EventPSR: Surface Normal and Reflectance Estimation from Photometric Stereo Using an Event Camera

Bohan Yu[1,2†] Jin Han[3,4] Boxin Shi[1,2*] Imari Sato[3,4]

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

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

[3] The University of Tokyo [4] National Institute of Informatics

ybh1998@pku.edu.cn, jinhan@nii.ac.jp, shiboxin@pku.edu.cn, imarik@nii.ac.jp

Abstract

Simultaneously acquisition of the surface normal and reflectance parameters is a crucial but challenging technique in the field of computer vision and graphics. It requires capturing multiple high dynamic range (HDR) images in existing methods using frame-based cameras. In this paper, we propose EventPSR, the first work to recover surface normal and reflectance parameters (e.g., metallic and roughness) simultaneously using an event camera. Compared with the existing methods based on photometric stereo or neural radiance fields, EventPSR is a robust and efficient approach that works consistently with different materials. Thanks to the extremely high temporal resolution and high dynamic range coverage of event cameras, EventPSR can recover accurate surface normal and reflectance of objects with various materials in 10 *seconds. Extensive experiments on both synthetic data and real objects show that compared with existing methods using more than* 100 *HDR images, EventPSR recovers comparable surface normal and reflectance parameters with only about* 30% *of the data rate.*

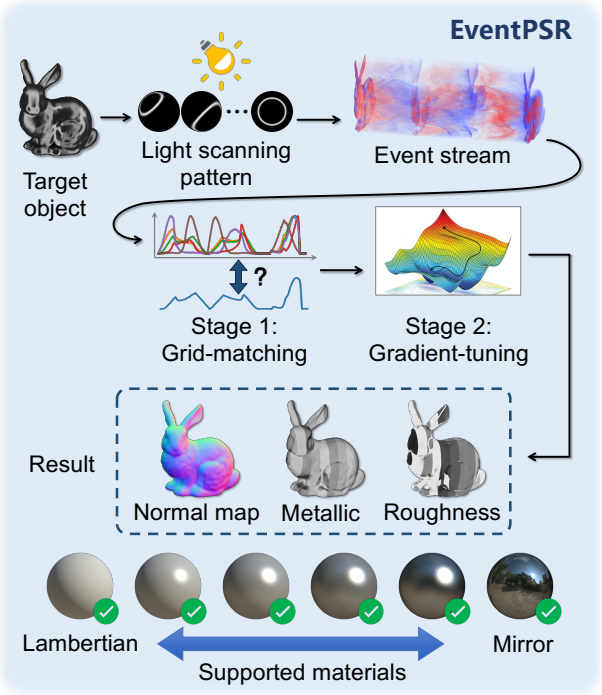

Figure 1. The EventPSR applies a sequence of light scanning patterns and an event camera to recover the accurate normal map and reflectance parameters (*i.e.*, metallic and roughness) simultaneously. A two-stage pipeline consisting of a grid-matching and a gradient-tuning stage is proposed for EventPSR, which supports shape and reflectance recovery for different materials.

1. Introduction

The fast and accurate acquisition of the shape and appearance information of an object is essential for many tasks in computer vision and computer graphics. This technique enables re-rendering realistic digital twins of real-world objects, supporting a wide range of downstream applications, such as virtual reality, augmented reality (AR), and digital content creation. However, dynamic range is a big issue when the material is composed of diffuse and highly specular properties. The time-consuming capturing process and difficulty recovering highly specular surfaces are two major challenges in this area.

Traditional shape and reflectance estimation methods [1, 3, 10, 15, 28, 31, 32] rely on multiple high dynamic range (HDR) images captured under varying lighting conditions or viewing angles. Different from photometric stereo (PS) that estimates the surface normal of diffuse dominant ob-

*Corresponding author: Boxin Shi

†This work was conducted while the first author, Bohan Yu, was doing internship at National Institute of Informatics.

[1]Code available: https://codeberg.org/ybh1998/EventPSR

jects from several images, shape and appearance estimation methods [1, 15] require more images to ensure the accurate observation of various reflections, which is time-consuming especially for shiny materials recovery. It is difficult to fully cover the specular highlights in shiny and metallic (*i.e.*, mirror-like) materials under directional light without complicated capturing setup [15] and multiple images to cover the dynamic range. Such a dilemma between speed and accuracy partly attributes to the insufficient capability of conventional frame-based cameras.

Event cameras characterized by extremely high temporal resolution and dynamic range, have shown promising advantages in many computer vision applications [9]. Researchers have explored applying event cameras in various photometric tasks [6, 19, 29, 33]. Particularly, EventPS [29] has demonstrated the superiority of event cameras for real-time surface normal estimation, showing the potential to accelerate and improve accuracy for shape and reflectance estimation tasks. However, it is still unclear *how to extract material-related parameters, including roughness and metallic, from the event stream only.*

In this paper, we propose **EventPSR**, a novel solution for simultaneous estimation of both surface normal and **R**eflectance properties from **P**hotometric **S**tereo using an **Event** camera. As shown in Fig. 1, EventPSR can simultaneously recover an object's shape and reflectance properties[2] with designed light scanning patterns and a two-stage algorithm, which supports the recovery of different materials. Directly applying the setting and algorithm from EventPS [29] (*i.e.*, capturing objects with a fixed event camera under a fast-moving point light source) is a non-trivial task. It is impossible to estimate the shape and reflectance at the same time without conquering the two challenges: 1) Design a suitable light scanning pattern that covers various materials and fully reveals the shape and reflectance properties of the objects. 2) Propose a new algorithm that converts the event signals to accurate normal and reflectance parameters simultaneously under the light scanning pattern.

To conquer the first challenge, we discover that there are three requirements that need to be satisfied for a suitable light scanning pattern: specular coverage, diffuse sensitivity, and event efficiency. Based on this, we design a pattern of "1D strip light" that covers the full sphere of the illumination direction space. It is designed not only for various materials ranging from ideal Lambertian to pure mirror, but also for efficient capture of event cameras with faster speed and low data rate. To conquer the second challenge, we propose a two-stage algorithm to estimate surface normal and reflectance parameters, consisting of a grid-matching stage and a gradient-tuning stage. We first render a database consisting of the intensity-changing curves, including the combinations of various surface normal directions and materials. In the grid-matching stage, we match the event stream triggered in each pixel to a specific curve in the database and get a group of coarse reflectance parameters. In the gradient-tuning stage, we finetune the estimated parameters to a more accurate sub-grid location. To summarize, the contributions of this paper are as follows:

- We propose EventPSR, the first solution to estimate surface normal and reflectance parameters simultaneously from continuous radiance differential event signals.
- We design a suitable light scanning pattern and an efficient two-stage algorithm with only 30% data rate compared with frame-based counterparts.
- We validate the effectiveness of EventPSR by a platform with multiple monitors. The results show that EventPSR can estimate accurate shape-related and material-related parameters of real objects.

2. Related Works

Our method is related to PS works, and comprehensive surveys can be found in [14, 25]. This section focuses on works addressing PS and surface reflectometry using conventional frame-based cameras, as well as recent advancements in photometry using event cameras.

Surface reflectometry. Measuring the detailed surface reflectance requires the use of controlled lighting to sample photometric stereo images. Romeiro *et al*. [23] inferred the bi-directional surface reflectance of materials using a light probe and a single HDR image of the known surface normal. Dupuy and Jakob [8] measured the reflectance spectra using a gonio-photometer to simultaneously handle BRDF acquisition, storage, and rendering. Learning-based BRDF acquisition and modeling approaches are summarized in the survey paper [17]. EventPSR differs from this category of research because we simultaneously recover the normal and reflectance parameters.

