Technische Universität München
Fakultät für Elektrotechnik und Informationstechnik
Lehrstuhl für Medientechnik

Haptic Material Acquisition, Modeling, and Display

Matti Strese, M. Sc.

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr.-Ing. Sami Haddadin
Prüfer der Dissertation:	1.	Prof. Dr.-Ing. Eckehard Steinbach
	2.	Prof. Dr.-Ing. Klaus Diepold

Die Dissertation wurde am 22.10.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 11.03.2021 angenommen.

Abstract

Technical solutions that capture and display visual and auditory information have evolved tremendously in recent decades and enable the development of high-definition (HD) applications nowadays. For example, a camera can capture an HD image of a stationary scene, which is then displayed and passively observed by a human sitting in front of a computer monitor. In contrast, the sense of touch requires active physical interaction from a user, and hence, imposes critical physical dependencies like scan force and speed during acquisition as well as during the haptic rendering of the required data. While great effort has been made to improve systems addressing the kinesthetic component of haptic rendering, i.e., the display of forces and torques, comparably less achievements have been reported in the cutaneous, or tactile, subdomain of haptic research. The construction of appropriate material models, the classification of materials, and the display of tactile subdimensions like roughness or hardness still pose several challenges on both the haptic data collection and display side. On the one hand, the design of an appropriate material scanner, a sufficiently representative feature space describing the perceptual dimensions of tactile touch, and a material classification engine have only been sparsely addressed so far. On the other hand, the electro-mechanical design and assembly of a display device that artificially recreates the perceptual dimensions is a likewise challenging task.

This thesis addresses both major challenges and presents a novel sensing framework (Texplorer) as well as a novel display system (Tactile Computer Mouse, TCM). Additionally, the connection between data acquisition and display is achieved by using a perceptually modeled and multimodal feature space. A novel material taxonomy is defined that unambiguously allows the identification of materials based on a supervised classification engine. A subset of the features as well as multidimensional input data traces further provide the required control signals for the TCM to display tactile sensations to a human user.

Kurzfassung

Technische Lösungen im Bereich der Bild- und Audioverarbeitungssysteme haben in den letzten Jahrzehnten einen fortgeschrittenen Entwicklungsstand erlangt und ermöglichen heutzutage hochqualitative Anwendungen. Beispielsweise kann eine Kamera eine statische Szene in entsprechend hoher Qualität aufnehmen, welche dann passiv für einen menschlichen Nutzer über einen Computerbildschirm angezeigt wird. Im Gegensatz dazu benötigt der Tastsinn eine aktive Interaktion des Nutzers mit seiner Umgebung. Folglich haben physikalisch-bedingte Abhängigkeiten wie Kraft oder Geschwindigkeit einen besonderen Einfluss auf einerseits die Aufnahme wie auch andererseits auf das haptische Darstellen solcher Daten. Die Forschung im Bereich der kinästhetischen Komponente, bestehend aus Kräften und Kraftmomenten, wurde in den letzten Dekaden intensiv vorangetrieben, währenddessen jedoch deutlich weniger Fortschritte für die kutane, also taktile, Domäne erreicht wurden. Die Aufnahme adequater Materialmodelle, die Klassifikation dieser und das taktile Darstellen verschiedener Subdimensionen wie Rauheit und Härte stellen besondere Herausforderungen dar. Einerseits mangelt es an entsprechenden Aufnahmegeräten, einem hinreichenden Merkmalsraum welcher alle taktilen Dimensionen abdeckt, sowie einem umfassenden Materialerkennungssystem für die Klassifizierung. Andererseits kann es als ambitionierte Aufgabe angesehen werden, ein elektromechanisches Gerät zu erschaffen, welches die taktilen Dimensionen künstlich erzeugt und währenddessen die Eingabeabhängigkeiten, wie Geschwindigkeit und Krafteinwirkung des Nutzers, modellbezogen beachtet.

Diese Dissertation addressiert beide Herausforderungen und präsentiert ein System zur Datenaufnahme (Texplorer) sowie ein taktiles Darstellungsgerät (Tactile Computer Mouse, TCM). Darüber hinaus wird die Verbindung zwischen Datenaufnahme und -darstellung basierend auf einen perzeptuellen Merkmalsraum geschaffen. Eine neue Material-Taxonomie wird vorgestellt, welche Materialien eindeutig einteilt. Ein Teil der identifizierten Merkmale sowie die aufgenommenen Datenspuren dienen als multidimensionale Eingabe für die TCM zur Darstellung taktiler Eigenschaften für den Nutzer.

Acknowledgements

The work presented in this dissertation was carried out as a member of the academic staff at the Chair of Media Technology (LMT) at the Technical University of Munich. Many people have supported me during the past years. First of all, I would like to express my deep gratitude to my supervisor Prof. Eckehard Steinbach for providing me with the opportunity to conduct research in his group. With his trust and scientific expertise, he managed to push me in the right direction while giving me ample space to develop my own ideas. I am particularly grateful for his extensive and very diplomatic feedback, and notably, his astonishing endurance during the countless review iterations of my papers. Despite his busy schedule, he also took his time for discussions. Furthermore, I would like to thank Prof. Klaus Diepold for agreeing to become the second examiner and Prof. Sami Haddadin for chairing the thesis committee.

I feel lucky to have met my former and current colleagues at LMT. Not only have they helped me professionally, but also made my time at LMT special and enjoyable on a personal level. I am particularly grateful to Dr. Rahul Chaudhari and Dr. Clemens Schuwerk for kick-starting my research with their experience. I highly value my colleagues and students at LMT, however, I feel obliged to highlight a few people. I thank Dr. Nicolas Alt, Dr. Dmytro Bobkov, Dr. Xiao Xu, Alexandra Zayets, Mojtaba Karimi, Edwin Babaians, Dr. Christoph Bachhuber, Andreas Noll, Martin Piccolrovazzi, and Jingyi Xu for interesting discussions about virtually anything from rather trivial to highly complex matters, and Andreas Noll also for the nice TCM1 design. I am also grateful to all my students, especially Simon Hofmann, Kevin Kuonath, Jonas Kirsch, Lara Brudermueller, Irem Öztürk, Yannik Böck, Yigitcan Özer, and Tamara Fiedler who helped me a lot and also contributed to our publications.

I appreciate the quick and reliable administrative support provided by Marta Giunta, Ingrid Jamrath, Simon Krapf, Dr. Martin Maier, Martina Schmid, and Brigitte Vrochte.

Finally, my utmost gratitude goes to my parents and my family members, who believed in my abilities and helped me in any situation, and to my wonderful wife Yana who supported me throughout my time at LMT, broadened my vision, and endured all of my flaws.

Contents

Notation **v**

1 Introduction **1**

2 Background and Related Work **5**
2.1 Human Haptic Perception 5
2.1.1 Receptors 6
2.1.2 Human Exploratory Procedures and Haptic Dimensions 9
2.2 Machine Haptics 13
2.2.1 Data Acquisition Devices 13
2.2.2 Comparison of Robotic and Human Operator Scan Approaches 15
2.2.3 Display Devices 16
2.3 Computer Haptics 21
2.3.1 Hardness 21
2.3.2 Friction 25
2.3.3 Macroscopic Roughness 28
2.3.4 Microscopic Roughness 29
2.3.5 Warmth 34
2.4 Haptic Communication 37
2.5 Chapter Summary 37

3 Data Acquisition **39**
3.1 Haptic Databases 39
3.1.1 Initial LMT 108 Material Database 39
3.1.2 LMT 184 Haptic Database 40
3.2 Texplorer Sensor Evolution 46
3.3 Texplorer2 47
3.3.1 Recording Procedure 47
3.3.2 TU1: Microscopic and Macroscopic Roughness 48
3.3.3 TU2: Warmth, Friction, and Stiffness 52
3.3.4 TU3: Visual Properties 56
3.3.5 TU4: Volume and Mass 57
3.4 Chapter Summary 58

4 Supervised Material Classification **59**
4.1 Handcrafted Features for Material Classification . . . 61
4.1.1 Texture-based Image Features . . . 62
4.1.2 Audio Features . . . 72
4.1.3 Tactile Features from Related Work . . . 79
4.1.4 Tactile Features based on Five-dimensional Feature Space . . . 80
4.1.5 TacTUM Tactile Features based on SynTouch Feature Space . . . 80
4.1.6 Features Exceeding Human Touch Capabilities . . . 98
4.2 Data Preparation . . . 99
4.2.1 Feature Suitability for Machine Learning Approaches . . . 99
4.2.2 Feature Optimization . . . 105
4.2.3 Feature Space Creation . . . 106
4.2.4 Feature Selection . . . 108
4.3 Model Selection and Classifier Fitting . . . 110
4.3.1 Linear Classifiers and Non-linear Kernel Support Vector Machines . . . 113
4.3.2 Nearest Neighbor Classifiers . . . 115
4.3.3 Probabilistic Models and Naïve Bayes Classifier . . . 116
4.3.4 Decision Trees . . . 118
4.3.5 Ensemble Classifier . . . 119
4.3.6 Classification Results . . . 121
4.4 Subjective Experiment . . . 126
4.4.1 Experimental Setup and Procedure . . . 127
4.4.2 Experimental Results . . . 128
4.4.3 Discussion . . . 131
4.5 Deep Learning-based Material Classification . . . 132
4.5.1 Background . . . 132
4.5.2 Network Design Components . . . 132
4.5.3 DL-based Texture Classification of LMT 184 Haptic Database . . . 138
4.5.4 Hybrid Model for Material Classification . . . 147
4.6 Chapter Summary . . . 151

5 Haptic Display **153**
5.1 Haptic Display Devices . . . 153
5.1.1 Augmentation of Traditional Force Feedback Haptic Devices . . . 153
5.1.2 Tactile Computer Mouse . . . 154
5.2 Tactile Dimension Rendering and Display . . . 160
5.2.1 Friction . . . 160
5.2.2 Macroscopic Roughness Rendering . . . 163
5.2.3 Microscopic Roughness Rendering . . . 167
5.2.4 Hardness Rendering . . . 172
5.2.5 Thermal Rendering . . . 174
5.2.6 Adhesive Tack . . . 176
5.3 Subjective TCM1 Experiment . . . 178

5.3.1 Experimental Setup and Procedure 178
5.3.2 Experimental Results 178
5.3.3 Discussion 179
5.4 Subjective TCM2 Experiment 179
5.4.1 Experimental Setup and Procedure 179
5.4.2 Experimental Results 181
5.4.3 Discussion 182
5.5 Chapter Summary 182

6 Conclusion and Future Work **185**
6.1 Summary of the Results 185
6.2 Limitations and Future Work 185

Bibliography **187**

List of Figures **207**

List of Tables **213**

Notation

Abbreviations

Abbreviation	Description	Definition
1D	One-dimensional	page 22
2D	Two-dimensional	page 29
3D	Three-dimensional	page 22
ANOVA	Analysis of Variance	page 104
CAD	Computer-aided Design	page 46
CNN	Convolutional Neural Network	page 132
CNS	Central Nervous System	page 5
DFT	Discrete Fourier Transform	page 30
DWT	Discrete Wavelet Transform	page 31
DL	Deep Learning	page 132
EP	Exploratory Procedure	page 9
FQ	Feature Quality	page 105
FSR	Force Sensitive Resistor	page 13
GNB	Gaussian Naïve Bayes	page 117
GO	God Object	page 22
HIP	Haptic Interface Point	page 22
JND	Just Noticeable Difference	page 5
k-NN	k-Nearest Neighbor	page 115
ML	Machine Learning	page 59
MLP	Multilayer Perceptron	page 132
MOS	Mean Opinion Score	page 128
MV	Majority Voting	page 109
NPM	Neodymium Permanent Magnet	page 51
PCA	Principal Component Analysis	page 108
PCB	Printed Circuit Board	page 49
PDF	Probability Density Function	page 117
PLA	Polylactic Acid	page 47
PWM	Pulse Width Modulation	page 161
ReLU	Rectified Linear Unit	page 133
SGD	Stochastic Gradient Descent	page 135
SLSM	Standard Linear Solid Model	page 22
SIBC	Semi-infinite Body-in-Contact	page 34
SRCC	Spearman Rank Correlation Coeff.	page 99
SVM	Support Vector Machine	page 114
TCM	Tactile Computer Mouse	page 154
UVFD	Ultrasonic Variable Friction Display	page 20
VCA	Voice Coil Actuator	page 18
VR	Virtual Reality	page 17

Symbols, scalars, and vectors

x	scalar value, variable
$\dot{x}$	derivative of x
$\|x\|$	absolute scalar value of x
$\hat{x}$	estimated/predicted value of x
$f(x)$	scalar function of x
$sgn(x)$	signum function of x
$sig(x)$	sigmoid function of x
$u(x)$	unit step function of x
$I(\omega)$	indicator function of event/condition ω
x[n], **x**	array (and short form notation)
x$[i]$	i^{th} entry of array **x**[n]
$\mathbf{x}_{\text{abs}}$	array holding the absolute scalar values of **x**
$\|\mathbf{x}\|$	euclidean norm of array **x**
$\overline{\mathbf{x}}$	mean of array **x**
$\sigma(\mathbf{x})$	standard deviation of array **x**
$\kappa(\mathbf{x})$	kurtosis of array **x**
$\mathbf{x}^{(l)}$	l−point moving average array of **x**
1	array filled with value 1
$< \mathbf{x}, \mathbf{y} >$	dot product of **x** and **y**
$\mathbf{X} = DFT1(\mathbf{x})$	1D discrete Fourier transform of array **x**
$\mathbf{D}^{(k)} = DWT_{D}^{(k)}(\mathbf{x})$	k^{th} detail level of 1D Discrete Wavelet Transform of **x**
$\mathbf{A}^{(k)} = DWT_{A}^{(k)}(\mathbf{x})$	k^{th} approximation level of 1D Discrete Wavelet Transform of **x**
$\mathbf{I}_{\text{g}}$	M-by-N-by-1 gray-scale image
$\mathbf{X} = DFT2(\mathbf{I}_{\text{g}})$	2D Discrete Fourier Transform of image $\mathbf{I}_{\text{g}}$
I	M-by-N-by-3 color image
$\mathbf{I}_{\text{adj}}$	adjusted contrast image **I**
$hist(\mathbf{I}_{\text{g}})$	histogram of $\mathbf{I}_{\text{g}}$

Chapter 1

Introduction

Haptic research addresses the human sense of touch, i.e., all physical properties such as forces, torques, or temperature differences, which humans perceive when interacting with objects [13]. A subdomain of this field, denoted as tactile haptics in the following, covers all challenges related to object material perception such as roughness or stickiness, and is of major interest for this thesis. Since the human skin is the largest organ, highly relevant applications will arise from advances in tactile haptics in the future. For example, sophisticated material scanners would allow for high-definition material classification and quality control, or, allow robotic systems an intuitive understanding of material properties for improved grasping applications. With respect to haptic display, high-definition tactile display systems would further enable augmented e-commerce shopping applications as well as immersive gaming. Likewise, enhanced controllers and joysticks with tactile feedback will improve existing teleoperation setups which currently rely on force and torque feedback only.

In order to understand why the tactile haptics field of research still has not reached the same level of sophistication like audio or image processing applications, an analogy to these domains is useful. For example, two typical tasks in the image processing domain are, on the one hand, image texture classification, and on other hand, the display of such images using computer monitors. Image textures generally have distinctive features such as the visual regularity or surface coarseness that enable humans to identify the material samples based on prior knowledge regardless of variations in observation distance or changes in ambient lighting. Moreover, these images are displayed and passively viewed by a human in a specific quality using a visual display system such as a computer monitor. Both of these tasks have reached high-definition quality during the last decades.

However, if the same tasks are transferred to address the sense of touch, state-of-the-art haptic scanners for material classification and haptic displays are not comparable to those devices from the image processing domain. The critical difference is the active physical interaction, which heavily complicates both the robust acquisition and the recreation of haptic data traces [14]. Recordings from real-world material samples based on a variety of sensors are required to evaluate how these material samples can be represented and also artificially be recreated in a computer system as illustrated in Fig. 1.1. This thesis investigates this research question and addresses the challenges listed in the following.

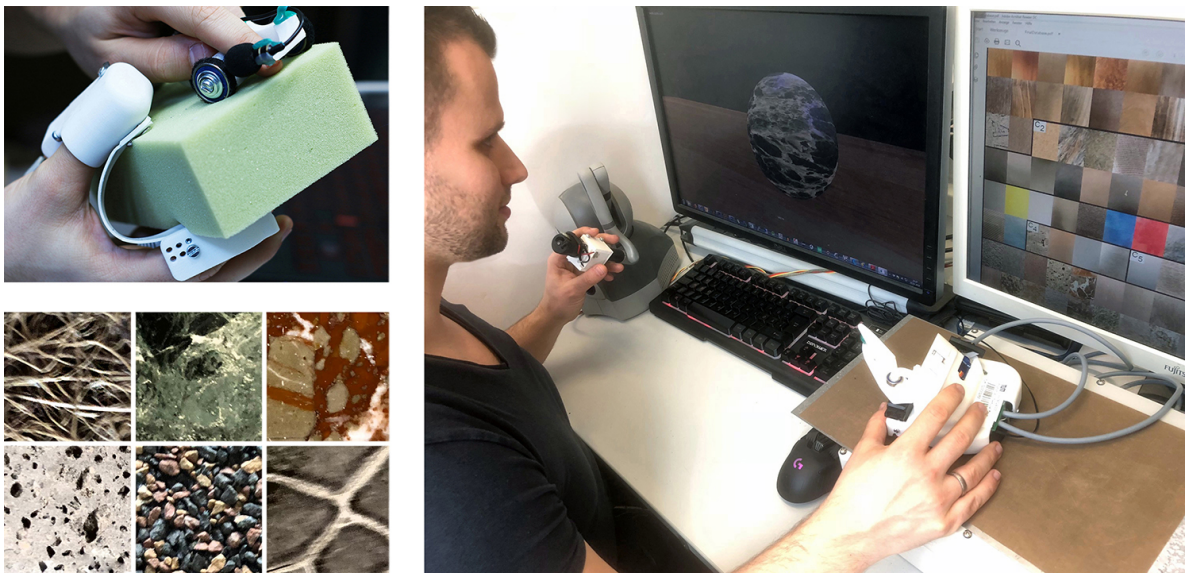


Figure 1.1: Haptic material acquisition, modeling, and display addresses the research questions of how to scan various material samples using novel recording devices (upper left), how to parameterize and store virtual material representations in a computer system (lower left and right), and how to recreate the touch sensations using novel haptic display devices (right).

1. There is no acquisition device available that can be used to robustly identify materials as well as to provide data traces for haptic display based on all relevant tactile dimensions. This point is addressed by the introduction of the Texplorer system [1], [2], [7] which comprises a complete device family consisting of different sensorized units.

2. Material classes are not uniquely labeled and an unambiguous taxonomy has not been reported so far. Consequently, only a few haptic databases have been recorded. During the creation of this thesis an interdisciplinary survey was conducted to establish a corresponding biological, geological, and chemical material taxonomy. A novel haptic database [15] is presented which contains 184 materials selected to cover a representative range of everyday material classes.

3. Whenever acquisition devices are wielded freely by human operators, the resulting data traces are susceptible to changes of scan-time parameters. Consequently, novel handcrafted features, deep learning-based approaches, and a fusion of these methodologies are presented to perform successful material classification in this thesis. The device design of the Texplorer units further targets to mitigate these dependencies.

4. The superposition of haptic impressions in a rendering device has been sparsely addressed so far. A novel grounded haptic device [3], [8] inspired by a computer mouse has been developed to jointly render all major relevant tactile dimensions.

Thesis Organization

This thesis follows the subsequent structure to address the aforementioned challenges.

- Chapter 2 covers relevant fundamentals of human haptic perception, tactile dimensions, state-of-the-art haptic acquisition, and existing display devices.
- Chapter 3 describes the acquisition of the haptic material databases and the properties of the novel recording devices.
- Chapter 4 covers the features designed in this thesis and their usage in a machine learning context for supervised classification.
- Chapter 5 presents the novel haptic device for the display of the tactile dimensions examined in this thesis.
- Eventually, Chapter 6 concludes this thesis and indicates possible directions for future research.

Figure 1.2 further summarizes the addressed fields of research in this thesis and highlights the corresponding sections in red font.

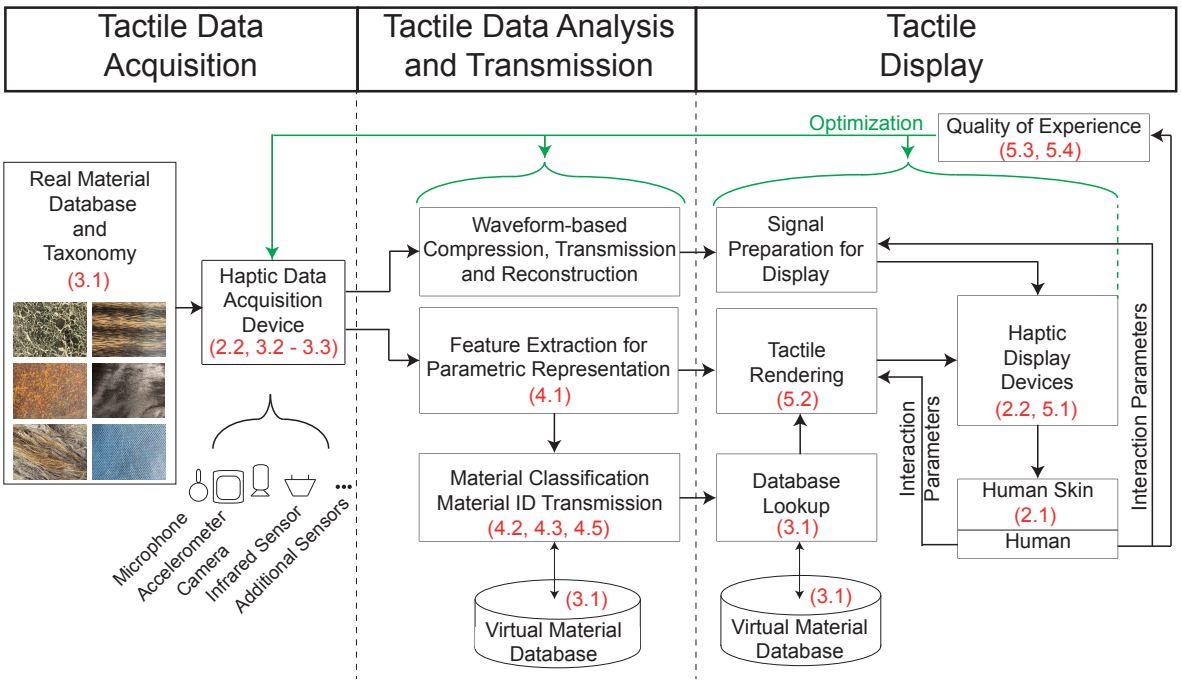


Figure 1.2: Overview of the studied haptic material acquisition, modeling, and display pipeline. Except the waveform-based compression, transmission and reduction part, all research areas are addressed within this thesis.

Parts of the work presented in this thesis have been published in international peer-reviewed scientific journals [1]–[6] and conferences [7]–[12].

Chapter 2

Background and Related Work

This chapter gives an introduction into the fundamental aspects of haptic research to enable an understanding of the topics discussed in the following sections. Haptics can broadly be categorized into **human haptic perception** (Section 2.1), **machine haptics** (Section 2.2), **computer haptics** (Section 2.3), and **haptic communication** (Section 2.4) according to the taxonomy introduced by Srinivasan et al. [16].

2.1 Human Haptic Perception

The human haptic perception system processes kinesthetic and tactile sensations passing sensory information to the central nervous system (CNS). These signals are combined with other modalities, like audition and vision, into an overall material perception [17]. The classification into the kinesthetic and tactile domains of haptic research is relevant from an engineering point of consideration since psychophysical parameters such as the temporal and spatial resolution or perceptual detection thresholds vary significantly and indicate different design and implementation guidelines for both acquisition and display device technologies. One relevant metric to quantify such thresholds is the just-noticeable difference (JND) κ, which was experimentally validated by Ernst Weber [18] and denoted as Weber fraction. It depends linearly on the ratio of change in stimulus intensity and absolute intensity as

$$\kappa = \frac{\Delta I}{I} \tag{2.1}$$

and has been studied for different physical parameters. These fractions are notably interesting in the context of haptic display, i.e., recorded signals and actuators need to possess a corresponding rendering resolution, which is provided by the majority of state-of-the-art sensors and actuators. For example, relevant Weber fractions are 7% – 10 % for forces between 0.5 N – 200 N [19], 15% – 27% for forces below 0.5 N [20], 15% – 22% for stiffness [21], and 19% -– 29% for viscous material perception [22], and for perceivable surface asperities amplitudes 5% – 12% [23]. The variety of these Weber fractions is related to the sensory input channels of the human skin, i.e., the receptors, which are discussed in the following.

Receptor	Sensitivity	Functionality	Related Perception
SA I Merkel	lower than 5 Hz	coarse texture, pressure	coarseness, hardness
SA II Ruffini	80 Hz – 500 Hz [24]	stretch sensing	friction, adhesion
FA I Meissner	8 Hz – 64 Hz	low-frequency vibration, stretch	coarseness, slippage
FA II Pacini	40 Hz – 1,000 Hz	high-frequency vibration	microscopic roughness
Thermal Receptors	5°C - 45°C	cold/warm sensing	thermal conductivity
Cold Nociceptors	lower than 15°C	painful coldness	thermal conductivity
Warm Nociceptors	higher than 45°C	painful warmth	thermal conductivity

Table 2.1: Cutaneous receptors in the glabrous (hairless) skin. Table adapted from [13], [24], [25].

2.1.1 Receptors

The kinesthetic sense, also known as proprioception, refers to the internal muscle, joint and tendon perception, and considers forces as well as torques acting on the human body. Primary and secondary spindle receptors are responsible for detecting the directionality and changes in muscle length, acceleration and speed, whereas Golgi tendon organs measure the forces generated by the muscles [20]. This thesis, however, mainly focuses on tactile sensations, which correspond to roughness, hardness, friction, temperature, as well as pain perception, and assist in many neuromuscular activities such as object identification while sliding the bare finger over the surface of an object [26], [27]. The majority of human studies have examined the mechano- and thermoreceptors located inside the hairless (glabrous) skin of the human hand [28], [29]. Fingertip tissue, for example, is densely populated with high-acuity receptors. These mechanoreceptors vary in their biological structure and distribution, but can be characterized by the two properties adaptation speed (slowly-adapting SA and fast-adapting FA) as well as by the size and sharpness of their receptive field (type I for small receptive fields, and type II for larger, blurry fields) in terms of vibrotactile perception. They are summarized in Table 2.1 and further explained in the following. Beforehand, it is worth highlighting that the mechanoreceptors are not associated with one particular tactile sensation, but cooperate to generate the neural signals that are fused by the CNS into human-made perception terms such as friction or roughness.

2.1.1.1 SA I - Merkel Disks

Slowly-adapting (SA I) Merkel disks are responsible for the intensity detection of curvature [30], edges [31] and coarse texture asperities [32]–[34] due to their large spatial acuity. They possess poor temporal resolution and are almost insensitive to vibrations [34], [35]. Weber et al. [34] used surface profilometry to verify that the SA I afferents are mainly activated within a hole diameter of embossed dots 2 mm – 5 mm. The smallest perceptible height difference was observed to be 2 μm, and a 10 - 15% Weber fraction is reported for the amplitude discrimination of sinusoidal gratings [36]. Hence, the Merkel disks mainly respond to surface structure, but also to hardness. Especially low-frequency hardness sensations such as static (smaller than 0.5 Hz), or partly, dynamic (between 0.5 Hz – 12 Hz) motions like pressing on, folding, pinching, or squeezing an object can activate the Merkel disks as shown by Weber et al. [34].

2.1.1.2 SA II - Ruffini Endings

Ruffini endings (SA II) sense skin stretch [37] and respond to the shear forces during incipient slip, i.e., the transition between static and dynamic friction, and presumably, adhesion during lifting a fingertip from a surface. SA II afferents are mainly distributed between the nails and fingers and are, consequently, most likely responsible for the sensation of stickiness, and hence, friction forces during texture exploration [38]. Tangential friction force has a Weber fraction of about 16% [39].

2.1.1.3 FA I - Meissner Corpuscles

The fast-adapting (FA I), or rapidly-adapting (RA I), Meissner corpuscles sense the rate of skin deformation within a 3 mm – 5 mm receptive field radius and are insensitive to static stimulation [37]. Their role in texture discrimination is not fully understood, since high-frequency vibrations are mainly perceived by the Pacinian corpuscles. However, they respond within a range of 8 Hz – 64 Hz [40] and stop firing when the skin movement stops, even under strong indentation [24]. Consequently, they sense low-frequency shear forces as well as fluttering sensations, i.e., lower frequencies during active sliding motions. The highest RA I response resolution of spike patterns is 4 ms according to Weber et al. [34]. Beside their role of low-frequency sensing, the Meissner corpuscles and Merkel disks neural activations are fused to identify coarse texture patterns during slow sliding. As the resulting device design guideline, coarse texture patterns need to be scanned and represented based on both spatial information and temporal texture patterns in order to address the Merkel disks and the Meissner corpuscles, respectively. Note that vibration Weber fractions are discussed in the next paragraph.

2.1.1.4 FA II - Pacinian Corpuscles

Pacinian corpuscles (FA II) respond most strongly to high-frequency vibrations [41] within 40 Hz – 1,000 Hz and exhibit a U-shape sensitivity curve with a peak at about 250 Hz [40]. By contrast, they are completely insensitive to static indentations and reveal almost no spatial acuity [42] since their receptive field is blurry across the human hand. They mainly detect high-frequency skin accelerations during sliding and high-frequency hardness information during tapping on material surfaces. This observation led to one of the fundamental device design guidelines of using accelerometers for vibration signal collection during, e.g., sensorized tool interactions [43]–[46].

Mahns et al. [47] measured frequency discrimination Weber fractions from 30% for 20 Hz sine wave signals to 14% for 200 Hz sine wave signals, respectively. Phase differences are well distinguished for lower frequencies (10 Hz – 40 Hz), but worse for higher frequencies, e.g., 100 Hz – 300 Hz [48]. The amplitude detection ranges from 13% – 16% (20 Hz – 300 Hz) [49]. The temporal resolution between two tactile touch events has been inferred to be about 2–5 ms during subsequent tappings on the human skin [34], [50]. In combination with the other mechanoreceptors, a display device representing sliding-based vibrations needs to cover a frequency range of 0 Hz – 450 Hz $\pm$ 50 Hz, depending on different human subject

thresholds. On the contrary, a data acquisition system may scan higher frequency ranges, but has to exclude lower frequencies less than 12 Hz in order to remove the motion-dependent components of the scanning system, e.g., the human operator hand [28].

2.1.1.5 Thermal Receptors

Thermal receptors like afferent fibers and thermal nociceptors [51] encode cold or warm stimuli of noxious temperature ranges below 15°C or above 45°C, respectively [52]. The temperature of the human skin ranges within 25°C – 36°C [53]. Ambient temperatures generally fall below these thresholds, and hence, object surfaces drain heat from the human finger depending on their thermal properties. The resulting sensation of coldness corresponds to the interface temperature, i.e., the finger pad temperature determined by the object surface temperature and the initial human skin temperature. Heat insulators or other insulating materials do not noticeably change this temperature level. Note that humans exhibit a neutral thermal zone of 30°C – 36°C where no thermal sensations are noticed for areas about 1,500 mm^2 [29], giving further design simplifications for haptic displays of such insulating materials. Outside this range, either the cold or warm afferent units reveal specific activity and need to be addressed correspondingly. The Weber fraction is remarkably small (0.5 % – 2 %) [54] for parts of the human hand (e.g., the thenar eminence), i.e., the base of the thumb is highly sensitive to changes in temperature of already 0.11°C (cooling) and 0.22°C (warming) at a rate of about 2°C/s. However, the index finger reveals a lower sensitivity, i.e., 0.5°C for warming and 0.3°C for cooling perception, and consequently, is easier to target for thermal display approaches. Cold afferent fibers conduct their signals ten-times faster than warm afferent units, i.e., with 10 m/s – 20 m/s [29] which gives further design indication for a thermal haptic display; if a thermal display can render about 3°C/s and more, the perceptual thresholds of humans [29] are perfectly covered.

2.1.1.6 Mechanical and Chemical Nociceptors

Mechanical and chemical nociceptors respond to excessive mechanical force (pain) or exposition to chemical ingredients such as capsaicin [55], respectively. Psychophysical behavior studies have shown that humans strongly avoid the sensation of potential pain; as a consequent design guideline, a haptic device should never expose a human subject to such a perception. Vibromechanical actuation principles generally work without sharp edges and are hence not harmful, whereas other device technologies based on electrical stimuli (see Section 2.2.3) may evoke electric shock sensations, and consequently, pain-related avoidance behavior [56] which can lead to the situation that the human user rejects to interact further with the device. This relevant observation supports the choice of vibromechanical actuation used in this thesis.

2.1.1.7 Hygroreceptors

The human fingertips do not contain hygroreceptors [57], [58] which enable the sensation of surface moisture, humidity, or wetness. However, humans perceive such effects as a com-

bination of thermal and tactile cues as well as during either static or dynamic surface interaction. Consequently, the human thermo- and mechanoreceptors covering slip detection need to be addressed to artificially recreate the sensation of surface moisture as examined by Shibahara et al. [59]. Note that this is one example of a typical haptic illusion (like the cutaneous rabbit [60]) which generally is exploited to simplify haptic device design. Nevertheless, this perception is not relevant for the majority of naturally occurring material samples, and hence, out of scope of this thesis.

2.1.2 Human Exploratory Procedures and Haptic Dimensions

The haptic perception of an object and its surface properties is a complex interaction of major relevant kinesthetic and tactile dimensions during exploratory procedures (EPs), i.e., stereotyped patterns which are described in the following. These patterns are notably relevant for this thesis since sensorized system design is supposed to mimic potentially descriptive actions, and likewise, a tactile display is supposed to match such patterns during haptic display.

Experiments with human participants have shown that humans use a broad range of adjectives to characterize their tactile sensations during object identification or recognition tasks. In terms of cutaneous perception, Hollins et al. [61] and Bergmann-Tiest [62] identified the four major tactile dimensions hardness, roughness, warmth, and friction. Okamoto et al. [63] further summarized a list of generally used tactile adjectives from such previous studies and derived five major perceptually relevant tactile dimensions based on Hollin's four dimensions with roughness being divided into macroscopic and microscopic roughness according to the duplex theory proposed by Katz [64]. Active discussion about the final feature space describing all relevant tactile dimensions [65] show that there is no common understanding on how to categorize these dimensions. As a promising attempt on standardization, Fishel et al. [66], [67] proposed a feature space originating from the major five tactile dimensions, which is denoted as SynTouch subdimensions, or features, in the following, and is shown in Fig. 2.1. The original definitions of the subdimensions from [67] are listed in the following.

- **Macrotexture** (mTX): The intensity of large features ($>$ 1 mm spacing) that creates the perception of texture ranging from smooth to textured.
- **Macrotexture Coarseness** (mCO): The perceived spacing of large features ($>$ 1 mm spacing), ranging from fine to coarse.
- **Macrotexture Regularity** (mRG): The perceived uniformity of large features, ranging from random to regular.
- **Microtexture Roughness** (uRO): The intensity of small features ($<$ 1 mm spacing) that creates the perception of roughness ranging from smooth to rough.
- **Microtexture Coarseness** (uCO): The perceived spacing of small features ($<$ 1 mm spacing), ranging from fine to coarse.

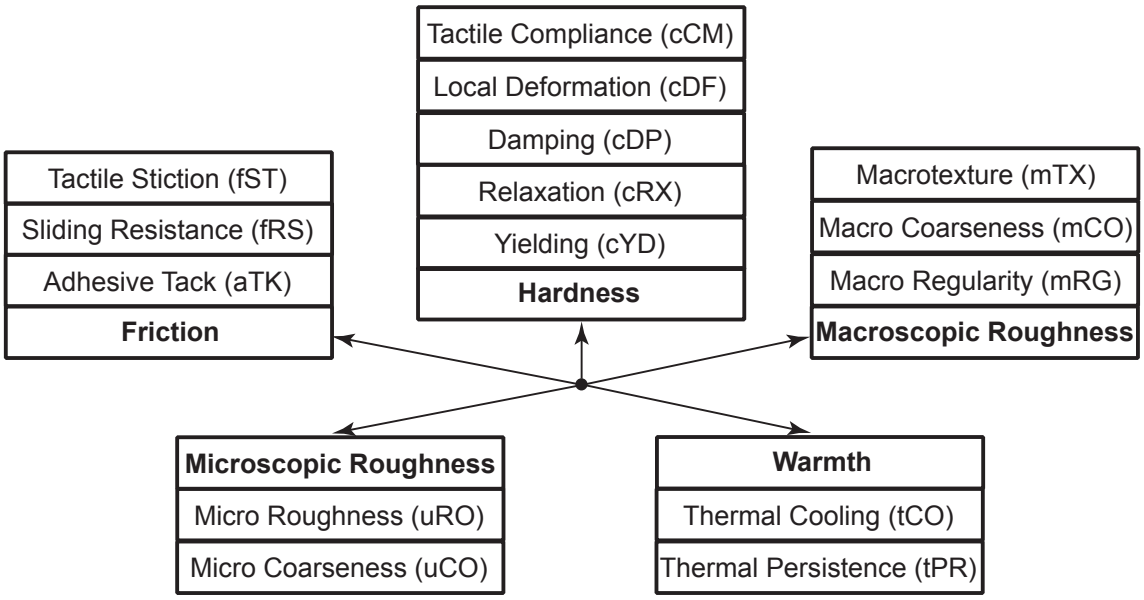


Figure 2.1: Five major tactile dimensions containing the 15-subdimensional SynTouch feature space. This thesis mathematically defines and implements these features for acquisition [2] and display [3]. The original naming convention is adopted with slight changes to credit the initial thoughts from Fishel et al. [66]. For example, cCP is changed to CCP^{TUM} in the following.

- **Tactile Stiction** (fST): The effort required to initiate sliding on a surface, ranging from low grip to high grip.
- **Sliding Resistance** (fRS): The effort required to continue sliding over a surface, ranging from slippery to sticky.
- **Adhesive Tack** (aTK): The effort required to break contact with a surface, ranging from no adhesion to sticky.
- **Tactile Compliance** (cCM): The degree that a surface deforms under pressure, from rigid to compliant.
- **Local Deformation** (cDF): The degree to which the surface wraps around the fingertip when being deformed, ranging stays flat to high wrap.
- **Damping** (cDP): The speed that a surface returns to its original shape after being deformed, ranging from springy to damped.
- **Relaxation** (cRX): The degree to which a surface stops pushing back after being deformed, ranging from maintains force to relaxes.
- **Yielding** (cYD): The degree to which a surface remains deformed after being pressed, ranging from recovers shape to stays deformed.
- **Thermal Cooling** (tCO): The initial rate that a surface draws heat from the fingertip, ranging from warm to cool.

- **Thermal Persistence** (tPR): The extent that a surface continues to draw heat from the fingertip, ranging from transient cooling to sustained cooling.

These tactile subdimensions are associated with the EPs in this thesis as discussed in the following. Note that kinesthetic dimensions are mentioned as well in order to give a complete overview about the haptic dimensions.

2.1.2.1 Enclosure

Before an object is grasped, humans rely on visible clues to create an understanding of its **volume** and further estimate tactile properties such as coarse structures by visual inspection [13], [68]. During the subsequent initial touch interaction, the forces and torques acting on different parts of the human body, e.g., on the fingers while grasping an object, finalize the kinesthetic understanding of the object's global volume.

2.1.2.2 Lifting

The existence of gravity leads to the kinesthetic impression of **weight** (in N), or **mass** (in kg), while lifting and holding an object. Weight can be perceived passively while holding an object, however, experiments have shown that active lifting and wielding motions increase the ability of estimating the mass of an object [69].

2.1.2.3 Static Touch

The initial contact with the material sample allows for warmth-related information, which is identified within the first seconds of passively holding an object. **Thermal cooling** (tCO) describes the initially perceived temperature gradient in °C when touching the surface of an object, and the ongoing drain of thermal energy over time is further covered by the **thermal persistence** tPR. The relevance of these thermal properties has been underscored compared to the other tactile dimensions, however, materials like glass and steel can already be discriminated by their different thermal conductivities as shown by Ho et al. [70] and Bergmann-Tiest et al. [71] absent visible information.

Another tactile subdimension, namely adhesive tack (aTK), is connected to the end of the static touch phase, i.e., during the release of the fingers from the surface [67] and is determined by the counteracting normal force between the surfaces of the finger and the material sample [66], [72].

2.1.2.4 Pressing

Dynamic tapping and continuous pressing motions on an unknown surface are cognitively linked to the stiffness of a material sample. The umbrella term for this tactile dimension is **hardness**, which corresponds to the area-based deformability under exerted pressure [73]–[75]. Hardness is considered a major tactile dimension [63], however, splits up into several perceptual subdimensions. For example, the SynTouch subdimension **compliance (cCM)** can be associated with the Young's Modulus (in $\mathrm{N/m^2}$) or the spring stiffness k_s (in $\mathrm{N/m}$).

Local deform (cDF) can be understood as a subdimension of compliance and relates to the degree of how strong or how likely the material wraps around the scan system or the human finger. **Damping (cDP)** denotes the temporal change of pressure force, and hence corresponds to the damping coefficient η. **Relaxation (cRX)** quantifies the counteracting force while actively pressing on a material sample, and can be represented as a fraction of the maximum pressure force. Lastly, **yielding (cYD)** describes the object state after interaction; referring to its original Boolean definition it either indicates a permanent deformation or the potential of the material sample to fully recover its shape.

2.1.2.5 Contour Following

The last two exploration patterns are contour following and sliding which are similar in their movement execution, however, put the focus on the different tactile dimensions **macroscopic roughness** for contour following and **microscopic roughness** and **friction** estimation during sliding.

Notably, the understanding of surface roughness has received great attention during the last decades [26], [48], [64], [76]. Most prominently, the duplex nature of roughness, consisting of microscopic and macroscopic roughness, is supported by different works like [76]–[79]. The perceptual threshold between macroscopic and microscopic roughness has been determined to be approximately 200 μm [79], and the upper bound of macroscopic roughness has been reported by Hollins et al. [77] to be within 3 mm – 8 mm. Weber et al. [34] show that embossed dots are best identified by the SA I afferents within 2 mm – 5 mm, and hence, can be associated with macroscopic roughness perception and contour information. Texture elements far above these thresholds are related to the global shape. A clear threshold has not been determined yet and further highlights the vague borderlines between kinesthetic and tactile research.

The SynTouch subdimensions **macrotexture (mTX)** corresponds to the perceived height, or analogously, perceived depth of texture structure elements, whereas **macrotexture coarseness (mCO)** relates to the surface's particle size and spacing, and **macrotexture regularity (mRG)** covers the spatial arrangement rule of surface patterns. Note that macroscopic roughness is partly identifiable by static touch alone, but active contour following aids in its certain identification.

2.1.2.6 Sliding

Sliding motions are similar to the aforementioned contour following EP, yet with the emphasis on the perception of roughness and friction. Notably, **microscopic roughness** sensations result from high-frequency vibrations during either surface-tool or surface-finger sliding actions [26], [64], [78], [80]–[82] and are covered by the SynTouch dimensions **micro coarseness (uCO)** and **micro roughness (uRO)** with the perceptual ranges of 8 Hz – 64 Hz (Meissner) and 40 Hz – 1,000 Hz (Pacinian), respectively [24], [25]. The perceived bare-finger roughness has been determined not to be a simple function of any mechanoreceptor alone, i.e., the complexity of roughness seems to cover all mechanoreceptor ranges [24]. Note that thermal and

roughness perception is strongly linked [83] and surface temperatures beyond the range of 5° C - 45° C makes roughness sensation imperceptible [29].

Object surfaces further cause the perceptual dimension of **friction**, or slipperiness [63] which compel the user, either holding a tool or using bare-finger interaction, to apply a different tangential force during the sliding motion over a surface. In this context, the SynTouch subdimensions **tactile stiction (fST)** and **sliding resistance (fRS)** can be understand as static (adherent) and kinetic (sliding) friction coefficients μ_0 and μ_k, respectively. Surface friction was initially understood as subdimension of roughness with strong positive correlation, but several studies such as the one performed by Skedung et al. [84] were able to nullify this hypothesis. Hence, friction is considered as an independent tactile dimension. Also, the transition (incipient slip) between sticking and sliding motion was found to be independent of the existence of surface asperities (gratings, particles, etc.) according to Srinivasan et al. [85]. Another critical influence on the surface friction is given by its hydrophilic or hydrophobic character [86], [87] and explains the significant friction coefficient difference of likewise smooth surfaces, e.g., porous glass versus polypropylene.

2.2 Machine Haptics

The field of machine haptics covers the hardware used to collect and to convert haptic data into signal representations which are used to generate computer-controlled signals addressing the human kinesthetic and cutaneous perception system. The corresponding umbrella terms in this context are haptic data acquisition and haptic data display.

2.2.1 Data Acquisition Devices

Two different approaches exist to capture the aforementioned tactile dimensions which are determined by the two major scan dependencies scan speed and scan force. A material scanner can either be wielded by a human operator in an unconstrained manner, or, be precisely controlled by a robotic scanning system. Single-sensor approaches that capture individual tactile dimensions have been examined initially by related work [44], [88]; then, handheld systems combining multiple domains were presented [7], [89], [90]. During the progress in automation and control engineering, such systems have been further miniaturized and equipped to more complex robotic scan systems in parallel. Nevertheless, the sophistication in controlled position or force schemes does not guarantee superiority to manually-wielded systems for material classification, and hence, both approaches are discussed in this thesis. To start with, the required individual sensors to capture dimension-specific haptic data are explained in the following.

