Robust Time-of-Flight-based Material Imaging using Three-Dimensional Deep Neural Networks on Spatial Neighborhoods of Pixels

Rajababu Udainarayan Singh and Miguel Heredia Conde
Center for Sensor Systems (ZESS), University of Siegen, Siegen, Germany
E-mail: rajababu.singh@student.uni-siegen.de and heredia@zess.uni-siegen.de
ORCID number: 0000-0003-4526-6999 and 0000-0001-5218-0822

***Abstract*—Time-of-Flight (ToF) cameras are active sensors for depth assessment between camera and object that measure the time traveled by the modulated light from an optical transmitter to a pixel array. The conventional methods of material imaging supported by RGB cameras fail in the dense mapping of look-alike materials. To date, the Material Impulse Response Function (MIRF), which yields valuable features for distinguishing materials using ToF cameras has been considered mostly in the temporal dimension. Our novel approach introduces material imaging based on spatial and temporal dimensions, exploiting three-dimensional features. Firstly, we propose an innovative approach to per-pixel material imaging using a set of features over a spatial neighborhood. Secondly, we introduce a bilateral weight matrix to boost the quality of the ToF feature set. Thirdly, an attempt has been made to avoid boundary region pixel misclassification while simultaneously reducing the computation by selecting the K-nearest neighbors in the ToF feature space within a patch. Finally, the above-mentioned approaches are validated on several datasets using newly-proposed three-dimensional deep learning models.**

***Index Terms*—Time-of-Flight, material imaging, spatial correlation, three-dimensional convolutional neural networks**

2022 IEEE Sensors | 978-1-6654-8464-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/SENSORS52175.2022.9967293

I. Introduction

Time-of-Flight (ToF) cameras perceive their environment using the ToF principle [1]–[3], akin to bat's range quantification behavior using ultrasonic vocalization. The ToF camera indirectly measures range by assessing the time delay of near-infrared (NIR) modulated light traveling from an optical source, reflecting off a scene, to an optical receiver by employing intelligent pixels, e.g., based on the Photonic Mixer Device (PMD) technology [4]–[6].

Solid-state ToF cameras were introduced in the mid-nineties. Despite initially being designed to estimate distances on a per-pixel basis, recent works [7], [8], have pointed out the possibility of performing material imaging with them. Classifying material categories, e.g., wood, foam, plaster, metal, and so on [9], [10], based on their texture using classical RGB cameras faces the fundamental problem of misidentifying materials having a similar appearance [11]. This highlights the importance of ToF-based material imaging in numerous applications, e.g., autonomous driving, robotics, manufacturing, and assembly, along with others [8].

In this paper, we focus on the methods explored for material recognition in [7], [8], which acquire Fourier samples of the Material Impulse Response Function (MIRF). The main contributions of our work are as follows:

- The major difference from the previous work is that we have taken advantage of correlations between pixels in the spatial domain, whereas till now, only the temporal dimension has been exploited (see Fig. 1).
- Introduction of a sliding window technique that modifies the direct Fourier samples with a spatially-varying bilateral filtering kernel.
- We propose an alternative way of exploiting local neighborhoods where pixels belonging to the same material are identified and extracted by using the well-known K-nearest neighbor (KNN) algorithm.
- Thorough experimental validation of the proposed methods and deep learning models, showing superior performance as compared to the prior art and ability to cope with boundary regions.

II. Related work

The earliest work on material recognition that uses ToF cameras is presented in [12], where new methodologies have been provided to acquire data employing indirect laser illumination.

The last decade has seen a steep rise in research related to ToF-based material imaging [1], [10], [13]. In [14], materials have been classified using Single-Photon Avalanche Diodes (SPADs), which can time-stamp individual photon arrivals. In [15], a modification of the Torrance-Sparrow model has been used to represent the surface reflectance component. These surface reflectance components have been used to identify the materials. An RGBD camera mentioned in [16], uses surface roughness for material classification. In [17], an optimization method has been used to reduce surface interreflection scattering. Normalization has been used to extract the Reflection Point Spread Function (RPSF).

In [18], an adaptive neighborhood convolutional neural network (AN-CNN) model (i.e., similar to our spatial window method) is proposed for terrain classification considering homogeneous and boundary regions and a maximum of 87% accuracy has been attained. Differently, we use the modified window prior to the CNN classifier.

