||_1. \tag{22}$$

Since Equation (21) represents a differential relationship, we use the radiance $\mathbf{R}_{\text{vol}}$ output from the first stage as an additional supervisory signal to constrain radiance $\mathbf{I}(\mathbf{u}, t)$.

$$\mathcal{L}_{\text{surf\_vol}} = ||\mathbf{I}(\mathbf{u}, t) - \mathbf{R}_{\text{vol}}(\mathbf{u}, t)||_1. \tag{23}$$

Optimization strategy. Like [78], our strategy is divided into two distinct optimization stages for those fields. Initially, we employ a combination of events and images to optimize the scene's geometric field $\mathcal{G}$. Following this, we enter the second stage, wherein $\mathcal{G}$ is freezed, allowing us to focus on the simultaneous optimization of the normal field $\mathcal{N}$, the BRDF field $\mathcal{B}$, and the light field $\mathcal{S}$.

4 EXPERIMENTS

In this section, we evaluate our proposed method using both synthetic and real-world datasets. We conduct ablation studies to validate the effectiveness of each module.

4.1 Datasets and Metric

Synthetic data. To compensate for the absence of a dataset comprising events, images (including overexposed, blurred, and low-light), base color GT, and normal GT. We augment the synthetic dataset from NeRF synthetic dataset [47] by extending it with seven synthetic scenes, *i.e.*, chair, drums, ficus, hotdog, lego, materials, and mic. First, we utilize Blender [50] to synthesize 1000 continuous HDR images, and composite events using the same method as [74].

[3] More details are provided in the supplementary material.

[4] We adopt similar implementation of D, F, and G as in previous works [75, 78] and details are provided in the supplementary material.

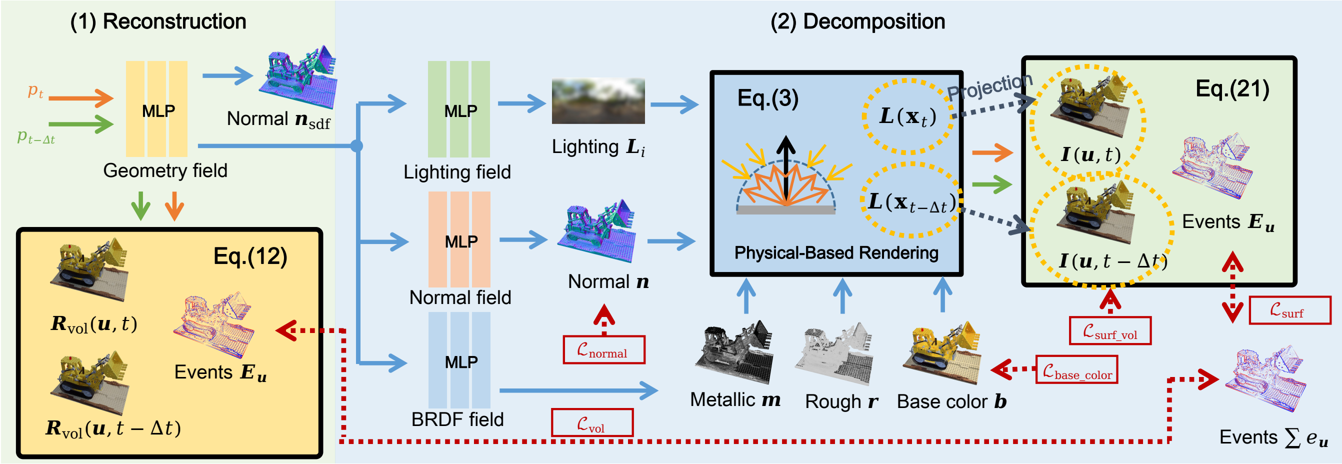
(1) Reconstruction
(2) Decomposition
p_t
$p_{t-\Delta t}$
MLP
Normal $\boldsymbol{n}_{\text{sdf}}$
Geometry field
Eq.(12)
$\boldsymbol{R}_{\text{vol}}(\boldsymbol{u}, t)$
Events $\boldsymbol{E}_{\boldsymbol{u}}$
$\boldsymbol{R}_{\text{vol}}(\boldsymbol{u}, t-\Delta t)$
MLP
Lighting field
Lighting $\boldsymbol{L}_i$
MLP
Normal field
Normal $\boldsymbol{n}$
MLP
BRDF field
$\mathcal{L}_{\text{normal}}$
$\mathcal{L}_{\text{vol}}$
Eq.(3)
$\boldsymbol{L}(\mathbf{x}_t)$
$\boldsymbol{L}(\mathbf{x}_{t-\Delta t})$
Physical-Based Rendering
Projection
Metallic $\boldsymbol{m}$
Rough $\boldsymbol{r}$
Base color $\boldsymbol{b}$
$\mathcal{L}_{\text{base\_color}}$
Eq.(21)
$\boldsymbol{I}(\boldsymbol{u}, t)$
$\boldsymbol{I}(\boldsymbol{u}, t-\Delta t)$
Events $\boldsymbol{E}_{\boldsymbol{u}}$
$\mathcal{L}_{\text{surf\_vol}}$
$\mathcal{L}_{\text{surf}}$
Events $\sum e_{\boldsymbol{u}}$

