Quantitative InfraRed Thermography Journal

ISSN: 1768-6733 (Print) 2116-7176 (Online) Journal homepage: https://www.tandfonline.com/loi/tqrt20

Material classification with laser thermography and machine learning

Tamas Aujeszky, Georgios Korres & Mohamad Eid

To cite this article: Tamas Aujeszky, Georgios Korres & Mohamad Eid (2019) Material classification with laser thermography and machine learning, Quantitative InfraRed Thermography Journal, 16:2, 181-202, DOI: 10.1080/17686733.2018.1539895

To link to this article: https://doi.org/10.1080/17686733.2018.1539895

Published online: 20 Dec 2018.

QUANTITATIVE INFRARED THERMOGRAPHY JOURNAL
2019, VOL. 16, NO. 2, 181–202
https://doi.org/10.1080/17686733.2018.1539895

Taylor & Francis
Taylor & Francis Group

Material classification with laser thermography and machine learning

Tamas Aujeszky, Georgios Korres and Mohamad Eid

Engineering Department, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

ABSTRACT

Robotic devices that can perform teleoperation work in complex and unstructured environments (e.g. healthcare, education, automotive, telepresence, defence, etc.) rely mainly on visual and/or auditory feedback to interact with a remote operator. There is a vital need for measuring the physical properties of ambient environments when performing precise manipulation tasks over a distance, to improve the quality of performance. In this paper, we propose an approach to combine infrared thermography with machine learning for a fast, accurate way to classify objects of different material. A laser source stimulates the surface of the object while an infrared camera captures its thermal signature, extracts features about these signatures and feeds them into a machine learning-based algorithm for classification. Results demonstrated that a classification accuracy of around 97% can be achieved with majority vote Decision Tree classifier, even in the presence of data from multiple data acquisition sessions.

ARTICLE HISTORY
Received 14 June 2018
Accepted 21 October 2018

KEYWORDS
Infrared imaging; measurement by laser beam; machine learning; image classification; haptic interfaces

1. Introduction

Teleoperation systems allow a user to physically interact with remote environments (real or virtual). However, performing precise manipulation tasks over a distance has proven more difficult and complicated than expected, as it involves multiple modalities of interaction (audio, visual and physical). The difficulties of teleoperation are mainly due to the unbalanced relation between vision and physical sensing. Teleoperation systems 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 colour, geometry, motion and texture). However, these systems have very limited knowledge about the physical properties of the manipulated scene (such as stiffness or friction properties).

The acquisition, storage, communication and display of the physical properties improve the quality of performance [1] or immersion [2] 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 [3]. Meanwhile, the spread of breakthroughs in artificial intelligence, robotics and virtual/augmented reality will enhance computers with a higher level of

CONTACT Tamas Aujeszky tamas.aujeszky@nyu.edu New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

This article has been republished with minor changes. These changes do not impact the academic content of the article.

awareness, making them cognizant of their environment. It will also provide humans with natural, immersive experiences, which will have a positive impact on education, training, industrial design and entertainment.

While there are various mechanisms to acquire some information about objects that is dependent on their material properties, it is usually not possible to derive an explicit equation for the relation between these properties. However, some applications of machine learning are designed precisely to establish a relationship between different forms of data [4]. Machine learning itself is already found in a wide range of applications, including computer-assisted diagnosis in healthcare [5], material characterisation [6] and fault diagnosis in machinery systems [7–9].

This paper is an extended version of our previous work [10]. We present a system that uses active excitation infrared thermography to capture the thermal response of objects, then employs machine learning to perform classification on various samples based on these thermal signatures. The system works in a fast and contactless way, with high accuracy. As a result, this system serves as a basis of a potential three-dimensional mapping tool for material properties.

The rest of the paper is organised as follows: we present an overview of the related work in thermography and machine learning in Section II. The methodology is introduced in Section III. Section IV contains an experimental study to validate the proposed methodology and Section V presents and discusses the experimental results. Conclusions are drawn in Section VI, along with stating our future research plans.

2. Related work

2.1. Audiovisual-based classification and machine learning

The earliest and most widely used method to perform material classification based on audiovisual signals is through spectroscopy, which studies the interaction between light and the molecules that make up the samples. However, spectroscopic measurements require careful and time-consuming preparations of the samples and can be very sensitive to changes in environmental lighting conditions, making it incompatible with a fast and mobile setup [11]. With ever-growing audiovisual data acquisition speed and storage capabilities, machine learning approaches are becoming more effective in estimating material properties. Vision sensors have been extensively used to recognise the surface materials [12]. The authors in [13] recognised material categories using images and the estimated 3D points. More recently, Bell et al. [14] and Wang et al. [15] developed a deep learning technique for material recognition and segmentation of images in the wild. Brando et al. [16] developed a method for predicting surface friction in environments with diverse surfaces to improve the performance of a biped robot in locomotion planning. Even though a vision camera provides extensive information for surface material recognition, vision-based material characterisation has intrinsic limitations, due mainly to the diversity in surface material appearance (material with similar look but different material composition, e.g. a plastic wall might have the same look as a wooden wall). This motivates the researchers to adopt multi-modal and contact-based sensory data to improve the quality of material characterisation.

2.2. Contact-based material characterisation

Physical contact between a sensing instrument and an explored object provides significant information about the material properties. For instance, stroking a rigid tool over the surface of an object or tapping onto the object induces vibrations that can easily be measured using an accelerometer. Several contact properties are utilised to recognise materials, including vibrations, contact forces and thermal interactions.

The authors in [17] attached an accelerator to a tail probe touching the ground to efficiently classify multiple surfaces via the produced vibration patterns. A multifunctional tactile sensor for detecting material hardness using several piezoelectric ceramic transducers arranged on the surface of a semi-elliptic structure imitating the human finger is proposed in [18]. Liu et al. [19] proposed an exploration strategy for recognising the physical properties (friction coefficient) of unknown objects by measuring contact location, normal and tangential force, and the vibrations generated from the contact in real time. The authors in [20] investigated material recognition from heat transfer given varying initial conditions and short-duration contact. In [21], an intelligent prodder is proposed to stimulate the material surface by a mechanical solicitation and read an elastic response, which depends on the physical characteristics of the material.

Acoustic sensing has been successfully used to perceive robot-terrain interactions [22]. Christie and Kottege [23] developed a real-time terrain classification system for legged robots using acoustic features. A Support Vector Machine classifier was trained on seven different terrain types with an overall recognition rate of 95.1% with noise removal. Several researchers utilised ultrasound for learning about material properties: Shevchik et al. [24] performed failure prediction in lubricated surfaces using acoustic responses, Watanabe et al. [25] developed a robot-assisted acoustic inspection system of infrastructures, and Fukuda et al. [26], developed a softness measurement system by force-type tactile sensor using acoustic reflection.

Since different sensors can be used to capture different attributes about material characterisation, combining multiple modalities could improve the classification performance. Furthermore, recent advances in robotics make it possible for robots to acquire multimodal measurements for an improved understanding of their environment. For instance, Taniguchi et al. [27] proposed an active perception method for recognising object categories based on the multimodal hierarchical Dirichlet process model. This model enabled a robot to form object categories using multimodal information (including visual, auditory and haptic information), which could be observed by interacting with the object. Very recently, the authors in [28] presented an approach for tool-mediated surface classification that combined multiple modalities (such as sound and acceleration signals, which are captured by impact and movement) to classify textured surfaces under variable freehand movement conditions.

