Thermography-based Material Classification using Machine Learning

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

Georgios Korres
Engineering Department
New York University Abu Dhabi
Abu Dhabi, United Arab Emirates
george.korres@nyu.edu

Mohamad Eid
Engineering Department
New York University Abu Dhabi
Abu Dhabi, United Arab Emirates
mohamad.eid@nyu.edu

***Abstract*—Infrared thermography has been widely used today for nondestructive evaluation and testing of materials and other qualitative approaches. However, the field of thermography is much less developed. Most of the existing research uses a relatively simple model, while more realistic models are currently in development. One interesting scenario for thermography is determining the material composition of objects based on their thermal response to excitation, which could lead to applications such as multimodal human-computer interaction, teleoperation and non-contact haptic mapping. This paper presents a system that is capable of classification between a range of different materials in real time, using laser excitation step thermography and a set of machine learning classifiers. Experimental results demonstrate a consistently high accuracy in determining the label of the material, even when the dataset is composed of multiple different sessions of data acquisition.**

***Index Terms*—Laser thermography, machine learning, material characterization, haptic mapping**

I. Introduction

Over the past few years we have seen the rise of immersive experiences. With the advent of virtual reality, mixed reality and augmented reality, it has been more and more important to capture three-dimensional spaces accurately, for various purposes. These include immersion for entertainment purposes, for accurate modeling with regards to industrial needs, for teleoperation in robotics, for personal communication and various other applications that will become essential in the next decade.

While optical 3D scanning has seen vast advancements lately and we are able to capture the structure of our environment with high fidelity, some applications require more than the acquired visual information. This is usually the case when there is a need for interaction with the recorded objects. These include multimodal human-computer interaction, where the computer might need to perform haptic interaction with the human counterpart, or teleoperation, where a robot has to know about the object it is interacting with, as well as a virtual reality environment in which the user can manipulate their surroundings. These scenarios are dependent on accurate mapping of the physical properties of objects in a scene. This need is also apparent in the emergence of haptic rendering technologies.

Machine learning is an emergent technology that is already widely utilized across different fields for various purposes. From optical character recognition, to online financial fraud detection and self-driving vehicles, there are enormous benefits of using machine learning, in that it enables automation of certain processes that previously required a human counterpart, and given enough training data it is easily capable of outperforming humans in a variety of use cases.

The purpose of this work is to present a novel system that is capable of differentiating between objects of different material composition using laser thermography and machine learning, thereby acting as a potential material recognition component of a 3D-scanning system. This setup uses laser excitation step thermography to examine the thermal response of an object, and machine learning classifiers are applied to decide which material class it belongs to.

The rest of the paper is organized as follows: Section II presents an overview of the related work in this field, and Section III specifies the methodology of data acquisition and evaluation. Section IV presents and discusses the experimental results, while conclusions and future work are detailed in Section V.

II. Related Work

Infrared thermography is the process where certain physical properties (such as thermal, electrical or mechanical) of an object are revealed through its heat signature. [1] [2]. Thermography as a field has been around since the 1980s and it has been used by the industry as a nondestructive testing and evaluation tool since then [3] [4] [5] [6] [7] [8] [9] [10] [11] [12]. Certain special use cases for infrared thermography include examining paintings and frescoes [13] [14] [15] [16], or finding leakage current in submicron electronic circuits [17] as well as characterizing the thermal behavior of their components [18] [19]. Thermography is a preferred method of inspection in these use cases due to its contactless and nondestructive nature, which is essential when examining delicate objects. A further overview of the range of applications is listed by Usamentiaga et al. [20].

