A Framework for Thermographic Material Characterization Using Multi-Channel Neural Network

Tamas Aujeszky, Georgios Korres, Mohamad Eid

***Abstract*—Research on material characterization has received an increasing amount of attention recently. In several application scenarios, it is essential to effectively estimate the physical properties of objects without coming into contact with them (e.g. Tele-operated or autonomous robotics). This paper presents the Haptic Eye: a framework using active thermography that uses a custom multi-channel neural network approach to perform classification between samples and regression towards their thermal properties. This neural network structure is uniquely suited for effective processing of thermographic data. The framework is realized, implemented, and evaluated with a set of 10 samples with diverse thermal/physical properties. Experimental results on a realization of the framework validate this approach, with a 92.20% classification accuracy using multi-channel neural network with majority vote, as well as more than 99.6% R^2-fit with respect to three different thermal properties, namely thermal conductivity, thermal diffusivity, and thermal effusivity (which is very useful to define thermal exchange during physical interaction).**

***Index Terms*—Active thermography, thermal properties, material characterization, machine learning.**

I. Introduction

Over the past decade we have seen tremendous progress in fields such as robot autonomy, tele-operation, 3-D computer vision, human-computer interaction and virtual reality. These advancements are in part driven by increases in sensing capabilities of three-dimensional environments [1]. Accurate mapping of these scenes allows for precise manipulation of objects in remote areas, error-free navigation of unknown environments, interacting with humans in a harmless and effective way, and even capturing or recording these scenes to be used in virtual models or recreated in reality.

While current audiovisual sensing technologies offer fast and detailed audio and visual data acquisition, the physical sensing capability is lagging far behind [2]. Tele-operated and autonomous robots are typically equipped with high resolution cameras, whose images are processed by well-established computer vision algorithms to learn the visual properties of the manipulated scene (such as color, geometry, motion and texture). However, these systems have very limited knowledge about the material properties of the manipulated scene (such as stiffness or friction properties) [3]. The acquisition, storage, communication and display of the physical properties improves the quality of performance [4] or immersion [5] in the remote environment. For instance, a system able to provide feedback to surgeons in the form of real-time quantitative measurements of soft tissue properties, force and stiffness, is still missing and highly desirable [6].

Recording the physical properties of an unknown scene is based on three approaches: vision-based, contact-based, and thermography-based [7]. Existing audiovisual methods for physical properties characterization are limited in application and accuracy due to the fact that the visual properties are not fundamentally related to the underlying physical properties of an object [8]. On the other hand, contact-based approach is both cumbersome and time-consuming, and may not be practical for most applications [9]. Finally, thermography-based approach is currently focused on qualitative evaluation of material defects [10].

Therefore, the Haptic Eye framework is proposed to provide quantitative physical properties characterization based on active thermography. The proposed system relies on active infrared thermography to capture the thermal response of the object under examination. This signal is processed and fed into a custom multi-channel variation of a convolutional neural network that allows for two different modes of operation: classification between families of materials or regression to thermal properties. This unique neural network structure is introduced in order to take maximum advantage of the characteristics of the thermographic recordings which serve as the raw data input. The output of the neural network can serve as the basis for estimation of the physical properties of the object.

The contributions of this paper include the followings:

- The conceptualization of the Haptic Eye framework for contactless material characterization.
- The realization, experimental evaluation, and the optimization of the proposed framework for material classification based on a modified convolutional neural network architecture.
- The realization, experimental evaluation, and the optimization of the proposed framework for material characterization using a regression-based algorithm for a list of three thermal properties: thermal conductivity, thermal diffusivity, and thermal effusivity.

The rest of this paper is organized as follows: Section II gives an overview of the existing approaches for material characterization. Section III contains a conceptual presentation of our approach, including a description of the Haptic Eye Framework. A realization of this framework is described in section IV, where experiments are presented for both classification and regression. Results are presented in section V,

and these are discussed in section VI. Finally, conclusions are drawn in section VII, along with future directions for research.

II. Related Work

Three approaches for material characterization are explored in the literature: audiovisual-based classification and machine learning [11], [12], thermography [13], [14], and contact-based material classification [15], [16].

With the advent of machine learning, a set of new approaches utilized audiovisual input for material characterization. Degol et al. [17] used images to estimate 3-D point clouds in order to recognize material categories. The authors in [18] and [19] used deep learning for image segmentation and material recognition in the wild. [12] presents a method to predict surface friction of a diverse set of surfaces in order to improve the movement of a biped robot. Other researchers combined different modalities, such as auditory, visual and tactile/kinesthetic in order to improve the quality of material characterization [20] [21]. However, this approach is fundamentally limited by the fact that the sensing element captures only visual properties that may not necessarily well correlate with the physical properties.

Another line of research uses the properties of physical interaction between a pen-shaped tool and the material for material characterization [22]. The tool is equipped with different sensors that provide signals which can be related to the stiffness or texture of the object. Depending on the type of sensors, the tool can acquire acceleration, force and torque data, for example. An example to this approach is [22], in which the authors perform classification using acceleration data acquired as a response to vibration patterns introduced into the ground to categorize haptic textures. In [23], the authors propose detecting the hardness of materials using a multi-functional tactile sensor composed of piezoelectric transducers designed to imitate a human finger. The authors of [24] propose a so-called intelligent prodder, which measures the elastic response of samples by providing a mechanical stimulation to them, and relates this information to the physical characteristics of the material.

Although thermography has been used for material characterization in the past, existing research [25], [26] is focused on using a closed form solution to the heat equation that is only applicable under assumptions for the excitation that it is instantaneous in time and has a Gaussian spatial distribution on the surface of the object. Another commonality is that the used setups are fundamentally unsuitable to be used for scanning objects in real scenes, and are therefore unable to form a basis for a realization of the Haptic Eye framework. In our previous work we have proposed and validated a novel mathematical model for infrared thermography with realistic conditions to be used for material classification [7]. In a previous work [27], we investigated realizing the Haptic Eye framework using classical machine learning methods (including including Support Vector Machine, K-Nearest Neighbors and Decision Tree). This paper demonstrates an evolution of this approach that uses a custom designed multi-channel neural network, advanced data processing, and the ability to perform regression in addition to classification, all performed on a larger sample set.

III. Contactless Material Characterization Approach

This section details the approach behind the Haptic Eye system, including the framework of the thermography system, the considerations behind the design of the neural networks used in this approach, as well as creating ground-truth data for training, testing and evaluating the neural network.