2.2.1.1 Friction

The friction properties of an object surface, e.g., the static and kinetic friction coefficient, are typically recorded using optical force sensors [91] or force sensitive resistors (FSRs) [9] which are either in direct surface contact or are laterally attached to the sensing device. Tribome-

ters, as used by Colgate et al. [92], further allow higher resolution scans of such friction measurements and are used to examine slip-stick phenomena such as stick-slip chatter.

Surface **adhesion**, which correlates to friction in general, has been rarely addressed in the context of either haptic data acquisition or display. Spinner et al. [72] report that bare-finger adhesion forces have never been quantified before, and examined the range of finger adhesion forces under natural conditions, which resulted in a narrow force range of 2.3 mN – 167.6 mN depending on a normal load range from 0 mN – 1,000 mN. Despite using a high-precision force sensor, the observed data spread is heavily dependent on the different surface microstructure, the moisture content, or the effective skin areas of the human participants. As one alternative data acquisition option, pure subjective ratings can be averaged across a broad range of different human subjects based on n-point Likert scales as performed in [2].

2.2.1.2 Hardness

The sensation of hardness is associated with area-based deformations, and hence, forces acting on a sensing element. Force sensitive resistors and optical force sensors are common ways to quantify this dimension during touch interaction. Note that the different underlying EPs of hardness exploration, mainly tapping and pressing, require different sensing approaches. For example, **low-frequency hardness** is generally measured using normal force measurement values [91] that are divided by the indentation depth values according to Hooke's law. This concept is used for measuring rigid body hardness estimation, e.g., according to the Brinell or Vickers hardness metric [93], or elastic compliance, e.g., by using the Shore hardness level measured with durometers as used by Aujeszky et al. [94].

By contrast, **high-frequency hardness** is perceived by tapping on the material sample and is generally measured using accelerometers as extensively examined by Okamura et al. [95] and Kuchenbecker et al. [96].

2.2.1.3 Macroscopic Roughness

The first macroscopic roughness acquisition techniques in haptic research used the pixel intensities of gray-scale images as the data source to create a surface height profile [97]–[100]. Macroscopic roughness information has also been inferred using stereoscopic [101] or distance-based measurements, and additionally by evaluation of low-frequency components of accelerometer-based vibrotactile signals. Alternatively, Dulik et al. [102] use infrared light during non-contact scans to infer the surface height profile.

2.2.1.4 Microscopic Roughness

The acquisition of fine roughness data is linked to dynamic sensor – surface interaction. In this context, force sensors are generally too large, fragile, and need to be mounted between two mechanical components to provide useful measurements whenever high-frequency vibration analysis is required. Acceleration sensors are preferable instead and have been first used by Howe et al. [43] in their work on dynamic texture sensing in 1989. Small, robust, and affordable 3-axis micro-electrical-mechanical systems (MEMS) accelerometers were rigidly

mounted to tools as demonstrated by de Boissieu et al. [44] and McMahan et al. [45]. Data-driven models [14], [91], [103]–[106] are proposed which collect large amount of vibrotactile data depending on different pairs of scan speed and scan force. Note that the primary purpose of these approaches is not to identify the material class of a given material sample, but to parameterize the roughness information for later realistic tactile display. Beyond using tool-mediated accelerometer scanning setups, other researchers collect the bare-finger vibrotactile signals and examine the propagation through the human skin as done by Tanaka et al. [107], Sano et al. [108], Visell et al. [109] and Vardar et al. [110]. Such bare-finger recordings are intuitively relevant for the haptic display of vibrotactile signals, and hence, relevant for this thesis as well.

2.2.1.5 Warmth

The identification of surface temperature changes is relevant to quantify the dimension of warmth. Typical sensors employed in this domain are thermistors or thermal IR cameras as used by Ho and Jones [29], [70], Aujeszky et al. [94], or thermal flux sensors as used by Choi et al. [111].

2.2.2 Comparison of Robotic and Human Operator Scan Approaches

The sensors mentioned in the previous section are typically combined into multimodal scanners to enable the simultaneous acquisition of multiple tactile dimensions. One of the domains, which increased the systematic need for touch-based data, was the identification of surface properties aiding robotic manipulation tasks. Notably, the introduction of biomimetic tactile sensors to the field of robotic grasping accelerated the research field of material surface analysis. Tada et al. [112] introduced a tactile finger containing one strain gauge and several polyvinylidene flouride (PVDF) films to measure pressure and fine vibrations, respectively, and first successfully distinguished between paper and wood samples. Several subsequent classification approaches have been proposed by Oddo et al. [113], Sinapov et al. [114], Jamali et al. [115], Fishel et al. [66], [116], and Romano et al. [117]. Notably, Fishel et al. introduced the BioTac sensor [116], [118] which combines multiple tactile sensing domains in a miniaturized system design and is frequently used by other researchers in the field. The BioTac sensor is able to extract pressure distribution information based on 19 electrodes, thermal flow based on a thermistor embedded in a liquid-filled part of the sensor, and static force as well as vibrotactile signals based on a pressure sensor. Its capability of material identification even exceeds human performance as proven in [116] and further supports the idea that the acquisition system does not necessarily need to match human perception ranges and limits for the specific task of material classification. The BioTac can be mounted at a linear stage to measure traction forces as done in [116], or, is mounted to robotic hands and grippers as done by Hoelscher et al. [119] or Kaboli et al. [120]. The most recent state-of-the-art scanning platform is the BioTac Toccare, which allows for a controlled scan of a predefined material sample to assess its tactile properties and can be considered as the gold standard for automated planar tactile sensing. The level of sophistication in the design of

the sensor and the controlled scan conditions, however, led to the fact that the majority of proposed features, discussed in Section 4.1.3, are less complex in their definition, and hence, depending on the scan conditions.

The operator-wielded Texplorer sensor [1], [7] introduced in this thesis constitutes an alternative direction of research. It also combines several sensors working together in terms of material classification, is low-cost, and easily extendable. The following arguments show that both approaches have their advantages and provide solutions for different applications.

1. Robotic systems can reproduce desired recording procedure patterns and generate more stationary material interaction signals on planar material samples. A human operator-based approach instead reveals a specific variability in the two parameters scan speed and force. For example, a sensorized tool generally is moved within a speed range of about 20 mm/s up to 240 mm/s and a force range of 0.2 N - 3 N as shown by Culbertson et al. [91]. However, a human operator can easily adapt the scan procedure to complex object geometries and freely scan any material sample without prior preparation of the scanned object. Humans can evaluate much faster where and how a scan should be performed, especially for very complex object geometries. Observations of Smith et al. [121] further show that human subjects are relatively consistent in reproducing exploratory speed and force within a typical range of 10 mm/s – 100 mm/s and 0.2 N – 2 N scan force. By contrast, robots likewise may not be able to keep contact forces constant for uneven or heavily structured material surfaces.

2. Robotic solutions require great effort to achieve comparable human movement precision and to reduce their motor vibrations which affect the accelerometer recordings during the robot-surface interaction. The cost and effort to set up and run such a system are significantly larger than a human-operated scanner. This point might be leveraged by cost scalability in the future, however, similarly applies for operator-wielded systems.

3. Whenever haptic data acquisition for haptic display comes into consideration, a human-operator-based approach becomes favorable in terms of the most suited tool tip for data acquisition, namely, the human finger. If any other tool tip material is used, e.g., plastic, rubber, or steel as commonly done in robotic setups, the collected data traces can not be used for haptic rendering of vibration signals at the human skin without complex and nonlinear adjustments. Moreover, human skin generally recovers from rough material interaction, whereas the sensor material silicone of the BioTac is prone to abrasion and needs to be replaced after several scans [90].

2.2.3 Display Devices

Haptic display devices are classified into grounded and wearable devices [122] depending on a rigid link to the environment, their effective workspace, mobility, and the methodology of routing signals and supply cables. Several prototypes for the latter approach have been presented recently such as the Dexmo project from Dexta Robotics [123], the Wolverine and

Grabity glove from Choi et al. [124], [125], or the DextrEs glove from Hinchet et al. [126]. Alternatively, the Ultrahaptics concept [127] is used to recreate mid-air sensations for VR applications using finger tracking techniques and aims to reproduce various shapes of virtual objects in front of a user. However, rich tactile feedback including all relevant dimensions, notably the parallel display of convincing hardness, roughness, and thermal sensations rendered at the glove fingertips still pose an unsolved challenge for all current wearable devices. Consequently, this thesis does not further discuss specific wearable approaches; the comprehensive survey of Pacchierotti et al. [128] gives a broad overview about the current state-of-the-art wearable haptic technologies.

Grounded haptic devices have been frequently used both for research and commercial applications and are most prominently represented by the haptic devices Sigma.7, Omega.3, or Delta.3 (Force Dimension, Switzerland), and are denoted as FD devices in the following. Note that preceding versions like the well-known and widely used PhantomOmni and Falcon devices are also comprised by this abbreviation. These kinesthetic force feedback devices can be extended with further actuation components. Alternatively, novel grounded devices can be specifically designed to include all tactile dimensions which have been sparsely addressed so far. Section 5.1.2 presents such a novel display device and discusses the reasons for this choice with respect to the rendering of the 15 SynTouch subdimensions. In the following, general actuation principles for the individual major tactile dimensions are introduced beforehand.

2.2.3.1 Friction Display

The FD devices like the PhantomOmni are commonly used in haptic research to render different friction forces to the stylus held by the human hand [91], [106], [129]. Other custom-made devices in [130] change the perceived friction between a tool and the underlying material by applying braking forces to a rolling steel ball, or, use electromagnets and iron plates [3], [5] to simulate surface friction.

2.2.3.2 Hardness Display

Similar to friction forces, low-frequency hardness impressions have been displayed using the earlier FD devices like the PhantomOmni [100], [131]. Additionally, Bianchi et al. [132] presented a custom softness display device which is able to change the perceived softness and fine roughness of displayed surface textures, or, Young et al. [133] introduced a device enhancing this concept by providing force and torque display at human fingertips using servo motors. Whenever hardness feedback is targeted at the human fingers, servo motors generally outperform the traditional FD devices in locally applied feedback strength, and are hence considered further in this thesis in Section 5.2.4.

2.2.3.3 Macroscopic and Microscopic Roughness Display

The force feedback capabilities of the commonly used FD haptic devices are also suitable to render low-frequency macroscopic roughness and contour information. Likewise, linear

actuators [134] and servo motors [128], [133] are generally included to generate physical displacement, indentation, or material contour impressions depending on the current height map value of a corresponding image representing such values.

By contrast to the efficient display of macroscopic roughness, state-of-the-art haptic devices reveal low-pass signal transmission characteristics [135], and hence, cannot display the high-frequency components of vibrotactile signals [136] required to address the Pacinian corpuscles. The resonant frequency of a pure oscillator (i.e., only modeled by spring stiffness) is

$$f_{\mathrm{r}} = \frac{1}{2 \cdot \pi} \cdot \sqrt{\frac{k_{\mathrm{s}}}{m}} \tag{2.2}$$

with the spring stiffness parameter k_{s} and the overall system mass m comprising the device and human hand mass. Considering the Phantom Omni device with a maximum stiffness of $k_{\mathrm{max}} \approx$ 1,000 N/m and an end-effector handle mass m = 0.1 kg, the resonant frequency is about 15 Hz and hence not sufficient for displaying high-frequency roughness feedback. The motor drive circuitry and friction in the devices limits their ability to accurately reproduce high-frequency vibrations, meaning that virtual surfaces often do not include fine texture, and hence, feel smooth and slippery [135]. However, several tactile display technologies can augment such traditional force feedback devices in order to superimpose vibrotactile information. The three major trends in current research are based on vibromechanical actuation, electrovibration, and ultrasonic actuation, and are shortly discussed in the following.

Vibromechanical Actuation Mechanical oscillatory displacement is the most frequently encountered vibrotactile display concept in both research and commercial applications. The survey conducted in [134] gives a broad overview about the current vibromechanical actuators which come in different implementation forms like linear electromagnetic actuators, eccentric rotating mass motors, piezoceramic actuators, pneumatic actuators, or tactile pattern displays.

Linear electromagnetic actuators use an alternating current (AC) source to generate a magnetic field with likewise oscillating orientation [137]. If the inner coil encloses a movable and ferromagnetic material, the actuator is called a solenoid; if it encloses a permanent magnet it is a voice coil actuator (VCA). State-of-the-art devices have a minimum size of about 10 mm and typically use a spring to return the moving element to its resting position in the case of the solenoid actuator. This combination of spring and moving mass has a natural resonant frequency, at which the input is strongly amplified [134], [138], [139]. The VCA ideally has built-in low-friction linear bushings instead to compensate this effect. Still, external springs need to be mounted to align a VCA properly. If the entire device casing is supposed to vibrate, the VCA is a suitable solution due to its highly relevant ability to represent a broad vibrotactile spectrum between 0 Hz – 1,000 Hz. This configuration is, e.g., used by Kuchenbecker et al. [14] for their tool-mediated acquisition and rendering setup. If the vibrotactile feedback is targeted for the human fingers, several companies manufactured VCAs in different form factors to avoid the necessity for external assemblies. A representative example is the C2 tactor (Engineering Acoustics, USA) which is highly suitable for this specific task of direct bare finger feedback, and hence used in this thesis for vibrotactile rendering.

Rotatory electromagnetic actuators generate vibrations when a direct current (DC) is applied. They are widely used in common electronic devices, however, cannot be used to render arbitrary combinations of frequency, amplitude and direction since the frequency and amplitude are linked due to the rotated mass, i.e., the inertia of the system [134]. Another significant disadvantage is the longer spin-up time up to 3 ms even when enhanced with specific control algorithms [138]. Compared to linear actuators, such a response time makes them inappropriate to render high-frequency vibrations to address the Pacinian corpuscles.

Non-electromagnetic actuators such as ceramic piezoelectric actuators generate arbitrary wave forms, have short responding times [134], and hence, are an alternative to VCAs. Their major drawback, however, is the more challenging system integration and more complex voltage supply. Another type of actuators are metallic shape memory alloys that change their shape according to the current temperature. Although these actuators have a high power-to-weight ratio, they require much energy and have a large hysteresis [134], [140]. Lastly, pneumatic actuators (air or fluid) are compact and light-weight, but cannot output high-frequency vibrations due to their inherent low-pass characteristics.

Electrovibration Mallinckrodt et al. [141] discovered the electrovibration effect when touching the brass socket of a lamp during operation, which felt vibrating, less smooth, and perceptually comparable to be covered in a layer of resin due to the attached 110 AC voltage. The periodic change of voltage in a capacitance-like configuration let fingertips vibrate depending on the applied voltage signal during active motions. The TeslaTouch display [142] is the first setup that used this effect directly on a touch screen, and ever since, electrovibration has become a major research trend for microscopic roughness rendering. The input data was initially based on images only, comparable to the macroscopic rendering approaches like the sandpaper system from Minsky et al. [99] or the approach from Wu et al. [143]. Samur et al. [144] propose to use prerecorded accelerometer data, and hence, vibrotactile signals, to drive such displays, and Colgate et al. [145]–[147] extensively investigated electrovibration properties and dependencies. The commercially available Tanvas Electrostatic Tablet Device [148] TanvasTouch currently represents the state-of-the-art for electrovibration-based rendering. Despite the high-resolution of electrostatic displays, and hence, the ability to precisely generate high-fidelity vibrotactile signals, several disadvantages pose challenges that make the use of vibromechanical actuators more preferable in this thesis as discussed in the following.

1. Humans generally do not perceive the same sensation similarly. Differences in magnitude or complete lack of haptic feedback individually results from each users' finger skin composition and dermal properties [145]. Influences like skin dryness and thickness as well as medical conditions affecting the skin cause strong variances in each individuals touch sensation [149] which strongly reduces the vibrotactile reproducibility of electrovibration displays.

2. The trade-off between user safety and vibrotactile perception is another concern for these displays. The insulation layer, which is several microns thick, is the only iso-

lation between the user's hand and up to 1 kV in the first prototypes [145] in order to generate a significant effect. As a consequence to the higher voltages, the current is generally limited to 0.5 mA, which is below a critical human threshold [142]. Recently, the output voltage applied to the touch surface is limited to the range of 80 V – 115 V peak-to-peak. This necessary limitation, however, noticeably decreases the overall intensity of the fine roughness sensation. However, varying fingertip wetness and properties between users still can lead to disturbing electric shocks which may evoke pain-related avoidance behavior as mentioned in Section 2.1. Consequently, the user may refuse to touch the device again at all. On the other hand, mug or dirt can completely dampen the roughness sensation, and hence, results in imperceptible rendering.

3. Electrovibration strictly requires active motion to work, which complicates the design and form factor of the tactile device, notably when combined with the other tactile dimensions.

Ultrasonic Variable Friction Displays Ultrasonic variable friction displays (UVFDs), also known as variable friction devices, use a vibrating glass plate to cause over-pressure, which leads to a reduction of the friction coefficient known as squeeze film effect [150], [151]. This effect is used to generate spatial haptic illusions of textures and shapes, comparable to the electrovibration effect. By reducing the friction force between the display and the fingertip, the user gets the impression of a more slippery surface. Based on the traced position of the fingertip, various textures and shape illusions can be generated [152], [153]. UVFDs can be compared to electrovibration devices in terms of high-resolution roughness display, however, share similar disadvantages in terms of the users finger constitution (medical condition, dirt, or wetness) which strongly influence the perceived effect. Further note that the maximum displayable friction coefficient is determined by the used glass layer. As mentioned in Section 2.1.1, glass is hydrophilic, and hence, the friction display heavily depends on different human fingers, which is noticeably critical for UVFDs compared to electrovibration setups.

The disadvantages of electrovibration and ultrasonic approaches make the usage of vibromechanical elements preferable, which are hence used as the primary display technology in this thesis. Note that vibrotactile actuators potentially are miniaturized in future research, and hence, making them even more suitable and versatile as tactile display actuators.

2.2.3.4 Thermal Conductivity Display

Peltier elements are commonly used to display thermal effects in [70], [111], [154]–[158] and in the thesis-related publications [3] and [5]. If the design of the thermal element does not allow any additional sensor placements [5], Peltier elements can be driven with empirical voltages to achieve an approximate target temperature. Otherwise, the surface temperature of the Peltier element can be precisely controlled in a closed-loop approach based on surface-mounted thermal sensors [111], [158].

2.3 Computer Haptics

Computer haptics covers algorithms that generate artificially kinesthetic and tactile feedback for the interaction with simulated virtual objects [159]. Figure 2.2 visualizes the generally used basic haptic rendering loop for purely kinesthetic feedback and has been intensively investigated in the context of haptic teleoperation [160]–[162]. The following section instead

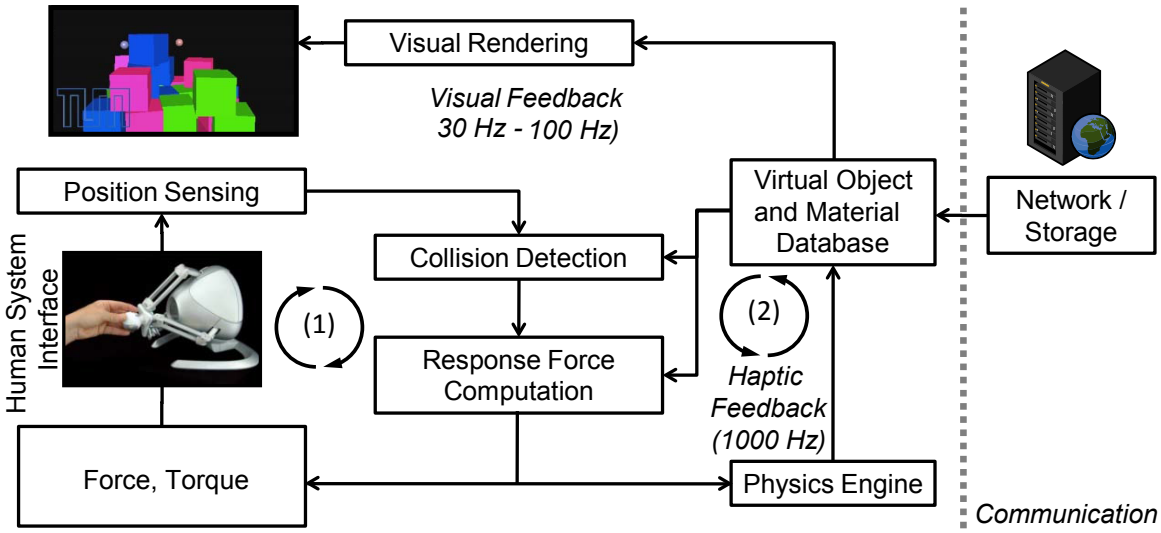


Figure 2.2: The general haptic rendering loop needs to run at least at 1 kHz to ensure realistic hardness display, i.e., kinesthetic force feedback. Further haptic effects such as Coulomb friction or macro-texture models have been included to this concept. However, thermal and high-fidelity microscopic roughness feedback were not conceptualized in the traditional haptic rendering pipeline.

focuses on tactile rendering models. In this context it is worth highlighting that the kinesthetic dimension volume, or shape, can be understood as the global result of all single-point tactile interactions during rendering and is notably associated with the tactile dimension hardness at each interaction point. That is why the first rendering algorithms, e.g., from Basdogan et al. [100] already contain hybrid kinesthetic and tactile feedback capabilities such as hardness and image texture-based macroscopic roughness feedback as well as basic friction effects. Consequently, the rendering of kinesthetic and tactile impressions is strongly interdependent.

2.3.1 Hardness

Hardness was one of the first rendered tactile impressions in haptics research. Since different actuators (Section 2.2.3) are required to address the different mechanoreceptors (Section 2.1.1), it is also required to address both hardness components using different models. The selected model notably depends on the fact whether the interacting bodies, i.e. the human hand and virtual material sample, are in contact or not. In the first case, only low-frequency rendering is required due to the limited motion range of the hand [28]. In the second case, high-frequency responses are elicited during tapping on a virtual material sample.

2.3.1.1 Low-Frequency Hardness Models - Pressure, Folding, Pinching, Squeezing

In-contact rigid body deformations can be classified into shape-changing actions such as shearing and bending, or both shape- and volume-changing actions such as dilating, expanding, and compressing [163]. Since tactile rendering in this thesis focuses on single points of interaction, the volume is not considered to be changed.

One of the most intuitive EPs, pressing, is based on Hooke's law and is defined in a single dimension (1D) as

$$\sigma = \frac{F}{A} = E \cdot \epsilon = E \cdot \frac{\Delta y}{l_0} \tag{2.3}$$

with σ being the stress, i.e., exerted force F per area A, the material-characteristic Young's Modulus E, and $\epsilon = \frac{\Delta y}{l_0}$ being the strain, i.e., the achieved change in length to the initial length l_0. Note that it is likewise common in haptic research to use the equation

$$F = \frac{E \cdot A}{l_0} \cdot \Delta y = k_{\text{spring}} \cdot \Delta y \tag{2.4}$$

which associates the spring stiffness (in N/m) to the Young's modulus (in N/m^2) and vice versa. This spring-based model is generally used for elastic deformation force feedback between the object-penetrating haptic interface point (HIP) and the visual interaction point (god object) defined by Zilles et al. [164].

The simplistic assumption of Hooke's law has strong limitations whenever the hardness properties of elastic and soft materials like sponges or foams need to be modeled. In this context, the rheological model viscoelasticity [165] covers the time-dependency and energy dissipation, i.e., internal conversion of mechanical stress into thermal energy, and is generally represented by springs and Newtonian dashpots in various configuration. A set of two springs and one dashpot is specifically relevant for this thesis and shown in the right part of Fig. 2.3, which is denoted as standard linear solid model (SLSM).

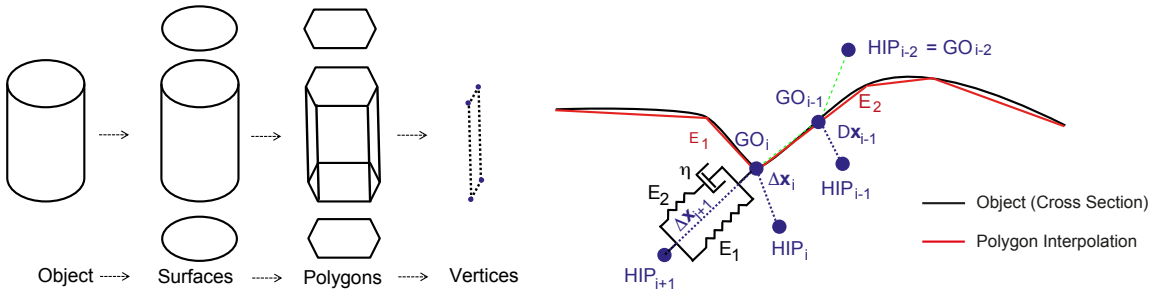


Figure 2.3: Haptic rendering is derived from 3D computer graphic algorithms and reveals dependencies between kinesthetic and tactile dimensions. For example, shapes are represented by polygon meshes visually and can be made tangible using the GO approach [164]. Based on these meshes, the HIP history and calculated GO are used to calculate interaction forces for haptic display such as hardness, friction, or macroscopic roughness height profiles. Shape is hence connected to the local perception of hardness, which is modeled by viscoelastic systems made of springs and dashpots. In this thesis, the shown configuration (standard linear solid model) of two springs and one dashpot is notably relevant for hardness rendering.

The introduction of viscoelasticity enables the rendering of several relevant hardness properties. Three notable differences characterize viscoelasticity.

1. The application of force may lead to permanent deformation when the force is removed, i.e., plastic deform can be modeled qualitatively representing the SynTouch dimension yielding cYD. The resulting σ-ϵ diagram forms a hysteresis loop, i.e., energy is dissipated, and hence irreversibly lost during the process. Note that forces in material exploration are generally less than 10 N, and hence, plastic deformations rarely occur for the majority of material samples.

2. Damping (cDP) behavior is introduced, i.e., the viscoelastic stretching of the material depends on the duration of force interaction.

3. Likewise, relaxation (cRX) during actively pressing on the material is represented, since a fraction of the input force can be irreversibly absorbed by the virtual material sample.

The modeling components of viscoelasticity are the linear elastic spring (Eq. 2.3) and the continuous linear viscous dashpot

$$\sigma = \eta \cdot \dot{\epsilon} \tag{2.5}$$

with η being the damping coefficient and $\dot{\epsilon}$ the first derivative of the strain, i.e., the speed of compressing the dashpot. Note that the stress σ can be converted into the force per area, in this thesis, bare finger area. The dashpot covers viscous hardness effects, i.e., the speed-depending force resulting on pushing an object through a liquid, and stays deformed after the interaction. Both elements (E and η) in series define the Maxwell arm model with

$$\epsilon_1 = \frac{1}{E} \cdot \sigma, \quad \dot{\epsilon}_2 = \frac{1}{\eta} \cdot \sigma, \quad \epsilon = \epsilon_1 + \epsilon_2 \tag{2.6}$$

and

$$\sigma + \frac{\eta}{E} \cdot \dot{\sigma} = \eta \cdot \dot{\epsilon} \tag{2.7}$$

This model considers damping phenomena and also models persistent deformations, since the dashpot does not recover to its initial strain. However, this also causes the fact that the Maxwell arm model is still not representative enough for the majority of viscoelastic materials [165]. Instead, the standard linear solid model [165], [166] further includes another spring (E_1) which is modeled in parallel to the Maxwell model (see right schematic in Fig. 2.3) leading to the difference equation

$$\sigma + \frac{\eta}{E_2} \cdot \dot{\sigma} = E_1 \cdot \epsilon + \frac{\eta \cdot (E_1 + E_2)}{E_2} \dot{\epsilon} \tag{2.8}$$

In this context, it is advisable to measure the stress, i.e., the force per area, instead of the strain. Force sensors have a predefined flat area to measure the currently applied user force and consequently adjust the position, and hence, the simulated strain of a moving element. The human EP pressing actually resembles a rheological model behavior, denoted as test [165], and likewise, other user EPs during hardness exploration can be represented by such tests.

1. **Creep-recovery response test**: An external constant stress σ_0 step function is applied, held, and completely released for specific time frames. This pattern reflects the typical EP (pressing) on a material sample.

2. **Stress-relaxation test**: A target indentation ϵ_0 step function is applied and held infinitely long. This procedure allows to determine, if the material stays deformed (yielding) and how it relaxes (relaxation). Note that this common test is further relevant for the tactile feature relaxation (cRX) discussed in Section 4.1.5.4.

3. **Cyclic test**: A periodic low-frequency indentation with a fixed amplitude and frequency is imposed. The frequency range is between 0 Hz – 12 Hz for a human user and quantifies the dynamic and complex-valued spring stiffness k* which has been investigated by Caldiran et al. [167] for the Maxwell arm model.

Notably the first and second test are relevant for this thesis since they resemble the typical pressing EP. The qualitative plots in Fig. 2.4 show the input of the creep-recovery and stress-relaxation tests. Section 5.2.4 discusses real material test results for haptic display.

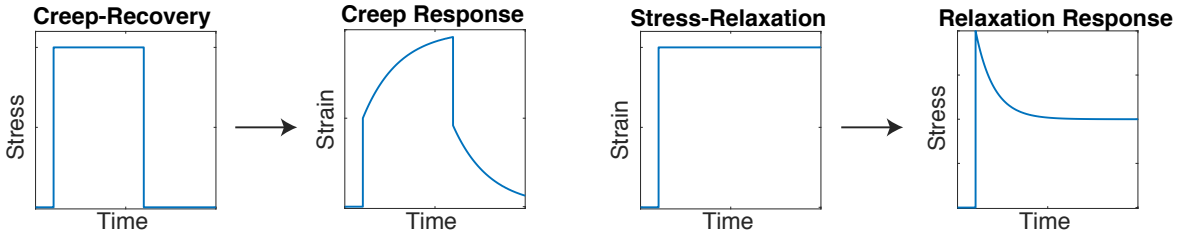


Figure 2.4: Idealized viscoelastic tests and responses. For the creep-recovery test, constant stress (force per area) is applied on the material and the strain is observed, which is comparable to pressing on a material sample with a constant force for a short period of time. Noticeably, if the material deforms, the motion penetrates the material sample. During the relaxation test, the material instead is indented and held to a specific position and the relaxation stress is observed over time, which helps in material classification in Section 4.1.5 and parameter identification for display in Section 5.2.4.

The standard linear solid model sufficiently represents viscoelastic materials; higher-order differential equations result from the addition of more dashpots and springs, as e.g. shown in the Wiechert model [165], and can even further model different stages of viscoelasticity. For the purpose of rendering the tactile subdimensions, the standard linear solid model, however, contains the three model parameters E_1, E_2, and η which are suitable to display the three tactile hardness subdimensions compliance, damping, and relaxation, and are further discussed during the haptic hardness model rendering in Section 5.2.4.

2.3.1.2 High-Frequency Hardness Models - Tapping

Whenever a human finger or a tool impacts on a material surface, high-frequency vibrations are elicited in the skin or held device. Soft and deformable object samples reveal a low-pass filtered response, i.e., viscoelastic materials excellently absorb tapping impulses [166]. Rigid bodies, by contrast, reveal an extensive and material-characteristic frequency response [96]. Culbertson et al. [106] identified that the response from tool-mediated tappings can be modeled by scaling the amplitude linearly, regardless whether the material is soft or deformable, since the mass of the users hand which wields the haptic device is constant. Figure 2.5 shows a series of accelerometer-based tap measurements on compressed

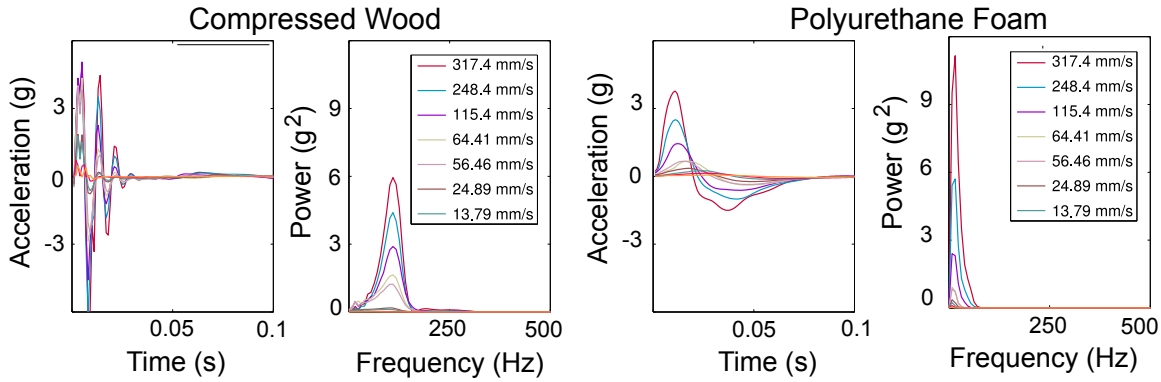


Figure 2.5: High-frequency hardness information during tapping can be modeled linearly according to the impact speed while leading to a sufficient perceptual similarity to real tapping on a material surface (adapted from [106] ©2017 IEEE).

wood (left) and polyurethane foam (right). Consequently, whenever the modeling of high-frequency taps is required, differently fast taps on a material sample can be recorded and their amplitudes are linearly interpolated for missing speeds.

2.3.2 Friction

Friction is modeled as the tangential force that resists motion when the finger or a tool is dragged across a surface with a certain normal force and tangential speed [163]. As core component, the Coulomb friction model is used to simulate the resistive interaction force F_f between two bodies without lubrication, known as dry friction, and depending on the normal force F_n as

$$F_\mathrm{f} = \mu \cdot F_\mathrm{n} \cdot sgn(\dot{x}) \tag{2.9}$$

with $sgn(\dot{x})$ being the sign of the current speed, which is the first temporal derivative of the position x. Several fundamental friction properties of this model as summarized in [168] are relevant for haptic display, which are listed in the following.

- Dry friction during sliding is approximately independent of speed and completely independent of the contact area.
- It is well-known in tribology that friction is not clearly correlated with surface roughness [168]. Microscopic roughness partly decreases surface adhesion, but intense roughness also increases required traction forces. Consequently, there is no obvious correlation between friction and roughness.
- The static friction coefficient μ_0 is approximately equal to the kinetic friction coefficient μ_0 for most material surfaces. However, for specific materials like glass, μ_0 is heavily depending on the contact conditions such as pressure force [169]. This effect is specifically critical for electrostatic and ultrasonic displays which currently still rely on finger-glass plate interactions [146], further complicates consistent display on such devices, and hence motivates the application of vibromechanical actuators instead.

Several friction models, shown in Fig. 2.6, have been used in haptic research for a wide application range to obtain different levels of achieved rendering realism. For example, the common implementation of Coulomb friction, the friction cone model has been introduced by Ruspini et al. [170] (right image in Fig. 2.6) to render static and dynamic friction. Viscous and lubricant interactions further are displayed using the viscous and Stribeck friction model. Note that this thesis focuses on the friction between solid bodies, and not lubricants;

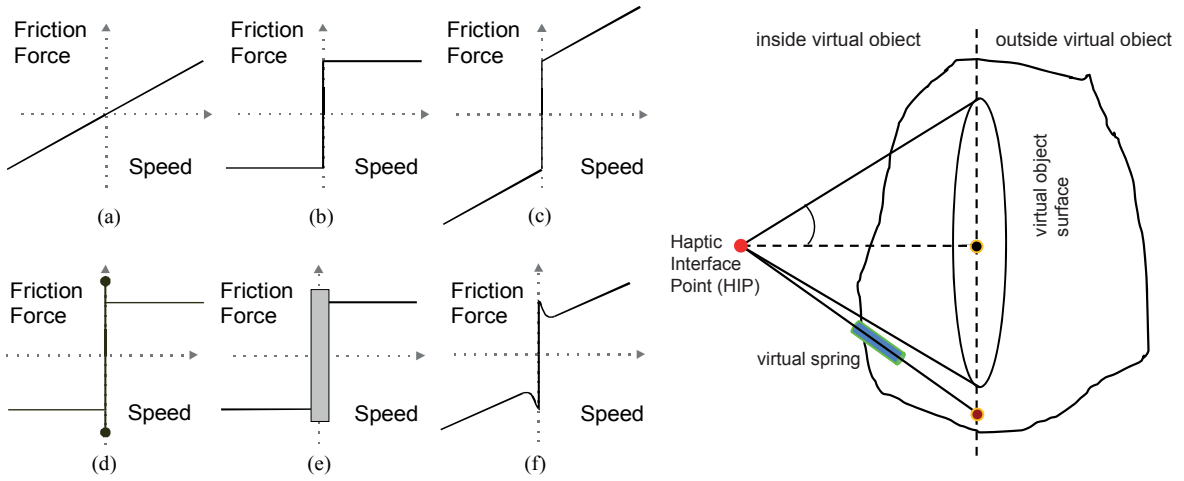


Figure 2.6: Basic friction models. (a) Viscous friction model for ideal fluids. (b) Basic Coulomb friction. (c) Combination of (a) and (b). (d) Coulomb friction with static friction and the commonly used implementation of the friction cone, shown in the right part of this figure. (e) Coulomb friction with deadband-based speed zone. (f) Combination of (d) and Stribeck friction for lubricant modeling (adapted from [171] ©1999 IEEE and [129] ©2002 IEEE).

consequently, viscous friction and Stribeck friction play a negligible role in friction rendering of dry materials and can be omitted. Instead, local stiction properties of dry friction are addressed by Dahl's friction model [172] who introduced the bristle analogy for small displacements, i.e., before the sliding motion enters the Coulomb friction sliding phase, the interacting surfaces are elongated by a tangentially-elongated spring with the spring stiffness k_0. The general Dahl model is defined as

$$\dot{F}_\mathrm{f} = k_0 \cdot \dot{x} \cdot \left| \left(1 - \frac{F_\mathrm{f}}{F_\mathrm{c}} \cdot sgn(\dot{x}) \right) \right|^i \cdot sgn\left(1 - \frac{F_\mathrm{f}}{F_\mathrm{c}} \cdot sgn(\dot{x}) \right) \tag{2.10}$$

with i being a material-depending empirical parameter within 0.5 and 2 as introduced by Dahl [172], and $F_\mathrm{c} = F_\mathrm{n} \cdot \mu$ the Coulomb friction, which is the maximum friction force in this context. As approximation for different materials, Dahl simplifies this expression by setting $i = 1$, leading to

$$\dot{F}_\mathrm{f} = k_0 \cdot \dot{x} \cdot \left(1 - \frac{F_\mathrm{f}}{F_\mathrm{c}} \cdot sgn(\dot{x}) \right) \tag{2.11}$$

which captures hysteresis and displacement behavior before two bodies relatively slide to each other. An analogy is the mass-spring behavior consisting of a spring attached between two points of the interacting bodies. The friction forces F_f is equivalent to the spring stiffness

k_0 times the achieved elongation z, and hence

$$z = \frac{F_\mathrm{f}}{k_0} \tag{2.12}$$

as well as

$$z_\mathrm{max} = \frac{F_\mathrm{c}}{k_0} \tag{2.13}$$

with z_max being the maximum displacement achieved by applying the Coulomb force $F_\mathrm{f,max} = F_\mathrm{c}$. Hence, equation (2.11) can be expressed [173] by the displacement z as

$$\dot{z} = \dot{x} \cdot \left(1 - \frac{z}{z_\mathrm{max}} \cdot sgn(\dot{x})\right) = \dot{x} - \dot{w} \tag{2.14}$$

with w being the surface adhesion point, and hence,

$$z = x - w \tag{2.15}$$

referring to the overall spring elongation. This simplified model has been used for haptic rendering by Culbertson et al. [106] and Choi et al. [101] due to the trade-off between implementation effort and sufficient friction modeling behavior compared to more subtle friction models (e.g., LuGre model [174]). Culbertson et al. [106] make use of the fact that the adhesion point w either is updated by $w = x \pm z_\mathrm{max}$ during sliding, or, stays unchanged by using the discrete-value model at the n^th iteration with

$$\mathbf{w}[n] = \begin{cases} \mathbf{x}[n] - sgn(\mathbf{x}[n] - \mathbf{x}[n-1]) \cdot z_\mathrm{max} & \text{if } |\mathbf{x}[n] - \mathbf{w}[n-1]| > z_\mathrm{max} \\ \mathbf{w}[n-1] & \text{otherwise} \end{cases} \tag{2.16}$$

The corresponding friction force can be inferred from the ratio

$$\frac{\mathbf{F}_\mathrm{f}[n]}{\mathbf{F}_\mathrm{c}[n]} = \frac{\mathbf{z}[n]}{z_\mathrm{max}} \tag{2.17}$$

note that $\mathbf{z}[n]$ does not exceed z_max, and consequently,

$$\mathbf{F}_\mathrm{f}[n] = \mathbf{F}_\mathrm{c}[n] \cdot \left(\frac{\mathbf{z}[n]}{z_\mathrm{max}}\right) = \mu_\mathrm{k} \cdot \mathbf{y}[n] \cdot k_0 \cdot \left(\frac{\mathbf{x}[n] - \mathbf{w}[n]}{z_\mathrm{max}}\right) \tag{2.18}$$

with the current penetration depth $\mathbf{y}[n]$ which is included in the Coulomb force calculation $F_\mathrm{c} = \mu_\mathrm{k} \cdot y \cdot k_0$. Culbertson et al. [106] empirically determined the maximum presliding parameter $z_\mathrm{max} = 4 \text{ mm} \cdot \mu_\mathrm{k}$ as tradeoff between rendering stability and perceived friction realism. Note that the static friction and slip-stick is not covered by this approach, but it solely requires one parameter, the kinetic friction coefficient, and is still able to display two phases of friction, namely, sticking and sliding.

Whenever the static friction coefficient μ_0 is also available, this simplified model can be extended to the generalized Dahl model as introduced by Hayward et al. [173]. Equation (2.16) is adjusted to

$$\mathbf{w}[n] = \begin{cases} \mathbf{x}[n] - sgn(\mathbf{x}[n] - \mathbf{x}[n-1]) \cdot z_\mathrm{max} & \text{if } |\mathbf{x}[n] - \mathbf{w}[n-1]| > z_\mathrm{max} \\ \mathbf{w}[n-1] + (\mathbf{x}[n] - \mathbf{x}[n-1]) \cdot \alpha(\mathbf{x}[n] - \mathbf{w}[n-1]) \cdot (\mathbf{x}[n] - \mathbf{w}[n-1]) & \text{otherwise} \end{cases} \tag{2.19}$$

with

$$\alpha(z) = \frac{1}{z_{\text{max}}} \cdot \frac{z^8}{z_{\text{stick}}^8 + z^8} \tag{2.20}$$

and $z_{\text{max}} = 4$ mm $\cdot \mu_{\text{k}}$ and $z_{\text{stick}} = 4$ mm $\cdot \mu_0$. Note that the higher order of the polynomial expression in Eq. (2.20) is an empirical finding of Hayward et al. [173] to smooth the sharp transition between sticking and sliding. Since $\mu_0 > \mu_{\text{k}}$, two new phases can be observed between the transition, namely, creeping and stick-slip oscillation which are also known as stick-slip-chatter or stick-slip-effect. Figure 2.7 shows the generalized Dahl model using the $\mu_{\text{kin}} = 0.4$ and the static friction coefficient $\mu_0 = 0.5$ in order to illustrates these effects.

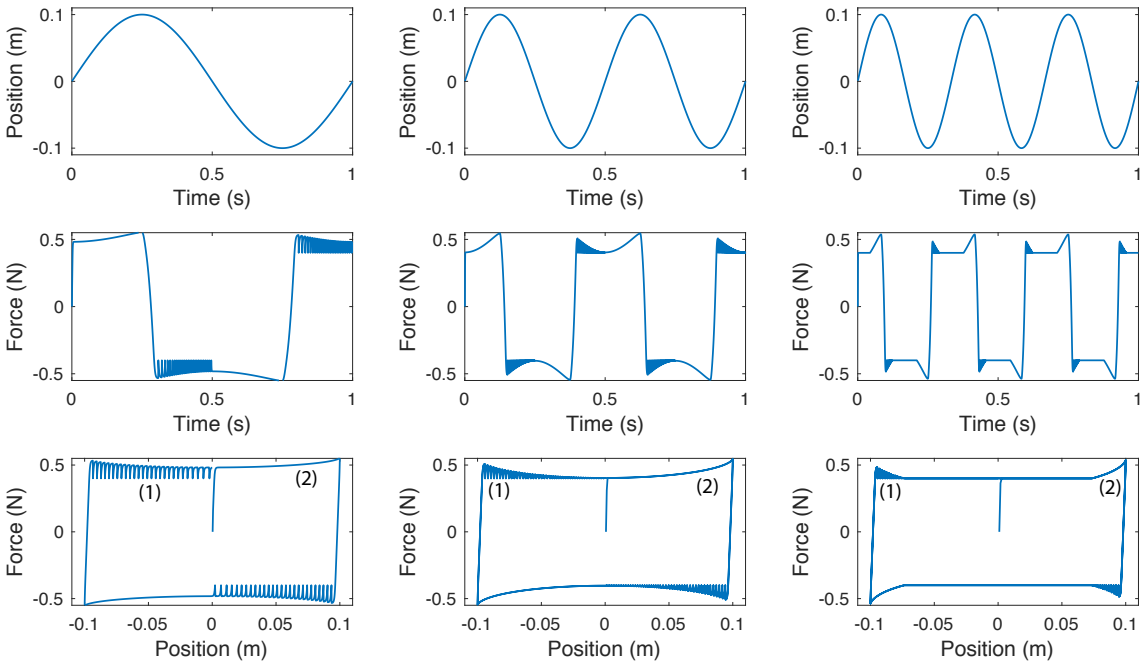


Figure 2.7: Generalized Dahl friction model. The first row shows the input periodical sliding (from left to right: 1 Hz, 2 Hz, 3 Hz) and the second and third row the resulting force-time and force-displacement curves. Note that the typical stick-slip chatter (1) occurs whenever speed becomes faster and is notably strong for lower frequencies whenever the static and kinetic friction coefficients are different, and creeping (2) occurs whenever the motion becomes slower, i.e., before the finger gets stuck again.