Authorized licensed use limited to: UB Siegen. Downloaded on December 09,2022 at 22:20:23 UTC from IEEE Xplore. Restrictions apply.

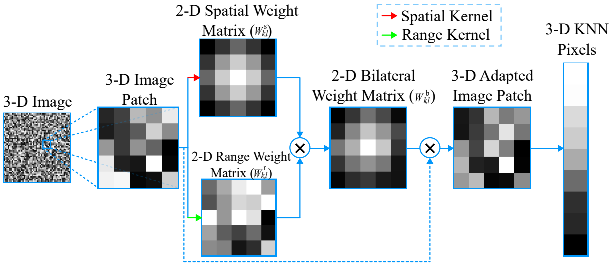

2-D Spatial Weight Matrix (W^{s}_{kl})
Spatial Kernel
Range Kernel
3-D KNN Pixels
3-D Image
3-D Image Patch
2-D Bilateral Weight Matrix (W^{b}_{kl})
3-D Adapted Image Patch
2-D Range Weight Matrix (W^{r}_{kl})

Fig. 1: An illustration of our proposed methods. The method starts by sliding a window of size $5 \times 5 \times 8 \times 2$ on the 3D image, simultaneously getting modified using a 2D bilateral weight matrix. Finally, congruent pixels have been extracted using KNN for training and validation.

III. Methodology

A. Generation of Three-Dimensional Data Cube

ToF sensors collect unique information about materials in a set of images taken at multiple frequencies. In prior work [7], [8], only the temporal frequency dimension has been exploited, ignoring the correlations of pixels in spatial domain, which is two-dimensional. For this reason, now we bring both worlds together, and hence we have three-dimensional features, where each pixel in the spatial domain can be expressed with a temporal-frequency feature vector (see Fig. 1). We propose to stack the complex phasor images $I \in \mathbb{C}^{n \times m \times K}$ taken at K = 8 different temporal frequencies (uniformly distributed, from 20 MHz to 160 MHz) in ascending order to form a three-dimensional multi-frequency data cube for processing, where n and m represent the two spatial dimensions and K represents the temporal frequency dimension.

B. Three-Dimensional ToF Image Patch Extraction

We stack the $\mathbb{R}$ and $\mathbb{I}$ components of the complex pixels in the fourth dimension featuring two channels ($c_{\text{ch}} = 2$). In prior state-of-the-art [8], per-pixel material imaging has been observed to provide relatively low accuracy due to insufficient features. To address this problem, considering that pixels belonging to the same material will typically appear grouped together in regions, we propose a 25-fold extension of feature vector in spatial domain by using a finite window $W \in \mathbb{R}^{i \times j \times K \times c_{\text{ch}}}$, where i and j represent the size of window in spatial dimensions, which will slide over the entire padded image $I^{\text{p}} \in \mathbb{R}^{n^{\text{p}} \times m^{\text{p}} \times K \times c_{\text{ch}}}$ (see Fig. 1) with a stride S = 1 to acquire data for per-pixel material imaging using an image patch supported on both spatial and temporal domain. Choosing $i = j = 5$ and the padding to be either symmetric or replicate yielded superior performance.

C. Feature Adaptation

As described in [18], while predicting the class of the central pixel of the image patch using a CNN, each pixel within the image patch will exhibit the same influence on the classification result. However, both the distance in the spatial domain and feature space should have an influence on how each pixel in the patch contributes to the classification result. In fact, pixels with smaller bilateral distances should exhibit a stronger influence on the classification result compared to pixels having larger bilateral distances (see Fig. 1). Due to this reason, we propose the use of a bilateral weight matrix ($W^{\text{b}}_{k,l}$) (see (3)) similar to the weight matrix used in [19], [20], on our ToF image patch (see Fig. 1). Let (k, l) be the coordinate of a neighboring pixel and (r, c) be the coordinates of the pixel that needs to be classified, then our $W^{\text{b}}_{k,l}$ is:

$$W^{\text{s}}_{k,l} = \exp\left\{-\frac{(k-r)^2 + (l-c)^2}{2\sigma_D^2}\right\} \tag{1}$$

$$W^{\text{r}}_{k,l} = \exp\left\{-\frac{||(\vec{I_{k,l}} - \vec{I_{r,c}})||^2}{2\sigma_R^2}\right\} \tag{2}$$

$$W^{\text{b}}_{k,l} = W^{\text{s}}_{k,l} \times W^{\text{r}}_{k,l} \tag{3}$$

In (1) and (2), σ_D and σ_R, are the smoothing parameters. In particular, the spatial weight matrix ($W^{\text{s}}_{k,l}$) (see (1)) should favor close neighborhoods of pixels, whereas range weight matrix ($W^{\text{r}}_{k,l}$) (see (2)) should suppress the pixels belonging to a class other than that of the central pixel in the image patch.