Figure 5: The overview of our method. The first part uses events and images to reconstruct a geometry field $\mathcal{G}$. The second part involves freezing the geometry field and using information from the geometry field along with events to reconstruct three fields: Lighting field $\mathcal{S}$, Nomral field $\mathcal{N}$ and BRDF field $\mathcal{B}$.

We uniformly sample 100 HDR images and adjust the exposure via method [30] to synthesize overexposed images. By employing method [40], we reduce the exposure and add noise to the images to synthesize underexposed images (low-light). Using method [54], we synthesized blurred images.

In addition to the aforementioned training data, we have also synthesized 200 testing images from different viewpoints than the training data, along with the corresponding ground truth for base color and normals. In summary, our synthetic dataset contains images under four different conditions, as well as the corresponding events. We will evaluate our results and SOTA methods on these three datasets.

Real-world data. To test our approach's robustness, we collected real data using the DAVIS 346C sensor [57] from three scenes, capturing 16-second monocular videos (each video is about 16 seconds and 500 frames) with events and frames. Frame poses were estimated using COLMAP [60]. Although the poses may not be perfectly accurate due to potential blurring, they were sufficiently estimated for our purposes. We uniformly selected 100 images to serve as the training data. Additionally, we randomly selected 100 images for testing.

Metric. Our evaluation encompasses normal quality (measured by MAE [9]), novel view synthesis, and base color (measured by PSNR [28], SSIM [28], and LPIPS [82])

4.2 Comparison with State-of-the-Art Methods

Due to significant differences in the scenes compared to those used in previous methods, some methods did not work in our scenarios. In the main text, we primarily focus on comparing state-of-the-art (SOTA) neural field-based intrinsic decomposition methods GD-IR [41] and its integration with other methods. The results of other methods will be presented in the supplementary materials.