Normal and reflectance estimation from conventional cameras. Alldrin *et al*. [1] proposed to employ novel bivariate approximations of isotropic reflectance functions to simultaneously recover the shape and reflectance of a surface from photometric stereo images. Chung *et al*. [7] optimized the geometry and spatially-varying BRDFs through a point-based rendering. Learning-based methods applied convolutional neural networks on single image [2, 21, 30] or multiple images [15, 24] to estimate the surface normal and reflectance parameters. These methods are limited in capturing speed due to the low dynamic range and high latency of conventional cameras.

Event camera-based photometry. Event cameras with high temporal resolution and high dynamic range properties have been applied to various photometry tasks in recent years. Muglikar *et al*. [19] put a linear polarizer rotating at high speeds in front of an event camera to reconstruct

[2]In this work, we recover reflectance-related properties specifically with roughness and metallic.

relative intensities at multiple polarizer angles, which were used for surface normals estimation. Han *et al*. [11] computed the frequencies of the event triggering during the transient process of global lighting changes for depth sensing and object iso-contour reconstruction. Zhou *et al*. [33] proposed to use the differential signals captured from an event camera to separate the direct and global illumination components in RGB videos by applying a sweeping occluder. EventPS [29] leveraged an event camera and a fast-rotating light source to achieve real-time surface normal reconstruction, which demonstrated the superiority over the conventional cameras in speed and data bandwidth. EventPSR can further estimate surface normal and reflectance parameters with high robustness and efficiency by designing light scanning patterns and estimation algorithms.

3. Method

As shown in Fig. 1, the overall workflow of EventPSR includes the design of suitable lighting scanning patterns (Sec. 3.2) and a two-stage algorithm that recovers surface normal and reflectance parameters (Sec. 3.3). In Sec. 3.1, we start from the mathematical formulation of event cameras, parameterized reflectance models, and target function.

3.1. Problem Formulation

Event formation model. Event cameras measure the scene brightness changes asynchronously on a logarithmic scale. An event $e = \{t, x, y, p\}$ is triggered when the changes of logarithmic brightness reach a triggering threshold C, where t is the timestamp, (x, y) is the pixel coordinate, and $p \in \{-1, +1\}$ is the polarity which represents the decrease or increase of brightness. Assuming there are totally K events triggered at pixel (x, y) during a short period of time, these events are represented as $\mathcal{E}_{x,y} = \{x, y, p_k, t_k\}$, where $k = \{1, 2, ..., K\}$. The irradiance value I in pixel (x, y) changing from t_{k-1} to t_k is represented as:

$$\log(I_{x,y}(t_k) + \epsilon) = \log(I_{x,y}(t_{k-1} + \eta) + \epsilon) + p_k C, \quad (1)$$

where ϵ is a small offset value to avoid taking the logarithm of zero, and η is the refractory time of the pixel [9].

Reflectance model and inverse problem. We use a simplified version of Disney Principled BRDF [4] model to represent the reflectance ($\mathbf{f}_{r,x,y}$) of the surface. Given an opaque and non-emissive object with a distant light source, the irradiance value $I_{x,y}(t)$ at timestamp t is as follows:

$$I_{x,y}(t) = \int_{\mathbf{w}_i \in \Omega} (\mathbf{N}_{x,y} \cdot \mathbf{w}_\mathrm{i}) \mathbf{f}_{r,x,y}(\mathbf{w}_\mathrm{i}, \mathbf{w}_\mathrm{r}) \mathbf{L}(\mathbf{w}_i, t) \mathrm{d}\mathbf{w}_i, \quad (2)$$

where Ω is the whole sphere direction space, $\mathbf{N}_{x,y}$ is the surface normal, $\mathbf{L}$ is the light source function, $\mathbf{w}_i$ and $\mathbf{w}_r$ are the incident and reflection direction, respectively. The light source function $\mathbf{L}(\mathbf{w}_i, t)$ represents a distant light with the same brightness at an angle of $\mathbf{w}_i$. For each pixel (x, y), its reflectance $\mathbf{f}_{r,x,y}$ is represented as:

$$\mathbf{f}_{r,x,y}(\mathbf{w}_\mathrm{i}, \mathbf{w}_\mathrm{r}) = \frac{1 - M_{x,y}}{\pi} + \frac{M_{x,y} DG}{\pi (\mathbf{N}_{x,y} \cdot \mathbf{w}_\mathrm{i})(\mathbf{N}_{x,y} \cdot \mathbf{w}_\mathrm{r})}, \quad (3)$$

where D, G are parameters that describe specular reflection property.[3] The variation of diffuse albedo has been canceled by the log difference mechanism in event cameras [29], so we set it as 1 in our method. Other parameters and Fresnel term from Disney Principled BRDF [4] are removed for simplicity.

Target function. To recover the surface properties,*i.e*., $\mathbf{N}_{x,y}$ (Surface normal), $R_{x,y}$ (Roughness, in the supplementary material), and $M_{x,y}$ (Metallic), we design a light scanning pattern $\mathbf{L}(\mathbf{w}_i, t)$ and apply an event camera to solve the inverse reflectance estimation problem. We represent the shape and reflection estimation algorithm as $\tilde{\mathbf{N}}(\mathcal{E})$, $\tilde{\mathbf{R}}(\mathcal{E})$, and $\tilde{\mathbf{M}}(\mathcal{E})$. Our target is to design these algorithms and optimize light scanning patterns $\mathbf{L}$ to estimate the surface normal and reflectance parameters:

$$\begin{aligned} \underset{\mathbf{L}, \tilde{\mathbf{N}}, \tilde{\mathbf{R}}, \tilde{\mathbf{M}}}{\operatorname{argmin}} \sum_{\mathbf{N}, M, R} & |\tilde{\mathbf{N}}(\mathcal{E}) - \mathbf{N}|_2 + \\ & |\tilde{\mathbf{M}}(\mathcal{E}) - M|_2 + |\tilde{\mathbf{R}}(\mathcal{E}) - R|_2, \end{aligned} \quad (4)$$

where the events $\mathcal{E}$ are triggered from $\mathbf{L}, \mathbf{N}, R, M$.

3.2. Light Scanning Pattern Design

It is necessary to design a suitable light scanning pattern to cover different materials and the shape of target objects. There are three requirements when designing the light scanning pattern: specular coverage, diffuse sensitivity, and event efficiency. In the following, we consider an effective use of a ring-shaped light source as the base scanning pattern. Taking a ring moving along the Z axis as an example, the formula of $\mathbf{L}(\mathbf{w}_i, t)$ is:

$$\begin{aligned} \mathbf{L}(\mathbf{w}_i, t) &= \max_u \max(0, \mathbf{w}_i \cdot \mathbf{R}(u, t))^\alpha, \\ \mathbf{R}(u, t) &= [\sin(t)\sin(u), \sin(t)\cos(u), \cos(t)], \end{aligned} \quad (5)$$

where $\mathbf{R}(u, t)$ represent all light directions on a ring. α is the width of the ring. In order to illustrate that our "moving ring light" pattern has better properties to overcome all these challenges, we compare it with two other light source patterns: "point light" and "structured environment map".

For the "point light" pattern, we combine two spiral movement trajectories to cover the whole sphere: the horizontal spiral and vertical spiral. In this way, we obtain a continuous moving light source pattern. We also add a

[3] Detailed formulas are shown in the supplementary material.

