Pixel-Wise Material Classification Using Deep Learning by Utilizing Transient Responses

Yohanssen Pratama[1,a)] Kazuya Kitano[1] Takahiro Kushida[2]
Yuki Fujimura[1] Takuya Funatomi[1] Yasuhiro Mukaigawa[1]

Abstract

This study introduces a novel method for material classification using the transient responses obtained via a Single-Photon Avalanche Diode (SPAD) sensor. SPAD possesses the capability to perform temporally associated single photon counting, achieving a resolution on the order of pico-seconds. The recorded data from such a time-resolved detector are referred to as 'transient,' wherein the count of photons is recorded for each arrival time and depicted in a transient histogram. As each material exhibits a unique transient response, our approach leverages the distinctive transient signatures of each material, allowing for the creation of feature vectors within the pico-second time range. This method accomplishes pixel-wise material classification by employing one-dimensional deep-learning models.

1. Introduction

Material classification is vital in distinguishing scenes in numerous computer vision and robotics tasks, such as object segmentation, recognition, and acquisition. However, RGB information alone is insufficient to differentiate materials. To address this challenge, there is work in computer vision that has proposed a solution using time-resolved imaging sensors such as the Time-of-Flight (ToF) camera to capture the correlation between a reference signal and the temporal response of the material. Then, the information is used to differentiate visually similar materials with a structural difference [1].

In this study, we used a Single-Photon Avalanche Diode (SPAD) sensor based on the time-correlated single-photon counting (TCSPC) technique to build a transient histogram [2] instead of a ToF camera for its higher time resolution of approximately 10 ps. Measuring the photon arrival time using a ToF camera in small intervals would be challenging due to its limited resolution in the nanosecond range [3]. The SPAD sensor is known for its remarkable ability to detect individual photons with exceptional timing precision. This sensor also has a wide range of applications in computer vision, including low-light image classification, 3D imaging,

[1] Nara Institute of Science and Technology
[2] Ritsumeikan University
[a)] yohanssen.pratama.yl0@is.naist.jp

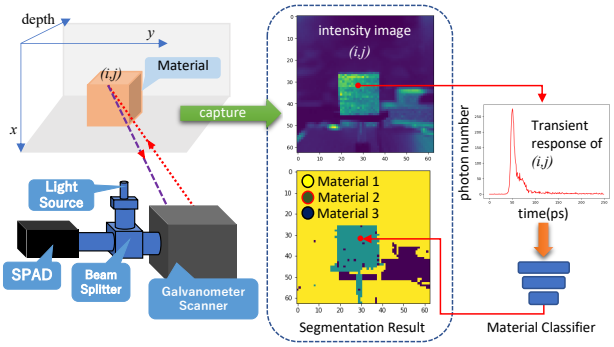

Fig. 1 Overview of the imaging system designed for this study. The galvanometer scanner performs raster scanning to collect the histogram. Then, the transient response is used to perform pixel-wise material classification.

Non-line-of-sight (NLOS) imaging, and colour classification using distinct wavelengths [4], [5], [6], [7].

Becker *et al.* [8] utilized the AMS TMF8801 direct ToF sensor to capture the transient response for material classification. Motivated by the approach that exploited the transient response of materials, we aim to expand it to enable pixel-wise material classification. Due to the complex and nonlinear nature of the transient response, traditional image analysis methods may not be suitable for this classification task. To tackle this problem, we proposed one-dimensional deep learning to perform pixel-wise material classification based on the transient response to handle this. This paper presents a novel approach to achieve pixel-wise classification of transient response with multiple materials using deep learning, specifically a Convolutional Neural Network (CNN). The CNN-based pixel-wise classification enables direct analysis of subsurface scattering from transient responses, whereas spatial CNN relies only on visual features in regular RGB images.