The striking argument for the application of this model is the fact that only two parameters, i.e., the friction coefficients, need to be fed into the model which considers the majority of relevant dry friction properties during real material interactions; these two coefficients coincide with the two SynTouch features tactile stiction fST and sliding resistance fRS.

2.3.3 Macroscopic Roughness

Minsky [98] proposed the first haptic texture mapping based on height images and a gradient-based algorithm in 1995. Srinivasan et al. [16] extended this approach to 3D and further examined the different techniques of force shading, force perturbation (bump-mapping),

and displacement mapping which are adapted from computer graphics. Notably force perturbation is relevant in the context of this thesis. The original algorithm from Max et al. [175]

$$\vec{M} = \vec{N} - \nabla h + (\nabla h \cdot \vec{N}) \cdot \vec{N} \tag{2.21}$$

changes the original normal vector $\vec{N}$ according to the height map intensity gradient ∇h and leads to a new normal vector $\vec{M}$. Figure 2.8 shows that a flat virtual surface can be extended by overlaying such predefined force vectors resulting from a height map which are perceived as surface structure perturbations.

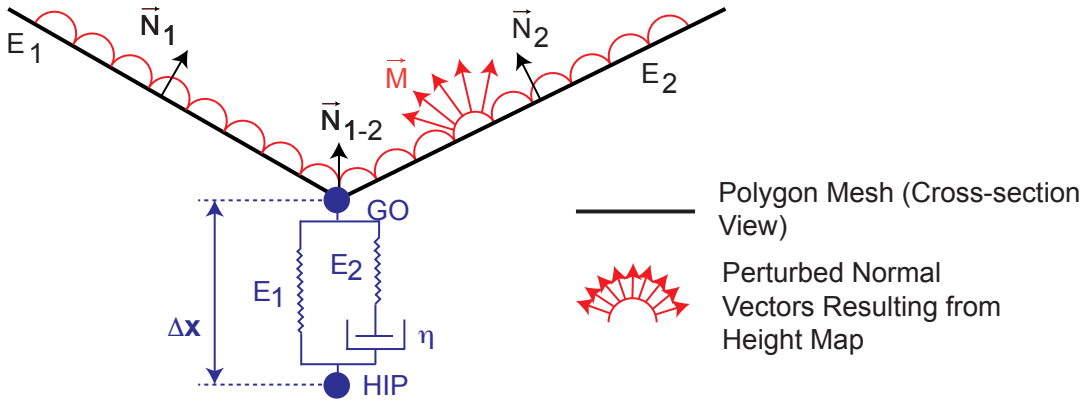


Figure 2.8: Macroscopic structure patterns can be added to completely flat virtual material surfaces. Note that macroscopic roughness is linked to hardness modeling, represented by the two springs and the dashpot from the aforementioned SLSM.

The height map value may either originate from a corresponding 2D height/depth map image (image-based haptics), or, from a mathematical function (procedural haptics [16]). In terms of real material display, the first approach is applied in this thesis based on texture images. Depending on the current position of the HIP and its current speed, a linear prediction can be made about the future position to estimate the next height gradient. This extension allows for a more realistic rendering of the gradient-based surface shapes whenever a tactile display can reproduce not only different heights, but also, different inclinations. In order to generate the height gradient in x and y directions, e.g., the Sobel filter can be applied as done by Hausberger et al. [176]. Note that the SynTouch feature macrotexture (mTX) is encoded in the actual height, or depth, of the image, whereas the regularity (mRG) and coarseness (mCO) are represented by the locations and spacing of the image patterns.

2.3.4 Microscopic Roughness

Both data acquisition and tactile display of microscopic roughness are based on 1D signals, denoted as vibrotactile signals, which need to be adjusted for the purpose of feature extraction as well as for later haptic display. In the following, the required fundamentals are summarized.

2.3.4.1 Required 1D Signal Transforms

A continuous vibrotactile signal $v(t)$ resulting from a tool-surface interaction is recorded with an acceleration sensor. The signal is quantized using an analog-digital converter

$$Q\{v(t)\} = \Delta \cdot \lfloor \frac{v(t)}{\Delta} + \frac{1}{2} \rfloor \tag{2.22}$$

with the quantization step size

$$\Delta = \frac{v_{\text{max}} - v_{\text{min}}}{2^B} \tag{2.23}$$

of word length B (in bit) and is then sampled at ΔT time steps [177] using

$$\mathbf{v}[\text{n}] = \sum_{n=-\infty}^{\infty} Q\{v(t)\}\delta(t - n \cdot \Delta T) \tag{2.24}$$

resulting in a time-discrete and quantized vibrotactile signal $\mathbf{v}[\text{n}]$. If N data points are recorded, the recording is $T = \frac{N}{f_\text{s}} = N \cdot T_\text{s}$ seconds long. Apart from being considered in the time domain, such signals can be converted into other representations for further analysis. A mathematical transform of $\mathbf{v}[\text{n}]$ is defined as

$$\mathbf{T}[u] = \sum_{n=1}^{N} \mathbf{v}[n] \cdot \mathbf{r}[n, u] \tag{2.25}$$

and its inverse as

$$\mathbf{v}[n] = \sum_{u=1}^{N} \mathbf{T}[u] \cdot \mathbf{s}[n, u] \tag{2.26}$$

with the kernels $\mathbf{r}[n, u]$ and $\mathbf{s}[n, u]$, which are the weighted averages of the original signal $\mathbf{v}[\text{n}]$ and the transform $\mathbf{T}[u]$ from a family of fundamental signals. In the context of this thesis, two different 1D transforms are of interest.

Vibrotactile signals resulting from material surface scans do not contain temporally dependent content, and hence, do not require short term analysis techniques. Consequently, their pseudo-stationarity is suitable for frequency analysis based on sinusoidal waves. Similar to its 2D application in image signal processing, the 1D discrete Fourier transform (DFT) is hence of crucial importance for digital signal processing, and in this context, for vibrotactile signals. The 1D DFT is calculated [178] for a specific frequency f as

$$\mathbf{V}[f] = \sum_{n=1}^{N} \mathbf{v}[n] \cdot \mathbf{r}[f, n] = \sum_{n=1}^{N} \mathbf{v}[n] \cdot e^{-j \cdot \frac{2 \cdot \pi \cdot f \cdot n}{N}} \tag{2.27}$$

$\mathbf{V}$ is a complex number array and each value can be written as

$$\mathbf{V}[f] = |\mathbf{V}[f]| \cdot e^{j \cdot \angle \mathbf{V}[f]} \tag{2.28}$$

represented by the magnitude $|\mathbf{V}[f]|$ and phase $\angle \mathbf{V}[f]$ component. Another related transform is the discrete cosine transform type 2 (DCT-II) with

$$\mathbf{V}[f]_{\text{DCT}} = \sum_{n=1}^{N} \mathbf{v}[n] \cdot \mathbf{r}[f, n] = \sum_{n=1}^{N} \mathbf{v}[n] \cdot cos\left[\frac{\pi}{N} \cdot \left(n + \frac{1}{2} \cdot k\right)\right] \tag{2.29}$$

which is generally used in compression tasks and is a possible alternative to the DFT in feature design. Figure 2.9 shows an example vibrotactile signal $\mathbf{v}$ and its magnitude $|\mathbf{V}|$ (material sample sandpaper). The phase information $\angle\mathbf{V}[f]$ does not change the microscopic roughness modeling, but is stored for later haptic display.

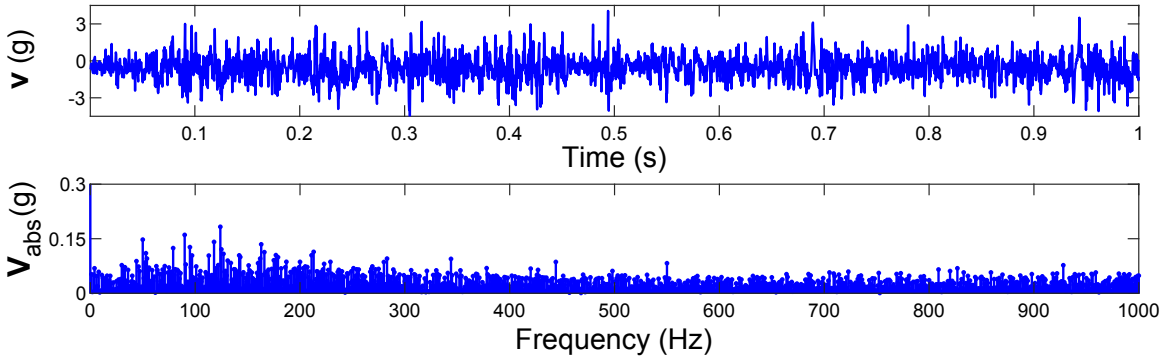


Figure 2.9: Example vibrotactile signal recording (upper figure, time domain and lower figure, absolute values of spectral domain) of a material sample (sandstone). Accelerometer-recorded signals cover a broader frequency range (i.e., 0 Hz – 1400 Hz) than the human perception receptors ranges (see Section 2.1.1), and hence, provide further potential information for the identification of the material surface.

Another relevant transform is the 1D discrete Wavelet transform (DWT) [179] which can represent an input signal at more than one resolution, which is denoted as multiresolution theory. This property is useful in the context of this thesis since features calculated from a single data trace may vary stronger due to operator-induced scan speed variations compared to a feature calculated from multiple components stemming from one and the same signal. Wavelets are categorized into wavelet families (Haar, Daubechies, etc.) and are represented by a scaling father basis function φ and a mother wavelet function ψ. The advantages of wavelets is the exact temporal and frequency-based analysis using scaled and shifted versions of these basis functions with the time discrete father scaling function

$$\boldsymbol{\varphi}[j,k,x] = 2^{j/2} \cdot \boldsymbol{\varphi}[2^j \cdot x - k] \tag{2.30}$$

and the mother wavelet function

$$\boldsymbol{\psi}[j,k,x] = 2^{j/2} \cdot \boldsymbol{\psi}[2^j \cdot x - k] \tag{2.31}$$

Wavelet basis functions like Daubechies (db) and Coiflet (coif) have been applied by Okeheim et al. [180] in the context of audio signal decomposition. Due to the similarity of audio and vibrotactile data, such an approach is useful for vibrotactile modeling as well. Notably, the Daubechies wavelets resemble impulse patterns generated by sliding a sensorized device over a surface, i.e., the acceleration spikes during the tool–surface physical interaction. The exact type of the db wavelet (i.e., db4 or db16, etc.) needs to be identified during optimization nevertheless.

Based on the scaling j and x-axis shifting variable k, the DWT [177] decomposes a time-discrete signal **v**[n] into a non-shifted and non-scaled approximation (A) part

$$\mathbf{T}_{\varphi}[0,0] = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathbf{v}[n] \cdot \boldsymbol{\varphi}[0,0,n] \tag{2.32}$$

and a detail (D) part

$$\mathbf{T}_{\psi}[j,k] = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \mathbf{v}[n] \cdot \boldsymbol{\psi}[j,k,n] \tag{2.33}$$

The inverse transform (IDWT) is

$$\mathbf{v}[n] = \frac{1}{\sqrt{N}} \cdot \left[\mathbf{T}_{\varphi}[0,0] \cdot \boldsymbol{\varphi}[0,0,n] + \sum_{j=1}^{J} \sum_{k=0}^{2^j-1} \mathbf{T}_{\psi}[j,k] \cdot \boldsymbol{\psi}[j,k,n] \right] \tag{2.34}$$

One major application of the 1D DWT is the multiresolution analysis (MRA), also denoted as wavelet decomposition, and is specifically relevant for this thesis. Its main concept is the step-wise filtering and downsampling (by 2) of the input signal **v**[n]. Note that if the input signal is cropped or zero-padded to a length of $N = 2^J$ the DWT is called the Fast Wavelet Transform (FWT) which is computationally more efficient. In this case, J filtering and downsampling steps can be performed in a FWT filterbank defined as

$$\mathbf{T}_{\varphi}[j,k,n] = \sum_{m=-\infty}^{+\infty} \mathbf{T}_{\varphi}[j+1,m] \cdot \mathbf{h}_{\varphi}[n-m] \;\text{ with } n = 0,2,...2^j - 2 \tag{2.35}$$

and

$$\mathbf{T}_{\psi}[j,k,n] = \sum_{m=-\infty}^{+\infty} \mathbf{T}_{\psi}[j+1,m] \cdot \mathbf{h}_{\psi}[n-m] \;\text{ with } n = 0,2,...2^j - 2 \tag{2.36}$$

using convolutions with the time discrete scaling and wavelet functions as filters $\mathbf{h}_{\varphi}(-n)$ and $\mathbf{h}_{\psi}(-n)$, respectively. Figure 2.10 shows two levels of such a filterbank (left) and the digital filters $\mathbf{h}_{\varphi}[n]$ and $\mathbf{h}_{\psi}[n]$ (right), of a Daubechies 8 (db8) wavelet; note its presumable similarity to vibrotactile spike patterns occurring during sensor-surface interaction. The MRA is

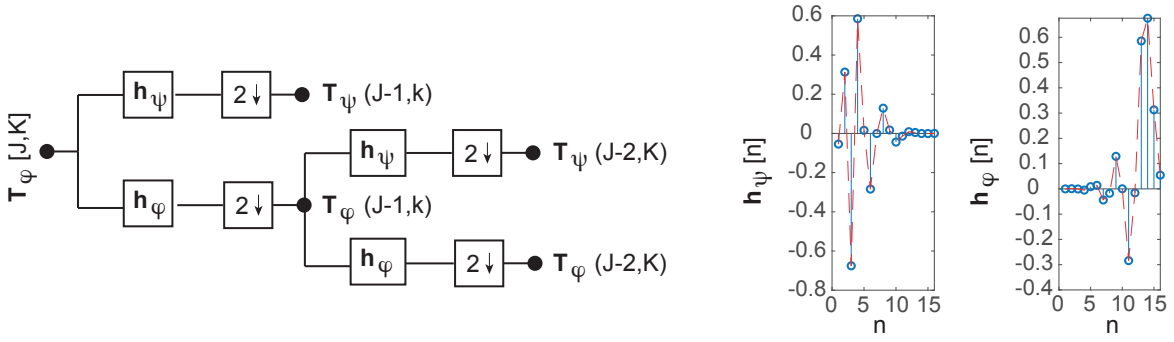


Figure 2.10: Exemplary two-level FWT filterbank (left) and time-discrete db8 filters (right). The input signal **v** = T_{φ}(J,k) is subsequently filtered and downsampled to obtain different signal representations of one and the same input signal.

achieved by using the Matlab function wavedec. The detail levels of the differently scaled FWTs are used for the calculation of the microscopic roughness feature in Section 4.1.5.2.

2.3.4.2 Digital Filters

Digital filters are applied to remove specific frequencies during the modeling of vibrotactile signals. A digital filter is defined [178] as

$$H(z) = \frac{y(z)}{x(z)} = \frac{\mathbf{b}[z]}{\mathbf{a}[z]} = \frac{b_1 + b_2 \cdot z^{-1} + ... + b_J \cdot z^{-1}}{a_0 + a_1 \cdot z^{-1} + ... + a_K \cdot z^{-1}} \tag{2.37}$$

with **a** and **b** being the filter coefficients. Their parametrization defines whether the filter is a finite impulse response (FIR) or infinite impulse response filter (IIR). The n-th output of an IIR filter, for example, is

$$\mathbf{y}[n] = \sum_{k=1}^{K} \mathbf{a}[k] \cdot \mathbf{y}[n-k] + \sum_{j=1}^{J} \mathbf{b}[j] \cdot \mathbf{x}[n-j] \tag{2.38}$$

In the case of FIR filters, the denominator reduces to a_0 only, which leads to stable system behavior, but comparably higher filter parameters necessary to get the same filter result as an IIR filter. Different filter behaviors (e.g., low-pass) are implemented using the Matlab filter design tool fdesign. Specifically bandpass filters, which remove lower frequencies (below 12 Hz) and larger frequencies (above 1,000 Hz), are relevant for this thesis in terms of human hand motion influence and sensor noise removal, respectively. In terms of haptic display, notably, the influence of hand motion (lower frequencies) and audible noise (higher frequencies) during recording has to be excluded. The filter used in this thesis is a 8^{th} order IIR Butterworth filter, which has suitable characteristics [178] for vibrotactile sliding signals. First, it reveals a smooth and monotonic amplitude response in the passband area. Second, it has a sufficiently steep roll-off around the cut-off frequency which increases with the filter order. Notably, the sliding signal analysis is an offline task, and hence, no temporal constraints are set for the filter order in this context. Third, the sliding signals are filtered first and, thereafter, the most stationary segment is extracted. Consequently, any initial impulses, and hence, the overshooting and ringing of the filter are removed. Note that the filter order also increases the overshooting and settling time of the step response and thereby imposes an upper limit for the maximum filter order.

2.3.4.3 Vibrotactile Model Considerations for Display

Recorded vibrotactile signals used for material analysis applications are likewise suited for vibrotactile signal display based on a set of adjustments. It is worth highlighting beforehand that any attempts of physics-based simulation are too computationally intensive for haptic rendering [181] due to the complex biomechanics of the human fingertip and the high effort of modeling the microstructure of a material surface. Any presently used model is hence a strong simplification of the real finger-surface interaction. Instead, microscopic roughness is typically modeled based on the aforementioned 1D vibrotactile signals following one of the two concepts.

On the one hand, vibrotactile signals are recorded at a specific speed as well as interaction force, and consequently, are repeated in a playback loop during haptic rendering as done by

Takeuchi et al. [182]. Weber et al. [34] further identified that a linear adaptation of the pitch ranges within 40 mm/s – 120 mm/s, and hence, it is a recommended option to change the pitch of the vibrotactile signals correspondingly.

On the other hand, data-driven approaches, as presented for example in [14], [91], or, [104], [105] use hundreds of prerecorded vibrotactile signals at specific scan speed – scan force pairs to calculate filter parameters (e.g., ARMA coefficients [91]). During display, the current user input speed and force are measured to interpolate the filter model coefficients using Delaunay triangulation as shown in [183]. Gaussian white noise is then fed into the interpolated filter to synthesize a vibrotactile signal in real-time which is displayed using another tool-mediated setup containing a voice coil actuator. The work in [105] further considers the speed components v_{x} and v_{y} to take anisotropic (i.e., direction-depending) surface textures into account. The applied scan force is concentrated on the tip of the device in tool-mediated setups and changes the vibrotactile signals [14], [106].

Note that both models have the potential to display vibrotactile signals realistically to a human user. The choice of model in this thesis is discussed in Section 5.2.3.

2.3.5 Warmth

Two different thermal models are relevant in the context of this thesis.

2.3.5.1 Contact-based Thermal Models

Whenever two bodies with different surface temperatures come into contact, thermal energy is exchanged via heat flux. The fundamental bioheat equation from Pennes et al. [184] describes this process for the human skin and a touched object as

$$\rho \cdot c \cdot \frac{dT}{dt} = \nabla(k \cdot \nabla T) + (\rho \cdot c)_{\mathrm{blood}} \cdot \omega_{\mathrm{blood}} \cdot (T_{\mathrm{a}} - T) + q_{\mathrm{m}} \tag{2.39}$$

with k being the thermal conductivity, ρ the density, c the warmth capacity, the term $(\rho \cdot c)_{\mathrm{blood}}$ the heat capacity of the human blood, ω_{blood} the perfusion rate of the blood, T_{a} the arterial blood temperature before entering the capillaries, and q_{m} being the metabolic heat generation. This equation reveals fundamental challenges for thermal-based rendering whenever living human skin is involved. Consequently, subsequent research with focus on haptic display simplified this equation by assuming that the skin is inanimate and has a well-defined geometry. Note that this assumption is based on the fact that the outermost skin layer (stratum corneum) is composed of dead cell tissue [24]. This is specifically relevant in the context of this thesis, since thermal rendering only at the index finger is intended. Moreover, Sarda et al. [185] identified that the two-dimensional heating process can be further simplified to a one-dimensional process. Hence, the bioheat equation becomes

$$\frac{dT^2}{dx^2} = \frac{\rho \cdot c}{k}\frac{dT}{dt} = \frac{1}{\alpha}\frac{dT}{dt} \tag{2.40}$$

The thermal interaction between two inanimate bodies is governed mainly by heat conduction, and consequently can be modeled by the two semi-infinite body-in-contact (SIBC)

model [186]. The line of research started by Jones and Ho [29], [70], [154] links thermal models to haptic display using controllable Peltier elements and has been further validated by [111], [158]. As an initial version, Ho and Jones [70] describe the resulting mixture interface temperature T_∞ between the skin and the surface as

$$T_\infty = \frac{T_{\text{mat},0} \cdot \sqrt{(k_{\text{mat}} \cdot \rho_{\text{mat}} \cdot c_{\text{mat}})} + T_{\text{skin},0} \cdot \sqrt{(k_{\text{skin}} \cdot \rho_{\text{skin}} \cdot c_{\text{skin}})}}{\sqrt{(k_{\text{mat}} \cdot \rho_{\text{mat}} \cdot c_{\text{mat}})} + \sqrt{(k_{\text{skin}} \cdot \rho_{\text{skin}} \cdot c_{\text{skin}})}} \tag{2.41}$$

with $T_{\text{mat},0}$ being the initial material surface temperature and $T_{\text{skin},0}$ the initial skin temperature, respectively. The material parameters relevant for thermal rendering have been made available in material databases like [187], [188]. Ho and Jones [157] determined the contact coefficient of the human skin to be $(k\rho c)^{\frac{1}{2}}_{\text{skin}} = 1.181 \frac{J}{m^2 K s^{\frac{1}{2}}}$ [70]. Note that the resulting temperature does not depend on the time in this initial model and that it is valid only if the Fourier number

$$Fo = \frac{\alpha_{\text{m}} \cdot t}{l^2} \tag{2.42}$$

is smaller than 0.05 [189] with $\alpha_{\text{m}} = \frac{k}{\rho \cdot c}$ being the thermal diffusivity of the material, l being the thickness, and t the touch time. For example, a highly-conductive material sample like a thin copper sheet violates this condition already after one second [70], and hence, the SIBC thermal rendering model is less preferable for such material samples. For less conductive or thicker material samples, however, this condition leads to contact times of generally up to 10 seconds which matches the relevant thermal conditions for, e.g., thermal persistence (tPR) [66]. If such a constant interface temperature can be assumed, the relevant conclusion for thermal modeling is that a technical cooling component (e.g., a Peltier element) merely has to be controlled to achieve a specific temperature based on the contact coefficient of the desired material [29].

The initial SIBC model does not consider the critical aspect that both the finger skin temperature or the touched material surface temperature may change over time and that both the skin and the material heat flux are influenced by the thermal contact resistance R. Consequently, Jones and Ho [29] use the simplified bioheat equation (2.40) to infer the contact condition of the heat flux as

$$q'' = \frac{T_{\text{skin}}(t) - T_{\text{mat}}(t)}{R_{\text{skin-mat}}} \tag{2.43}$$

Note that the thermal resistance also depends on several material parameters, and notably, from the contact force as further examined by Jones and Ho [157]. For common contact forces up to 2 N, the skin-material resistance $R_{\text{skin-m}}$ of smooth materials has been determined to follow

$$R_{\text{skin-mat}} = \frac{0.37 + k_{\text{mat}}}{1,870 \cdot k_{\text{mat}}} \tag{2.44}$$

with k_{mat} being the thermal conductivity of the material. For example, copper ($k = 401 \frac{W}{m^2 K}$) has an approximated resistance of $R_{\text{skin-copper}} = 5.35 \cdot 10^{-4} \frac{m^2 K}{W}$. Note that Jones and Ho [29] further determined $R_{\text{skin-display}} = 5.341 \cdot 10^{-4} \frac{m^2 K}{W}$ for the thermal resistance of Peltier elements (ceramics), and other $R_{\text{skin-mat}}$ can be inferred from Eq. (2.44). The first relevant

conclusion from this approximation is that the required cooling power (in W/m^2) of Peltier elements can be derived. For example, if a temperature gradient of 11^oC needs to be created by a 4 cm^2-sized Peltier element (length and width of 2 cm), equation (2.43) leads to $P = q'' \cdot A_{\text{peltier}} = \frac{11^o}{5.341 \cdot 10^{-4}} \approx 20,000 \frac{W}{m^2} \cdot 0.0004 m^2 = 8$ W.

The second relevant conclusion considers the artificial display of virtual material surfaces using Peltier elements. The real material temperature in Eq. (2.43) is replaced by the adjustable display temperature $T_{\text{disp}}(t)$ and becomes

$$q'' = \frac{T_{\text{skin}}(t) - T_{\text{mat}}(t)}{R_{\text{skin-mat}}} = \frac{T_{\text{skin}}(t) - T_{\text{disp}}(t)}{R_{\text{skin-disp}}} \tag{2.45}$$

and leads to the display temperature rendering equation

$$T_{\text{disp}}(t) = T_{\text{skin}}(t) \cdot \left(1 - \frac{R_{\text{skin-disp}}}{R_{\text{skin-mat}}}\right) + T_{\text{mat}}(t) \cdot \left(\frac{R_{\text{skin-disp}}}{R_{\text{skin-mat}}}\right) \tag{2.46}$$

Ho and Jones further collected time-depending data traces for differently conductive materials in [154]. The required temporal relationships of $T_{\text{skin}}(t)$ and $T_{\text{mat}}(t)$ are both inferred in [29], [154]. The simplified bioheat Equation (2.40) is solved with the initial condition that $q''_{\text{skin,s}} = q''_{\text{mat,s}}$. Consequently, Eq. (2.40) becomes

$$T_{\text{skin}}(t) = \frac{A}{B}\left(1 - e^{\alpha_{\text{skin}} \cdot B^2 \cdot t} \cdot erfc\Big(B \cdot \sqrt{\alpha_{\text{skin}} \cdot t}\Big)\right) + T_{\text{skin,0}} \tag{2.47}$$

with

$$A = \frac{-(T_{\text{skin,0}} - T_{\text{mat,0}})}{k_{\text{skin}} \cdot R} \quad B = \frac{1}{k_{\text{skin}} \cdot R}\left(1 + \frac{\sqrt{(k\rho c)_{\text{skin}}}}{\sqrt{(k\rho c)_{\text{mat}}}}\right) \tag{2.48}$$

and for $T_{\text{mat}}(t)$

$$T_{\text{mat}}(t) = \frac{C}{D}\left(1 - e^{\alpha_{\text{mat}} \cdot D^2 \cdot t} \cdot erfc\Big(D \cdot \sqrt{\alpha_{\text{mat}} \cdot t}\Big)\right) + T_{\text{mat,0}} \tag{2.49}$$

and

$$C = \frac{-(T_{\text{mat,0}} - T_{\text{skin,0}})}{k_{\text{mat}} \cdot R} \quad D = \frac{1}{k_{\text{mat}} \cdot R}\left(1 + \frac{\sqrt{(k\rho c)_{\text{mat}}}}{\sqrt{(k\rho c)_{\text{skin}}}}\right) \tag{2.50}$$

with the known skin and material parameters k, ρ, and c as well as $\alpha = \frac{k}{\rho \cdot c}$. The temporal evolution of the skin and the material surface can be solely determined by known material constants and the initial temperatures of the skin and material surface. In [29] the resulting temperature curves are shown for various material samples (polyurethane foam, granite, ABS plastic, steel, and copper) which effectively range between the ambient temperature (about 23^oC) and the finger skin temperature (about 32^oC). Both of these initial temperatures are measured using the novel tactile display presented in this thesis and are further discussed and visualized in Section 5.2.5. Since the model adjust the display temperature over time, the initial cooling (tCO) is displayed by the initial gradient, and the persistent cooling (tPR) is based on the persistent temperature gradient.

2.3.5.2 Non-contact-based thermal models

Another relevant thermal model approach in this thesis is based on Newton's laws of cooling [163] and applies whenever a surface is locally heated and emits the thermal energy into the ambient air. This thermal signature can be examined using active thermography. Aujeszky et al. [94] use a laser to actively heat up a material sample and locally measure the thermal gradient using infrared sensing devices, i.e., the thermal response to external excitation. Ultimately, all surfaces obtain ambient temperature after an amount of indefinite time.

The temporal change of temperature is described as

$$\frac{dT(t)}{dt} = -\frac{1}{\tau} \cdot (T(t) - T_{\text{amb}}) \tag{2.51}$$

with

$$\tau = \frac{c \cdot m}{\alpha \cdot A} \tag{2.52}$$

and A being the effectively heated area, c and m the heat capacity and mass of the material, α the heat transfer coefficient for non-turbulent heat transmission [186] with $\alpha = 6\ \frac{W}{m^2 \cdot K}$, and T_{amb} the ambient temperature. This differential equation is solved to

$$T(t) = T_{\text{amb}} + (T_{\text{max}} - T_{\text{amb}}) \cdot e^{-\frac{t}{\tau}} \tag{2.53}$$

The maximum temperature T_{max} is imposed on the material surface and the resulting cooling curve is highly characteristic depending on the material. If a metal is heated, a larger part of the thermal energy is conducted inside the metal compared to an insulating wooden material sample which maintains a larger surface temperature.

2.4 Haptic Communication

Since this work mainly focuses on the acquisition, offline storage, and display of tactile information, the research field of haptic communication is not further discussed. The interested reader can refer to extensive surveys from Steinbach et al. about haptic communications for the kinesthetic [162], [190] and tactile [4] domain, respectively.

2.5 Chapter Summary

This chapter provided fundamentals in the areas of haptic perception (Section 2.1), haptic acquisition and display devices (Section 2.2), and the rendering of haptic models (Section 2.3) to introduce the reader to the related work in the research area of haptic material analysis and display. This thesis mainly focuses on the challenges and limitations of the state-of-the-art in machine haptics and haptic modeling which were identified and summarized in the following.

1. There is no uniform taxonomy of material categorization across interdisciplinary fields, which leads to high ambiguity among different haptic databases from related work.

2. No acquisition device has been reported that can parameterize the most important tactile dimensions and is likewise easily extendable, versatile, and low-cost.

3. The mathematical representation of features in related work generally uses single domains only and do not fully represent a broad set of tactile features.

4. Likewise, no tactile display device has been reported that attempts to display all 15 SynTouch subdimensions simultaneously.

This thesis addresses all of these challenges. Chapter 3 introduces the taxonomy and novel Texplorer data acquisition device, Chapter 4 describes related work and custom-made features and a complete material classification pipeline based on handcrafted features as well as deep learning-based classification, and Chapter 5 introduces and discusses a novel tactile display device targeting the reproduction of the 15 SynTouch subdimensions.

Chapter 3

Data Acquisition

3.1 Haptic Databases

A major contribution of this thesis is the collection and labeling of a large haptic database which is described in the following. So far, only few haptic databases have been made publicly available. The most relevant one is provided by Culbertson et al. [191], which contains vibrotactile tool-surface interaction data in combination with the scan speed and force for 100 material classes. Several arguments, however, led to the conclusion that the collection of an own haptic material database was inevitable.

- A large set of tactile dimensions, e.g., friction coefficients, or thermal properties, were not recorded. The sensorized tool presented in [191] mainly focuses on the acquisition and display of vibrotactile information, and hence, is not necessarily representative neither for the acquisition of all five tactile dimensions nor the 15-subdimensional SynTouch feature space.
- It is possible to mitigate scan dependencies of haptic data acquisition devices by specific mechanical constructions before scanning a material sample. For example, the device in [191] is depending on the scan force and scan angle; both dependencies can be compensated by the usage of multiple bearings, and hence, a constant scan angle, in a novel sensing device.
- The presented 100 material samples in [191] often belong to the same material class, e.g., different types of paper belong to the material class celluloid which leads to critical ambiguities. Additionally, several material classes may not be represented sufficiently if a clear taxonomy was applied.

3.1.1 Initial LMT 108 Material Database

The first collections of material samples led to the initial LMT 108 haptic database (LMT-108-HDB) which was mainly used for the evaluation of novel tactile features and classification approaches as presented in [1] and [7]. Figure 3.1 shows the 108 material classes and the subjectively chosen taxonomy. During the creation of this thesis, however, the necessity of

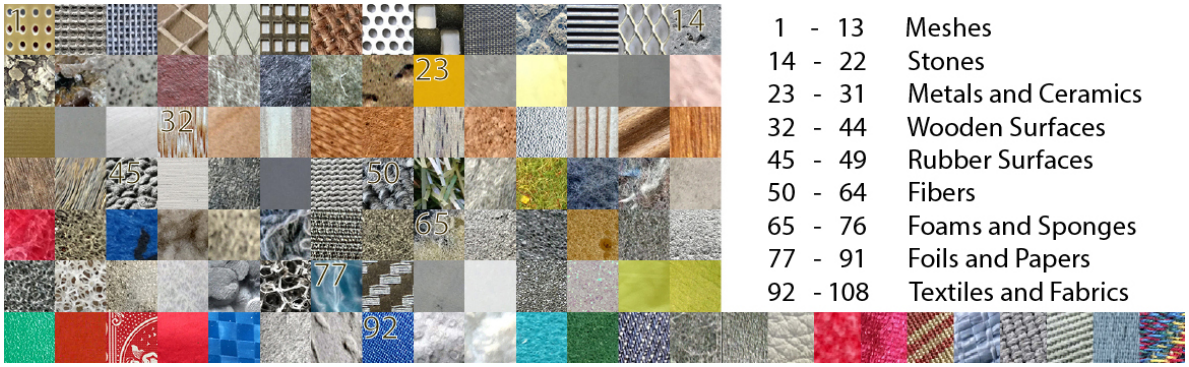


Figure 3.1: Initial LMT 108 haptic database based on a loose subjective taxonomy and made publicly available on the website [15] (adapted from [7] © 2017 IEEE).

an unambiguous taxonomy was identified and lead to a complete redefinition of the material class labels.

3.1.2 LMT 184 Haptic Database

The haptic databases from Culbertson et al. [191], Hassan et al. [192], or the LMT-108-HDB are based on colloquial material names (e.g., stone tile) which can lead to ambiguities across different industrial or commercial branches. For example, a pure granite brick can be pointed (left image in Fig. 3.2), or polished (right image in Fig. 3.2) which leads to completely different perception of the material surface, and hence, different scan results of the same material.

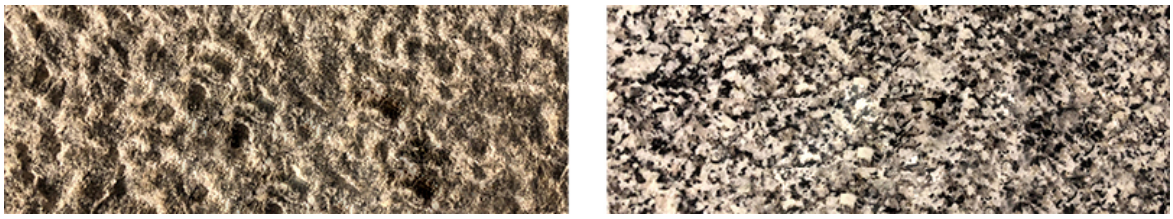


Figure 3.2: Granite is an igneous stone which is, for example, either pointed (left) or polished (right). The difference in the surface structure heavily influences its haptic feel, and comparably, recorded haptic data traces (adapted from [2] © 2019 IEEE).

The collection of a much larger haptic material database was facilitated by another circumstance during the creation of this thesis, namely, the TUM Department of Architecture gave access to its collection of building materials [193]. A large set of the available material samples, in addition to the materials from the previous LMT 108 haptic database [15], was scanned to set up the more extensive LMT 184 haptic database (**LMT-184-HDB**) which follows a novel taxonomy. In the current version of the database, two different human operators scanned each material five times and stored the data traces as .mat files which are made available on the website [15]. Other material archives, e.g., the one presented in [194], indicate which materials still can be scanned in the future.

The proposed naming convention works as follows. C denotes the major material class, e.g., metallic or wooden materials. The identifier S considers the material subclass, e.g.,

hardwood or softwood. M denotes a specific material, e.g., lead, tin, or aluminum on a fine classification level. Many materials come along with different processing shapes P. For example, wood can be veneered or naturally cut from the corresponding tree; aluminum can contain differently shaped holes or gratings, and sandpaper possesses different grit sizes. The applied processing procedures can lead to completely different visual and tangible perception, although the actual material is the same. Another identifier X denotes class-specific attributes. For example, the moisture of woods needs to be taken into account, which influences their density, or metals may contain more or less percentages of their components, especially steel alloys. Figure 3.3 shows a montage image containing each material class.

The taxonomy is mainly inferred from Weber et al. [193] and [195] and allows to categorize existing material recordings as well as to flexibly add new materials. The database currently consists of 88 wooden, 24 organic, 18 plastic, 4 ceramic, 4 glass, 28 stone, 12 metallic, and 6 composite material classes, in total 184. Table 3.1 shows the proposed material classification scheme. A complete list describing each material-characteristic properties is uploaded to the website [15]. Table 3.1 further shows an overview of the taxonomy.

Figure 3.3: Final haptic material database image montage. The different classes C have been separated by black lines. The class names are C_1 woods, C_2 biodegradables, C_3 plastics, C_4 ceramics, C_5 glasses, C_6 stones, C_7 metals, C_8 composites. The class imbalance toward woods actually reflects the true circumstance that this class has the most real world representatives, i.e., types of trees (adapted from [2] © 2019 IEEE).

Table 3.1: Material taxonomy table of the LMT-184-HDB [15] (adapted from [2] © 2019 IEEE).

C_1 woods (natural, organic)					
S	S_1 softwood	S_2 hardwood	S_3 bark	S_4 celluloid	S_5 comp.
P	P_1 veneered, tangential section P_2 tion P_3 veneered, cross section P_4 veneered, sliced/peeled P_5 tangential section P_6 cross section sliced/peeled		P_1 pressed P_2 natural	P_1 pressed P_2 finished P_3 crumpled P_4 corrugated P_5 creped	P_1 pressed P_2 laminated
X	moisture content X_1 less then 10% (dry) X_2 saturated (100%) X_{60} fresh (60%) $X_{10...99}$ moisture in %		bark comp. X_1 pure $X_{2...99}$ cork in %	color X_1 unprocessed brown X_2 bleached X_3 single-colored X_4 multi-colored	
M_1	spruce	alder	cork	cardboard	compressed
M_2	Douglas fir	poplar	-	paper	shavings
$M_{...}$	...	...	-	...	...

C_2 biodegradables (natural, organic)			
S	S_1 plant fiber	S_2 fur	S_3 animal skin
P	P_1 molded P_2 woven	P_1 woven P_2 woven, flaked P_3 natural	P_1 front (fur) side P_2 back (skin) side
X	composition X_1 pure (100%) X_y percentage of main component in y %		originality X_1 original (100%) X_2 artificial
M_1	hemp	sheep wool	cow
M_2	flax	cashmere wool	reindeer
$M_{...}$	...	...	...

C_3 polymers (natural, organic)					
S	S_1 thermopl.	S_2 thermoset	S_3 elast.	S_4 S_1-S_3.	S_5 copoly.
P	P_1 inj.-m., flat P_2 inj.-m., proc. P_3 extruded P_4 blown- extr. P_5 woven P_6 embossed P_7 mesh $>$ 5 mm P_8 mesh $<$ 5 mm	P_1 pressed P_2 molded	P_1 molded P_2 foamed	P_1 extruded P_2 injection-molded P_3 blow-molded	P_1 woven P_2 injection-molded P_3 extruded
X	composition X_1 (100%), X_{51} ... X_{99} (51% – 99%)				
M_1	polyethylene (PE)	epoxy resin (EP)	synthetic isoprene	TPE-O (TPE olefine)	PE/PP
M_2	polypropylene (PP)	melamine formaldehyde (MF)	polychlor-oprene	TPE-V (PE vulcanite)	PE/ABS
$M_{...}$	...	...	...	-	...

C_4 ceramics (mineralic, anorganic)			
S	S_1 earthenware	S_2 sinter	S_3 hightech ceramic
P	P_1 sandblasted (not glazed) P_2 polished P_3 laserblasted P_4 slip-casted		
X	composition X_1 pure (100%) X_y percentage of main ceramic in y %		
M_1	crockery	fine stoneware	high-temperature ceramic
M_2	refractory mixes	porcelain	ferro-electro-ceramic
$M_{...}$	...	...	...

C_5 glasses (mineralic, anorganic)				
S	S_1 crafted	S_2 vulcanite	S_3 tektite	S_4 fulgurite
P	P_1 floated P_2 spun P_3 foamed P_4 flung P_5 corrugated P_6 cut P_7 with ornaments	P_1 unprocessed P_2 sanded		
X	opacity X_1 (transparent), X_2 (nontransparent), X_{51} ... X_{99} (51% – 99%)			
M_1	float glass (soda-lime)	obsidian	moldavite	sand fulgurites
M_2	borosilicate glass	bims	indochinite	rock fulgurite
$M_{...}$	...	...	...	...

C_6 Stones (mineralic, anorganic)					
S	S_1 igneous	S_2 sediment	S_3 metamorph	S_4 cement	S_5 resin
P	P_1 stacked P_2 polished P_3 pointed P_4 sanded P_5 embossed	P_1 embossed P_2 polished P_3 sandblasted P_4 split P_5 sanded	P_1 flamed P_2 sandblasted P_3 polished P_4 pointed	P_1 sandblasted P_2 pointed P_3 embossed P_4 axed P_5 charringed P_6 corrugated	P_1 unprocessed
X	mixture X_1 (100%), X_{51} ... X_{99} (51% – 99%)			ISO DIN EN 197-1 X_1 ...X_5	natural X_1
M_1	granite	congl.	marble	concrete	amber
M_2	syrenite	sandstone	quarzite	mortar	rosin
$M_{...}$	...	...	...	-	...

C_7 metals (metallic, anorganic)				
S	S_1 ferrous	S_2 non-ferrous	S_3 alloy	S_4 metal fiber
P	P_1 steamrolled P_2 polished P_3 casted	P_1 steamrolled P_2 polished P_3 deformed P_4 casted P_5 perforated P_6 corrugated		P_1 woven
X	% iron X_1 (100%), X_{51} ... X_{99} (51% – 99%) X_2 ... X_{49} relevant industrial compositions: - X_2 ... Cr-V, - X_3 Cr-V-Si,...	% of main component X_1 (100%), X_{51} ... X_{99} (51% – 99%)	alloy mixture X_1 (100%) X_{51} - X_{99}	fineness X_1 ... X_9 fine (0000) - coarse (5)
M_1	basic steel	copper	brass	steelwool
M_2	stainless steel	lead	bronce	-
$M_{...}$	...	...	...	-

C_8 composites					
S	S_1: C_3–C_2	S_2: C_3–C_6	S_3: $C_3 - C_7$	S_4: $C_6 - C_1$	S_5: $C_5 - C_1$
P	P_1 pressed P_2 blend P_3 steamrolled P_4 woven			P_1 laminated	P_1 woven
X	mixture X_1 (100%), X_{51} ... X_{99} (51% – 99%)			FEPA grid size/10 X_4 40 μm ... X_{60} 600 μm	composition X_1 100%, X_{51} ... X_{99} (51% – 99%)
M_1	linoleum	tar	alu. polymer	sandpaper	glass fiber paper
M_2	-	carbon fiber reinforced plastic	electric conductive tissue	-	-

3.2 Texplorer Sensor Evolution

As discussed in Section 2.2.2, operator-based handheld tools or robotic setups can act as haptic cameras to scan a material sample, however, no standardized device has been made commercially available so far. The motivation of high flexibility, ease of integrating new sensors, and the possibility of reducing scan-time parameters are the fundamental reasons for this thesis to follow the operator-based concept. Over the past years, an entire device family of texture explorers, or short, **Texplorers** has been developed, assembled and constantly redesigned as shown in Figure 3.4.

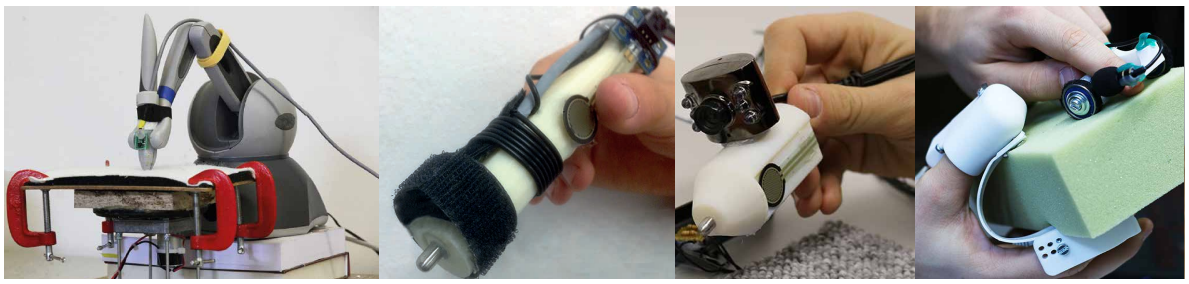


Figure 3.4: Design evolution of the Texplorer device. From left to right: Phantom-based Texplorer version 1.a, Texplorer version 1.b, Texplorer version 1.c, and two units of the final version Texplorer2. All models were designed by using the CAD software Autodesk Fusion 360. Note that the setups also gradually decreased in their component costs and increased in their flexibility of usage. The very first version requires, e.g., a Phantom Omni device, whereas the final version comes along with a low-cost USB data acquisition card NI USB-6002 (National Instruments, USA) attached to a laptop, and is mobile like the Proton Pack presented by Burka et al. [90].