2.3. Thermography

Infrared thermography studies the thermal radiation emitted by objects. In particular, most thermographic inspections examine thermal gradients for qualitative evaluation [29,30]. Passive thermography includes the study of naturally occurring thermal gradients, such as thermal leakage in poorly insulated houses, erroneous amount of blood flow in certain parts

of the human body (thermology) or movement of living subjects in an area under surveillance. On the other hand, active thermography includes the process of excitation of objects to introduce thermal contrast and investigates its features and evolution overtime. Excitation can take place electromagnetically, optically or even mechanically. Depending on the temporal form of excitation, three approaches are common: flash thermography (single burst of energy deposit), step thermography (continuous energy deposit) or lock-in (sinusoidal or otherwise periodical energy deposit) thermography.

Thermography has been used for industrial purposes since the 1980s. Its primary use case has been non-destructive testing and evaluation [31–36], but it has also been utilised to study paintings, frescoes and other works of art [37–40]. A more recent application has been the examination of submicron electronic circuits for leakage current [41] or thermal behaviour characterisation of components [42,43]. The commonality in these applications is a preference for contactless and non-destructive analysis, making infrared thermography the ideal choice of technology to handle delicate objects. Usamentiaga et al. present a more detailed survey of the field in [44].

Thermographic data also have the ability to provide information about the material properties of the object under examination. This can be achieved by examining how a thermal gradient on the surface of an object evolves overtime and space. The spread of heat is governed by the heat equation [45], which contains a proportionality constant between the time derivative of temperature and the Laplacian of its spatial distribution. This constant is the thermal diffusivity of the material (the ratio of its thermal conductivity and volumetric heat capacity), and refers to the capability of the material to conduct heat compared with storing it. In recognition of the anisotropy of most real-life material samples, it is customary to make a distinction between in-plane and through-plane thermal diffusivity, depending on which direction one is interested in measuring heat flow, similarly to other thermal properties such as thermal conductivity. A popular method for measuring through-plane diffusivity is laser flash analysis, where a laser source provides instantaneous energy to one side of the sample and the half-rise time of the temperature is measured at the other side [46]. In-plane diffusivity can be measured using the modified transient plane source method, where a short heat pulse is applied at the same location on the surface of the object where a heat reflectance sensor measures an increase in voltage [47]. The disadvantages in these methods are that they both require physical contact with the samples, and the measuring equipment is large in size, deeming them impossible to be used in scenarios where a mobile setup is a necessity. Another thermal property, thermal effusivity, is also inextricably linked to haptics, as it can be described in layman's terms as whether the material feels cold to touch. Therefore, it is true that by examining the thermal response of an object we can potentially determine its thermal properties, which could provide insight into its material composition.

There have been various studies to use thermography for estimating thermal diffusivity [48–54], but they tend to be limited in their scope by applying a closed form solution that applies to a special case of the heat equation [55], where the excitation takes place over an infinitesimal amount of time and with a Gaussian distribution on the surface of the object. However, many realistic thermography experiments might have different excitation characteristics. A case in point is laser excitation step thermography, where excitation takes place over a finite but non-zero amount of time, and where the spatial distribution of the incident beam is not

Gaussian. Other derivations of solution formulas under different sets of circumstances have been provided in [56] and [57]. A model-based estimation with more general and realistic conditions has been proposed in [58].

As mentioned above, thermography has been mainly used for non-destructive evaluation and testing, where mostly qualitative examinations like fault detection take place. As such, the need for quantitative applications is new, and this explains why all model-based approaches are either limited or still in early development. On the other hand, current non-contact thermal sensors are capable enough to conduct quantitative examinations, given their ability to acquire thermograms in real time with relatively high frame rate, spatial resolution and thermal sensitivity.

2.4. Summary

This paper presents a system that uses laser excitation step thermography to capture the thermal signature of various experimental samples. The acquired data are processed and fed into a machine learning classification algorithm that successfully differentiates between these objects. The non-destructive nature of the system and the lack of need for a contact force make it perfectly suitable for examining delicate objects. Since the system is also contactless, it is applicable for measurement over a distance, and also for switching examination targets dynamically within its working envelope. The working envelope itself is only bounded by the specifications of the individual components, such as the power of the laser diode or the resolution of the infared camera. As a result of these properties, the system serves as a basis for a 3D haptic scanner for real-life scenes.

The contributions of this paper to quantitative thermography research are threefold. First, we propose an approach combining machine learning and active thermography for contactless material classification. Second, we evaluate this approach on a sample set of five materials that represent different material families. Finally, we examine the parameters of four common machine learning classifiers that allow for maximisation of classification accuracy.

3. Methodology

An overview of the proposed approach is shown in Figure 1. The Software Controller instructs the Flash Controller to define the scheme for the laser beam (Gaussian, rectangular, etc.), and it specifies the temporal shape of the beam it expects the Flash Controller to produce. The Software Controller also retrieves the physical properties from the machine learning-based material estimation component that can be used to construct haptic models for the explored environment. The Flash Controller generates the electric pulses necessary to activate the Laser Source. The Laser Source converts the electric signal into light signal. The generated light signal is steered at the sampling point using the Steering Stage. The Steering Stage is controlled by the Software Controller to make sure the entire scene is scanned to estimate the physical properties of every object in it.

A machine learning approach is proposed to utilise the spatio-temporal profile of material cooling to estimate the material type and properties since the cooling pattern forms a unique signature for a material. Specifically, upon heating a small patch of the

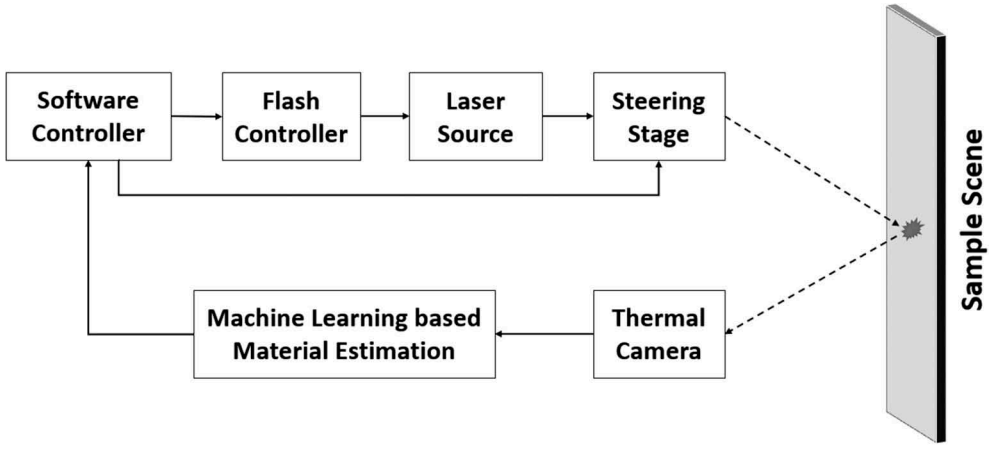


Figure 1. Proposed methodology.

target object, the cooling pattern is observed as a function of time and location (in a region including and around the patch that was heated). The temporally ordered sequence of images from the infrared camera over a time window (e.g. a few seconds in length) forms the raw data, which, after feature extraction, serve the input to the machine learning system as shown in Figure 1. The desired semantic output of the system is an estimate of the discrete label (e.g. classification among specific material types such as wood, glass, steel, etc.) and specific material properties as quantitative estimates (e.g. regression for density, thermal diffusivity).

4. Experimental study

4.1. Experimental setup

The setup of the experiment contained the following elements: a laser diode, a thermal camera, a desktop workstation, an Arduino board and various material samples. The setup is shown in Figure 2, while a corresponding system diagram is visible in Figure 3. The light emitted by the laser diode has a wavelength of 405 nm, and the diode is rated for 400 mW. The spatial distribution of the incident laser beam on the material samples is close to uniform over a disk with a diameter of approximately 1 mm. The infrared camera (FLIR T450sc) has a resolution of 320x240, a frame rate of 30 fps, a noise-equivalent temperature difference of up to 30 mK and is sensitive to the Long Wave Infrared Range (8–14 μm). The distance between the laser diode and the sample is approximately 14 cm, and the distance between the camera and the sample is approximately 22 cm. The Arduino board is responsible for connecting the laser diode to the desktop workstation and acts as a relay for turning it on and off. The desktop is running a script in the MATLAB R2017a software environment that allows control of both the camera and the laser diode.

