EventUPS: Uncalibrated Photometric Stereo Using an Event Camera

Jinxiu Liang[1,2,3#] Bohan Yu[1,2#] Siqi Yang[1,2,4] Haotian Zhuang[5]
Jieji Ren[6] Peiqi Duan[1,2] Boxin Shi[1,2*]

[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University

[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

[3] National Institute of Informatics [4] Institute for Artificial Intelligence, Peking University

[5] Tsinghua University [6] Shanghai Jiaotong University

cssherryliang@gmail.com,{ybh1998,yousiki,duanqi0001,shiboxin}@pku.edu.cn,
zht23@mails.tsinghua.edu.cn,jiejiren@sjtu.edu.cn

Abstract

We present EventUPS, the first uncalibrated photometric stereo (UPS) method using an event camera—a neuromorphic sensor that asynchronously detects brightness changes with microsecond resolution. Traditional frame-based UPS methods are hindered by high bandwidth demands and limited use in dynamic scenes. These methods require dense image correspondence under varying illumination and are incompatible with the fundamentally different sensing paradigm of event data. Our approach introduces three key innovations: an augmented null space formulation that directly relates each event to joint constraints on surface normals and lighting, naturally handling ambient illumination; a continuous parameterization of time-varying illumination that connects asynchronous events to synchronized lighting estimation; and a lighting fixture with known relative geometry that reduces ambiguity to a convex-concave uncertainty. We validate EventUPS using a custom-built LED lighting system. Experimental results show that our method achieves accuracy surpassing its frame-based counterpart while requiring only 5% of the data bandwidth.

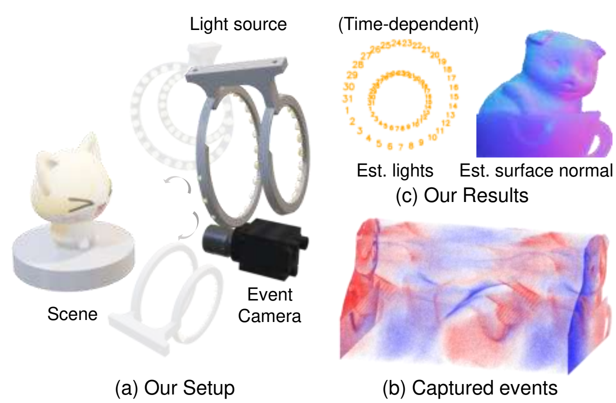

Figure 1. We propose an uncalibrated photometric stereo method designed for event cameras, EventUPS. Our approach employs (a) a portable lighting system with known relative geometry (*e.g.*, dual LED rings) but an unknown global pose, removing the need for precise calibration. From (b) sparse, asynchronous events triggered by sequentially activated lights, it jointly recovers (c) dense surface normals and the time-varying lighting trajectory[1].

1. Introduction

Photometric stereo (PS) recovers shape from images under varying illumination [14, 42]. For decades, it has been a cornerstone for high-fidelity 3D reconstruction, enabling detailed surface normal recovery for applications from cultural heritage to industrial inspection. The uncalibrated variant of PS (UPS)—where lighting directions are unknown—offers particular practical appeal by eliminating the need for precise calibration apparatus, facilitating deployment in uncontrolled environments. However, this flexibility incurs significant costs: conventional uncalibrated PS methods demand dense correspondences across multiple images under varying illumination, imposing high data bandwidth and rendering them ill-suited for dynamic scenes or resource-constrained systems.

Meanwhile, event cameras represent a paradigm shift in visual sensing [10]. Instead of capturing intensity frames at fixed intervals, these bio-inspired sensors asynchronously detect per-pixel logarithmic brightness changes. This provides microsecond temporal resolution, an exceptional dynamic range (140dB *vs*. 60dB), and drastically reduced data

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

[1] Est. Lights: The estimated time-varying light trajectory (x, y components on the unit sphere), with numbers indicating temporal sequence, colors denoting relative intensity (yellow higher, green lower).

rates—all critical for platforms like mobile robots. These unique attributes have transformed high-speed tasks such as SLAM [39] and object tracking [7].

Recent advances in event-based PS—exemplified by the EventPS [45]—have shown promise by leveraging these unique capabilities. By using a calibrated, continuously rotating light source, it computes surface normals directly from brightness change events, naturally accommodating rapid variations in surface orientation even in dynamic scenes. This approach dramatically reduces both latency and data redundancy compared to frame-based methods. However, EventPS [45] requires precise lighting calibration and darkroom conditions, which are challenging outside labs. In many real-world scenarios—such as robotic deployments, motion capture, or industrial settings—maintaining calibrated lighting and negligible ambient illumination becomes impractical due to environmental constraints and operational demands.

A natural question arises: *Can established frame-based UPS approaches adapt directly to event cameras?* Our analysis reveals fundamental incompatibilities: *i) Inability to handle ambient illumination.* Frame-based PS methods often address ambient illumination by capturing a dark frame without designed lighting. Unfortunately, this strategy is not applicable to events as they are generated asynchronously by logarithmic brightness changes—encoding brightness ratios rather than absolute intensities. *ii) Absence of synchronous correspondence.* Non-learning frame-based UPS [2, 46] typically performs matrix factorization to decompose observed intensities into surface normals and lighting directions, establishing constraints through pixel correspondences under shared illumination conditions. However, when applied to events—which correspond to different lighting directions and occur asynchronously—the establishment of such correspondences becomes challenging. *iii) Inapplicability of conventional ambiguity resolution.* Existing UPS resolves the remaining ambiguity by leveraging effective priors on albedo characteristics [28], which are not applicable to events since the event generation equation is inherently invariant to albedo.

In this paper, we present EventUPS—the first uncalibrated photometric stereo method using an event camera. Our approach rethinks PS for the event domain, leveraging the unique properties of event cameras to recover detailed surface geometry while simultaneously estimating the time-varying illumination, all without explicit lighting calibration. The core of our contribution is a theoretical framework that formulates each event as a constraint in the null space of an augmented normal vector, enabling joint recovery of surface normals and lighting directions despite the asynchronous nature of events and the presence of ambient illumination. To address the incompatibilities, we make the following technical contributions:

- An augmented null-space formulation *able to handle ambient illumination* by modeling it as an unknown variable within event-based constraints on normals and lighting.
- A continuous parameterization of the time-varying illumination that *establishes correspondence* by linking asynchronous events to a shared lighting trajectory.
- A geometric UPS ambiguity resolution technique *applicable to event data*, which uses a lighting fixture with known relative geometry but an unknown global pose.

Along with the proposed EventUPS framework, we develop a joint optimization framework that effectively integrates integrability constraints with our event-based null space formulation. To validate our approach in real scenes, we build a hardware prototype around an LED-based lighting system, which implements the lighting configurations with known geometry of dual ring (see Figure 1) and trefoil curve patterns. Experiments on semi-real and real data demonstrate that our method achieves accuracy outperforming the leading frame-based non-learning UPS [28] while requiring only 5% of the data bandwidth.

2. Related Works

PS and event-based vision have advanced independently for decades. Our work lies at their intersection, specifically addressing the challenges of UPS with event cameras. Here, we review developments in both domains that inform ours.

2.1. Uncalibrated Photometric Stereo

PS estimates surface normals from images captured under varying illumination directions [15, 42]. While classical approaches assume known lighting, the uncalibrated setting—where lighting is unknown—presents greater challenges but offers superior practical advantages. For comprehensive reviews, see [30, 33, 34, 41].

A fundamental challenge in UPS is the ambiguity in simultaneously recovering lighting and surface normals. This was first identified as a general 3×3 linear ambiguity [14], later reduced to the 3-parameter generalized bas-relief (GBR) ambiguity by imposing integrability constraints [2, 46]. Resolving the ambiguity typically relies on physical cues such as specular reflections [8, 9, 11, 38, 43], statistical priors on albedo [1, 21, 28, 32], shadows [36], interreflections [3], or perspective effects [27]. The light source configuration is also critical; strategic arrangements like ring lights with multi-view information can provide additional constraints to resolve ambiguity without strong surface assumptions [48]. Recent learning-based methods have shown promising results [4, 5, 20], but they often require extensive training data and can struggle with generalization.

Accelerating PS acquisition remains an active research area, with solutions like high-speed cameras [18, 22, 40] or multi-tap CMOS [44] with synchronized illumination, though these dramatically increase bandwidth. Single-shot multi-spectral systems [19, 26, 35, 37] offer an alternative but face challenges like limited spectral bands, crosstalk, and intensity inconsistencies across bands introduce additional challenges [12, 16, 17].

2.2. Photometric Methods with Event Cameras

With their high temporal resolution, wide dynamic range, and asynchronous operation, event cameras are powerful tools for photometric measurements. These advantages make them particularly suitable for capturing dynamic lighting effects crucial for photometric applications. Their microsecond resolution is especially valuable for active lighting techniques. MC3D [23] pioneered event-based structured light for high-speed 3D scanning, while ESL [24] improved noise robustness by exploiting spatial correlations. Researchers have explored diverse illumination patterns for specific tasks, such as capturing events during light activation to reconstruct iso-contours [13] and address intensity-distance ambiguity by using a diffuse sphere [6], estimating surface normals using rotating polarizers [25], and separating direct/global illumination components [47].

Most relevant to our work, EFPS-Net [31] fused sparse events with RGB images for normal estimation, while EventPS [45] pioneered purely event-based PS. EventPS showed that the differential nature of event cameras aligns perfectly with PS requirements, but it assumes calibrated lighting and cannot handle ambient illumination. The sparse, asynchronous nature of event data presents unique challenges in uncalibrated settings, particularly in establishing pixel correspondences under consistent lighting. This fundamentally different sensing paradigm makes direct application of non-learning UPS methods [14, 46] infeasible. Our work, EventUPS, is the first to solve these challenges.

3. Method

3.1. Problem Formulation

Image formulation. We consider a scene captured by an orthographic projection camera[2]. The surface is assumed to be Lambertian and is illuminated by a combination of a distant, time-varying directional light source and constant ambient illumination. For a pixel $\boldsymbol{x}$ in the image plane $\Omega \subset \mathbb{R}^2$, the reflected radiance $I_{\boldsymbol{x}}(t)$ at time t is given by:

$$I_{\boldsymbol{x}}(t) = \max\left[0, a_{\boldsymbol{x}} \boldsymbol{n}_{\boldsymbol{x}} \cdot \boldsymbol{s}(t)\right] + b_{\boldsymbol{x}}, \tag{1}$$

where $\boldsymbol{n}_{\boldsymbol{x}} \in \mathbb{R}^3$ is the unit surface normal ($\|\boldsymbol{n}_{\boldsymbol{x}}\|_2 = 1$), $a_{\boldsymbol{x}} > 0$ is the diffuse albedo, and $b_{\boldsymbol{x}} \geq 0$ is the ambient illumination term. The lighting is described by the vector $\boldsymbol{s}(t) \in \mathbb{R}^3$ with intensity $\|\boldsymbol{s}(t)\|^2$ and unit direction $\hat{\boldsymbol{s}}(t) = \boldsymbol{s}(t)/\|\boldsymbol{s}(t)\|$. The max operator models attached shadows, where the dot product becomes negative.