A. Haptic Eye Framework

Figure 1 shows a schematic diagram for the Haptic Eye framework. It comprises the interaction module, the excitation module, the acquisition module and the material characterization module. The interaction module is responsible for launching the characterization process and using the gained insight to control the actuator component. The excitation module is responsible for performing the required excitation at the proper time and location, with proper waveform and duration. The acquisition module consists of a set of sensors that record the scene containing the object under examination, with a schedule that is synchronized with the excitation module. The data processing module handles the acquired data by applying a series of data processing steps and feeding the processed data forward to the material characterization module. Finally, the material characterization module takes the processed data and estimates the material properties of the examined object. Each of these modules can be further broken down into components. Some of these, such as the laser source, are essential in any realizations of the framework. Others, such as an RGB camera, are optional but have the potential to make the system more robust or enable additional functionalities.

B. Multi-Channel Neural Network Design

There are several factors that were taken into consideration prior to designing the neural networks for classification and regression. One such factor was the amount of data available to be captured. The complexity of neural networks, measured by the amount of trainable parameters they have, can grow very large depending on the number of layers and hidden units they contain. Our goal was to keep the number of parameters low enough that we can capture enough data points so that they are at least an order of magnitude more than the number of parameters. Current state-of-the-art convolutional neural networks typically have tens of millions of parameters [28], and are trained with similarly sizable data sets. This is done to eliminate the potential risk of overfitting resulting from using a highly generalizable model on a limited data set [29].

The source of information for this research is the sequence of radiometric or thermal images (data frames) captured by the thermal camera, but using each sequence as a single input would have necessitated the recording of tens of thousands of sequences even for a rather shallow neural network design. Our idea was to design a neural network that is capable of using an individual frame as input and performing classification

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2978572, IEEE Transactions on Instrumentation and Measurement

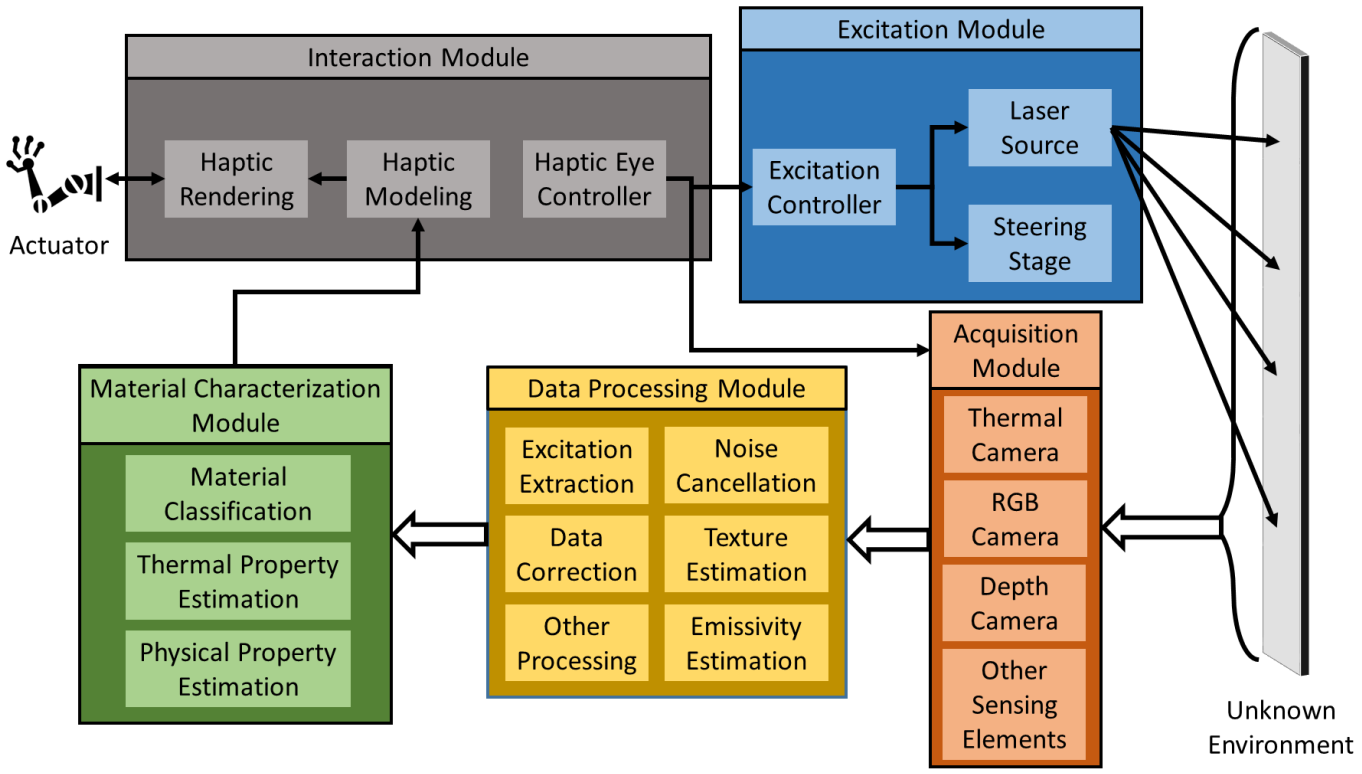

Fig. 1. The Haptic Eye framework.

or regression based on that one frame alone. Capturing a cooldown process then corresponds to acquiring a multitude of input data points, each corresponding to a single frame. Therefore our network is based on a convolutional neural network architecture, with a low number of layers to ensure an upper bound on the number of trainable parameters. It consists of two convolutional layers, each using a 3x3 kernel and each is immediately followed by a max-pooling (2x2) layer and a batch normalization layer. A diagram of the convolutional and fully connected layers is displayed in Figure 2. The number of input image channels is 1, whereas the outputs of the convolutional layers have 3 image channels. The output of the convolutional stage is flattened and fed into a fully connected stage composed of a 16-node fully connected layer and an output layer. For classification, this layer is followed by an output layer of size 10 (using soft-max activation), corresponding to the number of material samples (classes). For regression, the output layer is a single scalar, indicating the target (thermal property) value. All convolutional and dense layers except for the output layers use sigmoid activation.

The challenge with this approach is that the cooldown image captured 0.1 s after the end of the excitation period is different from the image captured 0.2 s later in the same sequence. Therefore the network needs to account for this difference and still map these two different images to the same material sample (classification) or thermal property value (regression). The solution to this challenge is to record an additional numerical value for each frame, which we call the time stamp of the frame. This indicates the amount of time that has passed between the end of the excitation period and the time that the frame was taken. The time stamp feature needs to be supplied to the neural network as an auxiliary input a second channel of information), making it a multi-channel neural network. This is accomplished by feeding it in at the end of the convolutional stage. Once the output of the convolutional stage is flattened, it is concatenated with the time stamp scalar. This extension of the regular convolutional neural network is what makes our multi-channel network uniquely suited for the task at hand: single frame inference ability from a processed thermographic recording and additional temporal data through the auxiliary input. Given that the evolution of the temperature distribution on the surface of an object is a continuous process, the time stamp value is a useful addition that allows the network to learn a robust representation of the cooldown process by evaluating single recorded frames relatively to each other based on how long after the end of the excitation period they were captured. In order to avoid problems arising from the different numerical scale of convolutional outputs and the time stamp, the entire feature vector undergoes batch normalization before being fed into the fully connected stage.