Depending on the presence of a heat source, infrared thermography can either be categorized as active or passive. Passive thermography relies on naturally occuring thermal gradients around its area of interest. In contrast, an active thermography setup contains an energy source that is responsible

978-1-5386-0979-8/17/$31.00 ©2017 IEEE

for supplying the examined object with a thermal contrast. This source can provide mechanical, optical or electromagnetic excitation to the material. Distinctions can also be made based on the temporal properties of the excitation process (flash, step or lock-in thermography). While important observations can be deduced about a single thermal image, it is much more informative to examine the temporal and spatial evolution of a thermal gradient, which is related to the material properties of the object, such as thermal diffusivity, volumetric heat capacity or thermal conductivity. In particular, there have been several studies on estimating thermal diffusivity of a material [21] [22] [23] [24] [25] [26] [27], which is the measure of how capable the material is of conducting heat with respect to storing it. However, most methods rely on an equation derived from a solution of a special case of the heat equation with limited potential to generalize. More recently, a model-based estimation of thermal diffusivity was proposed [28].

While model-based approaches of using infrared thermography certainly have potential, they are either very limited in their scope of application, or at very early stages of development. This is due to the fact that thermography has been primarily used in the industry for qualitative purposes in nondestructive evaluation and testing, such as fault detection. However, current state-of-the-art thermal sensors are capable of real-time acquisition with sufficient spatial, temporal and thermal resolution that makes it possible to perform quantitative research. The use of machine learning for classification is motivated by a few considerations. Firstly, that machine learning has been successfully applied in a wide range of scenarios, as mentioned above. Secondly, that for most applications, it might be sufficient to arrive at a categorical result (i. e. the material class of an object), rather than a numerical one. This bypasses the need for a specific equation and opens up the door to a classification approach, which is the hallmark feature of machine learning algorithms. Nevertheless, performing regression analysis between certain physical properties of a material and the features of its thermal response holds ample potential as well.

The aim of this paper is to present a system combining machine learning classification with laser excitation step thermography that is capable of differentiating between different materials based on their thermal response to excitation. This system is contactless and non-destructive, therefore it is suitable for a wide range of application scenarios, including ones where delicate objects are present. This is made possible by a lack of need for any contact force. Moreover, the range of these measurements is bounded by practical considerations only, such as the resolution of the thermal camera or the power of the laser source. This also means that such a system is also capable of dynamically switching between objects at different distances (withing its envelope), and is thereby suitable for 3D scanning of an entire scene.

III. Methodology

A. Experimental Setup

The experimental setup contains a laser source, an infrared thermal camera, a control circuit, and a material sample. A snapshot of the setup is shown in Figure 1. A diagram of the components is shown in Figure 2. The laser source is rated at 400 mW and emits a violet beam with a wavelength of 405 nm. The thermal camera is a FLIR450sc model, capable of 480x320 resolution and 30 fps frame rate. The laser source is approximately 14 cm away from the surface of the sample, while the camera is approximately 22cm away from it. The control circuit involves an Arduino board that is instructed through the serial port, as well as a desktop workstation running the Matlab R2017a software. This enables both the laser diode as well as the thermal camera to be both controlled by a script.

Fig. 1. The experimental module.

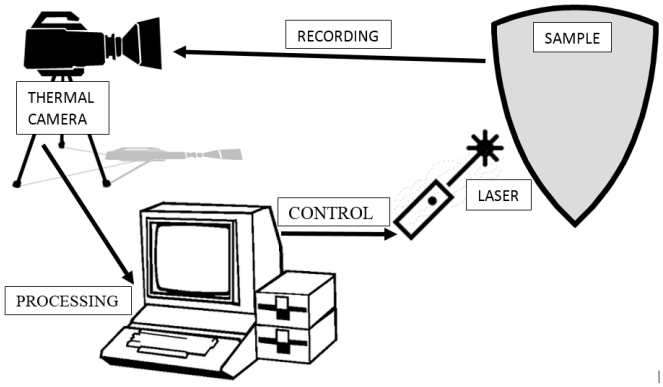

Fig. 2. The modeled experiment.

All samples in the experiment were chosen to have a smooth black surface, implying an emissivity value of approximately 1. The sample materials are: coal, machining polyethylene, black acrylic (matte), black marble and a sample of Sorbothane

with a durometer level of 70 on the Shore OO hardness scale. These samples are displayed in figure 3.