[2]Orthographic projection is a widely used assumption in PS [33] when the object-to-camera distance (approximately 20 cm in our setup) is significantly larger than object depth variations (approximately 3 cm).

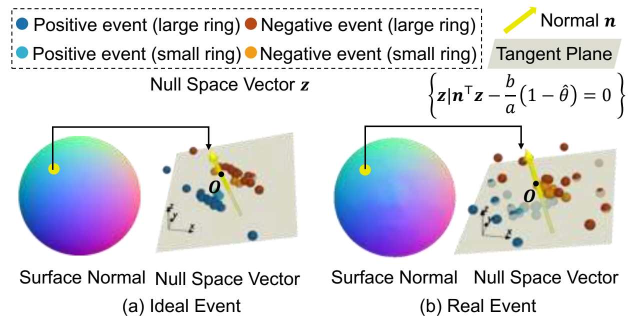

Figure 2. The null space vectors and estimated normals: (a) Lambertian sphere under ideal conditions, (b) non-Lambertian sphere with real events. EventUPS models ambient and albedo effects via a residual term, enhancing robustness.

Event generation. Event cameras are bio-inspired sensors that asynchronously report logarithmic changes in brightness. An event $(\boldsymbol{x}, t_{\boldsymbol{x}}^{(k)}, p_{\boldsymbol{x}}^{(k)})$ is triggered at pixel $\boldsymbol{x}$ when the logarithmic change in brightness exceeds a predetermined contrast threshold $\theta > 0$:

$$\log(I_{\boldsymbol{x}}(t_{\boldsymbol{x}}^{(k)})) - \log(I_{\boldsymbol{x}}(t_{\boldsymbol{x}}^{(k-1)})) \geq p_{\boldsymbol{x}}^{(k)} \theta, \tag{2}$$

where $t_{\boldsymbol{x}}^{(k)}$ denotes the timestamp of the k-th event at pixel $\boldsymbol{x}$, and $p_{\boldsymbol{x}}^{(k)} \in \{+1, -1\}$ indicates whether the brightness increased or decreased. The relationship between consecutive brightness measurements can be approximated as:

$$I_{\boldsymbol{x}}(t_{\boldsymbol{x}}^{(k)}) \approx I_{\boldsymbol{x}}(t_{\boldsymbol{x}}^{(k-1)}) \cdot \exp(p_{\boldsymbol{x}}^{(k)} \theta). \tag{3}$$

For notational simplicity, we define $\tilde{\theta}_{\boldsymbol{x}}^{(k)} = \exp(p_{\boldsymbol{x}}^{(k)} \theta)$.

UPS with events. Unlike prior event-based works [45] that assume calibrated lighting, we aim to jointly estimate the per-pixel surface normals $\boldsymbol{n}_{\boldsymbol{x}}$ and the continuous, time-varying lighting function $\boldsymbol{s}(t)$ from only the event stream $\mathbb{E} = \{(\boldsymbol{x}, t_{\boldsymbol{x}}^{(k)}, p_{\boldsymbol{x}}^{(k)})\}_{\boldsymbol{x} \in \Omega}$ across all pixels $\boldsymbol{x} \in \Omega$. This inverse problem is formulated as the minimization of the joint loss function that measures consistency between the observed events and the image formation model:

$$\mathrm{argmin}_{\{\boldsymbol{n}_{\boldsymbol{x}}\}_{\boldsymbol{x} \in \Omega}, \boldsymbol{s}(t)} \mathcal{L}(\{\boldsymbol{n}_{\boldsymbol{x}}\}, \boldsymbol{s}(t); \mathbb{E}). \tag{4}$$

3.2. Joint Estimating Normal and Lighting

The augmented null-space formulation. Conventional photometric stereo often handles ambient light by subtracting a dark frame. This is infeasible for event-based methods where measurements are relative and asynchronous. We propose an alternative by substituting the image model Eq. (1) into the event model Eq. (3):

$$\begin{aligned} &\max\left[0, a_{\boldsymbol{x}} \boldsymbol{n}_{\boldsymbol{x}} \cdot \boldsymbol{s}(t_{\boldsymbol{x}}^{(k)})\right] + b_{\boldsymbol{x}} \\ &\approx \tilde{\theta}_{\boldsymbol{x}}^{(k)} \cdot \left(\max\left[0, a_{\boldsymbol{x}} \boldsymbol{n}_{\boldsymbol{x}} \cdot \boldsymbol{s}(t_{\boldsymbol{x}}^{(k-1)})\right] + b_{\boldsymbol{x}} \right). \end{aligned} \tag{5}$$

For most observed events, the surface point is illuminated at both timestamps (*i.e.*, it is not in an attached shadow). This allows us to remove the $\max$ operator. The albedo $a_{\boldsymbol{x}}$ can be eliminated by dividing it out, as it appears on both sides. This makes our formulation inherently albedo-invariant. Rearranging the terms yields a linear constraint on the normal $\boldsymbol{n}_{\boldsymbol{x}}$ and the "residual" pixel-wise constant ambient-albedo ratio $r_{\boldsymbol{x}} = b_{\boldsymbol{x}}/a_{\boldsymbol{x}} \geq 0$:

$$\boldsymbol{n}_{\boldsymbol{x}} \cdot (\boldsymbol{s}(t_{\boldsymbol{x}}^{(k)}) - \tilde{\theta}_{\boldsymbol{x}}^{(k)} \boldsymbol{s}(t_{\boldsymbol{x}}^{(k-1)})) + r_{\boldsymbol{x}}(1 - \tilde{\theta}_{\boldsymbol{x}}^{(k)}) = 0. \quad (6)$$

Shadow events, which occur when one of the terms under the max operator becomes zero, are relatively rare and can be identified through temporal consistency checks. When they do occur, they provide constraints at shadow boundaries that can be incorporated into our framework.

Let $\boldsymbol{z}_{\boldsymbol{x}}^{(k)} = \boldsymbol{s}(t_{\boldsymbol{x}}^{(k)}) - \tilde{\theta}_{\boldsymbol{x}}^{(k)} \boldsymbol{s}(t_{\boldsymbol{x}}^{(k-1)})$. We define an *augmented normal vector* $\tilde{\boldsymbol{n}}_{\boldsymbol{x}} = [\boldsymbol{n}_{\boldsymbol{x}}^\top, r_{\boldsymbol{x}}]^\top \in \mathbb{R}^4$, which is constant for each pixel $\boldsymbol{x}$. For each event, we define a corresponding *augmented null space vector* $\tilde{\boldsymbol{z}}_{\boldsymbol{x}}^{(k)} = [\boldsymbol{z}_{\boldsymbol{x}}^{(k)\top}, (1 - \tilde{\theta}_{\boldsymbol{x}}^{(k)})]^\top$. This recasts Eq. (6) into an geometric relationship:

$$\tilde{\boldsymbol{z}}_{\boldsymbol{x}}^{(k)\top} \tilde{\boldsymbol{n}}_{\boldsymbol{x}} = 0. \quad (7)$$

This means the augmented normal $\tilde{\boldsymbol{n}}_{\boldsymbol{x}}$ must lies in the null space of the vectors $\{\tilde{\boldsymbol{z}}_{\boldsymbol{x}}^{(k)}\}$. Please see Fig. 2 for an illustration. With at least three linearly independent null space vectors from events at a pixel, we can determine the augmented normal $\tilde{\boldsymbol{n}}_{\boldsymbol{x}}$ up to a scale factor. By stacking all augmented null space vectors $\tilde{\boldsymbol{z}}_{\boldsymbol{x}}^{(k)}$ for a pixel $\boldsymbol{x}$ into a matrix $\boldsymbol{Z}_{\boldsymbol{x}} \in \mathbb{R}^{4 \times K_{\boldsymbol{x}}}$ where $K_{\boldsymbol{x}}$ is the number of events at that pixel and the k-th row is $\tilde{\boldsymbol{z}}_{\boldsymbol{x}}^{(k)\top}$, we solve for the normal by finding the vector that minimizes the projection:

$$\tilde{\boldsymbol{n}}_{\boldsymbol{x}}^* = \underset{\tilde{\boldsymbol{n}}_{\boldsymbol{x}}}{\operatorname{argmin}} \|\boldsymbol{Z}_{\boldsymbol{x}}^\top \tilde{\boldsymbol{n}}_{\boldsymbol{x}}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{n}_{\boldsymbol{x}}\|_2 = 1. \quad (8)$$

This constrained optimization problem has an elegant solution via Singular Value Decomposition (SVD) of the matrix $\boldsymbol{Z}_{\boldsymbol{x}} \boldsymbol{Z}_{\boldsymbol{x}}^\top \in \mathbb{R}^{4 \times 4}$. The solution is the eigenvector corresponding to the smallest eigenvalue, appropriately scaled to ensure the unit normal constraint.

Continuous lighting parameterization. In the uncalibrated setting, the lighting direction $\boldsymbol{s}(t)$ is unknown and varies continuously over time. The asynchronous nature of events makes it challenging to establish correspondences across pixels under identical lighting conditions, which is a prerequisite for non-learning uncalibrated PS methods. To address this challenge, we model $\boldsymbol{s}(t)$ as a B-spline curve. This provides a smooth, local, and efficient parameterization that is ideal for asynchronous data. By selecting M control points $\{\boldsymbol{s}(t_j)\}_{j=0}^{M-1}$ at uniform times $t_j = t_0 + j\Delta t$,

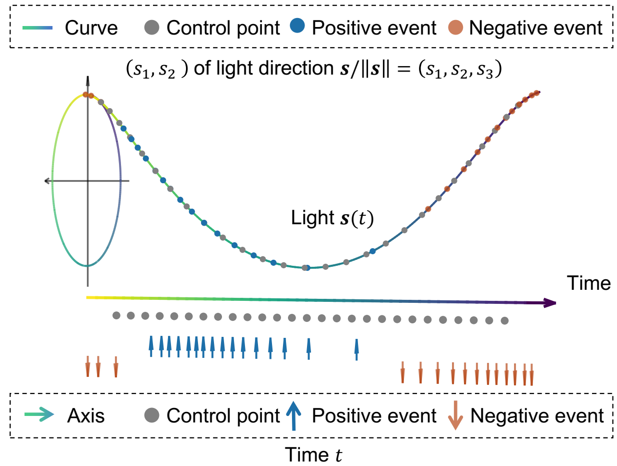

Figure 3. Continuous illumination model. We represent the time-varying lighting $\boldsymbol{s}(t)$ as a B-spline controlled by a set of points. This continuous representation links asynchronous events to a globally synchronized lighting estimate, shown here for a rotating ring pattern projected onto a 2D plane.