The total number of trainable parameters (the measure of computational complexity) of the current realization of the neural network framework stage is 5577 for regression and 5730 for classification. This number measures the degrees of freedom related to the structure of the neural network, being equal to the sum of the number of weights and biases, the values of which are set during the training process. Since the auxiliary input channel is only responsible for 16 added parameters (less than 0.3% of the total) under the presented network structure, and given that its addition does not increase

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2978572, IEEE Transactions on Instrumentation and Measurement

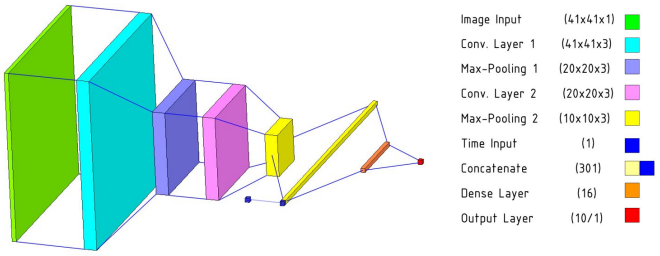

Fig. 2. The architecture of the multi-channel neural network, including the auxiliary time input. Additional branch normalization layers are applied after all layers. The network output is a vector of length 10 for classification and a scalar for regression.

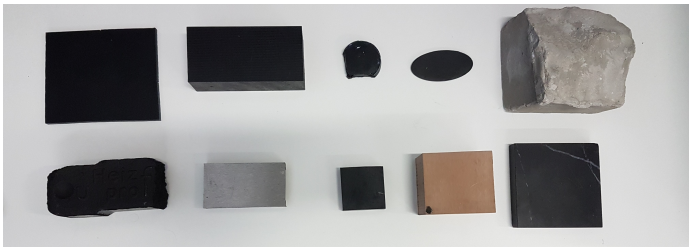

Fig. 3. The experimental samples. Top row, left to right: acrylic glass, machining polyethylene, silicone, sorbothane and concrete. Bottom row, left to right: coal, steel, high-pressure laminate, low-pressure laminate, black marble

the depth of the network, it can be stated with confidence that its computational complexity is only negligibly higher than that of an equivalently sized CNN. As indicated before, this is orders of magnitudes lower than state-of-the-art convolutional networks for image recognition (typically measured in millions of trainable parameters [28]). This means using this network requires much less computational power (allowing for a wider range of hardware to be deployed onto), and its training and inference processes are quicker as well.

C. Sample Material Selection

The literature about material science reports four families of physical material: polymers, ceramics, metals, and composites [30]. A sample set of 10 materials is designed with three selection criteria: (1) samples must represent the four families of material, (2) samples entail a large range of thermal properties (thermal conductivity, diffusivity, and effusivity ranges), and (3) samples are highly available in every day's life. Based on these criteria, five polymer samples are selected, namely silicone, acrylic glass, sorbothane, polyethylene, and coal (as a polymer composite). Similarly, two composite material samples are considered, namely low pressure laminate and high pressure laminate. Concrete and marble (as ceramic composites) are going to represent the ceramic family. Finally, steel is included as a widely available metal. A snapshot of these samples is shown in Figure 3.

D. Synthetic target data generation

The classification task requires labels for each frame, which can be easily provided by encoding each of the samples into a one-hot coded vector and providing the appropriate sample vector for each frame input. On the other hand, the regression task requires a numerical value to be provided for each frame to serve as the target data. This scalar data corresponds to the thermal property that the system will be trained to predict based on the data frame (and the time stamp). The challenge in providing this data is that it is not possible to carry out a direct measurement of this thermal property at the same time as each data frame is recorded. The way to overcome this challenge is to acquire thermal data separately and use it as the ground truth for generating the target values. To that effect, a two-step approach was designed: in the first step, a series of thermal measurements are taken of each sample on the same surface region where the thermal excitation takes place during the thermography process. These thermal measurements were conducted by the TCi Thermal Conductivity Analyzer, and returned a series of thermal conductivity and thermal effusivity values. The TCi device performs contact-based heating on the surface of the object and performs parametric model fitting on the measured thermal response. The calculated conductivity and effusivity values correspond to the numeric values of the parameters in the best-fit models. Based on a pair of these values, thermal diffusivity (the measure at the heart of the heat equation) can also be calculated using the following formula (where α represents thermal diffusivity, k stands for thermal conductivity and e is thermal effusivity):

$$\alpha = \frac{k^2}{e^2} \tag{1}$$

A total of 3 measurement sessions, each comprising of 25 measurements have been performed on each material sample. The coefficient of variation (the ratio of standard deviation to mean) for each sample was found to be no more than 2.3% for effusivity, 4% for conductivity and 4.4% for diffusivity. The results are summarized in Table I.

The second step in the synthetic data generation process was to generate target data values for each data frame. Given the small but nonzero variance of the measurement results, the approach to provide a single target data for each frame of a certain sample was dropped in favor of an alternative approach that captured this variance. To that effect, for each material sample and each material property, a Gaussian random variable was generated, having the same mean and variance as the measurements of that thermal property on the related material sample. For each data frame, a separate outcome of the corresponding random variables is assigned as a triplet of target values for thermal diffusivity, conductivity and effusivity.

IV. Framework Realization

The following section presents a realization of the Haptic Eye Framework that was created with the aim to validate our approach (Figure 4). Experiments using this system are carried out for both classification between samples and regression towards thermal properties.

A. Experimental Setup

In order to acquire the data frames, 10 data acquisition sessions of an active laser thermography experiment were

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2978572, IEEE Transactions on Instrumentation and Measurement

TABLE I
MEAN AND STANDARD DEVIATION VALUES PER SAMPLE USED FOR SYNTHETIC TARGET DATA GENERATION (SORBOTHANE HAS A SHORE OO DUROMETER VALUE OF 70).