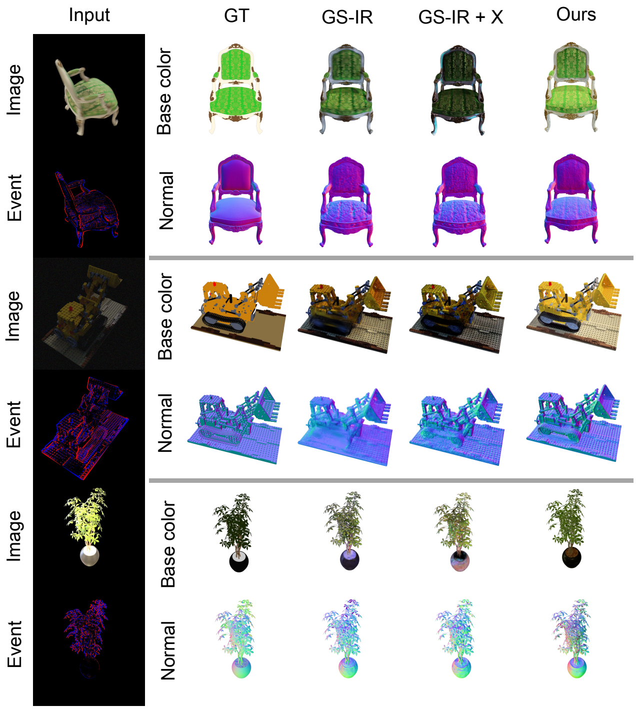

Figure 6: Qualitative comparison on our synthetic dataset. We visualize the estimated normal, base color results of our method and SOTA methods on three scenes (From top to bottom, they are blurred, and low-light, and overexposed respectively). In these three scenarios, X represents methods WZ23 [69] , WY21 [67] , and YH23 [74] respectively.

The results of the SOTA methods. Table 1 provides a quantitative comparison of synthetic data, highlighting that our approach maintains stable performance even under conditions of input image degradation. This resilience is attributed to the high temporal resolution and dynamic range benefits afforded by events.

Figure 6 displays the visualization results under three different types of image degradation. The degradation of images leads to inconsistencies in multi-view information, making it challenging for the network to extract useful information. However, our method circumvents the issues associated with image degradation by extracting scene information directly from events, thereby avoiding the pitfalls of degraded image quality.

The results of the combination of SOTA methods. To ensure a fair comparison given that our method simultaneously inputs events and images, we selected three event-guided methods to enhance images degraded by different factors: YH23 [74] for correcting overexposure [5], WZ23 [69] for deblurring, and WY21 [67] for enhancing low-light images. This approach allows us to evaluate the effectiveness of our framework against specialized methods tailored to address specific types of image degradation. Additionally, we compared methods that combine event-based NeRF method LL23 [42] with GS-IR [41][6]. These approaches are similar to ours, as they both extract valuable information from events and consider the consistency across multiple viewpoints.

Table 1 displays the quantitative results of the combined methods. It becomes clear that merely combining events with images straightforwardly fails to capitalize on the unique advantages of events effectively. This inefficiency stems from the event-enhanced method of applying modifications on a frame-by-frame basis, which leads to the loss of multi-view consistency in the enhanced images. Additionally, the methods that combine event-based NeRF with GS-IR [41] perform better than those that simply use events to enhance images. This improvement is due to event-based NeRF's ability to extract scene information from events while maintaining multi-view consistency. However, this approach only utilizes event information during the reconstruction of the 3D scene and does not fully exploit it during the decomposition phase. In contrast, our method makes full use of event information in both the 3D scene reconstruction and decomposition stages, resulting in superior outcomes.

Figure 7 showcases the visual results, clearly illustrating that our method achieves the best performance.

The results of real-world data. Figure 8 presents the results on real-world data which, unlike synthetic data, are constrained by device capabilities, resulting in images with a resolution of only 346x260 and significantly poorer image quality. Despite the low quality of the images, our results still outperform comparative methods. For the results of the remaining real-world data, please refer to the supplementary materials.

4.3 Ablation Studies

To validate the effectiveness and necessity of the components of our method, we compare it with its two variants:

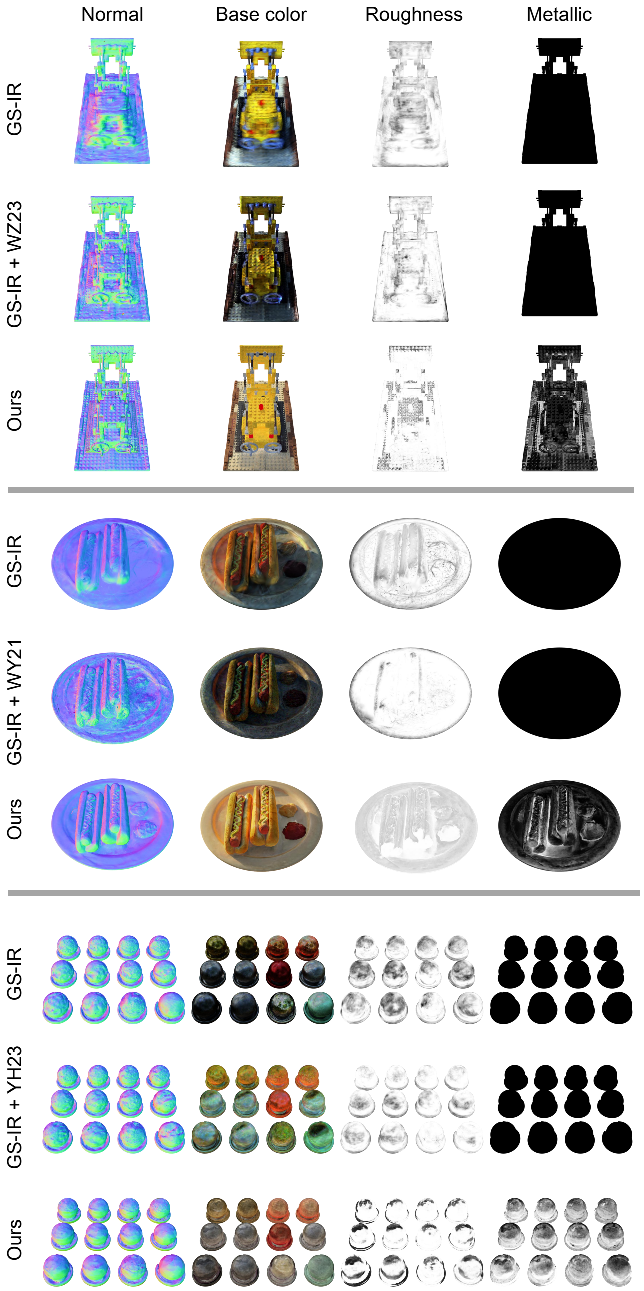

Figure 7: Qualitative comparison on synthetic dataset. We visualize the estimated normal, base color, roughness and metallic of our Event-ID and baseline methods. (From top to bottom: scenes with blur, low light, and overexposure.)

- **Variant A:** Without event surface rendering loss $\mathcal{L}_{\text{surf}}$.
- **Variant B:** Without specular-related cues M_s.

Table 2 presents the quantitative results for different variants. It is clear that removing any component of our method results in decreased performance.

[5] We use HDR images as the input for the intrinsic decomposition method.
[6] The images output by LL23 [42] are used as input for GS-IR [41].

Table 1: Quantitative comparison on synthetic dataset.

Conditions	Methods	Normal	Base Color			Novel View		
		MAE(°) ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Over-exposure	GS-IR [41]	28.1769	15.3427	0.8498	0.1675	18.5523	0.9156	0.0901
	GS-IR [41]+YH23 [74]	28.2794	15.7839	0.8558	0.1594	15.1489	0.8822	0.1259
Low light	GS-IR [41]	44.7076	16.2018	0.8461	0.1745	16.3147	0.8875	0.1120
	GS-IR [41]+WY21 [67]	33.0106	15.3934	0.8273	0.1748	23.4842	0.8987	0.1066
Blur	GS-IR [41]	35.5682	15.6107	0.8588	0.1832	22.6323	0.8916	0.1196
	GS-IR [41]+WZ23 [69]	28.9436	15.5308	0.8452	0.1727	23.5555	0.9026	0.0943
	GS-IR [41]+LL23 [42]	27.3428	17.1431	0.8403	0.1694	22.1696	0.8993	0.0899
	Ours	**20.1804**	**20.4238**	**0.8930**	**0.1140**	**25.3671**	**0.9222**	**0.0726**

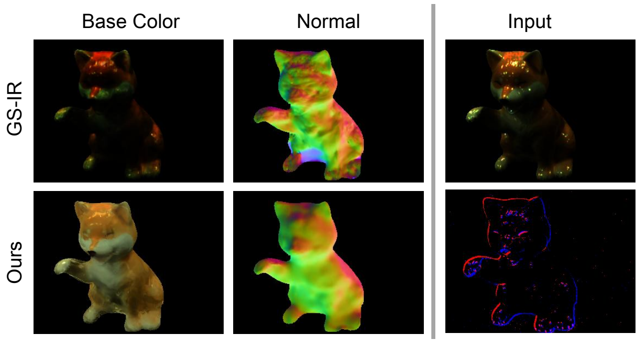

Figure 8: Qualitative comparison on real-world dataset in low-light scene. We visualize the estimated based color, normal, and input of our Event-ID and GS-IR [41].