In our approach, illustrated in Fig. 1, the first step is to gather pixel-wise transient responses, which are interpreted as a histogram. A galvanometer scanner with a built-in two-axis mirror is employed to do raster scanning, and the resulting intensity image contains transient response information of every pixel (Figure 2 shows an example of a histogram of different materials obtained from a SPAD sensor). Then, the transient response is used as input for the one-dimensional CNN classifier. Finally, the segmentation result image would display each material class as a unique colour

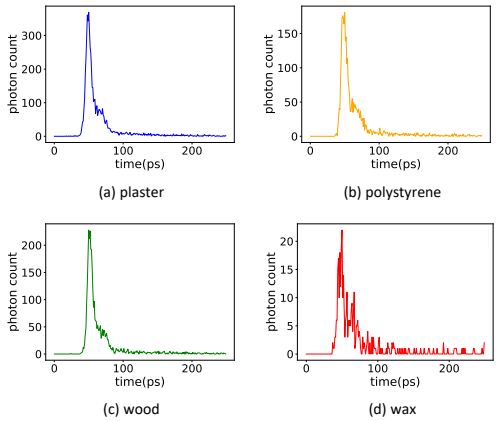

Fig. 2 Each material has a unique signature in its transient response shape, but differentiating them can be challenging due to the nonlinear nature of the transient shape.

based on the classifier's output.

2. Method

2.1 Single-Photon Avalanche Diode

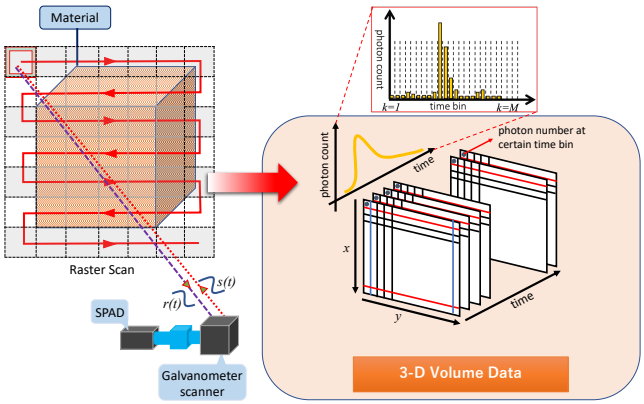

Fig. 3 The 3D coordinates represent the volume obtained from the raster scanning result. The x and y represent spatial coordinates, and the z coordinates represent temporal resolution.

The SPAD sensor works by utilizing the TCSPC principle [2], and it employs a Time-to-Digital Converter (TDC) for measuring time. TDC acts as a stopwatch to measure the time interval between a laser excitation pulse and the arrival of the corresponding photon. Then, a transient response in the form of a histogram is created based on the arrival time of each new pulse. In Fig. 3, a particular pixel on the material is illuminated by the photon-flux of waveform $s(t)$ and would reflect the waveform $r(t)$ that is incident at the SPAD $r(t)$ can be expressed as:

$$r(t) = as\left(t - \frac{2d}{c}\right) + b, \quad t \in \{1, \cdots, T_r\}, \tag{1}$$

where T_r is the repetition period of the pulse illumination. Assuming that a laser is illuminating a material with reflectance a and d represents the distance or depth to a certain point on the object. At the same time, c is the speed of light, and b is the background flux caused by external light [9]. The SPAD quantum efficiency η, which depends on the flux of the incoming light and the physical properties of the SPAD, can be expressed as:

$$\eta = (1 - R)e^{\alpha D}\left(1 - e^{\alpha W}\right), \tag{2}$$

if we let α be the absorption coefficient of silicon (the used SPAD is based on a reliable silicon avalanche photodiode sensitive in the visible spectral range), W be the thickness of the depletion region, D be the depth of the junction, and R be the power reflection coefficient for an interface between air and silicon [10], then assuming a constant background flux and the use of a single detector, we can express the rate function λ using equations Eq. (1) and Eq. (2) as follows:

$$\lambda(t) = \eta\left(as\left(t - \frac{2d}{c}\right) + b\right), \quad t \in \{1, \cdots, T_r\}, \tag{3}$$

