Imprint
Dr.-Ing. habil. Miguel Heredia Conde
Center for Sensorsystems (ZESS)
University of Siegen
57076 Siegen
Germany

University of Siegen
Faculty of Natural Science and Engineering
Department of Electrical Engineering and Computer Science
Master of Science - Mechatronics

Studienarbeit

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

Author:
Rajababu Udainarayan Singh (1530580)

Supervised by
Dr.-Ing. habil. Miguel Heredia Conde

Date of Submission: January 2, 2023

University of Siegen
Center for Sensorsystems (ZESS)
Paul-Bonatz-Straße 9-11
H-57076 Siegen
Germany

Acknowledgment

I would like to express my greatest gratitude towards **Dr.-Ing. habil. Miguel Heredia Conde** for the vision of exploiting Time-of-Flight camera-based material sensing in multi-dimensions and for trusting me with such an exciting mission. Without him and his support, the results in this studienarbeit would not have been achieved.

I would like to express my special thank to **Mr. Peter Hof** for his technical assistance and cooperation during the experimental setup.

I would like to thank the Center for Sensorsystems (ZESS) at the University of Siegen for providing a nice environment accommodated with important experimental setups.

Finally, I want to thank my parents, especially my brother for their contribution in every aspect along the years. Without their support it won't be possible for me to finish this work.

Abstract

Time-of-Flight cameras are active sensors for depth assessment between the 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 classification 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 both 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 three-dimensional features. 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.

Abbreviation

Contents

List of Figures

Chapter 1

Introduction

A use of Time-of-Flight (ToF) sensor that you may not know! Engineers Alfred Fielding and Marc Chavannes originally invented "bubble wrap" in 1957, in Hawthorne, New Jersey, with the intention of selling it as a textured wallpaper by sealing two shower curtains together, creating a disorganized cluster of air bubbles. However, Fielding and Chavannes quickly discovered a very successful *second use* for it as a packing material for fragile products. In a similar manner, the optical ToF sensor despite being initially developed to determine an object's proximity using the modulated light that is reflected back after it, in more recent years, a desire to identify various materials using ToF sensors has emerged drastically, which will be deeply covered in this studienarbeit/student work.

Chapter Outline: In this chapter, first, we begin with a brief discussion on the background of the vision sensors, in Section 1.1, as we need the concept to set up our problem. Here we explore the fundamentals of the digital image sensor in Subsection 1.1.1 and then discuss ToF technology in Subsection 1.1.2. Following that, we extend our discussion on the working principle of the so-called array of smart pixels (i.e., PMD pixels) in Subsection 1.1.3. Then, we provide detailed information about the main topic of this student work, which is enhancing the material sensing using a ToF camera in Section 1.2, where we provide a detailed description of the research purposes in Subsection 1.2.1 and finally, we provide our contribution in short words in Section 1.3 and then, will close our discussion by providing the detailed outline about the coming chapters in Section 1.4.

1.1 Background

1.1.1 Fundamentals of the Digital Image Sensor

One of the most crucial elements of any camera used for vision is the image sensor, whose primary job is to convert the light (photons) into an electrical signal, however, not all sensors are built the same. Numerous categories exist for classifying sensors, including chroma type (i.e., color or monochromatic), shutter type (i.e., global or rolling shutter), and structure type (i.e., Charge-Coupled Device (CCD) or Complementary Metal–Oxide–Semiconductor (CMOS)). On the other hand, the resolution, frame rate, pixel size, and sensor form factor can also be used to categorize them. Here, we go over some of the fundamentals of the image sensor technology used in vision cameras and how it relates to various classes. The purpose of image sensors is the same; to convert incoming light (photons) into an electrical signal that can be processed, analyzed, viewed, or stored for future application. The image sensor is a solid-state device made up of micro-lenses, light-sensitive elements, and micro-electrical elements in chips that are first taken from raw wafers, and divided into numerous sections, each of which contains a single sensor die.

Sensor Functions Inside a Camera The incident light (photons) received by the image sensor is usually focused through micro-lenses or other optics. Each pixel in a color sensor is sensitive to a particular color wavelength thanks to an additional layer composed of an array of color filters that sits below the micro-lens and absorbs light of undesirable wavelengths. There is no color filter present in monochromatic sensors, hence every pixel is sensitive to all visible light wavelengths. The pixel size, which is frequently represented in micrometers (µm), includes the whole photo-surface diode as well as any adjacent electronics. For increased light sensitivity, larger pixel sizes are often preferred because there is more photo-diode surface area to receive light. On the other side, smaller pixels are needed for the same sensor format with a higher resolution, which can lower sensor sensitivity. An essential part of any image sensor is the shutter type, which controls when each pixel starts accumulating charge and when the accumulated charge is to be read out and the pixel reset. The two main forms of electrical shutters are rolling shutters and global shutters, which function differently and yield distinct picture results, especially when the camera or target is moving.