the lighting at any time t is a linear combination:

$$\boldsymbol{s}(t) = \sum_{j=0}^{M-1} C_{j,l}(t) \boldsymbol{s}(t_j), \quad (9)$$

where $\{C_{j,l}(t)\}$ are the B-spline basis functions of order l. An important property of this formulation is that the value of $\boldsymbol{s}(t)$ for $t_j \leq t < t_{j+1}$ depends only on l consecutive control points $\boldsymbol{s}(t_j), \boldsymbol{s}(t_{j+1}), \ldots, \boldsymbol{s}(t_{j+l-1})$, providing locality in the temporal domain. This links the null space vector $\boldsymbol{z}_{\boldsymbol{x}}^{(k)}$ directly to the control points:

$$\boldsymbol{z}_{\boldsymbol{x}}^{(k)} = \sum_{j=0}^{M-1} \widetilde{C}_{j,l}(t_{\boldsymbol{x}}^{(k)}) \boldsymbol{s}(t_j), \quad (10)$$

where

$$\widetilde{C}_{j,l}(t_{\boldsymbol{x}}^{(k)}) = C_{j,l}(t_{\boldsymbol{x}}^{(k)}) - \tilde{\theta}_{\boldsymbol{x}}^{(k)} C_{j,l}(t_{\boldsymbol{x}}^{(k-1)}), \quad (11)$$

This expression depends only on the observed event timestamps $t_{\boldsymbol{x}}^{(k)}$ at pixel $\boldsymbol{x}$, providing a crucial bridge between pixel-specific asynchronous events and a global, continuous lighting model (see Fig. 3). In matrix form, we stack the control points $\{\boldsymbol{s}(t_j)\}_{j=0}^{M-1}$ into a matrix $\boldsymbol{S} \in \mathbb{R}^{3 \times M}$, and the time-dependent coefficients $\widetilde{C}_{j,l}(t_{\boldsymbol{x}}^{(k)})$ into a matrix $\boldsymbol{C}_{\boldsymbol{x}} \in \mathbb{R}^{M \times K_{\boldsymbol{x}}}$. The augmented measurement matrix for a pixel $\boldsymbol{x}$ with $\tilde{\boldsymbol{\theta}}_{\boldsymbol{x}} = [(1 - \tilde{\theta}_{\boldsymbol{x}}^{(1)}), \ldots, (1 - \tilde{\theta}_{\boldsymbol{x}}^{(K_{\boldsymbol{x}})})] \in \mathbb{R}^{K_{\boldsymbol{x}}}$ becomes:

$$\tilde{\boldsymbol{Z}}_{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{S}\boldsymbol{C}_{\boldsymbol{x}} \\ \tilde{\boldsymbol{\theta}}_{\boldsymbol{x}}^\top \end{bmatrix} \in \mathbb{R}^{4 \times K_{\boldsymbol{x}}}. \quad (12)$$

The data term becomes:

$$\mathcal{L}_{\text{data}}(\{\tilde{\boldsymbol{n}}_{\boldsymbol{x}}\}, \boldsymbol{S}) = \sum_{\boldsymbol{x} \in \Omega} \|\tilde{\boldsymbol{Z}}_{\boldsymbol{x}}^\top \tilde{\boldsymbol{n}}_{\boldsymbol{x}}\|_2^2. \quad (13)$$

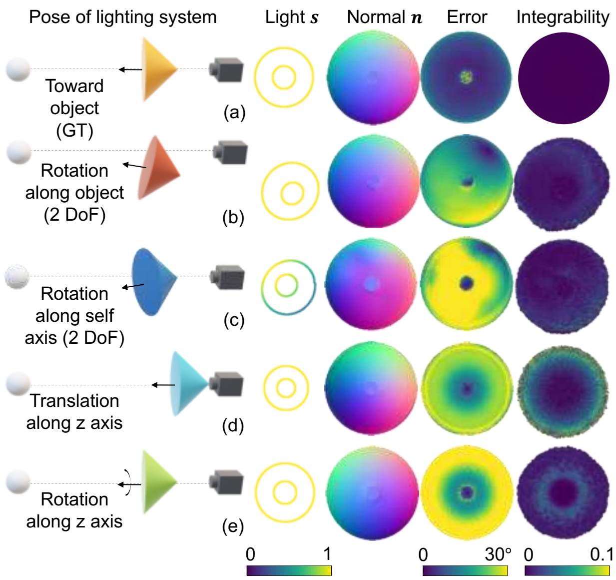

Figure 4. Resolving Ambiguity. With ground-truth lighting (a), the normal is recovered accurately. Assuming incorrect lighting transformations (b-e), estimating normals without constraints leads to large errors. The strong correlation between normal error and integrability cost validates the effectiveness of enforcing integrability with a known lighting fixture to resolve ambiguity.

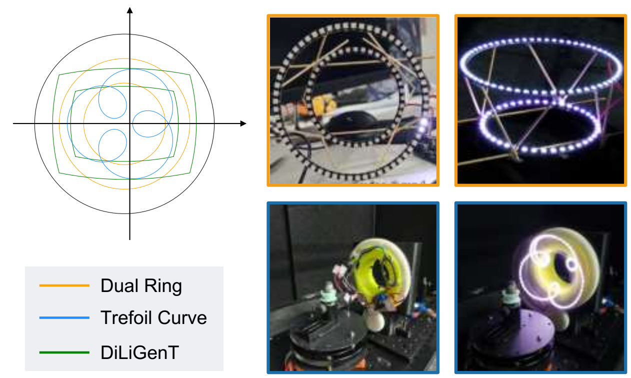

Figure 5. The lighting configurations employed in our experiments. We built physical prototypes for the "dual ring" and the "trefoil curve" patterns using programmable LEDs, shown both statically (left) and as dynamic trajectories captured with long exposure (right). We also test on the standard "DiLiGenT" pattern.

This bilinear structure naturally lends itself to an alternating optimization scheme.

3.3. Resolving Ambiguity

Integrability. UPS suffers from a fundamental ambiguity where multiple combinations of surface normals and lighting directions can produce identical observations. Physical constraints, particularly the integrability of surface normals, require that the normals be derivable from the gradient of an underlying surface. For non-learning frame-based UPS, this constraint reduces the general linear ambiguity to the GBR transformation [2, 46], which maps the surface $z(\boldsymbol{x})$ to $\lambda z(\boldsymbol{x})+\mu x+\nu y$, where $\lambda \neq 0$. For event-based UPS, imposing integrability similarly reduces the ambiguity to the GBR transformation, as the constraint acts on the normal field independently of the observation modality. Specifically, we use an integrability term. It ensures the normal field $\{\boldsymbol{n}_{\boldsymbol{x}}\}$ corresponds to a valid, continuous surface by penalizing the curl of the corresponding gradient field:

$$\mathcal{L}_{\text{int}} = \sum_{\boldsymbol{x}} \left\| \frac{\partial}{\partial y} \left(\frac{(\boldsymbol{n}_{\boldsymbol{x}})_x}{(\boldsymbol{n}_{\boldsymbol{x}})_z} \right) - \frac{\partial}{\partial x} \left(\frac{(\boldsymbol{n}_{\boldsymbol{x}})_y}{(\boldsymbol{n}_{\boldsymbol{x}})_z} \right) \right\|^2. \tag{14}$$

Constraints from lighting fixture configurations. To further resolve this ambiguity, we leverage a portable lighting system with a *known 3D structure* undergoing *unknown 6-DoF transformations*. The system consists of light sources at known positions $\boldsymbol{p}_i$ (for $i = 1, 2, \dots, N$) in a local reference frame. Assuming the camera is at the origin, the world positions become $\boldsymbol{p}_i' = \boldsymbol{R}\boldsymbol{p}_i + \boldsymbol{t}$ under a rigid transformation with orthogonal matrix $\boldsymbol{R} \in O(3)$ and translation vector $\boldsymbol{t} \in \mathbb{R}^3$. The light vector for the i-th source is $\boldsymbol{s}_i = u_i \boldsymbol{p}_i'$ with unknown intensity $u_i > 0$. A GBR transformation matrix $\boldsymbol{G}$ acts on the recovered normals, and an inverse transformation $\boldsymbol{G}^{-\top}$ acts on the light vectors. The observed lighting $\boldsymbol{s}_i'$ is related to the true lighting $\boldsymbol{s}_i$ by $\boldsymbol{s}_i' = \boldsymbol{G}^{-\top}\boldsymbol{s}_i$. Combining these, we obtain a relationship between the known local positions $\boldsymbol{p}_i$ and the estimated, transformed light vectors $\boldsymbol{s}_i'$:

$$\boldsymbol{s}_i' = u_i \boldsymbol{G}^{-\top}(\boldsymbol{R}\boldsymbol{p}_i + \boldsymbol{t}). \tag{15}$$

This system of equations constrains the possible transformations. As formally proven in the supplement, these constraints reduce the GBR ambiguity to a discrete set. Specifically, for a set of at least seven light positions in general position (*i.e.*, not all lying on any nontrivial quadric surface of a specified form passing through the origin), under physical constraints, the ambiguity can be resolved up to the classic convex/concave reflection. This eliminates the need for albedo or reflectance priors. An illustration of integrability's role in resolving positional ambiguity in our setup is shown in Fig. 4.

Lighting configurations. For event cameras that asynchronously detect brightness changes with high temporal resolution, we sequentially activate each light (*e.g.*, on millisecond timescales) to generate an event stream, with timestamps correlating to lighting changes. We designed two physical patterns (see Fig. 5) that satisfy the geometric requirements: a coaxial dual-ring consisting of LEDs arranged on two concentric rings of different radii, separated by a known axial distance; and a trefoil curve with a moving ring of lights along a non-planar, asymmetric 3D trefoil knot trajectory. Both configurations meet the conditions for ambiguity resolution. Details of the theorem, proof, and lighting configurations are provided in the supplement.

3.4. Optimization

We solve for normals and lighting via alternating optimization, minimizing a joint loss function:

$$\mathcal{L} = \mathcal{L}_{\text{data}}(\{\tilde{\boldsymbol{n}}_{\boldsymbol{x}}\}, \boldsymbol{S}) + \lambda_{\text{int}}\mathcal{L}_{\text{int}}(\{\boldsymbol{n}_{\boldsymbol{x}}\}), \tag{16}$$

where the weight λ_{int} balances data fidelity Eq. (13) against surface smoothness Eq. (14). The optimization proceeds by alternating between two main steps until convergence:

- Normal update: With the lighting control points $\boldsymbol{S}$ fixed, the problem decouples per pixel. We solve for each augmented normal $\tilde{\boldsymbol{n}}_{\boldsymbol{x}}$ from its observation matrix $\tilde{\boldsymbol{Z}}_{\boldsymbol{x}}\tilde{\boldsymbol{Z}}_{\boldsymbol{x}}^{\top}$. The integrability term is also incorporated.
- Lighting update: With the normals $\{\tilde{\boldsymbol{n}}_{\boldsymbol{x}}\}$ fixed, the data term $\mathcal{L}_{\text{data}}$ becomes a quadratic function of the lighting control points $\boldsymbol{S}$. This step reduces to solving a global linear least-squares system to update $\boldsymbol{S}$.

