Supplementary:
Event-based bispectral photometry using temporally modulated illumination

Tsuyoshi Takatani[†] Yuzuha Ito[†] Ayaka Ebisu[†] Yinqiang Zheng[‡] Takahito Aoto[†]

[†]University of Tsukuba, Japan [‡]The University of Tokyo, Japan

In this supplemental document, we additionally report statistical analysis of the event generating mechanism and experiments of distance and 3D shape reconstructions in water.

S1. Statistical analysis

As explained in the manuscript, event cameras record events triggered by changes in brightness. Ideally, if conditions of a change in brightness, such as brightnesses at the beginning and ending of the change, are determined, the number and timing of triggered events are theoretically calculated. However, the unstable factors of event cameras, such as dead time and limited bandwidth, lead to loss and delay in events. This is often modeled as a statistical threshold, which is useful to reasonably synthesize event dataset from a series of conventional video frames [S1]. We also assume a statistical threshold following a normal distribution, *i.e.*, $h \sim \mathcal{N}(\mu_h, \sigma_h^2)$, and analyze the threshold of each pixel based on the recorded events.

In the setup of the validation in the manuscript, we take event data when one of the light sources is turned off. Since a triangle wave is employed as the modulation signal, the incident signal to the event camera is linearly proportional to time in a half of the period, as below;

$$S_p(\boldsymbol{x}, t) = a(\boldsymbol{x})t, \tag{S1}$$

where $a(\boldsymbol{x})$ is a variable of a linear function at a pixel $\boldsymbol{x}$. When a positive event is triggered, the differential signal is supposed to be equal to or greater than the threshold, according to Eq. (11). Since event cameras digitally sample an analog signal based on an operation clock, the differential signal can be greater than the threshold depending on the frequency of clock. Now, for simplicity, the differential signal is assumed to be equal to the threshold when an event is triggered. From Eqs. (10) and (S1), the threshold can be

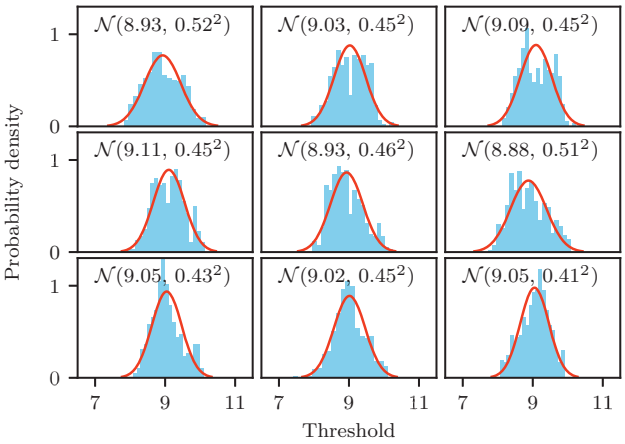


Figure S1: Histograms of the polarity threshold for 3×3 pixels when using the Prophesee EVK. The red lines are estimated normal distributions which are described as $\mathcal{N}(\mu_h, \sigma_h^2)$.

calculated, as follow;

$$h = \ln S_p(\boldsymbol{x}, t) - \ln S_p(\boldsymbol{x}, \tau(t)) \tag{S2}$$

$$= \ln \frac{t}{\tau(t)}. \tag{S3}$$

As well as the validation, the white board is located at 100 mm from the wall of the glass tank, and the frequency of modulation signal is set to 10 Hz. The threshold of each pixel is estimated from the recorded events using Eq. (S3). The average and standard deviation are analyzed for 100 periods of the modulation signal. The experimental result shows that each pixel has a different distribution of the threshold. As an example, the distributions of 3×3 pixels are shown in Fig. S1. And the total average and standard deviation of the threshold in the region illuminated by the temporally modulated illumination are 9.03 and 0.49, respectively. When replacing the event camera into the iniVation DAVIS 346, the total average and standard deviation of