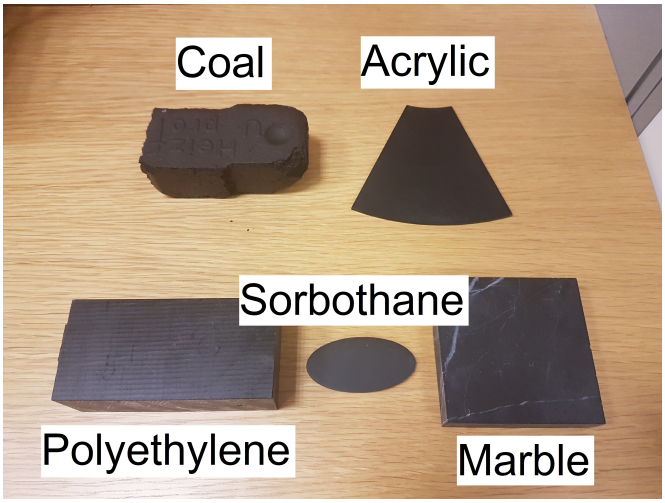

Fig. 3. Experimental samples.

B. Data Acquisition and Processing

The experimental procedure consists of three parts: data acquisition, data processing and lastly, training and evaluating machine learning classifiers. During the data acquisition phase, the camera first records 40 frames of thermal image, which are later used to remove the ambient temperature component of the thermal data. The laser source is then turned on for a fixed amount of 10 seconds, heating up the material at a specific location. As soon as this period ends, the camera starts recording again, this time to acquire the cooldown process around the excitation area. The camera measures the thermal radiation emitted from the surface of the material and deduces its temperature with a specified spatial resolution and frame rate. 100 frames are captured during this period, lasting for approximately 4 seconds. This is the raw data to be used.

In total, 150 experimental runs were completed over 3 sessions (50 runs per session). In each session, each of the five samples underwent laser excitation and subsequent data acquisition 10 times, with sufficiently large time gaps in between these experimental runs to allow for the material to reach thermal equilibrium. Every experimental session was completed over a single day, but different sessions were performed over different days to investigate the robustness of the system with respect to slight possible changes in ambient temperature and other uncontrollable factors.

Processing the data happens in multiple steps. Having checked for irreal temperature spikes (happening about 0.01% of the time due to a peculiarity in the thermal sensor), the frames are smoothened to counter thermal noise (the sensor inside the camera is rated at a noise equivalent temperature difference of up to 50 mK). Next, the set of frames is truncated so that only the initial sequence corresponding to the first 80% of cooldown (the temperature difference between the start and the end of the sequence, based on the maximum temperature per frame) is kept. This way only the part with the highest signal-to-noise ratio remains. Finally, the pre-recorded frames are combined to calculate the ambient temperature of the material before excitation, and this is subtracted from the data.

C. Classification

Once each frame is processed as mentioned above, three features are extracted from each frame. These are: the temperature at the center of excitation, the full width at half maximum around the center of excitation (measured in pixels, averaged in x-, and y-directions on the frame), and the time difference between the end of excitation and the time the frame itself was captured. Therefore, each frame of each experimental run is a data point in the three-dimensional feature space. Each point also has a label based on which sample it corresponds to.

The classifier training and validation procedure is as follows. First, we assemble the dataset, which consists of data from an equal amount of experimental runs for each of the five samples. Depending on the procedure, this dataset involves either a single experimental session (single-session classification, 50 experimental runs) or all three experimental sessions (multi-session classification, 150 experimental runs). From this dataset, one experimental run corresponding to each of the samples is taken to form the testing (validation) dataset, while the remaining runs form the training dataset. This way it is ensured that each sample is proportionally represented in both sets, and it is also ensured that frames belonging to the same experimental run are not simultaneously present in both datasets.