The experimental samples used are: coal, matte black acrylic, machining polyethylene, a Sorbothane sample (having a durometer value of 70 on the Shore OO hardness scale) and black marble. These experimental samples all have smooth black surfaces,

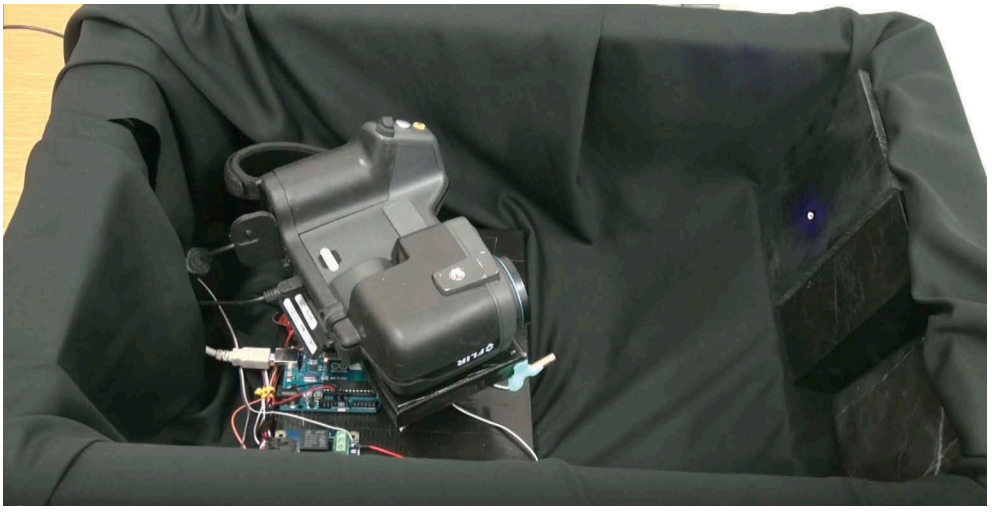


Figure 2. The experimental module.

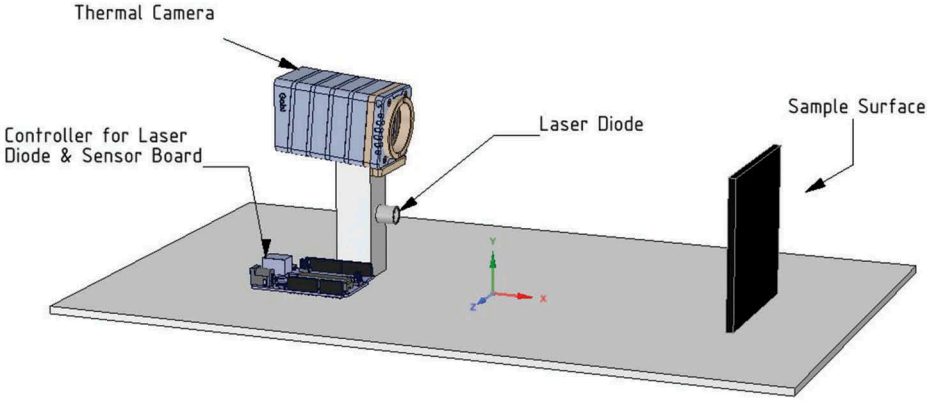


Figure 3. Diagram of the experiment.

which indicates that their emissivity values are all close to 1. Figure 4 shows these samples next to each other.

4.2. Data acquisition and processing

The data acquisition process consists of multiple steps, as follows: First, a total of 40 frames are recorded with the laser source turned off. The goal of this part of the recording is to serve as a baseline in the absence of excitation. In the next step, the laser diode is turned on for 10 s, without recording. At the end of the 10 s period, the laser diode is turned off and the infrared camera starts recording the cooldown phase,

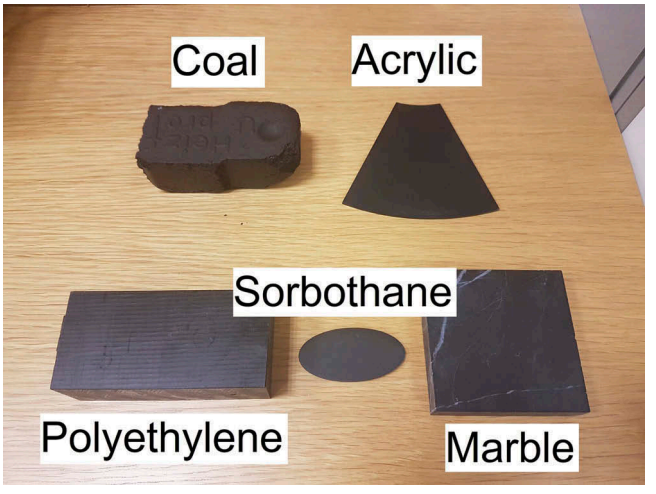
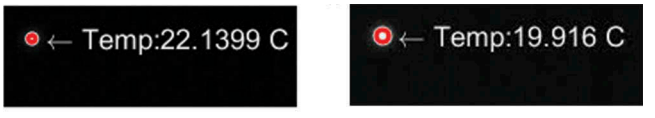


Figure 4. Experimental samples.



(a) Start of cooldown recording. (b) End of cooldown recording.

Figure 5. Illustration of the evolution of selected features at different times in the same cooldown recording. The red circle displays the width of the distribution and its diameter is the *full width at half maximum*. The displayed temperature is the *centre point temperature*, measured at the centre of the circle. These images have undergone ambient removal for better visibility.

for a total of 100 frames. Since the frame rate of the camera can vary slightly between 25 and 30 frames per second, the entire data acquisition process takes about 15 s.

The above-described procedure constitutes a single experimental run. An experimental session contains 10 consecutive runs for each of the five samples (for a total of 50 runs in a session). A break, lasting for a minimum of 3 min, is observed between consecutive experimental runs to ensure that the energy deposed in any experimental run will not interfere with the following ones. Recordings belonging to the same experimental session were performed over the same day, and a total of three experimental sessions ($3 \cdot 50 = 150$ experimental runs) were conducted over different days to examine how robust the system is with respect to slight expectable deviations in environmental factors such as ambient temperature and lighting.

Data processing is carried out over a sequence of steps. Due to a rare glitch in the software for the infrared camera, a single-frame temperature spike appears with a frequency of around 0.01%. If such a frame is present, it is removed. Next, the frames are spatially smoothened using a Gaussian filter to reduce the effect of thermal noise. In the following step, a limit is placed on the number of frames so that only those are kept which correspond to the first 80% of the amount of cooldown in the sequence. The reason for this exclusion of later frames is that they correspond to a stage where the thermal signal is much smaller than in the earlier frames, yielding a significantly worse

signal-to-noise ratio. In the final step, an ambient temperature image is formed from the initial 40 frames that were taken in the absence of excitation, and this image is subtracted from the data frames.

4.3. Classification

The processed data serves as the basis for building the feature space through feature extraction. A total of three features were selected for this classification task (illustrated in Figure 5):

- The temperature at the centre of excitation (measured in K)
- The full width at half maximum (FWHM) of the temperature distribution around the centre of excitation (measured in pixels), as the average of x- and y-directional cross-sections.
- The elapsed time between the start of the recording (the end of the excitation) and the frame capture time.

As a result, our feature space is 3D, and each frame in the processed data yields a single feature point. A single experimental run contributes as many feature points as frames it has in its processed form. Each feature point is assigned a class label depending on the sample object that it was acquired from.