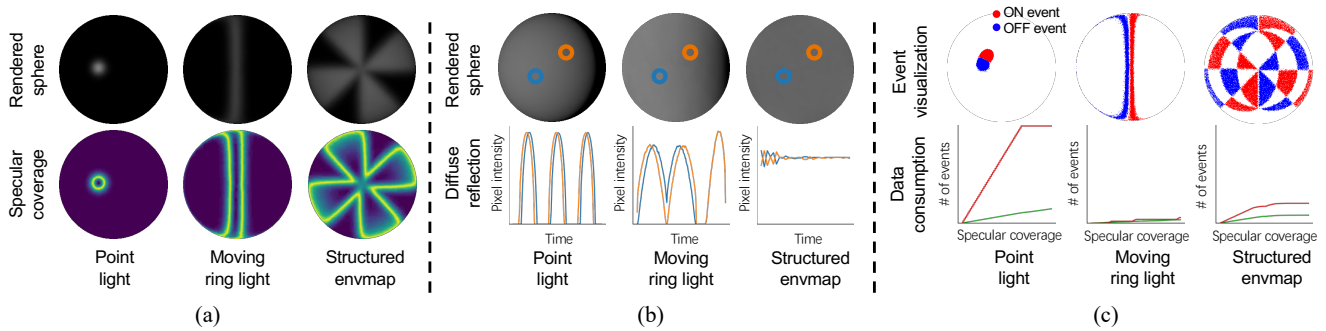

Figure 2. The design of light scanning pattern. (a) Illustration of specular coverage requirement. The first row shows a shiny ball under three lighting patterns. The second row shows where specular parameters can be recovered (the brighter the better). (b) Illustration of diffuse sensitivity requirement. The first row shows the rendered frames of a diffuse ball under three lighting patterns. The second row shows the profiles of intensity changes of two different points on the sphere over time. (c) Illustration of event efficiency requirement. The first row shows the event signals triggered under three lighting patterns. The second row plots the average (green) and max (red) number of events triggered over time when covering the whole specular reflection.

smoothness term ε (the size of the point light), so the lighting will cover the whole sphere lighting direction gradually. For the horizontal spiral, the formula $\mathbf{L}(\mathbf{w}_i, t)$ is:

$$\begin{aligned} \mathbf{L}(\mathbf{w}_i, t) &= (\mathbf{R}(t) \cdot \mathbf{w}_i)^{\alpha}, \\ \mathbf{R}(t) &= [\sin(\beta t)\cos(t), \cos(\beta t)\cos(t), \sin(t)], \end{aligned} \tag{6}$$

where β determines the number of spirals. For the vertical spiral, the formula $\mathbf{L}(\mathbf{w}_i, t)$ is:

$$\begin{aligned} \mathbf{L}(\mathbf{w}_i, t) &= (\mathbf{R}(t) \cdot \mathbf{w}_i)^{\alpha}, \\ \mathbf{R}(t) &= [\cos(\beta t)\sin(t), \cos(\beta t)\cos(t), \sin(\beta t)]. \end{aligned} \tag{7}$$

For the "structured environment map" pattern, we project structured light along three axes alternatively. Taking structured light along the Z axis as an example, the formula of this light source pattern $\mathbf{L}(\mathbf{w}_i, t)$ is:

$$\begin{aligned} \mathbf{L}(\mathbf{w}_i, t) &= \left\lfloor \mathrm{fmod}(\mathbf{R}(2^t \mathbf{w}_i), 2) \right\rfloor, \\ \mathbf{R}(\mathbf{w}_i) &= \frac{\mathrm{atan2}(\mathbf{w}_i \cdot [0,1,0], \mathbf{w}_i \cdot [1,0,0])}{\pi} + 1, \end{aligned} \tag{8}$$

where "fmod" is the modulo operator. t should be integer and the "structured environment map" pattern is discrete.

Specular coverage requirement. Given metallic (*i.e.*, without diffuse part) or even mirror-like objects, it is difficult to use single distant point light for surface normal estimation. Because the specular surface only responds to lighting from a small lobe. For the specular dielectric material (*i.e.*, with a diffuse part), the surface normal can be estimated from the diffuse reflection. However, the material properties are still only recoverable in the small specular highlight part. We show rendered images of a shiny ball under all three lighting setups in the top row of Fig. 2 (a). The bottom row shows that the solvable regions for "point light" is a small circle around the lighting direction. In contrast, the "moving ring light" and "structured environment map" have larger solvable regions with a strip around the ring or structure edges, which is more efficient for specular coverage. The larger specular coverage results in faster data-capturing speed for challenging specular surface materials.

Diffuse sensitivity requirement. Diffuse reflection has the opposite property from specular reflection: it spreads the input light to all the upper hemisphere directions instead of only a small highlight region. In order to estimate surface normal from the reflection, the light source pattern needs to contain high and diverse low-frequency parts to make it distinguishable from different surface normal. As shown in Fig. 2 (b), we render the images of a rough diffusive ball under all three lighting setups in the top row, and plot the reflection intensity of two pixels in the bottom row. The diffuse reflection of two pixels is shifted in time for "point light" and "moving ring light" patterns. Since event cameras have an extremely high temporal resolution, the time shift can be well-captured by the camera and converted to surface normal accurately. In contrast, the diffuse reflection intensity profiles for the two pixels are similar for "structured environment map", making it difficult to distinguish the surface normal of two points.

Event efficiency requirement. Since the event camera measures the brightness changes, the bandwidth consumption is mainly determined by the brightness changing magnitude in the scene. We show the synthetic event signals and the accumulated number of events from rendered a mirror-like ball in Fig. 2 (c). Both "point light" and "moving ring light" have a similar number of events triggered by different surface normal (observed by two pixels). The "moving ring light" pattern consumes fewer events for specular coverage than the "point light", indicating a better bandwidth efficiency. In contrast, the "structured environment map"

leads to a variable number of events across different surface normal. In the worst case, where there is a mirror plane facing at the worst surface normal with a maximum number of events generated, the whole scene is flickering, and the read-out will be congested.

Based on the comparisons and analysis above, the designed "moving ring light" achieves relatively larger coverage of various materials while preserving higher efficiency of event triggering.

3.3. EventPSR Algorithm

As shown in Fig. 1, we propose a two-stage solving pipeline, consisting of a grid-matching stage and a gradient-tuning stage. The proposed algorithm provides a global minimum result with high value resolution.

Stage-1: Grid-matching. To estimate the surface normal and reflectance parameters from the captured event streams, we first render a database containing various intensity-time curves using a physically-based rendering engine. We assume that the rendered temporal resolution is high enough so that the intensity curves can be represented by linear interpolation of a series of frames. There are 4 parameters to be estimated for each pixel, including 2 parameters for surface normal, 1 parameter for roughness, and 1 parameter for metallic. The database is on a 5-dimension grid (1 additional dimension for time), which is defined as $\mathcal{B} = \{\mathbf{N}, R, M, t\}$.

Given a series of event signals $\mathcal{E}_{x,y}$ at a pixel (x, y), the goal in this stage is to find the most matching curve in the database. Therefore, we define the loss function between the observed event stream and curve candidates as follows:

$$\begin{aligned}\mathcal{L}_1(x, y, \mathbf{N}, R, M) = \sum_k |\log(\mathcal{B}(\mathbf{N}, R, M, t_k) + \epsilon) \\ - \log(\mathcal{B}(\mathbf{N}, R, M, t_{k-1} + \eta) + \epsilon) - p_k C|^2.\end{aligned} \tag{9}$$

By designing the loss function on each pair of events, we prevent the event integral process, which has an integration drifting issue (More details about the drifting issue are shown in the supplementary material). Furthermore, our loss function averages errors of all events in the same pixel, implicitly suppressing event triggering noise.