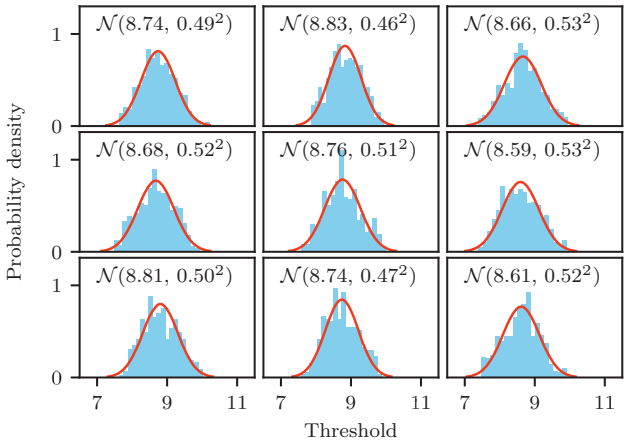


Figure S2: Histograms of the polarity threshold for 3×3 pixels when using the iniVation DAVIS 346. The red lines are estimated normal distributions which are described as $\mathcal{N}(\mu_h, \sigma_h^2)$.

the threshold are 8.71 and 0.51, respectively. Also, an example of the distributions of 3×3 pixels is shown in Fig. S2. Currently, our method assumes a constant threshold for all pixels. Introducing the statistical threshold into our method could improve the reconstruction of bispectral difference, which will be future work.

S2. Additional experiments

Distance reconstruction We additionally evaluate the distance reconstruction using the Prophesee EVK in our method and the FLIR BFS-U3-51S5M-C in the existing method. The Prophesee EVK is another event camera. Also, the conventional camera has better properties than the APS inside the iniVation DAVIS 346. The FLIR BSF-U3-51S5M-C has a lower readout noise of $3\,\mathrm{e}^-$, a higher dynamic range of $71\,\mathrm{dB}$, and a higher bit depth of $12\,\mathrm{bit}$, than the iniVation DAVIS 346 APS ($55\,\mathrm{e}^-$, $57\,\mathrm{dB}$, $8\,\mathrm{bit}$). However, the silicon photodiodes in them are so different that their evaluations cannot be directly compared. On the other hand, comparisons among different event camera and conventional cameras are helpful to predict the performance of our method when using a higher performance event camera.

In the same setup as the distance reconstruction in the manuscript except for the camera, we reconstruct a distance to the white board in water. Before experiments, spectral and geometric calibrations are performed. The white board is linearly translated every $10\,\mathrm{mm}$ in the range of 10 to $250\,\mathrm{mm}$. The frequency of modulation signals is set to $1\,\mathrm{Hz}$. The conventional camera captures HDR images. The reconstructed results by both the methods are shown in Fig. S3. When a measurable range of distance is defined as where an error is less than $15\,\mathrm{mm}$, the range of our

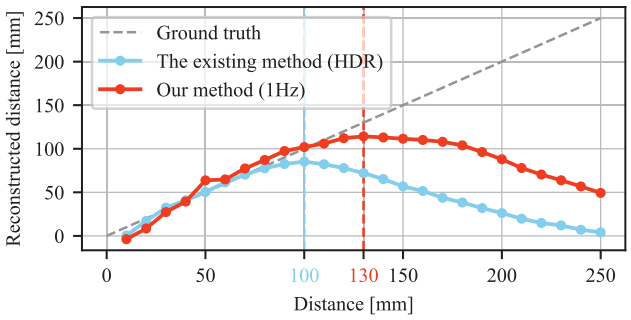


Figure S3: Experimental results of distance reconstruction in water by our method using the Prophesee EVK and the existing method using the FLIR BFS-U3-51S5M-C.

method is until $130\,\mathrm{mm}$ and that of the existing method is until $100\,\mathrm{mm}$. The average error and its standard deviation in the measurable range are 6.9 and $4.2\,\mathrm{mm}$ for our method, and 4.2 and $4.5\,\mathrm{mm}$ for the existing method, respectively. The experimental results show that the measurable range of our method is wider than that of the existing method, while the accuracy of the existing method is a bit higher than that of our method. This may be because of different measurable ranges and different photodiodes. The performance of our method does not largely vary even with different event cameras in comparison with the distance reconstruction in the manuscript. This can be explained by that the standard deviations of both the event cameras are comparable, as described in Sec. S1. The accuracy of our method finally relies on the standard deviation of threshold.

3D shape reconstruction We perform 3D shape reconstruction with a lot of target objects. In the manuscript, only three of them are shown because of limited space. The additional 15 target objects are shown on the first column in Figs. S4 and S5. Also, the depth images reconstructed by our method and its 3D views are shown on the second column. Those by the existing method using the conventional camera are shown on the third column. Additionally, we apply the existing method for brightness images individually reconstructed from events at both the wavelengths by using a general brightness reconstruction technique [21]. Those results are on the fourth column in the figures. The experimental results show that all the methods can reconstruct a global shape of an object but not a local shape, such as Pottery Cat Plate and Pottery Squid Place. This is because the resolution in depth is not enough high to reconstruct such a detailed surface. It seems difficult to reconstruct a depth image in a case where interreflections of light occur, such as Plastic Statue's chin, which is a well known problem in photometry based analysis.