Time bins are time units with a width that typically corresponds to the resolution of a stopwatch, which is usually some picoseconds. To create a histogram that represents photon numbers over time, we can map them to the corresponding time bins. Let's assume that the time bin size is Δ, and the repetition period pulse illumination is T_r. The pulse width is T_p for the sequence of N_s pulses. Then, we can define the total number of time bins that captured the photon as $M = T_r/\Delta$. Let the histogram h be the vector of size $M \times 1$ that contains the photon counts at each time bin after we illuminate the pixel N_s times with pulse waveform $s(t)$. Then, from the low-flux photon-counting theory [11], for the specific pixel i, j we have that:

$$h(i,j)_k \sim Poisson\left(N_s \int_{(k-1)\Delta}^{k\Delta} \lambda(t)dt\right), k = 1, \cdots, M. \tag{4}$$

The full-time bins of the respected pixel h are illustrated in Fig. 3 as a histogram. A histogram for the specific pixel $h(i, j)$, could be expressed as below:

$$h(i,j) = [h_1, \cdots, h_k, \cdots, h_M] \in \mathbb{R}^M, \tag{5}$$

because the same process could be applied to other pixels, we can omit i, j.

$$h \in \mathbb{R}^M. \tag{6}$$

Light transport (scattering) also differs depending on each material, and this affects a histogram's shape due to the different number of photons acquired per time bin.

2.2 Classification

A one-dimensional Fully Convolutional Neural Network (FCN) is a type of architecture used for material classification. Each layer of the FCN contains a one-dimensional time series. We utilize transient response histograms as input for the CNN. Afterwards, the function f maps each transient histogram h to corresponding material classes C. The 1-D FCN architecture used can be seen in Fig. 4.

$$f(h) : \mathbb{R}^M \to \mathbb{R}^C. \tag{7}$$

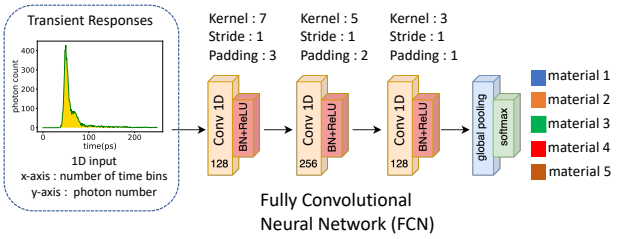

Fig. 4 1-D FCN architecture uses a transient histogram as an input. The result would be the class with the highest probability from the softmax function.

3. Experiment

3.1 Lab Prototype

The layout of our experiment can be seen in Fig. 5. We used SPAD ID100 and time controller ID900, which were manufactured by ID Quantique company, for the experiment. ID900 has a function generator and a Time-to-Digital Converter (TDC). A pulse laser and the TDC can be synchronized by sending a synchronization signal from the internal function generator. The TDC has 10,000 bins, and the shortest time resolution (Δ) of 13 ps from the TDC was used. We used a pulsed laser light source manufactured by Tama Denshi Co., Ltd. LDB-160C-639nm, with a pulse profile width of around 300 ps. An external trigger from the function generator drove the laser. In this experiment, a computer-controlled galvanometer mirror will perform raster scanning, while the target material will be illuminated by pulsed laser light transmitted through the mirror. The distance between the SPAD and the target is approximately 40 cm. To achieve picosecond-level accuracy in measuring the arrival times of photons, a timing histogram is constructed using a time-correlated single photon counting (TCSPC) technique. This method utilizes a time-to-digital converter (TDC) to correlate a synchronization signal from the laser as the start signal, and the stop signal is provided by photons detected by the single-photon avalanche diode (SPAD).