Table 2: Quantitative comparison of ablation study on synthetic dataset.

Variants	Normal	Base Color			Novel View		
	MAE(°) ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
w/o $\mathcal{L}_{\text{surf}}$	26.0321	18.4492	0.8669	0.1261	23.9775	0.9076	0.0921
w/o M_s	23.7485	19.0964	0.8653	0.1194	24.8082	0.9147	0.0720
Ours	**20.1804**	**20.4238**	**0.8930**	**0.1140**	**25.3671**	**0.9222**	0.0726

Effectiveness of the clues from events. To verify the effectiveness of the specular-related cues obtained from events, we remove the M_s from the loss $\mathcal{L}_{\text{normal}}$ and $\mathcal{L}_{\text{base\_color}}$ as a part of our comparative experiment. Figure 9 and Figure 10 respectively show the results on base color estimation and normal estimation. It can be observed that utilizing specular-related clues to separate specular highlights is beneficial for the reconstruction of base color and normals in specular regions. The reason is that for base color, the specular regions reflect the color of the light source rather than the color of the object itself. Masking out the pixels in the specular positions is tantamount to reducing the input of incorrect information, allowing the network to be undisturbed by specular information. As for normal estimation, the normal is constrained by the output of the first part, and the normal output from the first stage is inaccurate in the specular areas due to not considering the effect of specular reflections. Lowering the weight of the normals in these regions also equates to reducing the input of incorrect information. Therefore, in the experimental results with M_s, the highlighted box areas exhibit a relatively smooth surface of normal.

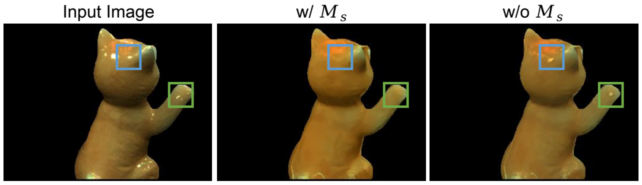

Figure 9: Visualization of ablation study on the effectiveness of specular-related clues (M_s) to base color estimation. The placement of the boxes highlights the differences between the two outcomes.

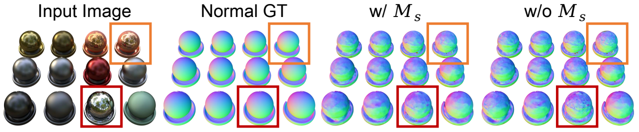

Figure 10: Visualization of ablation study on the effectiveness of specular-related clues (M_s) to normal estimation. The placement of the boxes highlights the differences between the two outcomes.

5 CONCLUSION & LIMITATIONS

In conclusion, our research presents a intrinsic decomposition framework that effectively harnesses event-based data to overcome the limitations of conventional imaging under adverse conditions. By integrating an event-based reflectance model with specular-related event clues, we have enabled robust 3D scene reconstruction and intrinsic decomposition in challenging scenarios such as low-light, blur, and overexposure. Our method not only maintains consistency across multiple viewpoints but also improves upon the reliability of intrinsic decomposition, paving the way for practical applications in fields requiring precise scene interpretation and manipulation. This work signifies a considerable advancement in the utilization of event cameras for complex imaging tasks and sets the stage for future explorations in the domain.

Limitations. Our method does not model shadows, and the current resolution of event cameras limits the application of this technique. However, these issues do not hinder the exploration of this scientific question, and we believe that with advancements in event camera technology and its commercial development, the resolution and imaging quality of event cameras will improve.

REFERENCES

[1] Himanshu Akolkar, Sio-Hoi Ieng, and Ryad Benosman. 2020. Real-time high speed motion prediction using fast aperture-robust event-driven visual flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 1 (2020), 361–372.

[2] Jonathan T Barron and Jitendra Malik. 2014. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence* 37, 8 (2014), 1670–1687.