- The first version (version 1.a, left-most image in Fig. 3.4) collected vibrotactile signals during controlled tapping and sliding motions on texture samples by using a gear motor attached to a rotatory plate and a Phantom Omni device (Geomagic Touch) in a controlled procedure.
- Because of limiting constraints in operating workspace and rotatory plate motor noise influence, the next step was to use a custom-made operator-wielded device which included an accelerometer, FSR-based friction force sensor, and an embedded microphone (version 1.b, second-left image in Fig. 3.4).
- The next improvement step of the Texplorer (version 1.c, second-right image in Fig. 3.4) was to add all sensors required to parameterize the five relevant tactile dimensions from Okamoto et al. [63] performed in the thesis-related publication in [7]. Also, the form factor and handling was improved. Figure 3.5 shows the various sensors used for the Texplorer version 1.c and the resulting multidimensional data traces.
- The final Texplorer version (right-most image in Fig. 3.4) considers relevant human exploratory pattern, e.g., folding a material sample, and allows for the characterization of the 15 SynTouch subdimensions from Fishel et al. [66]. It is denoted as Texplorer version 2.0, or in short Texplorer2, and consists of four independent units further explained in the next Section 3.3.

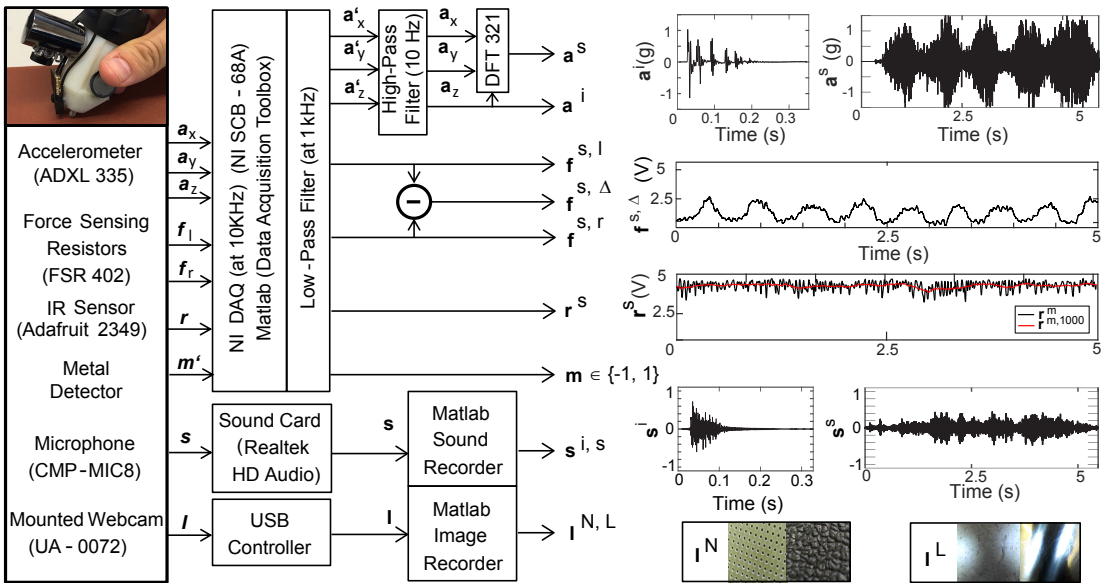


Figure 3.5: Texplorer (version 1.c) system overview. Six sensors provide the input data traces that are used for surface classification based on the five major perceptual dimensions from [63]. The resulting signals (right part of the figure) originate from parallel recordings while tapping on and sliding over the material surface. Tapping on and sliding over the material surfaces leads to acceleration and audio tapping and sliding signals, reflectance scans reveal surface height information, FSR scans are used to infer frictional forces, camera-captured images to visual features, and a metal detector robustly identifies metal surfaces (adapted from [7] © 2017 IEEE).

3.3 Texplorer2

The Texplorer2 was conceptualized as an extendable, versatile, and low-cost operator approach to provide data representing all 15 tactile subdimensions contrasting to expensive robotic scan systems following a similar purpose. The scan system is embedded into a bag to allow for mobile scanning using a laptop, a laptop supporter/tray (Rocket Packs, Germany) and a power bank (Anker, USA) as further inspired by the design of the Proton Pack by Burka et al. [90]. A USB data acquisition card USB-6002 (National Instruments, USA), attached to a Windows 7 laptop Lenovo (Lenovo, USA) 80VR (i7-7700HQ CPU at 2,800 MHz, 16 GB RAM) collects the multimodal sensor data. In addition to the lower cost and simple scanning procedure, the system may scan material surfaces of arbitrary shape, geometry, and size. The operator performs multiple scanning steps with four mobile devices, which are denoted as Texplorer2 units (TUs). All 3D-printed components are made of polylactic acid (PLA).

3.3.1 Recording Procedure

The Texplorer2 units are sequentially used to collect data traces while performing the EPs like static touch or sliding. Figure 3.6 shows the four units of the Texplorer2 and the following list summarizes the general operating principle and the target haptic dimension of each unit. The corresponding human EP is marked as bold.

Figure 3.6: Overview of the Texplorer2 units. The interchangeable components such as iPhone8, operator bag, and laptop tray are not depicted (adapted from [2] © 2019 IEEE).

- TU1 records acceleration and audio signals to calculate microscopic roughness features as well as reflectance data for macroscopic roughness definition during **sliding** and **contour following**. Comparable to the scan dependency ranges in [66] or [183], the typical scan speed of recording devices range from 5 mm/s to 240 mm/s [66] or, is on average 130 mm/s for tool-mediated setups [183]. The operator contact force can be expected to range within 0.2 N and 4.0 N [66].
- TU2 records normal and tangential forces to describe the surface friction, pressure and folding FSR information to describe hardness, and thermal cooling data to describe warmth. Beyond, it acts as metal and magnetic detector. This unit is first placed on the material sample to collect **static touch** information, and thereafter, used to press on and squeeze the material sample (**pressing** EP).
- TU3 captures surface images, magnified surface images, and images with additional illumination using an iPhone 8 (Apple, USA) camera (**enclosure** for visual features and macroscopic roughness).
- TU4 measures the mass (**lifting**) and the volume (**enclosure**) to estimate the density of an object sample by placing the sample on top of the scale and inside the measuring jug.

3.3.2 TU1: Microscopic and Macroscopic Roughness

The first Texplorer2 unit (TU1, shown in Fig. 3.7) is slid by the operator over the surface of an object as indicated in Fig. 3.8. Three bearings ensure that the influence, i.e., the variance, of the applied force is mitigated and the distance between the surface and the IR sensor is held constant.

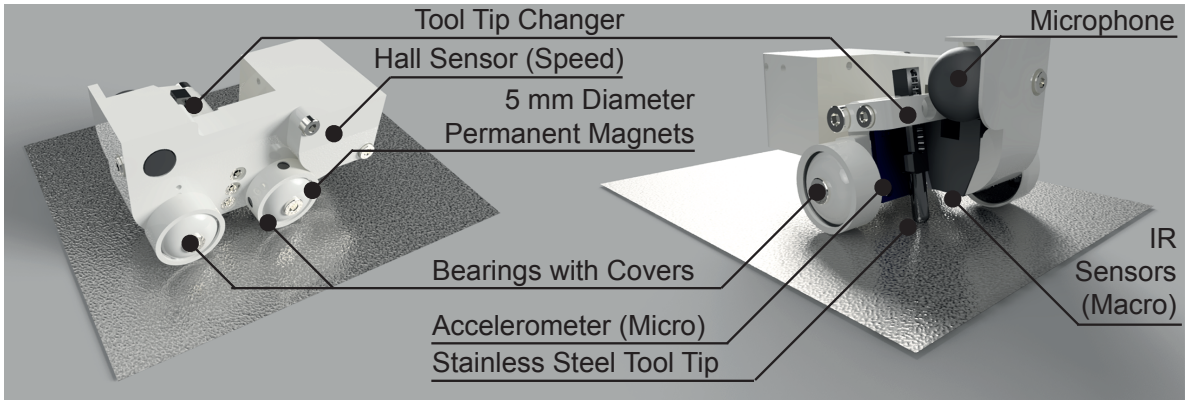


Figure 3.7: Texplorer2 unit 1. The accelerometer can either be equipped with a stainless steel tool tip or be used while sliding the bare finger directly over the material surface to collect vibrotactile signals. In parallel, the microphone captures bare finger- or tool-surface audio signals. The infrared distance sensor collects height information based on the material structure and reflectivity. The scan force is compensated by the bearings during a slide.

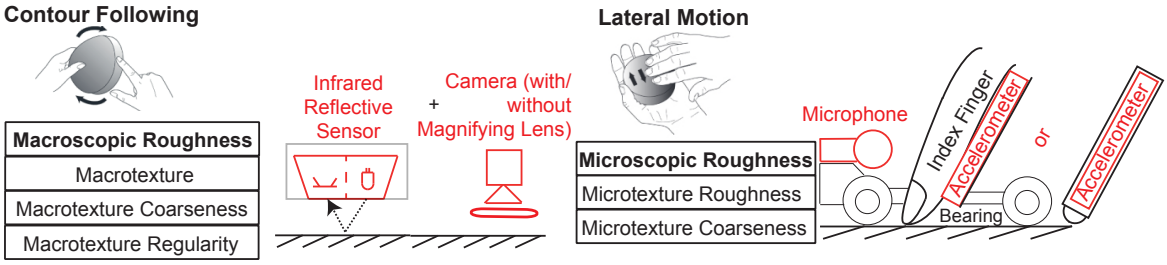


Figure 3.8: The TU1 captures vibrotactile, audio, and height profile data while sliding over a material surface (adapted from [2] © 2019 IEEE).

3.3.2.1 Vibrotactile Signals

A three-axis acceleration sensor ADXL335 PCB (Adafruit, USA) with a range of $\pm$ 3 g (Earth's gravitational constant g = 9.81 m/s^2) records vibrotactile signals to quantify **microscopic roughness** information at a sampling rate of 3,000 Hz. The sensitivity is 300 mV per g and the cut-off frequency $F_{-3\,dB}$ can be manually set by external capacitors following the data sheet guidelines as

$$F_{-3dB} = \frac{1}{2 \cdot \pi \cdot 32k\Omega \cdot C_{x,y,z}} \tag{3.1}$$

Three external capacitors with $C_{x,y,z} = 10$ nF are soldered to the PCB leading to $F_{-3\,dB}$ of 500 Hz, i.e., frequencies beyond are damped to avoid the influence of potential sensor noise.

Bare finger recordings $\mathbf{v}_{f,t}$ during tapping on and $\mathbf{v}_{f,s}$ during sliding over the material surface are recorded. The operator holds the acceleration unit between the thumb and index finger as shown in Fig. 3.8 (upper middle illustration). The index finger is placed on top of the accelerometer and simultaneously touches the material surface directly. Delhaye et al. [196] report that the amplitude of such skin-based vibrations decreases with the square of the distance traveled; consequently, the ADXL335 PCB was rasped, i.e., the unnecessary distance to the accelerometer on the PCB was shortened, and the attachment point is as close as possible to the surface-finger interaction point.

Additionally, tool tip tapping $\mathbf{v}_{t,t}$ and sliding $\mathbf{v}_{t,s}$ signals were collected beside the bare-finger recordings. The operator can switch to a 3D-printed stainless steel tool tip (hemispherical tip with a diameter of 6 mm) that is pressed directly on the ADXL335 and repeats the same scanning procedure. The operator force is compensated by the bearings during these scans. The two types of vibrotactile signals (bare finger-based and steel tool tip-based) are captured for one specific reason. Before data collection was made, it was not examined which signal source is more suitable for robust material classification, and, which signals are more appropriate for haptic display. During the progress of this thesis it was observed that stainless steel tool tip recordings perform better for classification, presumably because they amplify the tool-surface interaction, and the bare finger recordings are intuitively more suitable to be used as vibrotactile signals during haptic display.

3.3.2.2 Audio Signals

Sensor-surface interactions lead to audible information [197], which can be captured by microphones as shown by Kroemer et at. [198]. An omnidirectional capacitive Boya microphone (BY-M1, Shenzhen Jiayz Photo Industrial, China) is mounted on the TU1 to capture audio signals of a frequency range between 10 Hz – 15,000 Hz using a sampling rate of 44,100 Hz both for bare finger $\boldsymbol{a}_{f,s}$ as well as for the stainless steel tool tip $\boldsymbol{a}_{t,s}$ audio recordings, which are recorded in parallel to the vibrotactile signals.

3.3.2.3 Infrared Reflectance Signals

Dulik et al. [102] and Aytac et al. [199] employed infrared reflectance (IR) scans to successfully classify five and eight material samples in pilot studies, respectively. In this thesis, a comparable IR sensor (Adafruit AF2349) is employed consisting of an emitter, a detector, and a 10 kΩ resistor in a voltage divider circuit. One-dimensional reflectance signals $\mathbf{r}$ are captured to quantify macroscopic roughness information. Note that the IR sensor has a fixed distance of 7 mm to the surface which is ensured due to the stable rolling of the three bearings during a slide. For completely flat surfaces, this distance is below the sensing range, and hence, the IR values for flat surfaces are constant which is highly relevant to distinguish between macroscopically and microscopically rough surfaces. However, if a material surface contains coarse structures or holes, the measured distance varies accordingly. Also, the different reflectivity of metals, or the color and brightness information of surfaces can provide useful information for surface classification based on such IR measurements. For example, the average reflectivity $\bar{\mathbf{r}}$ of the IR sensor values is obtained without movement, and hence, completely independent of any scan dependency, since the IR sensor has a constant distance to a material surface. It has been validated by Aytac et al. [199] that matt surfaces follow Lambert's law of reflection in terms of IR signals and that the specular reflections of mirror-like surfaces, especially for metals, can be modeled by parametric curves depending on the scanning angle of the IR sensor. Consequently, IR sensors should always be mounted perpendicular to a surface and device rotations during a scan need to be avoided as ensured by the TU1.

3.3.2.4 TU1 Speed Estimation

The contact of the bearings with any underlying material surface allows for the acquisition of scan speed. The TU1 bearing covers are 3D-printed and contain two excavations holding different numbers of neodymium permanent magnets (NPMs). Specifically, one excavation contains only one magnet, whereas the other embeds five magnets to generate clearly different magnetic field intensities. A hall sensor MAG3110 (Sparkfun Electronics, USA) is attached close to the bearing and measures the magnetic flux density. Due to the different numbers of attached PMs, their significantly stronger magnetic field (compared to the Earth's magnetic field of $\approx 20\ \mu$T), and the well-defined circumference of the bearing (69.1 mm), the scan speed is calculated as

$$v = \frac{2 \cdot \pi \cdot r_{\text{bearing}}}{\Delta t} \tag{3.2}$$

with Δt being the temporal difference between two peaks in the magnetic measurement as shown in Fig. 3.9.

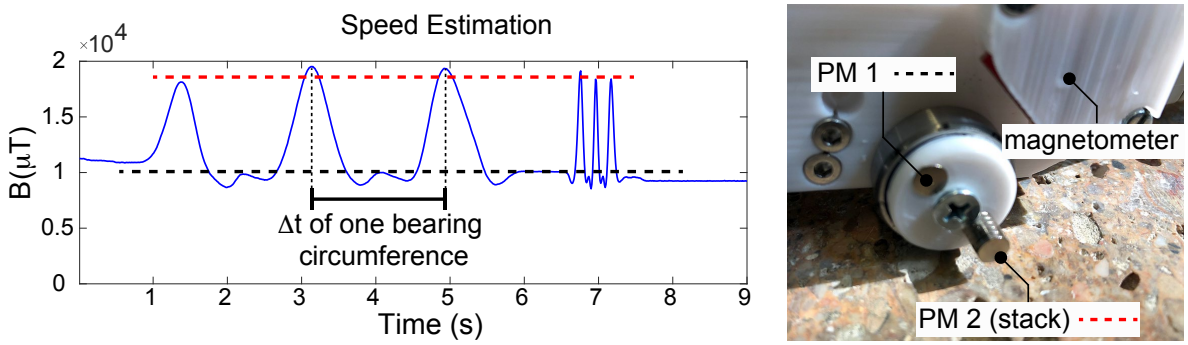


Figure 3.9: TU1 speed estimation using a MAG3110 magnetometer and bearing-embedded permanent magnets. The different magnet stack sizes induce a characteristic magnetic flux pattern for each full bearing rotation (69.1 mm bearing circumference), which is related to the device speed.

3.3.2.5 Data Processing and Segmentation

A vibrotactile scan with TU1 contains a sliding motion signal which is effectively 0.5 s – 2 s long depending on the operator speed. Note that variations in scan speed cannot be avoided, however, the operator is instructed to perform the scans consistently and without unnecessary speed fluctuations. After the recording begins, only accelerometer noise is present until the operator starts moving the TU1. A 0.1 ms frame is shifted over the raw vibrotactile signal to identify the start of the sliding motion. Compared to the signal noise energy, the window frame energy rises during this procedure which leads to the start index. One-second-long acceleration signals are extracted and stored as 1D vibrotactile data. Alternatively, the signals can be segmented manually according to visual stationarity. The audio data indices are the same due to the parallel recording procedure. The vibrotactile signals are band-limited between 10 Hz and 1,000 Hz to remove the effects of human hand motion, aliasing [178], and accelerometer sensor noise (above 1,400 Hz according to the ADXL335 data sheet). The three acceleration signal components are combined into one using the DFT321 algorithm with the

absolute value for each frequency

$$|\mathbf{V}[f]| = \sqrt{|\mathbf{V}_x[f]|^2 + |\mathbf{V}_y[f]|^2 + |\mathbf{V}_z[f]|^2} \tag{3.3}$$

and the phase

$$\angle\mathbf{V} = \angle\sum_{i=1}^{3}\mathbf{V}_i \tag{3.4}$$

as proposed by Landin et al. [200] and used by Culbertson et al. in [46] and subsequent works [106], [201].

3.3.3 TU2: Warmth, Friction, and Stiffness

The second Texplorer2 unit (TU2), shown as a computer rendering in Fig. 3.10, embeds the sensors to capture the data for the tactile dimensions friction, hardness, and warmth as depicted in Fig. 3.11.

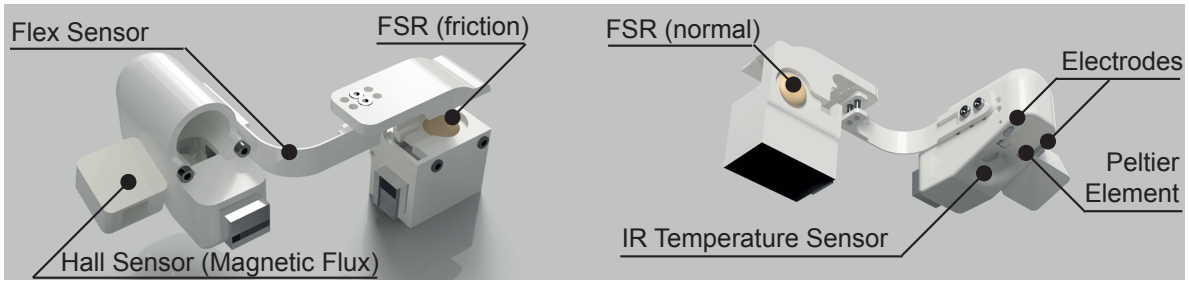


Figure 3.10: Texplorer2 unit 2. The flex sensor and normal force FSR are used to infer hardness-related properties, the normal and friction force FSRs to infer friction characteristics, and the Peltier element and infrared sensor to quantify thermal behavior. Additionally, the metal and magnetic flux sensors provide further independent metal classification.

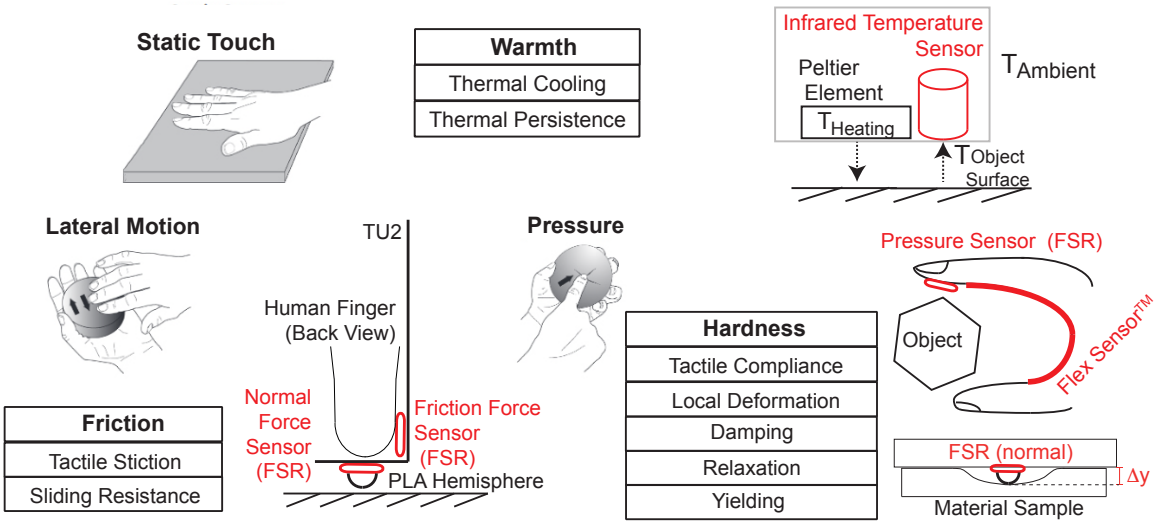


Figure 3.11: Acquisition of thermal, friction, and hardness material information using TU2. Sensing components are shown in red (adapted from [2] © 2019 IEEE).

3.3.3.1 Thermal Signals

Ho and Jones [70] have shown that humans can identify specific materials by their thermal properties alone. Consequently, a technical material scanner potentially benefits from such data traces as well. Two different approaches are conceptually relevant for data acquisition which are explained in the following and are shown in Fig. 3.12. Note that the more intuitive contact-based approach was examined first, but the non-contact-based is preferred for material classification, as explained in the following.

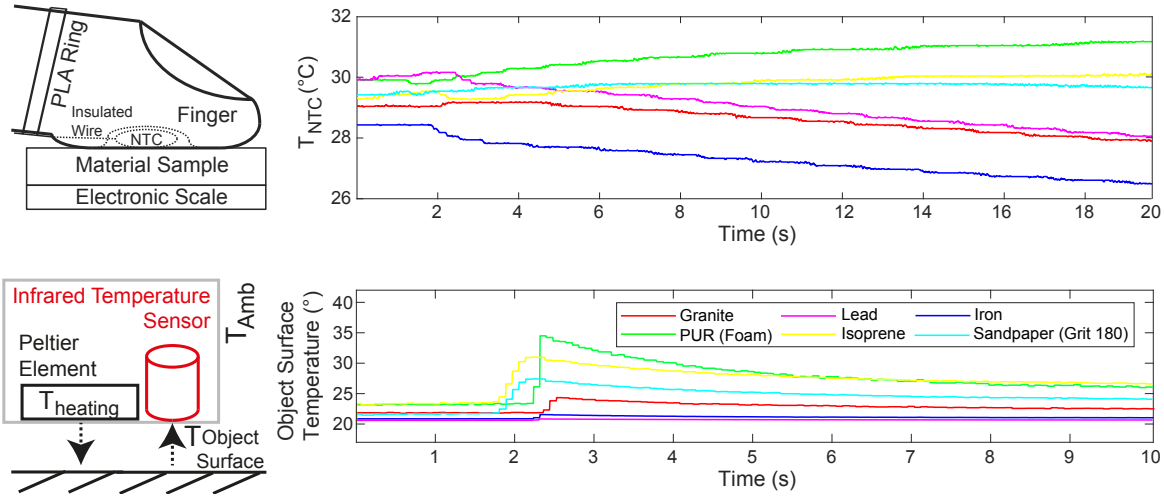


Figure 3.12: Top: Intuitive thermal contact-based approach with human finger involved in the data acquisition, showing thermal data traces for six materials. Note that the initial drop or increase also refers to thermal cooling, and the subsequent sustained cooling or heating (e.g., for the foam sample) is related to thermal persistence. The electronic scale is used to upkeep a constant pressure on the surface, since the normal force effectively influences the contact area, and hence, the thermal conductivity, as shown by Ho and Jones [157]. Bottom: Non-contact thermal sensing approach used for material classification. The Peltier element heats up a specific spot on the material surface, and the subsequent thermal cooling is measured using an IR thermal sensor.

Contact-based Thermal Recordings Contact-based thermal sensor recordings measure the heat flow through the sensor and material surface. However, the precise measurements of the skin temperature, the object surface temperature, and the thermal contact resistance according to the thermal model presented in Section 2.3.5 lead to major device design challenges. Choi et al. [111] attempt to solve this challenge by measuring the heat flux between the finger and surface directly by introducing a flux sensor (FHF01, Hukseflux, Netherlands). However, this sensor constitutes another thermal contact resistance which is critically large (e.g., five-times larger than copper) and hence, introduces a systematic bias to the measurements. Instead, this thesis examines the contact-based approach shown in the lower left part of Fig. 3.12, which consists of an NTC thermistor (LittleFuse, NTC KC502J2K) mounted at a 3D-printed PLA ring and is in contact with the material surface and the bare finger during a scan. When in contact with a surface, heat is drained from the finger into the surface, and hence, bypasses the NTC thermistor while measuring the interface temperature. Two major drawbacks, however, make this approach critical considering data acquisition and robust

feature calculation. Different finger sizes, skin properties, normal force variations which influence the contact area, and effectively, the thermal contact resistance [29], [154], cause a strong dependency on the operator finger constitution. Consequently, many recordings need to be done in similar data-driven approach as for microscopic roughness across different operators and normal forces. Secondly, the operator finger temperature needs to regain its initial value before touching another material sample, which can take minutes if a thermally conductive material was scanned before. This fact can already been observed in Fig. 3.12 since the temperature curves at $t = 0$ vary greatly among the material samples, yet the ambient room temperature was nearly constant, indicating that the scan time needs to be much larger which becomes impractical for larger datasets.

Non-contact-based Infrared Recordings As a result of the issues mentioned with contact-based recordings, non-contact based approaches based on Newton's law of cooling (as introduced in Section 2.3.5) are considered in the following. In this thesis, a thermography approach comparable to the one applied by Aujeszky et al. [94] infers thermal properties of the material samples. The corresponding setup is shown in the upper right part of Fig. 3.12. It actively heats the material surface with a constant input thermal energy Q and measures the subsequent local thermal cooling. A Peltier element is heated for three seconds while TU2 is being placed on the material surface. The maximum surface temperature of the Peltier ceramic is about 44°C at an ambient room temperature of 21°C. Subsequently, the unit is moved to measure the thermal cooling rate using an infrared temperature sensor MLX90614 (Melexis, Belgium) for ten seconds. The recorded temperature data array is denoted as **t**. The remaining effect of the ambient temperature, which is the only thermal scan dependency, on these measurements is further investigated in Section 4.2.2.

Note that the contact-based approach still has potential applicability for thermal display approaches since the data is highly interpretative for each material. However, the proposed thermal model by Jones et al. [29] incorporates the relevant material properties comparably without any human-related dependencies for display; and on the acquisition side the shown Texplorer2 non-contact approach only depends on the ambient temperature, and is hence preferable for material classification.

3.3.3.2 Force Sensitive Resistor Signals

The TU2 contains two FSR400 sensors (Interlink, USA) for normal and tangential sliding force measurement as shown in the middle part of Fig. 3.11. Surface friction information is recorded by placing the index finger inside the TU2 and performing a sliding motion across the object surface. The force arrays $\mathbf{f}[\mathrm{n}]$ from a scan are denoted as $\mathbf{f}'_{\mathrm{fr,n}}$ and $\mathbf{f}'_{\mathrm{fr,t}}$, respectively. The raw FSR sensor values $\mathbf{f}'$ grow logarithmically with increasing pressing force and each voltage value $f' = \mathbf{f}'[n]$ is converted to a force value f using the expression

$$f = 9.81 \cdot \mathrm{m/s^2} \cdot 0.01\ \mathrm{kg} \cdot e^{f' \cdot 1.33\ \frac{1}{\mathrm{V}}} \tag{3.5}$$

according to the inverse force-voltage relation diagram (Interlink FSR integration guide). This conversion leads to the normal and friction force arrays $\mathbf{f}_{\mathrm{fr,n}}$ and $\mathbf{f}_{\mathrm{fr,t}}$ collected when

the operator performs a slide over the material surface. Note that the normal FSR, which contacts the surface, is equipped with a 14 mm diameter flattened PLA hemisphere which ensures firm contact and a constant area with any scanned material surface as well as a constant distance Δy = 4 mm between a material sample and the entire TU2 casing.

The FSR which measures the normal force for friction data collection is also used to infer **hardness** properties. An additional Flex sensor FS-L-0095-103-ST (Spectra Symbol, USA) collects the folding angle $\boldsymbol{\alpha}_\mathrm{f}$ while the normal force FSR records the folding force $\mathbf{f}_\mathrm{f}$ in parallel.

3.3.3.3 Metal and Magnet Detector Signals

Material classification is not limited to specific human perception capabilities, and hence, the Texplorer 1.c [7] and Texplorer2 [2] are equipped with electric conductivity and magnetic flux measurement components. Two electric conductive tissues are used as electrodes (see Fig. 3.13) during the static contact with a material to identify metals as performed in [7]. This procedure leads to a binary array $\mathbf{m}[\mathrm{n}] \in \{0\ \mathrm{V}, 5\ \mathrm{V}\}$, which is mapped to the values $\mathbf{m}[\mathrm{n}] \in \{1, -1\}$. Only electrically conductive materials, e.g., metals, pull down the DAQ input to ground (0 V) during scan time, leading to an $\mathbf{m}[\mathrm{n}] = \mathbf{1}$ array. A binary value $isMetal = 1$ is set if $\overline{\mathbf{m}} > 0.5$.

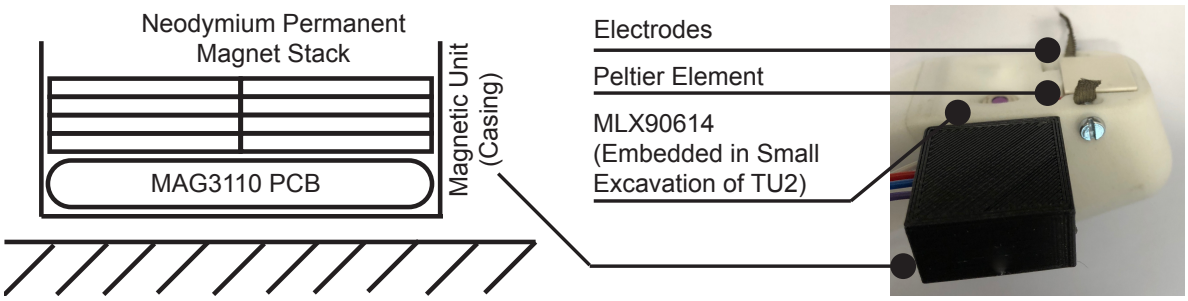


Figure 3.13: Close-up view of the static touch part of TU2. The electric conductive tissues reliably detect metals, the Peltier element and MLX90614 measure thermal cooling, and an additional magnetometer unit (magnified in left part of the image) instantly detects the change of the external magnetic field caused by the neodymium stack through a material sample (adapted from [2] © 2019 IEEE).

TU2 further measures the change of a magnetic field through a material sample. More specifically, a stationary external magnetic field is applied in a constant distance (thickness of casing) which is influenced by the underlying material sample. A 3D-printed casing contains a 3-axis digital magnetometer MAG3110 (Sparkfun Electronics, USA) PCB and a stack of eight neodymium permanent magnets, shown in Fig. 3.13 (right). The MAG3110 is zero-calibrated based on these permanent magnets which expose a strong and constant magnetic field to rule out the influence of the Earth's magnetic field. The operator places this unit on a surface, and the resulting material-specific magnetic flux data traces are denoted as $\Delta\mathbf{B}$, and are shown in Fig. 3.14 for selected material sample recordings. Although not a tactile property, the change of the magnetic field characterizes several metals and their alloys by reliably distinguishing between ferro-, non-, and diamagnetic materials. Features based on metal or magnetic material detection exceed human material classification capabilities, are

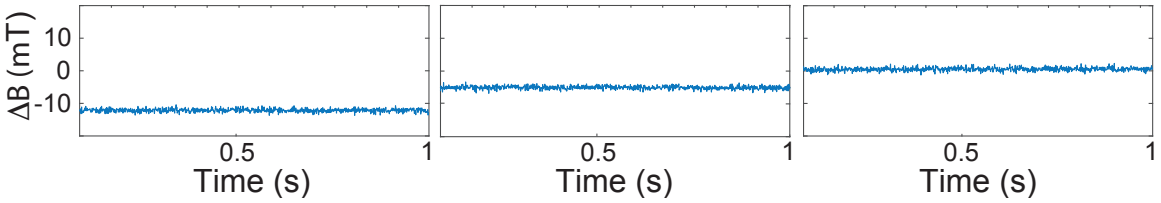


Figure 3.14: Recorded magnetic data during static touch. Cast iron (left) reveals significant magnetic properties (ferromagnetic), steel wool (middle) is paramagnetic, and aluminum (right) does not change the magnetic field. Diamagnetic materials, which are currently not part of the database, would generate positive values. All other materials lead to a mean value of $\overline{\Delta \mathbf{B}} = 0$ (adapted from [2] © 2019 IEEE).

not influenced by other sensing domains (such as vibrotactile signals), and hence, lead to further independent characteristic features.

3.3.4 TU3: Visual Properties

Visual information is linked to human haptic perception as summarized by Lacey et al. [68] and contains information describing relevant material properties such as regularity or macroscopic roughness. The Texplorer2 system relies on a third unit based on a (smartphone) camera denoted as TU3.

3.3.4.1 Camera

The images of the materials, denoted as $\mathbf{I}_{\text{disp}}$ for potential visual display during haptic rendering, are captured under varying distance and rotation conditions using an iPhone 8 (Apple, USA) camera with a resolution of 3,024 x 4,032 pixels. The camera flashlight is further used to freely capture illuminated images $\mathbf{I}_{\text{illu}}$ which provides information about the object's surface reflectivity. To further infer macroscopic details, and hence, macro images $\mathbf{I}_{\text{macro}}$, a ten-times zoom magnifying lens (TaoTronics, China) is attached to the iPhone 8. Note that the distance between each surface and the camera was held constant using a 3D-printed bumper (2.5 cm distance) to ensure the same recording distance and perspective.

All images ($\mathbf{I}_{\text{disp}}$, $\mathbf{I}_{\text{macro}}$, $\mathbf{I}_{\text{illu}}$) were collected based on a free choice of the scan parameters rotation angle and ambient illumination to deliberately maximize the variance caused by operators and during subsequent scans. This allows to test the robustness of the image features in Section 4.1.1. As another minor contribution, this thesis presents an improved TU3 camera setup. An ambient light LED stripe (Inion White, China) is embedded into a truncated cone-like 3D-printed distance holder case shown in Fig. 3.15. This improvement guarantees a constant illumination and 10-cm-distance condition, and hence, removes the major image dependencies; only the recording rotation of the images is left as dependency as shown in Fig. 3.16. Section 4.1.1.3 covers the rotation dependencies for a subset of the material images during a pilot study to further show that the proposed improvement is beneficial for image-based surface classification. Nevertheless, the LMT-184-HDB images are used without the TU3 extension for the material classification to further show that features already can account for a specific operator-induced data variance.

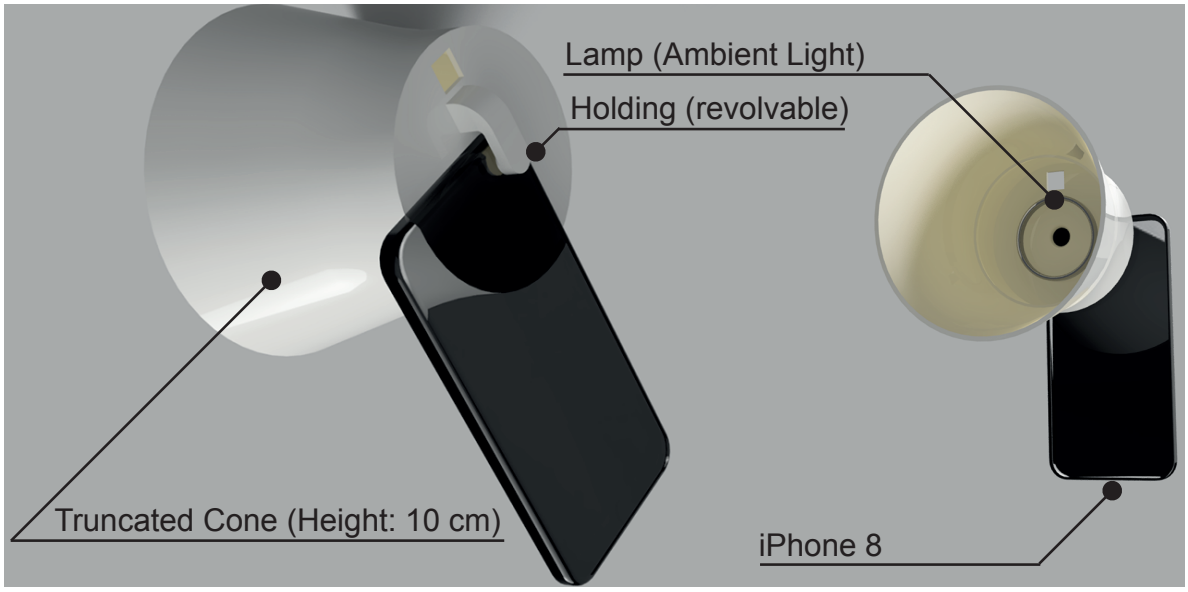


Figure 3.15: Extension of Texplorer2 unit 3. A smartphone camera captures 4K resolution surface images in a constant distance of 10 cm and fixed ambient illumination conditions.

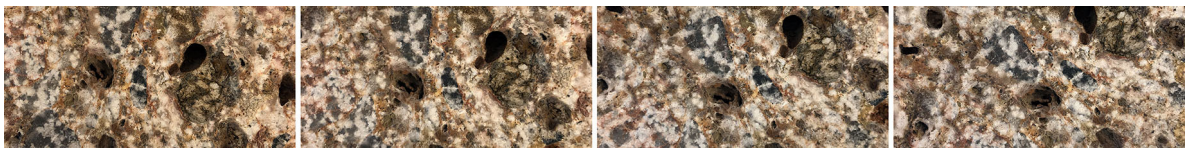


Figure 3.16: Images captured by the TU3 extension considering different rotation angles. The ambient lighting, fixed distance, and perspective angles ensure similar image conditions.

3.3.4.2 Data Processing

100 pixels are removed from each border of the image to reduce potential lens artifacts. Blurry pictures due to focal point inconsistencies were recaptured.

3.3.5 TU4: Volume and Mass

The density

$$\rho = \frac{m}{V} \tag{3.6}$$

is one of the most characteristic material properties and has been published online for the majority of all available materials (e.g., in [202]). However, for very complex object geometries and undocumented materials, a density measurement approach is required.

The acquisition of its first component mass m is trivial. The different material samples are placed on an electronic scale (Smart Weigh Scale II, Brewista, USA) with a measurement range of 0.1 g – 2 kg in order to measure the mass m.

The estimation of the volume V instead is challenging and heavily depends on the object shape. Generally, laser scanners and stereoscopy-based methods are used to extract point cloud mesh data of large-scale objects. However, hollow and reflective material samples pose significant challenges to robust data acquisition. Instead, a measurement approach based on

volume displacement (i.e., Archimedes' principle) can determine the overall volume of the objects regardless of shape, the presence of holes and concavities, and reflectivity. For example, a set of measuring jugs can be filled with water and the object is placed inside. Instead of using water which may damage certain objects, the thesis related work [2] uses micro-pearls as fill material to infer the object volume. The displacement is proportional to the volume of the object. If the object could be damaged by water, fine granular micro-pearls are used instead. To read the measuring scale accurately and to avoid uneven micro-pearl distribution, a vibration plate (NewGen, USA) is placed below the jugs and the operator applies a constant 20 Hz vibration before each manual reading of the scale as shown in Fig. 3.17.

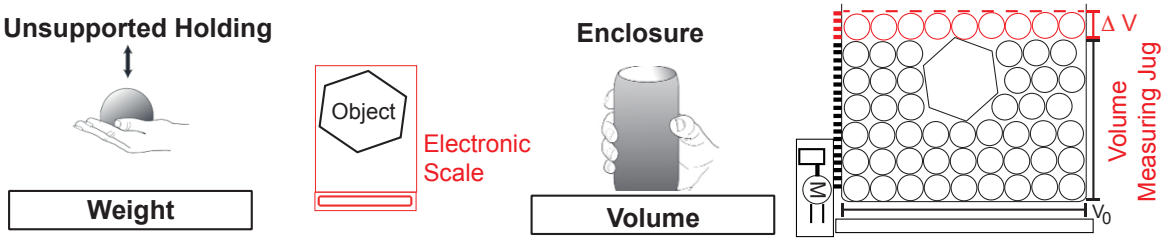


Figure 3.17: Volumetric approach based on Archimedes' principle. Different measuring jugs are used to infer the volume of arbitrarily-shaped objects by displacing either water or 1-mm micropearls (adapted from [2] © 2019 IEEE).

3.4 Chapter Summary

The presented Texplorer device family was specifically designed and constantly improved to collect a multimodal set of data traces relevant for both material classification and haptic display. Figure 3.18 shows all current recording modalities for one material sample. These

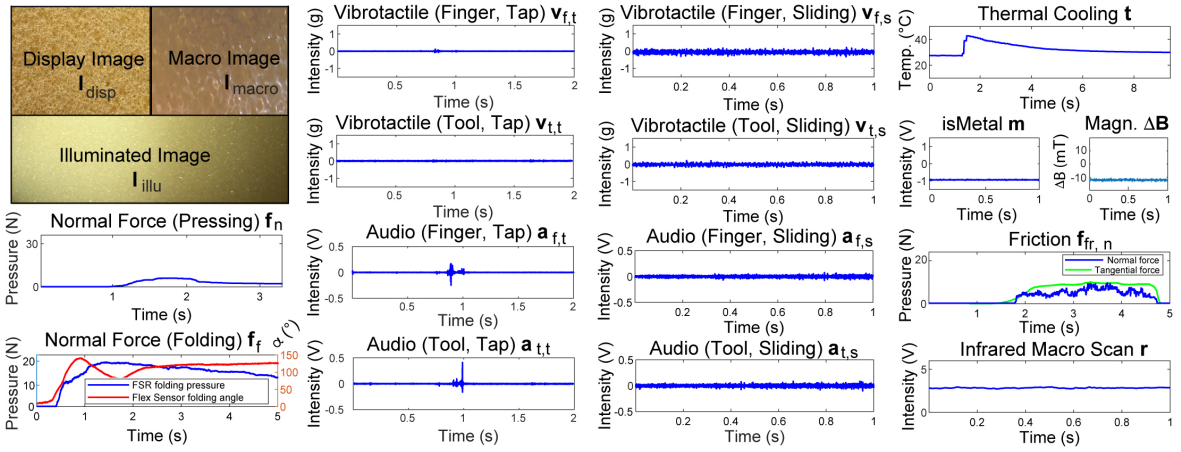


Figure 3.18: Example of one recorded data entry. Overall, twenty different modalities provide the feature data source for material classification (adapted from [2] © 2019 IEEE).

database entries are passed to the machine learning-based classification pipeline in Section 4 as raw input data. Future work based on the Texplorer2 may introduce further sensing domains such as capacitance or infrared spectroscopy.

Chapter 4

Supervised Material Classification

One of the primary contributions of this thesis is to predict a material based on the sensory output from the presented Texplorer devices. The umbrella term for this task is Machine Learning (ML) which is categorized into three major domains according to [203]:

- Supervised Learning predicts classes based on a previously labeled training dataset.
- Unsupervised Learning identifies hidden structures in unlabeled datasets.
- Reinforced Learning approaches learn actions to improve an interactive decision process based on a penalty or reward strategy.

Supervised classification is the most common task in machine learning [203]. Since the recorded materials have unique names to learn from, this thesis focuses on the application of supervised learning algorithms which were successfully applied in other engineering domains such as object recognition or audio genre classification. The subsequent sections address the components of supervised material classification and are structured as follows.

- Multimodal features from various domains (image, audio, tactile) are described in Section 4.1.
- Data preparation consisting of feature space creation, data partitioning, reduction, scaling, missing and inconsistent data handling, and single feature optimization is shown in Section 4.2.
- The choice of classifiers and model creation is examined in Section 4.3. Subsequently, the model evaluation of the handcrafted (HC) features and optimization of hyperparameters is explained. An alternative approach based on human subjective ratings and the corresponding subjective material classification is discussed in Section 4.4 and compared to the sensorized Texplorer approach.
- Deep learning-based (DL) material classification is evaluated in Section 4.5 and further compared to the handcrafted feature results. Moreover, a fusion of these different domains that outperforms the individual approaches is presented.

Figure 4.1 further visualizes the customized workflow for material classification as conducted in this thesis.