3.2 Dataset

The data used to train the model was stored in UCR format [12]. In this dataset, the target class was the type of material, and the variable was the number of photons per time bin, represented by a transient histogram. We generated a histogram of the transient response for seven materials by scanning each pixel. Five materials are used as a database: plaster, polystyrene, wood, wax, and leather, which are shown in Fig. 6. The time variable was fixed at a length of 10,000, which is the same as the transient histogram time bin.

We conducted our experiment in an environment free from external light and collected the measurement data for each material separately. During the training phase, we utilized a 63×63 pixels resolution dataset. However, we only trained the model using representative pixels belonging to the material, not the background. We determined which pixels were

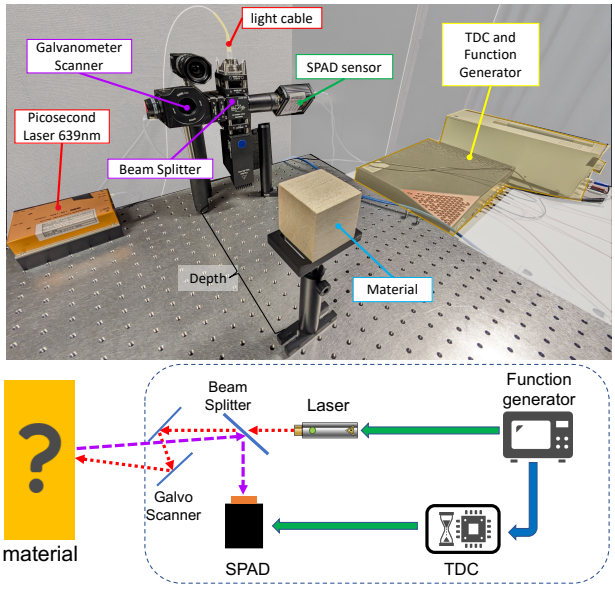

Fig. 5 Experimental Layout: pulsed illumination light is provided by a picosecond laser. Galvanometer scanners perform raster scanning across the target, and the SPAD is used to detect the scattered photons from the material. Photons detected by the SPAD are correlated with a sync signal from the laser using a TDC to measure the photon arrival timestamps.

Fig. 6 In our materials database, we exclude the background region to collect the reference of the transient histogram from each material.

representative by analyzing their intensity. The material intensity is typically higher than the background intensity since we used a black diffused background. This approach enabled us to train our model using transient data specific to each material. Table 1 presents the number of training datasets (number of transient responses from the several pixels of the material database) and the respective colourmap assigned to each material class.

Table 1 Material Class Datasets.

Class	Material	Colourmap	Datasets
0	plaster	blue	1217
1	polystyrene	orange	1217
2	wood	green	1217
3	wax	red	925
4	leather	brown	816

The reference transient histogram for each material can be seen in Fig. 2, and the curve plot colour represents the shape and colourmap of the transient histogram from each material.

3.3 Preprocess

In the preprocessing step, we reduce the time bin from 10,000 to 250 due to an important feature represented by the pulse. The pulse begins from the change in the initial