We alternate between updating the normals $\{\tilde{\boldsymbol{n}}_{\boldsymbol{x}}\}$ and the lighting parameters $\boldsymbol{S}$ until convergence. This framework robustly recovers accurate surface normals from event data, even with ambient light and without calibration.

4. Experiment

4.1. Implementation Details

Our EventUPS implementation uses a hybrid CPU/GPU approach: event pre-processing is GPU-accelerated with OpenCL and Rust bindings, while the PyTorch-based optimization framework handles the core algorithm. For the B-spline lighting representation, we use cubic splines (order $l = 4$) with 31 control points for each 360° rotation cycle unless otherwise specified. The integrability weight is $\lambda_{\text{int}} = 0.1$ for all real data and $\lambda_{\text{int}} = 0.01$ for all synthetic data.

We developed a custom lighting system using programmable IC LEDs. For the “dual ring” configuration, we mounted concentric LED rings (WS2818B protocol, 800kHz fixed rate) with different radii sharing a common axis. The inner ring contains 40 LEDs, while the outer ring has 60 LEDs. One complete scanning cycle takes approximately 4 seconds. For the “trefoil curve” configuration, we attached an LED ring (APA102, up to 20MHz SPI clock frequency, 24 LEDs) to a rotating mechanism. The LEDs switch at a frequency three times faster than the rotation mechanism, creating a trefoil curve pattern. One complete cycle takes approximately 1/3 second, limited by the rotation speed of the mechanism.

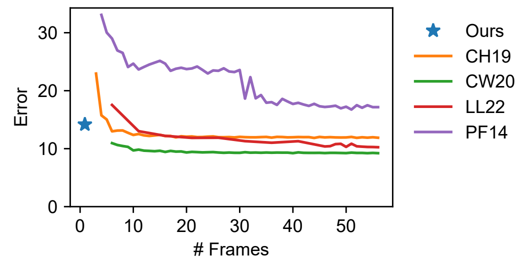

Figure 6. Performance *vs*. data rate comparison. This plot shows the average MAE of EventUPS against several frame-based methods on the DiLiGenT-Ev dataset as a function of input frames.

We captured data using a Prophesee EVK4 HD camera with an IMX636 event sensor. The contrast sensitivity threshold biases were set to -20 and the dead time bias to -20, resulting in an approximately refractory period of 580 µs between consecutive events at a pixel.

4.2. Comparison with Frame-based UPS methods

Data. To facilitate comparison with frame-based methods, we created a semi-real dataset named DiLiGenT-Ev using images from the DiLiGenT dataset [34]. This procedure, which follows the protocol of EventPS [45], allows for a direct comparison between event-based and frame-based techniques on identical underlying scenes and lighting conditions. Specifically, we selected images captured under the outermost rectangular ring of light directions and converted them to event streams using an event simulator [29].

Results. We benchmark EventUPS against state-of-the-art frame-based UPS methods, including the top-performing non-learning method PF14 [28] from the original DiLiGenT benchmark, and several deep learning-based approaches: CW20 [5], CH19 [4], and LL22 [20]. We evaluate accuracy (Mean Angular Error, MAE) and data efficiency. For data rate calculations, we use Total Events $\times$ 16 bits/event (encoded in the Prophesee EVT 3.0 format) for event-based methods, and Frames (lighting directions) $\times$ Resolution $\times$ 8 bits/pixel $\times$ 3 exposures (bracketing for the limited dynamic range) for frame-based methods.

Fig. 6 illustrates the trade-off between data rate and accuracy for frame-based methods—using more images improves accuracy but increases data requirements. Table 1 provides the detailed quantitative breakdown of comparisons with PF14 and CW20. In sharp contrast, EventUPS (marked by a star) demonstrates remarkable data efficiency. EventUPS achieves an MAE of 14.14°, a result that is not only superior to the best performance of the non-learning baseline PF14 (17.16° MAE) but does so using a data volume equivalent to roughly two frames. Our method operates at just 1.6% of the data rate of the 56-frame deep learning setup and less than 5% of the data rate of PF14. This ef-

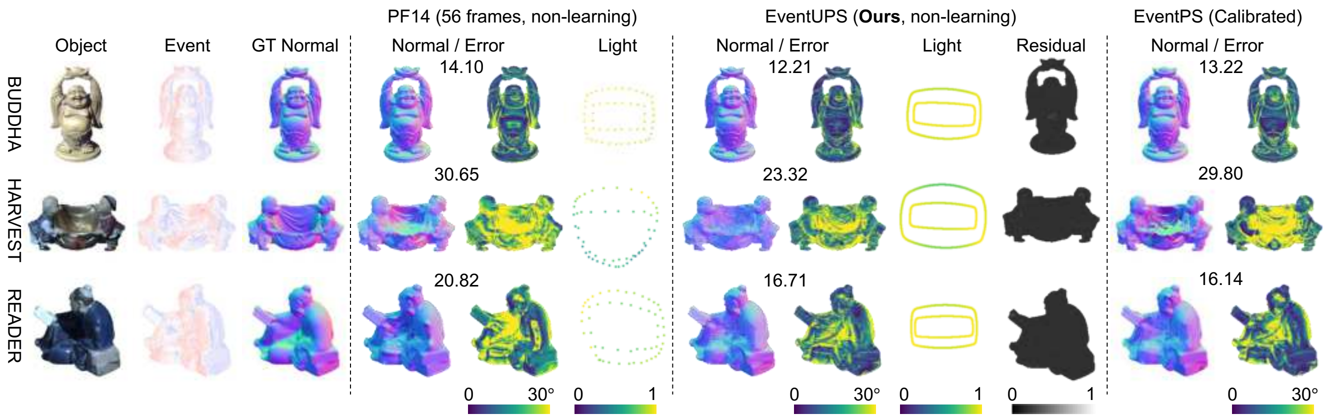

Figure 7. Qualitative results on the DiLiGenT-Ev dataset. Our uncalibrated method produces reconstructions superior to the non-learning, frame-based PF14 [28] and comparable to the calibrated EventPS [45], but without requiring lighting calibration. The "Light" columns visualize the estimated trajectories (x, y components on the unit sphere), with colors denoting relative intensity.

Table 1. Quantitative comparison (MAE in degrees) on the DiLiGenT-Ev dataset. Our method outperforms the frame-based method PF14 [28] (best-performing non-learning method in the original DiLiGenT dataset [34]) while using less than 5% data rate. We include results from a learning-based method CW20 [5] and the only event-based method EventPS [45] (a calibrated baseline of ours, upper bound) as reference. The last column shows the percentage of data rate compared to CW20 with the best performance. The last row indicates the number (#) of events per round for each data.

Uncali-brated	Non-learning	Input	Method	BALL	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READER	Average	Data Rate
✓	✗	6 frames	CW20	5.31	12.10	9.27	9.64	10.97	20.99	9.19	7.61	13.50	10.95	11%
✓	✗	56 frames	CW20	4.96	10.23	7.20	6.42	9.47	18.55	7.43	6.57	12.17	9.22	100%
✓	✓	56 frames	PF14	**5.23**	14.10	**10.03**	19.32	28.63	30.65	9.89	15.79	20.82	17.16	33%
✓	✓	Events	Ours	8.86	**12.21**	10.78	**18.58**	**12.97**	**23.32**	**8.70**	**15.13**	**16.71**	**14.14**	**1.6%**
✗	✓	Events	EventPS	3.45	13.22	7.32	23.33	13.54	29.80	8.34	12.42	16.14	14.17	1.6%
# Events				470 k	306 k	354 k	458 k	200 k	327 k	246 k	334 k	393 k	343 k	/

ficiency stems from the event camera's inherent ability to capture only meaningful brightness changes.

Fig. 7 shows the visual comparison between different non-learning methods, that is, the frame-based UPS method PF14, the proposed EventUPS, and the calibrated event-based method EventPS. EventUPS reconstructs clean and detailed normals that closely match the ground truth, and accurately recovers the circular lighting trajectory. The normals from PF14 appear visibly noisier and less defined, especially for the challenging HARVEST object, aligning with the error metrics in the table. This demonstrates our method's ability to robustly disentangle geometry, lighting, and non-ideal effects from sparse event data alone. EventUPS's superiority over non-learning, frame-based PF14 shows particular strength on objects with complex specularities and textures like HARVEST (23.32° *vs.* 30.65°) and READER (16.71° *vs.* 20.82°). Most notably, our uncalibrated method's performance is on par with, and even slightly surpasses, the average performance of the fully calibrated event-based method EventPS (14.17° MAE). For objects like HARVEST and BUDDHA, our method is more accurate. We attribute this to our model's explicit residual term, which effectively isolates and accounts for non-ideal, real-world effects like ambient illumination.

4.3. Evaluation on Real Camera

Data. To demonstrate the practical applicability of our method, we fabricated five physical objects with ground truth normal maps and captured real events using our custom lighting systems. These data have varying characteristics: simple geometry (BALL), spatially-varying albedo (BALLICCV), moderate geometric details (BUNNY), and complex details (HORSE, TIGER).

Results. Fig. 8 shows results on our captured real dataset, which presents additional challenges including sensor noise, non-idealities in event generation. Our method produces high-quality normal reconstructions with clear

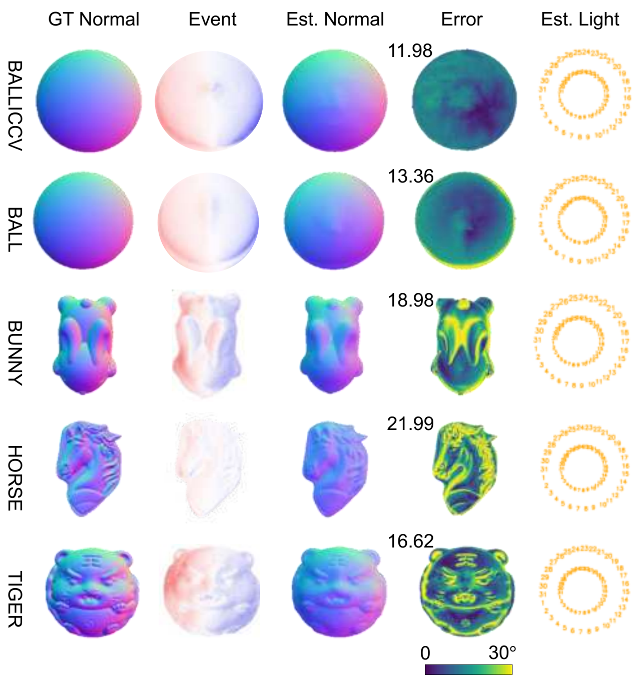

Figure 8. Results on real data. EventUPS successfully recovers detailed normals despite sensor noise and non-ideal lighting (light trajectories are visualized as in Fig. 1).

preservation of fine details, achieving an average MAE of 16.58° across all objects. The spherical object BALLICCV with "ICCV" text highlights an important property: the estimated normal map is invariant to ambient-albedo ratio variations, as the text is barely visible in both the event and the reconstructed normal map. The BUNNY and TIGER objects demonstrate EventUPS's ability to recover intricate geometric details. We observe slightly higher angular errors near object boundaries, which we attribute to near-light effects (the distance between the big ring light and the object is approximately 10 cm) and imperfect lighting alignment. Nevertheless, the overall quality of reconstruction is high. The estimated lighting configurations accurately reflect the dual ring structure used during capture.