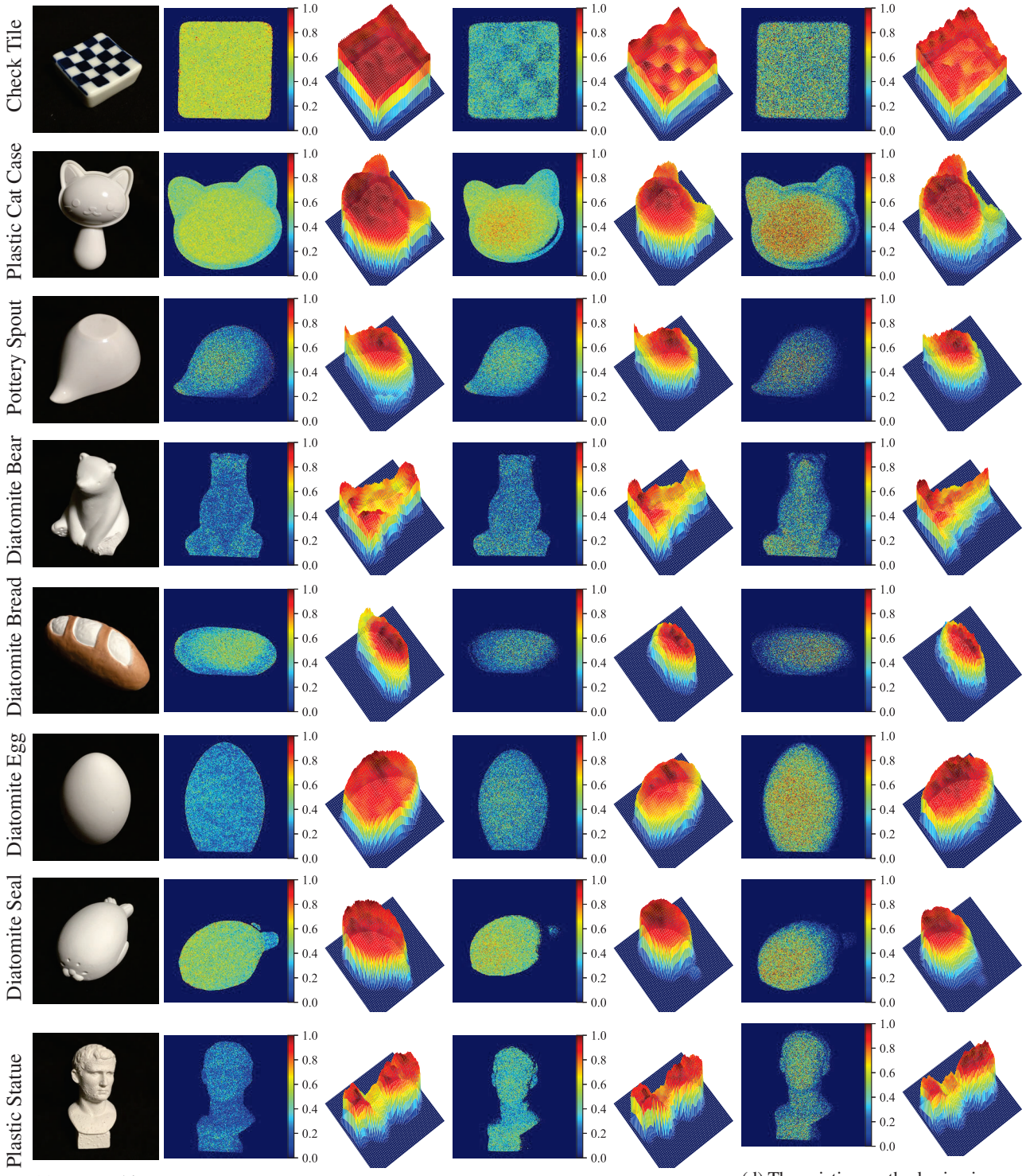
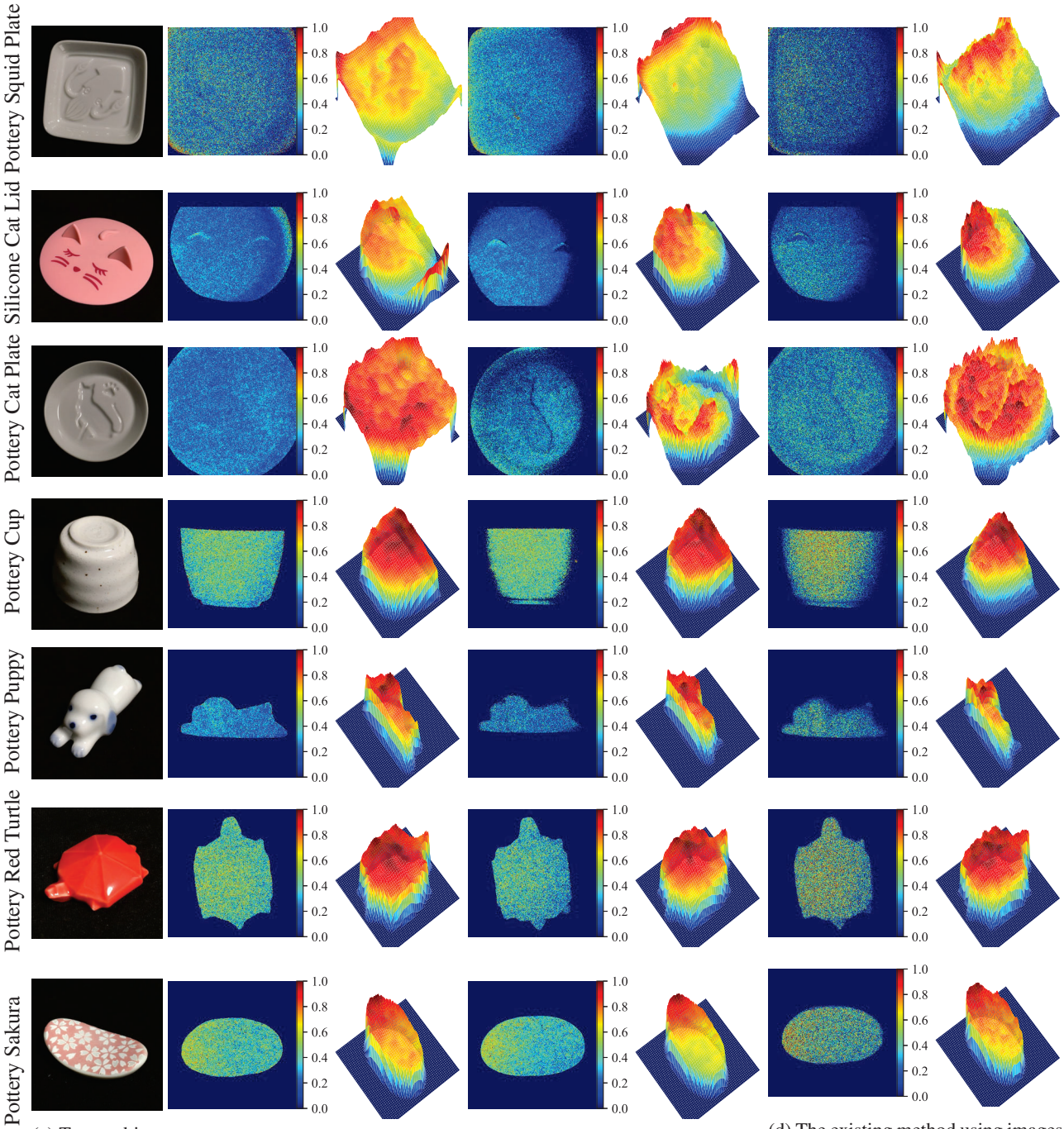


Figure S4: Experimental results of the depth image reconstruction. (1 of 2)

References

[S1] Henri Rebecq, Daniel Gehrig, and Davide Scaramuzza. ESIM: an Open Event Camera Simulator. In *Proc. of Conf. on Robot Learning*, pages 1–14, Zurich, Switherland, Oct 2018.



(a) Target object (b) Our method (c) Asano *et al.* [3] (d) The existing method using images reconstructed from events

Figure S5: Experimental results of the depth image reconstruction. (2 of 2)