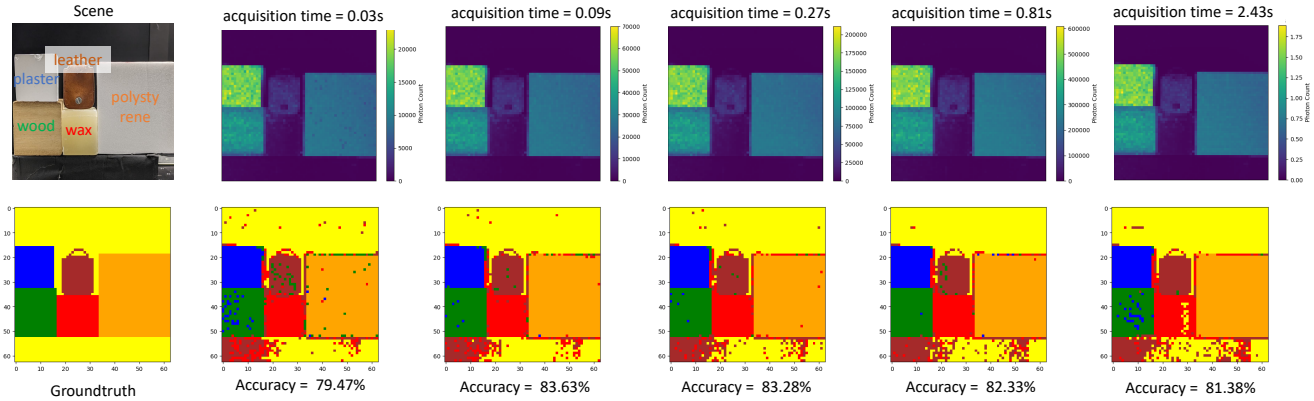

Fig. 7 Pixel-wise material classification comparison is used for different acquisition times.

slope of the transient histogram pulse and ends with the tail, which could be captured in the range of 250 time bins. To reduce the peak difference for each material object, we can ignore the difference in object depth Δd (Fig. 8). Here, we try to use the arg-max function to identify the peak of each object, which is an axis with a maximum number of photons. In Fig. 8, we can compare the peak for each object's material before and after alignment. We also perform h_z-normalization; as the intensity of the histogram obtained by SPAD is changed depending on the number of accumulation photons that are captured (sometimes the number of photons that are reflected by the same materials is varied and could give misinterpretation due to different peak heights). The transient histogram for each class is already going through a z-normalization process in Eq. (8):

$$h_z = \frac{h - \mu}{\sigma}, \tag{8}$$

where h_z is the histogram that already normalized, μ is the mean of the photon count rate, σ is the standard deviation of the photon count rate. We choose z-normalization due to its robustness to outliers, such as the common hot pixel in the measurement. Also, z-normalization could be used to remove offset and scaling in the time series classification [13].

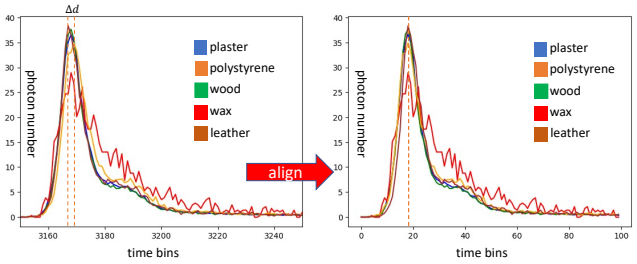

Fig. 8 We can ignore the difference in object depth Δd by trying to align the pulse peak for every material

4. Result

For the classifier, we used CNN architectures for 1-dimensional time series classification, namely FCN [14]. The Tsai [15] framework for the time series classification is being used with a learning rate of $l_r = 1e - 3$ for a total of 100 epochs. The classifier tries to distinguish between transient histograms in pixel-wise order, and the result would be a class represented by different colours for each material. In the process, the classifier will form one hot encoding to translate the result from the classification into a class number, and each class has a different colour label. Then, the results of pixel-wise segmentation based on the transient histogram will be given as a pixel map, which can be seen in Fig. 7.

Table 2 Accuracy Result

Acquisition Time (s)	Accuracy (%)
0.03	79.47
0.09	83.63
0.27	83.28
0.81	82.33
2.43	81.38

Table 2 presents the accuracy results for various acquisition times. On average, the accuracy is approximately 83(%) when the background (yellow colourmap) is not considered. The noise contributed to misclassification due to the acquisition time being too short at 0.03 seconds. Longer acquisition time may reduce noise but could also lower accuracy. Temporal resolution is also a factor that can improve classification accuracy, assuming that finer time resolution would lead to better accuracy. We intend to further investigate the impact of time resolution on classification accuracy in the future.

5. Conclusion

In our paper, we introduced a new method for classifying materials at the pixel level using deep learning and transient responses. We used the features extracted from the transient responses dataset as inputs for the deep learning models. Our experiments showed that adjusting the acquisition time improved the accuracy of material segmentation. In future work, we aim to investigate how temporal resolution affects accuracy.