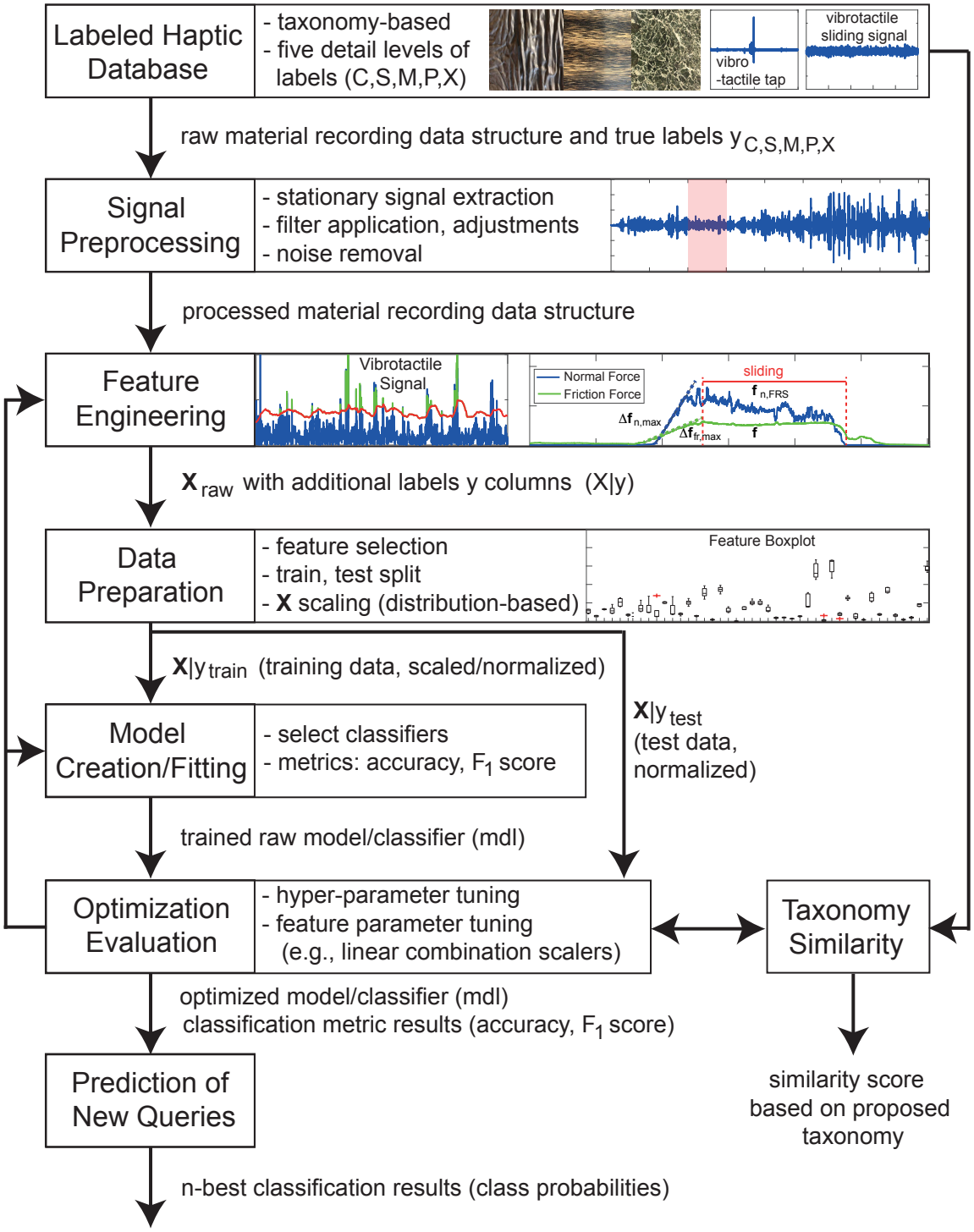


Figure 4.1: Machine learning workflow for material classification. The accuracy, F_1 score, and similarity of the misclassified labels to the taxonomy quantify the system performance.

4.1 Handcrafted Features for Material Classification

The multimodal variety of sensing data introduces specific domain-related dependencies which need further consideration during feature design. For example, vibrotactile signals recorded by pen-like devices [91], [105] vary heavily with respect to the scan parameters speed and force. To the current state of the art, it is infeasible to model the microscopic contact conditions of the tool tip - surface interaction [181]. Instead, this thesis focuses on reducing the operators variations of the data acquisition steps as well as to design more robust features. Other recording modalities are likewise prone to scan parameters, for example, thermal measurements depend on the ambient temperature, or surface images depend on camera rotation, distance, and illumination conditions. It is not possible to remove all dependencies, however, the improved scan conditions of the Texplorer2, introduced in Section 3.3, mitigate their influence on the individual scans, and hence, facilitates robust feature design as shown in the following list.

1. **Vibrotactile and audio signals:** Since the operator-introduced force is mainly compensated by the TU1, the scan speed remains the only major dependency. However, the operator is instructed to wield the device in a reasonable speed range of human material exploration, i.e., within 40 mm/s – 120 mm/s [34], [183], which further decreases the signal speed dependency.

2. **Reflectance signals:** Since the TU1 embeds a reflectance sensor in a constant distance, this sensor has no remaining dependency anymore. Nevertheless, the distance conditions may slightly change for soft material samples if the TU1 deforms the material during a slide. This undesired material indentation can be prevented if two rigid and flat bars, comparable to rails, are placed on the material sample to allow fluent sliding of the TU1 over such soft samples.

3. **Thermal signals:** The ambient temperature mainly determines the object surface temperature and needs to be recorded before each scan session starts.

4. **Friction-related force signals:** The overall applied normal force and finger contact area with the FSR sensor placed inside TU2 differs between different operators.

5. **Hardness-related force signals:** The Texplorer2 allows for a fixed indentation depth (see stress-relaxation test from Section 2.3.1), and hence, the resulting force signals for soft materials are invariant of the operators overall pressure force. Moreover, the folding motion allows for an unambiguous identification if the material is rigid or not with respect to common human surface exploration pressure forces.

6. **Image data:** Several algorithms applied in computer vision are robust to variations in, e.g., image rotation or scaling. Ideally, the operator captures sharp high resolution surface images.

7. **Metal and magnet sensor signal:** These signals are not influenced by any environmental parameter and are completely independent of the aforementioned sensing domains.

Historically, image texture features were examined first, and hence, are discussed prior to the audio and tactile features.

4.1.1 Texture-based Image Features

Material surfaces reveal visual clues to humans and allow for material identification even before the material sample is touched. Hence, textural features resulting from a visual data collection device such as a camera are of interest to allow for vision-based surface classification. Visible surface texture is a repetitive pattern in which geometric primitives are arranged following a placement rule [204]. Two classification schemes - statistical and structural - have been identified to describe image texture features. Note that the structural features partly depend on image statistics as well, and on the other hand, a part of the statistical features revealed an intuitive association in terms of structural elements. Hence, a strict separation between both approaches is not given.

Liu et al. give a comprehensive overview of the last two decades of image-based texture classification in [205]. The description of all presented approaches goes beyond the scope of this thesis. One of the main observation in [205] is that handcrafted feature approaches as well as deep learning-based approaches enable a robust classification of the existing image texture databases. However, these databases only contain a small set of all available materials and their processing shapes. Based on the interdisciplinary taxonomy from Section 3.1.2, the LMT-184-HDB is consequently a suitable candidate for further evaluation of image texture features for both handcrafted and deep learning-based approaches. Note that the deep learning-based approach are discussed after all handcrafted features in Section 4.5. As for the handcrafted features, the most promising surface texture feature definitions from Haralick et al. [206] and Tamura et al. [204] using the corresponding Matlab implementations from Monzel et al. [207] and Sdhir et al. [208] have been used and adapted to the surface images. Additional features like the dominant color (I_{col}) or edginess (I_{edg}) from the thesis-related work in [1] and the rotation-invariant local binary pattern (RLBP) descriptor proposed by Ojala et al. [209] are also considered. In total, 37 appropriate image features are selected beforehand and applied to $\mathbf{I}_{\text{disp}}$, $\mathbf{I}_{\text{macro}}$, and $\mathbf{I}_{\text{illu}}$ respectively, leading to 111 handcrafted image features in total for the creation of the initial feature space.

4.1.1.1 Statistical Image Texture Features

Statistical texture is defined as a functional combination of a texture element and its placement rule in an image which can be measured by statistics of local properties. Notably, the importance of pixel neighbor statistics for texture classification has been identified by Haralick et al. [206] in 1973. In general, k-adjacency of a pixel $\mathbf{p}$ at coordinates (x,y) is given by the k neighbors horizontally (x $\pm$ d,y), vertically (x, y$\pm$ d), and diagonally (x $\pm$ d,y $\pm$ d) with a usual distance of $d = 1$ pixel. If a M-by-N grayscale image $\mathbf{I}_{\text{g}}$ contains K possible intensity levels, each entry of the resulting K-by-K gray-level co-occurrence matrix (GLCM)

G is calculated as

$$\mathbf{G}^{1,0}[i,j] = \sum_{n=1}^{N}\sum_{m=1}^{M} I(\mathbf{I}_{\mathrm{g}}[m,n] = i \text{ and } \mathbf{I}_{\mathrm{g}}[m+1,n] = j) \tag{4.1}$$

for a specific neighbor direction condition, e.g., the neighbors of p to the right with (x + 1, y) and step size 1. Note that the sums over the indicator function accumulate the occurrences of specific pixel pairs [177]. These entries further can be represented as pixel-pair probability matrix $\mathbf{P}^{1,0}$ using the normalization

$$\mathbf{P}^{1,0}[i,j] = \frac{\mathbf{G}^{1,0}[i,j]}{\sum_{i=1}^{K}\sum_{j=1}^{K}\mathbf{G}^{1,0}[i,j]} \tag{4.2}$$

Based on normalized GLCMs, Haralick et al. [206] defined a set of 14 statistical features such as energy content or local homogeneity. For example, images with homogeneous surface texture generally reveal large values at the main-diagonal of **G**, whereas high-frequency spatial patterns lead to larger off-diagonal entries. If other directions are considered, the influence of rotation can be mitigated when the GLCMs are averaged [177]. It was further observed by Haralick et al. [206] that a subset of their defined statistical image texture features provide an intuitive understanding in terms of textural pattern description. For example, the angular momentum was observed to be related to the size of homogeneous areas in the image, the average correlation is an indicator for linear dependencies (e.g., line-like structures), or the average entropy describes the presence of complex structures (e.g., coarse structures). Note that this thesis identified to use such GLCMs as input for deep learning-based classification to learn hidden features from the pixel neighbor relations as well in Section 4.5.3.2.

Another statistical image feature approach was developed and extensively discussed by Ojala et al. [209] which relies on adjacent patterns for the definition of local image texture features. Their concept of rotation-invariant local binary patterns (RLBPs) has proven to be robust against image rotations, and hence, is suitable for texture identification, notably, if the Texplorer2 unit 3 extension is used. This thesis uses the Matlab function extractLBPFeatures() which implements the algorithm from [209]. The rotation-invariance has been proven for the OuTex dataset, and is also evaluated on different samples of the LMT database in this thesis as shown in Section 4.1.1.3.

4.1.1.2 Structural Image Texture Features

Structural features represent human descriptions of visual texture patterns. Tamura et al. [204] summarized six basic texture features, namely, coarseness, contrast, directionality, line-likeness, regularity, and roughness, which follow this concept. Note that regularity and roughness, however, were defined as combinations of the other features, and hence, their definitions are highly correlated and not used in this thesis. Tamura et al. [204] admitted that features based on human tactile perception such as roughness are challenging to quantify with pure vision-based data and thereby indicated the necessity of touch-related tactile features.

Image Contrast and Edginess The first structure-related feature related to human perception considers the visual sharpness of texture edges. Tamura et al. [204] defined four factors for image contrast that go beyond image quality, i.e., these conditions also have a perceptual meaning in terms of texture. In detail, the histogram variance σ^2 of a grayscale texture image alone is not sufficient to capture the sharpness of edges, i.e., a plateau-like distribution can have the same variance as two extensive histogram peaks; that is why Tamura et al. further included the kurtosis $\alpha_4 = \frac{\mu_4}{\sigma^4}$ in their contrast calculation which leads to

$$I_{\text{con}} = \frac{\sigma^2}{\alpha_4^n} \tag{4.3}$$

with n = 0.25 being an empirical value which was chosen to minimize the feature correlation in their work. A similar contrast-detecting feature defined in this thesis is denoted as edginess I_{edg}, which is correlated to the presence of sharp edges in a texture image. The M-by-N input image is transformed to the frequency domain and multiplied with a 2D Gaussian low-pass filter $\mathbf{H}$ [177]

$$\mathbf{H}[m,n] = e^{-\frac{\sqrt{(m-M/2)^2+(n-N/2)^2}}{2\cdot\sigma^2}} \tag{4.4}$$

and thereafter transformed back to the spatial domain to obtain $\mathbf{I}_{\text{low}}$. The variance σ^2 is set to a small value to achieve a specific low-pass filtering effect, e.g., to 1% of the image width. The final value of σ^2 is determined using a grid-based search approach on the dataset during feature optimization. The low-pass filtered image $\mathbf{I}_{\text{low}}$ is subtracted from the original grayscale image, and the overall mean of the absolute difference of both images determines the rotationally-robust edginess feature as

$$I_{\text{edg}} = \frac{1}{M \cdot N} \sum_{m=1}^{M} \sum_{n=1}^{N} |\mathbf{I}_{\text{g}}[m,n] - \mathbf{I}_{\text{low}}[m,n]| \tag{4.5}$$

If this feature is calculated on the illuminated images $\mathbf{I}_{\text{illu}}$, the illumination dependency of the feature is inherently addressed as well. The same considerations are observable for I_{con}, since the flashlight illuminance intensity of a smartphone camera generally exceeds environmental lighting conditions. This is an example of how existing features can be improved by not changing their definition, but their underlying data, which depend on an improvement in the capturing conditions, i.e., the the application of an external light source in this case.

Anisotropy, Directionality, and Linelikeness Anisotropy, or directionality, defines the global property of a dominant visual orientation in an image that is characteristic for woods or man-made structures containing straight lines. Tamura et al. [204] applied the two directional Prewitt filters

$$\boldsymbol{\Delta}_h = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{\Delta}_v = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \tag{4.6}$$

on the images to obtain the gradient images $\mathbf{G}_h$ and $\mathbf{G}_v$. The presence of large pixel intensities in the averaged gradient image $\mathbf{G} = 0.5 \cdot (|\mathbf{G}_h| + |\mathbf{G}_v|)$ with the major orientation $\theta = tan^{-1}(|\mathbf{G}_v|/|\mathbf{G}_h|)$ is associated with the presence of a dominant direction. Consequently, the histogram $hist(\mathbf{G})$ over the bins 0°...180° is examined. The amount of large peaks then indicates the type of directionality (e.g., bidirectional for two significant peaks) and their relative size quantifies its strength. In terms of natural textures, further filters seem to be relevant, however, heavily increase the computational complexity. Instead, another relevant image transform applicable in this context is the Hough transform [210]. A point $\mathbf{p}$ at the image indices (m,n) can be passed through by an infinite set of lines that are perpendicular (i.e., normal) to the line l(ρ, θ) which always crosses the origin of the x-y plane, leading to the normal representation

$$\rho = m \cdot cos(\theta) + n \cdot sin(\theta) \tag{4.7}$$

Other points also possess these infinite amounts of lines passing through them. In the transformed $\rho - \theta -$ space, the resulting sinusoidal curves have intersections whenever multiple points lie on a straight line in the spatial x – y – domain. Note that the $\rho - \theta -$ space can be limited between $\theta = 0°...\ 90°$ and $\rho = \pm\sqrt{M^2 + N^2}$ (original image diagonal length). The Hough transform is hence also suited to determine anisotropy, which is a synonymous term for directionality. In this thesis, the grayscale conversions of surface material images $\mathbf{I}_\mathrm{g}$ are

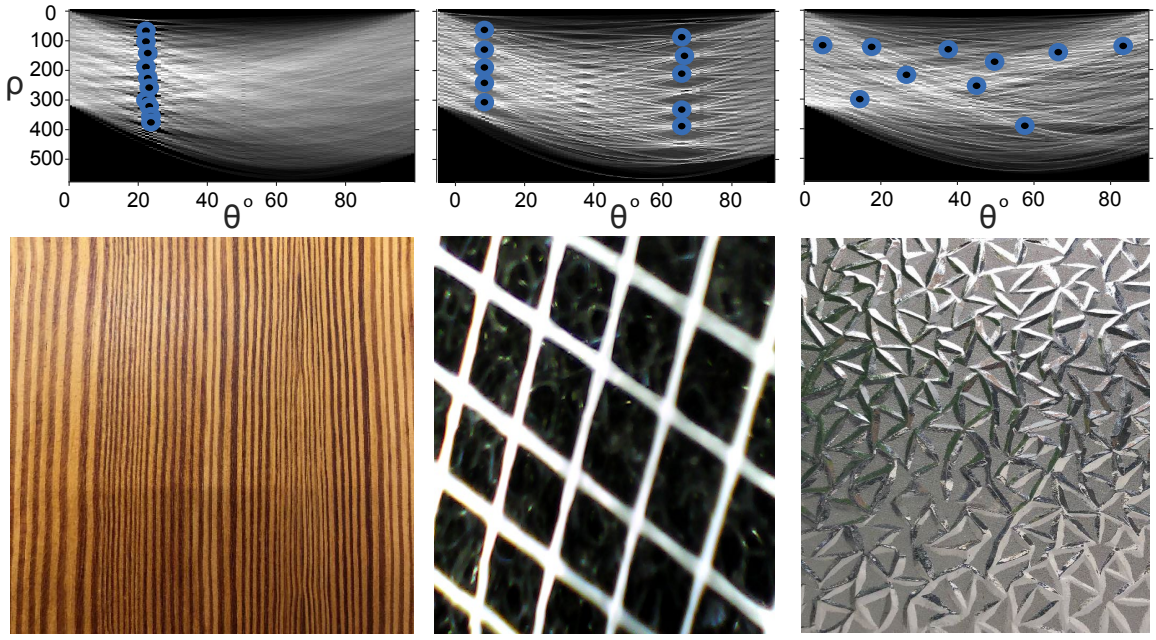


Figure 4.2: Custom image anisotropy rotation-invariant feature based on Hough transform (adapted from [7] © 2017 IEEE).

calculated and a Canny edge detection ($\sigma = 3$) algorithm using the Matlab image processing toolbox function edge() is used to obtain the binary edge-enhanced images $\mathbf{I}_\mathrm{edge}$. Then, the Hough-Transform $H(\cdot)$ is applied to identify the indices of the ten largest peaks in $H(\mathbf{I}_\mathrm{edge})$, which are stored in $\mathbf{P}_{10}(\rho, \theta)$ containing 2-by-10 entries of ten (ρ, θ)-pairs. Figure 4.2 shows the Hough transforms for three surface images $\mathbf{I}_\mathrm{disp}$. The θ values (x-axis) of $\mathbf{P}_{10}(\rho, \theta)$, de-

noted as vector θ_{10}, encode the anisotropy of the texture elements. Note that bidirectional structures like meshes (middle plot) reveal peaks at two dominant values for θ (8° and 63°), whereas anisotropic surfaces (left plot) reveal peaks only at one θ value (at 20°). Irregular or isotropic surface images reveal peaks at arbitrary θ values. This feature is denoted as image anisotropy as

$$I_{\text{an}} = \frac{1}{1 + \sigma(\theta_{10})} \cdot sig(I_{\text{con}}) \tag{4.8}$$

Higher visual anisotropy leads to smaller $\sigma(\boldsymbol{\theta}_{10})$ values. The contrast feature from Tamura et al. [204] is range-limited between 0 and 1 using the sigmoid function $sig()$. This term decreases the I_{an} value for visually isotropic surface material images, such as metals, papers or plastics, but on the contrary, increases the feature value for strongly visual line-like texture elements.

Extensions of the Hough transform to other geometrical primitives are similarly relevant for textural feature definition. Another visual texture dimension, Line-likeness [204], characterizes the shape of the texture elements themselves. Specifically relevant for natural textures is the blob-likeness, which characterizes the presence of blob-like, or circular, patterns. Consequently, circle detection based on the circular Hough transform [177] is applicable in this context. The image in the spatial domain is transformed into the circular Hough domain following

$$c_3^2 = (x - c_1)^2 + (y - c_2)^2 \tag{4.9}$$

and leads to a three-dimensional parameter space. This circular version of the Hough transform has notable potential for texture classification, because it is by definition independent of the image rotation. The number of identified circular patterns is accumulated on ten different scaling levels determined stepwise by the original image size using the Matlab implementation $imresize()$ and $imfindcircles()$. The average number of all circles found on the ten scale levels then determines the blob-likeness feature I_{blob}.

Image Coarseness In its original definition from Tamura et al. [204] and as illustrated in Fig. 4.3, textural coarseness is defined by the difference of adjacent subwindow mean values of size $2^k \times 2^k$, with $k \in \{1, ..., n\}$ in the horizontal, vertical and diagonal directions. These mean values of a subwindow are denoted as average gray levels (AGLs). In the horizontal direction, for example, the difference of two AGLs is computed as $E_{k,h}(x, y) = |AGL_k(x + 2^{k-1}, y) - AGL_k(x - 2^{k-1}, y)|$. If only small-sized texture elements are present, both AGLs are similar for larger k, however, coarse texture elements in the original image lead to large differences even for larger subwindows. For each pixel, the k$^{\text{th}}$ value that corresponds to the maximum difference of AGLs in all directions leads to a coarseness map $\mathbf{I}_{\text{coarse}}$ holding the values 2^k. Image coarseness is then determined as the average of $\mathbf{I}_{\text{coarse}}$ as

$$I_{\text{coa}} = \frac{1}{M \cdot N} \sum_{m=1}^{M} \sum_{n=1}^{N} \mathbf{I}_{\text{coarse}}[m, n] \tag{4.10}$$

where M and N are the width and height of the picture, respectively. Two relevant adjustments are necessary in this context.

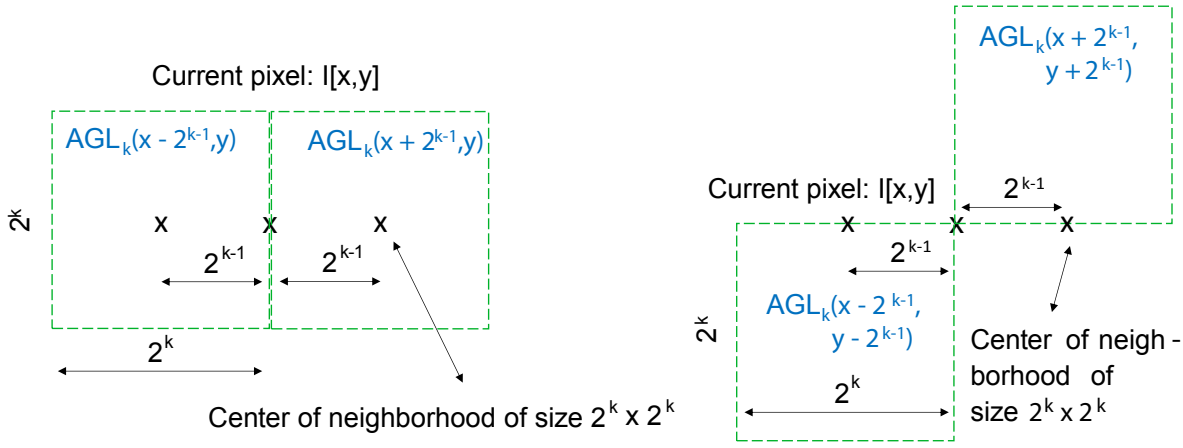


Figure 4.3: Graphical illustration of the adjacent windows with respect to the current pixel in the horizontal (left) and right-diagonal direction (right) for image coarseness calculation (adapted from [7] © 2017 IEEE).

First, the original definition is only partly robust to image rotations. Since computational speed is not a relevant objective required in offline material classification, the images are rotated in small degree steps (e.g., 3^o) to effectively find coarse texture segments in any direction.

Second, the window size $2^k \times 2^k$ requires a relation to the possible occurring size of real material coarse texture elements. The focal length f of the iPhone8 camera is $f = 26$ mm, and the recording distance typically ranges within $Z_C \approx 100$ mm – 140 mm. A coarse structure element can be represented as a square with a side length of X_C in camera coordinates; note that the world coordinate system is set to be identical with the camera coordinate system. Weber et al. [34] consider the maximal X_C of a particle of about 5 mm – 10 mm belonging to the tactile dimension of macroscopic roughness. As one possibility to determine the maximum side length X_C of a texture element in pixel coordinates $\mathbf{x}_p$, a well-defined visual pattern (e.g., a checkered paper with 5 mm distances between the squares) with a black-painted 10-by-10 mm^2 square serves as reference pattern. Ideally, the images are captured perfectly perpendicular to the surface within the aforementioned distance Z_C. Two images at the distances (100 mm and 140 mm) have been captured as shown in Fig. 4.4 (left and middle). A third image (right) shows the magnified checkered paper using the lens for $\mathbf{I}_{\text{macro}}$ at a fixed recording distance of 25 mm.

This approach reveals that the black square refers to either 300 pixels (100 mm recording distance) or 240 pixel (140 mm recording distance) for the display images $\mathbf{I}_{\text{disp}}$, and about 1700 pixels for $\mathbf{I}_{\text{macro}}$ without requiring to determine the internal camera parameters of the iPhone8. Also note that the potential lens distortion effects are negligible in the center of the images, and hence, no radial correction is required. Consequently the algorithm of calculating I_{Coa} can be adjusted to account for this highly relevant information. In this case, it is reasonable to resize the image to a smaller size, e.g., 20% of the original dimension, leading to a maximal AGL size of 48 - 60 pixels, which is set to the next power of 2, in this case, 64 pixels.

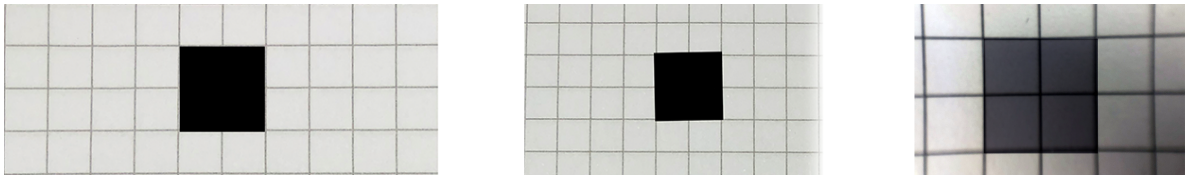


Figure 4.4: Well-defined squared patterns for the identification of maximum coarse texture element sizes. A structural element of 10 mm^2 is 240 – 300 pixel large given the usual recording distance of 100 mm – 140 mm and 4K resolution. Magnified images (right) are about 1700 pixel at a fix recording distance of 25 mm. These values are relevant to determine appropriate sizes of image feature parameters, notably, the coarseness feature. Also note that these parameters help in determining the relevant image size for deep learning-based image classification in Section 4.5. If coarse materials occupy 300 pixels in width and height, and if an overlap of 100 pixels is considered for data generation, an overall width/height of 500 pixels is chosen as image patch input size.

Color and Brightness Color and brightness are not textural features, however, become highly relevant for material classification to distinguish between different types of woods or metals. In fact, even the human tactile perception system is not able to distinguish such materials by touch alone, but by their dominant color and surface gloss during visual inspection. Humans tend to describe image color information by the specific components such as hue, saturation and brightness instead of RGB components [177]. Hence, the HSI (hue, saturation, intensity) color model is suitable, which decouples the intensity from the color-carrying information. Most specifically, the hue component H is related to the dominant color in an image with

$$H = \begin{cases} \theta & \text{if } B \leq G \\ 360 - \theta & \text{otherwise} \end{cases} \tag{4.11}$$

and

$$\theta = cos^{-1}\left(\frac{0.5 \cdot (R-G) + (R-B)}{\sqrt{(R-G)^2 + (R-B) \cdot (G-B)}}\right) \tag{4.12}$$

and describes the conversion of an RGB pixel to an angle of the color cone. The saturation is defined as

$$I_{\text{sat}} = 1 - \frac{3}{(R+G+B)} \cdot \min(R,G,B) \tag{4.13}$$

and the intensity, or brightness I_{br} is

$$I_{\text{br}} = \frac{(R+G+B)}{3} \tag{4.14}$$

If all RGB pixels are converted to HSI pixels $\mathbf{I}_{\text{hsv}}$ and all hue values are extracted as array $\mathbf{h}_{\text{hue}}$, the histogram $hist(\mathbf{h}_{\text{hue}})$ with a fixed bin size of 360 values (i.e., the color resolution if the color cone) can be computed; the index (bin number x) of the maximum value in $hist(\mathbf{h}_{\text{hue}})$ is the dominant color

$$I_{\text{col}} = \arg\max_x hist(\mathbf{h}_{\text{hue}}) \tag{4.15}$$

Note that special care is necessary to avoid feature value discontinuities for reddish surfaces (from 360^o to 0^o). As observed for real material samples, pink-colored surfaces are actually

most rarely encountered. Consequently, the discontinuity is shifted to 345^o which sets a clear threshold between blueish (less than 345^o) and reddish (345^o – 15^o) surfaces.

Multi-color images are also examined for their most relevant color. As special case, the absence of color (e.g., a whitish or completely black surface image) needs to be addressed as well. The image contrast and the brightness act as filters to identify this case. Whenever this condition is valid, the dominant color is equal to the intensity I and set as value $I_{\text{col}} = I - 255$. In this way, the feature ranges between [-255 and +345] covering all shades of gray and the hue values. The absence of any color and illumination, i.e., black surfaces, consequently leads to the lowest feature value.

4.1.1.3 Image Feature Rotation Dependency

All the materials in the LMT-184-HDB were recorded under varying conditions to ensure notable data spread, and hence, to show if the aforementioned features are robust enough for classification. As introduced in Section 3.3.4.1, the TU3 can be improved with a custom-made extension which allows for constant distance and ambient illumination conditions. As remaining dependency, the image rotation can distort the feature values and is examined in a smaller evaluation in the following. Features resulting from three heavily textured material surface images (hemp M_1, makrolon M_2, teak M_3) are evaluated under increasing rotation angles, which is shown as example in Fig. 4.5 for the material sample hemp. The

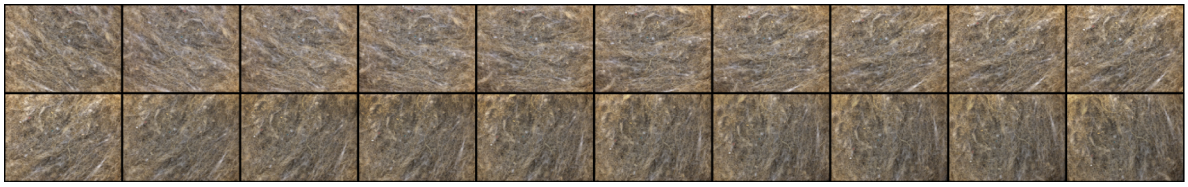


Figure 4.5: Montage of surface images (hemp) captured with the TU3 extension under subsequent rotations of approximately 8^o. Note that the illumination and distance conditions do not noticeably change within the images.

three samples have been selected according to their different visual appearance; hemp contains line-like fibers, makrolon (polycarbonate, PE/PP) contains irregular structure patterns, and teak contains man-made line elements (gratings). Twenty pictures with about 8^o stepwise increase in rotation angle have been captured per material. Each of the aforementioned image features has been calculated and normalized between 0 and 1, leading to 20 rotation-dependent image feature vectors $\mathbf{f}_{\text{rot}}$ for each material sample. Ideally, the 20 entries for each feature are similar for a specific material sample and have low variance to guarantee rotation independence. As two appropriate metrics, the interpolated slope $\text{m}(\mathbf{f}_{\text{rot}})$ and the standard deviation $\sigma(\mathbf{f}_{\text{rot}})$ are evaluated; ideally, both values are close to zero in order to prove rotation invariance. Table 4.1 shows all slopes and standard deviations for the three material samples. Figure 4.6 further shows three exemplary features that are robust as well as descriptive in terms of feature value difference among the three selected material samples.

Table 4.1: Image feature dependencies with respect to rotation of images captured with the TU3 extension. The first three entries represent the slope, and the last three the standard deviations of three different material feature vectors consisting of 20 feature values. The closer an entry is to zero, the more invariant the feature is to arbitrary rotations. Features like blob-likeness, edginess, the RLBPs, color, or brightness are least affected by rotations.

Feature	M_1 $m(\mathbf{f}_{rot})$	M_2 $m(\mathbf{f}_{rot})$	M_3 $m(\mathbf{f}_{rot})$	M_1 $\sigma(\mathbf{f}_{rot})$	M_2 $\sigma(\mathbf{f}_{rot})$	M_3 $\sigma(\mathbf{f}_{rot})$
ASM	-0.008	-0.009	-0.005	0.045	0.050	0.064
Contrast	0.001	0.005	0.020	0.017	0.037	0.125
Correlation	-0.001	0.001	-0.001	0.016	0.005	0.008
Variance	-0.000	0.003	-0.009	0.008	0.015	0.054
Homogeneity	-0.005	-0.001	-0.004	0.032	0.012	0.051
Sum Average	-0.001	0.002	-0.006	0.008	0.010	0.036
Sum Variance	-0.001	0.003	-0.010	0.010	0.018	0.063
Sum Entropy	0.001	0.000	0.000	0.005	0.003	0.002
Entropy	0.002	0.001	0.001	0.014	0.005	0.011
Diff. Variance	0.005	0.004	0.018	0.048	0.033	0.111
Diff. Entropy	0.004	0.001	0.004	0.026	0.006	0.026
Inf. Measure Corr. I	0.001	-0.000	0.002	0.010	0.004	0.016
Inf. Measure Corr. II	-0.000	0.000	-0.000	0.010	0.003	0.003
Max. Corr. Coeff.	-0.000	0.001	-0.001	0.016	0.005	0.008
Tamura Coarseness	-0.015	-0.001	-0.007	0.083	0.007	0.061
Tamura Contrast	0.006	0.003	0.002	0.036	0.015	0.027
Tamura Linelikeness	-0.006	0.003	-0.006	0.032	0.019	0.036
Tamura Directionality	0.005	0.001	-0.001	0.039	0.011	0.060
FFT2 75	0.033	0.010	0.000	0.191	0.066	0.000
FFT2 80	0.017	0.002	0.010	0.096	0.046	0.086
FFT2 85	0.003	0.013	-0.001	0.036	0.091	0.073
1^{st} RLBP	0.001	-0.000	0.000	0.012	0.010	0.012
2^{nd} RLBP	0.001	-0.000	-0.000	0.019	0.009	0.009
3^{rd} RLBP	0.000	0.000	0.001	0.002	0.004	0.012
4^{th} RLBP	-0.001	0.000	0.001	0.013	0.007	0.015
5^{th} RLBP	0.000	0.000	-0.000	0.012	0.018	0.028
6^{th} RLBP	-0.000	0.000	-0.000	0.012	0.009	0.009
7^{th} RLBP	-0.000	-0.000	-0.000	0.003	0.001	0.008
8^{th} RLBP	0.000	-0.000	0.000	0.014	0.008	0.007
9^{th} RLBP	-0.000	-0.000	0.000	0.009	0.009	0.008
10^{th} RLBP	0.000	-0.000	-0.000	0.018	0.017	0.014
Color	0.010	-0.000	0.001	0.086	0.000	0.006
Brightness	-0.001	0.001	-0.005	0.006	0.008	0.028
Saturation	0.007	0.001	0.002	0.037	0.012	0.014
Edginess	0.006	0.001	0.004	0.031	0.006	0.023
Glossiness	-0.003	0.001	-0.004	0.020	0.006	0.022
Blob-Likeness	0.000	0.002	0.003	0.001	0.010	0.017

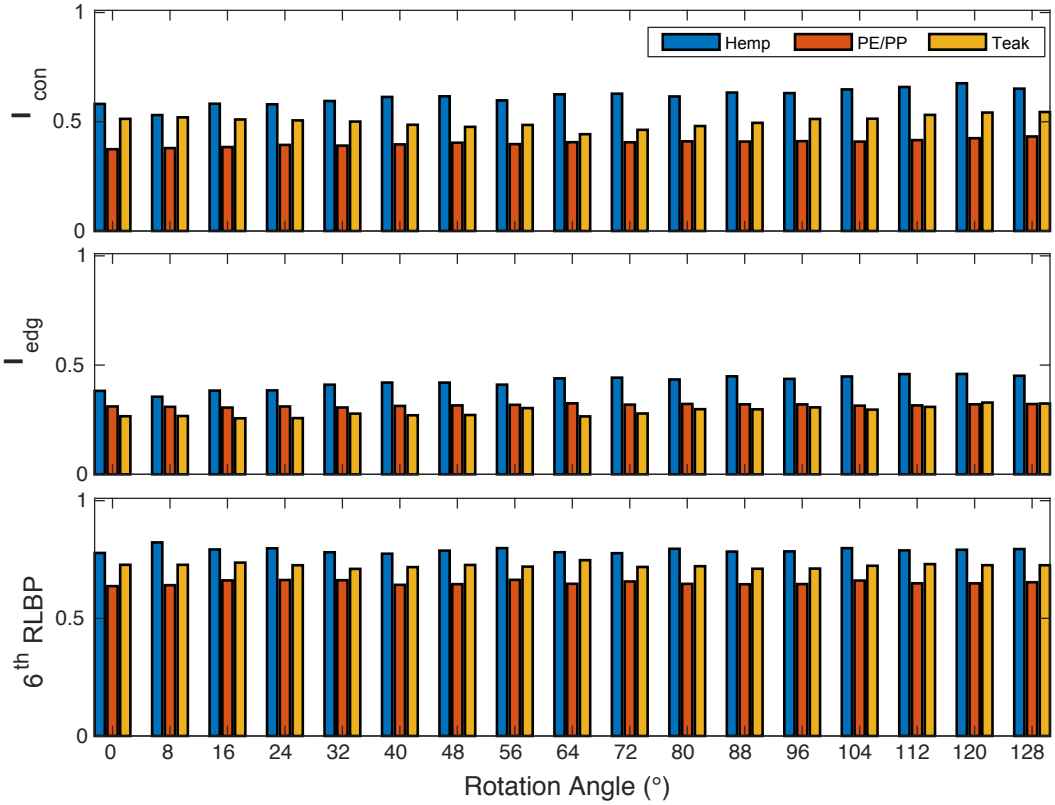


Figure 4.6: Subset of image features (image contrast, edginess, and 6^{th} RLBP) calculated for the three material images under stepwise rotation. The image features have been min-max-normalized according to the range of values of all material samples from the LMT 184 haptic database. As expected, these features lead to very robust and sufficiently distinctive feature values and can be considered as robust to image rotation thanks to the TU3 extension.

4.1.2 Audio Features

The field of audio signal processing offers a variety of features [211] that can be used in the analysis of vibrotactile signals as performed in the thesis-related work in [12]. Audio features can either be applied on vibrotactile signals **v**, or on the corresponding microphone-recorded audio signals **a** during tapping on and sliding over a material surface. The majority of the potential audio features which have been identified as being descriptive for material classification follow the implementations from Kim et al. [211] and Giannakopoulos et al. [212]; 35 well-known features like the spectral centroid, zero crossing rate, and the Mel frequency cepstral coefficients are applied on the audio and vibrotactile signals (both finger and steel tool tip) recordings. Note that the following feature definitions are preceded by (V) for being calculated based on the vibrotactile signals, but are likewise valid for audio (A) data traces.

4.1.2.1 Time Domain Features

A smaller set of audio features suited for material classification stems from time domain definitions.

Signal Power The vibrotactile signal power (VSP) is defined as

$$VSP = \frac{1}{N}\sum_{n=1}^{N} |\mathbf{v}[n]|^2 \tag{4.16}$$

and is heavily depending on the signal amplitude values. Nevertheless, VSP gives an indication of the silence or loudness of a material surface scan, and hence, is related to macro- and microscopic roughness. Note that the Texplorer2 mitigates the force dependency, and hence, such basic features still can lead to high discrimination among different material classes.

Zero-Crossing Rate The vibrotactile zero-crossing rate (VZCR) counts the numbers of sign changes between subsequent signal values and is defined as

$$VZCR = \frac{1}{2 \cdot N}\sum_{n=2}^{N} |sgn(\mathbf{v}[n]) - sgn(\mathbf{v}[n-1])| \tag{4.17}$$

VZCR is considered as approximation for the fundamental frequency, and is hence suitable for a clear separation of coarse and fine materials. The finer the surface is, the more high-frequent zero-crossings are expected, and vice versa, very coarse patterns presumably lead to lower ZCRs.

Entropy of Energy The vibrotactile entropy is a metric that measures abrupt changes in signals, and hence, is related to irregularities of coarse structures on a material surface. It is calculated as

$$VE = -\sum_{k=1}^{K} \mathbf{e}[k] \cdot log_2(\mathbf{e}[k]) \tag{4.18}$$

with

$$\mathbf{e}[k] = \frac{\mathbf{E}_{\text{sub}}[k]}{E_{\text{signal}}} \tag{4.19}$$

The subframe energy is

$$\mathbf{E}_{\text{sub}}[n] = \sum_{n=1}^{W} |\mathbf{v}[n]|^2 \tag{4.20}$$

and the total signal energy

$$E_{\text{signal}} = \sum_{k=1}^{K} \mathbf{E}_{\text{sub}}[k] \tag{4.21}$$

based on the length of window W and K frames. Their values are determined by the sample rate and the expected duration of the abrupt changes. For example, a tap on a surface can last up to 100 ms. Likewise, sliding motions over coarse materials lead to comparable spike patterns, only laterally, and hence, 100 ms indicates the window width for entropy.

Periodicity and Harmonic Ratio Surface scans lead to quasi-periodic audio and vibrotactile signals. Consequently, the fundamental period and its intensity, also known as harmonic ratio, can be characteristic for the surface structure. The auto-correlation function $\mathbf{\Psi}_{\text{vv}}[m]$ of the vibrotactile signal is calculated using

$$\mathbf{\Psi}_{\text{vv}}[m] = \sum_{n=1}^{N} \mathbf{v}[n] \cdot \mathbf{v}[n-m] \tag{4.22}$$

which based on different values for the lag m, and $\mathbf{\Psi}_{\text{vv}}$, and is min-max-normalized between 0 and 1. Note that the auto-correlation is largest for m = 0. A minimal shift of $m_{\text{min}} = 20$ points is applied to exclude self-repetition.

Finally, the position of the first largest peak (i.e., lag $m = F_0$) in the auto-correlation function and its intensity $\Psi_{\text{vv}}[m = F_0]$ are the fundamental period and harmonic ratio, respectively, which are identified using the peak detection algorithm findpeaks() from the Matlab signal processing toolbox.

4.1.2.2 Frequency Domain Features

One important difference of material classification to audio processing, notably to speech recognition, is the potential pseudo-stationarity of material surface recordings. Consequently, time-dependent shift operations of subframes can be avoided in general, and hence, the resulting trade-off between temporal and spectral resolution shifts to a higher frequency resolution based on the calculation over the entire audio or vibrotactile segment. The same procedure is also known as long-term averaged feature extraction in audio processing [213], which fully discards the temporal evolution of a signal, but has shown notable performance in music genre classification [214]. That is why a complete audio or vibrotactile frame can be analyzed at full frequency resolution, making the use of spectral features attractive and simultaneously averages the temporal effects of operator speed variations within one recording of audio or vibrotactile data traces. Figure 4.7 shows that the spectral properties like

envelope shape, spread, and distribution) are characteristic and motivate the usage of such frequency domain features even if speed and force variations between different scans are present.

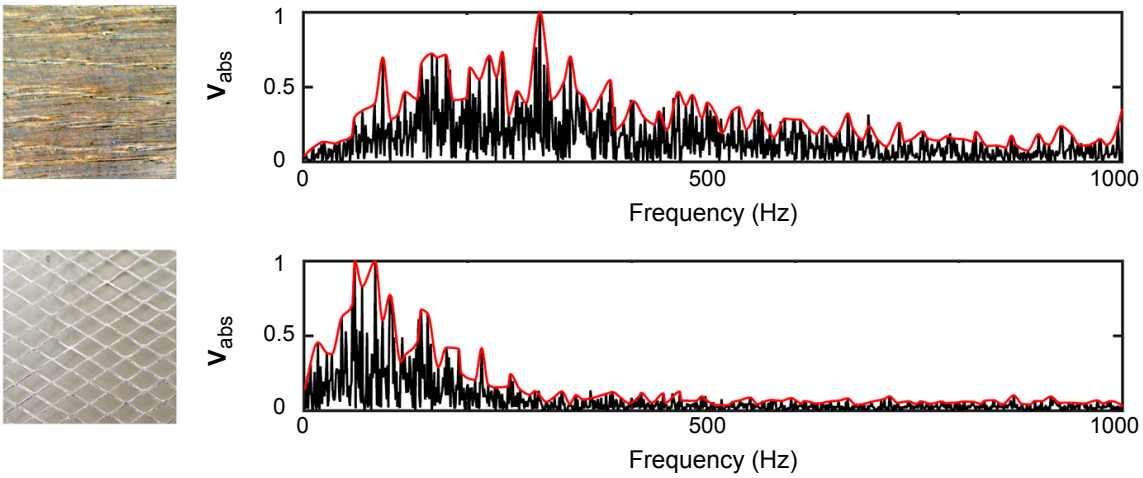


Figure 4.7: Spectral plots for two different surfaces during the acceleration movement phase. The acceleration spectra of the surfaces reveal characteristic shapes, especially with regard to their spectral envelope (adapted from [7] © 2017 IEEE).

In the following, relevant features are listed first, and thereafter explained in the context of material classification.

Spectral Centroid and Spectral Spread The vibrotactile signal spectra shown in Fig. 4.7 show that their centroids and spreads around the centroids [214] are characteristic. Given a vibrotactile signal, the vibrotactile spectral centroid (VSC) is defined as

$$VSC = \frac{\sum_{i=1}^{N} \mathbf{f}[i] \cdot |\mathbf{V}[\mathbf{f}[i]]|}{\sum_{i=1}^{N} |\mathbf{V}[\mathbf{f}[i]]|} \tag{4.23}$$

Likewise, the vibrotactile spectral spread (VSS) around the VSC captures the distribution around the centroid and is calculated as

$$VSS = \frac{\sum_{i=1}^{N} (\mathbf{f}[i] - VSC)^2 \cdot |\mathbf{V}[\mathbf{f}[i]]|}{\sum_{i=1}^{N} |\mathbf{V}[\mathbf{f}[i]]|} \tag{4.24}$$

Spectral Roll-off Another spectral shape descriptor is the vibrotactile spectral roll-off (VSR) [214], which describes the frequency until the majority of the spectral content is concentrated. A 90% spectral roll-off threshold of an absolute vibrotactile signal spectrum $|\mathbf{V}[\mathbf{f}]|$ leads to the condition

$$\sum_{i=1}^{SRoF} |\mathbf{V}(\mathbf{f}[i])| = 0.90 \cdot \sum_{i=1}^{N} |\mathbf{V}[\mathbf{f}[i]]| \tag{4.25}$$

which is fulfilled by the roll-off frequency SRoF. Note that this feature is specifically robust to frequency shifts within the 90% of lower frequencies.