D. KNN Pixel Refinement

In prior work [7], per-patch material imaging has been observed to give poor results in the boundary regions. For this reason, we propose extraction of a set of pixels exhibiting the smallest distance in the Fourier-feature vector space with respect to the center pixel within the image patch using Euclidean distance metric [21], [22] and concatenation, as depicted in Fig. 1. Our analysis has shown that choosing KNN = 9, which means an increase of 9-fold in spatial domain feature vector compared to previous work [7], [8], has yielded the highest classification accuracy (see Fig. 2 and 3).

E. Deep Learning Model Development

To date, most material imaging has been carried out using conventional machine learning algorithms, such as Support Vector Machine (SVM), Fully Connected Neural Network (FCNN), along with others in [7], [8], [17]. The voxel information from adjacent slices in a three-dimensional data cube can provide crucial information for ToF-based material imaging. For this reason, in addition to FCNN, we have exploited several 3-D deep learning algorithms, where each neural network counts with one 3-D input, one softmax, and one final classification layer, while the Residual Neural Network (ResNet) features two addition layers (see Table I). We have developed a small 3-D CNN (see Table I) consisting of 25 layers, which allowed reducing the number of parameters without losing performance. Moreover, we have constructed a 3-D ResNet (see Table I) consisting of 29 layers, which helps in avoiding the exploding and vanishing gradients during training. An adam optimizer with an initial learning rate of 0.08 has yielded the best result for our models.

Authorized licensed use limited to: UB Siegen. Downloaded on December 09,2022 at 22:20:23 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Deep Learning Models and Test Accuracy

		Deep Learning models		
		3-D CNN	*3-D ResNet*	*FCNN*
Layers	BatchNormalization	7	8	9
	Convolution3D	4	5	-
	ReLU	6	7	8
	Dropout	2	1	1
	Fullyconnected	3	3	9
Accuracy	12M dataset	95%	93%	95%
	14M dataset	99%	97%	99%
	5M dataset	97%	96%	98%

IV. Experimental results

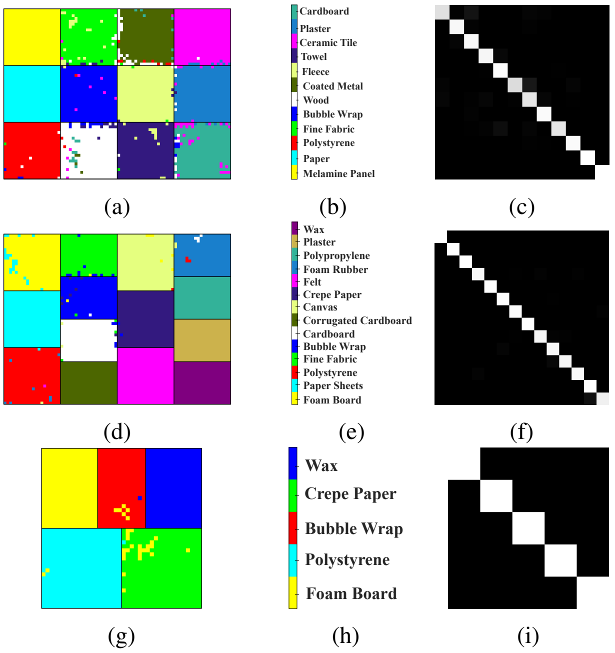
We have evaluated our proposed methods on three datasets (see Tab. I), which have been already used for processing and analysis using state-of-the-art ToF material imaging methods [7], [8]. The 12M (i.e., 12 Materials from the 15M dataset) and 14M datasets are having three instances per material taken at 3 different distances. Due to the fewer number of instances in the previous two datasets, a new dataset has been acquired in the year 2021 [23], which consists of five materials, each taken at 10 different distances and 7 different orientations. For all datasets, the ratio between training and validation is kept at 70 : 30.