We enumerate 4D surface normal $\mathbf{N}$ and material properties R, M on a discrete grid and select the one with minimum loss. The results $\tilde{\mathbf{N}}, \tilde{R}, \tilde{M}$ at this stage are as follows:

$$\tilde{\mathbf{N}}_{x,y}, \tilde{R}_{x,y}, \tilde{M}_{x,y} = \underset{\mathbf{N},R,M}{\operatorname{argmin}} \mathcal{L}_1(x, y, \mathbf{N}, R, M). \tag{10}$$

Stage-2: Gradient-tuning. The grid-matching stage provides us with a coarse result close to the global minimum, but this result can be further improved for the following two reasons: 1) The resolution of grid-matching stage depends on the searching grid size. Due to the computational cost, this resolution is not high enough. 2) In real setup, there are some regions of the environment map that do not have a display. If the material is shiny, metallic, and the surface normal is reflecting to the dark region, there is no brightness change, which leads to ambiguity in these regions. More details are shown in the supplementary material.

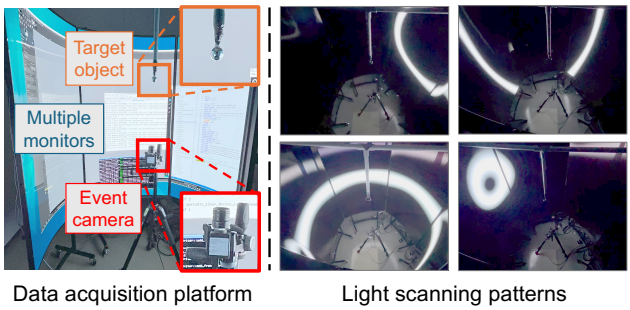

Figure 3. Data acquisition platform for capturing real data. There are 5 commodity monitors surrounding the target object to illuminate the designed pattern from any incidence angle. We show the capture device setup on the left side. The target object is hung in the middle and an event camera is pointing at the target from the bottom. Three projected patterns are shown on the right.

To solve the first problem, we calculate the gradient of surface normal $\mathbf{N}$ and material properties R, M, and interpolate the database on these parameters. In this way, we obtain an approximated rendered intensity and its gradient in a continuous parameter space. Therefore, we minimize the $\mathcal{L}_1(x, y, \mathbf{N}, R, M)$ using gradient descent to get the optimized result.

To solve the second problem, we introduce a small inter-pixel smoothness regularization term:

$$\begin{aligned}\mathcal{L}_2 = \beta \sum_{x,y} \big(&|\mathbf{N}_{x,y} - \mathbf{N}_{x-1,y}|^2 + |\mathbf{N}_{x,y} - \mathbf{N}_{x,y-1}|^2 + \\ &|R_{x,y} - R_{x-1,y}|^2 + |R_{x,y} - R_{x,y-1}|^2 + \\ &|M_{x,y} - M_{x-1,y}|^2 + |M_{x,y} - M_{x,y-1}|^2\big).\end{aligned} \tag{11}$$

In this way, for shiny and metallic surface, the parameters in ambiguous regions are filled with interpolated values, while the diffusive or dielectric surfaces are not affected. The optimized parameters of this gradient-tuning are the final results of EventPSR.

4. Experiments

4.1. Implementation Details

Data acquisition platform. To illuminate the target object using our designed light scanning pattern, we build a data acquisition platform with multiple commodity monitors, as illustrated in Fig. 3. The target object is hanging at the center of the device and surrounded by several monitors. There is a Prophesee EVK4 event camera beneath the target object, pointing above at the object and capturing the event stream as the light scanning pattern changes.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Average
Ours Data Ratio	0.198	0.146	0.131	0.406	0.345	0.598	0.185	0.262	0.432	0.634	0.164	0.282	0.296	0.255	0.260	0.250	0.303
Ours Normal	4.082	6.941	6.970	13.807	11.649	5.105	1.163	7.556	1.199	0.994	6.135	10.015	12.511	15.934	6.594	8.876	7.471
NeILF Normal	6.259	12.356	Failed	Failed	Failed	5.775	Failed	Failed	8.449	5.115	Failed	Failed	9.740	44.471	Failed	13.795	N/A
SDM Normal	6.238	5.889	9.313	9.604	8.222	3.812	25.141	6.068	13.026	4.388	18.579	8.224	9.856	8.825	7.360	7.717	9.516
Ours Roughness	0.002	0.003	0.003	0.005	0.004	0.003	0.001	0.002	0.001	0.001	0.003	0.004	0.004	0.005	0.003	0.004	0.003
NeILF Roughness	0.005	0.005	Failed	Failed	Failed	0.018	Failed	Failed	0.242	0.005	Failed	Failed	0.005	0.009	Failed	0.237	N/A
SDM Roughness	0.027	0.007	0.011	0.009	0.007	0.020	0.008	0.010	0.026	0.020	0.016	0.020	0.018	0.009	0.018	0.010	0.015
Ours Metallic	0.012	0.013	0.011	0.021	0.031	0.014	0.004	0.013	0.004	0.006	0.011	0.010	0.025	0.026	0.012	0.015	0.014
NeILF Metallic	0.051	0.042	Failed	Failed	Failed	0.017	Failed	Failed	0.704	0.046	Failed	Failed	0.032	0.060	Failed	0.673	N/A
SDM Metallic	0.025	0.027	0.022	0.024	0.034	0.026	0.017	0.027	0.221	0.080	0.027	0.036	0.086	0.023	0.027	0.062	0.048

Figure 4. Quantitative comparison on synthetic data. We show the relative data consumption and per-object averaged error for normal, roughness, and metallic estimation. The color heatmap for each parameter (normal, roughness, and metallic) is normalized independently. The results from NeILF [31] are empty for some objects, which are marked as "Failed".

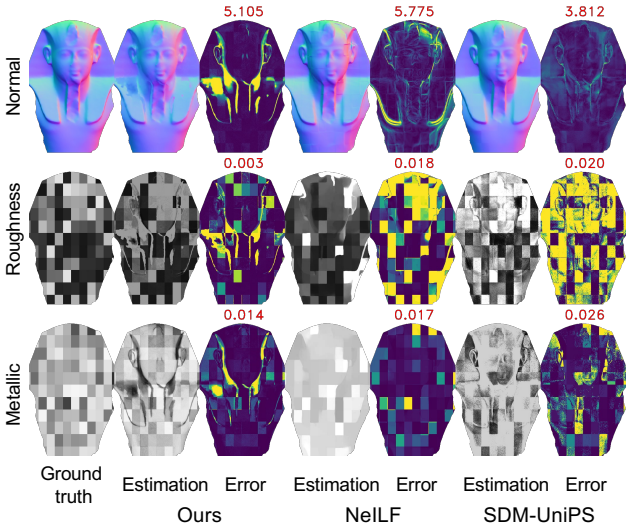

Figure 5. Quantitative comparison results on synthetic data. We show the estimated normal, roughness, metallic, and error maps of EventPSR and two comparing methods. The averaged error maps are labeled in red numbers.

The light scanning pattern received by the target object is slightly different from the designed one because of some non-ideal factors. The uncovered area at the top and bottom will cause some missing illumination, and the inter-reflection of the monitors will cause some extra light from the wrong direction. To alleviate the influences caused by non-ideal light scanning pattern projection, we capture the received illumination by a mirror ball in the scene. The database $\mathcal{B}(\mathbf{N}, R, M, t)$ is rendered using the calibrated environment maps. In addition, we flash all panels to full white at the beginning and end of the illumination pattern for temporal alignment.