Acknowledgements This work was supported by JSPS KAKENHI Grant Number JP20K20629.

References

[1] Su, S., Heide, F., Swanson, R., Klein, J., Callenberg, C., Hullin, M. and Heidrich, W.: Material Classification Using Raw Time-Of-Flight Measurements, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).

[2] Wahl, M., Bülter, A., Buller, G., Collins, R., Lauritsen, K., Riecke, S., Schönau, T., Dertinger, T., Ruettinger, S., Ishii, K., Otosu, T., Tahara, T., Barth, A., Voithenberg, L., Lamb, D., Grußmayer, K., Herten, D.-P., Ruedas-Rama, M., Alvarez-Pez, J. and Erdmann, R.: *Advanced Photon Counting: Applications, Methods, Instrumentation* (2015).

[3] Kitano, K., Okamoto, T., Tanaka, K., Aoto, T., Kubo, H., Funatomi, T. and Mukaigawa, Y.: Recovering temporal PSF using ToF camera with delayed light emission, *IPSJ Transactions on Computer Vision and Applications*, Vol. 9, No. 1, p. 15 (2017).

[4] Morimoto, K., Ardelean, A., Wu, M.-L., Ulku, A. C., Antolovic, I. M., Bruschini, C. and Charbon, E.: Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications, *Optica*, Vol. 7, No. 4, pp. 346–354 (2020).

[5] Sun, Q., Dun, X., Peng, Y. and Heidrich, W.: Depth and Transient Imaging with Compressive SPAD Array Cameras, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 273–282 (2018).

[6] Fujimura, Y., Kushida, T., Funatomi, T. and Mukaigawa, Y.: NLOS-NeuS: Non-line-of-sight Neural Implicit Surface, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10532–10541 (2023).

[7] Yao, D., Connolly, P. W. R., Sykes, A. J., Shah, Y. D., Accarino, C., Grant, J., Cumming, D. R. S., Buller, G. S., McLaughlin, S. and Altmann, Y.: Rapid single-photon color imaging of moving objects, *Opt. Express*, Vol. 31, No. 16, pp. 26610–26625 (2023).

[8] Becker, C. N. and Koerner, L. J.: Plastic Classification Using Optical Parameter Features Measured with the TMF8801 Direct Time-of-Flight Depth Sensor, *Sensors*, Vol. 23 (2023).

[9] O'Toole, M., Heide, F., Lindell, D. B., Zang, K., Diamond, S. and Wetzstein, G.: Reconstructing Transient Images from Single-Photon Sensors, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2289–2297 (2017).

[10] Zappa, F., Tisa, S., Tosi, A. and Cova, S.: Principles and features of single-photon avalanche diode arrays, *Sensors and Actuators, A: Physical*, Vol. 140, pp. 103–112 (2007).

[11] Shin, D., Xu, F., Venkatraman, D., Lussana, R., Villa, F., Zappa, F., Goyal, V. K., Wong, F. N. and Shapiro, J. H.: Photon-efficient imaging with a single-photon camera, *Nature Communications*, Vol. 7 (2016).

[12] Dau, H. A., Bagnall, A. J., Kamgar, K., Yeh, C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A. and Keogh, E. J.: The UCR Time Series Archive, *CoRR*, Vol. abs/1810.07758 (2018).

[13] Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J. and Keogh, E.: Addressing Big Data Time Series: Mining Trillions of Time Series Subsequences Under Dynamic Time Warping, *ACM Trans. Knowl. Discov. Data*, Vol. 7, No. 3 (2013).

[14] Wang, Z., Yan, W. and Oates, T.: Time series classification from scratch with deep neural networks: A strong baseline, *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1578–1585 (2017).

[15] Oguiza, I.: tsai - A state-of-the-art deep learning library for time series and sequential data, Github (2022).