With the features extracted and the feature space created, the next step is forming the training and validation datasets. For single-session-based classification, only the data from experimental runs in the same session is considered. For multi-session-based classification, data from each of the three sessions are assembled. Two decisions were made during the partition of training and validation datasets. The first decision is that each sample should be represented in the same proportion in both datasets. This is done to optimise for an objective function that is equally sensitive to the features of all samples. The second decision is that frames from the same experimental run should not be included in both datasets at the same time. The reason for this decision is that there might be some correlation between frames from the same recording that is independent of the sample, including minor artefacts produced by the environment, the camera, etc. The chosen scheme upholds both of these principles by selecting one experimental run per sample, and combining these to form the validation dataset. The feature points corresponding to the remaining experimental runs form the training dataset. In the case of single-session classification, this scheme can be executed in $10^5 = 100{,}000$ different ways, depending on which experimental run we choose for each sample to be included in the validation dataset. For multi-session classification, this number of possibilities grows to $30^5 = 24{,}300{,}000$.

The next step is the training of the classifiers. To eliminate bias based on the magnitude of feature values, each feature is rescaled to have a mean of 0 and a standard deviation of 1. A set of four different multi-class classifiers have been selected that are based on radically different assumptions and models:

- Multi-class support vector machine (SVM) using error-correcting output codes (ECOC) and a linear kernel

- K-nearest neighbour algorithm (KNN)
- Linear discriminant analysis classifier (LDA)
- Decision tree classifier (DT)

These classifiers are all commonly used for classification purposes in contemporary research [59–62], when training a custom neural network is unnecessary, undesirable or otherwise prohibitive. Once the classifiers are trained, each feature point in the validation dataset is classified by each of the classifiers and receives a separate classification label from each of them. There are two methods employed to determine the accuracy of the classifier. The first method uses these returned class labels and compares them to the labels of the actual objects that we know each frame corresponds to. Accuracy is then given as the ratio of correct label assignments over all label assignments and is a separate value for each of the classifiers.

The second method is named majority-vote rule. In this method, we compare the labels returned by the classifier in the group of feature points that are known to belong to the same object (in other words, this means these frames are captured in the same experimental run, and the system safely assumes that the material composition of the object did not change during that process). The majority (or plurality) of these returned labels will constitute the final label for all feature points in this group. Accuracy is then determined similarly to the first method: by the ratio of correctly assigned final labels over all assigned labels. This majority-vote rule is designed to yield increased accuracy when there is only a small portion of misclassified points, but the disadvantage is that it needs to have all feature points from the same experimental run as inputs.

5 Experimental results

5.1. Single-session results

This section details the results that were acquired when the datasets were composed of experimental runs in the same session (session A, B or C). Each session comprises 50 experimental runs, of which five are used for validation. The feature space for session A is displayed in Figure 6. It is visible that (other than the exception of marble and the centre point temperature feature) neither of the features are enough by themselves to classify the materials. It is up to the classifiers to combine them effectively, which certainly seems possible given the apparent separation of these classes in some areas.

The resulting accuracies can be seen in Table 1. These numbers are the result of averaging all 100,000 possible cases, as noted in the Classification subsection. Results are consistently high, especially when the majority-vote rule is used, in which case some classifiers achieve perfect accuracy in certain sessions.

5.2. Multi-session results

This section details the results that were acquired when a single dataset was composed of experimental runs in the all sessions (sessions A, B and C) combined. This means that the dataset comprises all 150 experimental runs, of which five are selected to form the validation dataset. The feature space representation of the complete dataset can be seen

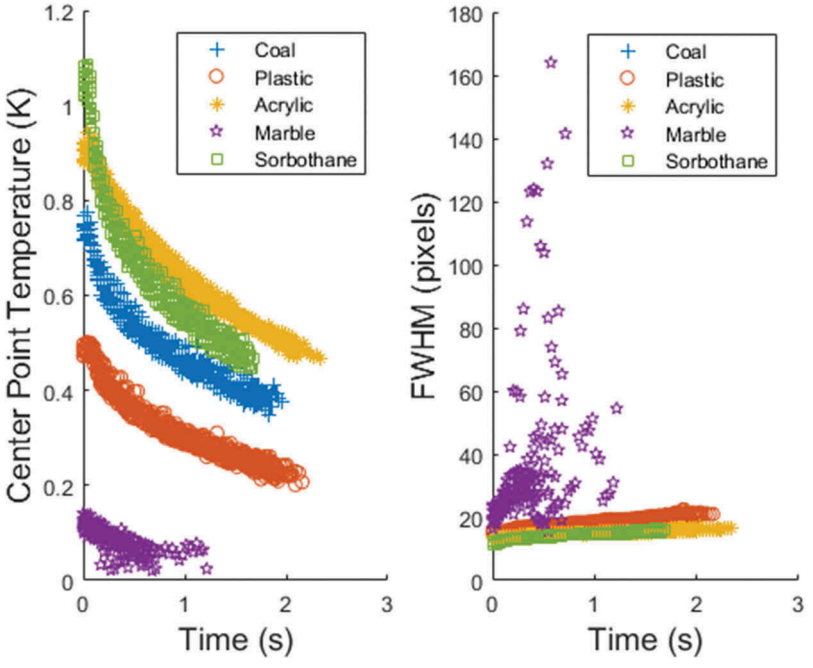


Figure 6. Feature space for session A.

Table 1. Single-session classification accuracies for Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Discriminant Analysis (LDA) and Decision Tree (DT) classifiers on sessions A, B and C.

	SVM	KNN	LDA	DT
With majority-vote rule				
Session A	96.86%	99.12%	90.39%	97.97%
Session B	89.92%	94.24%	87.65%	91.14%
Session C	95.06%	98.13%	94.45%	96.65%
With majority-vote rule				
Session A	100%	100%	99.00%	100%
Session B	94.80%	100%	94.00%	96.96%
Session C	100%	100%	100%	100%

in Figure 7. This figure displays much more overlap than in the single-session case, making classification a harder task.

The resulting accuracies can be seen in Table 2. Due to the large number of total cases (24.3 million) depending on which experimental runs are selected to form the validation dataset, these accuracies are calculated as the average of a random selection of $100,000$ cases. This random selection was evaluated multiple times, and the results differed in 0.01% or less, which suggests that results for the entire dataset are very close or identical.

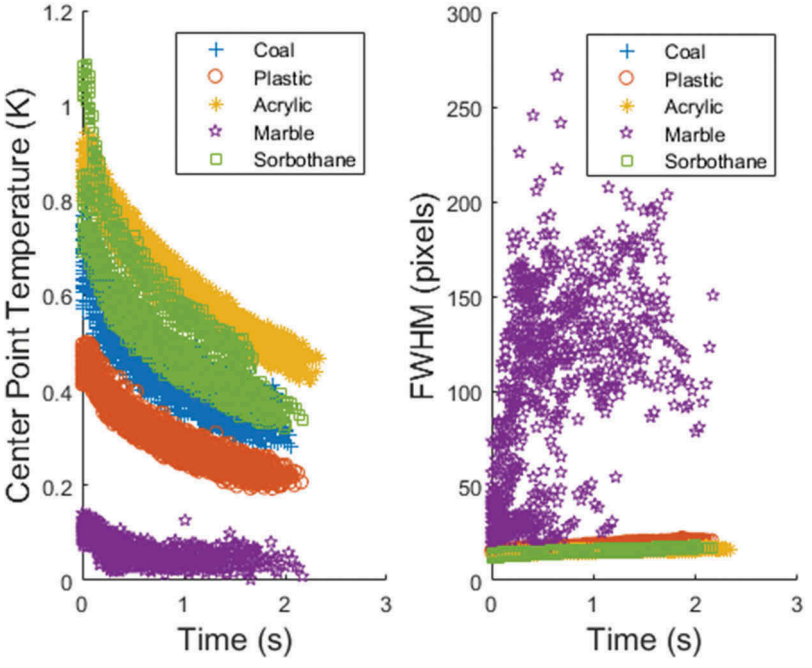


Figure 7. Feature space for sessions A, B and C.