Chroma Vector In its original definition from [215], the energy of spectral bins is denoted as chroma vector if the bin size is chosen to follow the 12 equal-tempered pitch classes of Western-type music. Audio and vibrotactile scans from tool-surface interaction, however, resemble unvoiced sounds, and hence, a nonlinear bin sizing is not required for material analysis. Consequently, the energy of 12 equal-sized bins, adjusted to either the audio or vibrotactile sample frequency are calculated with one entry being

$$\mathrm{c}[k, \mathbf{f}_l, \mathbf{f}_u] = \sum_{f=\mathbf{f}_l[k]}^{\mathbf{f}_u[k]} \frac{\mathbf{V}[k]}{\mathbf{f}_l[k] - \mathbf{f}_u[k]} \tag{4.26}$$

where $\mathbf{f}_{l,u}$ denote the lower and upper limits of each equally-distant frequency band.

Mel Frequency Cepstral Coefficients Mel-Frequency cepstral coefficients (MFCCs) are known to capture the spectral envelope of a given data frame by calculating the discrete cosine transform of the audio signal spectra [216]. This thesis adapts and modifies the MFCC implementation from [217], [218] to vibrotactile signals, denoted as vibrotactile cepstral coefficients (VTCCs), following the similar argumentation as for the Chroma feature, i.e., audio parameters are removed whenever they are not required for vibrotactile data. Likewise, material-related audio data leads to audio-tactile cepstral coefficients (ATCCs). Figure 4.8

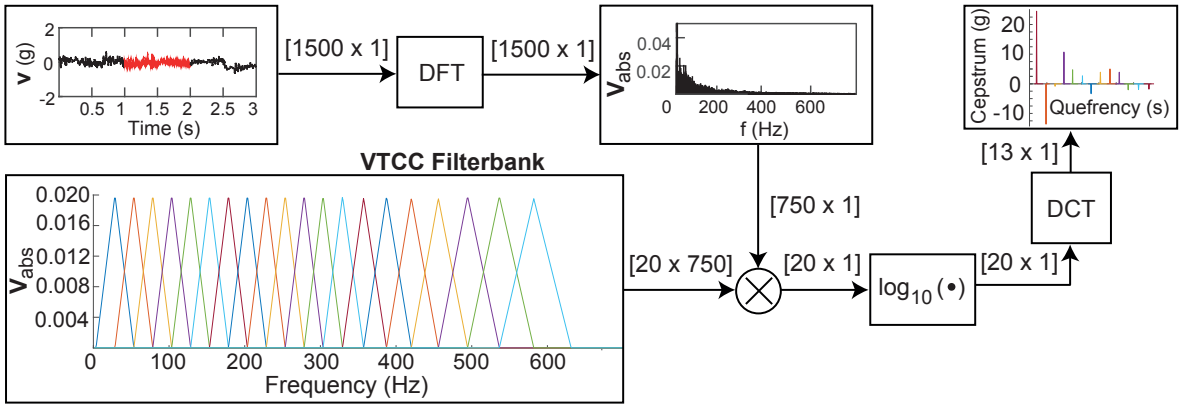


Figure 4.8: VTCC feature vector calculation of a 1-s-long surface signal. The filterbank uses 20 filters for the cepstral analysis. Their shape, spacing, and amplitude result from feature optimization. Note that the coarse triangle binning structure reduces slight speed-depending variations in the spectrum of the signal, e.g., shifts (adapted from [7] © 2017 IEEE).

shows the processing pipeline for a 1-s-long vibrotactile signal segment, which is extracted from the sliding phase with TU1. The applied filterbank has to be adjusted to the acceleration sensor bandwidth. In addition, the audio specific processing steps such as the preintensification of the audio signal and the liftering procedure (cepstral filtering) [218] are removed. The algorithm calculates the 1D DFT and the corresponding filterbank entries, then applies the logarithm to decrease the dynamic range, and finally calculates the cepstrum using the 1D DCT [216]–[218]. The first entry denotes the constant component of the spectrum and is hence removed. The remaining lower components describe the overall shape of the spectrum and the higher components its smoothness. Consequently, subsequent feature

selection is relevant to remove components with less descriptive power. Several components of the algorithm are also subject to further optimization. For example, the filterbank amplitudes and shapes (constant, linearly -, exponentially decaying) were examined. No improvement has been observed in terms of applying different intensities to the filters, i.e., all filterbank entries have constant height, however, lower frequencies are captured with more narrow triangle-shaped filterbank entries. Note that a rectangular filter shape with overlappings overemphasizes frequencies; instead a triangular shape with overlappings ensures equal weighting among frequencies.

Note that the aforementioned spectral features fully describe the shape of an audio or vibrotactile spectrum and thereby determine whether the signal is voiced or unvoiced, i.e., reveals characteristic peaks or resembles an equal distribution with the SC centered in the spectrum. The latter is mainly encountered for microscopically rough material samples like sandpaper or the sparkling sound of a plastic foil. It can further be observed that audio-damping materials have the largest part of their spectral distribution allocated in low-frequency energy bands. By contrast, glasses, ceramics, and stones can have characteristic high-frequency components, which is also quantified by the chroma vector and spectral roll-off. The overall shape, i.e., in terms of appearance is described by the VTCCs, and hence, specifically characteristic for individual materials.

After the definition of the audio features relevant for this thesis, the next section examines their speed dependency in a subset-based pilot study.

4.1.2.3 Audio Feature Speed Dependency Analysis

Thanks to the design of TU1, the audio features are mainly affected by the possible variations in operator speed. The three materials (M_1 PLA 2 mm gratings, M_2 PLA 8 mm gratings, M_3 unpolished granite) shown in Fig. 4.9 (upper part) are selected to examine the expected variations in their respective recorded audio data traces $\mathbf{a}$ (lower part). The selected material samples deliberately reflect extreme examples examined under a broad range of scan speeds; the majority of materials contains less distinct surface structures (e.g., compared to the 8 mm PLA gratings) and the operator speed can be assumed to be in a narrower scan speed range. Note that the same material scans also include tactile-related vibrotactile signals which are evaluated in the tactile features section 4.1.5.6 and that the audio features introduced beforehand are interchangeably usable for both vibrotactile and audio signals. Comparable to the image rotation analysis in Section 4.1.1.3, each feature value has been calculated on five audio data traces, leading to a vector $\mathbf{f}_{\text{speed, audio}}$. Example values for temporal and frequency-based audio features are shown in Fig. 4.10. Likewise, the interpolated slope and standard deviation from appropriate metrics for quantification and consequently are calculated and shown in Table 4.2.

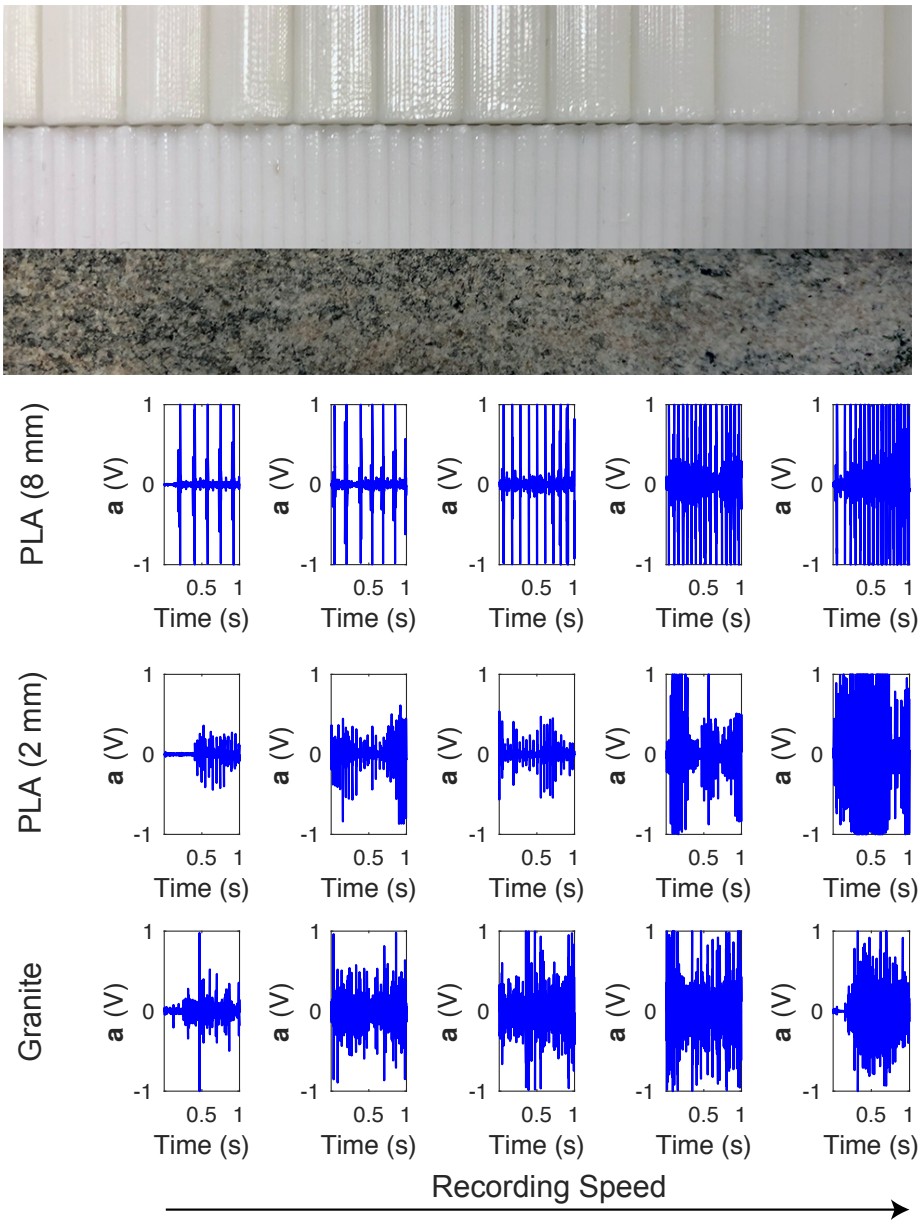


Figure 4.9: Three material samples (top: PLA with 8 mm grating distance, middle: PLA with 2 mm grating distance, bottom: unpolished granite tile) used for speed evaluation for both audio and vibrotactile signals. Strong feature value variations are expected due to the characteristically coarse or rough surface structure of each sample. The scan speed increased from about 20 mm/s – 160 mm/s.

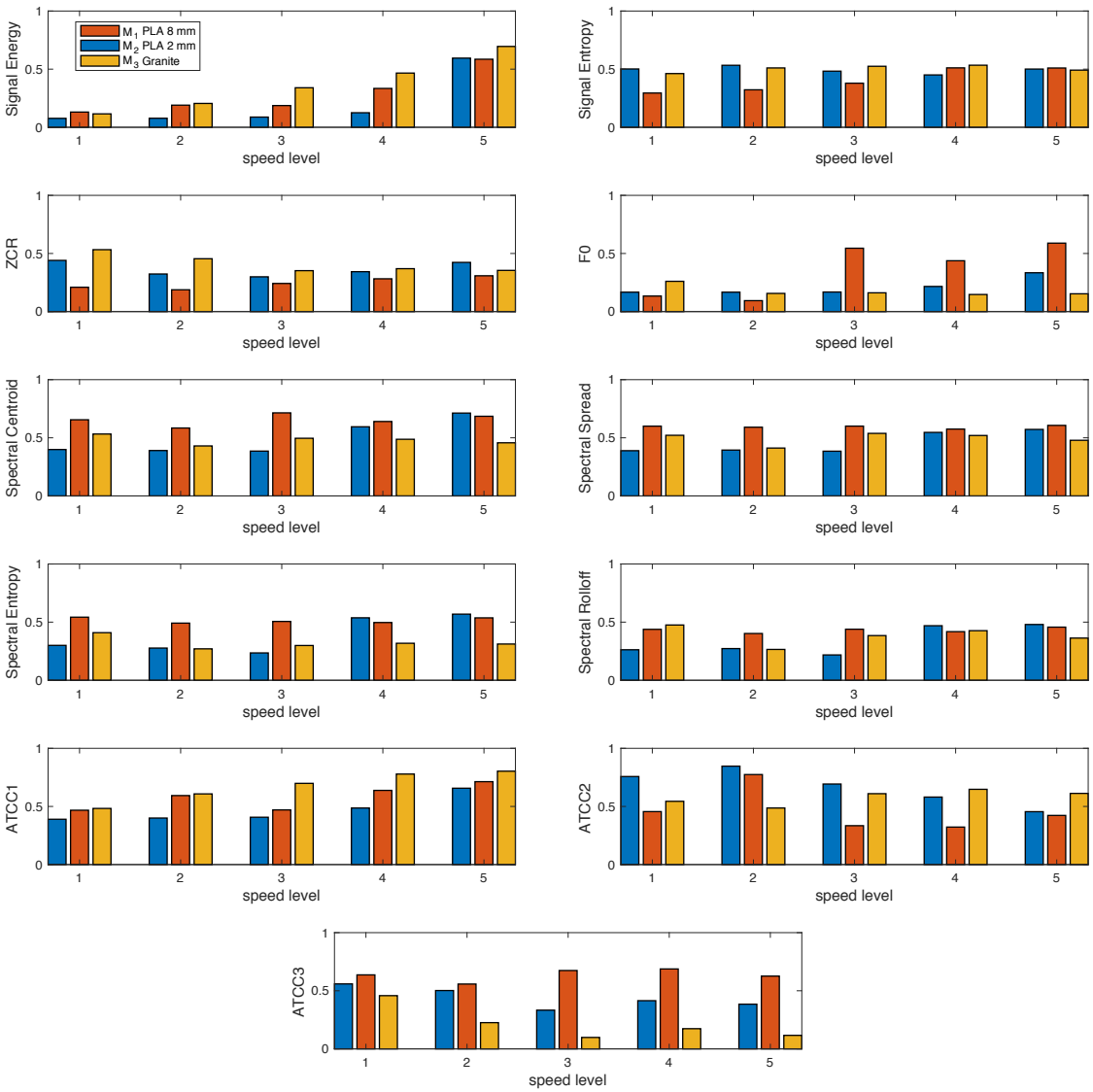


Figure 4.10: Comparison of different audio features for the five increasing speed audio recordings **a** (x-axis) and three materials. Intuitively, the audio signal energy (first row, left) heavily depends on the scan speed, whereas other features like the ZCR (second row, left) remain relatively constant and specific for the three materials.

Table 4.2: Speed dependency result table for audio features under increasing recording speed conditions. The material samples are M_1 PLA 8mm, M_2 PLA 2mm, and M_3 Granite. The first three entries denote the slope of the feature vector consisting of five entries (five different speeds), and the last three the standard deviations of the material feature vectors. Ideally, all values are close to zero to indicate independence of different recording speeds. Note that the audio signal energy (ASE) heavily depends on the recording speed, but other features like the zero-crossing rate (ZCR) are more robust to different speed levels. Nevertheless, these plots further show that any scanning system should avoid unnecessarily high scan speeds.

Feature	M_1 m($\mathbf{f}_{\text{speed}}$)	M_2 m($\mathbf{f}_{\text{speed}}$)	M_3 m($\mathbf{f}_{\text{speed}}$)	M_1 σ($\mathbf{f}_{\text{speed}}$)	M_2 σ($\mathbf{f}_{\text{speed}}$)	M_3 σ($\mathbf{f}_{\text{speed}}$)
ZCR	-0.001	0.024	-0.025	0.035	0.041	0.045
ASE	0.108	0.105	0.142	0.226	0.184	0.228
ASEn	-0.008	0.062	0.008	0.030	0.102	0.029
Sp.Centroid	0.083	0.012	-0.009	0.150	0.049	0.039
Sp. Spread	0.052	-0.000	0.002	0.094	0.012	0.051
Sp. Entropy	0.080	-0.001	-0.015	0.157	0.024	0.052
Sp.Rolloff	0.063	0.005	-0.006	0.124	0.021	0.079
ATCC1	0.062	0.054	0.081	0.112	0.107	0.131
ATCC2	-0.087	-0.052	0.030	0.153	0.184	0.064
ATCC3	-0.044	0.011	-0.073	0.091	0.051	0.145
F0	0.038	0.125	-0.022	0.072	0.232	0.048

Variations in scan speed have an observable influence on the audio features, however, features like the ZCR or the ATCCs stay relatively robust in their feature value. Note that these scans were conducted to check if audio features are overly influenced by changes in scan speed, which is not observable. Consequently, audio features are very likely to be a reasonable choice for material classification approaches.

4.1.3 Tactile Features from Related Work

Since the sensorized tool scanning a surface for audio and vibrotactile data traces is a similar source of information, the same considerations from the audio signal section are valid for vibrotactile-based feature design and has been partly examined in the past. For example, de Boissieu et al. [44] used vibrotactile signals to calculate signal statistics features such as the variance and kurtosis as well as spectral features like the spectrum mean, slope, and the spectral form. Likewise, Dallaire et al. [89] use the vibrotactile signal variance, skewness, kurtosis, fifth moment, and sum of variations over time. Jamali et al. [115] also use PVDF sensors and strain gauge scans under controlled conditions to calculate the largest peaks in the spectral domain, denoted as Fourier peaks (FPs), and the mean pressure, respectively. They further show that the largest five FPs lead to a classification performance of 95% for well-controlled scans. Comparably, Fishel et al. [116] calculated three features based on common signal statistics like the signal power (roughness), the spectral centroid (fineness), and the coefficient of friction (traction) which are sufficient to classify 117 material samples with 95.4% classification accuracy. They further introduce the relevance of selecting the correct exploratory procedure (EP) for specific materials; it suffices for some materials to simply tap

on and statically press on the surface for successful material classification. Another objective of Fishel et al. [116] was feature orthogonality resulting from their different sensing origin, which is also conceptually followed in this thesis by employing multiple sensors. Table 4.3 summarizes these vibrotactile signal statistic features which have frequently been applied by related work, and hence, are also added to the initial feature space for the sake of comparison in this thesis. It is worth highlighting that all these features perform well under controlled

Table 4.3: Commonly used vibrotactile features from related work which have also been examined in this thesis.

VSP	$= \frac{1}{2 \cdot N} \cdot \sum_{i=1}^{N} \lvert\mathbf{v}[i]\rvert^2$	Vibrotactile Signal Power $\overline{\mathbf{v}}$ [44], [89]
VSSD	$= \frac{1}{N} \cdot \sum_{i=1}^{N} (\mathbf{v}[i] - \overline{\mathbf{v}})^2$	V. Std. $\sigma(\mathbf{v})$ [44] [89], [219]
VSK	$= (\frac{1}{N} \cdot \sum_{i=1}^{N} (\frac{\mathbf{v}[i] - \overline{v}}{\sigma(\mathbf{v})})^4) - 3$	V. Kurtosis $\kappa(\mathbf{v})$ [44], [89]
SC	$= \frac{\sum_{i=1}^{N} \mathbf{f}[i] \cdot \mathbf{P}^2_{\text{AC, abs}}[\mathbf{f}[i]]}{\sum_{i=1}^{N} \mathbf{P}^2_{\text{AC, abs}}[\mathbf{f}[i]]}$	Fineness [116]
R	$= log_{10}\big(\frac{1}{N} \sum_{n=1}^{N} (\mathbf{p}_{\text{A,C,filt}}[n]^2)\big)$	Roughness [116]
FP_5		Five Dominant Fourier Peaks [115]

scan conditions as reported in the original papers. This thesis further investigates whether they are descriptive when applied on unconstrained operator-based scans.

4.1.4 Tactile Features based on Five-dimensional Feature Space

The majority of the aforementioned related work approaches focused on the evaluation of vibrotactile signals, which are known to be depending on scan-time parameters such as speed. Other signal domains have been sparsely addressed, but presumably further aid in the classification of materials. It is also worth noting that it is not the major objective to completely rule out the scan dependency for the various data traces, but to determine multimodal features that allow for a robust material classification nevertheless. Based on the five relevant tactile dimensions summarized in [63] a set of tactile features, denoted as TacTUM-5, was designed and evaluated in [7]. Burka et al. [220] cross-validated these features on the U-Penn database [191] and verified that they are more robust to scan variations then the signal statistic features introduced in Section 4.1.3. Note that several features like microscopic or macroscopic roughness acceleration strength are comparably used for the more extensive SynTouch-based feature space, and hence, are explained in the next section.

4.1.5 TacTUM Tactile Features based on SynTouch Feature Space

The aforementioned TacTUM-5 features have shown considerable classification accuracy among the 108 LMT Haptic Database. As introduced in Section 2.1.2, Fishel et al. [66], [67] proposed a set of even more representative tactile features which can be understood as a detailed description of the five major tactile dimensions from Okamoto et al. [63]. During the creation of this thesis, there was no explicit mathematical definition of these features publicly available, and hence, own interpretations were made as a major contribution of this thesis.

In order too avoid the potential confusion with the original feature names defined by Fishel et al. [66], this thesis adapts their naming convention with specific adjustments to credit the original qualitative description of these tactile features. The first letter is capitalized (which indicates the five major tactile dimension from [63]: Friction F, Microscopic Roughness Mi, Macroscopic Roughness Ma, Compliance C, Thermal Conductivity T), and TUM is added as superscript. For example, the feature addressing the SynTouch-inspired subdimension of damping cDP is denoted as CDP^{TUM}. The equations for the calculation of these features is presented in the following.

4.1.5.1 TacTUM Macroscopic Roughness

The macrotexture (MaTX^{TUM}), macrotexture coarseness (MaCO^{TUM}), and macrotexture regularity (MaRG^{TUM}) define the macroscopic roughness intensity, sizing, and regularity, of a material surface, respectively.

Macrotexture MaTX^{TUM} The first feature, MaTX^{TUM}, describes the intensity of coarse structures on a material surface. Several sensors such as the accelerometer and IR sensor in TU1, or, the friction force FSRs are applicable to capture the physical existence of such surface structural patterns and are examined in the following.

The first component is related to spikes in the vibrotactile signals. Fishel et al. [66] propose to use the lower frequency components within 5 Hz – 100 Hz of vibration sensor measurements for macroscopic roughness definition. However, very coarse material surfaces may let the sensor probe laterally impact during a scan, which possibly leads to high-frequency taps (comparable to tappings on a surface), and hence, to vibrotactile signal spikes. A spike detection algorithm on the absolute values of the vibrotactile sliding signals $\mathbf{v}_{\text{abs}}$ is defined which further uses a 100–point window simple moving average array $\mathbf{v}_{\text{abs}}^{(100)}$ of $\mathbf{v}_{\text{abs}}$. The moving average window length is an optimization parameter. The standard deviation $\sigma(\mathbf{v}_{\text{abs}})$ and the mean of $\mathbf{v}_{\text{abs}}$ are included to obtain an adjusted array which mitigates amplitude variations for different recording speeds of one vibrotactile recording by using the expression

$$\tilde{\mathbf{v}} = (\sigma(\mathbf{v}_{\text{abs}}) + \overline{\mathbf{v}}_{\text{abs}}) \cdot \mathbf{1} + \mathbf{v}_{\text{abs}}^{(100)} \tag{4.27}$$

Spikes are identified using

$$\mathbf{v}_{\text{abs},\Delta} = \mathbf{v}_{\text{abs}} \cdot u(\mathbf{v}_{\text{abs}} - \tilde{\mathbf{v}}) \tag{4.28}$$

based on the unit step function $u(\mathbf{x})$ that returns 1 element-wise if the corresponding array entry is greater than 0, and 0 otherwise. The vibrotactile spikiness (VS) component then is

$$VS = \log_{10}(1 + \overline{\mathbf{v}}_{\text{abs},\Delta}). \tag{4.29}$$

Friction signal variations (FV) during a slide are another component for MaTX^{TUM}. Rapid changes in the friction force signal $\mathbf{f}_{\text{fr,t}}$ with respect to its moving average $\mathbf{f}_{\text{fr,t}}^{(100)}$ indicate a change in the material's underlying physical structure, e.g., while sliding over a coarse or bumpy surface profile.

$$FV = \sigma(\mathbf{f}_{\text{fr,t}} - \mathbf{f}_{\text{fr,t}}^{(100)}) \tag{4.30}$$

As a third component, the IR sensor reflectance data is related to the height profile of a material surface. During a scan, variations in the IR signals capture the change of the current surface-sensor distance. The moving average $\mathbf{r}^{(100)}$ is calculated and the standard deviation of the difference $\mathbf{r} - \mathbf{r}^{(100)}$ denotes the reflectance variation (RV) value

$$RV = \sigma(\mathbf{r} - \mathbf{r}^{(100)}) \tag{4.31}$$

Finally, MaTX^{TUM} is calculated as a linear combination

$$\textcircled{1} \quad MaTX^{\text{TUM}} = \lambda_1 \cdot VS + \lambda_2 \cdot FV + \lambda_3 \cdot RV \tag{4.32}$$

using the weighting parameters $\lambda_{1,2,3}$ which are determined during feature optimization (Section 4.2.2) as (3/9, 2/9, 4/9). Figure 4.11 shows the three signal components for compressed coconut fibers which have a naturally coarse surface profile.

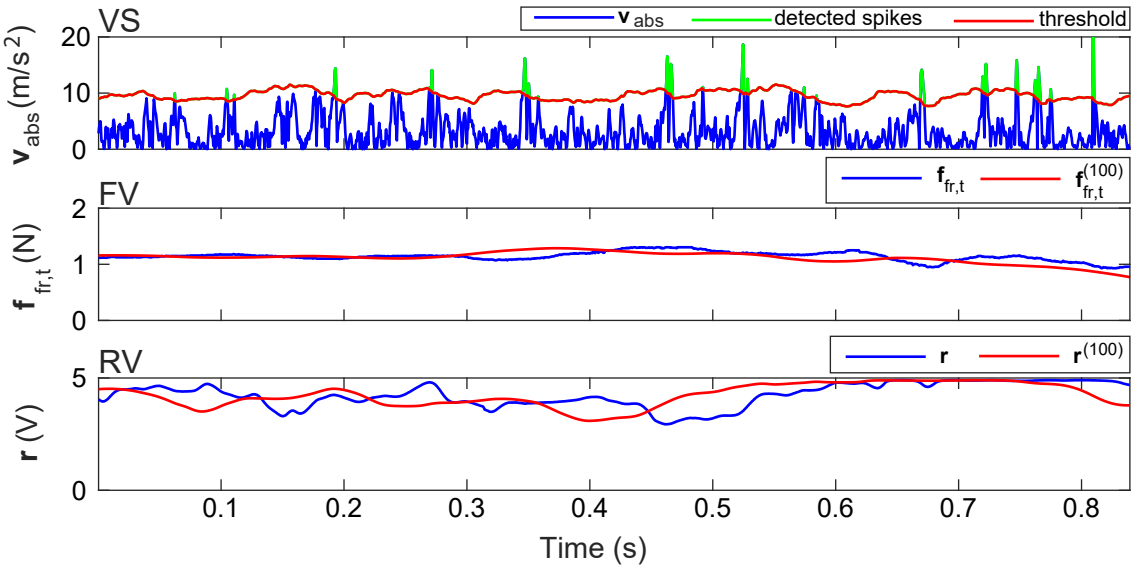


Figure 4.11: Macrotexture (MaTX^{TUM}) feature components for compressed coconut fibers material sample. A spike detection algorithm based on vibrotactile signal statistics filters significant spikes which are most likely occurring due to coarse surface obstacles (top). Likewise, these surface obstacles can lead to tangentially recorded friction changes (middle) around the smoothed friction signal (red line). Independently, the IR sensor measures the distances while sliding over the material sample (below) and is likewise compared to its moving average (red line).

Macro Coarseness MaCO$^{\text{TUM}}$ The size of visual patterns determines the visual coarseness of a material surface. This thesis combines a visual and tangible feature component to define MaCO$^{\text{TUM}}$ instead of using only vibration data as proposed by Fishel et al. [66].

The image coarseness feature I_{Coa} of Tamura et al. [204] from Section 4.1.1.2 is applicable to quantify the visual coarseness. Considering progresses in image processing as well as the fact that the position of coarse structure elements in the image is irrelevant, the bag of visual word pipeline [205] also provides invariant image descriptors to quantify macro coarseness. Feature detectors are considered to be more robust to slight image transformations which can occur during the surface image capturing step, e.g., potential changes of perspective or rotations. The speeded-up robust feature (SURF) descriptor [221] first detects blob-like texture elements as shown in Fig. 4.12, and then extracts and sorts the scale property s in a descending order array $\mathbf{s}$ using the Matlab function detectSURFFeatures() on the detected corner pixels. Natural occurring structure elements are also more likely to be blob-like than squarish compared to the coarseness feature definition in Section 4.1.1.2.

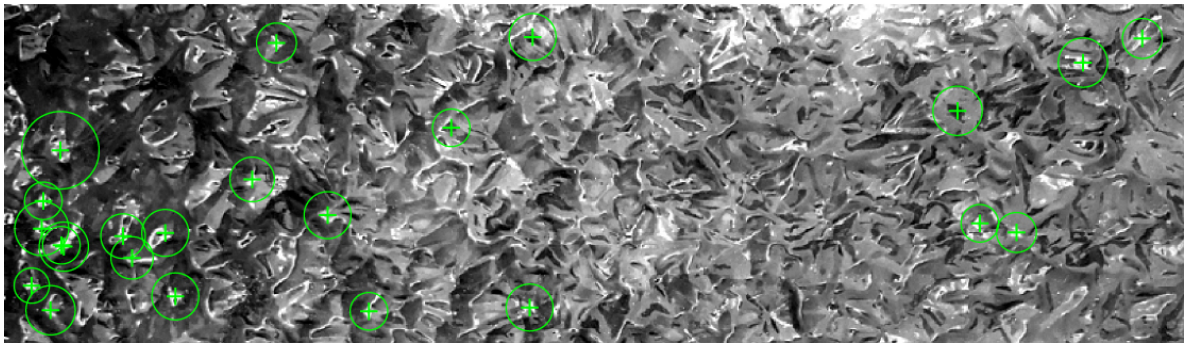


Figure 4.12: MaCO$^{\text{TUM}}$ visual feature component, shown for makrolon material sample. The average of the n largest green circles (scale property) is related to the visual coarseness of the surface image.

The visual component of macrotexture coarseness is considered as the mean value of the n largest scale factors $\mathbf{s}_{\text{n}}$ identified in $\mathbf{I}_{\text{disp}}$ multiplied by the contrast feature value I_{con} from Tamura et al. [204] from Section 4.1.1.2. $I_{\text{con}}(\mathbf{I}_{\text{disp}})$ is the image contrast of the material images; low-contrast images are more likely to be completely flat material surfaces, e.g., of plastics or glasses.

These purely visual-motivated components are then multiplied the mean value of the absolute vibrotactile signal. This step increases the likelihood that MaCO$^{\text{TUM}}$ becomes more variant to operator speed changes, but it transforms this image feature into a tactile feature; otherwise, painted structures on flat materials potentially distort the feature value. Hence,

$$\textcircled{2} \quad MaCO^{\text{TUM}} = \overline{\mathbf{v}}_{\text{abs}} \cdot I_{\text{con}}(\mathbf{I}_{\text{disp}}) \cdot \overline{\mathbf{s}}_{\text{n}} \tag{4.33}$$

with n being 10 in this thesis resulting from a feature optimization grid search (Section 4.2.2).

Macrotexture Regularity MaRG$^{\text{TUM}}$ The spatial positions of texture elements for MaCO$^{\text{TUM}}$ calculation is irrelevant, but by contrast, visual regularity strongly depends on such spatial information and is related to the intensity and arrangement of the image spectral compo-

nents of the 2D DFT [177]. Regular structures reveal specific frequency peaks, whereas irregular image spectra show an incoherent noisy pattern. The absolute part of the DFT2 for each uth and vth entry

$$\mathbf{X}[u,v] = \sum_{m=1}^{M}\sum_{n=1}^{N} \mathbf{I}[m,n] \cdot e^{-j\cdot 2\cdot \pi \cdot \left(\frac{u\cdot m}{M}\cdot \frac{v\cdot n}{N}\right)} \tag{4.34}$$

is calculated to obtain a spectral representation which is then centered for each entry using $\mathbf{X}[u,v] = \mathbf{X}[u-\frac{M}{2}, v-\frac{N}{2}]$. Thereafter, a threshold τ is defined

$$\tau = \alpha \cdot \underbrace{\max_{\substack{1\leq u\leq M\\ 1\leq v\leq N}}}\ \mathbf{X}[u,v]$$

with $\alpha = 0.75$ being an optimization parameter. Consequently, all values smaller than 75% of the maximum value are removed using

$$\mathbf{X}_{\text{thresh}}[u,v] = \begin{cases} \mathbf{X}[u,v] & \text{if } \mathbf{X}[u,v] \geq \tau \\ 0, & \text{otherwise} \end{cases} \tag{4.35}$$

The remaining peaks represent the most characteristic spatial frequencies of the surface image. Irregular structures have several peaks around the center of the real part of the image spectrum, whereas regular images reveal repetitions with larger spatial frequencies. That is why a w-by-w sized Gaussian high-pass filter $\mathbf{H}_\sigma = 1.0 - \frac{1}{2\cdot\pi\cdot\sigma^2}\cdot e^{\frac{u^2+v^2}{2\cdot\sigma^2}}$ with σ = 25 and w = 200 pixels is elementwise applied to the spectral domain of $\mathbf{X}_{\text{thresh}}$ to strongly underweight closer peaks to the center using the Hadamard product

$$\mathbf{X}_{\text{filt}} = \mathbf{X}_{\text{thresh}} \circ \mathbf{H}_\sigma. \tag{4.36}$$

Figure 4.13 visualizes these steps for the image spectrum of a regularly perforated epoxy plate ($C_3S_4M_2P_2X_1$).

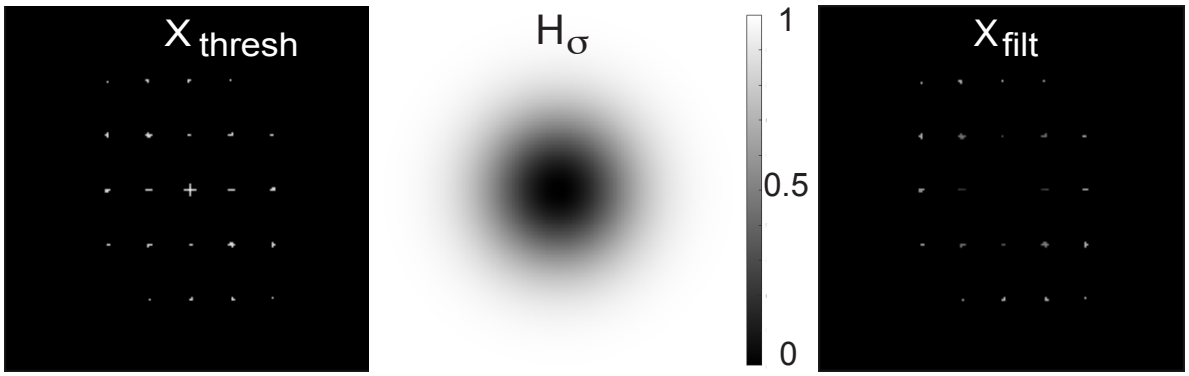


Figure 4.13: Image spectrum before (left) and after (right) applying the Gaussian high-pass filter. The spectral peaks close to the center are removed, only those farther from the center indicate actual regularity (adapted from [2] © 2019 IEEE).

Spectra of irregular surface images scatter close to the center, whereas regular surface images can reveal specific peaks far from the center. The threshold τ and the filter dimensions are result of the feature optimization (Section 4.2.2). Finally, macrotexture regularity

(MaRG$^{\text{TUM}}$) is calculated as the product of the min-max range $r_{\text{min-max}} = r_{\max} - r_{\min}$ of the IR sensor values and the mean value of all pixels in $\mathbf{X}_{\text{filt}}$ as

$$\text{③} \quad MaRG^{\text{TUM}} = r_{\text{min-max}} \cdot (\overline{\mathbf{X}}_{\text{filt}} - C(\mathbf{I}_{\text{disp}})) \tag{4.37}$$

The image contrast is subtracted for two reasons. First, homogeneous images with low contrast are perceived as regular in a haptic context. Second, high-contrast irregular surface images should lead to a very small MaRG$^{\text{TUM}}$ value. The multiplication of $r_{\text{min-max}}$ follows the similar motivation as for MaCO$^{\text{TUM}}$ and aids to prevent that irregularly painted structures on actually flat, homogeneous surfaces distort the feature value. Note that MaCO$^{\text{TUM}}$ uses the vibrotactile energy following a similar motivation, but MaRG$^{\text{TUM}}$ uses the independent IR-based value in order to create two less correlated features MaCO$^{\text{TUM}}$ and MaRG$^{\text{TUM}}$, respectively.

4.1.5.2 TacTUM Microscopic Roughness

The microtexture roughness (MiRO$^{\text{TUM}}$) and microtexture coarseness (MiCO$^{\text{TUM}}$) describe the microscopic roughness features of a material surface.

Microtexture Roughness MiRO$^{\text{TUM}}$ The majority of the microscopic roughness feature design steps are based on signal statistics and frequency-domain features as shown in Section 4.1.3, which has been proven to be successful for controlled robotic scan approaches. However, larger operator-induced variances in speed and normal forces potentially influence the robustness of such features. The Texplorer2 is designed to mitigate the force variations, and the operator generally slides over a material sample within reasonable speed ranges. Consequently, even the tactile features from Section 4.1.3 potentially have distinctive classification ability among different scans when the data is collected with the Texplorer2. Nevertheless, more elaborated features can prove to be more robust compared to those from related work. For example, the Daubechies or Coiflet family wavelets as introduced in Section 2.3.4 reveal notable visual similarity to vibrotactile signals. In order to address the speed dependency for operator-based approaches, this thesis introduced a DWT algorithm in [7] based on two, instead of one, transformed signals which is explained in the following.

The wavelet decomposition of a vibrotactile signal $\mathbf{v}_{\text{t,s}}$ resulting from, e.g., a stainless steel tool tip recording, leads to approximation (A) and detail (D) signals which change comparably for different recording speeds. The detail signals are related to the higher frequencies in the signal, and hence, represent microscopic roughness. The corresponding feature from [7] is based on the difference of the two detail levels $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(5)}$ and is calculated using

$$\text{④} \quad MiRO^{\text{TUM}} = \log_{10}\left(1 + \overline{\mathbf{D}^{(1,5)}}\right) \tag{4.38}$$

with

$$\mathbf{D}^{(1,5)} = \mathbf{D}^{(1)}(\mathbf{v}_{\text{t,s}}) - \frac{\overline{\mathbf{D}^{(1)}(\mathbf{v}_{\text{t,s}})}}{\overline{\mathbf{D}^{(5)}(\mathbf{v}_{\text{t,s}})}} \cdot \mathbf{D}^{(1)}(\mathbf{v}_{\text{t,s}}). \tag{4.39}$$

$\mathbf{D}^{(k)}(\mathbf{v})$ contains the absolute values of the k^{th} detail level of the 1D DWT of $\mathbf{v}$. The logarithm is used to achieve evenly distributed feature values, since roughness-related vibrotactile signal amplitudes scale exponentially between completely smooth and very rough surfaces [62], [116]. The used wavelet type is a Daubechies 8 (db8) wavelet resulting from the feature optimization (Section 4.2.2). Figure 4.14 visualizes the feature calculation for the rough material surface of volcanic tuff.

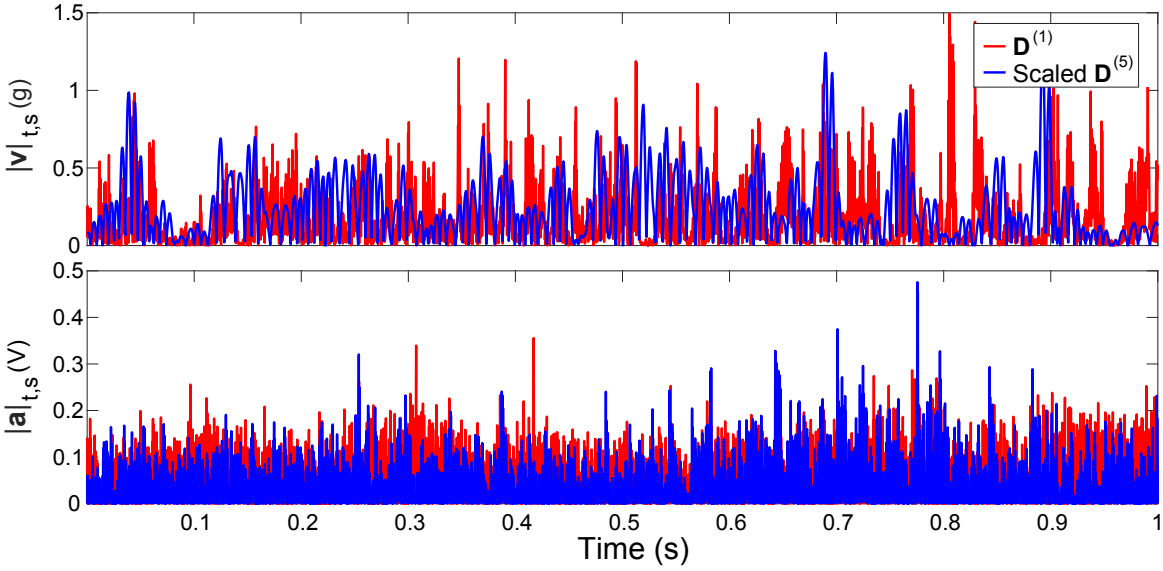


Figure 4.14: Microtexture roughness feature visualization (top: vibrotactile, bottom: audio) for volcanic tuff material sample. Note that both the detail levels (red and blue) have similar signal energy on average, but reveal differences in their spikiness. The average of all elementwise-calculated differences is characteristic for a specific material recording (adapted from [7] © 2017 IEEE).

Microtexture Coarseness MiCO$^{\text{TUM}}$ Fishel et al. [116] define fineness, the opposite of coarseness, based on the SC definition. This thesis uses the vibrotactile steel tool tip recordings and applies this definition on the real part of the vibrotactile signal spectrum $\mathbf{V}_{\text{abs,t,s}}$ as

$$\text{⑤} \quad MiCO^{\text{TUM}} = f_{\text{s}} - \frac{\sum_{i=1}^{N} \mathbf{f}[i] \cdot |\mathbf{V}_{\text{abs,t,s}}[\mathbf{f}[i]]|}{\sum_{i=1}^{N} |\mathbf{V}_{\text{abs,t,s}}[\mathbf{f}[i]]|} \tag{4.40}$$

with f_{s} being the sample rate of the accelerometer. Smaller MiCO$^{\text{TUM}}$ values (i.e., large spectral centroids) indicate the presence of a fine (i.e., smooth) material surface, e.g., a polished steel surface. Figure 4.15 shows the absolute spectrum of the material sample scan volcanic tuff representing a slightly more coarse than fine material in the context of this feature.

Fishel et al. [116] further examined the conversion between spatial and temporal frequency, which is defined as

$$f = \frac{v}{\lambda} \tag{4.41}$$

Intuitively, the absolute scan speed v has a notable impact on the dominant frequency f during a scan over a material surface with λ-sized surface particles. However, Fishel et al. [116] have reported that frequencies do not necessarily relate linearly to the spatial wavelength

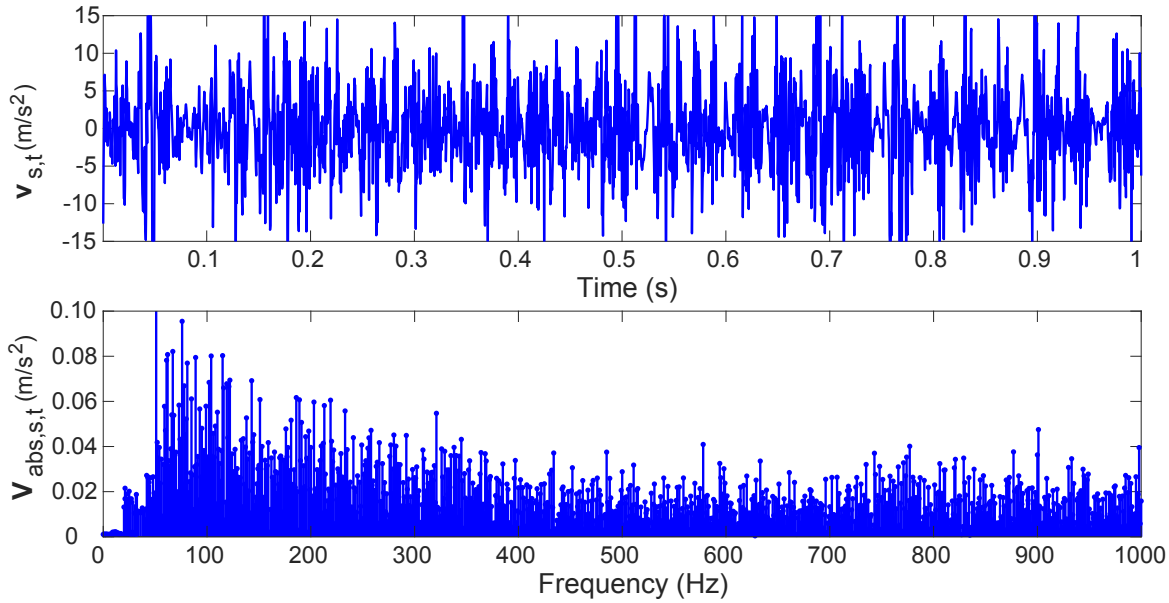


Figure 4.15: Microtexture coarseness MiCO^{TUM} feature data source (volcanic tuff material sample scan). The spectral centroid is located left to the center of the spectrum, and hence, the material structure can be considered as microscopically coarser than fine.

of a texture. The spectral centroid has been demonstrated to remain even constant for the majority of their 117 sample material database. Materials with coarsely-shaped surfaces reveal an increase in SC, but others such as graphite actually decrease with higher scanning speed. This highly relevant observation is further validated by this thesis using the larger database of 184 material classes. The initial assumption of linear speed – frequency relation does not hold for natural material samples, which reveal a highly complex spectral response while sliding a sensorized tool over a surface. MiCO^{TUM} is presumably a robust feature, consequently, and should be further investigated in material classification.

