

# PS-EIP: Robust Photometric Stereo Based on Event Interval Profile

## Supplementary Material

Kazuma Kitazawa<sup>1</sup> Takahito Aoto<sup>2</sup> Satoshi Ikehata<sup>3</sup> Tsuyoshi Takatani<sup>1</sup>

<sup>1</sup>University of Tsukuba <sup>2</sup>Optech Innovation, LLC. <sup>3</sup>National Institute of Informatics (NII)

In this supplementary material, we provide implementation details and important results that couldn't be included in the main manuscript due to space constraints.

### 9. Details of Data Acquisition

In Sec. 5, we discuss the implementation and calibration of the prototype system. In this section, we provide additional details about the implementation process.

#### 9.1. Bias Setting for Event Camera

The current event cameras generally offer a range of adjustable settings that enable tuning the sensor performance to meet various application requirements and conditions, such as higher speed, lower background activity, and a higher contrast sensitivity threshold. These customizable sensor settings are referred to as *bias setting*.

The event camera used in this paper (Prophesee EVK Gen4) has five biases: positive and negative thresholds, low-pass and high-pass filters, and dead time. The biases for the positive and negative thresholds literally determine the thresholds  $h_p$  and  $h_n$ , respectively. In the proposed method, it is preferable for these biases to be as small as possible. However, decreasing the bias for the threshold tends to increase the deviation of the statistical threshold. Refer to Fig. 6 in [34]. The bias for low-pass filter determines the maximum rate of change in intensity that can be detected. In the prototype system, we adjust this bias to ensure that no events derived from the PWM, which is used to control LEDs, are detected. The bias for high-pass filter determines the minimum rate of change in intensity. In the prototype system, we set this bias as small as possible to cover all frequencies. The bias for dead time determines the duration during which a pixel is unable to detect a new event after each event. In the proposed method, we prefer this bias to be small; however, a minimal dead time bias may result in an increased number of events exceeding the event recording rate. In the proposed problem, the spatio-temporal density of events is relatively high, compared to other applications, such as object tracking. Thus, we adjust this bias as small as possible to ensure that the profile reconstruction

does not collapse. Finally, we show the biases used in the experiments in Tab. 1, along with the default biases.

#### 9.2. Calibration Process

In the calibration process, intrinsic parameters of the event camera is also estimated to remove lens distortion, similar to conventional cameras. After calibrating the event thresholds for all pixels, the intrinsic parameters are estimated from 10 poses of an ArUco board under all the LEDs modulated by a triangular wave. Similar to the light trajectory calibration, event accumulation images are reconstructed to facilitate the camera calibration process.

As mentioned in the main manuscript, we apply Santo's method [50] to estimate the positions of LEDs. Here, we provide an example of the pin shadow detection in Fig. 10. It is important to note that the offset light should be zero to detect any pattern based on albedos in an event accumulation image.

#### 9.3. Effect of the Offset Light

In the main manuscript, we demonstrate how the offset light can reduce the albedo dependency in the event-generating mechanism of the prototype system. Here, we present pixel values in the event accumulation images for different color chips, as shown in Fig. 11. The pixel value indicates the number of recorded events. A small offset results in the number of events being strongly influenced by the albedos of the color chips, as depicted in the blue plot. Conversely, with a sufficiently large offset, the number of events be-

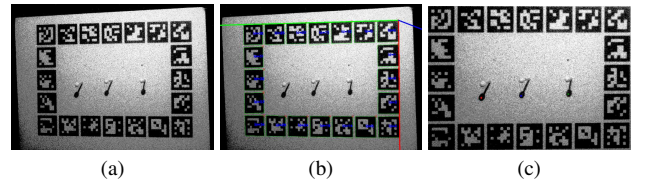


Figure 10. An example of the pattern detection in the calibration. (a) Event accumulation image. (b) ArUco detection. (c) Pins' shadows detection.