In the next step, a set of multi-class classifiers are trained on the data points of the training dataset (which correspond to individual frames from experiments). Each data point consists of the three features mentioned above and a label representing the class of the corresponding sample. In order not to bias the classifiers towards features with larger numerical values, feature scaling is carried out before training, where each feature is scaled to have zero mean and unit standard deviation. The classifiers used are:

- Multi-class support vector machine (SVM) using error-correcting output codes (ECOC) and a linear kernel
- K-nearest neighbor algorithm (KNN)
- Linear discriminant analysis classifier (LDA)
- Decision tree classifier (DT)

These classifiers are very different from each other in terms of any underlying assumptions, complexity and the treatment of dimensionality.

Having trained the classifiers, the testing dataset is used for validation. Each classifier returns a predicted class label for all the points in the testing dataset, which are then compared with their actual labels. Accuracy of a classifier is then measured as the the ratio of the number of correct predictions and the size of the testing dataset. This approach is taken further by introducing a majority-vote rule. In this case, while each validation data point has its own set of predicted labels by

each classifier, the points belonging to the same experimental run are grouped together and a single final label is assigned to them based on the predicted label with the highest frequency in a group. This step is motivated by the fact that each experimental run corresponds to a single sample, and most real-life applications of this proposed system would also allow for this assumption, which in turn reduces misclassification (i. e. when a single frame is misclassified, but the majority is classified well, this method improves classification accuracy).

IV. Results

A. Single-session Classification

Figure 4 represents the feature space for session A. It can be seen that other than the center point temperature corresponding to marble frames, none of the other samples have a single feature that would allow for classification purely on its own. However, an appropriate combination of the three features clearly displays potential for accurate classification, as the datasets are somewhat separated.

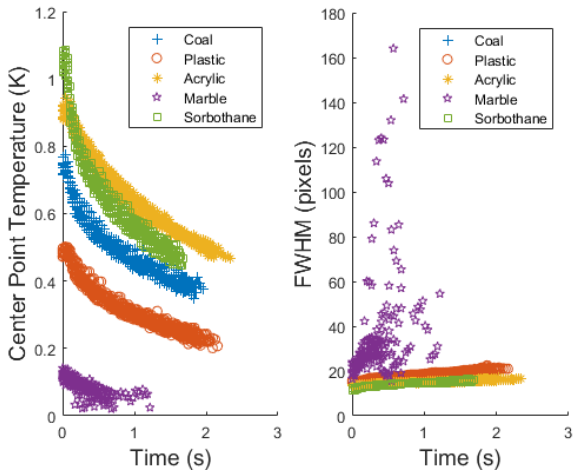

Fig. 4. Feature space for session A.

During single-session classification, a dataset is formed from a single experimental session, which corresponds to 10 experimental runs per sample (50 runs in total). The training set consists of 9 runs per sample, while the testing set comprises 1 run per sample. There are 10^5 possible combinations for selecting the testing dataset. Each of these combinations were tested out and the accuracies were averaged, separately for each classifier and session. The resulting accuracies can be seen in table I.

The previous results were further improved by the introduction of the majority-vote rule (discussed above) into the classification procedure. Table I also presents these results for each classifier under this approach.

B. Multi-session Classification

In the multi-session case, the dataset contains all 150 experimental runs from sessions A, B and C. The feature space representation of the complete dataset can be seen in figure 5.

Without majority-vote rule	SVM	KNN	LDA	DT
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%
Without majority-vote rule	**SVM**	**KNN**	**LDA**	**DT**
Session A	100%	100%	99.00%	100%
Session B	94.80%	100%	94.00%	96.96%
Session C	100%	100%	100%	100%

TABLE I
CLASSIFICATION ACCURACIES FOR SUPPORT VECTOR MACHINE (SVM), K-NEAREST NEIGHBOR (KNN), DISCRIMINANT ANALYSIS (LDA) AND DECISION TREE (DT) CLASSIFIERS ON SESSIONS A, B AND C.