Due to each image consisting of homogeneous material for the above-mentioned datasets, we have developed several artificial heterogeneous test setups to evaluate our proposed methods, where Table I depict the attained test accuracy using our developed deep learning models. From Fig. 2a and 2c, we can see that cardboard, coated metal, and wood from the 12M dataset are getting confused with each other. Nevertheless, our model is able to attain a maximum of 95% accuracy. From the 14M dataset, Fig. 2d and 2f depict that corrugated cardboard, felt, paper sheets, plaster, polypropylene, and wax are achieving 100% classification accuracy, while the overall maximum accuracy is 99%. The maximum accuracy of 98% obtained for the 5M dataset (see Fig. 2g and 2i) has proved that our methods perform well on a dataset exhibiting high variations in terms of distances and orientations.

Finally, to evaluate our proposed methods on a real heterogeneous target [24], we have constructed a demonstrator setup [25], (see Fig. 3a and 3b), consisting of five materials. Each image has been taken at 5 different distances (uniformly distributed between 82 cm to 47 cm) and at $[-10°, 0°, 10°]$, using the PMD Selene module mentioned in [2], [4]. Based on the performance of our models on artificial heterogeneous targets, we adopt the 3-D CNN classifier mentioned in Section III-E and attained 99% validation accuracy (see Fig. 3c).

V. Conclusion

In this paper, we have provided a series of improvements over prior work dealing with the novel idea of using ToF sensors for per-pixel material sensing using machine learning techniques. We have presented a technique that takes advantage of the fact that pixels observing the same material are typically grouped together in regions. Experimental evaluations have shown improvements in classification accuracy in challenging real ToF datasets with respect to a baseline not considering spatial neighborhoods, reaching 95% to 99% accuracy. The results of this research are expected to constitute a step forward towards robust material imaging, unleashing new application domains for ToF cameras. Future work contemplates evaluating the proposed method on a much higher number of materials. In addition, an attempt can be made on adaptive KNN, such that sudden changes in temporal frequency-feature space can be automatically detected.

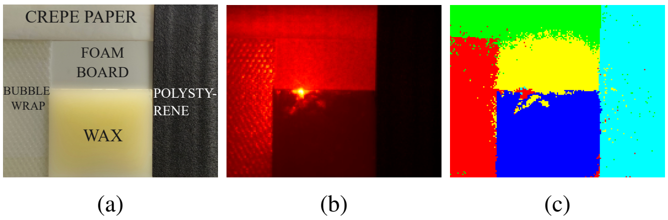


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 2: Per-pixel material imaging of artificial heterogeneous targets using our proposed methods on three datasets: (a), (d), and (g) show the results of the 3-D CNN on the 12M (b), 14M (e), and 5M (h) datasets, respectively, while (c), (f), and (i) depict the corresponding confusion charts (known class per columns and predicted class per rows, white: 100% accuracy). The black line highlights the boundaries between materials, while the dominant color corresponds to the correct material.



(a) (b) (c)

Fig. 3: The demonstrator setup (a) consists of five different materials (see Fig. 2h) placed at approximately the same distance to the camera. (b): NIR amplitude image, (c): material image obtained using the proposed approach. The coloring is according to Fig. 2h.

Authorized licensed use limited to: UB Siegen. Downloaded on December 09,2022 at 22:20:23 UTC from IEEE Xplore. Restrictions apply.

References

[1] R. Lange and P. Seitz, “Solid-state Time-of-Flight range camera,” *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390–397, 2001.

[2] M. Heredia Conde, *Compressive Sensing for the Photonic Mixer Device: Fundamentals, Methods and Results*, ch. 2, pp. 11–49. Springer Vieweg Wiesbaden, 2017.

[3] B. Langmann, K. Hartmann, and O. Loffeld, “Increasing the accuracy of Time-of-Flight cameras for machine vision applications,” *Computers in Industry*, vol. 64, no. 9, pp. 1090–1098, 2013. Special Issue: 3D Imaging in Industry.

[4] X. Luan, *Experimental Investigation of Photonic Mixer Device and Development of TOF 3D Ranging Systems Based on PMD Technology*. PhD thesis, Dept. Elect. Eng. and Comput. Sci., Univ. of Siegen, Siegen, NRW, Germany, 2001.

[5] M. Heredia Conde, K. Hartmann, and O. Loffeld, “Subpixel spatial response of PMD pixels,” in *2014 IEEE International Conference on Imaging Systems and Techniques (IST) Proceedings*, pp. 297–302, 2014.