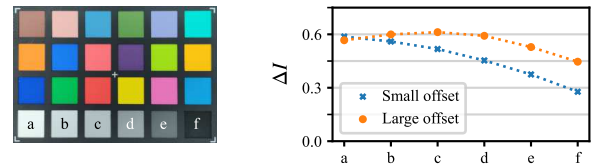


Figure 11. Effect of the offset light. The number of events with a larger offset is relatively equal for different color chips.

Table 1. Bias setting.

Bias	Positive threshold	Negative threshold	Low-pass filter	High-pass filter	Dead time
Default	112	52	23	48	45
Ours	99	57	20	0	50

comes relatively equal for all the color chips, as shown in the orange plot, albeit with a decrease in the number of events for darker chips.

In our experiments, the offset light intensity was empirically set to approximately 24 % of the main light. While there is no theoretical optimal value due to device-specific properties, a rigorous calibration could be performed using a color chart, ensuring stable event generation irrespective of surface albedo.

#### 9.4. Linear Interpolation of Reconstructed EIP

Events occur when changes in intensity exceed a threshold, meaning the samples used for EIP reconstruction are not evenly spaced along the time axis. This results in a temporal distribution bias during fitting with the theoretical EIP, making the outcome overly dependent on specific samples rather than the overall shape of curve. Additionally, when capturing events over multiple lighting cycles, the sampling timing is not always consistent, leading to complications when integrating results across cycles. To address these issues, we adopted a simple interpolation approach.

Specifically, we reconstructed a continuous EIP from sparse samples using the interpolation, then resampled it at predefined, evenly spaced time intervals to compare these points with the theoretical EIP. In our setup, the intervals between two consecutive events are generally small enough that the segment of the profile between them can be reasonably approximated as a linear function. This assumption simplifies the interpolation process, enabling to model each segment using a linear function. From this continuous representation, we uniformly sampled EIP values.

#### 9.5. Examples of Reconstructed EIP

Here, we show some examples of the actual reconstructed EIPs obtained by our prototype system with *Diffuse* to visually demonstrate the actual shapes of the profiles. In Fig. 12a, we show both the reconstructed and theoretical EIPs at four surface points with varying azimuth angles, while keeping their zenith angles constant. The EIPs have similar overall shapes, but their phases are shifted according to the azimuth angles. In contrast, Fig. 12b shows the reconstructed and theoretical EIPs at four different points with varying zenith angles, while keeping azimuth angles constant. In this case, the amplitudes of the profiles vary depending on the zenith angles.

#### 9.6. Details of the Optimization Process

The proposed method employs a multi-stage approach. Initially, optimization is performed without considering outliers such as shadows and specular reflections. Then, the presence of outliers is assessed, and a mask is generated accordingly for re-optimization. The process of the surface normal recovery is illustrated in Fig. 13. Initially, the mask

for collapsed events is applied to all surface points. If the minimized cost exceeds a threshold,  $c$ , the profile on that point may contain non-Lambertian effects. For instance, maps of costs after the first minimization are shown in the first column of Fig. 14 when targeting *Glossy* and *Pole*. The recovered normal maps without the second mask process exhibit significant errors in surface points affected by the non-Lambertian effects.

In the second mask process, the mask selection is determined based on the temporal distance between the top and bottom peaks, denoted as  $|t_t - t_b|$ . For specular highlights, the temporal distance is typically very short, while cast shadows tend to have a relatively longer temporal effect on the surface point. Therefore, if the temporal distance exceeds a threshold,  $t_{tb}$ , the mask for cast shadows is applied; otherwise, the mask for specularity is used.

We generate a single mask per outlier segment, regardless of the number of outlier types (e.g., multiple cast shadows, as *2Poles* in Fig. 16). Empirically, multiple masks often reduce inlier observations, causing optimization instability. In future work, we will explore more robust methods, such as L1 loss or low-rank priors, to handle multiple outliers simultaneously.