1.1.2 Range Imaging

The term "range imaging" refers to a group of methods used to create a 2-D image whose pixels indicate the distance between the camera and each point of the scene. If properly calibrated, the pixel values can be represented directly in physical units, such as meters. In this Subsection, we will present some of the most widely used methods of range imaging.

Stereo Triangulation A stereo triangulation technology [3], is a passive method of extracting depth data on each pixel by finding the disparity between the corresponding points from different images acquired using a multiple-camera setup system, that has been positioned at known distances. Since only some of the spots are visible to all cameras, range imaging based on stereo triangulation often only yields accurate depth estimates for a small part of the spots.

Sheet of Light Triangulation Sheet of Light Triangulation, as opposed to stereo, is an active method of determining depth data on each pixel using a single high-resolution camera and a light source capable of illuminating the scene with a sheet of light. The sensor is able to detect the angle of the reflected light in relation to the plane of the projected light, allowing for accurate measurement of the distance to the surface. This method is particularly useful for measuring the distance to flat or slightly curved surfaces, as the sheet of light can be aligned with the surface to ensure accurate measurement. A series of depth features of the scene can be generated by moving the light source, the camera, or the scene keeping the remaining two stationary.

Structured light Depth can be determined by illuminating the scene with a specially designed light pattern, which can be, for instance, in the form of horizontal and vertical lines, points, or checkerboard patterns.

Time-of-Flight The ToF cameras perceive their environment using ToF technology [4–6], akin to bat's range quantification behavior using ultrasonic vocalization. The ToF camera indirectly measures range or depth d (see (1.1)), by assessing the time delay of Near-Infrared (NIR) modulated light traveling from an optical source, reflecting off a scene, to an optical receiver. This is done by employing intelligent pixels, e.g., based on the PMD technology [7–9]. A phase shift θ_{depth} can be measured using a periodic modulation signal, which can be in the form of either pulsed modulation or continuous wave modulation.

$$d = \frac{c}{4\pi f}\theta_{depth} \tag{1.1}$$

Where $c = 299792458\ m/s$ represents the speed of light in the medium and f represents the modulation frequency.

1.1.3 Photonic Mixer Device (PMD)

The PMD consists of an array of smart pixels, which sample incoming light synchronously with the emission of modulated light. Unlike the basic technology mentioned in Section 1.1.1, the PMD pixel allows for an alternating voltage inside each pixel's photodiode, creating drift fields that divide and pull converted electrons to different detector junctions (see Fig. 1.1). This concept is based on the creative idea of **Prof. Rudolf Schwarte**, from *University of Siegen, Zentrum für Sensorsysteme (ZESS)*. Due to its unique operation principle of performing concurrent mixing and charge integration procedure in the photosensitive area of the PMD, it offers an excellent solution to overcome the most typical difficulties in the conventional 3-D ranging system design with continuous wave modulation. Each smart PMD pixel is an independent ranging unit, having the ability to deliver range-related information. The ability to easily integrate the PMD pixels into a PMD line or PMD matrix allows the realization of compact, flexible, fast, and robust low-cost 3-D ToF solid-state ranging systems.

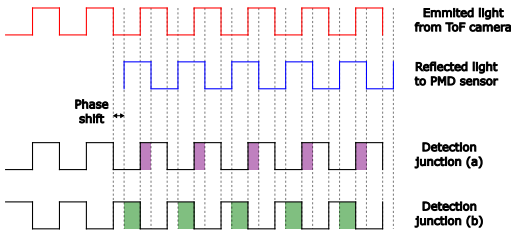
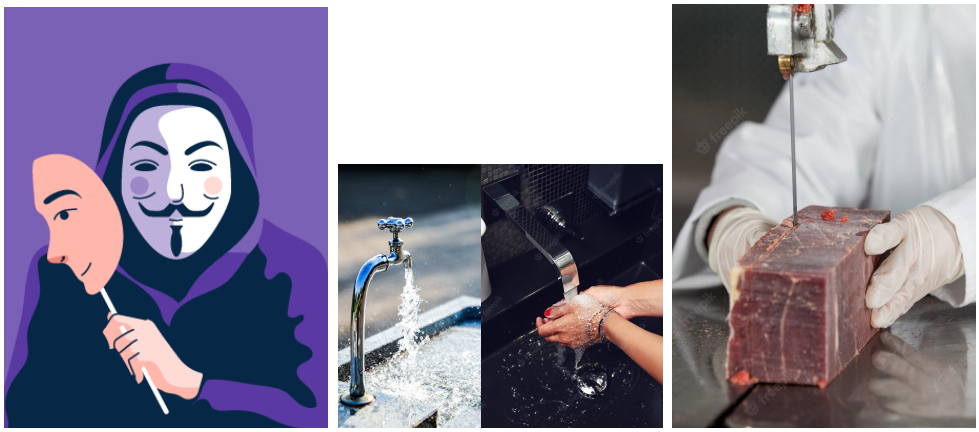


Figure 1.1: An illustration of the timing diagram of a ToF camera.