[6] S. Foix, G. Alenya, and C. Torras, “Lock-in time-of-flight (tof) cameras: A survey,” *IEEE Sensors Journal*, vol. 11, no. 9, pp. 1917–1926, 2011.

[7] M. Heredia Conde, T. Kerstein, B. Buxbaum, and O. Loffeld, “Near-infrared, depth, material: Towards a trimodal Time-of-Flight camera,” in *2020 IEEE SENSORS*, pp. 1–4, 2020.

[8] M. Heredia Conde, “A material-sensing time-of-flight camera,” *IEEE Sensors Letters*, vol. 4, no. 7, pp. 1–4, 2020.

[9] S. Su, F. Heide, R. Swanson, J. Klein, C. Callenberg, M. Hullin, and W. Heidrich, “Material classification using raw Time-of-Flight measurements,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3503–3511, 2016.

[10] K. Tanaka, Y. Mukaigawa, T. Funatomi, H. Kubo, Y. Matsushita, and Y. Yagi, “Material classification using frequency-and depth-dependent Time-of-Flight distortion,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2740–2749, 2017.

[11] J. Holloway, T. Priya, A. Veeraraghavan, and S. Prasad, “Image classification in natural scenes: Are a few selective spectral channels sufficient?,” in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 655–659, 2014.

[12] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala, “Single view reflectance capture using multiplexed scattering and Time-of-Flight imaging,” *ACM Trans. Graph.*, vol. 30, p. 1–10, dec 2011.

[13] G. Agresti and S. Milani, “Material identification using RF sensors and convolutional neural networks,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3662–3666, 2019.

[14] C. Callenberg, Z. Shi, F. Heide, and M. B. Hullin, “Low-cost SPAD sensing for non-line-of-sight tracking, material classification and depth imaging,” *ACM Trans. Graph.*, vol. 40, jul 2021.

[15] A. M. Mannan, H. Fukuda, L. Cao, Y. Kobayashi, and Y. Kuno, “3D free-form object material identification by surface reflection analysis with a Time-of-Flight range sensor,” *Conference on Machine Vision Application*, pp. 227–230, 2011.

[16] J. Kim, H. Lim, S. C. Ahn, and S. Lee, “RGBD camera based material recognition via surface roughness estimation,” in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1963–1971, 2018.

[17] S. Lang, J. Zhang, Y. Cai, and Q. Wu, “Classification of materials using a pulsed Time-of-Flight camera,” *Machine Vision and Applications*, 2021.

[18] A. Zhang, X. Yang, L. Jia, J. Ai, and Z. Dong, “SAR image classification using adaptive neighborhood-based convolutional neural network,” *European Journal of Remote Sensing*, vol. 52, no. 1, pp. 178–193, 2019.

[19] F. Banterle, M. Corsini, P. Cignoni, and R. Scopigno, “A low-memory, straightforward and fast bilateral filter through subsampling in spatial domain,” *Computer Graphics Forum*, vol. 31, pp. 19–32, February 2012.

[20] S. M. Aswatha, J. Mukhopadhyay, and P. Bhowmick, “Image denoising by scaled bilateral filtering,” in *2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*, pp. 122–125, 2011.

[21] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[22] F. Nigsch, A. Bender, B. van Buuren, J. Tissen, E. Nigsch, and J. B. O. Mitchell, “Melting point prediction employing k-nearest neighbor algorithms and genetic parameter optimization,” *Journal of Chemical Information and Modeling*, vol. 46, no. 6, pp. 2412–2422, 2006. PMID: 17125183.

[23] S. K. Kasam and M. Heredia Conde, “Multi-channel near infrared ToF response images of five materials.” https://dx.doi.org/10.21227/0142-7561, 2022.

[24] R. U. Singh and M. Heredia Conde, “Heterogeneous target Time-of-Flight dataset.” https://dx.doi.org/10.21227/e9ex-by73, 2022.

[25] M. Heredia Conde and R. U. Singh, “Live demonstration: a trimodal time-of-flight camera with enhanced material imaging,” in *2022 IEEE SENSORS*, pp. 1–1, 2022. to appear.

Authorized licensed use limited to: UB Siegen. Downloaded on December 09,2022 at 22:20:23 UTC from IEEE Xplore. Restrictions apply.