#### 9.7. Time Consumption

Though not our top priority, computational speed remains important. Our current implementation has larger time consumption compared to EventPS in terms of measurements and inferences.

The total measurement time is determined by the time for the light source rotation and the number of measurements, which can be reduced as suggested in Fig. 6e and Fig. 5b, respectively. If future event cameras demonstrate improved fidelity to the event model, it would become possible to

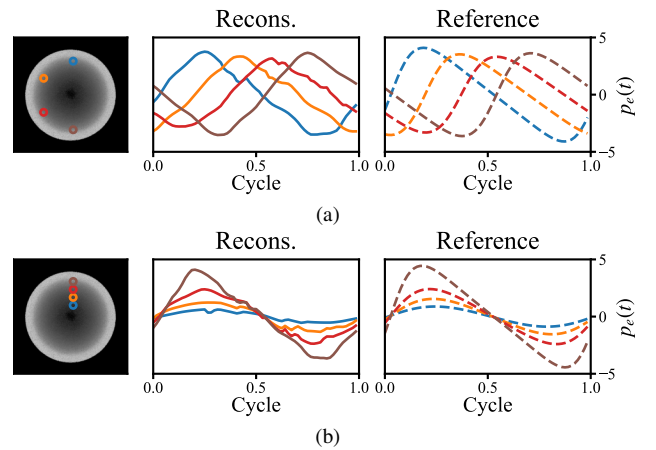


Figure 12. Shapes of the EIPs at various surface points. From left to right: surface points on *Diffuse*’s  $\Delta I$ , reconstructed EIPs for each point, and theoretical EIPs (reference). (a) Different azimuth angles under the identical zenith angle. (b) Different zenith angles under the identical azimuth angle.

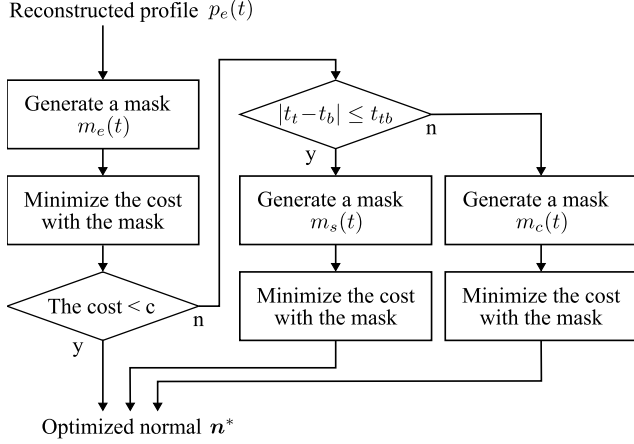


Figure 13. Process of the cost minimization.

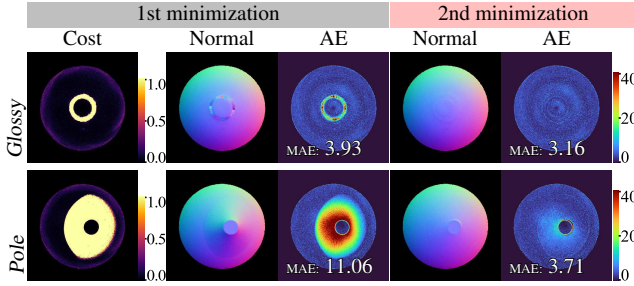


Figure 14. Effect of the second mask process. From left to right: cost map, recovered normal map, and angular error map at the first minimization, and normal map and angular error map at the second minimization.

achieve a higher rotation frequency and require fewer measurements.

The long inference time is due to the complexity of our pipeline. Algorithmically, the full pipeline (*i.e.*, EIP reconstruction, mask generation, and surface normal optimization) performs at 0.07 fps. This is slower than EventPS-FCN (2 fps) but comparable to the pixel-based EventPS-CNN (0.1 fps). Notably, our method runs on a CPU, while the others use GPU-based implementations, suggesting that GPU parallelization could significantly boost performance.