Sample	Conductivity $(W/m*K)$		Diffusivity mm^2/s		Effusivity $W*\sqrt{s}/(m^2*K)$	
	Mean	SDev	Mean	SDev	Mean	SDev
Concrete	0.565	0.01087	0.370	0.00681	927.7	10.37
Polyethylene	0.539	0.00362	0.356	0.00193	902.4	3.61
Acrylic glass	0.231	0.00704	0.161	0.00707	575.2	5.68
Coal	0.409	0.00349	0.283	0.00334	768.9	3.40
Marble	3.312	0.12961	1.610	0.05320	2609.6	59.34
Sorbothane	0.351	0.01352	0.246	0.00907	707.9	14.48
Low-pressure laminate	0.208	0.00427	0.144	0.00528	546.9	1.75
High-pressure laminate	0.382	0.00427	0.265	0.00502	742.5	2.05
Steel	34.672	0.47857	9.668	0.14052	11150.5	72.83
Silicone	0.230	0.00554	0.161	0.00515	573.0	6.07

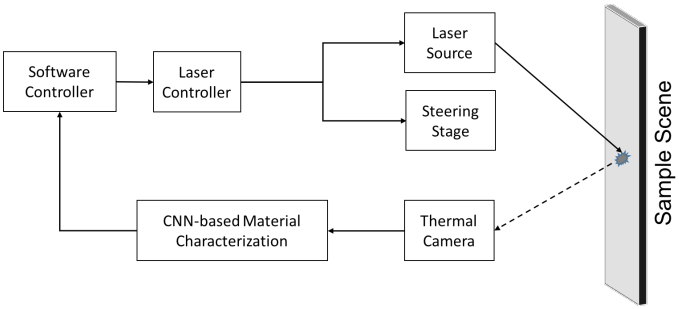

Fig. 4. Realization of the framework used for the experimental validation.

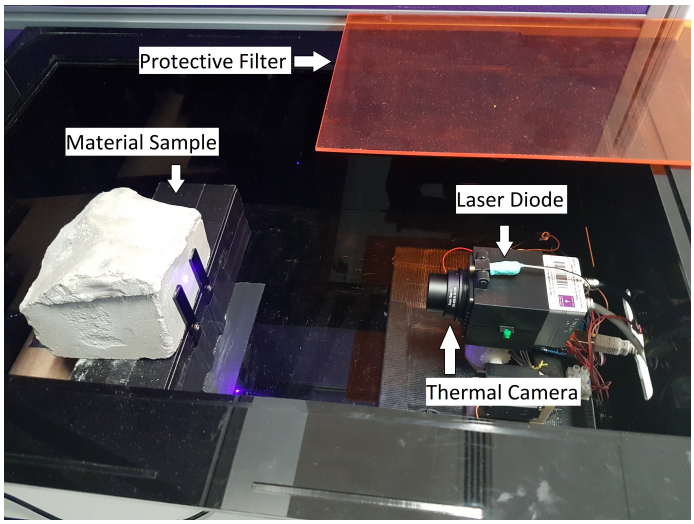

Fig. 5. The experimental setup.

conducted. The excitation is provided by a 405 nm laser diode (US-Lasers Inc. D405-120), operating at 120 mW of power. The process is recorded by a Xenics Gobi-640-GigE camera operating in radiometric mode, which produces a 16-bit grayscale image with VGA (640x480) resolution and 50 Hz frame rate. The laser is controlled by a relay operated by an Arduino board, which is instructed through a serial connection. The camera is instructed through a Gigabit Ethernet connection and it sends the captured frames through the same. The distance from the sample to the camera is 195 mm and to the laser diode is 225 mm. Completing the experimental setup is a desktop computer running the MATLAB R2018a software environment. A MATLAB script is responsible for controlling the laser diode and the camera, as well as processing and saving the acquired data. An image of the experimental setup is shown in Figure 5.

B. Data Acquisition

The entire process of acquiring the data set used to train the neural networks is divided into experimental sessions, which are further divided into experimental rounds. A single experimental round consists of the following steps: first, without the presence of any excitation, a series of 40 'ambient' frames is captured. In the next step, the laser diode is turned on for a period of 5 seconds. As soon as the laser diode is turned off, a series of 100 'cooldown' frames are captured. Given the 50 Hz frame rate of the camera, this procedure captures the first 2 seconds after the excitation. Lastly, a minimum of 2 minutes is observed during which no action is taken to ensure that the sample can return to its original thermal state and that subsequent measurements are not affected by previous ones.

The above described procedure for an experimental round is repeated 10 times for the same sample, after which it is repeated 10 times for each of the remaining samples. This set of 100 experimental rounds makes up an experimental session, which is conducted in its entirety over a single day. A total of 10 experimental sessions were conducted over a period of 10 days.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2978572, IEEE Transactions on Instrumentation and Measurement

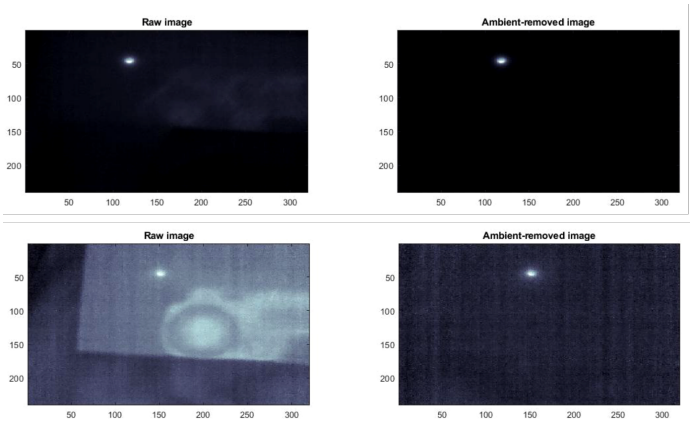

Fig. 6. Ambient noise removal example for Polyethylene and Marble (before and after ambient noise removal).

C. Data Processing

Processing the acquired data consists of a number of steps that are taken separately for each experimental round. First, both the ambient and cooldown frames are cropped to a fixed 320x240 region that contains the excitation area and its close surroundings. This ensures that even when a sample is relatively small compared to the field of view of the camera, the processing will not be affected by any parts of the background. Secondly the 40 ambient frames are averaged over time into a single average ambient frame of the same resolution. The pixel values of this average ambient frame are then subtracted separately from every cooldown frame. This results in a set of 100 'ambient-removed' data frames, where non-transient thermal differences are all removed. This improves the image contrast for processing and enables the system to potentially work in environments where there is a small pre-existing thermal variation on the sample surface in steady state (an example is shown in Figure 6). The last step in the process is to localize the excitation center. This is done by taking the sequence of the first 10 ambient-removed frames, applying a 3D moving average smoothing on them and finding the pixel with the highest value. The 41x41 neighborhood centered around the location of this pixel is extracted from all ambient frames, and serves as the input frames to the neural network. This process is repeated separately for each experimental round.