Algorithm implementation. Both of the two stages are implemented using PyTorch [20]. For the grid-matching stage, the resolution of rendered database $\mathcal{B}(\mathbf{N}, R, M, t)$ are as follows: 64×64 for normal resolution, 12 for roughness resolution, 8 for metallic resolution, and 600 for temporal resolution. According to Eq. (3), the reflectance model is a linear interpolation from pure diffuse to pure mirror. To reduce GPU VRAM consumption, we replace the metallic dimension using 1 pure diffuse component and 1 pure mirror component. Different material metallic properties are constructed on the GPU on the fly. Both of the grid-matching and gradient-tuning stages are conducted on a single NVIDIA GeForce RTX 3060 graphics card.

4.2. Dataset

Synthetic data. To evaluate the accuracy of our method compared with other methods, we render 16 objects using the geometries from the sculpture dataset [27]. The material parameters including albedo, roughness, and metallic, are set randomly per patch. We use a ray-tracing-based rendering engine, with 1024 samples per pixel and 3 maximum bounces. The rendered images are of 512×512 resolution. An example is shown in the first column in Fig. 5. Each of the objects is rendered under 3 different setups: Our illumination with fixed camera position, DiLiGenT10^2 [22] light directions with fixed camera position (for the comparison with SDM-UniPS [12]), and fixed light direction with different camera positions (for the comparison with NeILF [31]). For the first setup, the event streams are simulated from high-frame-rate image sequences. The event triggering threshold noise is set to 0.01. For the other two setups, there are total 100 HDR images rendered per object. Besides, the independent Gaussian noise of 0.01 standard variance is added to each image.

Real data. We capture three types of real data using our data acquisition platform. The captured data are used for material roughness accuracy evaluation, surface normal accuracy evaluation, and complex object generalization evaluation, respectively.

- **Surface normal evaluation data**: We select several pairs of objects with almost identical geometry but extremely different materials (*i.e.*, rough and shiny), as shown in the first row in Fig. 6. The ground truth of surface normal is captured by a structured light 3D scanner.[4]
- **Material roughness evaluation data**: We create several

[4] https://www.shining3d.com/professional-solutions/desktop-3d-scanner/einscan-se-v2

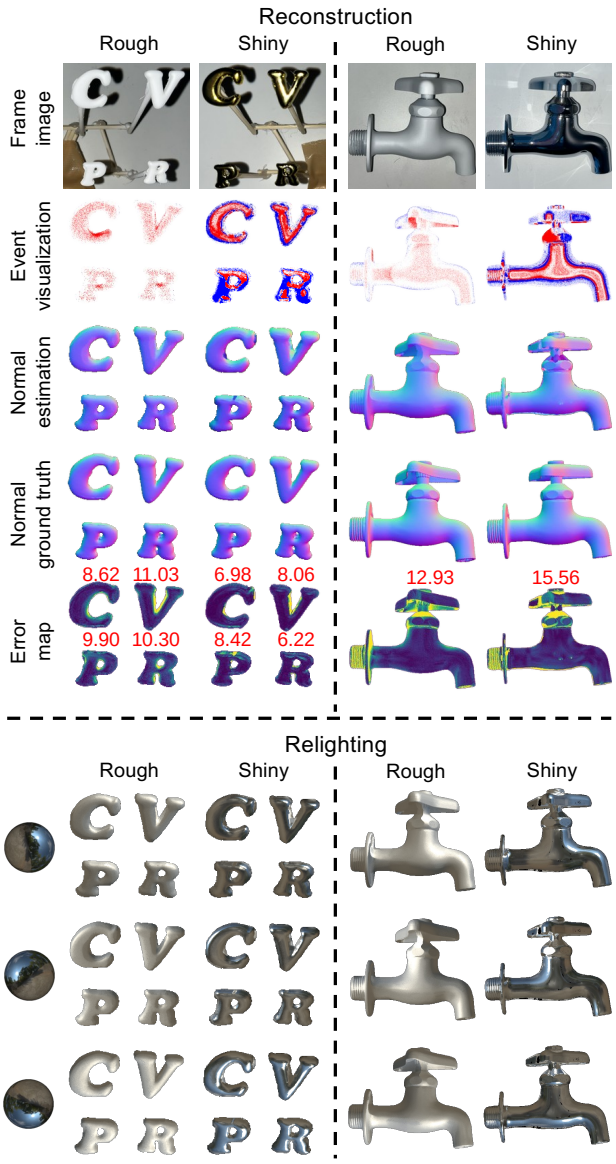

Figure 6. Top: Quantitative evaluation for surface normal estimation on real objects. The error value of each object is labeled in red number. Bottom: Relighting results based on the estimated surface parameters. The environment maps are labeled on the left. The albedo of rendered objects are set to uniform white.

balls with varying surface roughness. A polished metal ball serves as the low roughness reference. For the high roughness surface, we apply a spray coating to achieve a diffuse surface. The ground truth values of roughness are measured using a frame-based camera with HDR exposure under ideal parallel light.

- **Generalization evaluation data**: We select multiple objects whose surface are composed of multiple different materials, as shown in the first row in Fig. 8. The complexity of the shape and materials will test the generalization ability of EventPSR on real-world objects.

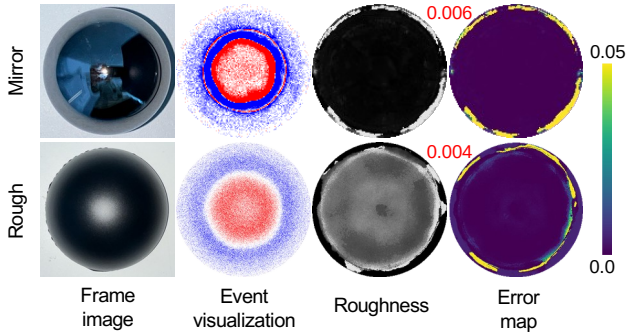

Figure 7. Quantitative demonstration for robustness of material estimation on real objects. The first row is a metal ball with mirror-like surface material. The second row is a ball sprayed with a non-glossy coating.

4.3. Evaluation on Synthetic Data

We conduct quantitative evaluations of the surface normal and reflectance parameters estimation with the synthetic dataset. The comparison includes two frame-based methods [13, 31] that represent the state-of-the-art NeRF-based [31] and PS-based [13] normal and reflectance estimation approaches. For data rate usage comparison, we estimate the date consumption in the following way: For EventPSR, we assume that the input event streams employ 16-bit Prophesee EVT 3.0 format. For the other two comparison methods, we assume that the HDR input are captured by merging 3 multi-bracketing 8-bit grayscale images. We use mean angular error to measure the accuracy of recovered normal maps. For the recovered roughness and metallic, we first standardize the data to avoid representing differences between models, then we apply mean squared error (MSE) to measure the accuracy. More details are described in the supplementary material.

The data consumption and parameters error of each object are shown in Fig. 4. On average, our method consumes about 30% of the data rate, achieving comparable accuracy of normal estimation and superior accuracy of roughness and metallic estimation. We visualize the result of one example from the synthetic dataset in Fig. 5. We can see that the error of our method concentrates mainly on the high geometry change area for all three parameter maps, where the inter-reflection and cast shadow are most severe. For the other two comparing methods [13, 31], the errors of normal estimation are more evenly distributed. However, the estimated roughness and metallic from the comparing methods [13, 31] can hardly reflect the high frequency of different material patches of the target object. More examples are shown in the supplementary material.