Table 2. Multi-session classification accuracies for Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Discriminant Analysis (LDA) and Decision Tree (DT) classifiers on sessions A, B and C.

	SVM	KNN	LDA	DT
With majority-vote rule Session A + B + C	85.06%	89.13%	80.89%	90.44%
With majority-vote rule Session A + B + C	89.07%	93.89%	86.00%	97.08%

5.3. Classifier tuning

It can be seen from the classification results that the two classifiers with the best performance are the K-nearest neighbour (KNN) and decision tree (DT) classifiers. They consistently perform better than the other two classifiers and are better suited for this dataset for reasons discussed in the Discussion subsection. However, both KNN and DT have a set of parameters that can be tuned to achieve higher accuracy. This section is dedicated to investigating the effect of changing these parameters.

5.3.1. K-nearest neighbour classifier

Although KNN can have various parameters, the two most important ones are the number K of closest feature points to be examined by the classifier in the training set for each point to be classified, and the distance metric that the classifier uses to calculate distances between points to select the closest ones. The results in the previous

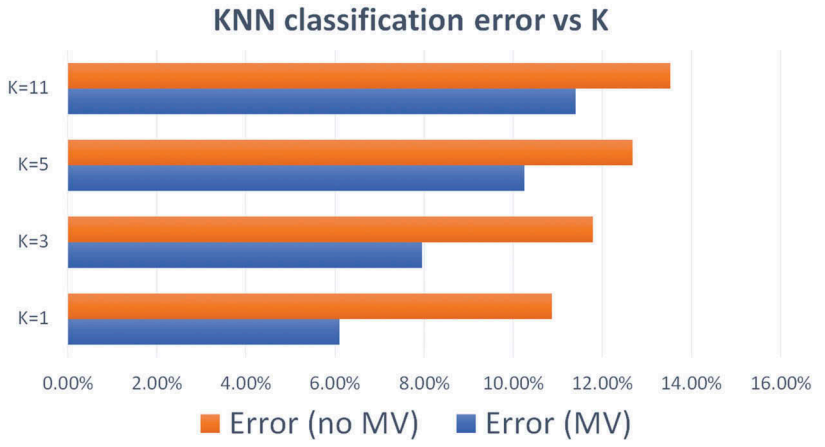


Figure 8. K-Nearest Neighbour multi-session classification error vs. K.

subsections correspond to the default setting in our application with $K = 1$ and Euclidean distance metric.

We have selected the values of $K = 1$, $K = 3$, $K = 5$ and $K = 11$ to investigate how the classification accuracy (or the classification error) changes for different values of K. Figure 8 displays the resulting classification error rate with each of these values, using the default Euclidean metric. What can clearly be seen here is that the error rate increases along with K, for both the regular and the majority-vote approaches. Therefore, the classifier seems to be optimally tuned at $K = 1$.

The second part of tuning the KNN classifier involved varying the distance metric it uses. Four options were used, they are: the default Euclidean metric, the Cityblock metric (sum of absolute values of coordinate differences, also known as Manhattan metric or Taxicab metric), the Chebyshev metric (highest absolute value of any coordinate difference) and a custom Minkowski distance with the exponent value of 4. The Minkowski metric in a general form for points $x = (x_1, \ldots x_n)$ and $y = (y_1, \ldots y_n)$ is:

$$d(x, y) = \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p} \tag{1}$$

It is true that $d(x, y)$ is a valid distance metric for any $p \geq 1$, and the Euclidean, Cityblock and Chebyshev metrics are special Minkowski distances with corresponding exponents of $p = 2$, $p = 1$ and $p = \infty$, respectively. For the custom case, the exponent value of $p = 4$ was used. The value of K used for these classifiers was the default value of $K = 1$. The resulting error rates can be seen in Figure 9.

While there is no clear trend between the magnitude of the exponent value and the classification error, it is clear that the Cityblock distance metric performs optimally in both approaches. The resulting error rate drops from 10.86% to 10.50% for the regular approach and from 6.11% to 4.30% for the majority-vote approach when switching from the default Euclidean distance metric to Cityblock, which is a marked improvement in both cases and is very significant compared with the 0.01% maximum deviation of the results when ran on

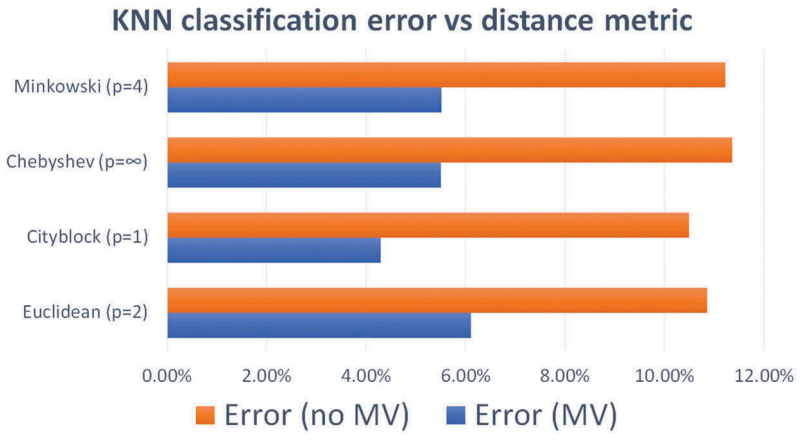


Figure 9. K-Nearest Neighbour multi-session classification error vs distance metric.

different random subsets of the combined dataset. To conclude, the optimally tuned KNN classifier for this task has a *K*-value of 1 and a Cityblock distance metric.

5.3.2. Decision tree classifier

The decision tree classifier similarly has a number of parameters that can be tuned to optimise the classification accuracy, including the minimum split size and the split criterion. Decision trees are prone to overfitting in some cases, given that they can grow branches and keep splitting themselves until every value in the training set is correctly classified. However, this can translate to worse accuracy when presented with a general case from the validation dataset. This can be controlled for by placing a limit on either the depth of the tree or the minimum size of leaves. The minimum split size is a value that essentially prevents the decision tree from branching from a node that has fewer points assigned than the limit. By default, the minimum split size has a value of $N = 1$.

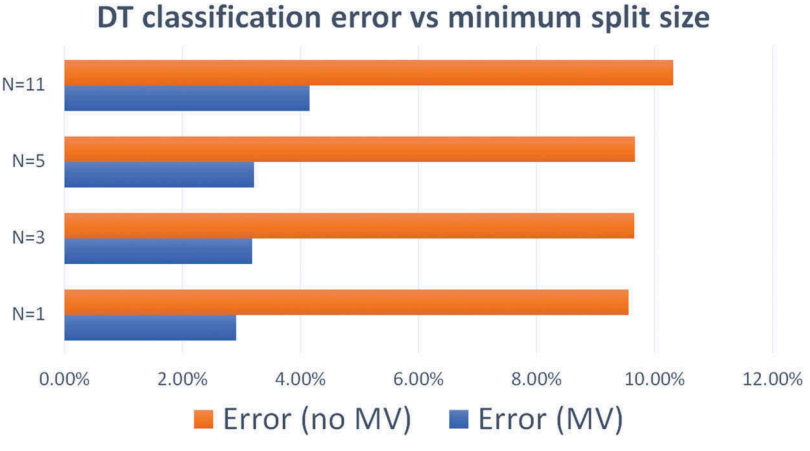


Figure 10. Decision Tree multi-session classification error vs minimum split size.