1.2 Time-of-Flight-based Material Sensing

Solid-state ToF cameras were introduced in the mid-nineties. Despite initially being designed to estimate distances on a per-pixel basis, recent works [2, 10], have pointed out the possibility of performing material imaging with them. Classifying material categories, e.g., wood, foam, plaster, metal, and so on [11, 12], based on their texture using classical RGB cameras faces the fundamental problem of misidentifying materials having a similar appearance. Moreover, in the current digital world, where secure identification increasingly relies on face recognition with conventional cameras, spoofing can pose a real threat to people (see Fig. 1.2a). Additionally, to avoid wastage of our limited natural resources, such as water, a ToF sensor for material imaging can be a useful tool by supplying water only when

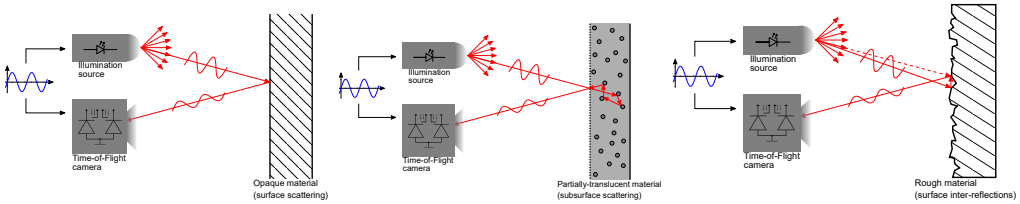


(a) (b) (c)

Figure 1.2: Some real-world application scenarios for material imaging using ToF sensor. Image (a) from [1].

it detects skin (see Fig. 1.2b). In production facilities, such as the food industry, a minor error during handling food products can lead to serious injuries (see Fig. 1.2c). This highlights the importance of ToF-based material imaging in numerous applications, e.g., autonomous driving, robotics, manufacturing, and assembly, along with others [10].

Though since traditional material imaging methods supported by RGB cameras perform poorly in the dense classification of look-alike materials (see Fig. 1.2). Thanks to the unique behavior of the Material Impulse Response Function (MIRF) for each material, which yields valuable features for distinguishing material samples using ToF cameras, by considering various scattering phenomena (see Fig. 1.3).



(a) Surface scattering (b) Subsurface scattering (c) Surface inter-reflections

Figure 1.3: Schematics of the suggested material imaging concept illustrating various scattering phenomena caused by the light source. Image from [2]

1.2.1 Research Purposes

The purpose of this research project is to exploit the behavior of ToF sensor-based material imaging using spatial neighborhoods of pixels with various three-dimensional deep neural networks. The prior state-of-the-art [2, 10] in ToF based per-pixel material imaging has exploited only in the temporal frequency dimension, and experimental evaluation shows that it is still far from a real-world application. Hence, one way to have better material recognition is to exploit the pixel neighborhood in the spatial domain. The sub-tasks for this student work are as follows:

1. Literature review on theoretical foundations of ToF related cameras.
2. Literature review on ToF based material imaging.
3. Handling the dataset imbalance.
4. Developing algorithms that exploit spatially correlated pixels for ToF-based material imaging.
5. Benchmark our method against prior state-of-the-art material imaging.

1.3 Contributions

In this paper, we focus on the methods explored for material recognition in [10], where direct Fourier samples of the MIRF have been acquired and processed for per-pixel material imaging. Additionally, a voting scheme has been adopted for super-pixel classification. Meanwhile in [2], due to variation in temperature of the PMD sensor, which is having a direct effect on the acquired samples, a temperature calibration method has been proposed.

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. 3.1).
- 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.
- The other important contribution of our work is to handle class-imbalanced classification tasks.

1.4 Student Work Outline:

This student work is organized into five chapters. At first, we introduced the background and objective of this research together with outlining our contributions in Chapter 1. In Chapter 2, we briefly summarize various material imaging techniques that have been so far introduced and we describe how our proposed methods differ from them. In Chapter 3, we propose a general image processing technique with the deep learning framework that we followed to solve the material sensing classification problem. Then, we report the results of applying the algorithm to several benchmark datasets to prove the effectiveness of our method in Chapter 4. Besides, we compare our experimental results with the previous state-of-the-art material sensing techniques. Finally, in Chapter 5, we state the summary and discuss the results. Moreover, we indicate our research limitations and suggest some further works.

Chapter 2

Related Works

2.1 Overview

We begin by reviewing some of our own (i.e. ZESS) prior work on which this student work builds, and then move on to other work that considers specific material sensing problems using machine learning methods.