To demonstrate the robustness of EventPSR in recovering various materials, we evaluate the performance of the three methods on a cow object with 6 different surface materials, from ideal Lambertion to pure mirror. More details

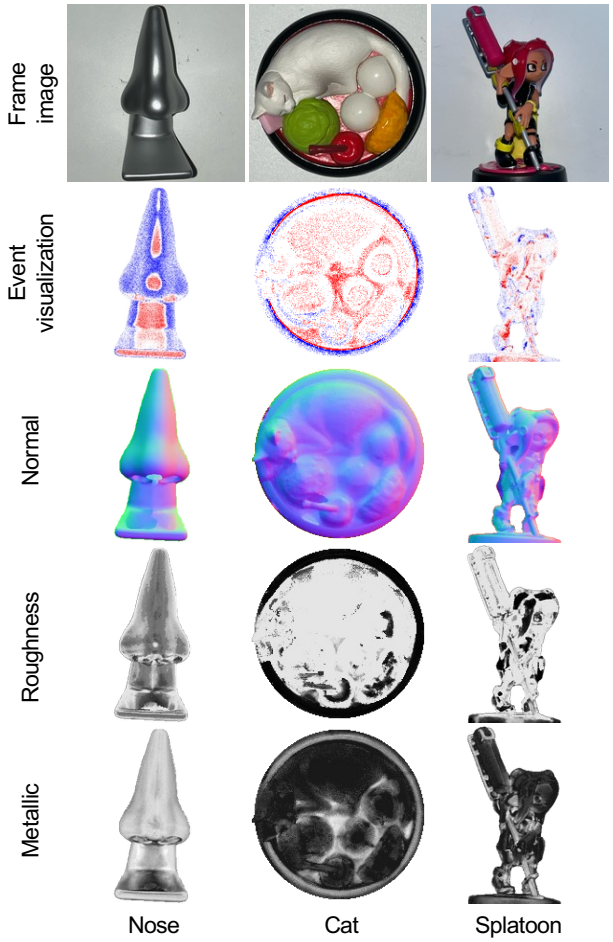

Figure 8. Qualitative demonstration of real-world objects. From left to right, the complexities of shape and material are increasing. EventPSR recovers parameter maps consistently, and preserves high-frequency details for simple and complex objects.

are shown in the supplementary material.

4.4. Evaluation on Real Data

Quantitative evaluation on surface normal estimation. To demonstrate that EventPSR can estimate the surface normal of objects with various materials, we select some objects with identical shape but extremely different surface materials. As shown in the top part of Fig. 6, the surface normal estimation results across different materials remain consistent, with only minor errors. The bottom part of Fig. 6 are the relighting results rendered with the estimated normal and reflectance parameters under different lighting conditions.[5] The results demonstrate that EventPSR effectively handles material variation and maintains accuracy across different surface types.

Quantitative evaluation on reflectance estimation. To verify the accuracy of roughness estimation using EventPSR, we conduct experiments with two semi-spheres featuring distinct roughness levels, as illustrated in Fig. 7. EventPSR achieves an average MSE of 0.0067 in roughness estimation, showing its high precision and robustness among different surface materials. The error maps highlight EventPSR's capability to differentiate subtle variations in material roughness, underscoring its robustness across diverse reflectance properties.

Qualitative evaluation on real objects. The results of surface normal and reflectance parameters estimation of three real-world complex objects are shown in Fig. 8. The nose model in the left column has spatially uniform specular material. The estimated surface normal is mostly unaffected by specular reflection. The estimated roughness and metallic are mostly uniform on the surface. In the middle column, there is a diffuse cat lying on a black and metallic (mirror-like) plate. From the estimated result, we can see the low roughness and high metallic on the outer rim. There is a splatoon model with complex materials in the right column. From the image captured by a frame-based camera, we can see that different parts are in different materials. The eyes, hair, and backpack have low roughness, while the rough skin has high roughness. The recovered roughness is consistent with the observation. Besides, the high metallic in the pipes of the ink gun is also recovered by EventPSR.

5. Conclusion

We propose EventPSR, an efficient and robust approach for simultaneous estimation of surface normal and reflectance parameters using an event camera. EventPSR is the first to use continuous radiance differential signals for joint estimation of shape and material properties, reducing data rate by 30% compared to frame-based methods.

Limitations. The capturing speed is limited by the refresh rate of the current monitors. The accuracy of real object results is affected by cast shadow and inter-reflection. Despite the robustness of albedo variation, we cannot recover diffuse albedo. These factors can be improved by employing a faster device and designing more robust algorithms, which are described in the supplementary material.

Acknowledgement

This work is supported by National Natural Science Foundation of China under Grant No. 62136001, 624B2006, 62088102, Beijing Natural Science Foundation under Grant No. L233024, JST-Mirai Program under Grant No. JPMJM123G1, and JSPS Kakenhi under Grant No. 20H05953. We thank NHK (Japan Broadcasting Corporation) Science & Technology Research Laboratories (STRL) for supporting the illumination devices for real-world experiments. PKU-affiliated authors thank openbayes.com for providing computing resource.

[5] Objects are rendered with a uniform albedo of white.

References

[1] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2008. 1, 2

[2] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik Lensch, and Jan Kautz. Two-shot spatially-varying BRDF and shape estimation. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2

[3] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. NeRD: Neural reflectance decomposition from image collections. In *Proc. of Int. Conf. Comput. Vis.*, 2021. 1

[4] Brent Burley and Walt Disney Animation Studios. Physically-based shading at disney. In *ACM SIGGRAPH*, pages 1–7. vol. 2012, 2012. 3, 1

[5] 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. 3

[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. 2

[7] Hoon-Gyu Chung, Seokjun Choi, and Seung-Hwan Baek. Differentiable point-based inverse rendering. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2024. 2

[8] Jonathan Dupuy and Wenzel Jakob. An adaptive parameterization for efficient material acquisition and rendering. *ACM Trans. Graph.*, 37(6):1–14, 2018. 2

[9] 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. 2, 3

[10] Dan B Goldman, Brian Curless, Aaron Hertzmann, and Steven M. Seitz. Shape and spatially-varying BRDFs from photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010. 1

[11] 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

[12] Satoshi Ikehata. Universal photometric stereo network using global lighting contexts. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 6

[13] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 7, 2

[14] Yakun Ju, Kin-Man Lam, Wuyuan Xie, Huiyu Zhou, Junyu Dong, and Boxin Shi. Deep learning methods for calibrated photometric stereo and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024. 2, 3

[15] Kaizhang Kang, Cihui Xie, Chengan He, Mingqi Yi, Minyi Gu, Zimin Chen, Kun Zhou, and Hongzhi Wu. Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM Trans. Graph.*, 2019. 1, 2

[16] M. Katzin. The scattering of electromagnetic waves from rough surfaces. *Proceedings of the IEEE*, 52(11):1389–1390, 1964. 1

[17] Behnaz Kavoosighafi, Saghi Hajisharif, Ehsan Miandji, Gabriel Baravdish, Wen Cao, and Jonas Unger. Deep SVBRDF acquisition and modelling: A survey. In *Computer Graphics Forum*, 2024. 2

[18] Zongrui Li, Qian Zheng, Boxin Shi, Gang Pan, and Xudong Jiang. DANI-Net: Uncalibrated photometric stereo by differentiable shadow handling, anisotropic reflectance modeling, and neural inverse rendering. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2023. 3

[19] 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. 2

[20] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Proc. of Adv. Neural Inform. Process. Syst.*, 2019. 6

[21] Konstantinos Rematas, Tobias Ritschel, Mario Fritz, Efstratios Gavves, and Tinne Tuytelaars. Deep reflectance maps. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2016. 2

[22] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi. Diligent102: A photometric stereo benchmark dataset with controlled shape and material variation. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 6

[23] Fabiano Romeiro, Yuriy Vasilyev, and Todd Zickler. Passive reflectometry. In *Proc. of Eur. Conf. Comput. Vis.*, 2008. 2

[24] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo networks for determining surface normal and reflectances. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2020. 2

[25] 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

[26] K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces∗. *J. Opt. Soc. Am.*, 57(9): 1105–1114, 1967. 1