The values selected to be tested were $N = 1$ (the default case), $N = 3$, $N = 5$ and $N = 11$ for the minimum split size. The resulting classification error rates are presented in Figure 10. As in the case for KNN, it is visible that increasing the minimum split size increases the error rate (decreases accuracy), which results in the default minimum split size value of 1 being optimal.

The second parameter to be examined is the split criterion, which governs the algorithm that the decision tree uses to perform branching. Three different options were examined, they are: Gini's diversity index (GDI), the Twoing rule (Twoing) and maximum deviance (or cross-entropy) reduction (Deviance). The previously presented results all used the default GDI criterion. The results for the different split criteria (all using a minimum split size of 1) can be seen in Figure 11.

The best-performing split criterion based on these results is the maximum deviance reduction. For the default approach, it achieves an error rate of 9.29% compared with 9.56% for the default GDI criterion, while for the majority-vote approach, it results in a 2.45% error rate compared with 2.92% for GDI. Therefore, it can be concluded that the optimally tuned decision tree classifier has a minimum split size of 1 and a maximum deviance reduction split criterion.

5.4. Running time analysis

While classification accuracy is the key measurement for all classifiers, it is also important to ensure that both the training time (fitting the classifier on the training data) and the inference time (obtaining the class label for new feature point from a trained classifier)

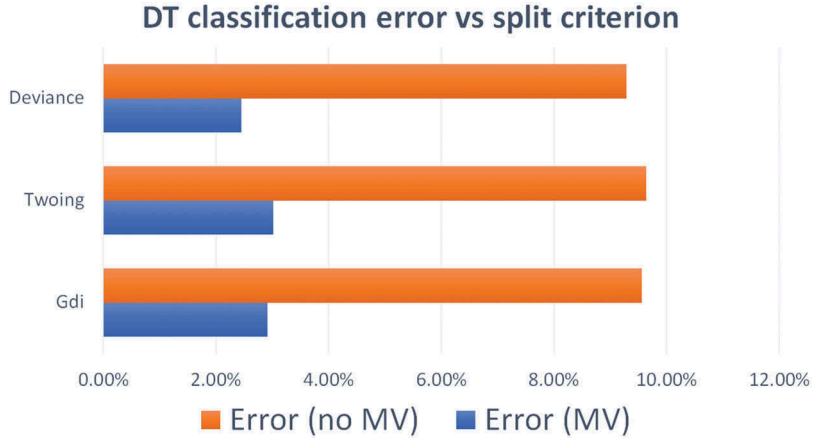


Figure 11. Decision Tree multi-session classification error vs split criterion.

Table 3. Running Time Analysis for KNN and DT Classifiers.

Running times	Training time	Inference/frame	Inference/run
KNN-Euclidean	17.3 ms	39.5 μs	1.99 ms
KNN-Cityblock	16.9 ms	42.1 μs	2.12 ms
DT-GDI	44.7 ms	14.5 μs	0.727 ms
DT-Deviance	57.0 ms	14.4 μs	0.723 ms

are sufficiently low that they do not prevent the system to be used in accordance with the goals and requirements of the application. These times were measured on the desktop workstation in the experimental setup. In particular, the regular approach for determining the classification accuracy performs inference for every single training frame (or feature point) independently, while the majority-vote approach requires all frames in an experimental run to be classified. Therefore, both of these processes were measured. The results can be seen in Table 3. It is clear from the table that the inference times are low enough to be considered negligible for real-time applications.

5.5. Discussion

The results presented in this article validate the proposition that infrared thermography is a reasonable method to be used for classification based on material properties. This is evident from the high accuracy values obtained, even when both the training and the validation dataset were formed from experimental runs belonging to different experimental sessions. A closer look at the results also confirms that the accuracy values largely depend on how well the theoretical basis of each classifier fits with the data. In all cases, the linear discriminant analysis (LDA) classifier has the worst performance. In fairness, this is likely explained by the fact that LDA is designed for classes that have Gaussian distribution in all dimensions. This is certainly not true here, as the 'elapsed time' feature has an obvious uniform distribution, while the other two features ('full width at half maximum' and 'centre point temperature', as described in Section 4.3) are also not Gaussian. Thus, the role of the LDA classifier was to provide insight into whether it is true that a wide range of classifiers perform well on the dataset or if there is a need for a more careful approach in selecting and optimising our classification tool.

The support vector machine (SVM) classifier is similar to the LDA case, namely in that the feature space representation of the data in Figure 7 suggests that there is unlikely to be a good linear separator plane between the different classes (using radial basis function kernel would have resulted in the same assumption of Gaussian distribution as for the LDA classifier). However, SVM performs consistently better than LDA, even though it is also in turn consistently outmatched by both the KNN and the DT classifiers. However, the SVM classifier, relying on a set of binary learners, suffers greatly from an exponential increase in training time as the number of classes is increased.

The two best-performing classifiers (KNN and DT) offer remarkable classification accuracy. This proves that, although there is some apparent overlap between different classes in the feature space representation, this is small enough that the amount of independent misclassification is around 9–10%. When the majority-vote rule is applied, the amount of misclassification is reduced to 2–4%. This is considered exceptional in contemporary real-life machine learning applications reaching or even surpassing the accuracy of contact-based classification systems [28,63]. The increase in classification accuracy for the majority-vote approach also confirms that most of the classification errors occur as outlier frames in an experimental run where the majority of frames are otherwise classified correctly. In other words, it can be viewed as more like a variation in the classifier output rather than a systematic bias. The increase in classification accuracy when moving to the majority-vote approach means that using more than one frame

captured of the same object and evaluating their predicted labels collectively increases classification accuracy.

A rather important topic when discussing machine learning classifiers is the phenomenon of overfitting and the concept of bias-variance trade-off. Overfitting is a fairly common, but undesirable phenomenon in machine learning. Normally, a classifier tries to capture every last bit of variance in the training dataset, and it tries to train itself so that every single data point in the training set is classified correctly. However, in some cases, the boundary between different classes in the feature space is rather soft, and certain feature points belonging to a class can manifest slightly on the wrong side of them (this is especially true for noisy data). In this case, the classifier 'learns' very complicated decision boundaries and this can transfer poorly to the validation set, as it is affected differently by random noise than the training set. The solution to this problem is to restrict the classifier to only learn classification rules that are not more complex than a specified level. This is called the bias-variance trade-off. The two classifiers examined in section 5.3 both have parameters to control this limit. Adjusting the numerical parameter values of K (for KNN) or the minimum split size (for DT) is a valid way to control the bias-variance trade-off, but the results show that there is no overfitting in this dataset even with the default values of 1 for both parameters. It is still preferable though to keep these as optional parameters as the number of classes and/or data points increase. On the other hand, the classifier tuning results show that the accuracy values can be slightly increased by tweaking the way the classifier handles distances (for KNN) or split criteria (for DT), which is a different way to tune these classifiers.

A final, worthy point of discussion relates to the running time of these classifiers. The average training times on a conventional desktop workstation of less than 58 ms (for DT) and less than 18 ms (for KNN) are negligible, and is also promising for two reasons. First, because training times of balanced decision trees and k-dimensional trees (for KNN) increase roughly linearly with respect to sample size. Therefore, even multiplying the training data by a thousand would result in only a few seconds of training. Second, this also allows for implementation on a single-board computer with significantly less power. While in most cases, training of a classifier does not need to be performed on the same device that uses the classifier, this result allows for systems that can potentially retrain themselves in light of new data on the spot, which is especially appealing for application scenarios in robotics. It should also be noted that there is no difference in training time between the regular and the majority-vote approach, since they both use the same classifiers.