Conference Paper Our strategy for attaining robust ToF-based material imaging using the spatial correlation of the acquired MIRF was first presented in our conference paper, [13]. This student work extends that work with more discussion on how a broad set of goals can be expressed and achieved using our proposed methods, much more comprehensive experiments, some additional theoretical perspectives, and more advice for the practitioners.

2.2 Material Sensing at ZESS

Conference paper [2], showed that multiple different types of scattering phenomena can be modeled by the MIRF, whose number of Fourier coefficients for each pixel can be increased by broadening the frequency band from 120 MHz [10], to 160 MHz. One can then train a classifier to use features based on these Fourier coefficients to achieve an improved classification accuracy, which is around 94%, as mentioned in [2]. This method played a significant role in improving our classification result by extracting crucial information from the various materials. Conference paper [10] has adopted a voting scheme method based on superpixels to reduce the overall classification time. Additionally, a Support Vector Machine (SVM) with a quadratic kernel was used as a classifier, which performed poorly in the boundary regions. In contrast, in our work, we propose to use a CNN as a classifier for per-pixel material sensing, as suggested in [2].

2.3 Other related Work

The earliest work on material recognition that uses ToF cameras is presented in [14], 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 [4, 12, 15]. Similar to our prior work [10], and its follow-up work [2], [11] has made an attempt to remove the extrinsic parameters by normalizing the Fourier coefficients with respect to the depth and amplitude measurements. They have used the classical machine learning models, which are Radial Basis Function (RBF)-SVM, Nearest Neighbor, Decision Tree, Random Forest, and so on, to classify the materials. Nevertheless, the maximum accuracy they achieved was 80.9%. Conference paper [12] suggests that using a non-parametric classifier (i.e., Nearest Neighbor) or any such classifier with a relatively small number of parameters can be useful in cases where the feature vector is typically high-dimensional while the materials are low-dimensional. They removed the data consisting of specular reflection from the target surface, contributing negatively to the classifier, and attained a maximum accuracy of 90.5% using a non-parametric classifier. Similarly, we have implemented some deep learning models with a limited number of learnable parameters. Moreover, an attempt has been made to examine the thickness of the material by utilizing the ToF measurements. The energy spectra of various materials obtained after processing the raw measurements of a vision camera can exhibit a unique pattern, which can be learned by an artificial neural network classifier, as mentioned in [16]. On homogeneous material, their methodology achieved a maximum accuracy of 95.1%. In [17], materials have been classified using Single-Photon Avalanche Diodes (SPADs), which can time-stamp individual photon arrivals. In [18], 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. Methods in [19] include an RGBD camera that uses surface roughness for material classification. In [20], an optimization method has been used to reduce surface interreflection scattering. Normalization has been used to extract the Reflection Point Spread Function (RPSF). In [21], an Adaptive Neighborhood Convolutional Neural Network (AN-CNN) model (i.e., similar to our spatial window method) is proposed for terrain classification considering multispectral data for homogeneous and boundary regions and a maximum of 87% accuracy has been attained. Differently, we use a modified window prior to the CNN classifier.

Chapter 3

Methodology

Since our work deals with data in multi-dimensions, in this Section, first, we briefly explain the generation of a three-dimensional data cube in Section 3.1 and then, ToF image patch extraction for per-pixel material classification in Section 3.2. Afterward, we introduce a bilateral weight matrix to improve the quality of the ToF three-dimensional image patch in Section 3.3. Then, since we are interested in solving the problem of miss-classification of pixels in the boundary regions, we introduce the KNN method with detail in Section 3.4. In the end, we provide a detailed description of our developed three-dimensional deep learning models that have been used for classification in Section 3.5.

3.1 Generation of 3-D Data Cube

ToF sensors collect unique information about the material of the observed objects in a set of images taken at multiple frequencies. In prior work, [2,10], only the temporal frequency dimension has been exploited, ignoring the correlations of pixels in the spatial domain, which is two-dimensional. In order to leverage these correlations, 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. 3.1). 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 represent the temporal frequency dimension.

Handling imbalanced dataset The classifier fails to classify the minority population in some classification problems. Methodologies, such as [22], have been proposed to deal with this kind of problem. However, to address this issue, we utilized a much simpler approach, which is by