[3] Alexis Baudron, Zihao W Wang, Oliver Cossairt, and Aggelos K Katsaggelos. 2020. E3d: event-based 3d shape reconstruction. *arXiv preprint arXiv:2012.05214* (2020).

[4] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 294–311.

[5] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. 2021. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12684–12694.

[6] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. 2021. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems* 34 (2021), 10691–10704.

[7] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik Lensch, and Jan Kautz. 2020. Two-shot spatially-varying brdf and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3982–3991.

[8] Brent Burley and Walt Disney Animation Studios. 2012. Physically-based shading at disney. In *Acm Siggraph*, Vol. 2012. vol. 2012, 1–7.

[9] Tianfeng Chai and Roland R Draxler. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development* 7, 3 (2014), 1247–1250.

[10] Yakun Chang, Chu Zhou, Yuchen Hong, Liwen Hu, Chao Xu, Tiejun Huang, and Boxin Shi. 2023. 1000 FPS HDR Video With a Spike-RGB Hybrid Camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22180–22190.

[11] Haosheng Chen, David Suter, Qiangqiang Wu, and Hanzi Wang. 2020. End-to-end learning of object motion estimation from retinal events for event-based object tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 10534–10541.

[12] Zehao Chen, Qian Zheng, Peisong Niu, Huajin Tang, and Gang Pan. 2021. Indoor lighting estimation using an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14760–14770.

[13] François Darmon, Bénédicte Bascle, Jean-Clément Devaux, Pascal Monasse, and Mathieu Aubry. 2022. Improving neural implicit surfaces geometry with patch warping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6260–6269.

[14] Graham D Finlayson, Mark S Drew, and Cheng Lu. 2009. Entropy minimization for shadow removal. *International Journal of Computer Vision* 85, 1 (2009), 35–57.

[15] Qiancheng Fu, Qingshan Xu, Yew Soon Ong, and Wenbing Tao. 2022. Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. *Advances in Neural Information Processing Systems* 35 (2022), 3403–3416.

[16] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. 2020. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence* 44, 1 (2020), 154–180.

[17] Guillermo Gallego, Mathias Gehrig, and Davide Scaramuzza. 2019. Focus is all you need: Loss functions for event-based vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12280–12289.

[18] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. 2018. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3867–3876.

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

[20] Wenhang Ge, Tao Hu, Haoyu Zhao, Shu Liu, and Ying-Cong Chen. 2023. Ref-neus: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4251–4260.

[21] Dan B Goldman, Brian Curless, Aaron Hertzmann, and Steven M Seitz. 2009. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 6 (2009), 1060–1071.

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

[23] Cheng Gu, Erik Learned-Miller, Daniel Sheldon, Guillermo Gallego, and Pia Bideau. 2021. The spatio-temporal poisson point process: A simple model for the alignment of event camera data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13495–13504.

[24] Tom Haber, Christian Fuchs, Philippe Bekaer, Hans-Peter Seidel, Michael Goesele, and Hendrik PA Lensch. 2009. Relighting objects from image collections. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 627–634.

[25] Richard Hartley and Andrew Zisserman. 2003. *Multiple view geometry in computer vision*. Cambridge university press.

[26] Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. 2022. Shape, light, and material decomposition from images using monte carlo rendering and denoising. *Advances in Neural Information Processing Systems* 35 (2022), 22856–22869.

[27] Javier Hidalgo-Carrió, Daniel Gehrig, and Davide Scaramuzza. 2020. Learning monocular dense depth from events. In *2020 International Conference on 3D Vision (3DV)*. IEEE, 534–542.

[28] Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*. IEEE, 2366–2369.

[29] Berthold KP Horn and Robert W Sjoberg. 1979. Calculating the reflectance map. *Applied optics* 18, 11 (1979), 1770–1779.

[30] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. 2022. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18398–18408.

[31] Inwoo Hwang, Junho Kim, and Young Min Kim. 2023. Ev-NeRF: Event based neural radiance field. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 837–847.

[32] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. 2023. Tensoir: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 165–174.