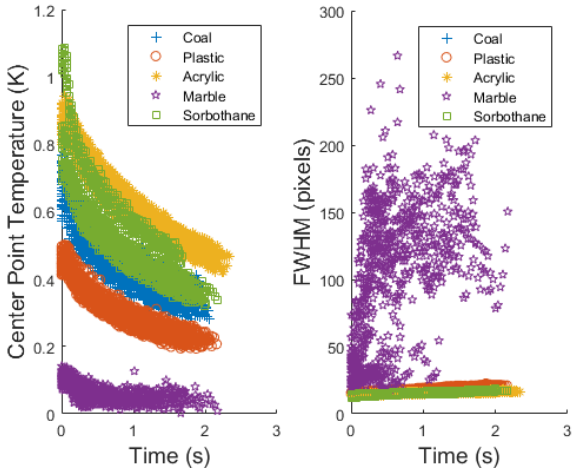

Fig. 5. Feature space for sessions A, B and C.

The testing set again contains 1 run per sample, while the training set consists of the remaining 29 runs per sample. This results in 30^5 (24.3 million) combinations for selecting the testing set, which could not all be tested out. However, a random selection of 10^3, 10^4 and 10^5 combinations were tested out, with differences in resulting accuracies less than 0.1%. Table II displays these results for 10^5 combinations.

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

TABLE II
MULTI-SESSION CLASSIFICATION ACCURACIES FOR SUPPORT VECTOR MACHINE (SVM), K-NEAREST NEIGHBOR (KNN), DISCRIMINANT ANALYSIS (LDA) AND DECISION TREE (DT) CLASSIFIERS ON SESSIONS A, B AND C.

C. Discussion

Based on these results, the classification algorithms have yielded a very high accuracy in general. We have selected these four algorithms because of how differently they work and it is mirrored in the results. For example, the linear discriminant analysis classifier assumes a Gaussian distribution of data points belonging to the same class, which is definitely not the case with respect to the time stamp feature (which has uniform distribution). Having known about this conflict in advance, this classifier was used as a control element to determine how well classifiers do overall on this dataset, even when their usage is not ideal. Similarly for the support vector machine classifier with linear kernel, there does not seem to be an obvious set of hyperplanes to serve as a classifier based on the feature set representation. The results are in line with these thoughts, seeing that even though all classifiers perform well, the discriminant analysis consistently performs worse than the rest, while the SVM classifier is a constant third.

It can be concluded that the mathematically least complex classifiers performed the best, namely KNN and the Decision tree. These classifiers both achieved well over 90% accuracy in the multi-session classification procedure, which is very convincing, given the amount of overlap present in figure 5. This is likely due to the contribution of the majority-vote rule, which rejects outliers when they are present in a sufficiently low ratio.

These results bode well for a larger amount of classes (samples), since they are achieved without specifically tuning the classifiers. This includes changing the value of N for KNN, or the minimum split size for a decision tree. These parameters would allow for controlling the bias-variance tradeoff of the classifier, which would be very useful when the number of samples is increased beyond the current selection.

Finally, it is also worth mentioning that the runtime of these algorithms is remarkable. The procedure involving the training of a hundred thousand of each of the four classifiers took less than a day on a regular desktop workstation, and most of the time was devoted to training the SVM classifier, which relies on a set of binary learners and is therefore increasingly time consuming to train as the number of classes increases. Meanwhile, the best performing classifiers (KNN and Decision tree) take less than $0.1s$ second each to train. It is equally important to mention that evaluation takes even less time, and this is the key process that would be part of a real-life application scenario, where a pre-trained classifier is applied to an unknown sample. This enables real-time decision making. However, using the majority-vote rule requires more than a single captured frame, and is therefore dependent on these frames being captured in the first place. This presents two possible choices to the end user, depending on the use case: use classification without the majority-vote rule for real-time results, or use the majority-vote rule for soft-real time results with a higher accuracy.