Robustness to lighting patterns. To evaluate the robustness of our method to different lighting configurations, we captured the same objects using both the "dual ring" and "trefoil curve" lighting patterns. As shown in Fig. 9, EventUPS produces consistent normal reconstructions regardless of the lighting pattern used. This demonstrates that our approach effectively recovers both surface normals and lighting directions without requiring a specific lighting arrangement, providing flexibility for different practical scenarios. The trefoil pattern yields better results for objects with complex geometry due to its more varied lighting directions, while the dual ring configuration provides more uniform coverage for smoother objects.

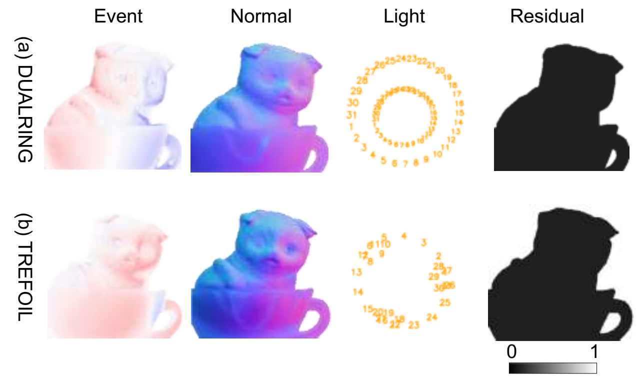

Figure 9. Performance across illumination patterns. EventUPS delivers consistent normal and lighting estimates for both the (a) dual-ring and (b) trefoil curve patterns, showcasing its robustness.

5. Conclusion and Discussion

We have presented EventUPS, the first method to solve UPS using an event camera. By introducing a novel augmented null space formulation, a continuous lighting model, and leveraging known lighting fixture geometry, we resolve the fundamental challenges of applying UPS to the event domain. Our experiments show that EventUPS achieves accuracy comparable to or better than its traditional frame-based counterpart while requiring only a fraction (5%) of the data bandwidth. Our method is robust to ambient illumination and different lighting configurations, making it practical for deployment in less controlled environments.

While our approach represents an advance in uncalibrated event-based PS, several avenues for future work remain. First, extending our framework to handle non-Lambertian surfaces would enhance its applicability in real-world scenarios with complex materials. Second, integrating learning-based components could further improve robustness to noise and outliers while maintaining the physical interpretability of our current model. Finally, exploring applications in dynamic scene reconstruction, where both the camera and objects may be in motion, represents an exciting direction for future research.

Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant Nos. 62136001, 62302019, 624B2006, 62088102), Beijing Natural Science Foundation (Grant No. L233024), and Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Grant No. Z241100003524012). PKU-affiliated authors thank openbayes.com for providing computing resources.

References

[1] Neil G. Alldrin, Satya P. Mallick, and David J. Kriegman. Resolving the Generalized Bas-Relief Ambiguity by Entropy Minimization. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2007. 2

[2] Peter N. Belhumeur, David J. Kriegman, and Alan L. Yuille. The Bas-Relief Ambiguity. *Int. J. Comput. Vis.*, 35(1):33–44, 1999. 2, 5

[3] M.K. Chandraker, F. Kahl, and D.J. Kriegman. Reflections on the generalized bas-relief ambiguity. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2005. 2

[4] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. Self-Calibrating Deep Photometric Stereo Networks. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 2, 6

[5] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K. Wong, and Yasuyuki Matsushita. What Is Learned in Deep Uncalibrated Photometric Stereo? In *Proc. of Eur. Conf. Comput. Vis.*, 2020. 2, 6, 7, 5

[6] Zehao Chen, Qian Zheng, Peisong Niu, Huajin Tang, and Gang Pan. Indoor Lighting Estimation Using an Event Camera. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 3

[7] Jason Chui, Simon Klenk, and Daniel Cremers. Event-Based Feature Tracking in Continuous Time with Sliding Window Optimization. *arxiv preprint arxiv:2107.04536*, 2021. 2

[8] O. Drbohlav and M. Chaniler. Can two specular pixels calibrate photometric stereo? In *Proc. of Int. Conf. Comput. Vis.*, 2005. 2

[9] Ondřej Drbohlav and Radim Šára. Specularities Reduce Ambiguity of Uncalibrated Photometric Stereo. In *Proc. of Eur. Conf. Comput. Vis.*, 2002. 2

[10] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. 1

[11] Georghiades. Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In *Proc. of Int. Conf. Comput. Vis.*, 2003. 2

[12] Heng Guo, Fumio Okura, Boxin Shi, Takuya Funatomi, Yasuhiro Mukaigawa, and Yasuyuki Matsushita. Multispectral photometric stereo for spatially-varying spectral reflectances: A well posed problem? In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 3

[13] Jin Han, Yuta Asano, Boxin Shi, Yinqiang Zheng, and Imari Sato. High-Fidelity Event-Radiance Recovery Via Transient Event Frequency. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 3

[14] Hideki Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 11:3079–3089, 1994. 1, 2, 3

[15] Micah K Johnson and Edward H Adelson. Shape estimation in natural illumination. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2011. 2, 4

[16] Yakun Ju, Xinghui Dong, Yingyu Wang, Lin Qi, and Junyu Dong. A dual-cue network for multispectral photometric stereo. *Pattern Recogn.*, 2020. 3

[17] Dongyeop Kang, Yu Jin Jang, and Sangchul Won. Development of an inspection system for planar steel surface using multispectral photometric stereo. *Optic. Eng.*, 2013. 3

[18] Hyeongwoo Kim, Bennett Wilburn, and Moshe Ben-Ezra. Photometric Stereo for Dynamic Surface Orientations. In *Proc. of Eur. Conf. Comput. Vis.*, 2010. 2

[19] Leonid L Kontsevich, AP Petrov, and IS Vergelskaya. Reconstruction of shape from shading in color images. *J. Opt. Soc. Am. A*, 1994. 3

[20] Junxuan Li and Hongdong Li. Self-calibrating Photometric Stereo by Neural Inverse Rendering. In *Proc. of Eur. Conf. Comput. Vis.*, 2022. 2, 6

[21] Feng Lu, Yasuyuki Matsushita, Imari Sato, Takahiro Okabe, and Yoichi Sato. Uncalibrated Photometric Stereo for Unknown Isotropic Reflectances. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2013. 2

[22] Tom Malzbender, Bennett Wilburn, Dan Gelb, and Bill Ambrisco. Surface Enhancement Using Real-Time Photometric Stereo and reflectance transformation. In *Proc. of Eurographics Symp. Render. Tech.*, 2006. 2

[23] Nathan Matsuda, Oliver Cossairt, and Mohit Gupta. MC3D: Motion contrast 3D scanning. In *Proc. of Int. Conf. Comput. Photogr.*, 2015. 3

[24] Manasi Muglikar, Guillermo Gallego, and Davide Scaramuzza. ESL: Event-based structured light. In *Proc. of Int. Conf. 3D Vis.*, 2021. 3

[25] Manasi Muglikar, Leonard Bauersfeld, Diederik Paul Moeys, and Davide Scaramuzza. Event-based shape from polarization. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 3

[26] Giljoo Nam and Min H Kim. Multispectral photometric stereo for acquiring high-fidelity surface normals. *IEEE Comput. Graphics Appl.*, 2014. 3

[27] Thoma Papadhimitri and Paolo Favaro. A New Perspective on Uncalibrated Photometric Stereo. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2013. 2

[28] Thoma Papadhimitri and Paolo Favaro. A Closed-Form, Consistent and Robust Solution to Uncalibrated Photometric Stereo Via Local Diffuse Reflectance Maxima. *Int. J. Comput. Vis.*, 107(2):139–154, 2014. 2, 6, 7, 5

[29] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. ESIM: an open event camera simulator. In *Proc. of Conf. on Robotics Learning*, 2018. 6, 4

[30] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi. DiLiGenT10^2: A photometric stereo benchmark dataset with controlled shape and material variation. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 2

[31] Wonjeong Ryoo, Giljoo Nam, Jae-Sang Hyun, and Sangpil Kim. Event fusion photometric stereo network. *Neural Networks*, 2023. 3

[32] Boxin Shi, Yasuyuki Matsushita, Yichen Wei, Chao Xu, and Ping Tan. Self-calibrating photometric stereo. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2010. 2

[33] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2016. 2, 3

[34] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019. 2, 6, 7

[35] Melvyn L. Smith and Lyndon N. Smith. Dynamic photometric stereo - a new technique for moving surface analysis. *Image Vision Comput.*, 2005. 3

[36] Kalyan Sunkavalli, Todd Zickler, and Hanspeter Pfister. Visibility Subspaces: Uncalibrated Photometric Stereo with Shadows. In *Proc. of Eur. Conf. Comput. Vis.*, 2010. 2

[37] Tsuyoshi Takatani, Yasuyuki Matsushita, Stephen Lin, Yasuhiro Mukaigawa, and Yasushi Yagi. Enhanced photometric stereo with multispectral images. In *Proc. of Int. Conf. on Mach. Vis. Appl.*, 2013. 3

[38] Ping Tan, Satya P. Mallick, Long Quan, David J. Kriegman, and Todd Zickler. Isotropy, Reciprocity and the Generalized Bas-Relief Ambiguity. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2007. 2

[39] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Ultimate SLAM? combining events, images, and imu for robust visual slam in HDR and high-speed scenarios. *IEEE Robot. Autom. Lett.*, 2018. 2

[40] Daniel Vlasic, Pieter Peers, Ilya Baran, Paul Debevec, Jovan Popović, Szymon Rusinkiewicz, and Wojciech Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM Trans. Graph.*, 28(5):174, 2009. 2

[41] Feishi Wang, Jieji Ren, Heng Guo, Mingjun Ren, and Boxin Shi. DiLiGenT-Pi: Photometric stereo for planar surfaces with rich details-benchmark dataset and beyond. In *Proc. of Int. Conf. Comput. Vis.*, 2023. 2

[42] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 1980. 1, 2

[43] Zhe Wu and Ping Tan. Calibrating Photometric Stereo by Holistic Reflectance Symmetry Analysis. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2013. 2