[27] Olivia Wiles and Andrew Zisserman. SilNet: Single-and multi-view reconstruction by learning from silhouettes. In *Proc. of Brit. Mach. Vis. Conf.*, 2017. 6

[28] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. NeILF: Neural incident light field for physically-based material estimation. In *Proc. of Eur. Conf. Comput. Vis.*, 2022. 1

[29] 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

[30] Ye Yu and William AP Smith. InverseRenderNet: Learning single image inverse rendering. In *Proc. of IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 2

[31] Jingyang Zhang, Yao Yao, Shiwei Li, Jingbo Liu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. NeILF++: Inter-reflectable light fields for geometry and material estimation. In *Proc. of Int. Conf. Comput. Vis.*, 2023. 1, 6, 7, 2

[32] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. NeRFactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph.*, 2021. 1

[33] 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. 2, 3

EventPSR: Surface Normal and Reflectance Estimation from Photometric Stereo Using an Event Camera Supplementary Material

Bohan Yu[1,2†] Jin Han[3,4] Boxin Shi[1,2*] Imari Sato[3,4]

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

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

[3] The University of Tokyo [4] National Institute of Informatics

ybh1998@pku.edu.cn, jinhan@nii.ac.jp, shiboxin@pku.edu.cn, imarik@nii.ac.jp

6. Details for the parameterized BRDF model

The detailed formula of D and G in Eq.(3) are as follows:

$$
\begin{aligned}
D &= \frac{1}{\pi R\left(\mathbf{N}\cdot\mathbf{H}\right)^4}\cdot\exp\left(\frac{\left(\mathbf{N}\cdot\mathbf{H}\right)^2-1}{R(\mathbf{N}\cdot\mathbf{H})^2}\right),\\
G &= \min\left(1,\frac{2\left(\mathbf{H}\cdot\mathbf{N}\right)\left(\mathbf{w_i}\cdot\mathbf{N}\right)}{\mathbf{w_i}\cdot\mathbf{H}},\right.\\
&\qquad\left.\frac{2\left(\mathbf{H}\cdot\mathbf{N}\right)\left(\mathbf{w_r}\cdot\mathbf{N}\right)}{\mathbf{w_i}\cdot\mathbf{H}}\right),\\
\mathbf{H} &= \frac{\mathbf{w}_\mathrm{i}+\mathbf{w}_\mathrm{r}}{|\mathbf{w}_\mathrm{i}+\mathbf{w}_\mathrm{r}|_2},
\end{aligned}
\tag{12}
$$

where $\mathbf{N}$ is the surface normal, and R is the surface metallic. Both of them have the same definition as Eq.(4), with the pixel coordinates (x, y) omitted. $\mathbf{w}_i$ and $\mathbf{w}_r$ are the incident and reflection direction, respectively. The specular distribution term D describes the microfacet normal direction distribution. Here we use Beckmann distribution [16] as D. The geometry term describes the cast shadow of the microfacet model. We use the same formula in the Disney BRDF model [4] and Torrance-Sparrow model [26] as G.

7. Event Integral Drifting issue

We show a case of event integral curve in Fig. 9. Due to non-ideal triggering of real event camera, the integral of events (orange line) can be constantly rising or falling, causing mismatch compared with rendered database (green line). However, as shown in Eq. (9), we directly compare each individual pair of events with the rendered database without the event integral, to avoid the event integral drifting issue.

*Corresponding author: Boxin Shi

†This work was conducted while the first author, Bohan Yu, was doing internship at National Institute of Informatics.

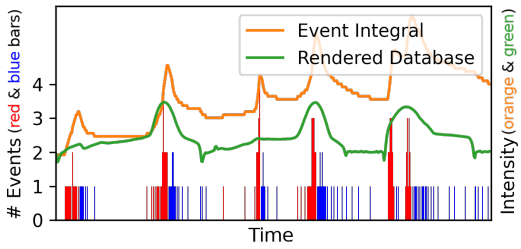

Figure 9. Demonstration of the event integral drifting issue. The integral of the event signal (orange line) is constantly rising in this example, while the actual intensity (green line) is returning to the initial value.

8. Device Coverage

The monitors in our data acquisition platform surround the objects horizontally, leaving two uncovered areas at the top and bottom. For a mirror ball, the coverage map, shown in Fig. 10, reveals these uncovered regions at the center and border of the object. Within the coverage region, surface normal and material properties are accurately estimated, while in the uncovered areas, they are interpolated during the second gradient-tuning stage. We achieved a coverage area of 69.25%, using five monitors in our data acquisition platform. This coverage can be further improved with a more compact monitor arrangement.

9. Roughness and Metallic Standardization

The definitions of roughness and metallic vary across different models used by comparison methods. To ensure consistency, we normalize the roughness and metallic outputs for all results in the logarithmic domain. Let the original output of a method on the entire synthetic dataset be represented as a vector $\mathbf{X}$ (where $\mathbf{X}$ can denote either roughness or metallic), and let the ground truth values be represented

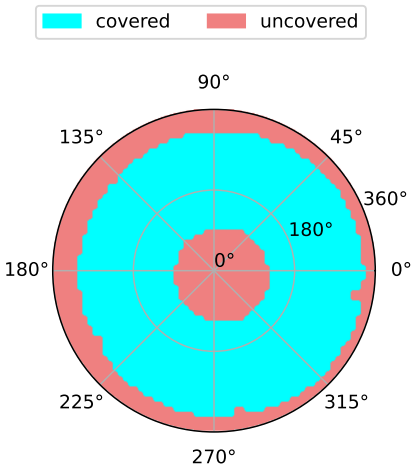

Figure 10. The coverage map on data acquisition platform. The cyan on the sphere represents the area that can be covered by the light scanning pattern. The red area is the uncovered area.

as $\hat{\mathbf{X}}$. The normalized method output $\mathbf{X}^*$ is defined as:

$$\begin{aligned} S &= \frac{\text{std}(\log(\hat{\mathbf{X}}) - \text{mean}(\log(\hat{\mathbf{X}})))}{\text{std}(\log(\mathbf{X}) - \text{mean}(\log(\mathbf{X})))}, \\ T &= \text{mean}(\log(\hat{\mathbf{X}})) - \text{mean}(\log(\mathbf{X})) * S, \\ \mathbf{X}^* &= \exp(\log(\mathbf{X}) * S + T). \end{aligned} \tag{13}$$

This process is similar to applying batch normalization to the logarithmic roughness and metallic values across the entire synthetic dataset. Consequently, the mean and standard deviation of the log-transformed roughness and metallic are aligned with those of the ground truth. By normalization in this manner, discrepancies in scale and power across different material property definitions are eliminated, ensuring fair and consistent quantitative comparisons.

10. Robustness on Various Materials

To demonstrate the robustness of EventPSR in recovering various materials, we evaluate the performance of the three methods on a cow object with 6 different surface materials, from ideal Lambertion to pure mirror, as shown in Fig. 11. Since the PS-based method [13] is designed from the Lambertian reflection model, the error of normal estimation is becoming larger as the surface material is becoming shiny. The shiny materials are also challenging for NeRF-based method [31], which even failed to reconstruct the object after the "rough metallic" material in Fig. 11. Because highly specular and metallic materials are too challenging for the geometry initialization stage in NeILF++ [31].