[33] Haiyan Jin, Qiaobin Wang, Haonan Su, and Zhaolin Xiao. 2023. Event-guided low light image enhancement via a dual branch GAN. *Journal of Visual Communication and Image Representation* 95 (2023), 103887.

[34] Daniel R Kepple, Daewon Lee, Colin Prepsius, Volkan Isler, Il Memming Park, and Daniel D Lee. 2020. Jointly learning visual motion and confidence from local patches in event cameras. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 500–516.

[35] Taewoo Kim, Jeongmin Lee, Lin Wang, and Kuk-Jin Yoon. 2022. Event-guided deblurring of unknown exposure time videos. In *European Conference on Computer Vision*. Springer, 519–538.

[36] Simon Klenk, Lukas Koestler, Davide Scaramuzza, and Daniel Cremers. 2023. E-nerf: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters* 8, 3 (2023), 1587–1594.

[37] Junxuan Li and Hongdong Li. 2022. Neural reflectance for shape recovery with shadow handling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16221–16230.

[38] Junxuan Li, Hongdong Li, and Yasuyuki Matsushita. 2021. Lighting, reflectance and geometry estimation from 360deg panoramic stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10591–10600.

[39] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. 2020. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2475–2484.

[40] Jinxiu Liang, Yixin Yang, Boyu Li, Peiqi Duan, Yong Xu, and Boxin Shi. 2023. Coherent Event Guided Low-Light Video Enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10615–10625.

[41] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. 2023. Gs-ir: 3d gaussian splatting for inverse rendering. *arXiv preprint arXiv:2311.16473* (2023).

[42] Weng Fei Low and Gim Hee Lee. 2023. Robust e-NeRF: NeRF from Sparse & Noisy Events under Non-Uniform Motion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 18335–18346.

[43] Zhan Lu, Qian Zheng, Boxin Shi, and Xudong Jiang. 2024. Pano-NeRF: Synthesizing High Dynamic Range Novel Views with Geometry from Sparse Low Dynamic Range Panoramic Images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 3927–3935.

[44] Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, and Luc Van Gool. 2023. Deformable Neural Radiance Fields using RGB and Event Cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3590–3600.

[45] Satya P Mallick, Todd Zickler, Peter N Belhumeur, and David J Kriegman. 2006. Specularity removal in images and videos: A PDE approach. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9*. Springer, 550–563.

[46] Stephen R Marschner, Stephen H Westin, Eric PF Lafortune, and Kenneth E Torrance. 2000. Image-based bidirectional reflectance distribution function measurement. *Applied optics* 39, 16 (2000), 2592–2600.

[47] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance

fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

[48] Mohammad Mostafavi, Kuk-Jin Yoon, and Jonghyun Choi. 2021. Event-intensity stereo: Estimating depth by the best of both worlds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 4258–4267.

[49] Manasi Muglikar, Leonard Bauersfeld, Diederik Paul Moeys, and Davide Scaramuzza. 2023. Event-based shape from polarization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 1547–1556.

[50] Tony Mullen. 2011. *Mastering blender.* John Wiley & Sons.

[51] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. 2022. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 8280–8290.

[52] Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H Kim. 2018. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–12.

[53] Yeongwoo Nam, Mohammad Mostafavi, Kuk-Jin Yoon, and Jonghyun Choi. 2022. Stereo depth from events cameras: Concentrate and focus on the future. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 6114–6123.

[54] Yunshan Qi, Lin Zhu, Yu Zhang, and Jia Li. 2023. E2NeRF: Event Enhanced Neural Radiance Fields from Blurry Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 13254–13264.

[55] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. 2019. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence* 43, 6 (2019), 1964–1980.

[56] Why Physically-Based Rendering. 2015. Physically-Based Rendering. *Procedia IUTAM* 13 (2015), 127–137.

[57] Viktor Rudnev, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik. 2023. EventNeRF: Neural radiance fields from a single colour event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 4992–5002.

[58] Christoph Schlieder. 2010. Digital heritage: Semantic challenges of long-term preservation. *Semantic Web* 1, 1-2 (2010), 143–147.