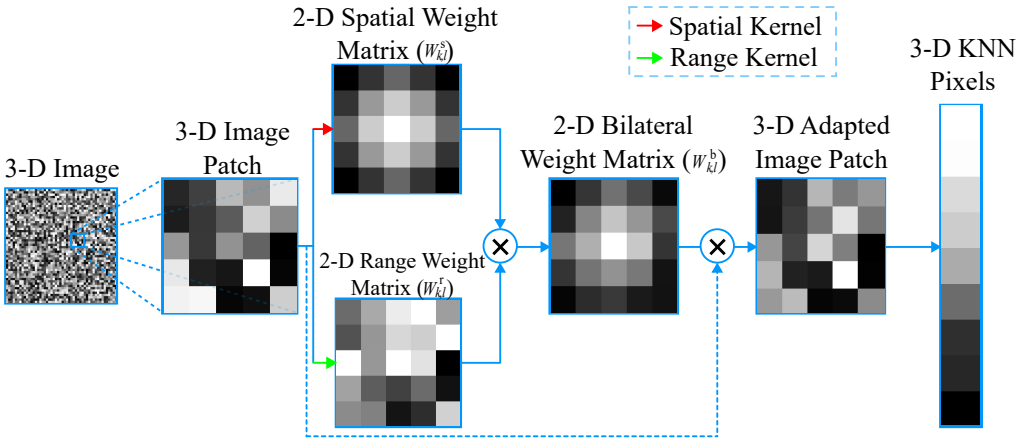


Figure 3.1: An illustration of our proposed methods. The method starts by sliding a window of size $5\times5\times8\times2$ on the 3D image, simultaneously getting modified using a 2D bilateral weight matrix. Finally, congruent pixels have been extracted from the weighted patches using KNN for training and validation.

down-sampling/undersampling each image with a fixed size of the random crop, which involves randomly selecting a portion of an image and using it as a new, smaller image. However, it is crucial to keep in mind that the size of the random crop should be large enough to generalize the entire region. Furthermore, it is important to note that random crop is just one technique that can be used to handle imbalanced data. Other techniques include oversampling (for example, using a method such as bicubic, bilinear, pixel shuffle, etc.) and synthetic data generation. It is often a good idea to try a combination of these techniques to find the one that works best for your particular dataset.

3.2 Three-Dimensional ToF Image Patch Extraction

We stack the $\mathbb{R}$ and $\mathbb{I}$ components of the complex pixels in a fourth dimension featuring two channels (c_{ch}). In the prior state-of-the-art, [10], per-pixel material imaging has been observed to provide relatively low accuracy due to insufficient descriptiveness. To address this problem, considering that pixels belonging to the same material will typically appear grouped in regions, we propose an initial 25-fold increase in the spatial-feature vector 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 the window in the 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. 3.1), where n^{P} and m^{P} represents the two spatial dimensions of the padded image, with a

stride $S = 1$ to generate patches in both spatial and temporal frequency domains for robust material imaging.

$$I^{\mathrm{p}} = \mathrm{pad}(I), \qquad \text{where} : \mathrm{pad}(\cdot) \in \begin{Bmatrix} \mathrm{symmetric}(\cdot) \\ \mathrm{replicate}(\cdot) \\ \mathrm{circular}(\cdot) \end{Bmatrix} \tag{3.1}$$

Choosing $[i_{\mathrm{rows}},\, j_{\mathrm{cols}}] = 5$ and in (3.1) $\mathrm{pad}(\cdot)$ to be either $\mathrm{symmetric}(\cdot)$ or $\mathrm{replicate}(\cdot)$ yielded superior performance.

3.3 Feature Adaptation

As described in [21], a CNN, while predicting the class of the central pixel of the image patch, all the pixels in the image patch will exhibit the same influence on the classification result. However, both the distance in the spatial domain and feature space should influence 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. 3.1). Due to this reason, we apply the bilateral weight matrix $W^{\mathrm{b}}_{k,l}$, (see (3.4)) similar to the weight matrix used in [23, 24], to our ToF image patch (see Fig. 3.1). 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 the coefficients of the weight matrix $W^{\mathrm{b}}_{k,l}$ is:

$$W^{\mathrm{s}}_{k,l} = \exp\left\{-\frac{(k-r)^2 + (l-c)^2}{2\sigma_D^2}\right\} \tag{3.2}$$

$$W^{\mathrm{r}}_{k,l} = \exp\left\{-\frac{||(\vec{I_{k,l}} - \vec{I_{r,c}})||^2}{2\sigma_R^2}\right\} \tag{3.3}$$

$$W^{\mathrm{b}}_{k,l} = W^{\mathrm{s}}_{k,l} \times W^{\mathrm{r}}_{k,l} \tag{3.4}$$

In (3.2) and (3.3), σ_D and σ_R, are the smoothing parameters. In particular, the spatial weight matrix $W^{\mathrm{s}}_{k,l}$, (see (3.2)) should favor close neighborhoods of pixels, whereas the range weight matrix, $W^{\mathrm{r}}_{k,l}$, (see (3.3)) should suppress the pixels belonging to a class other than that of the central pixel in the image patch.