[44] Takuya Yoda, Hajime Nagahara, Rin-ichiro Taniguchi, Keiichiro Kagawa, Keita Yasutomi, and Shoji Kawahito. The dynamic photometric stereo method using a multi-tap cmos image sensor. *Sensors-basel.*, 2018. 2

[45] Bohan Yu, Jieji Ren, Jin Han, Feishi Wang, Jinxiu Liang, and Boxin Shi. EventPS: Real-Time Photometric Stereo Using an Event Camera. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 2, 3, 6, 7, 5

[46] A.L. Yuille, D. Snow, R. Epstein, and P.N. Belhumeur. Determining Generative Models of Objects Under Varying Illumination: Shape and Albedo from Multiple Images Using SVD and Integrability. *Int. J. Comput. Vis.*, 35(3):203–222, 1999. 2, 3, 5

[47] Xinyu Zhou, Peiqi Duan, Boyu Li, Chu Zhou, Chao Xu, and Boxin Shi. EvDiG: Event-guided Direct and Global Components Separation. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 3

[48] Zhenglong Zhou and Ping Tan. Ring-Light Photometric Stereo. In *Proc. of Eur. Conf. Comput. Vis.*, 2010. 2

EventUPS: Uncalibrated Photometric Stereo Using an Event Camera

Supplementary Material

Jinxiu Liang[1,2,3#] Bohan Yu[1,2#] Siqi Yang[1,2,4] Haotian Zhuang[5]
Jieji Ren[6] Peiqi Duan[1,2] Boxin Shi[1,2*]

[1] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University
[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
[3] National Institute of Informatics [4] Institute for Artificial Intelligence, Peking University
[5] Tsinghua University [6] Shanghai Jiaotong University

cssherryliang@gmail.com,{ybh1998,yousiki,duanqi0001,shiboxin}@pku.edu.cn,
zht23@mails.tsinghua.edu.cn,jiejiren@sjtu.edu.cn

6. Proof of GBR Resolving

We demonstrate that a lighting system with known 3D structure, defined by its 3D light positions, can resolve the GBR ambiguity up to the classic convex/concave reflection, provided certain generic conditions on the light positions are met.

Theorem 1. *Let P be a set of at least seven light source positions in $\mathbb{R}^3$, represented in homogeneous coordinates as $\mathbf{p} = [x, y, z, 1]^\top$. The corresponding light vector for each source is $\mathbf{s} = u \cdot [x, y, z]^\top$, where $u > 0$ is a light-specific scalar. Suppose the following relation holds for every $\mathbf{p} \in P$:*

$$\begin{bmatrix} v \cdot \mathbf{Gs} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{p}$$

where:

1. *$v \neq 0$ is a scalar that may vary for each light source.*
2. *$\boldsymbol{R}$ is a 3×3 orthogonal matrix ($\boldsymbol{R} \in O(3)$, $\det \boldsymbol{R} = \pm1$) and $\mathbf{t}$ is a 3×1 translation vector.*
3. *$\mathbf{G}$ is a 3×3 GBR transformation matrix:*

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & n & p \end{bmatrix}.$$

4. *The set P is in general position, meaning the points do not all lie on any quadric surface passing through the origin with equation:*

$$\pi_1 x^2 + \pi_2 xy + \pi_3 xz + \pi_4 x + \pi_5 y^2 + \pi_6 yz + \pi_7 y = 0$$

where not all π_i are zero.

Under these conditions, the only solutions for $\mathbf{t}$ is $\mathbf{t} = \mathbf{0}$, and the only solutions for $(\mathbf{R}, \mathbf{G})$ are the four coupled pairs:

- $(\mathbf{R}, \mathbf{G}) = (\mathrm{diag}(1, 1, 1), \mathrm{diag}(1, 1, 1))$,
- $(\mathbf{R}, \mathbf{G}) = (\mathrm{diag}(1, 1, -1), \mathrm{diag}(1, 1, -1))$,
- $(\mathbf{R}, \mathbf{G}) = (\mathrm{diag}(-1, -1, 1), \mathrm{diag}(1, 1, -1))$,
- $(\mathbf{R}, \mathbf{G}) = (\mathrm{diag}(-1, -1, -1), \mathrm{diag}(1, 1, 1))$.

Remark. *A common physical constraint in photometric stereo is that the lights must remain in the same hemisphere relative to the scene (*i.e.*, always in front of the camera, $z > 0$). This implies that the z-component of a light's position must have the same sign in both the true and ambiguous configurations. This prunes the four solutions to two: the first and the third solutions.*

Proof. Let $\mathbf{R} = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$ and $\mathbf{t} = \begin{bmatrix} j \\ k \\ l \end{bmatrix}$.

For a point $\mathbf{p} = [x, y, z, 1]^\top$, the left-hand side becomes:

$$\mathbf{Gs} = u \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ m & n & p \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = u \begin{bmatrix} x \\ y \\ mx + ny + pz \end{bmatrix}.$$

Thus:

$$\begin{bmatrix} v \cdot \mathbf{Gs} \\ 1 \end{bmatrix} = [vux, vuy, vu(mx + ny + pz), 1]^\top.$$

The right-hand side gives:

$$\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} = \begin{bmatrix} a & d & g & j \\ b & e & h & k \\ c & f & i & l \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} =$$

$$[xa + yd + zg + j, xb + ye + zh + k, xc + yf + zi + l, 1]^\top.$$

Equating the vector components of both sides gives three

#Equal contributions. *Corresponding author.

scalar equations that must hold for each point in P:

$$vux = xa + yd + zg + j, \tag{18}$$
$$vuy = xb + ye + zh + k, \tag{19}$$
$$vu(mx + ny + pz) = xc + yf + zi + l. \tag{20}$$

The scalar vu varies with each light source and must be eliminated to obtain constraints on the matrices. We can isolate vu from (18) and (19) (assuming at least one point has $x \neq 0$ and $y \neq 0$, which is guaranteed by the general position condition) by cross-multiplication. From equations (18) and (19), multiplying (18) by y and (19) by x:

$$vuxy = y(xa + yd + zg + j),$$
$$vuxy = x(xb + ye + zh + k).$$

Equating the right-hand sides:

$$y(xa + yd + zg + j) = x(xb + ye + zh + k).$$

Rearranging the terms produces the equation of a quadric surface:

$$bx^2 + (e-a)xy + hxz + kx - dy^2 - gyz - jy = 0. \tag{21}$$

This equation holds for every point in P and passes through the origin $(0, 0, 0)$ (substituting $x = y = z = 0$ yields $0 = 0$).

Note that this quadric has a special form: it contains no z^2 or standalone z terms, which is a direct consequence of our elimination process.

By the theorem's premise, the derived quadric (21) is precisely of the form excluded by the general position condition. The vector of its coefficients is $\boldsymbol{\pi} = [b, e - a, h, k, -d, -g, -j]^\top$. For each point $\mathbf{p}_i = [x_i, y_i, z_i, 1]^\top \in P$, we obtain a linear constraint:

$$[x_i^2, x_iy_i, x_iz_i, x_i, y_i^2, y_iz_i, y_i]\boldsymbol{\pi} = 0.$$

Stacking these constraints for seven points yields the homogeneous system $\mathbf{M}\boldsymbol{\pi} = \mathbf{0}$, where $\mathbf{M}$ is a 7×7 matrix with rows $[x_i^2, x_iy_i, x_iz_i, x_i, y_i^2, y_iz_i, y_i]$.

Since the seven or more points in P do not lie on such a surface (unless it is trivial), the coefficients must all be zero, *i.e.*, $\boldsymbol{\pi} = \mathbf{0}$, therefore,

$$b = 0, e - a = 0, h = 0, k = 0, d = 0, g = 0, j = 0.$$

This simplifies our matrices to:

$$\mathbf{R} = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ c & f & i \end{bmatrix}, \quad \mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ l \end{bmatrix}.$$

Since $\boldsymbol{R} \in O(3)$ is orthogonal, we have $\mathbf{R}^\top\mathbf{R} = \mathbf{I}_3$. Computing the rows: Row 1: $[a, 0, 0]$ has norm $|a| = 1$, so $a = \pm 1$.

Row 2: $[0, a, 0]$ has norm $|a| = 1$ (confirmed).

Row 3: $[c, f, i]$ must have unit norm and be orthogonal to rows 1 and 2:

- Orthogonality conditions:
  - With row 1: $ac + 0 \cdot f + 0 \cdot i = 0 \Rightarrow c = 0$ (since $a \neq 0$).
  - With row 2: $0 \cdot c + af + 0 \cdot i = 0 \Rightarrow f = 0$ (since $a \neq 0$).
- Unit norm constraint: $c^2 + f^2 + i^2 = 0^2 + 0^2 + i^2 = 1 \Rightarrow i = \pm 1$.

Therefore: $\mathbf{R} = \mathrm{diag}(a, a, i)$ where $a, i \in \{-1, 1\}$.

Now we use the simplified forms of $\mathbf{R}$ and $\mathbf{t}$ in the initial scalar equations. From equation (18) with $d = g = j = 0$: $vux = ax$.

This yields $vu = a$ for points with $x \neq 0$ or $y \neq 0$. Substituting $vu = a$, $c = f = 0$, and the simplified $\mathbf{R}$ into equation (20):

$$a(mx + ny + pz) = zi + l.$$

Rearranging gives the equation of a plane:

$$amx + any + (ap - i)z - l = 0.$$

This equation must hold for all points in P. However, the points in P cannot be coplanar. If they were, they would lie on a plane $Ax + By + Cz + D = 0$. This would imply they also lie on the quadric surface defined by $x(Ax + By + Cz + D) = Ax^2 + Bxy + Cxz + Dx = 0$, which is one of the forms forbidden by the general position condition.

Since the points are not coplanar, the only way the planar equation can hold for all of them is if all its coefficients are zero:

- $am = 0 \implies m = 0$ (since $a \neq 0$),
- $an = 0 \implies n = 0$ (since $a \neq 0$),
- $ap - i = 0 \implies p = i/a = ia$ (since $a^2 = 1$),
- $-l = 0 \implies l = 0$.

For points with $x = y = 0$ (if any), consistency is verified after deriving the coefficients: substituting into (20) gives $vup = i + l/z$, but after zeroing coefficients ($m = n = 0$, $p = ia$, $l = 0$), we obtain $vu = a$, consistent with other regions.

This fully determines the parameters of $\mathbf{G} = \mathrm{diag}(1, 1, ai)$ and $\mathbf{t} = \mathbf{0}$.

We have $\mathbf{R} = \mathrm{diag}(a, a, i)$ and $\mathbf{G} = \mathrm{diag}(1, 1, ai)$ where $a, i \in \{-1, 1\}$.