V. Conclusion and Future Work

Our results show that material classification is possible with high accuracy based on the measurement of thermal response to laser excitation. This experiment lays down the groundwork for developing an efficient and effective material characterization method that could be used for haptic mapping.

Our future plans include examining the proposed system on a larger number of material samples. Improving the signal processing algorithms is also possible to be even more robust with respect to environmental conditions. Finally, we are working on decoupling the system from the workstation to produce a mobile device. This would enable the proposed model to indirectly measure the physical properties of materials without physical contact, which has potential applications in tele-operation systems, haptic modeling and human-robot interaction.

References

[1] X. Maldague, *Theory and Practice of Infrared Technology for Nondestructive Testing*, Wiley: New York, NY, USA, 2001.

[2] V. Yefremenko, E. Gordiyenko, G. Shustakova, Y. Fomenko, A. Datesman, G. Wang, J. Pearson, E. E. W. Cohen, and V. Novosad, *A broadband imaging system for research applications*, Review of Scientific Instruments, vol. 80, Article 056104, pp. 1-3, 2009.

[3] T. S. Durrani, A. Rauf, F. Lotti, and S. Baronti, *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, pp. 598-601.

[4] T. S. Durrani, A. Rauf, K. Boyle, and F. Lotti, *Reconstruction Techniques For The Inspection of Composite Materials Using Thermal Images*, International Conference on Acoustics, Speech and Signal Processing (ICASSP 88), 1988, pp. 863-866.

[5] T. Kakuda, A. Limarga, A. Vaidya, A. Kulkarni, and T. D. Bennett, *Non-destructive thermal property measurement of an APS TBC on an intact turbine blade*, Surface & Coating Technology, vol. 205, no. 2, 2010, pp. 446-451.

[6] P. Bison, F. Cernushi, and S. Capelli, *A thermographic technique for the simultaneous estimation of in-plane and in-depth thermal diffusivities of TBCs*, Surface & Coatings Technology, vol. 205, no. 10, 2011, pp. 3128-3133.

[7] F. Cernushi, P. Bison, S. Marinetti, and E. Campagnoli, *Thermal diffusivity measurement by thermographic technique for the non-destructive integrity assessment of TBCs coupons*, Surface & Coatings Technology, vol. 205, no. 2, 2010, pp. 498-505.

[8] S. E. Burrows, S. Dixon, S. G. Pickering, T. Li, and D. P. Almond, *Thermographic detection of surface breaking defects using a scanning laser source*, NDT & E International, vol. 44, no. 7, 2011, pp. 589-596.

[9] I. Plotog, B. Mihailescu, I. Pencea, M. Branzei, P. Svasta, T. Cucu, and M. Tarcolea, *Methods for Pads Thermophysical Parameters Assessment in Terms of 4P Soldering Model*, IEEE 34th International Spring Seminar on Electronics Technology, 2011, pp. 320-326.

[10] N. Horny, J.-F. Henry, S. Offerman, C. Bissieux, and J. L. Beaudoin *Photothermal infrared thermography applied to the identification of thin layer thermophysical properties*, https://www.researchgate.net/publication/265667818, Retrieved October 6th, 2016.

[11] N. W. Pech-May, A. Oleaga, A. Mendioroz, and A. Salazar, *Fast Characterization of the Width of Vertical Cracks Using Pulsed Laser Spot Infrared Thermography*, Journal for Nondestructive Evaluation, vol. 35, article 22, 2016.

[12] S. E. Burrows, A. Rashed, D. P. Almond, and S. Dixon, *Combined laser spot imaging thermography and ultrasonic measurements for crack detection*, Nondestructive Testing and Evaluation, vol. 22, 2007, pp. 217-227.

[13] J.-L. Bodnar, J.-L. Nicolas, K. Mouhoubi, and V. Detalle, *Stimulated infrared thermography applied to thermophysical characterization of cultural heritage mural paintings*, The European Physical Journal Applied Physics, vol. 60, article 21003, 2012, pp. 1-6.