## 10. Additional Experiments

### 10.1. Mask Margin for Non-Lambertian Effects

In the second mask process, the margins of the masks related to specularities and cast shadows are considerable to enhance performance. Here, we analyze the performance of the proposed method in relation to these margins.

For *Glossy*, we analyze the MAE for varying margins,  $\Delta t_s$ , as shown in Fig. 15a. The margin for specularities is changed in a range of 8 to 30 % of the entire cycle. A small margin may allow specular lobes to still influence the profile, while a larger margin reduces the regions used for the

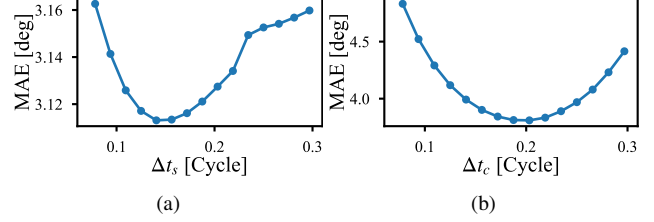


Figure 15. Analysis on the margins of masks. (a) Mean angular error with respect to different  $\Delta t_s$ . (b) Mean angular error with respect to different  $\Delta t_c$ .

cost minimization. In our prototype system, a margin of 14 % is identified as optimal for  $\Delta t_s$ . Similarly, we analyze the margin for cast shadows,  $\Delta t_c$ , for *Pole*, as shown in Fig. 15b. In this case, a 20 % margin is found to be optimal for  $\Delta t_c$ . However, since these findings are only based on *Glossy* and *Pole*, it is crucial to conduct further analysis of the margins using various objects with different types of specular characteristics and more complex shapes to ensure robustness. This aspect will be addressed in future work.

### 10.2. EventPS with Our Mask Mechanism

We evaluated the impact only on the removal of outlier regions by applying EventPS-FCN to event data (*Diffuse*, *Glossy*, *Pole*) masked by our method. Fig. 16 shows improvements over Fig. 7, particularly in shadowed regions near edges and areas with specularities and cast shadows. However, the mean angular error (MAE) increased: *Diffuse*:  $7.96^\circ \rightarrow 12.57^\circ$ , *Glossy*:  $10.12^\circ \rightarrow 12.41^\circ$ , *Pole*:  $12.02^\circ \rightarrow 14.78^\circ$ . This suggests that while filtering enhances robustness in non-Lambertian regions, it may reduce informative inliers, highlighting the need to refine the mask mechanism. The results also affirm PS-EIP’s superiority

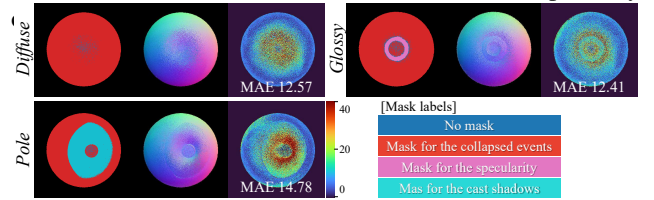


Figure 16. Quantitative evaluation of EventPS-FCN [67] with events filtered by our masks. From left to right: applied mask labels, recovered normal map, and its angular error map.

### 10.3. Additional Quantitative Comparisons

We additionally targeted 10 more 3D printed objects with different shapes and reflectances. The experimental results are shown in Fig. 17. *Glossy2C* is painted with two different colors and then given a glossy top coat. The MAE remains low, comparable to the case without the top coat, *Colors*. *DiffuseYR* and *DiffuseBG* are 3-color spheres with different colors. Recovering normals for blue and green surfaces is more difficult compared to yellow and red surfaces. This difficulty may arise due to the lower radiance of these surfaces and the spectral sensitivity characteristics of the event

camera. Despite *SemiGlossy* having wider specular lobes than *Glossy*, the mask for specularity functions effectively. *RefsBunny* contains two-color surfaces, and its top half is glossy. The MAE is quite comparable to the gray, diffuse case of *Bunny*.