The four possibilities are:

- $(a, i) = (1, 1)$: $\mathbf{R} = \mathrm{diag}(1, 1, 1)$, $\mathbf{G} = \mathrm{diag}(1, 1, 1)$,
- $(a, i) = (1, -1)$: $\mathbf{R} = \mathrm{diag}(1, 1, -1)$, $\mathbf{G} = \mathrm{diag}(1, 1, -1)$,
- $(a, i) = (-1, 1)$: $\mathbf{R} = \mathrm{diag}(-1, -1, 1)$, $\mathbf{G} = \mathrm{diag}(1, 1, -1)$,
- $(a, i) = (-1, -1)$: $\mathbf{R} = \mathrm{diag}(-1, -1, -1)$, $\mathbf{G} = \mathrm{diag}(1, 1, 1)$.

Existence of general position point sets.

To ensure the theorem is not vacuous, we must show that a set of points P satisfying the general position condition exists. The condition requires that the only solution to $\mathbf{M}\boldsymbol{\pi} = \mathbf{0}$ is $\boldsymbol{\pi} = \mathbf{0}$, where $\mathbf{M}$ is the 7×7 matrix whose rows are $[x_i^2, x_i y_i, x_i z_i, x_i, y_i^2, y_i z_i, y_i]$ for the seven points. This is equivalent to showing that we can find 7 points such that $\mathbf{M}$ is invertible.

We can construct such a set:

1. Choose 5 points on the plane $z = z_0$ (with $z_0 \neq 0$). Let their (x, y) coordinates be five points on a circle that does not pass through the origin. Let $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_7]^\top$ be the coefficient vector of the forbidden quadric. On this plane, any quadric of our form becomes:

$$\pi_1 x^2 + \pi_2 xy + \pi_5 y^2 + (\pi_3 z_0 + \pi_4)x + (\pi_6 z_0 + \pi_7)y = 0.$$

   This is a conic section in the xy-plane passing through the origin. However, five points uniquely define a conic. The unique conic passing through our five chosen points is the circle we constructed, which does not pass through the origin. This is a contradiction unless all the coefficients of the conic are zero. Therefore:

$$\begin{aligned} &\pi_1 = \pi_2 = \pi_5 = 0, \\ &\pi_3 z_0 + \pi_4 = 0 \implies \pi_4 = -\pi_3 z_0, \\ &\pi_6 z_0 + \pi_7 = 0 \implies \pi_7 = -\pi_6 z_0. \end{aligned}$$

   This means that for any quadric of the forbidden form to pass through these five points, its equation must simplify to: $\pi_3 xz - \pi_3 z_0 x + \pi_6 yz - \pi_6 z_0 y = 0$, which can be factored as $(\pi_3 x + \pi_6 y)(z - z_0) = 0$. This is satisfied for our first five points, as $z - z_0 = 0$.

2. Choose 2 additional points, $\mathbf{p}_6$ and $\mathbf{p}_7$, on a different plane (e.g., $z = z_1$, where $z_1 \neq z_0$ and $z_1 \neq 0$). For these points, $z - z_0 \neq 0$, so the equation reduces to $\pi_3 x + \pi_6 y = 0$:

$$\begin{aligned} \pi_3 x_6 + \pi_6 y_6 = 0, \\ \pi_3 x_7 + \pi_6 y_7 = 0. \end{aligned}$$

   We can choose $\mathbf{p}_6$ and $\mathbf{p}_7$ such that their projection vectors $[x_6, y_6]^\top$ and $[x_7, y_7]^\top$ are linearly independent (i.e., not collinear and not the zero vector). This ensures that the only solution to the 2×2 system above is the trivial one: $\pi_3 = 0$ and $\pi_6 = 0$. Substituting back, we find $\pi_4 = -\pi_3 z_0 = 0$ and $\pi_7 = -\pi_6 z_0 = 0$.

This construction forces all coefficients $\pi_1, \ldots, \pi_7$ to be zero. Therefore, the only quadric of the specified form passing through these seven points is the trivial one, meaning the points satisfy the general position condition. Such sets exist, and the proof is complete. $\square$

7. More Implementation Details

7.1. Lighting Configurations

We designed two configurations with known relative geometry but unknown global pose to resolve the UPS ambiguity.

Case 1: Coaxial dual-ring lights. This configuration uses two concentric rings of lights with known radii $r_1 > 0$ and $r_2 > 0$ and a known axial separation $\Delta d \neq 0$. The lights' canonical positions in the local frame are:

$$\boldsymbol{p}_i(t) = [r_i \cos t, r_i \sin t, z_i], \tag{22}$$

where $z_1 = 0, z_2 = \Delta d$ denotes the axial offset, and $t \in [0, 2\pi)$ (*e.g.*, equally spaced for M sources) parametrizes the angular position around the ring.

In this system, these points span two parallel planes, encoding depth via differences Δd in lighting directions between rings. Rings are circles centered on the z-axis, not passing through the origin. This mirrors the proof's construction: points on two $z =$ constant planes with circular distributions not through the origin.

We validate it satisfies the general position as follows, drawing on the proof's construction (points on two planes with circles not through origin, extended to non-coplanar sets): Restrict to one plane ($z = 0$); the forbidden quadric becomes a conic in xy passing through $(0, 0)$ (no constant term). With ≥ 5 points on a circle not through $(0, 0)$, the unique conic through them has a constant term, so no matching conic without constant exists unless trivial. Adding points on $z = \Delta d$ forces remaining coefficients to zero (by linear independence of $[x_k, y_k]$). The 3D distribution resolves translations (axial separation fixes z-shift; angular spread fixes xy-shifts). Thus, the general position holds for generic $r_1 \neq r_2, \Delta d \neq 0$, and non-collinear angular positions.

Case 2: Trefoil curve. This configuration uses a single ring of lights mounted on a tilted, rotating mechanism to trace a non-planar, asymmetric 3D trefoil curve. The light positions are defined by:

$$\begin{aligned} \boldsymbol{p}(t) = &\begin{bmatrix} (r_1 + r_2 \sin(\omega t + \phi)\cos(\alpha))\cos(t) \\ (r_1 + r_2 \sin(\omega t + \phi)\cos(\alpha))\sin(t) \\ -r_2 \sin(\omega t + \phi)\sin(\alpha) \end{bmatrix} + \\ &\begin{bmatrix} -r_2 \cos(\omega t + \phi)\sin(t) \\ r_2 \cos(\omega t + \phi)\cos(t) \\ 0 \end{bmatrix} \end{aligned} \tag{23}$$

where r_1, r_2 are radii, α is the tilt angle, ω is a frequency ratio, $t \in [0, 2\pi)$ parametrizes the curve, and ϕ is a phase shift. This structure combines circular motion in the xy-plane (radius r_1) and a secondary oscillation (amplitude r_2, tilt α, frequency ω, phase ϕ) that traces a 3D trefoil

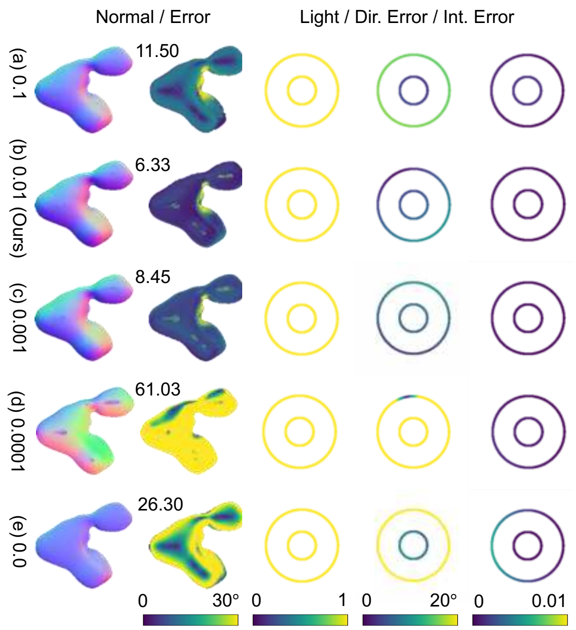

Figure 9. Effectiveness of the integrability constraint. As the weight λ_{int} decreases from (a) 0.1 to (e) 0, the MAE sharply increases. Without this constraint ($\lambda_{\text{int}} = 0$), the solution is ambiguous, leading to errors in both normal and lighting estimation. Our chosen value of (b) $\lambda_{\text{int}} = 0.01$ provides the best balance.

curve. The curve is non-planar (z-component varies with $\sin(\omega t + \phi)$), asymmetric (trefoil pattern breaks rotational symmetry), and spans $\mathbb{R}^3$. Sampled points (≥ 7, equally spaced in t) are in general position if parameters ensure no accidental alignment on low-degree surfaces.

We validate it satisfies the general position as follows: The curve avoids coplanarity (z-variation) and special quadrics. For generic parameters (*e.g.*, $r_1 \neq r_2 > 0$, $\alpha \neq 0$, $\omega \neq 1$, $\phi \neq 0$), sampled points do not satisfy any nontrivial forbidden quadric. This follows from perturbation: start with the proof's construction (non-coplanar points not on such quadrics), and note the trefoil is a continuous deformation preserving genericity. The asymmetry and 3D spread resolve rotations/translations. Numerically, for specific parameters (*e.g.*, $r_1 = 1$, $r_2 = 0.5$, $\alpha = \frac{\pi}{4}$, $\omega = 3$, $\phi = 0$), sampling 7 points yields a full-rank $\boldsymbol{M}$ (verifiable via SVD; rank deficiency would require unlikely parameter tuning). Thus, the general position holds.

Both configurations satisfy the requirements. They are continuous (parameterized by $t \in [0, 2\pi]$), but in practice, ≥ 7 discrete points are placed (equally placed). Both configurations are realizable with off-the-shelf components and are suitable for handheld operation.

7.2. B-Spline Coefficients

For the lighting estimation with continuous representation, we use temporal coefficients $C_{j,l}(t)$ defined by the standard

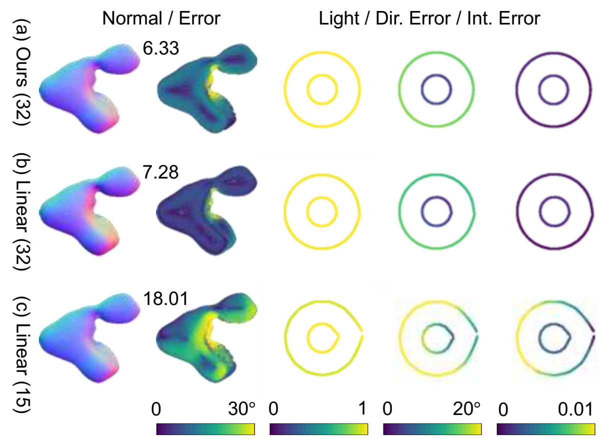

Figure 10. Ablation on continuous lighting representation. We compare (a) our cubic B-spline model with (b, c) a simpler linear interpolation model. Using (b) linear interpolation with the same number of control points (32) results in an increase in error. Reducing the knots to (c) 15 leads to a degradation in performance, as the model can no longer capture the smooth lighting variation.