4.1.5.3 Friction

Tactile stiction (FST^{TUM}) and sliding resistance (FRS^{TUM}) can be considered equivalent to the static and dynamic friction coefficient of a material surface following the Coulomb friction model.

Tactile Stiction FST^{TUM} and Sliding Resistance FRS^{TUM} The two FSRs introduced in Section 3.3.3.2 measure normal and friction force values during a sliding motion over the material sample. Note that the normal FSR touches the surface directly, whereas the operator finger internally presses against the FSR measuring the friction force inside TU2. As long as the normal force $\mathbf{f}_n$ is about zero, the operator does not noticeably press on the material surface. An increasing friction force $\mathbf{f}_{fr}$ then indicates the beginning of the static friction phase. Since the TU2 can not directly measure the transition from sticking to sliding over the surface, the ending segment of the static phase needs to be determined following another approach. The friction and normal forces increase differently whenever the operator starts

touching the material in order to perform a slide. Two 10 ms windows are shifted over both the normal and friction force values to identify the maximum positive slopes $\Delta\mathbf{f}_{\text{fr,max}}$ and $\Delta\mathbf{f}_{\text{n,max}}$, and to define tactile stiction as

$$⑥ \quad FST^{\text{TUM}} = \frac{\Delta\mathbf{f}_{\text{fr,max}}}{\Delta\mathbf{f}_{\text{n,max}}} \tag{4.42}$$

Figure 4.16 visualizes an example signal trace (linoleum material sample) and the corresponding normal and friction force signals.

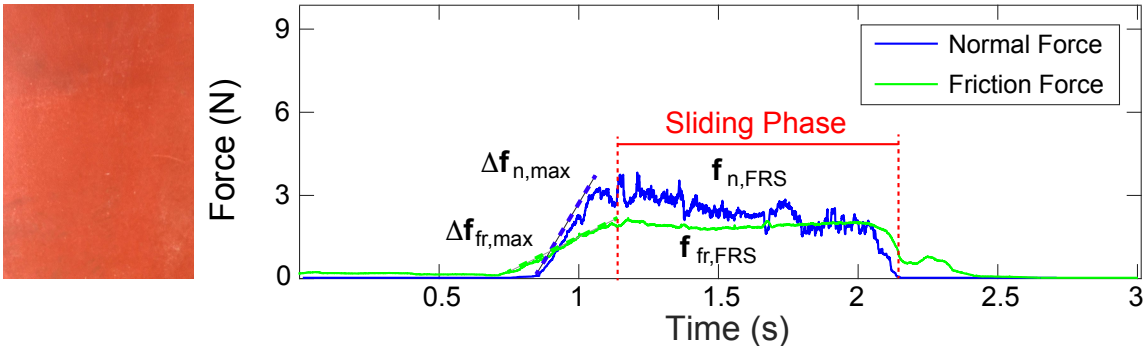


Figure 4.16: Normal and tangential (friction) forces relevant for the calculation of tactile stiction FST^{TUM} and sliding resistance FRS^{TUM}. Note that the increase of the friction force before the normal force increase may result from, e.g., operator finger movements inside the TU2 (adapted from [2] © 2019 IEEE).

The signal values after the ending of the static segment describe the sliding phase until the normal force again decreases to zero. The FSR signals in the sliding segment are denoted as $\mathbf{f}_{\text{fr,FRS}}$ and $\mathbf{f}_{\text{n,FRS}}$, respectively. Consequently, sliding resistance is calculated as

$$⑦ \quad FRS^{\text{TUM}} = \frac{\overline{\mathbf{f}_{\text{fr,FRS}}}}{\overline{\mathbf{f}_{\text{n,FRS}}}} \tag{4.43}$$

The FRS^{TUM} follows the idealized model of Coulomb friction, which has been shown by Fishel et al. [116] to have sufficient relevance in explaining a tactile dimension which cannot be quantified by any other sensing domain.

Adhesive Tack ATK^{TUM} Adhesion is experienced whenever the finger or sensing device is lifted from the surface, and hence, is potentially correlated to the aforementioned two friction features. However, several reasons suggest to not define adhesive tack based on measurements from a sensing device, but on the subjective rating of the human operator instead.

1. Notable adhesion forces are challenging to measure robustly between the human finger and a surface as introduced in Section 2.1.2. The majority of naturally occurring material surfaces are not noticeably adhesive and lead to small force values (in the range of mN) during high-effort measurements as shown by Spinner et al. [72].

2. The adhesion properties of any sensing device also differ from those of a human finger; a material surface scanned with a biomimetic sensor made of silicone can result in a completely different adhesive tack feature value compared to a bare finger interaction.

3. An operator can intuitively assess the stickiness of a material during several short touches, whereas an automated systems requires considerably longer scan time.

4. Any oily or sticky surface that actually is strongly adhesive may damage the sensing device, and the preceding visual detection of such surface conditions is challenging for state-of-the-art robotic systems, but obvious for human operators.

Due to these reasons, subjective ratings are preferable for at least this feature. A feasible trade-off in subjective rating time and subjective feature resolution is the judgment of adhesion based on a 5-point Likert scale defined as

$$\textcircled{8} \quad ATK^{\mathrm{TUM}} = \begin{cases} 0 & \text{if not adhesive} \\ 0.25 & \text{if slightly adhesive} \\ 0.5 & \text{if uncertain} \\ 0.75 & \text{if adhesive} \\ 1 & \text{if strongly adhesive} \end{cases} \tag{4.44}$$

The inclusion of such an operator judgment-based value represents another feature not depending on any other sensing domain. Specifically the correlation to the other friction features is causally decreased, since the interaction material, i.e., PLA vs. operator finger skin tissue, is different in terms of contact adhesion.

4.1.5.4 TacTUM Compliance

Operator-based material scanning approaches reveal an advantage over robot-based approaches for complex object structures in terms of hardness sensing. Humans intuitively identify the ideal location for grasping, folding, or pinching a material sample. The normal FSR and Flex sensor signals are used to infer hardness-related features from Fishel et al. [66] during comparable operator-material interactions. Note that the majority of materials in the LMT-184-HDB are not deformable by reasonable interaction forces applied by humans during the EPs. In detail, the folding angle values $\boldsymbol{\alpha}_{\mathrm{f}}$ of the TU2 act as a prefilter in the binary decision if a material is deformable at all. The operator grasps the material sample and attempts to fold it as strong as possible. If the expression $\sigma(\boldsymbol{\alpha}_{\mathrm{f}}) < 1^{\mathrm{o}}$ holds, then the material sample cannot be folded and is consequently identified as rigid. All five compliance features mentioned in the following are set to zero to avoid noisy feature values of such stiff material samples with respect to common human material exploration forces.

Tactile Compliance CCP$^{\mathrm{TUM}}$ Tactile compliance is intuitively linked to Young's modulus and spring stiffness as introduced in Section 2.3.1. Since the contact area A_{TU2} is constant, the tactile compliance is hence defined as the ratio of the required maximum normal force f_{n} during constant pressing motion and achieved indentation depth, or displacement, Δy. Thanks to the design of the Texplorer2 unit 2, Δy is fixed (as shown in Fig. 4.17), and the corresponding normal force is measured.

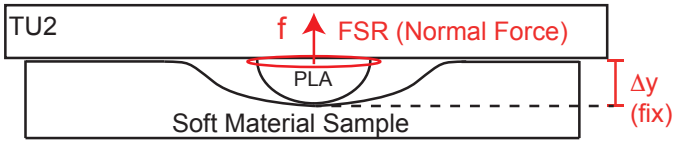


Figure 4.17: Texplorer Unit 2 while being pressed on a soft material sample, leading to a fixed indentation Δy. The normal force FSR (in red) is used to infer the tactile compliance. Note that the indentation depth is constant due to the PLA hemisphere after the operator presses Δ y = 4.0 mm into the soft material sample, and hence, this step resembles the stress relaxation test from Section 2.3.1 (adapted from [2] © 2019 IEEE).

The tactile compliance then is defined as

$$⑨ \quad CCP^{\text{TUM}} = \begin{cases} 0 & \text{if } \sigma(\boldsymbol{\alpha}_\text{f}) < 1^\text{o} \\ \frac{f_\text{max}}{\Delta y} & \text{otherwise} \end{cases} \tag{4.45}$$

with

$$f_\text{max} = \max_{1 \leq n \leq N} \mathbf{f}[n] \tag{4.46}$$

being the maximum force and Δy = 4.0 mm measured with a caliper. If the material sample is deformed up to this indentation depth, any additional operator force is compensated by the whole TU2 casing making this feature robust during subsequent scans.

Local Deformation CDF$^{\text{TUM}}$ The local deformation around TU2 is calculated as

$$⑩ \quad CDF^{\text{TUM}} = \sigma(\boldsymbol{\alpha}_\text{f}) \tag{4.47}$$

describing how the object can be folded during unconstrained interaction based on the achieved folding angle $\boldsymbol{\alpha}_\text{f}$. Note that this feature is determined during a free folding action with the operator being instructed to lift and fold the material sample at least one time to the maximal point. This feature is hence dependent on the operator motion, but allows to identify very specific material samples (papers, foils, etc.) which cannot be identified using the other hardness features. Notably, a robotic system which only presses on a material sample even may identify flexible materials as rigid due to the implicitly measured stiffness of the underlying pad, e.g., of a table.

Damping CDP$^{\text{TUM}}$ Damping is generally understood as the ratio of the counteracting force and the current speed [165]. The temporal change of the folding angle array $\boldsymbol{\omega}_\text{f} = \frac{\boldsymbol{\alpha}_\text{f}}{\Delta t}$ values is measured alongside with the corresponding folding force $\mathbf{f}_{\text{f,max}}(\omega_\text{f})$. Figure 4.18 illustrates how the TU2 provides the data for this definition. The $\boldsymbol{\omega}_\text{f}$ results from a smoothed derivative of the folding angle values. Then, the maximum omega is calculated using

$$\omega_\text{max} = \max_{1 \leq n \leq N} \boldsymbol{\omega}_\text{f}[n] \tag{4.48}$$

and its maximum index

$$n_\text{max} = \arg\max_{1 \leq n \leq N} \boldsymbol{\omega}_\text{f}[n] \tag{4.49}$$

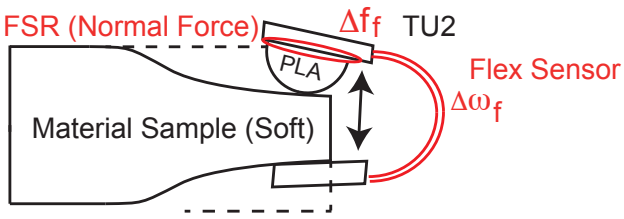


Figure 4.18: Texplorer Unit 2 while pinching a soft material sample. The FSR measuring the normal force and the Flex sensor are shown in red. Damping and relaxation can be defined by the ratio ΔF and ω_f (adapted from [2] © 2019 IEEE).

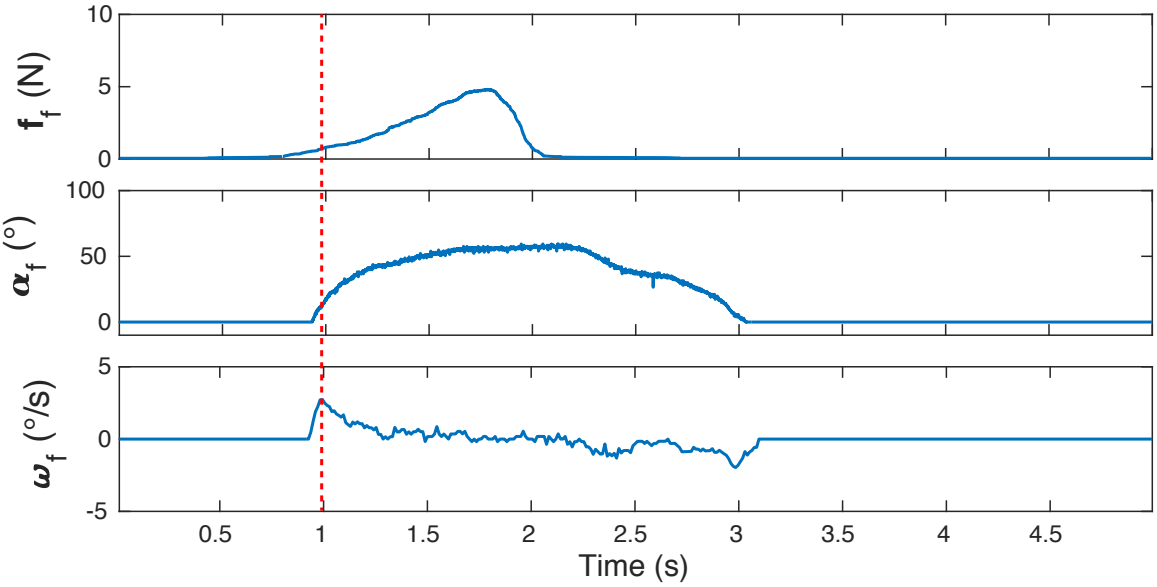


Figure 4.19: Damping-related data for material sample hemp. The current force at the maximum angular speed (dotted red line) divided by the maximum angular speed value is used to define the damping feature.

The corresponding force at the maximum ω_max is denoted as $f_\mathrm{n} = \mathbf{f}_\mathrm{f}[n_\mathrm{max}]$, and consequently, tactile damping is

$$\textcircled{11} \quad CDP^\mathrm{TUM} = \begin{cases} 0 & \text{if } \sigma(\boldsymbol{\alpha}_\mathrm{f}) < 1^\mathrm{o} \\ \frac{f_\mathrm{n}}{\omega_\mathrm{max}} & \text{otherwise} \end{cases} \tag{4.50}$$

Relaxation CRX$^\mathrm{TUM}$ Tactile relaxation is based on the FSR measuring the normal force $\mathbf{f}$ during pressing on the material sample as shown in Fig. 4.20. The difference of the maximum occurring force

$$f_\mathrm{max} = \max_{1 \le n \le N} \mathbf{f}[\mathrm{n}] \tag{4.51}$$

and the sustained force value after a fixed-distance time value, e.g., three seconds $\mathrm{n}_\mathrm{max,3s} = \mathrm{n}_\mathrm{max} + 3\,\mathrm{s} \cdot f_\mathrm{s}$

$$f_\mathrm{max,\,3s} = \mathbf{f}[\mathrm{n}_\mathrm{max,3s}] \tag{4.52}$$

with

$$\mathrm{n}_{\max} = \underset{1 \leq n \leq N}{\arg\max}\, \mathbf{f}[\mathrm{n}] \tag{4.53}$$

and f_s being the sample frequency of the FSR, determines the relaxation behavior of a material sample. Referring to Section 2.3.1, this procedure matches the common stress-relaxation

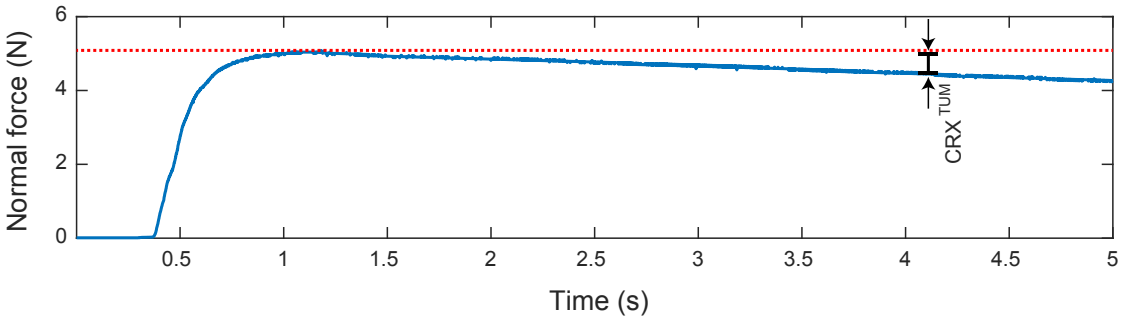


Figure 4.20: Tactile relaxation feature visualization. The red dotted line shows the maximum achieved pressure force during the EP pressure. It can be seen that the material sample relaxes until the end of the recording (i.e., slightly deforms) during the application of constant pressure of the TU2 unit 2. Consequently, $\mathrm{CRX}^{\mathrm{TUM}}$ obtains a value larger than zero.

test because the Texplorer2 imposes a fixed initial ϵ_0 due to its design. Moreover, the stress can be expressed as force because the contact area is constant for all measurements, leading to

$$⑫ \quad CRX^{\mathrm{TUM}} = f_{\max} - f_{\max,\,3\mathrm{s}} \tag{4.54}$$

The majority of material samples in the haptic database sustain their maximum force, and hence, this feature value is characteristic for viscoelastic material samples. It becomes further relevant for haptic display, since viscoelastic behavior is connected to the relaxation as will be discussed in Section 5.2.4.

Yielding $\mathrm{CYD}^{\mathrm{TUM}}$ As defined in Section 2.1.2, yielding is defined as "the degree to which a surface remains deformed after being pressed" [67]. The human operator identifies this feature value without using the Texplorer2 device. The motivation is similar to the adhesive tack feature definition. First, the majority of materials that are scanned are not deformable while being exposed to reasonable human pressure force, and even if deformable, they generally recover to their initial shape. A human operator can quickly assess this fact whereas a robot-based system requires more time and possibly introduces unnecessary scanning uncertainty. For example, a piece of paper placed on a table is identified as rigid by a planar-working robotic system, but the operator intuitively lifts it, recognizes that it is not rigid, and also identifies that it could stay deformed after interaction depending on the applied force. For such scenarios, yielding is defined on a 5-point Likert scale as

$$\textcircled{13}\quad CYD^{\text{TUM}} = \begin{cases} 0 & \text{if material recovers shape} \\ 0.25 & \text{if material state likely recovers} \\ 0.5 & \text{if uncertain or not obvious} \\ 0.75 & \text{if material state unlikely recovers} \\ 1 & \text{if material stays deformed} \end{cases} \tag{4.55}$$

to account for the state uncertainty for such materials after operator interactions. Only reasonable forces are considered during common human–surface interactions.

4.1.5.5 Warmth

Based on the non-contact Newton's law of cooling approach introduced in Section 2.3.5, this thesis evaluates the object surface thermal cooling curve after being constantly heated by a Peltier element. Figure 4.21 shows such a temperature cooling curve for hemp, which is a strongly insulating material.

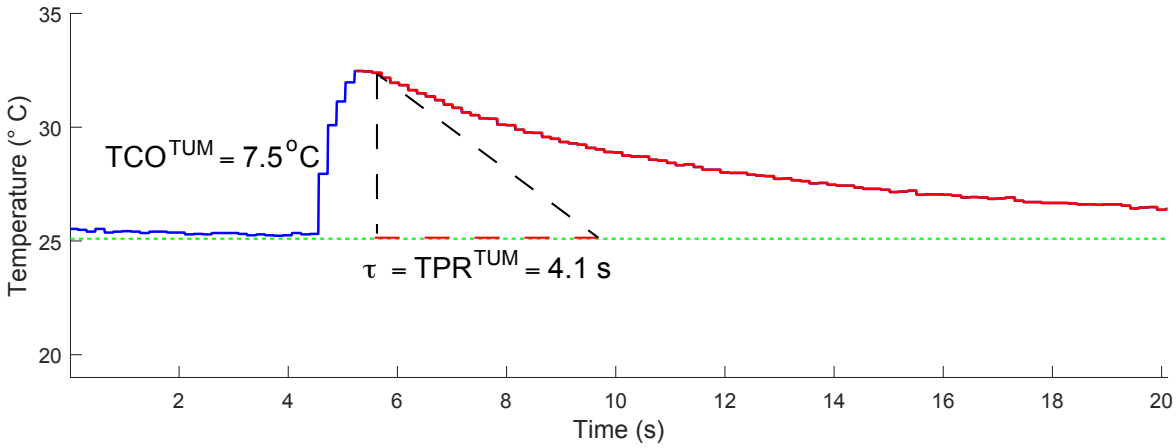


Figure 4.21: Thermal cooling curve for material sample hemp. The ambient temperature of the room is shown as dashed green line. The initial increase of temperature and ongoing decrease indicate thermal properties of the material sample. Note that the thermal cooling time τ is also characteristic for each material.

Thermal Cooling TCO^{TUM} The maximal temperature after being heated is characteristic for the material sample's insulation behavior. The time-discrete thermal signals $\mathbf{t}$ are sampled at $f_s = 200\,\text{Hz}$. Using

$$T_{\max} = \max_{1 \le n \le N} \mathbf{t}[n] \tag{4.56}$$

and measuring, e.g., $T_{\text{amb}} = 19$ °C during the beginning of the recordings, the first thermal feature TCO^{TUM} is inferred as

$$\textcircled{14}\quad TCO^{\text{TUM}} = T_{\max} - T_{\text{amb}} \tag{4.57}$$

This feature is comparable to the first thermal feature presented by Aujeszky et al. [94] (temperature at the center of excitation), which has been shown to successfully differentiate between a small subset of five materials. Insulating materials locally heat up to a significant temperature, whereas metals have almost no measurable difference compared to their initial surface temperature. Note that the ambient temperature needs to be recorded whenever a new recording session starts.

Thermal Persistence TPR^{TUM} The rate of heat transfer which takes place after 5–15 seconds after initial contact is defined as thermal persistence [66]. The thermal cooling curve follows an exponential decay

$$T(t) = T_{\max} \cdot e^{-\frac{t}{\text{TPR}^{\text{TUM}}}} \tag{4.58}$$

and the decay rate τ is visualized in Fig. 4.21 as the intersection of the temperature gradient at the maximum temperature and the constant ambient temperature for the material sample hemp. The thermal persistence TPR^{TUM} is equal to τ and quantified by fitting an exponential curve to the thermal data using the Matlab function fit (exp1).

4.1.5.6 Tactile Feature Speed Dependency Assessment

The final design of the Texplorer2 TU1 allows to capture the scan speed while applying a relatively constant normal force during a slide. Consequently, recordings in a reasonable range of operators' scan speeds lead to signal traces containing this relevant dependency for further assessment. Therefore, the same three materials from the audio speed analysis section 4.1.2.3 are selected to represent the more extreme condition of scan speed dependency, and their data traces are shown in Fig. 4.22 (bare-finger vibrotactile signals), Fig. 4.23 (steel tool tip vibrotactile signals), Fig. 4.24 (friction scans), and Fig. 4.25 (reflectance signals).

Table 4.4: Speed dependency result table for tactile features under increasing recording speed conditions. The material samples are M_1 PLA 8mm, M_2 PLA 2mm, and M_3 Granite. The first three entries denote the slope of the feature vector consisting of five entries (five different speeds), and the last three the standard deviations of the material feature vectors. Ideally, all values are close to zero to indicate independence of different recording speeds. For example, MaCO^{TUM} reveals higher dependency of scan speed, since its visual component is multiplied by the vibrotactile signal energy, which itself heavily depends on the scan speed. Other features, such as FRS^{TUM} are more robust to different scan speeds.

Feature	M_1 m($\mathbf{f}_{\text{speed}}$)	M_2 m($\mathbf{f}_{\text{speed}}$)	M_3 m($\mathbf{f}_{\text{speed}}$)	M_1 $\sigma(\mathbf{f}_{\text{speed}})$	M_2 $\sigma(\mathbf{f}_{\text{speed}})$	M_3 $\sigma(\mathbf{f}_{\text{speed}})$
VS	-0.009	-0.031	0.010	0.022	0.058	0.019
FV	0.007	0.010	0.009	0.018	0.030	0.021
RV	-0.021	-0.004	-0.016	0.043	0.052	0.027
MaTX	-0.101	-0.099	0.030	0.033	0.044	0.032
MaCO	0.116	0.174	0.123	0.191	0.294	0.201
MiRO	0.099	0.094	0.128	0.075	0.066	0.019
MiCO	-0.001	-0.023	0.070	0.029	0.060	0.128
FRS	-0.017	0.022	-0.030	0.032	0.045	0.053

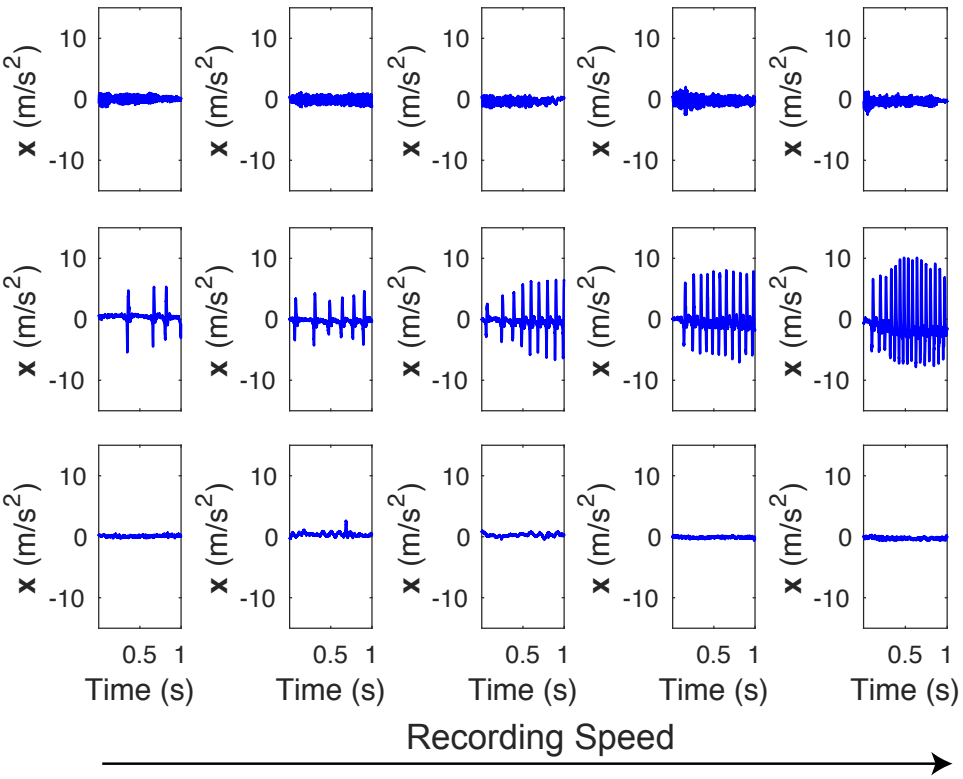


Figure 4.22: Five bare human finger-mounted vibrotactile signal recordings under increasing speed conditions of PLA (2mm, top row), PLA (8mm, middle row), and granite (bottom row).

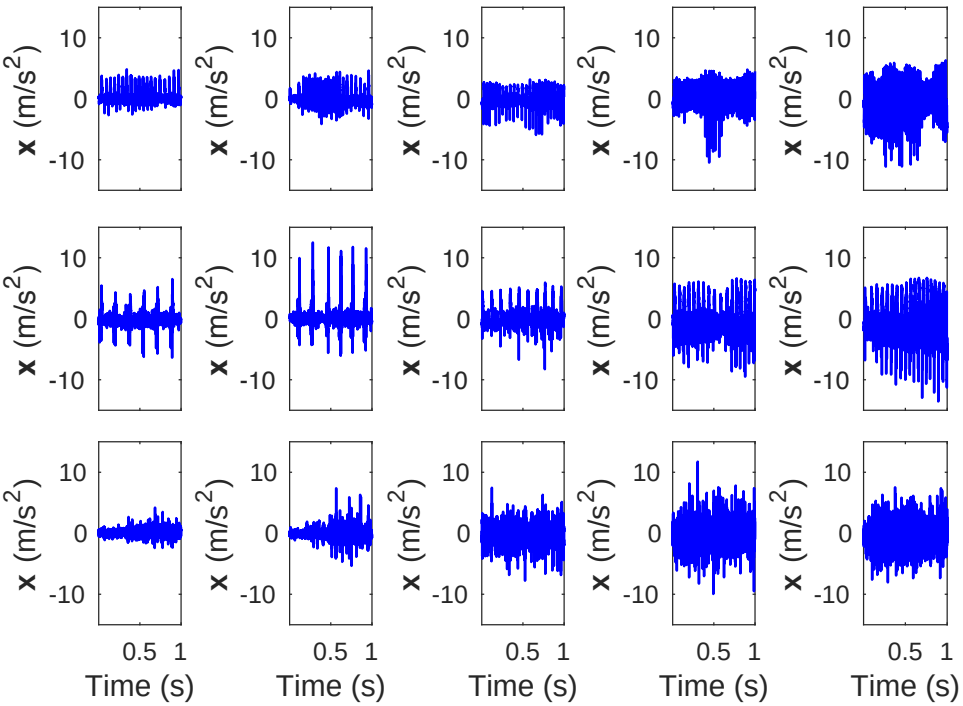


Figure 4.23: PLA (2mm), PLA (8mm), and granite vibrotactile signals recorded with steel tool tip.

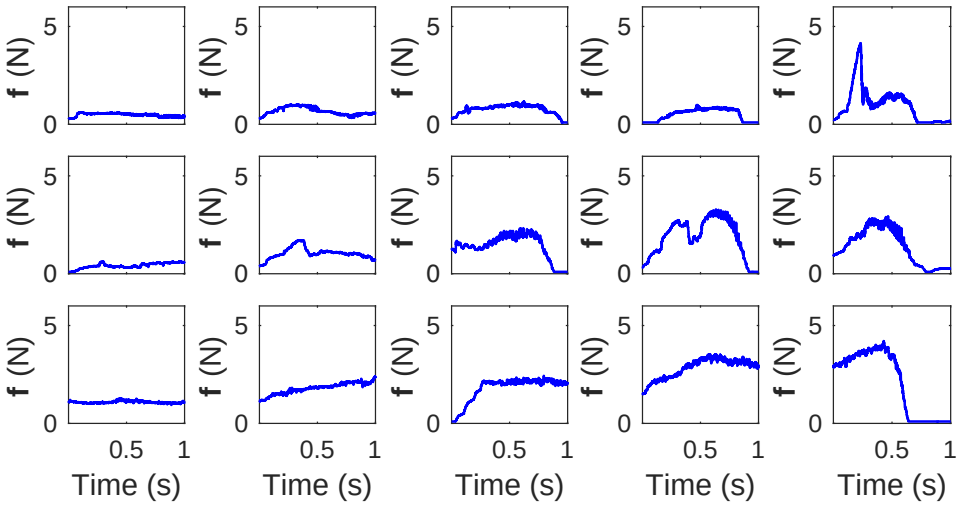


Figure 4.24: PLA (2mm), PLA (8mm), and granite friction signals recorded with TU2 friction unit.

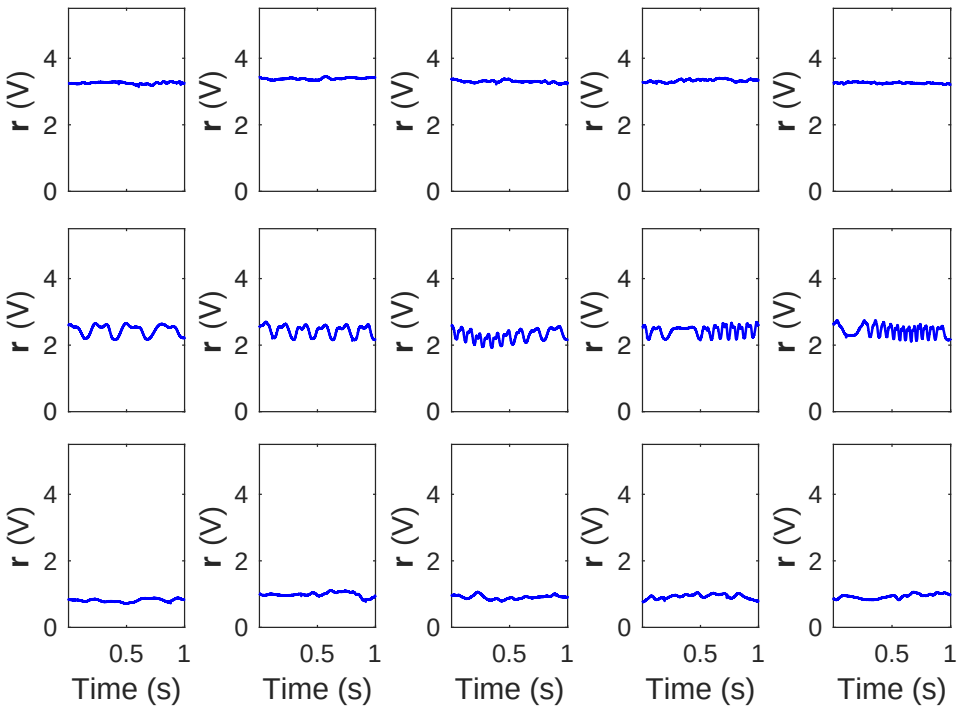


Figure 4.25: PLA (2mm), PLA (8mm) and granite reflectance signals obtained from TU1 sliding recordings. The mean value indicates overall reflectance, and the variations around it the possible presence of surface structure.

Comparable to the audio features in Section 4.1.2, the speed dependency of the tactile features from Section 4.1.5 has been calculated using the steel tool tip recordings. Likewise, the resulting five feature bar plot values ideally stay constant with increasing speed level and show low variance. Figure 4.26 shows the bar plots for the speed depending features, and Table 4.4 their stationarity metrics slope and standard deviation which ideally are constant. Similar to the audio features, the tactile features do not show significant feature variances according to the speed which makes them likewise a potentially suitable input for a material classification engine.

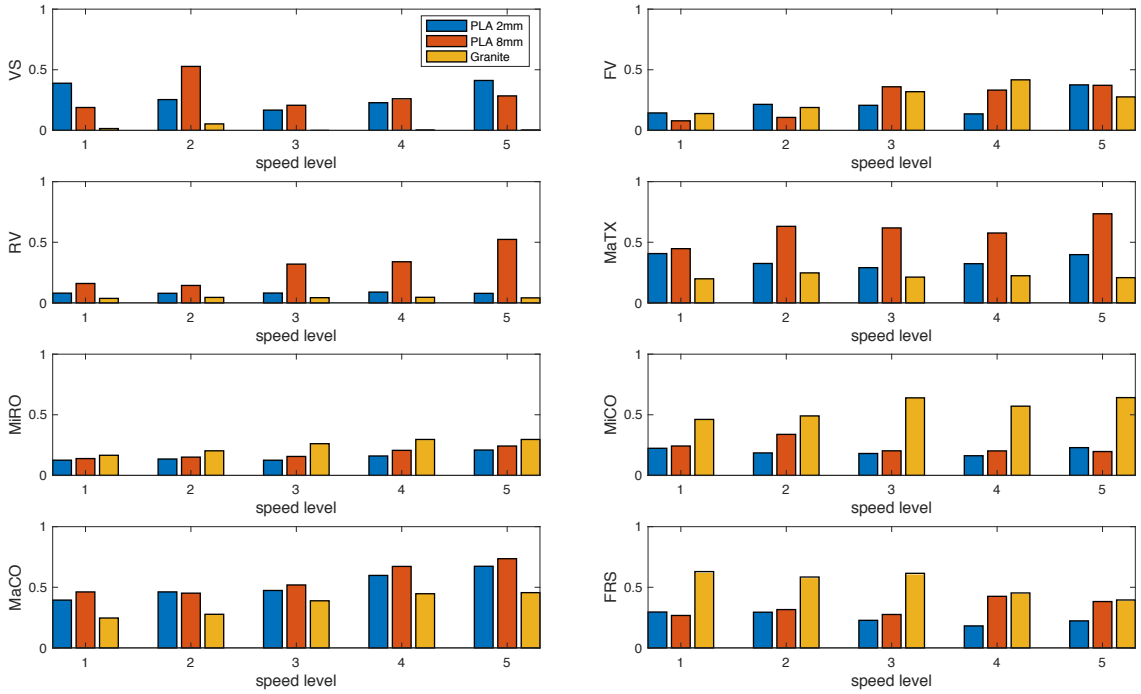


Figure 4.26: TacTUM feature scan speed dependencies. The three components of MaTX^{TUM} reveal speed dependency, however, their linear combination to MaTX^{TUM} reduces the variations. MiRO^{TUM} and MiCO^{TUM} also remain characteristic for the three material samples. Notably and similar to Fishel et al. [222] the fineness MiCO^{TUM} is not clearly increasing with scan speed, i.e, for the granite sample it even decreases. MaCO^{TUM} increases with speed due to the vibrotactile signal component. Sliding friction FRS^{TUM} is relatively independent of scan speed as expected from the Coulomb friction law.

4.1.5.7 Thermo-Tactile Feature Dependency Assessment

The TCO^{TUM} and TPR^{TUM} tactile features shown in Section 4.1.5.5 depend on the ambient temperature of the scan environment. In order to examine this dependency, a well-insulating material sample (hemp) was scanned multiple times between different days and reasonable ambient temperature ranges. Figure 4.27 shows TU2 thermal data recordings (left) and the resulting feature values (right). The first ten recordings (left image, shades of blue plots)

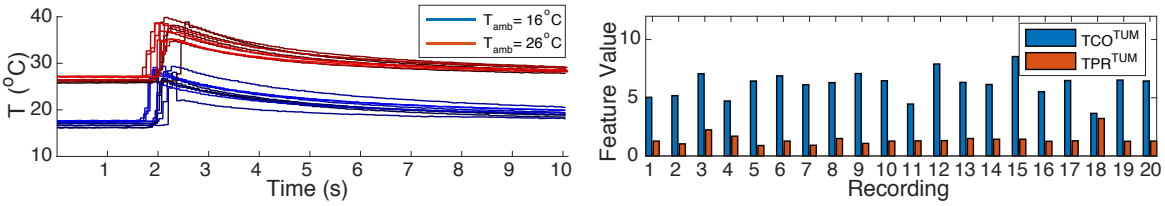


Figure 4.27: Ambient temperature thermal dependency of TCO^{TUM} and TPR^{TUM}. Left: thermal signals recorded at different days, i.e., different ambient temperatures. Right: Calculated feature values according to the day recordings. Both values stay robust regardless the very different ambient conditions.

were recorded at an ambient room temperature of about 16 ^{o}C, and the other ten (left image, shades of red plots) with ambient room temperatures of about 26 ^{o}C. The right bar plot shows that both features stay relatively constant (TCO^{TUM} = 6^{o}C $\pm$ 1.5^{o}C, TPR^{TUM} = 1.2^{o}C $\pm$ 0.4^{o}C) with respect to the ambient temperature. Note that, comparably to the audio and vibrotactile speed recordings, these extremer recording conditions were deliberately chosen to emphasize that under common conditions (i.e., ambient temperature $\approx$ 20^{o}C) both feature values are robust.

4.1.6 Features Exceeding Human Touch Capabilities

The previous sections covered the features connected to the human perception of visual, audible, and tangible information. Technical systems like the Texplorer, however, are equipped with additional sensors to record data traces leading to features that go beyond human sensing capabilities. For example, features related to magnetic or metallic material properties, or the surface's different ability to reflect infrared light can be quantified using the corresponding sensors and feature definitions. Table 4.5 shows such features which do not fall into any of the previously mentioned feature domains, but reveal additional and independent material classification abilities.

Table 4.5: Overview of features exceeding human touch discrimination capabilities, which are independent of the aforementioned audio, visual, and tactile data domains.

ρ	$= \frac{m}{V}$	Density
MR	$= \bar{\mathbf{r}}$	Mean reflectivity
isMetal	$= \overline{\mathbf{m}}$	Metal detection
deltaB	$= \overline{\Delta \mathbf{B}}$	Change of magnetic field

4.2 Data Preparation

The individual feature values need to be analyzed according to their statistics, suitability, and relevance for a ML pipeline. Suitable features are then used to create, select, and to optimize the feature space before a classifier is trained.

4.2.1 Feature Suitability for Machine Learning Approaches

The most important ingredients of ML are content-based, independent and descriptive features [203]. Several steps have been examined to extract useful features for the task of material classification in this thesis.

4.2.1.1 Content-based Features and Task-oriented Approach

The first important step in handcrafted feature-based machine learning is the identification of a large set of task-related and not redundant features [223], which has been conducted in the previous section. For example, the TacTUM features presented in this thesis are heavily inspired by the major tactile dimensions from Okamoto et al. [63] and Fishel et al. [66], and hence, qualitatively fulfill the first condition.

4.2.1.2 Feature Independence

The second prerequisite is expressed by the correlation between feature vectors. They are ideally as low as possible to guarantee feature independence [224]. In this thesis, the Spearman rank correlation coefficient (SRCC) [225]

$$\rho_{\text{SRCC}} = 1 - \frac{6 \cdot \sum_{i=1}^{N} d_i^2}{N \cdot (N^2 - 1)} \tag{4.59}$$

is calculated for two feature vectors $\mathbf{x}$ and $\mathbf{y}$ with d being the rank distance $d_i = R(\mathbf{x}_i) - R(\mathbf{y}_i)$ between their ranks, and N being their lengths. The smallest feature vector entry is associated with the lowest rank and the largest value with the highest rank. If a strong correlation exists, the ranks are also similar, and hence, their difference becomes small. If all rank differences are comparably small, a high correlation exist between both vectors. Note that the Pearson's correlation coefficient is not suitable due to the potential non-linearity in the data whereas the SRCC is non-parametric, and hence, needs no assumption about the underlying data distribution.

Whenever the absolute value of ρ_{SRCC} indicates stronger correlation, i.e., $|\rho_{\text{SRCC}}| \geq 0.5$, one of the two compared features should be excluded from the feature space. Larger correlations can be expected for subdimensions belonging to the same major tactile dimension. For example, FST^{TUM} and FRS^{TUM} belong to the dimension of friction and are intuitively correlated. According to Guyon et al. [224], however, stronger correlations are not automatically a criterion for strict exclusion, i.e., strongly correlated features may increase the classification performance. Nevertheless, non-correlated features are intuitively preferred [224] for any classification system. Figure 4.28 shows the SRCC between all 15 tactile features and the following Figs. 4.29 – 4.31 further show correlations among various domains.

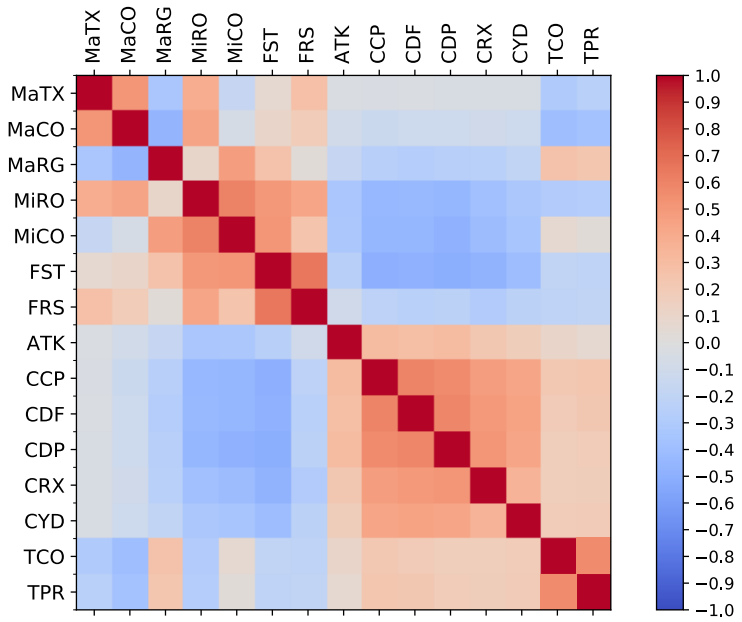


Figure 4.28: SRCCs for the 15 TacTUM features shown as correlation heatmap. Warmth (TCO^{TUM} and TPR^{TUM}) and friction (FST^{TUM} and FRS^{TUM}) features are more correlated, and hence, only one of each respective feature should be selected for the final feature space. Additionally, the hardness-related features are weakly correlated because the majority of the LMT-184-HDB materials are rigid, and hence, all five hardness feature values equal 0.

In general, causal independence is expected if different sensors collect data during different exploratory procedures, e.g., the IR sensor during static touch and the accelerometer during sliding. Figure 4.29 shows how all tactile features are correlated to the audio features, and Fig. 4.30 compares tactile features to those from related work and the ones exceeding human perception.

Only a few tactile features in Fig. 4.28 and 4.30 are correlated, notably FST^{TUM} and FRS^{TUM} as well as TCO^{TUM} and TPR^{TUM}. Note that features like yielding, adhesive tack, or the features going beyond human tactile perception (Fig. 4.30) are weakly correlated to the other features, and hence, are ideal candidates for a final feature space. Further note that audio signal-based and vibrotactile features are also not correlated (Fig. 4.29), and interestingly, the ATCCs are only weakly correlated compared to the chroma feature. The image features (see Fig.4.31) show stronger correlations within the Haralick and Tamura features as well as the RLBPs. Consequently, not all of these features should be jointly selected for a final feature space.