Even more important is the case of the inference times. The results indicate that inference for a single feature point takes less than 43 μs for KNN and less than 15 μs for DT. This means that the inference itself adds negligible overhead to the entire operation. Given that query time for both the decision tree [64] and the K-dimensional tree (for KNN) [65] increases logarithmically with the number of samples, this means that even a much larger number of samples would still yield less than 100 μs, or that even a significantly slower system, like a single-board computer could perform inference of a single frame in less than a millisecond. This is especially vital for real-time applications. In the case of the majority-vote approach, the inference is needed to be carried out on all of the frames in the experimental run, which takes less than 2.2 ms for KNN and less than 0.8 ms for DT. This is also bound to increase logarithmically with sample size, so this

task will take less than 0.01 s even for significantly larger data sets and less than 0.1 s on a significantly slower computer. This is negligible compared with the time it takes to acquire all the frames, which is a prerequisite of the majority-vote approach, and takes a few seconds as per our experimental methodology (this length of time can be adjusted by changing the length of excitation or the number of captured frames). As a result, the two approaches present a classical trade-off between accuracy and speed. It is possible to balance these factors out by evaluating the majority (or plurality) in a limited number of frames from an experimental run, which should take less time than the majority-vote approach, and it would also have higher accuracy than the default approach.

6 Conclusion and future work

The results published in this article present a convincing case that the thermal signature of objects serves as a valid input for high accuracy classification between their material classes. This work functions as the foundation for a fast, accurate and contactless material characterisation tool, with the potential application of mapping a 3D scene.

Our future work includes expanding the limits of this system by making it capable to classify a larger set of material samples. We are also planning on employing improved signal processing algorithms for more accurate feature extraction and a higher tolerance with respect to environmental conditions. Ultimately, we are developing a portable version of the system with a single-board computer to replace the workstation and a micro-thermal imaging sensor to replace the thermal camera. This fully mobile setup greatly extends the possible use cases, since it could be deployed rapidly, with the potential ability to perform in real-time, and without the need to make any physical contact with the examined sample. This can enable use cases such as a robot that can roam an environment and identify physical properties of unknown objects before physically manipulating them. Scenarios such as rescue missions, logistical operations or precise teleoperation (including telesurgery), human-robot interaction and autonomous robots can rely on this technology.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

T. Aujeszky is an Electrical Engineering Ph.D. candidate at New York University, where he has received his B.S. and M.S. degrees previously. His research interests include haptic modeling, machine learning and image processing.

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).

ORCID

Tamas Aujeszky http://orcid.org/0000-0002-6643-4480
Georgios Korres http://orcid.org/0000-0003-4426-3593
Mohamad Eid http://orcid.org/0000-0002-6940-7891

References

[1] Pacchierotti C, Tirmizi A, Bianchini G, et al. Enhancing the performance of passive teleoperation systems via cutaneous feedback. IEEE Trans Haptics. 2015 Oct-Dec 1;8(4):397–409.
[2] Ramsamy P, Haffegee A, Jamieson R, et al. Using haptics to improve immersion in virtual environments. Springer Lecture Notes Com Sci. 2006;3992(1):603–609.
[3] Advincula AP, Wang K. Evolving role and current state of robotics in minimally invasive gynecologic surgery. J Minim Invasive Gynecol. 2009;16(3):291 301.
[4] Angra S, Ahuja S. Machine learning and its applications: A review, 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC); 2017; Chirala. pp. 57–60.
[5] Ma M, Sun C, Chen X. Discriminative deep belief networks with ant colony optimization for health status assessment of machine. IEEE Trans Instrum Meas. 2017 Dec;66(12):3115–3125.
[6] Liu H, Sun F, Fang B, et al. Multimodal measurements fusion for surface material categorization. IEEE Trans Instrum Meas. 67(2):246–256.
[7] Turner C, Sari-Sarraf H, Hequet E. Training a new instrument to measure cotton fiber maturity using transfer learning. IEEE Trans Instrum Meas. 2017 Jul;66(7):1668–1678.
[8] Sun J, Yan C, Wen J. Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. IEEE Trans Instrum Meas. 2018 Jan;67(1):185–195.
[9] Ding X, He Q. Energy-fluctuated multiscale feature learning with deep convNet for intelligent spindle bearing fault diagnosis. IEEE Trans Instrum Meas. 2017 Aug;66(8):1926–1935.
[10] Aujeszky T, Korres G, Eid M. Thermography-based material classification using machine learning. IEEE International Symposium on Haptic Audio-Visual Environments and Games (HAVE '17); 2017; Abu Dhabi. pp. 1–6.
[11] Smith M, Thompson K, Lennard F. A literature review of analytical techniques for materials characterisation of painted textilesPart 2: spectroscopic and chromatographic analytical instrumentation. J Inst Conservation. 2017;40(3):252–266.
[12] Shirmohammadi S, Ferrero A. Camera as the instrument: the rising trend of vision based measurement. IEEE Instrum Meas Mag. 2014 Jun;17(3):41–47.
[13] DeGol J, Golparvar-Fard M, Hoiem D. Geometry-informed material recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jul; Las Vegas. pp. 1554–1562.
[14] Bell S, Upchurch P, Snavely N, et al. Material recognition in the wild with the materials in context database. Proceedings of the IEEE Conference on Computer Vision Pattern Recognition; 2015 Jun; Boston. pp. 3479–3487.

[15] Wang Q, Li P, Zuo W, et al. RAID-G: robust estimation of approximate infinite dimensional gaussian with application to material recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016 Jun; Las Vegas. pp. 4433–4441.
[16] Brandao M, Shiguematsu YM, Hashimoto K, et al. Material recognition CNNs and hierarchical planning for biped robot locomotion on slippery terrain. Proceedings of the IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids); 2016 Nov; Cancun. pp. 81–88.
[17] Giguere P, Dudek G. A simple tactile probe for surface identification by mobile robots. IEEE Trans Robotics. 2011 Jun;27(3):534–544.
[18] Kimoto A, Matsue Y. A new multifunctional tactile sensor for detection of material hardness. IEEE Trans Instrum Meas. 2011 Apr;60(4):1334–1339.
[19] Liu H, Song X, Bimbo J, et al. Surface material recognition through haptic exploration using an intelligent contact sensing finger, Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2012 Oct; Vilamoura. pp. 52–57.
[20] Bhattacharjee T, Wade J, Kemp C. Material recognition from heat transfer given varying initial conditions and short-duration contact. Proceedings of Robotics: Science and Systems; 2015;Rome. pp. 19–27.
[21] Baglio S, Cantelli L, Giusa F, et al. Intelligent prodder: implementation of measurement methodologies for material recognition and classification with humanitarian demining applications. IEEE Trans Instrum Meas. 2015 Aug;64(8):2217–2226.
[22] Libby J, Stentz AT. Using sound to classify vehicle-terrain interactions in outdoor environments 2012. IEEE International Conference on Robotics and Automation (ICRA 2012); 2012 May; St. Paul.
[23] Christie J, Kottege N. Acoustics based terrain classification for legged robots, Proceedings of the IEEE International Conference on Robotics and Automation (ICRA); 2016 May; Stockholm. pp. 3596–3603.
[24] Shevchik SA, Saeidi F, Meylan B, et al. Prediction of failure in lubricated surfaces using acoustic timefrequency features and random forest algorithm. IEEE Trans Ind Inform. 2017 Aug;13(4):1541–1553.
[25] Watanabe A, Even J, Morales LY, et al. Robot-assisted acoustic inspection of infrastructuresCooperative hammer sounding inspection, Proceedings of the IEEE/RSJ. International Conference on Intelligent Robots and Systems (IROS); 2015 Sep/Oct; Hamburg. pp. 5942–5947.
[26] Fukuda T, Tanaka Y, Fujiwara M, et al. Softness measurement by forceps-type tactile sensor using acoustic reflection, proceedings of the IEEE/RSJ. International Conference on Intelligent Robots and Systems (IROS); 2015 Sep/Oct; Hamburg. pp. 3791–3796.
[27] Taniguchi T, Takano T, Yoshino R. Multimodal hierarchical dirichlet process-based active perception. Cornell University Library Online; 2015 [cited 2017 Nov 21]. Available from: https://arxiv.org/abs/1510.00331
[28] Strese M, Schuwerk C, Iepure A, et al. Multimodal feature-based surface material classification. IEEE Trans Haptics. 2017 Apr./Jun;10(2):226–239.
[29] Maldague X. Theory and practice of infrared technology for nondestructive testing. New York, NY, USA: Wiley; 2001. ISBN: 978-0-471-18190-3.
[30] Yefremenko V, Gordiyenko E, Shustakova G, et al. A broadband imaging system for research applications. Review Sci Instr. 2009;80(5):1–3. Article 056104.
[31] Durrani TS, Rauf A, Lotti F, et al. Thermal imaging techniques for the non destructive inspection of composite materials in real time. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 87); 1987; Dallas. pp. 598–601.
[32] Durrani TS, Rauf A, Boyle K, et al. Reconstruction techniques for the inspection of composite materials using thermal images. International Conference on Acoustics, Speech and Signal Processing (ICASSP 88); 1988; New York. pp. 863–866.
[33] Kakuda T, Limarga A, Vaidya A, et al. Non-destructive thermal property measurement of an APS TBC on an intact turbine blade. Surface Coating Technol. 2010;205(2):446–451.
[34] Bison P, Cernushi F, Capelli S. A thermographic technique for the simultaneous estimation of in-plane and in-depth thermal diffusivities of TBCs. Surf Coat Technol. 2011;205(10):3128–3133.