3.4 KNN pixel refinement

In prior work [2], 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 between the Fourier-feature vectors with respect to the center pixel within the image patch

Table 3.1: 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%

using Euclidean distance metric [25, 26] and concatenation as depicted in Fig. 3.1. In this way, we identify the KNN of the central pixel in the feature space. Our analysis has shown that choosing $KNN = 9$, which means an increase of 9-fold of the feature vector size, as compared to previous work [2, 10], has yielded the highest classification accuracy (see Fig. 4.1 and 4.3).

3.5 Deep learning model development

To date, most ToF material imaging has been carried out using conventional machine learning algorithms, such as SVM, Fully Connected Neural Network (FCNN), along with others, in [2, 10, 20]. 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 additional layers (see Table 3.1). We have developed a small 3-D CNN (see Tab. 3.1) consisting of 25 layers, which allowed reducing the total number of learnable parameters to 41557, without losing performance. Moreover, we have constructed a 3-D ResNet (see Tab. 3.1) consisting of 29 layers, which helps in avoiding the exploding and vanishing gradients during training but did not yield any performance gain. An adam optimizer with an initial learning rate of 0.08 has yielded the best result for training our models. It should be noted, that examining all possible sets of hyper-parameters is an NP-hard problem.

Chapter 4

Experiments

4.1 Introduction to Datasets

We have evaluated our proposed methods on three datasets (see Table 3.1), which have been already used for processing and analysis using state-of-the-art ToF material imaging [2, 10]. The 12M (i.e., 12 Materials from the 15 Materials (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 [27], which consists of 5M, each taken at 10 different distances and 7 different orientations. Moreover, each image represents a material with a spatial resolution of 176×224 pixels. After processing the 15M dataset with our proposed methods, the total number of input data elements available for training and validation is 70778880. Whereas for 14M dataset, the total number of input parameters is 17547264, and for 5M dataset it is 154460160. For all datasets, the ratio between training and validation data is kept at $70 : 30$.

4.2 Results

Due to each image target consisting of homogeneous material for the above-mentioned datasets, we have simulated an artificial heterogeneous test target to evaluate our proposed methods. From Fig. 4.1a, 4.1c, we can see that cardboard, coated metal, and wood from the 12M dataset are getting confused with each other. Nevertheless, the proposed method is able to attain a maximum of 95% accuracy (see Table 3.1). From the 14M dataset, Fig. 4.1d, 4.1f depict that corrugated cardboard, felt, paper sheets, plaster, polypropylene, and wax are achieving 100% classification accuracy, while the overall maximum accuracy is 99% (see Table 3.1). The maximum accuracy of 98% obtained for the 5M dataset (see Fig. 4.1g, 4.1i and Table