D. Neural Network Training

Performing accurate single-frame classification or regression dictated the overall structure of the multi-channel neural network, and the size of our data set placed a limit on the total number of trainable parameters. As stated in Section III-B, this number is 5730 for classification and 5577 for regression. In addition to decisions about the dimensions of the network, there were also decisions that we could make for certain hyperparameters for the network and the training process based on how well they performed with the acquired data. These included the activation function (sigmoid or ReLU), the number of hidden layers in the fully connected layer, the training batch size, the number of training epochs, the choice of optimizer and the learning rate. The large number of factors and options precluded us from executing a complete sweep of all combinations, so we arrived at our selection based on a number of trial runs. This includes sigmoid activations after each convolutional and dense layer (except for softmax activation at the end of the classification output layer), 16 hidden units in the fully connected layer, a batch size of 2048 and the Adam optimizer with an initial learning rate of 0.01. Additional controlling of the training process was achieved with the use of callback functions.

The data acquisition process yielded a total of 100,000 data frames, each serving as a data point with a dimension of 41x41 (pixels) + 1 (time stamp) = 1682. This data set is partitioned into 3 sets. The training set contains 80% of the data, the validation set contains 10%, and the remaining 10% makes up the testing set. The network is trained on the training data set, and the validation data set is used as input for callback functions. The "reduce learning rate on plateau" callback reduces the learning rate to 40% of its previous value after every 6 consecutive training epochs without improvement in the validation accuracy. The "early stopping" callback stops the training process if the validation accuracy does not increase for 50 consecutive epochs, or if the maximum number of 1000 training epochs is reached (in practice the training stops around 120 epochs). Finally, the "restore best weights" callback ensures that the model weights corresponding to the highest validation accuracy are restored. When the training is finished, the network is evaluated on the testing data set. For classification, the evaluation metric is the classification accuracy on the testing data set, and for regression it is the R^2 value for the predicted thermal values compared to the actual ones in the testing data set.

This partition of data points is done according to the experimental sessions they were recorded in. This means that the model is trained on 8 sessions at a given time, validated on a separate 9th session and tested on the remaining 10th session. There are a total of 90 different partitions for training, validation and testing sessions. We have repeated the entire training process for each of these 90 partitions and the average of these results is shown in Section 5.

This session-based partition serves three purposes. Firstly, since each session contains an equal amount of frames for each material sample, it is ensured that each sample is represented in the same proportion in the training, the validation, and the testing data sets. Secondly, it is imperative that data points from a given experimental round are entirely in the training data set, entirely in the validation set or entirely in the testing data set, otherwise the model could exploit similarities between subsequent frames without learning meaningful features. Since data points of the same experimental round are also of the same experimental session, this condition is satisfied with this partition. Lastly, using entirely separate sessions as the validation and training data sets is more indicative of a device that is pre-trained on data captured at a different time, and used for examining new data (with a potentially slightly different set of loosely controlled environmental parameter values such as atmospheric temperature, humidity, etc.).

The neural network models (one model for classification,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2978572, IEEE Transactions on Instrumentation and Measurement

TABLE II
CLASSIFICATION AND REGRESSION RESULTS FOR THE TESTING SET

Task	Metric	Single-Frame Result	Majority-Vote Result
Classification	Accuracy	89.98%	92.20%
Regression	Conductivity R^2	99.17%	99.77%
	Diffusivity R^2	99.12%	99.75%
	Effusivity R^2	99.16%	99.67%

three models for regression with respect to the three thermal property values) are implemented in Python, using the Keras open source neural network library with TensorFlow as a back-end [31]. These models are trained using the Adam optimizer, with an initial learning rate of 0.01. The loss function is specified as categorical cross-entropy for classification, and the mean squared error for regression. The target values for regression (the thermal property values) are normalized for each partition before training the model. Specifically, the training data set is transformed to one with zero mean and unit standard deviation, and the original mean and standard deviation of the training data set are also used to normalize the validation and testing data sets.

This procedure generates predictions (class labels for classification, estimates of thermal conductivity/diffusivity/effusivity for regression) for each input, meaning a combination of a single processed frame and a time stamp value. Further to this default approach, we have also decided to implement another inference method that we call the majority-vote approach. In this procedure, we merge all the predictions for individual frames in a certain recording in the testing data set into a single combined prediction. This is valid based on the assumption that a single recording captures the same material sample with all of its frames. In practice, a combined prediction for classification is the mode of the predicted class labels by the individual frames. For the regression case, the combined prediction is the mean of the predictions for the corresponding thermal property value by the individual frames. In this latter case the fit is evaluated compared to the mean of the individual assigned target values generated in Section III-D.

V. RESULTS

A. Classification Results

Table II lists the classification accuracy and regression R^2 results. The average accuracy for all partitions is 89.98% for single-frame classification and 92.20% for the majority-vote approach. Figure 7 displays the confusion matrix that results from averaging the predictions over the testing data sets of all 90 partitions.

Comparison results with the state-of-the-art techniques show significant improvement for the proposed approach (shown in Table III. The classification accuracy of our model matches that of the best-performing Decision Tree classifier, even though this model needs to accurately classify samples belonging to twice the number of classes as in the previous case. This model also takes advantage of a much larger data set, which is important for two reasons: Firstly, the data in this study is coming from a more diverse set of sessions

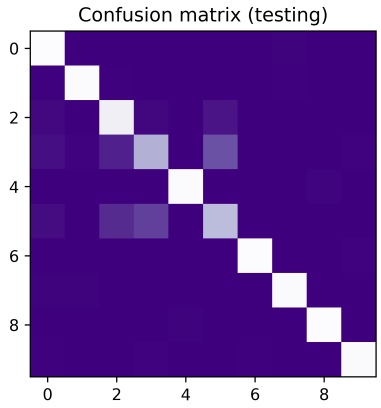

Fig. 7. Average confusion matrix for classification on the testing data set in all 90 partitions.

acquired over a longer period of time. This is more promising when taking into account slight environmental variations in physical parameters such as humidity or ambient temperature. Secondly, both the inference time and the complexity of a decision tree classifier increase with the size of the training set, whereas it is independent of that in the case of a neural network. In addition to this, we ran a separate baseline comparison test with a standard convolutional neural network that lacks the auxiliary input, but is otherwise identical. Results demonstrated that the multi-channel network achieved a larger than 5% improvement, which is quite significant.

B. Regression Results

The results for regression, including root-mean-square error and coefficient of determination (R^2) values can also be found in Table II. The average R^2 values for thermal conductivity, thermal diffusivity and thermal effusivity are 99.17%, 99.12% and 99.16%, respectively. Using the majority-vote approach, these numbers are further improved to 99.77%, 99.75% and 99.67%, respectively. Figures 8, 9 and 10 display the predicted thermal conductivity, diffusivity and effusivity values of all data points. This is done by evaluating the predicted value of each data point during all partitions where that point is part of the testing data set, and combining these results into a single plot. In order to effectively emphasize the concentration of such a large amount of points, each point was assigned an opacity value of 3% on the plots.