[59] Carolin Schmitt, Simon Donne, Gernot Riegler, Vladlen Koltun, and Andreas Geiger. 2020. On joint estimation of pose, geometry and svbrdf from a handheld scanner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 3493–3503.

[60] Johannes L Schonberger and Jan-Michael Frahm. 2016. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 4104–4113.

[61] Shintaro Shiba, Yoshimitsu Aoki, and Guillermo Gallego. 2022. Secrets of event-based optical flow. In *European Conference on Computer Vision.* Springer, 628–645.

[62] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. 2021. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 7495–7504.

[63] Timo Stoffregen and Lindsay Kleeman. 2019. Event cameras, contrast maximization and reward functions: An analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 12300–12308.

[64] Mariagrazia Tortora, Stefano Sfarra, Marco Chiarini, Valeria Daniele, Giuliana Taglieri, and Giorgio Cerichelli. 2016. Non-destructive and micro-invasive testing techniques for characterizing materials, structures and restoration problems in mural paintings. *Applied Surface Science* 387 (2016), 971–985.

[65] Fangjinhua Wang, Silvano Galliani, Christoph Vogel, Pablo Speciale, and Marc Pollefeys. 2021. Patchmatchnet: Learned multi-view patchmatch stereo. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 14194–14203.

[66] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689* (2021).

[67] Ziwei Wang, Yonhon Ng, Cedric Scheerlinck, and Robert Mahony. 2021. An asynchronous kalman filter for hybrid event cameras. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 448–457.

[68] Ziwei Wang, Liyuan Pan, Yonhon Ng, Zheyu Zhuang, and Robert Mahony. 2021. Stereo hybrid event-frame (shef) cameras for 3d perception. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE, 9758–9764.

[69] Wenming Weng, Yueyi Zhang, and Zhiwei Xiong. 2023. Event-Based Blurry Frame Interpolation Under Blind Exposure. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 1588–1598.

[70] Qingshan Xu and Wenbing Tao. 2019. Multi-scale geometric consistency guided multi-view stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5483–5492.

[71] Zexiang Xu, Sai Bi, Kalyan Sunkavalli, Sunil Hadap, Hao Su, and Ravi Ramamoorthi. 2019. Deep view synthesis from sparse photometric images. *ACM Transactions on Graphics (ToG)* 38, 4 (2019), 1–13.

[72] Zhenyu Xu, Yiguang Liu, Xuelei Shi, Ying Wang, and Yunan Zheng. 2020. Marmvs: Matching ambiguity reduced multiple view stereo for efficient large scale scene reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 5981–5990.

[73] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. 2018. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (ToG)* 37, 4 (2018), 1–13.

[74] Yixin Yang, Jin Han, Jinxiu Liang, Imari Sato, and Boxin Shi. 2023. Learning event guided high dynamic range video reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 13924–13934.

[75] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. 2022. Neilf: Neural incident light field for physically-based material estimation. In *European Conference on Computer Vision.* Springer, 700–716.

[76] Renjiao Yi, Ping Tan, and Stephen Lin. 2020. Leveraging multi-view image sets for unsupervised intrinsic image decomposition and highlight separation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 12685–12692.

[77] Youngho Yoon and Kuk-Jin Yoon. 2023. Cross-Guided Optimization of Radiance Fields with Multi-View Image Super-Resolution for High-Resolution Novel View Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 12428–12438.

[78] Jingyang Zhang, Yao Yao, Shiwei Li, Jingbo Liu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. 2023. Neilf++: Inter-reflectable light fields for geometry and material estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 3601–3610.

[79] Jingyang Zhang, Yao Yao, and Long Quan. 2021. Learning signed distance field for multi-view surface reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 6525–6534.

[80] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. 2022. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 5565–5574.

[81] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 2021. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5453–5462.

[82] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 586–595.

[83] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)* 40, 6 (2021), 1–18.

[84] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. 2022. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 18643–18652.

[85] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. 2019. Unsupervised event-based learning of optical flow, depth, and egomotion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 989–997.

[86] Todd E Zickler, Peter N Belhumeur, and David J Kriegman. 2002. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision* 49 (2002), 215–227.