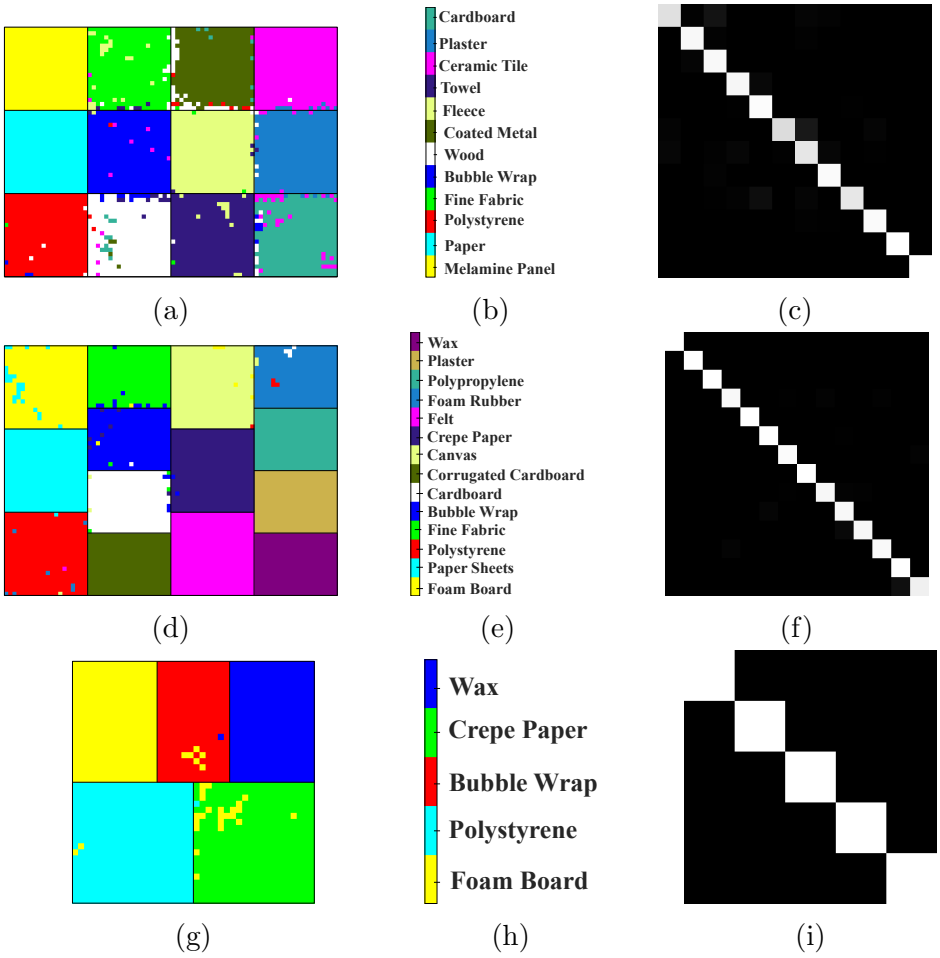


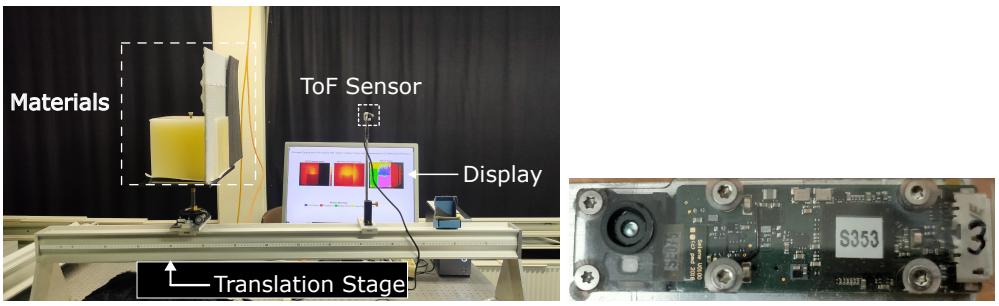
Figure 4.1: Per-pixel material imaging of artificial heterogeneous targets using our proposed methods on three datasets: (a), (d), and (g) show the results of the 3-D CNN on the 12M (b), 14M (e), and 5M (h) datasets, respectively, while (c), (f), and (i) depict the corresponding confusion charts (known 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.

3.1) has proved that our methods perform well on a dataset exhibiting high variations in terms of distances and orientations, which helps in a better generalization. One of the reasons for the 14M dataset to perform the best could be a very limited number of data samples to learn from (i.e., 4.1176 fold lesser compared to 15M dataset), which contributes in faster learning of the network. However, it is important to note that a limited number of samples typically lacks generalization, resulting in limited applicability in real-world applications. Finally, as we can see that the performance of 12M and 5M datasets is comparatively lower than the 14M dataset, we attempted

to close this gap by means of transfer learning [28], since it performs well for problems belonging to similar categories. We used the trained model from the 14M dataset on the 15M and 5M datasets. In our case, however, empirically we found that this method does not provide any significant improvement.

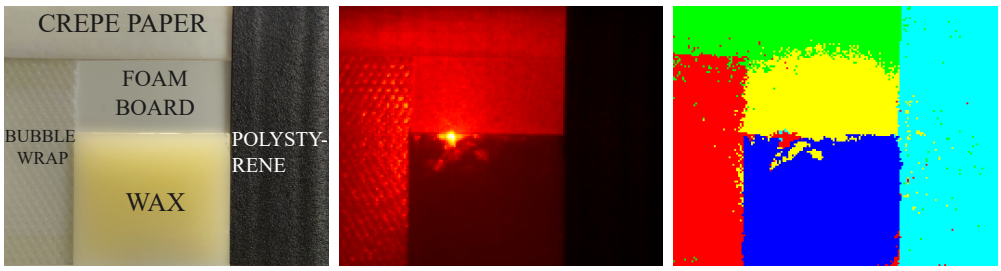
4.3 Demonstrator Setup and Result

Finally, to evaluate our proposed methods on a real heterogeneous target [29], we have constructed a demonstrator setup [30], (see Fig. 4.3), consisting of five materials. Each image has been taken at 5 different distances (uniformly distributed between 82 cm to 47 cm) and at 3 different orientations (uniformly distributed between $-10°$ to $10°$) (see Fig. 4.3b), using the PMD Selene module, which has been used to carry out various experiments that have been mentioned in [7,31]. The PMD Selene module is manufactured by *pmdtechnologies ag*. The module is having a large bandwidth of 160 MHz and features a spatial resolution of 176×224 pixels. The translation stage depicted in Fig. 4.2a has been used to collect multiple images at different locations, by providing a to-and-fro motion to either the target objects or our ToF camera. The base supporting the target objects in Fig. 4.2a, can be rotated with respect to the vertical axis, in order to acquire data from multiple orientations. Our developed algorithm takes a few milliseconds to capture a single image. Based on the performance of our models on artificial heterogeneous targets, we adopt the 3-D CNN classifier that has been mentioned in Section 3.5 and attained 99% validation accuracy (see Fig. 4.3c). Despite attaining a very high test accuracy on the 5M dataset, it can be seen from fig. 4.1g and 4.3c that crepe paper and foam board are getting confused with each other more compared to other materials. One possible reason could be the thickness of the crepe paper that has been used while acquiring the 5M dataset, which is approximately less than 1 mm.