In order to demonstrate the significance of these results, we have designed two additional regression methods for comparison. The first method is a standard linear regression that takes all pixel values of the thermal frame input as separate predictors and fits a linear model on them. The second method is identical to our multi-channel neural network but without the time stamp input channel: in effect this is a convolutional neural network. Given that there is no validation step in using the linear model, it was trained with a 90%-10% training-testing split split, while the neural network baseline model was trained using the previously described 80%-10%-10% training-validation-testing split. The results are shown in Table IV.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2978572, IEEE Transactions on Instrumentation and Measurement

TABLE III
COMPARISON OF AVERAGE CLASSIFICATION RESULTS WITH STATE-OF-THE-ART METHODS [27] AND BASELINE (MV REFERS TO THE MAJORITY-VOTE APPROACH)

Method	Number of samples	Number of sessions	Accuracy
Support Vector Machine	5	3	85.06%
K-Nearest Neighbor	5	3	89.13%
Linear Discriminant Analysis	5	3	80.89%
Decision Tree	5	3	90.44%
Convolutional Neural Network (baseline)	10	10	84.14%
Multi-Channel Neural Network	10	10	89.98%
Multi-Channel Neural Network (MV)	**10**	**10**	**92.20%**

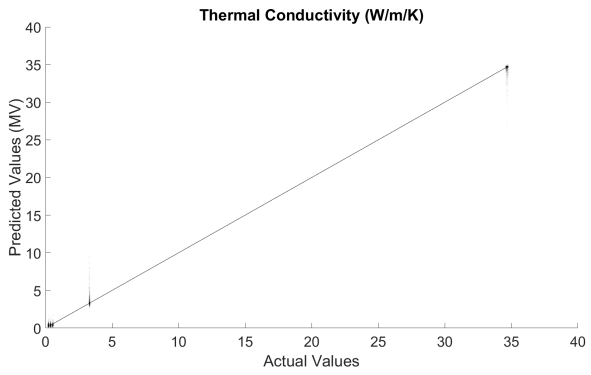

Fig. 8. Actual vs predicted values for thermal conductivity with 3% opacity.

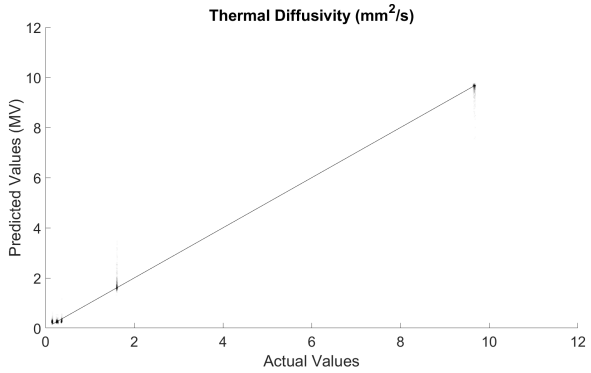

Fig. 9. Actual vs predicted values for thermal diffusivity with 3% opacity.

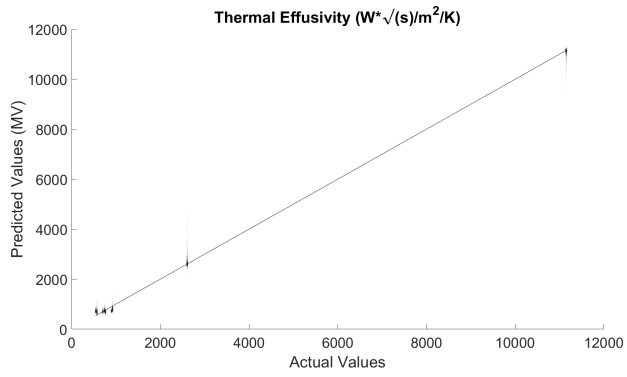

Fig. 10. Actual vs predicted values for thermal effusivity with 3% opacity.

These results show that using our neural network results in a marked improvement in the R^2-values over a standard linear regression. There is also a considerable benefit in using the time stamp input. While the improvement based on the average R^2-value over all three target values might not look like much (98.594% to 99.148%), this actually translates to reducing the residual error rate from 1.407% to 0.852%, which is a 39.5% relative residual error reduction. Furthermore, the majority-vote approach has further improved the results to 99.77%, 99.75% and 99.67%, for thermal conductivity, thermal diffusivity and thermal effusivity respectively.

VI. DISCUSSION

There are a number of important points to discuss based on the presented results. Firstly, the results for both classification and regression are highly significant. The 89.98% single-frame and 92.20% majority-vote testing accuracy for classification demonstrate that the system is capable of extracting meaningful features from the data frames and the method for partitioning the data set for the training, validation and testing process ensures these robust features can apply to a new session of thermal data that is completely disjoint from the training data set. Moreover, this result is produced using a relatively shallow neural network with only 2 convolutional and 2 fully connected stages. We designed this network with two principles in mind: that it has to serve as a proof of the concept of thermal or radiometric image based classification, and that its number of trainable parameters has to be at least an order of magnitude smaller than our training data set. State-of-the-art visual image-based classification is done by networks with hundreds of millions of trained parameters, and it is reasonable to expect that a network with a similarly complex structure would achieve even higher classification accuracy if it is demanded by the application scenario, as long as enough data is available to train, validate and test it. This expectation is backed up by results showing that our multi-channel approach performs significantly better than a standard convolutional neural network.

A similar argument can be made for the regression results: with a 99% average R^2 value for actual versus predicted thermal values, it is clear that there are features in the processed thermal data corresponding to all three examined thermal property values, and just as importantly, that the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIM.2020.2978572, IEEE Transactions on Instrumentation and Measurement

TABLE IV
COMPARISON OF REGRESSION RESULTS, WITH AND WITHOUT THE TIME STAMP INPUT, TO A BASELINE LINEAR REGRESSION, FOR THERMAL CONDUCTIVITY, THERMAL DIFFUSIVITY AND THERMAL EFFUSIVITY. (MV REFERS TO THE MAJORITY-VOTE APPROACH)

$R^2 values$ (testing)	Conductivity	Diffusivity	Effusivity
Linear Regression	37.69%	43.13%	46.57%
Convolutional Neural Network (no time stamp)	98.69%	98.45%	98.63%
Multi-Channel Neural Network (with time stamp)	99.170%	99.119%	99.156%
Multi-Channel Neural Network (MV, with time stamp)	**99.77%**	**99.75%**	**99.67%**

network is capable of being trained to find them. By extending the depth of the network, its number of trainable parameters and the size of the training data set, a further improvement can be made. The models trained to predict these thermal property values use the same network architecture, but their training data is different, so the final model weights are different. This demonstrates that this neural network is both adaptable and versatile. This is especially important, as it proves that the system can be used as a potential haptic sensing tool. There are two approaches to this use case: one is by directly correlating haptic properties such as stiffness, texture or friction to a predefined function of the three examined thermal property values. Given the accurate regression results this would present an accurate output. In absence of such a highly correlated function, the indirect approach is by using the three predicted thermal property values in a look-up-table of known values to deduce the material composition of the object and infer its haptic properties from known values.