*2Poles* poses a challenge as certain regions contain cast shadows from two distinct parts. The current mask for cast shadows assumes shadows derived from a single part in the scene. In future work, addressing such complex shapes will be a focus. *Black* is also challenging to recover, much like conventional photometric stereo methods. As mentioned earlier, the principle does not work well when the radiance is low due to the log-amp characteristics.

#### 10.4. Additional Qualitative Comparisons

Finally, we evaluate the proposed method on 12 more practical objects are targeted for qualitative assessment. The experimental results are shown in Fig. 18. The fine surface structures of objects such as *Owl*, *CactusB*, and *CactusC* are successfully recovered. Most of the practical objects exhibit specularity on their surface. For instance, *WDogA* and *WDogB* are pottery objects with strong specularity, yet their normal maps appear well-recovered. *Bear* also has specularity with a wider lobe. It seems that the texture colors do not significantly affect the normal map, but the impact of specularity remains noticeable. This discrepancy might be due to the mask margin not aligning well with this particular case, suggesting a need for adjustment tailored to the target. Additionally, *Panda*, being a rubber object, exhibits subsurface scattering, presenting a challenging non-Lambertian effect that warrants consideration in future work.

Since the ground truth shapes of these practical objects are unknown, we conduct a visual evaluation of depth structures reconstructed from the normal maps. Therefore, in this paper, we employ a state-of-the-art method for reconstructing depth from a normal map, as proposed by Cao *et al.* [7]. The code is accessible on GitHub<sup>1</sup>. Upon inputting the recovered normal map, its spatial foreground mask image, and the camera intrinsics, a depth structure is generated as polygons (PLY file). The depth images in Fig. 18 are rendered from these polygons with a MatCap (ceramic dark) in Blender [1].

#### 10.5. Extension to non-Lambertian BRDF

Our PS-EIP assumes an event generation model based on Lambertian diffuse reflection. However, the definition of EIP itself is more general, allowing for extensions to models other than Lambertian diffuse reflection (e.g., microfacet BRDF such as Cook-Torrance). Even if a non-Lambertian model is employed, the optimization process for EIP can still be achieved using the same methodology. Although we