(a) (b)

Figure 4.2: Our experimental setup (a) and the PMD Selene module (b).



(a) (b) (c)

Figure 4.3: The demonstrator setup (a) consists of five different materials (see Fig. 4.1h) placed at approximately the same distance to the ToF sensor. (b): NIR amplitude image, (c): material image obtained using the proposed approach.

Chapter 5

Conclusion

5.1 Summary and Discussion

In this student work, 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. We noticed that among all the materials, our model has always performed best on "wax", which is similar to the result mentioned in [11]. Similarly, the "crepe paper"/"paper" has performed the worst. Furthermore, because the image upsampling methods (i.e., bicubic and bilinear) performed poorly on the test set, we should avoid it when dealing with the class-imbalanced dataset. The results of this research are expected to constitute a step forward towards robust material imaging, unleashing new application domains for ToF cameras.

5.2 Research Limitation

As previously stated, our method achieves the research objectives. Thus, using our proposed methods, one can guarantee that the material detection will be at least 95% accurate. Nonetheless, this study has some limitations:

Computation Time: When compared to previous work, the use of spatial neighborhoods for a per-pixel ToF-based material classification consumes a significant amount of computation time. It will be difficult for a real-world application, especially when the image resolution is high. We fixed this problem using parallel computing for our live demonstration.

Limited Range: We have several ideas to generalize the work by solving the limited-range problem. The material imaging range is currently limited from 47 cm to 87 cm, creating a gap in a real-world deployment. This could be solved by acquiring the dataset over a much wider range of distances and orientations. Furthermore, with adequate resources, experiments could have been carried out on a much larger number of materials.

5.3 Future Work

In this research, we introduced a new method to boost the quality of the ToF measurements to classify different materials. We present experimental results on three benchmark datasets for the multi-class classification problem, where all of them are class-imbalanced datasets, and tested one of the models on a live demonstrator, which is based on our developed 3-D CNN model. As a result, for future work, we can use cascade models to generalize our methods to a much larger number of materials. Although we focused on classification tasks in the experiments due to time constraints, it is worth noting that regression tasks can be performed using the proposed methods to quantify intrinsic parameters of the material by training the model on appropriate datasets. The other improvement to this research would be to balance the datasets using different methods. Moreover, an attempt can be made on adaptive KNN, such that sudden changes in temporal frequency-feature space can be automatically detected. This might result in a better solution for real-world applications.

Bibliography

[1] S. Joshi, "What is spoofing? How to protect yourself against it." `https://www.g2.com/articles/spoofing`, 2022.

[2] M. Heredia Conde, T. Kerstein, B. Buxbaum, and O. Loffeld, "Near-Infrared, Depth, Material: Towards a Trimodal Time-of-Flight Camera," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 11271–11279, 2022.

[3] X. Ma, H. Chen, and Y. Zhao, "Stereo image coding method using stereo matching with difference-based adaptive searching windows," in *2009 IEEE International Workshop on Imaging Systems and Techniques*, pp. 373–376, 2009.

[4] 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.

[5] A. Medina, F. Gayá, and F. Del Pozo, "Compact laser radar and three-dimensional camera," *Journal of the Optical Society of America A*, vol. 23, pp. 800–805, Apr. 2006.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] 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.

[13] R. U. Singh and M. Heredia Conde, "Robust Time-of-Flight-based Material Imaging using Three-Dimensional Deep Neural Networks on Spatial Neighborhoods of Pixels," in *2022 IEEE SENSORS*, pp. 1–4, 2022. to appear.

[14] 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.

[15] 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.

[16] M. I. Saripan, W. H. Mohd Saad, S. Hashim, A. T. A. Rahman, K. Wells, and D. A. Bradley, "Analysis of photon scattering trends for material classification using artificial neural network models," *IEEE Transactions on Nuclear Science*, vol. 60, no. 2, pp. 515–519, 2013.

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] A. F. Cruz, C. Belém, J. Bravo, P. Saleiro, and P. Bizarro, "Fairgbm: Gradient boosting with fairness constraints," 2022.

[23] 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.

[24] 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.

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

[26] 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.

[27] 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.

[28] S. Bozinovski, "Reminder of the first paper on transfer learning in neural networks, 1976," *Informatica*, vol. 44, 09 2020.

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

[30] 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.

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

Statutory declaration

I declare that I have authored this studienarbeit independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material that has been quoted either literally or by content from the used sources.

Siegen, January 2, 2023

Rajababu Udainarayan Singh