Another factor that increases merit of these results is the variance of the synthetic target data set. This variance was introduced to account for the uncertainty of the measurements and the fact that these measurements could not be taken simultaneously with the data frames. As such, this variance has a meaningful purpose, but given that it does not correspond to any variance in the data frames, this means that it acts as a source of irreducible error that cannot be expected to be eliminated. The fact that the network still performs regression well despite this added error is a testament to its viability as a robust system for thermal property value estimation.

The significance of these results also validates the notion that the proposed multi-channel neural network is a compelling architecture, and that a single processed data frame contains enough information to be used for material recognition if its time stamp is known. Furthermore, using radiometric images as the raw data acquired by the thermal camera is also validated by the results. In general, different objects have different emissivity values, which dictates how much of the energy supplied by the laser diode is deposited into the material (to be re-emitted as thermal radiation) as opposed to directly reflected. However, there are a number of reasons why our system can be robust enough to overcome this (as shown by the experimental results as well). Firstly, it is true that most materials have a small and well-defined range of emissivity values [32]. That means to the system does not need to deal with a high variability of this hidden variable in most real-life scenarios. For instances in which the same material can have more than one narrow range of emissivity (such as polished or anodized aluminum), this can be taken care of by training the system on both of these samples. For classification they can be treated as different classes, but the system output will conform to the use case (i.e. whether it is desirable to differentiate between these cases for the end user or if it is better to combine them). For regression the same target value (thermal property) will be assigned for both cases. A second reason why the system performs very well without using emissivity values as inputs is that due to the linear nature of the heat equation, and in the presence of only negligible amount of advection during the data acquisition time window (as discussed in our earlier work [27]), emissivity acts as a uniform scaling factor to the input (data frames). Neural networks can learn to identify features that are complex enough that they are not affected by this uniform scaling, even with the realization presented in this article, where the networks for both classification and regression have more than 5000 trainable parameters. As stated above, having access to more data can increase this number and allow the learning of even more complex features that result in more robust systems.

In addition to stating the significance of the results, a handful of observations are also worth mentioning. One is related to the classification results. While the overall confusion matrix shows all materials classified correctly in a very high proportion of cases, examining the misclassification yields a noteworthy finding, which is that most of the misclassification cases are between coal and sorbothane frames. This can be convincingly explained by pointing out how similar they are in terms of their thermal property values. Defining a dissimilarity score between values a and b as a percentage ratio of $|a-b|/|a+b|$ (0% for $a = b$, close to 100% for $a << b$ or $a >> b$), this yields 7.63% for thermal conductivity, 6.99% for thermal diffusivity and 4.13% for thermal effusivity.

For regression, it is important to discuss the effect of the choice of loss function on the results. The loss function used for the network that yielded the results presented here is the mean squared error (MSE), which is the most commonly used loss function for regression. This error function heavily penalizes deviations with large absolute values, while it is less strict about forcing predicted values that are somewhat close to the actual values to be nearly identical. The resulting trade-off is quite clear in Figures 8, 9 and 10: all groups of sample values seem to have a similarly sized spread. The two rightmost groups in all figures (marble and steel) have a somewhat long vertical tail distribution of a few data points towards each other - this is due to the fact that these are the two materials where the laser excitation produces the lowest

signal-to-noise ratio, and frames captured in the latter half of an experimental round for these two materials are impossible to distinguish with the naked eye. The network seems to work better here than humans, but for the frames where it is slightly confused between whether they belong to marble or steel, the choice of mean squared error loss function incentivizes the prediction to be placed somewhere between the actual marble or steel values, rather than committing to either.

Although the results from the realization of the Haptic Eye framework are very promising, several limitations should be noted. First of all, the realization of the proposed framework did not take into consideration the effects of emissivity and surface texture. In more practical scenarios, these assumptions are not necessarily valid and thus the effects of emissivity and surface texture on characterization accuracy must be studied. These properties may be calculated using other sensory technologies (computer vision) and used as auxiliary inputs to the neural network to compensate and improve characterization. Furthermore, Due to the use of visible laser that might be harmful to humans, the application of the proposed system for characterizing human tissues could be a challenge.

VII. Conclusion and Future Work

This article presents a thermography-based contactless material recognition framework that uses a multi-channel neural network to make predictions based on individual thermal or radiometric frames. An experimental realization of this framework is described and the results show that the system is capable of performing classification with high accuracy (92.20%) on a data set of 10 different samples, and of providing accurate predictions for three different thermal properties.

Future work will expand the testing to include a large number of material samples. Furthermore, the use of deeper networks and the inclusion of additional features in order to improve the characterization accuracy beyond the 92% will be considered. In particular, we plan to experiment with autoencoders to derive an optimized set of features that maximize the characterization accuracy. Another interesting avenue involves exploring a direct mapping between thermal response and physical properties of objects such as stiffness and texture by relying on a versatile neural network architecture. We also plan to investigate using the system on samples that are outside of the training set and comparing the acquired regression results with the ground truth. Finally, we are developing a hardware prototype to encapsulate the system into a single-board computer setup for portability. Achieving these goals will ensure that the system can potentially be used in teleoperation, human-computer interaction and real-time physical interaction use cases.

References

[1] J. Zabalza, Z. Fei, C. Wong, Y. Yan, C. Mineo, E. Yang, T. Rodden, J. Mehnen, Q.-C. Pham, and J. Ren, "Smart sensing and adaptive reasoning for enabling industrial robots with interactive human-robot capabilities in dynamic environments—a case study," *Sensors*, vol. 19, no. 6, p. 1354, 2019.

[2] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1726–1743, 2013.

[3] F. Alambeigi, Z. Wang, R. Hegeman, Y.-H. Liu, and M. Armand, "A robust data-driven approach for online learning and manipulation of unmodeled 3-d heterogeneous compliant objects," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4140–4147, 2018.