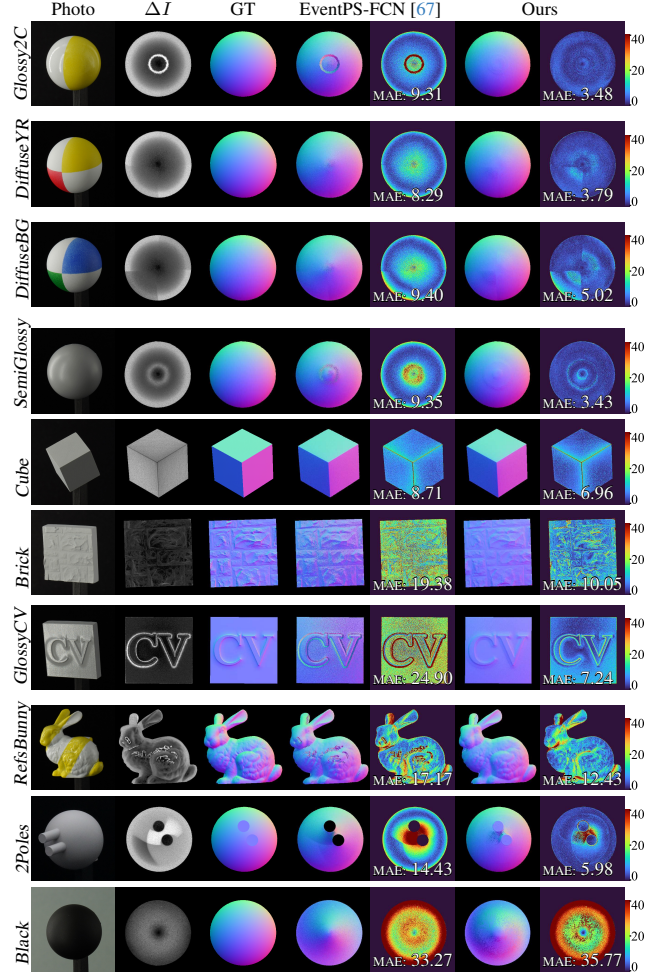


Figure 17. Additional, quantitative evaluation with 3D printed objects. From left to right: photograph, event accumulation image from the events, ground truth, normal map recovered by EventPS-FCN [67] with its angular error map, and those by ours.

have not yet explored this, we plan to examine it in future work.

<sup>1</sup>[https://github.com/xucan-42/bilateral\\_normal\\_integration](https://github.com/xucan-42/bilateral_normal_integration)



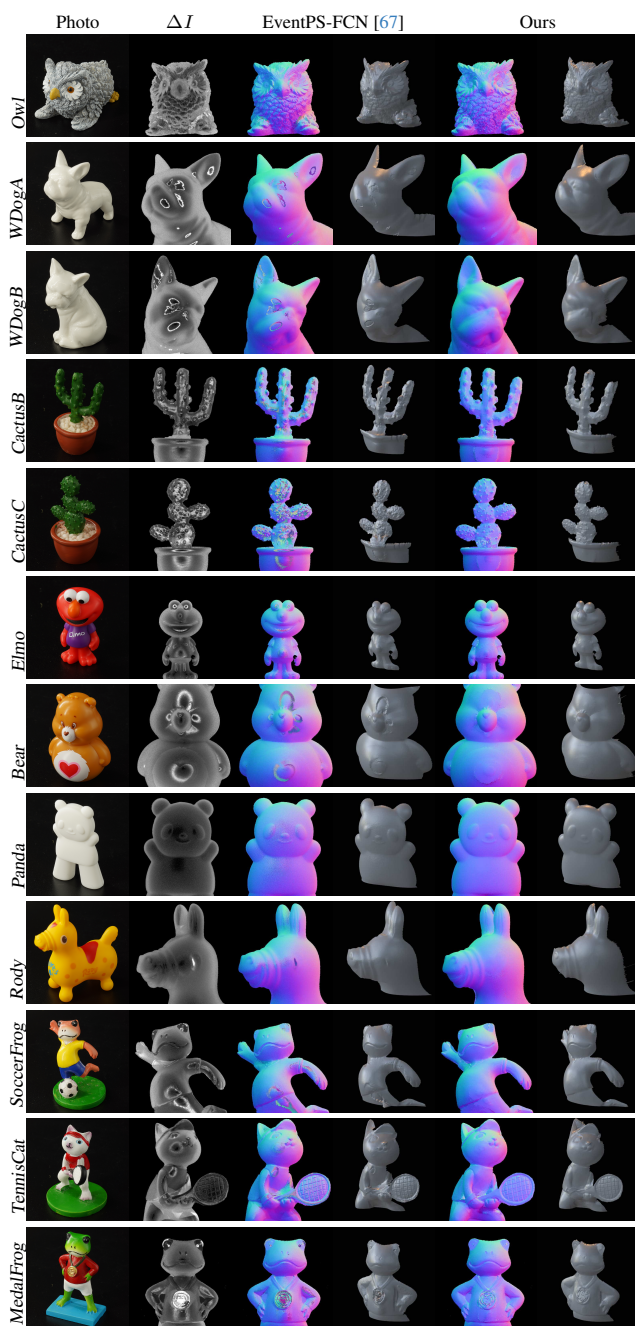


Figure 18. Additional, qualitative evaluation with the practical objects. From left to right: photograph, event accumulation image from events, normal map recovered by EventPS-FCN [67] and its depth [7], and those by ours.