[35] Cernushi F, Bison P, Marinetti S, et al. Thermal diffusivity measurement by thermographic technique for the non-destructive integrity assessment of TBCs coupons. Surf Coat Technol. 2010;205(2):498–505.
[36] Horny N, Henry J-F, Offerman S, et al. Photothermal infrared thermography applied to the identification of thin layer thermophysical properties. 5th International Conference on Quantitative InfraRed Thermography (QIRT); 2000; Reims. pp. 36–41.
[37] Bodnar J-L, Nicolas J-L, Mouhoubi K, et al. Stimulated infrared thermography applied to thermophysical characterization of cultural heritage mural paintings. Eur Phys J Appl Phys. 2012;60(2):1–6.
[38] Kamel M, Bodnar J-L, Detalle V, et al. Stimulated infrared thermography applied to the local characterization of fresco. Quantitative Infrared Thermography Conference (QIRT 16); 2016; Gdansk. pp. 135–143.
[39] Mouhoubi K, Bodnar J-L, Detalle V, et al. Non-destructive testing of works of art by stimulated by infrared thermography: PPT interest, Quantitative Infrared Thermography Conference (QIRT 16); 2016; Gdansk. pp. 144–151.
[40] Bodnar J-L, Nicolas J-L, Mouhoubi K, et al. Characterization of an inclusion of plastazote located in an academic fresco by photothermal thermography. Int J Thermophys. 2013;34 (8–9):1633–1637.
[41] Huth S, Breitenstein O, Huber A, et al. Lock-in IR-Thermography a novel tool for material and device characterization. Diffus Defect Data Part B Solid State Phenom. 2002;82–84:741–746.
[42] Raad PE, Komarov PL, Burzo MG. Non-contact surface temperature measurements coupled with ultrafast real-time computation. Twenty-Third Annual IEEE Semiconductor Thermal Measurement and Management Symposium; 2007; San Jose. pp. 57–63.
[43] Ionescu C, Branzei M, Mihailescu B, et al. Studies on thermal properties of substrates for electronics using IR thermography. IEEE 20th International Symposium for Design and Technology in Electronic Packaging (SIITME); 2014; Bucharest. pp. 45–49.
[44] Usamentiaga R, Venegas P, Guerediaga J, et al. Infrared thermography for temperature measurement and non-destructive testing. Sensors. 2014;14(7):12305–12348.
[45] Heat Conduction Equation. Wolfram Mathworld. [cited 2017 Nov 2]. Available from: http://mathworld.wolfram.com/HeatConductionEquation.html
[46] Corbin SF, Turriff DM. Thermal diffusivity by the laser flash technique. New York, NY, USA: Wiley; 2012.
[47] C-Therm TCi(TM) Principles of Operation. Axel Products Physical Testing Services. [cited 2018 Aug 18]. Available from: http://www.axelproducts.com/downloads/C-Therm_MTPS_Theory_of_Operation.pdf
[48] Krapez J-C, Spagnolo L, Friess M, et al. Measurement of in-plane diffusivity in non-homogeneours slabs by applying flash thermography. Int J Thermal Sci. 2004;43(10):967–977.
[49] Lakestani F, Salemo A, Volcan A. Modulated spot heating for the measurement of thermal diffusivity. J Appl Phys. 2005;97(1):1–5. article 013704.
[50] Dong H, Zheng B, Chen F. Infrared sequence transformation technique for in situ measurement of thermal diffusivity and monitoring of thermal diffusion. Infrared Phys Technol. 2015;73(1):130–140.
[51] Pech-May NW, Wilbur N, Mendioroz A, et al. Simultaneous measurement of the in-plane and in-depth thermal diffusivity of solids using pulsed infrared thermography with focused illumination. NDT & E Inter. 2016;77(1):28–34.
[52] Pandya SN, Peterson BJ, Sano R, et al. Calibration of a thin metal foil for infrared imaging video bolometer to estimate the spatial variation of thermal diffusivity using a photothermal technique. Review of Sci Instr. 2014;85(5):1–9. article 054902.
[53] Gfroerer T, Phillips R, Rossi P. Thermal diffusivity imaging. Am J Phys. 2015;83(11):923–927.
[54] Yeshurun L, Azhari H. Non-invasive measurement of thermal diffusivity using high-intensity focused ultrasound and through-transmission ultrasonic imaging. Ultrasound Med Biol. 2016;42(1):243–256.
[55] Cernushi F, Russo A, Lorenzoni L, et al. In-Plane thermal diffusivity evaluation by infrared thermography. Review of Sci Instr. 2001;72(10):3988–3995.

[56] Tinsley L, Chalk C, Nicholls J, et al. A study of pulsed thermography for life assessment of thin EB-PVD TBCs undergoing oxidation ageing. NDT & E Int. 2017;92(1):67–74.
[57] Cernushi F, Bison PG, Figari A, et al. Thermal diffusivity measurements by photothermal and thermographic techniques. International Journal of Thermophysics. 2004;25(2):439–457.
[58] Aujeszky T, Korres G, Eid M. Measurement-based thermal modeling using laser thermography. IEEE Trans Instrum Meas. 2018;67(6):1359–1369.
[59] Nugrahaeni RA, Mutijarsa K. Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification. International Seminar on Application for Technology of Information and Communication (ISemantic); 2016; Semarang. pp. 163–168.
[60] Bellingegni AD, Gruppioni E, Colazzo G, et al. NLR, MLP, SVM, and LDA: a comparative analysis on EMG data from people with trans-radial amputation. J Neuroeng Rehabil. 2017;14(82):1–16.
[61] Noi PT, Kappas M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. Sensors. 2018;18(1):1–20.
[62] Sousa RT, Marques O, Soares FA, et al. Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs. Procedia Computer Science. 2013;18(1):2579–2582.
[63] Zheng H, Fang L, Ji M, et al. Deep learning for surface material classification using haptic and visual information. IEEE Trans Multimedia. 2016;18(12):2407–2416.
[64] Complexity - Decision Trees. Scikit-learn documentation. section 1.10.4. [cited 2018 Aug 14]. Available from: http://scikit-learn.org/stable/modules/tree.html
[65] K-Dimensional Trees. Nearest neighbors, scikit-learn documentation. section 1.6.4.2. [cited 2018 Aug 14]. Available from: http://scikit-learn.org/stable/modules/tree.html