[4] C. Pacchierotti, A. Tirmizi, G. Bianchini, and D. Prattichizzo, "Enhancing the performance of passive teleoperation systems via cutaneous feedback," *IEEE transactions on haptics*, vol. 8, no. 4, pp. 397–409, 2015.

[5] P. Ramsamy, A. Haffegee, R. Jamieson, and V. Alexandrov, "Using haptics to improve immersion in virtual environments," in *International Conference on Computational Science*, pp. 603–609, Springer, 2006.

[6] A. P. Advincula and K. Wang, "Evolving role and current state of robotics in minimally invasive gynecologic surgery," *Journal of Minimally Invasive Gynecology*, vol. 16, no. 3, pp. 291–301, 2009.

[7] T. Aujeszky, G. Korres, and M. Eid, "Measurement-based thermal modeling using laser thermography," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 6, pp. 1359–1369, 2018.

[8] R. W. Fleming, "Visual perception of materials and their properties," *Vision research*, vol. 94, pp. 62–75, 2014.

[9] A. Abdulali, R. Rakhmatov, T. Ogay, and S. Jeon, "Data-driven modeling and rendering of force responses from elastic tool deformation," *Sensors*, vol. 18, no. 1, p. 237, 2018.

[10] F. Ciampa, P. Mahmoodi, F. Pinto, and M. Meo, "Recent advances in active infrared thermography for non-destructive testing of aerospace components," *Sensors*, vol. 18, no. 2, p. 609, 2018.

[11] S. Shirmohammadi and A. Ferrero, "Camera as the instrument: the rising trend of vision based measurement," *IEEE Instrumentation & Measurement Magazine*, vol. 17, no. 3, pp. 41–47, 2014.

[12] M. Brandao, Y. M. Shiguematsu, K. Hashimoto, and A. Takanishi, "Material recognition cnns and hierarchical planning for biped robot locomotion on slippery terrain," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pp. 81–88, IEEE, 2016.

[13] J.-C. Krapez, L. Spagnolo, M. Frieß, H.-P. Maier, and G. Neuer, "Measurement of in-plane diffusivity in non-homogeneous slabs by applying flash thermography," *International journal of thermal sciences*, vol. 43, no. 10, pp. 967–977, 2004.

[14] F. Lakestani, A. Salerno, and A. Volcan, "Modulated spot heating for the measurement of thermal diffusivity," *Journal of applied physics*, vol. 97, no. 1, p. 013704, 2005.

[15] H. Liu, X. Song, J. Bimbo, L. Seneviratne, and K. Althoefer, "Surface material recognition through haptic exploration using an intelligent contact sensing finger," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 52–57, IEEE, 2012.

[16] T. Bhattacharjee, J. Wade, and C. C. Kemp, "Material recognition from heat transfer given varying initial conditions and short-duration contact," Georgia Institute of Technology, 2015.

[17] J. DeGol, M. Golparvar-Fard, and D. Hoiem, "Geometry-informed material recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1554–1562, 2016.

[18] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3479–3487, 2015.

[19] Q. Wang, P. Li, W. Zuo, and L. Zhang, "Raid-g: Robust estimation of approximate infinite dimensional gaussian with application to material recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4433–4441, 2016.

[20] T. Taniguchi, T. Takano, and R. Yoshino, "Multimodal hierarchical dirichlet process-based active perception," *arXiv preprint arXiv:1510.00331*, 2015.

[21] M. Strese, C. Schuwerk, A. Iepure, and E. Steinbach, "Multimodal feature-based surface material classification," *IEEE transactions on haptics*, vol. 10, no. 2, pp. 226–239, 2016.

[22] P. Giguere and G. Dudek, "A simple tactile probe for surface identification by mobile robots," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 534–544, 2011.

[23] A. Kimoto and Y. Matsue, "A new multifunctional tactile sensor for detection of material hardness," *IEEE Transactions on instrumentation and measurement*, vol. 60, no. 4, pp. 1334–1339, 2011.

[24] S. Baglio, L. Cantelli, F. Giusa, and G. Muscato, "Intelligent prodder: Implementation of measurement methodologies for material recognition and classification with humanitarian demining applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2217–2226, 2015.

[25] S. N. Pandya, B. J. Peterson, R. Sano, K. Mukai, E. A. Drapiko, A. G. Alekseyev, T. Akiyama, M. Itomi, and T. Watanabe, "Calibration of a thin metal foil for infrared imaging video bolometer to estimate the

spatial variation of thermal diffusivity using a photo-thermal technique," *Review of Scientific Instruments*, vol. 85, no. 5, p. 054902, 2014.
[26] T. Gfroerer, R. Phillips, and P. Rossi, "Thermal diffusivity imaging," *American Journal of Physics*, vol. 83, no. 11, pp. 923–927, 2015.
[27] T. Aujeszky, G. Korres, and M. Eid, "Material classification with laser thermography and machine learning," *Quantitative InfraRed Thermography Journal*, vol. 16, no. 2, pp. 181–202, 2019.
[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[29] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016.
[30] M. F. Ashby, H. Shercliff, and D. Cebon, *Materials: engineering, science, processing and design*. Butterworth-Heinemann, 2018.
[31] F. Chollet *et al.*, "Keras." https://keras.io, 2015.
[32] "Emissivity table." https://www.thermoworks.com/emissivity_table, 2018.

Tamas Aujeszky is a PhD student in Electrical Engineering at New York University Polytechnic School of Engineering. He has a BS in Electrical Engineering from New York University Abu Dhabi (NYUAD) and an MS in Electrical Engineering from New York University (NYU). His research area involves human computer interaction, tangible interfaces and multimodal interaction.

Georgios Korres studied Applied Mathematics in the School of Science and Engineering at the University of Crete. Korres was involved for several years with the development of educational software and hardware regarding educational robotics. The last three years he has also dealt with industrial automation (mainly in the field of recycling industry). His research interests focus on development of new sensors and actuators as well as the use of these in human computer interaction.

Mohamad Eid received the PhD in Electrical and Computer Engineering from the University of Ottawa, Canada, in 2010. He is currently an assistant professor of electrical and computer engineering in the engineering division at New York University Abu Dhabi (NYUAD). He was previously a teaching and research associate at the University of Ottawa (June 2008-April 2012). He is the co-author of the book: "Haptics Technologies: Bringing Touch to Multimedia", Springers 2011, the co-chair of the 3rd International IEEE Workshop on Multimedia Services and Technologies for E-health (MUST-EH 2013), and has been involved in the organization of the Haptic-Audio-Visual Environment and Gaming (HAVE) workshop for the years 2007, 2008, 2009, 2010, and 2013. His academic interests include Multimedia haptics, with emphasis on affective haptics, tangible human computer interaction, and instrumentations (sensors and actuators).