De Boor–Cox recurrence relation for a spline of order l (degree $l-1$):

$$C_{j,0}(t) = \begin{cases} 1 & \text{if } t_j \leq t < t_{j+1}, \\ 0 & \text{otherwise}, \end{cases} \tag{24}$$

and

$$C_{j,l}(t) = \frac{t - t_j}{l\Delta t} C_{j,l-1}(t) + \frac{t_{j+l+1} - t}{l\Delta t} C_{j+1,l-1}(t). \tag{25}$$

8. Further Analysis and Ablation Studies

Synthetic data generation. To enable comprehensive evaluation, we created a synthetic dataset using objects from the Blobby dataset [15]. For each object, we rendered 500 dense images under rotating lighting conditions for 6 complete rounds, using the dual ring scanning pattern. These images were then converted to event streams using the ESIM simulator [29].

Effectiveness of the integrability constraint. As shown in Fig. 9, the integrability constraint is crucial for resolving ambiguity. Removing it ($\lambda_{\text{int}} = 0$) causes the average MAE to increase by over 18%, demonstrating its effectiveness in pruning incorrect solutions allowed by the GBR ambiguity.

Effectiveness of the continuous lighting representation. We ablated our B-spline model to validate its design. Fig. 10 shows that a cubic spline significantly outperforms linear interpolation, which yields higher MAE with the same number of control points. This validates our choice of a cubic spline representation for handling asynchronous data.

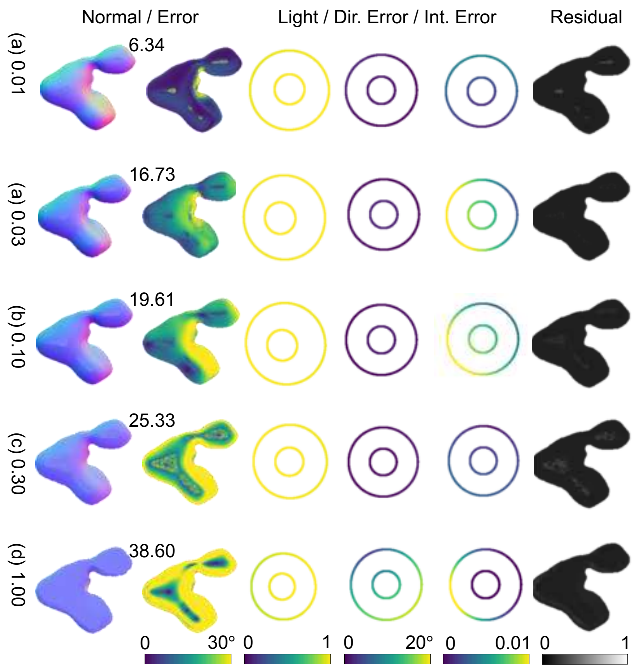

Figure 11. Robustness to ambient illumination. We test our method on synthetic data with varying ratios of ambient to direct illumination. Performance remains stable for ratios up to 0.10, and degrades gracefully thereafter.

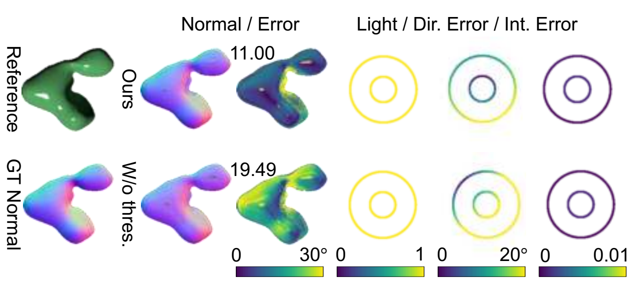

Figure 12. Robustness to specular highlights, a common non-Lambertian effect. Our EventUPS filters high-frequency events that often correspond to specularities. The top row shows the result with filtering, achieving a low MAE. The bottom row shows the result without filtering, where specular events introduce significant artifacts and increase the error.

Robustness to ambient illumination. Fig. 11 demonstrates our method's strong robustness to ambient light—a critical feature for real-world deployment. Our augmented null space formulation explicitly models this effect, allowing EventUPS to maintain high accuracy even when the ambient light is 10% as strong as the direct illumination. Performance degrades gracefully, far surpassing methods that assume dark-room conditions.

Robustness to non-Lambertian effects. Following [45], we filter high-frequency events (often due to specular highlights or sharp shadow boundaries) using thresholding. As shown in Fig. 12, this effectively mitigates artifacts from moderate specular highlights, demonstrating that our method's robustness.

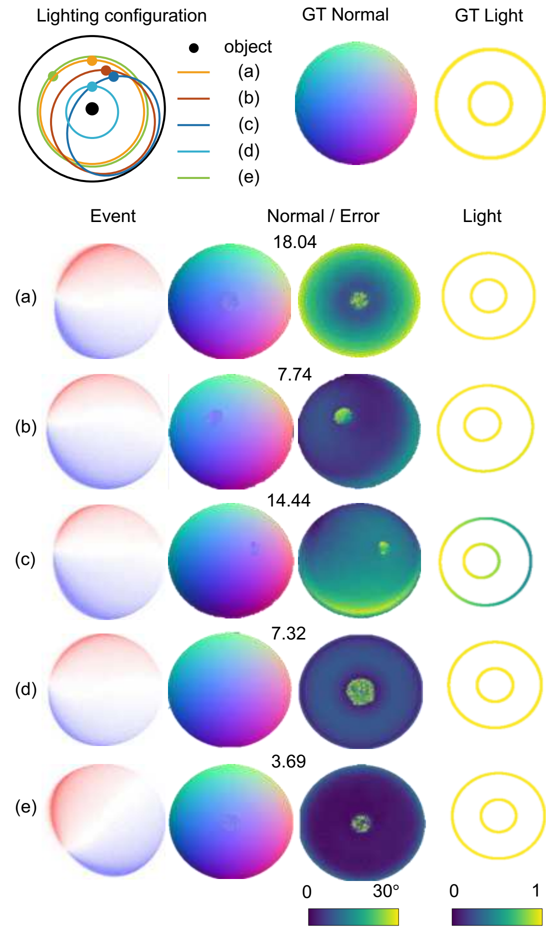

Figure 13. Robustness to the lighting system pose. We simulated various transformations relative to the camera: (a) a view-centered baseline, (b) rotation around the object, (c) rotation about the light's axis, (d) translation along the z-axis, and (e) rotation around the z-axis. The top-left plot shows the projected light path for each case. EventUPS maintains consistent, high-quality performance across all configurations.

Robustness to lighting system pose. Fig. 13 confirms that our method is resilient to the global pose of the lighting system. This robustness is critical for practical use cases, especially handheld operation, where precise alignment is not feasible.

9. Complete Evaluation Results

Complete visual results of our method, EventPS [45], PF14 [28] (56 frames), and CW20 [5] (6 frames) for all objects in the DiLiGenT-Ev dataset are shown in Fig. 14.

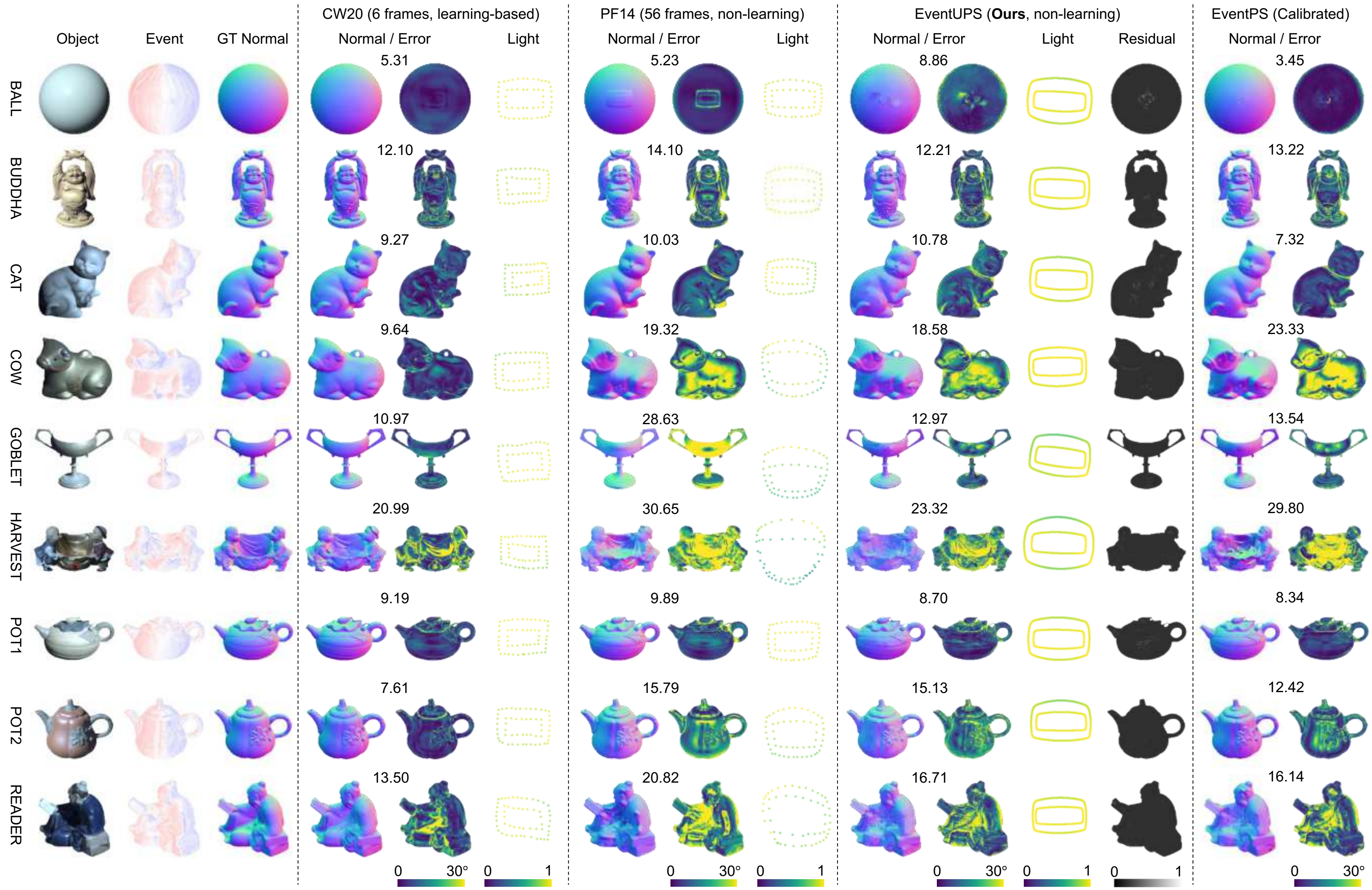

Figure 14. Complete visual results on the DiLiGenT-Ev semi-real dataset, showing the full set of reconstructions for all objects.