[14] M. Kamel, J.-L. Bodnar, V. Detalle and J.-M. Vallet, *Stimulated infrared thermography applied to the local characterization of fresco*, Quantitative Infrared Thermography Conference (QIRT 16), 2016, pp. 135-143.
[15] K. Mouhoubi, J.-L. Bodnar, V. Detalle, and J.-M. Vallet, *Non-destructive testing of works of art by stimulated by infrared thermography: PPT interest*, Quantitative Infrared Thermography Conference (QIRT 16), 2016, pp. 144-151.
[16] J.-L. Bodnar, J.-L. Nicolas, K. Mouhoubi, J. C. Candore, and V. Detalle, *Characterization of an Inclusion of Plastazote Located in an Academic Fresco by Photothermal Thermography*, International Journal of Thermophysics, vol. 34, no. 8-9, 2013, pp. 1633-1637.
[17] S. Huth, O. Breitenstein, A. Huber, D. Dantz, U. Lambert, and F. Altmann, *Lock-in IR-Thermography a novel tool for material and device characterization*, Diffusion And Defect Data Part B Solid State Phenomena, vols. 82-84, 2002, pp. 741-746.
[18] P. E. Raad, P. L. Komarov, and M. G. Burzo, *Non-Contact Surface Temperature Measurements Coupled with Ultrafast Real-Time Computation*, Twenty-Third Annual IEEE Semiconductor Thermal Measurement and Management Symposium, 2007, pp. 57-63.
[19] C. Ionescu, M. Branzei, B. Mihailescu, and D. Bonfert, *Studies on Thermal Properties of Substrates for Electronics using IR Thermography*, IEEE 20th International Symposium for Design and Technology in Electronic Packaging (SIITME), 2014, pp. 45-49.
[20] R. Usamentiaga, P. Venegas, J. Guerediaga, L. Vega, J. Molleda, and F. G. Bulnes, *Infrared Thermography for Temperature Measurement and Non-Destructive Testing*, Sensors, vol 14, 2014, pp. 12305-12348.
[21] J.-C. Krapez, L. Spagnolo, M. Friess, H.-P. Maier, and G. Neuer, *Measurement of in-plane diffusivity in non-homogeneours slabs by applying flash thermography*, International Journal of Thermal Sciences, vol. 43, no. 10, 2004, pp. 967-977.
[22] F. Lakestani, A. Salemo, and A. Volcan, *Modulated spot heating for the measurement of thermal diffusivity*, Journal of Applied Physics, vol. 97, article 013704, 2005, pp. 1-5.
[23] H. Dong, B. Zheng, and F. Chen, *Infrared sequence transformation technique for in situ measurement of thermal diffusivity and monitoring of thermal diffusion*, Infrared Physics & Technology, vol. 73, 2015, pp. 130-140.
[24] N. W. Pech-May, N. Wilbur, A. Mendioroz, and A. Salazar, *Simultaneous measurement of the in-plane and in-depth thermal diffusivity of solids using pulsed infrared thermography with focused illumination*, NDT & E International, vol. 77, 2016, pp. 28-34.
[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, article 054902, 2014, pp. 1-9.
[26] T. Gfroerer, R. Phillips, and P. Rossi, *Thermal diffusivity imaging*, American Journal of Physics, vol. 83, 2015, pp. 923-927.
[27] L. Yeshurun and H. Azhari, *Non-invasive Measurement of Thermal Diffusivity Using High-Intensity Focused Ultrasound and Through-Transmission Ultrasonic Imaging*, Ultrasound in Medicine & Biology, vol. 42, no. 1, 2016, pp. 243-256.
[28] T. Aujeszky, G. Korres and M. Eid, *Measurement-Based Thermal Modeling Using Laser Thermography*, in IEEE Transactions on Instrumentation and Measurement (unpublished).