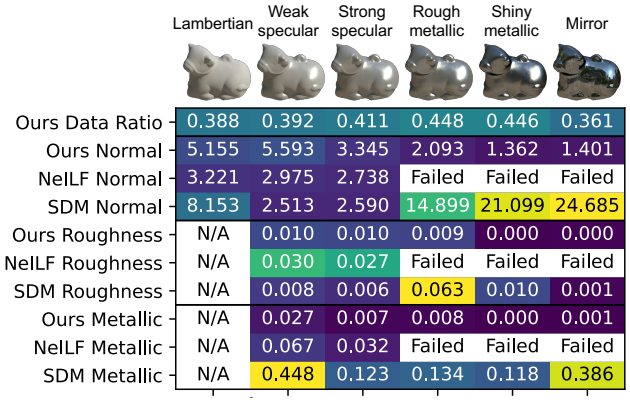

Figure 11. Case study on robustness. The materials of objects in the first row vary from ideal Lambertian to pure mirror. EventPSR performs consistently for different materials, while the comparing methods fail or have an increasing error as the specular part gets more and more dominant.

11. Additional Results on Synthetic Data

To evaluate the performance of EventPSR, we conduct extensive experiments on synthetic dataset, as shown in Fig. 13. The synthetic dataset includes objects with varying levels of roughness, metallicity, and geometric complexity to test the robustness and versatility of our method.

12. Ablation study

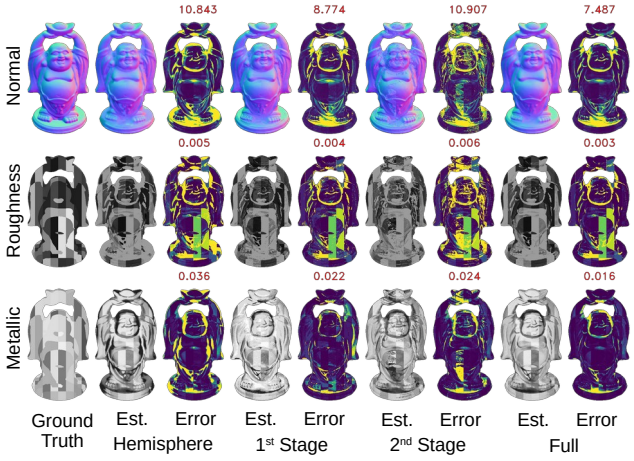

Figure 12. Ablation study shows that, high illumination coverage ratio and the full two-stage pipiline solution are both essetial to achieve high eccuracy.

We conduct ablation experiments on synthetic data with reduced light coverage ratio using a hemisphere (50% coverage ratio) and evaluate our two-stage estimation pipeline. As shown in Fig. 12, hemisphere illumination yields higher errors in bottom and boundary areas, likely due to specular

reflection in uncovered regions. The first-stage-only results exhibit high overall error, likely due to low parameter resolution, while the second-stage-only results show incorrect patches, possibly stuck in local minima. The full pipeline's superior accuracy underscores the importance of both high light coverage and our proposed two-stage solution.

13. Discussion and Potential Improvements.

The accuracy of our current solution is bottlenecked by non-Lambertian effects, and the speed is limited by the illumination device. We propose several potential extensions: First, to enhance robustness against cast shadows and inter-reflections, explicit modeling methods such as DANI-Net [18] or deep learning-based PS [5, 14] can be employed. Besides, applying a multi-view system can provide more comprehensive coverage of specular regions under limited illumination. For the diffuse albedo recovery, we can integrate a frame-based camera to create a hybrid camera system, or leverage alternative intensity estimation techniques like [11]. For the illumination system, we utilize multiple OLED panels for broad light coverage, uniform intensity, and mostly white color. A different illumination setup may have non-uniform intensity & color, or position misalignment. These artifacts can be reduced by the end-to-end calibration process. For real-time capturing, our high event efficiency design depicted in Fig. 2 (c) suggests that the event camera can accommodate illumination with increased speed. To achieve real-time processing, we can replace the grid search with more efficient algorithms, such as heuristic search or locality-sensitive hashing.

14. Light Pattern and Relighting Result Video

We provide a video including the working data acquisition platform, and the light scanning pattern illuminated by the monitors. We also render the object relighting results with white uniform albedo in the supplementary video.

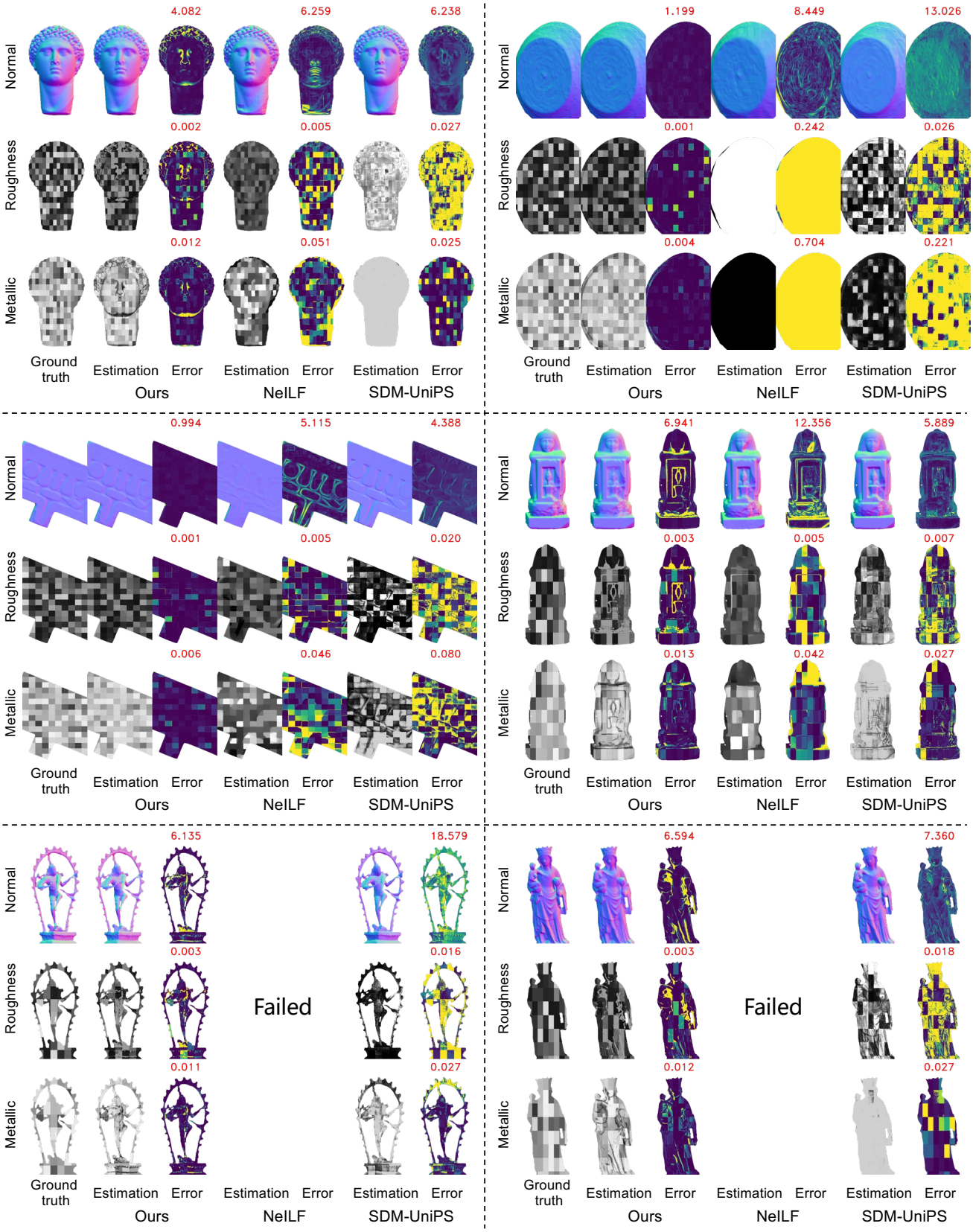

Figure 13. Additional evaluation results on the synthetic dataset.