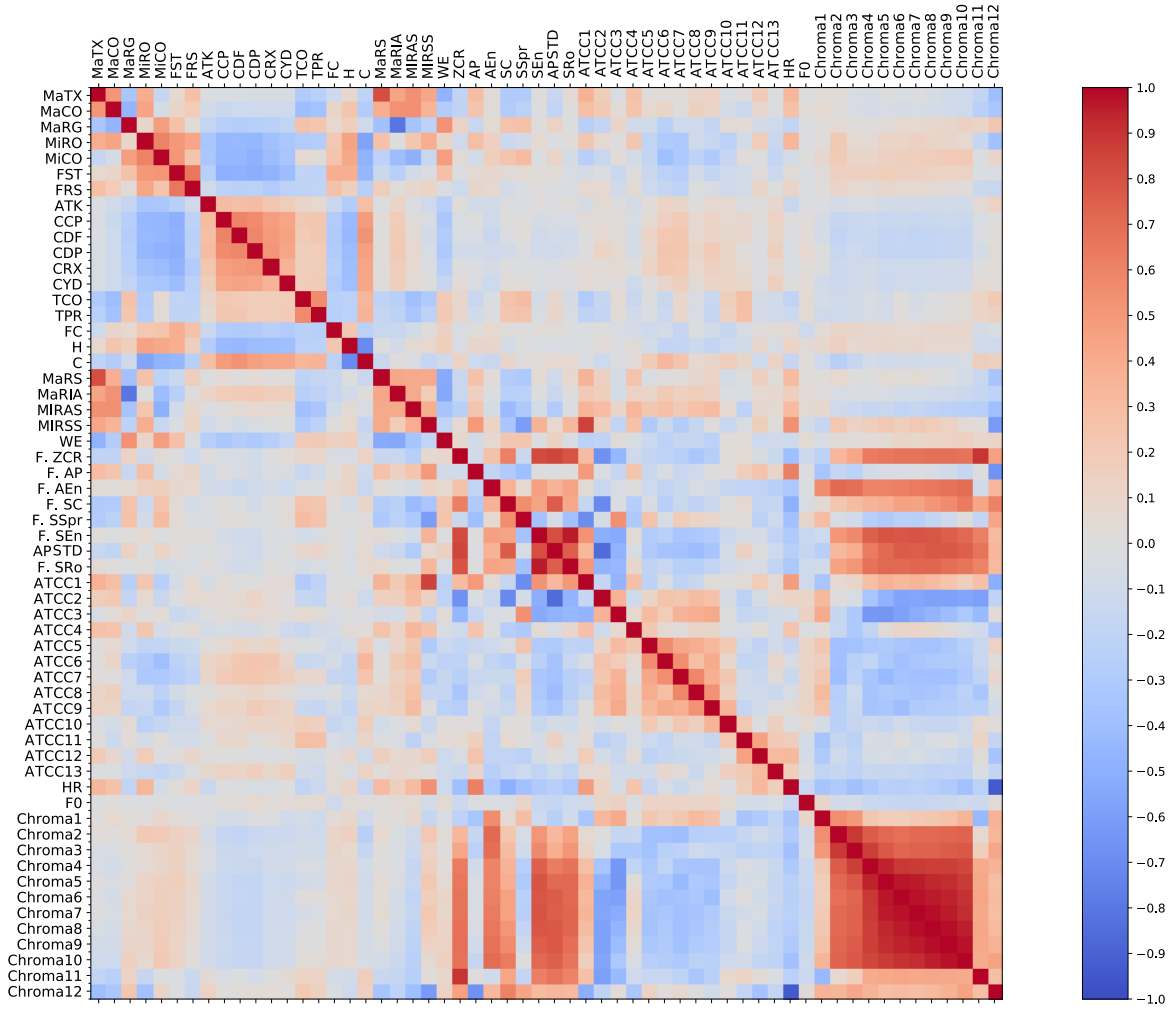


Figure 4.29: Correlation heatmap of tactile features compared to audio features. As important observation, the low correlation values between both domains allow for a combined feature space. Note that within the audio features, ATCCs are weakly and the adapted Chroma features are strongly correlated. Consequently, the ATCCs can be used as a combined feature vector for classification, whereas the Chroma features can be dimensionality-reduced without significant loss of information.

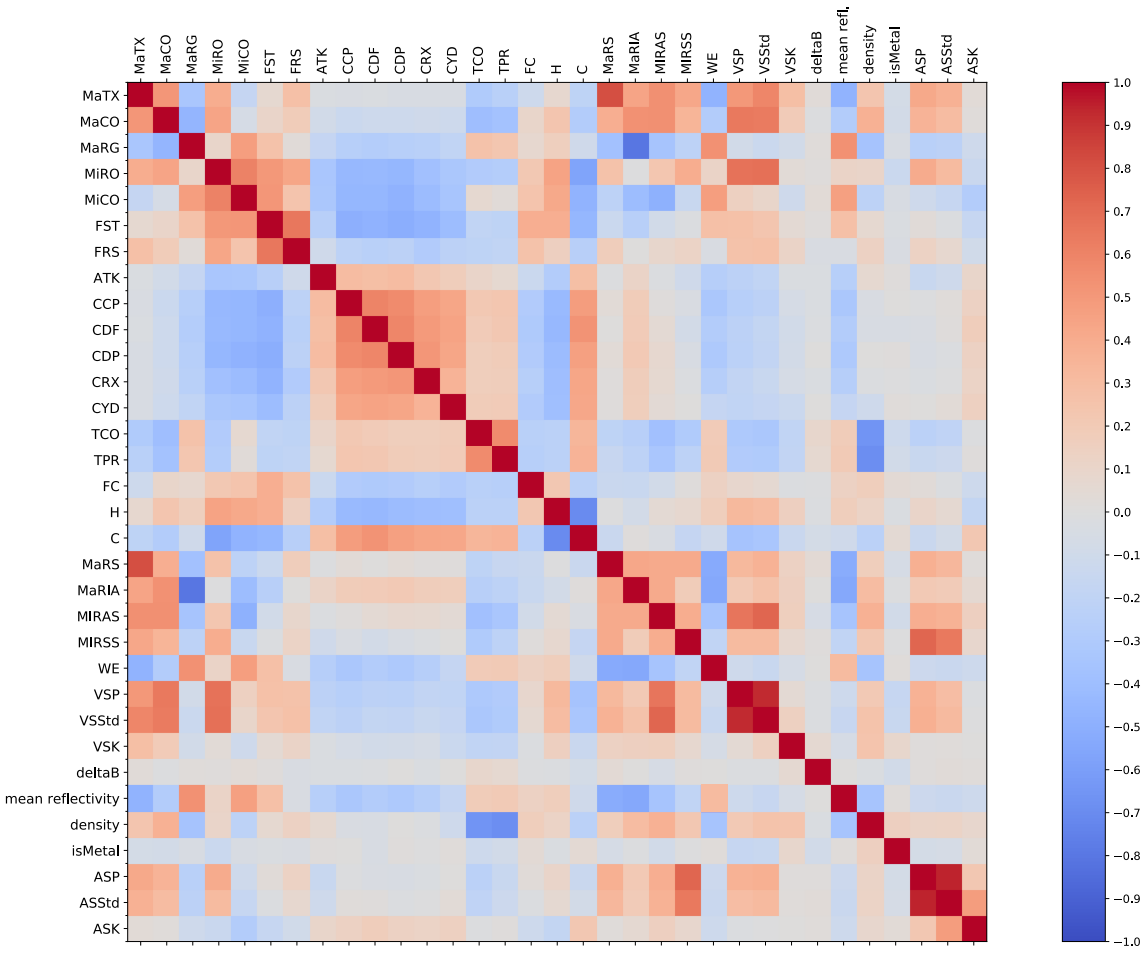


Figure 4.30: Correlation heatmap of tactile features and features going beyond human perception (see Section 4.1.6). Note that density, the Boolean metal detection feature, or the magnetic feature are completely independent, which is most favorable [224] for a classification pipeline and motivates the application of further features from such domains. Ongoing work in material classification should further investigate which other domains are relevant and likewise uncorrelated to the existing data traces.

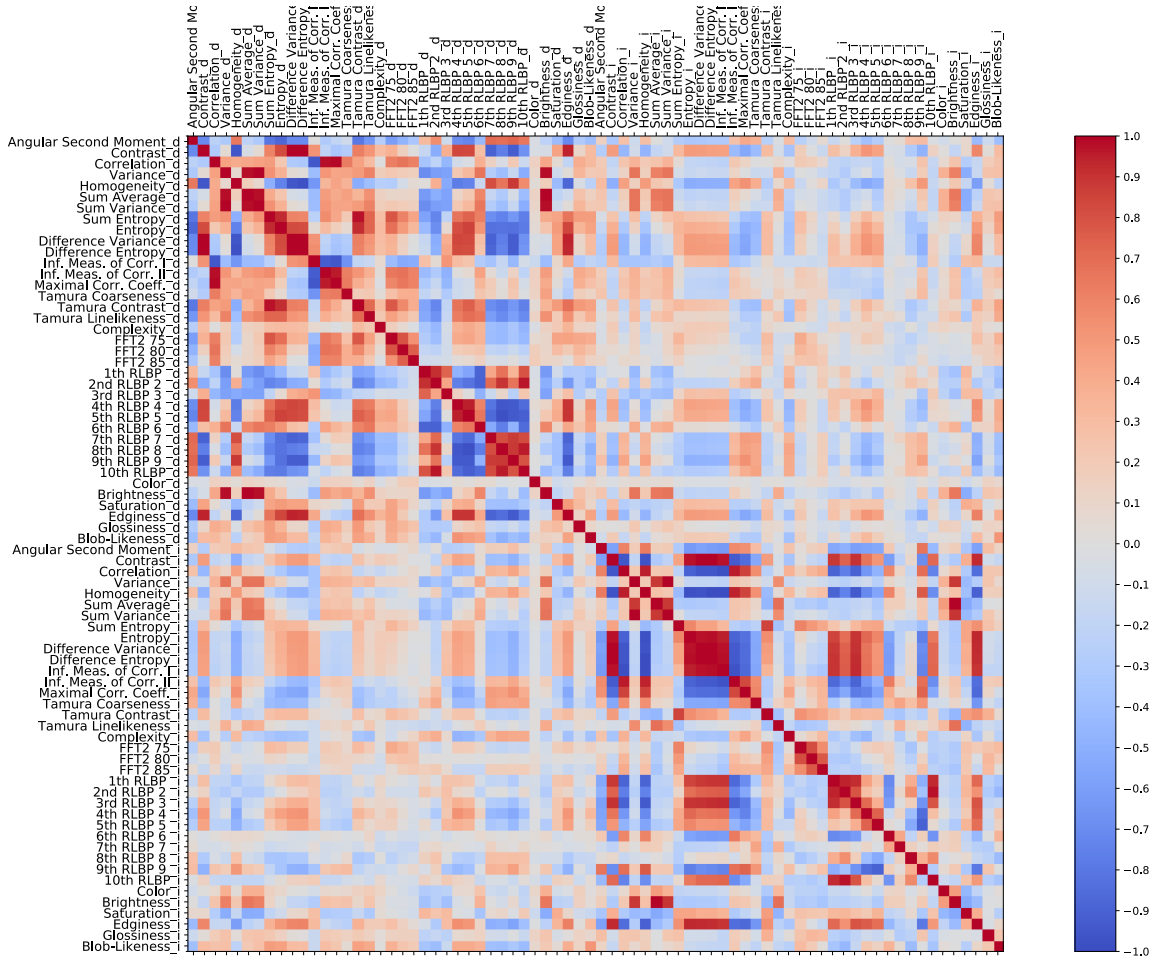


Figure 4.31: Correlation heatmap of surface texture image-based features (marked with subscript _d) compared to illuminated texture image features (marked with subscript _i). In contrast to the expectation of a high correlation among images captured with the same sensor, i.e., a camera, the measured correlations are generally lower than 0.5 between both subdomains. This observation implies that the same image features can be useful based on different image data recording conditions, in this case, by using a flashlight. However, within the image features, stronger correlations exist, e.g., for the ten RLBP or the Haralick-based texture features. Comparable to the audio features, either reduction techniques need to be applied or only individual features need to be selected to avoid potential redundancy in the final feature space.

4.2.1.3 Feature Quality

The third prerequisite of well-suited features for supervised classification is their ability to be distinctive among different material classes. In this case, each feature alone (one-way) is examined, i.e., a single column of the feature space is considered in the following. Since supervised classification is applied in this thesis, the observations (rows of the feature space) that belong to the same material, i.e. have the same label, are denoted as groups. Consequently, the objective of well-discriminating feature design is to generate groups that are significantly different compared to the other groups. This procedure is known as statistical hypothesis testing, which follows the steps mentioned next.

Identification of the Null Hypothesis The null-hypothesis of N groups in this case is $H_0 : \mu_1 \approx \mu_2 \approx ... \approx \mu_\mathrm{N}$, i.e., it is initially expected that the means of all groups are not significantly different, which is undesired in the context of the feature design, since all material classes then lead to a similar feature vector and cannot be distinguished. Consequently, the alternate hypothesis $H_1 : \exists i, j : \mu_\mathrm{i} \neq \mu_\mathrm{j}$ ideally leads to the rejection of the null hypothesis since at least one material group is different from the others. This fact alone is not sufficient for qualitative feature design; in the case of multiple material classes, post-hoc tests (multiple comparison tests [226]) among each group need to be performed to check and count how often the null hypothesis can be rejected for all pairwise group comparison.

Selection of an Appropriate Test to Evaluate the Hypothesis The underlying data distribution of each group, i.e., material class, determines which statistical test is applicable. First, it is relevant to test if the data follows a normal distribution $\mathrm{X} \approx \mathrm{N}(\mu, \sigma^2)$ or not by, e.g., using the Kolmogorov-Smirnov (KS) test [178]. Whenever a normal distribution is present, parametric one-way analysis of variance (Anova) is selected as test [227]; if any other distribution is present, the non-parametric Anova version Kruskal-Wallis test [228] is applicable. This test compares the medians of multiple groups without requiring a specific probability distribution in the data. The Kruskal-Wallis test is preferably selected since several features do not necessarily follow a normal distribution. With respect to the relation between different samples, the scans of one and the same material sample can be considered as independent, since the individual scan trajectories over the material surface can differ significantly during subsequent recordings.

Approval or Rejection of the Null Hypothesis The Kruskal-Wallis test returns a score (F-score) which is a ratio of the variance between and within the groups, and also a p-value which shows how reliable the F-score actually is. Larger p-values indicate that the result of the test occurred by chance, and hence, are not statistically significant; the null-hypothesis cannot be rejected since the ratio of within-group variation to between-group variation is not large enough. For descriptive features, however, it is necessary that this null-hypothesis is rejected for as many group comparisons (i.e., material classes) as possible. As initially mentioned, each group is pair-wise tested using post-hoc tests based on the Matlab function multcomp using the non-parametric Kruskal-Wallis (or alternatively, the Mann-Whitney-U)

test and Tukey's honest significant difference criterion to obtain a vector $\mathbf{p}$ containing the p-values of all comparisons. A threshold, denoted as significance level α, is generally set to be 0.01 or 0.05 and indicates whether a 1% or 5% chance exists that the test result is not reliable. The more p-values are smaller than α, the more likely it is that the results are reliable and did not occur by chance. Consequently, a feature quality (FQ) metric is defined as

$$FQ = 100\% \cdot (1 - \overline{\mathbf{p}}) \tag{4.60}$$

based on the average of all p-values $\mathbf{p}$ resulting from the pairwise comparisons $\frac{N \cdot (N-1)}{2}$ with N being the number of groups. For example, if all 184 materials from the LMT database are used as groups, and hence, N = 184, then 16,836 comparisons are made and lead to a similarly long vector $\mathbf{p}$. A descriptive feature shows a small intraclass and a large interclass variance, and hence, has a large FQ value. Note that this metric is used as one target parameter for feature optimization in the next Section 4.2.2 and also as filter for feature selection in Section 4.2.4 Feature quality in this context can be visualized using boxplots with the individual groups (i.e., material classes) represented as boxes and shown for CCP^{TUM} and a subset of materials in Fig. 4.32. For example, the feature box for material class 89 (hemp)

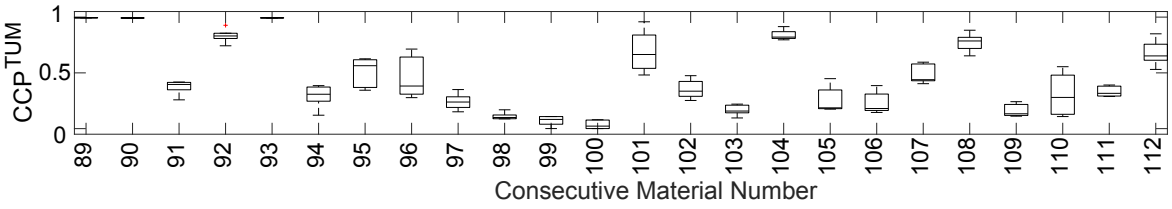


Figure 4.32: Example boxplot of feature CCP^{TUM} of class C_2 (biodegradables like furs or leather). The feature value is set to zero for stiff materials during its calculation, whereas it has a noticeable distinction performance for soft materials (adapted from [2] © 2019 IEEE).

may not be statistically different to the boxes of material class 90 (flax) or 93 (sheep wool), but is very likely to be different, and hence characteristic, to all other material classes shown in this boxplot. An ideal feature boxplot shows small boxes with mean values scattered over the whole feature value range.

Another highly relevant insight gained by feature quality and boxplots is the direct comparison of vibrotactile signals recorded with a stainless steel tool tip or the bare human finger. Figure 4.33 shows, for example, the MiRO^{TUM} feature evaluated for these different data traces. It can be observed that the steel tool tip recordings are more characteristic for the individual material groups due to a higher feature quality value as well as smaller and more scattered boxes, and are hence preferable in the context of material classification.

4.2.2 Feature Optimization

Several features from Section 4.1.5 contain parameters that need to be optimized in terms of feature quality. For example, the optimal wavelet type for the feature MiRO^{TUM} needs to be determined by testing different wavelet types followed by an evaluation on how the feature quality and its correlation to the other features changes. The optimization approach follows the same idea of high feature quality and low correlation between the features. The

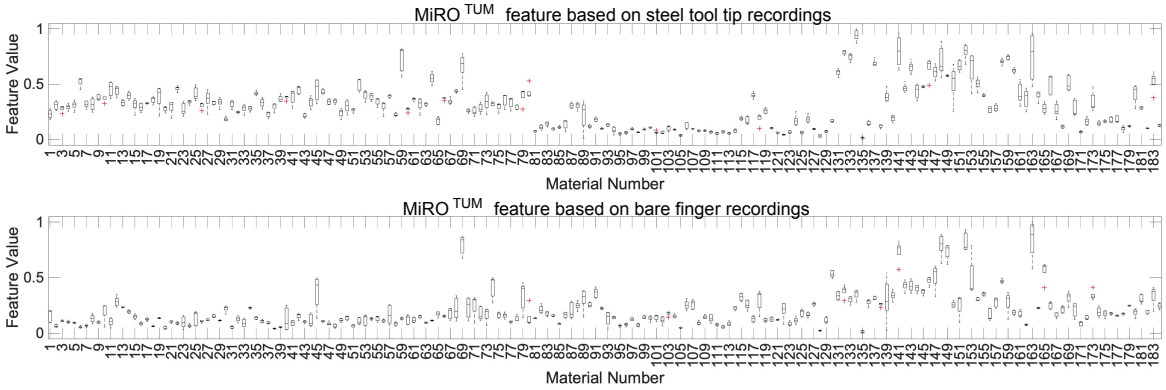


Figure 4.33: MiRO$^{\text{TUM}}$ boxplot for steel tool tip (top) and bare finger vibrotactile recordings. For example, the steel tool tip recordings are clearly different for the first 80 material classes (different woods) compared to the subsequent 40 material classes (e.g., furs and leathers) and thereby indicate that steel tool tip recordings have a larger distinction capability than bare finger recordings.

corresponding feature space columns need to be recalculated using different parameter sets. A grid search of potential parameters was performed to identify the maximum FQ and minimum correlation values. For example, the weighting parameters $\lambda_{1...3}$ of MaTX$^{\text{TUM}}$ result in a set of (3/9, 2/9, 4/9) indicating that the influence of the reflectance sensor data is most relevant and also causally independent since the reflectance sensor is not used for any other TacTUM feature.

4.2.3 Feature Space Creation

Each individual feature vector $\mathbf{x} \in \mathbb{R}^{\text{O x 1}}$ from the various aforementioned domains is concatenated to form an initial feature space $\mathbf{X} \in \mathbb{R}^{\text{O x F}}$. F denotes the number of features, which initially is 252 before feature selection takes place, and O is the number of observations, i.e., feature values or instances I. Additionally, the known label vector, generally denoted as true predictor values $\mathbf{y}_{\text{true}}$, or short $\mathbf{y}$, is concatenated as last column to the feature space $\mathbf{X}|\mathbf{y}$. As a specific contribution and different to comparable machine learning problems, this thesis does not only rely on a single label vector, but multiple ones each representing the different nuances of the material name taxonomy $C_{\text{x}}S_{\text{x}}M_{\text{x}}P_{\text{x}}X_{\text{x}}$ proposed in Section 3.1.2.

$$\mathbf{Y} = \begin{bmatrix} y_{\text{C}}^{(1,1)} & y_{\text{S}}^{(1,2)} & y_{\text{M}}^{(1,3)} & y_{\text{P}}^{(1,4)} & y_{\text{X}}^{(1,5)} \\ y_{\text{C}}^{(2,1)} & y_{\text{S}}^{(2,2)} & y_{\text{M}}^{(2,3)} & y_{\text{P}}^{(2,4)} & y_{\text{X}}^{(2,5)} \\ \dots & \dots & \dots & \dots & \dots \\ y_{\text{C}}^{(O-1,1)} & y_{\text{S}}^{(O-1,2)} & y_{\text{M}}^{(O-1,3)} & y_{\text{P}}^{(O-1,4)} & y_{\text{X}}^{(O-1,5)} \\ y_{\text{C}}^{(O,1)} & y_{\text{S}}^{(O,2)} & y_{\text{M}}^{(O,3)} & y_{\text{P}}^{(O,4)} & y_{\text{X}}^{(O,5)} \end{bmatrix} \tag{4.61}$$

For example, the major class labeling level ($\mathbf{y}_C$) comprises the root material class labels C_1 ... C_8 (woods, biodegradables, ..., composites) regardless the subclass S or the material M. Subclass classification further includes the subclass S, e.g., hardwood or softwood.

Material-level classification additionally considers the material M, such as aluminum, ABS, or oak, and is of major interest for this thesis. P and X further introduce different labels whenever the material is differently processed and composed, respectively. These five different levels of labeling lead to a label matrix $\mathbf{Y}$ containing $\mathbf{y}_{\mathrm{C}}$, $\mathbf{y}_{\mathrm{S}}$, $\mathbf{y}_{\mathrm{M}}$, $\mathbf{y}_{\mathrm{P}}$, $\mathbf{y}_{\mathrm{X}}$. Note that the used taxonomy provides another metric used for describing the performance of the classification system beside other commonly used metrics by considering the misclassified labels based on the true label array $\mathbf{y}$ and the predictions $\hat{\mathbf{y}}$. The denominator is the overall number N of test instances, and the numerator is computed using the misclassified major class labels $\mathbf{y}_{\mathrm{C}} \neq \hat{\mathbf{y}}_{\mathrm{C}}$ based on the class identifier $\hat{y}_{\mathrm{C}}$ which is extracted from each misclassification $\hat{y}$, leading to the novel metric taxonomy similarity

$$Sim_{\text{tax,C}} = \frac{\sum_{n=1}^{N}(1 - I(\mathbf{y}_{\mathrm{C}}[n] \neq \hat{\mathbf{y}}_{\mathrm{C}}[n]))}{N} \tag{4.62}$$

using the indicator function $I(\omega)$, which is equal to 1 if the given condition ω is true, and 0 otherwise. It is considered as critical whenever the predicted class label $\hat{y}_{\mathrm{C}}$ is not equal to the true class label y_{C} which does not increase the numerator by one for each misclassified observation. If all misclassified labels share at least the same major class (e.g., C_1 woods), this metric is desirably close to 1.0, i.e., 100%. The same considerations can be made for the subclass-based similarity $Sim_{\text{tax,S}}$.

4.2.3.1 Data Partitioning

The feature space is split according to a stratified k-fold cross-validation [229] using the Matlab function cvpartition. Leave-one-out cross-validation (LOOCV) is a special case of k-fold cross-validation with k being set to the number of observations. A comparison between both approaches was made and, as a result, the achievement in accuracy is comparable, but the computational time differs substantially. Consequently, k-fold cross-validation is applied in this thesis.

4.2.3.2 Missing and Inconsistent Data Handling

Only numerical values are part of the feature space. Further note that the feature space is initialized with the specific value $f_0 = 0 - \epsilon$ with ϵ being the machine precision constant, i.e., a very small number. If any value is not correctly calculated, this value is used instead without distorting the feature space.

4.2.3.3 Feature Space Scaling

In order to make the features comparable to each other [203], a standardized feature space $\mathbf{X}_{\text{std}}$ is created. Each feature vector entry, i.e. feature value $\mathbf{f}[i]$, is standardized [230] by applying

$$\mathbf{f}[i] = \frac{\mathbf{f}[i] - \bar{\mathbf{f}}}{\sigma(\mathbf{f})} \tag{4.63}$$

Note that features larger than three-times the standard deviation are outliers, and hence, are capped. The raw feature space $\mathbf{X}_{\text{std}}$ is denoted as $\mathbf{X}$ in the following.

4.2.4 Feature Selection

Feature selection becomes necessary whenever larger numbers of relevant and descriptive features are present in the initial feature space. It is of fundamental importance to select a set of features that leads to high accuracy and predictive capability for unseen test data [231]. As general practice, the number of features is kept as low as possible to avoid over-fitting while using a sufficient number of descriptive features to avoid under-fitting [232]. Consequently, feature selection is a trade-off between using various predictive and uncorrelated features, and on the opposite, reducing the curse of dimensionality [223]. If too many features are selected, the high-dimensional feature space gets sparse and classification algorithms are more likely to fail in establishing generalizable decision boundaries for their predictions.

Different approaches have been reported [224], [231] to create the necessary subset of features which are listed in the following.

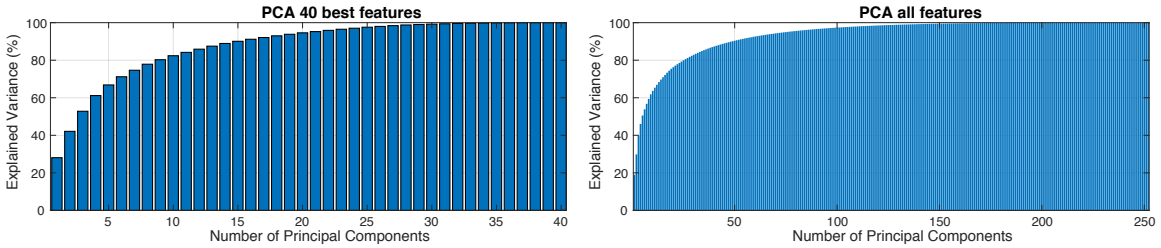


Figure 4.34: Explained variance over the principal components (PCs) for the best 40 features (left) and all features (right). Since many PCs are required to explain the majority of the variance in the data, both feature spaces should not be reduced by PCA.

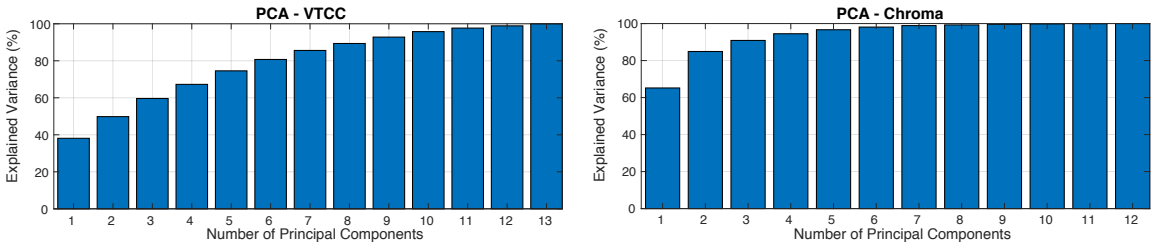


Figure 4.35: Explained variance over the PCs for VTCCs (stainless steel tool tip audio recordings, left) and the chroma vector (right). Since the majority of the variance is explained by three PCs for the chroma vector, the three PCs can be used as reduced set of chroma features instead.

1. **Principal Component Analysis** Principal component analysis (PCA) is performed on the initial feature space to reduce the dimensionality by summarizing strongly correlated features to principal component (PC) axes. The three major PCs only explain about 30% of the variance for the given feature space in this thesis, and the first ten about 62%, as shown in Fig. 4.34. An exhaustive test of PCA and different classifiers shows that 82 feature are necessary to represent 95% of the variance, which potentially leads to over-fitting if all PCs are used. Consequently, feature reduction by PCA is not preferred in this context. However, note that several features such as the Chroma values are highly redundant and can be reduced to fewer values as shown in the right

part of Fig. 4.35. By contrast, VTCCs (left part of Fig. 4.35) do not reveal significant principal components and are hence used as individual features.

2. **Filter Methods** assess the predictive power of each feature directly. Statistic metrics, such as the p-value or the correlation to the class variable are used to assess whether features are predictive. They do not depend on any classifier and exclude weak features that are not predictive. These methods further help in preselecting redundant features. The feature quality metric FQ from Section 4.2.1 was applied and the n-best standalone features have been determined.

3. **Wrapper Methods** make use of classifiers (which are explained in detail in Section 4.3) to follow a greedy-strategy in order to achieve the optimum in accuracy. In its most trivial form, a brute force feature selection approach (forward or backward [224]) can presumably find the optimum over a long period of calculation time after achieving a terminating condition, for example, by achieving a specific target accuracy. If different classifiers are used, a majority voting (MV) scheme can be applied on the features most frequently selected during this procedure. Since the number of features in this thesis is about 250, multiple runs can be performed in reasonable time. One advantage of exhaustive brute-force algorithms is that they have a higher probability in finding the actual global optimum.

4. **Embedded Methods and Implicit Feature Selection** based on, for example, the decision tree [233] and random forest [234] classifiers (see Section 4.3) inherently identify the most predictive features during their training procedure.

Table 4.6: Best selected features from different domains. The last column aggregates the best features from the individual domains such as image or audio features. All rows are ordered based on the feature importance for each domain. Note that the column Misc* consists of all features that were either used in related work or do not belong to the other domains such as image or perceptually tactile features (adapted from [2] © 2019 IEEE).

Image	Tactile	Audio	Misc*	All
I_{edg}^{illu} [9]	$MiCO^{TUM}$	2^{nd} ATCC [9]	Density (Table 4.5)	Density (Table 4.5)
I_{br}^{disp} [9]	$MiRO^{TUM}$	1^{st} ATCC [9]	MR (Table 4.5)	$MiCO^{TUM}$
I_{coa}^{illu} [204]	TCO^{TUM}	Sp. Spread [212]	VSP (Table 4.3)	MR (Table 4.5)
I_{entr}^{disp} [206]	$MaTX^{TUM}$	Sp. Centroid [212]	VSSD (Table 4.3)	$MiRO^{TUM}$
I_{com}^{disp} [206]	TPR^{TUM}	ZCR [212]	VSK (Table 4.3)	I_{edg}^{illu} [1]
I_{con}^{illu} [1]	$MaCO^{TUM}$	Sp. Entropy [212]	SC (Table 4.3)	2^{nd} VTCC [9]
5^{th} RLBP [209]	$MaRG^{TUM}$	6^{th} ATCC [9]	R (Table 4.3)	I_{br}^{disp} [9]
9^{th} RLBP [209]	FRS^{TUM}	4^{th} ATCC [9]	FP2 (Table 4.3)	TCO^{TUM}
4^{th} RLBP [209]	FST^{TUM}	3^{th} ATCC [9]	isMetal (Table 4.5)	I_{coa}^{illu} [204]
10^{th} RLBP [209]	CCP^{TUM}	5^{th} ATCC [9]	deltaB (Table 4.5)	$MaTX^{TUM}$

In this thesis, a hybrid approach of feature selection is applied. Beforehand, filter methods remove completely redundant features [224]. The results of all different approaches are

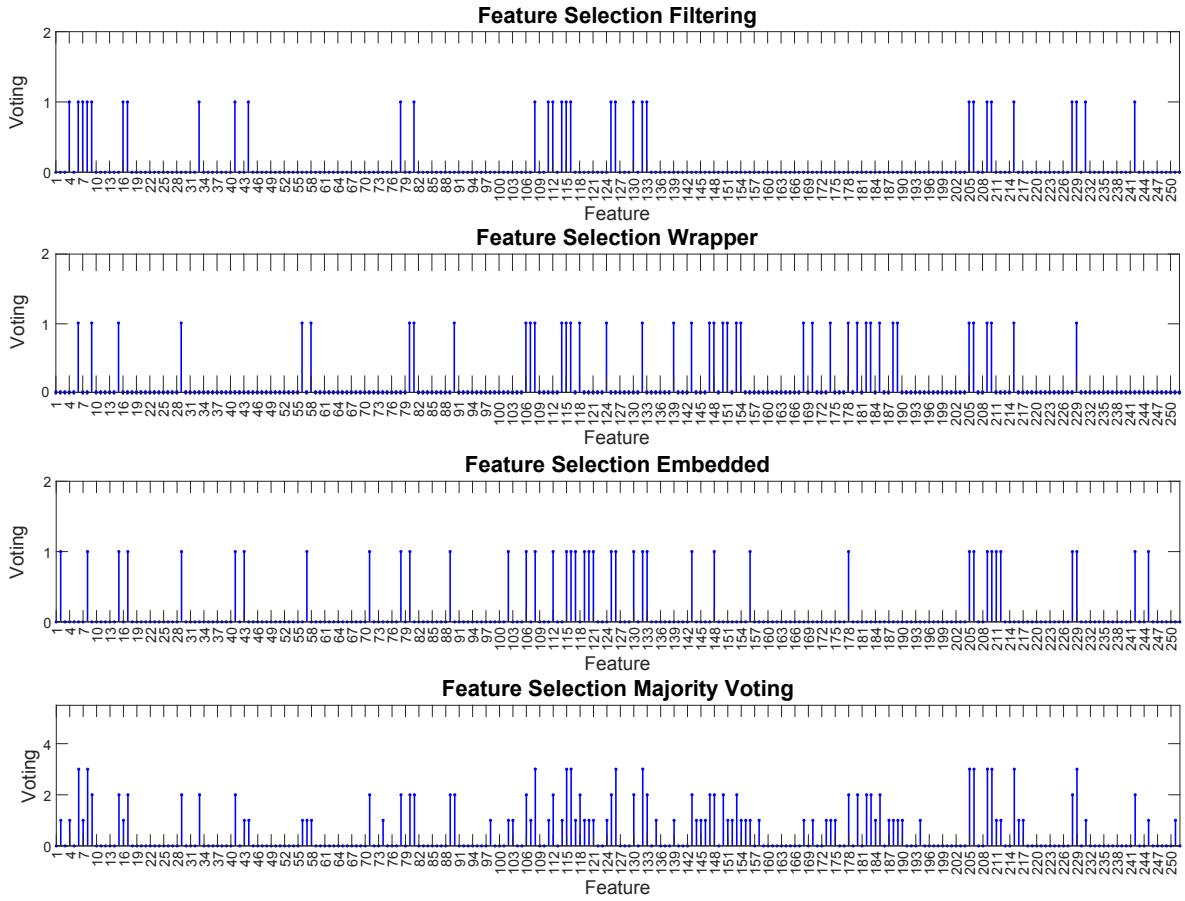


Figure 4.36: Results of different feature selection approaches to find the most frequently selected features in a majority-voting approach (bottom plot). Note that the SynTouch-related tactile features in this thesis (numbers 112 - 126) are frequently selected, which is another indicator for their suitability for material classification.

used in a majority voting scheme to create a final set of very predictive features. The individual feature selection results (wrapper, filter, embedded) are accumulated to identify the features which are most frequently selected. Subsequently, the SRCCs across the set of selected features are calculated. Whenever a critical correlation ($\rho > 0.75$) occurs, one of the two respective features was removed from the feature selection set (generally the one with the lower FQ value). A final majority voting scheme is shown in Fig. 4.36, and the resulting best features are considered as best features for classification. Table 4.6 lists the ten best features for different domains.

4.3 Model Selection and Classifier Fitting

A well-known problem [235] in handcrafted feature-based machine learning is the question on the selection of an appropriate model, denoted as classifier c. Based on a training feature space $\mathbf{X}_{\text{train}}$ and the corresponding true labels $\mathbf{y}_{\text{train}}$, c is fit using

$$c = train(\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}) \tag{4.64}$$

After this training step, the T instances of the test data $\mathbf{X}_{\text{test}}$ are used to make predictions $\hat{\mathbf{y}}$ with

$$\hat{\mathbf{y}} = predict(c, \mathbf{X}_{\text{test}}) \tag{4.65}$$

which are compared to the true label vector $\mathbf{y}_{\text{test}}$ in order to assess its classification performance. Various metrics can be calculated based on the true and predicted material labels. Their definition is originally tied to binary classification tasks, however, has been extended to multiclass classification tasks by the application of either the one-vs-all or one-vs-rest strategy [203]. Two classes, generally denoted as $\oplus$ and $\ominus$, define a binary classification scenario. The number of all T instances belonging to class $\oplus$ is $POS = \sum_{t=1}^{\mathrm{T}} I(y = \oplus)$, and of those belonging to class $\ominus$ similarly as $NEG = \sum_{t=1}^{\mathrm{T}} I(y = \ominus)$ using the indicator function $I()$ which equals one whenever its argument (condition) is true. Note that $POS + NEG = T$. If, for example, a single true label is $y = \ominus$, but the material classification system predicted $\hat{y} = \oplus$, an instance of class $\ominus$ was wrongly predicted as belonging to class $\oplus$. Confusion matrices, also known as contingency tables and shown in Table 4.7, contain the four possible output combination between true and predicted class labels. True positives (TP), for example, represent the amount of predicted $\oplus$ classes which actually are also true classes $\oplus$.

Table 4.7: Contingency table adapted from [203]. Note that each multiclass scenario, i.e., with more than two classes, can be reduced to a per-class binary confusion matrix (one-vs-all or one-vs-rest strategy [203]) allowing for the application of the same metrics.

	Predicted True	Predicted False
Actual True	$TP = \sum_{t=1}^{\mathrm{T}} I(y = \hat{y} = \oplus)$	$FN = \sum_{t=1}^{\mathrm{T}} I(y = \oplus, \hat{y} = \ominus)$
Actual False	$FP = \sum_{t=1}^{\mathrm{T}} I(y = \ominus, \hat{y} = \oplus)$	$TN = \sum_{t=1}^{\mathrm{T}} I(y = \hat{y} = \ominus)$

These values are further used to calculate the classification metrics which are summarized in Table 4.8.

Table 4.8: Classification metrics [203] based on contingency matrix entries.

Accuracy	Acc =	$\frac{1}{T} \cdot \sum_{t \in T} I(y = \hat{y}) = \frac{TP+TN}{TP+FP+FP+FN}$
Class Precision	Prec =	$\frac{\sum_{t \in T} I(y=\hat{y}=\oplus)}{\sum_{t \in T} I(\hat{y}=\oplus)} = \frac{TP}{TP+FP}$
Recall, True Positive Rate (TPR)	Rec =	$\frac{\sum_{t \in T} I(y=\hat{y}=\oplus)}{\sum_{t \in T} I(y=\oplus)} = \frac{TP}{POS}$
False Positive Rate (FPR)	FPR =	$\frac{\sum_{t \in T} I(y=\oplus,\hat{y}=\ominus)}{\sum_{t \in T} I(y=\oplus)} = \frac{FP}{NEG}$
F_1 Score	F_1 =	$2 \cdot \frac{Prec \cdot Rec}{Prec+Rec}$

Precision, recall and F_1 score are calculated class-wise (e.g., for each of the 184 material

classes during fine-level classification) and averaged. Note that the F_1 score is based on the geometric mean of precision and recall, and consequently leads to a lower value than the arithmetic mean whenever either precision or recall are considerably different. Since class imbalance is faced during material classification, the following example for coarse-level classification is discussed. Table 4.9 shows an example material classification engine output for the class C_3 (plastics) with $\hat{y} = \hat{C}_3$ denoting the predicted plastics from the classification engine. Considering only the third class (plastics with $\oplus = plastic, \ominus = other\ material$), the

Table 4.9: Multi-class contingency table for specific material classification example based on the overall class labels C.

	$\hat{y} = C_1$	$\hat{y} = C_2$	$\hat{y} = C_3$	$\hat{y} = C_4$	$\hat{y} = C_5$	$\hat{y} = C_6$	$\hat{y} = C_7$	$\hat{y} = C_8$
$y = C_1$	406	11	3	0	0	0	0	0
$y = C_2$	10	94	9	0	0	2	0	5
$y = C_3$	2	9	73	1	0	1	0	4
$y = C_4$	0	1	0	16	0	3	0	0
$y = C_5$	0	0	2	0	13	5	0	0
$y = C_6$	0	2	3	3	1	131	0	0
$y = C_7$	0	0	1	0	0	4	55	0
$y = C_8$	6	1	5	0	0	0	1	17

same table can be reduced to the binary Table 4.10 to calculate the metrics for class C_3.

Table 4.10: Converted Multi-class contingency table for class C_3 plastics.

	Predicted as true	Predicted as false
Actually true	73	2 + 9 + 1 + 1 + 4 = 17
Actually false	3 + 9 + 2 + 3 + 1 + 5 = 23	406 + 10 + 6 + ... + 5 + 4 = 807

The metrics for class C_3 can be calculated as

$$\text{Acc} = 100\% \cdot \frac{73 + 807}{920} = 95.65\% \tag{4.66}$$

$$\text{Prec} = 100\% \cdot \frac{73}{73 + 23} = 76.04\% \tag{4.67}$$

$$\text{Rec} = 100\% \cdot \frac{73}{73 + 17} = 81.11\% \tag{4.68}$$

$$\text{FPR} = 100\% \cdot \frac{23}{23 + 807} = \ 2.77\% \tag{4.69}$$

$$\text{F}_1 = 78.49\% \tag{4.70}$$

The comparison of the metrics Acc and F_1 score for C_3 shows a significant difference in overall classification accuracy and class C_3 F_1 score. This observation originates from the class imbalance caused by class C_1 and the overall number of NEG compared to POS. The classification system performs well on the overall classification of all materials, especially those from C_1, but performs worse on identifying materials from class C_3. The calculations of all per-class metrics are repeated for all other seven classes and their precisions, recalls (mean average sensitivity, MAS), and F_1 scores are averaged. In this context, macro-averaging (e.g., macro-precision) is equivalent to the arithmetic mean of all class precisions and is also known as mean average precision (MAP). On the other hand, micro-averaged metrics consider the class instance occurrences as weights, and consequently, lead to a weighted average of the metric. Since both recall and precision are equally relevant for the material classification, the F_1 score is used in this thesis alongside the accuracy. Note that in other scenarios, e.g., medical tests, the precision, recall, and other metrics like the FPR, are not necessarily equally relevant for the outcome of the classification task.

After the definition of metrics, individual classifiers are evaluated in the following. Fernandez-Delgado et al. [235] examined the vast amount of 179 different supervised classification approaches on 121 datasets and summarized their observations which classifiers generally perform better than others on average. The random forest classifier and the support vector machine with Gaussian kernel, also denoted radial base function (RBF) performed best across all datasets. It is, however, likewise relevant to test more simplistic classifiers first. These potentially tend to under-fit, i.e., lead to a lower accuracy, but likewise tend to generalize better. One the other hand, more complex classifiers are more likely to over-fit the results, but achieve a high overall accuracy on both the training and test set since they can learn more complex structures from the data. This thesis compares several classifiers and their characteristics and, whenever obvious, highlights the classifier properties that make them suitable for the specific task of material classification.

4.3.1 Linear Classifiers and Non-linear Kernel Support Vector Machines

Linear classifiers rely on the geometric concept of hyperplanes to separate the instances in the feature space [203] for binary classification. As aforementioned, the binary decision rule can be extended to any multiclass scenario by applying a one-vs-all or one-vs-rest strategy [203]. The type of hyperplane for the two-dimensional case is a line which is defined by its normal equation $w_1 \cdot x_1 + w_2 \cdot x_2 + w_0 = 0$, or in general notation, $\mathbf{w}^T \cdot \mathbf{x} + w_0 = 0$. The vector $\mathbf{w}$ is normal (i.e., oriented perpendicular in 2D) to its corresponding hyperplane. Given a set of N training data points $\mathbf{x}_i$ which contain a binary label $y_i \in \{-1, 1\}$ the objective is to identify the line (or in general a hyperplane) which optimally separates the training data by assigning the positive class to the one side of the trained linear separation and the negative class to the other side as

$$\hat{y} = \begin{cases} +1 \text{ if } \mathbf{w}^T \cdot \mathbf{x}_i + w_0 > t \\ -1 \text{ otherwise} \end{cases} \tag{4.71}$$

based on the decision threshold t, i.e., to identify the optimal $\mathbf{w}$ and offset w_0 (also known as bias b in the context of deep learning). This optimization problem on how to define the optimal $\mathbf{w}$ characterizes the core task of linear classifiers such as perceptrons [236], [237] or support vector machines (SVMs) [238]. Note that the perceptron classifier and logistic regression classification accuracies generally are inferior compared to SVMs [235], and hence, are not further discussed.

Support vector machines are the extension of the binary perceptron model and a frequently used classifier in comparable classification tasks. Instead of finding one of the infinite number of linear decision boundaries between the data points through iteration, SVMs are based on two equidistant hyperplanes $H_{-1,1} : \mathbf{w}^T \cdot \mathbf{x} + w_0 = \pm 1$ which separate two classes. Since SVMs are based on the linear separation of data points, they are not ideal to handle overlapping regions or to learn nonlinear decision boundaries. However, two techniques allow to bypass this strict constraint of linear separability.

First, this constraint can be mitigated by introducing slack variables ζ_i for each training sample as

$$\mathbf{w}, w_0 = \min_{\mathbf{w}, w_0} \frac{1}{2}||\mathbf{w}||^2 + C \cdot \sum_{i=1}^{N} \zeta_i \text{ subject to } y_i \cdot (\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1 - \zeta, 1 \leq i \leq N \tag{4.72}$$

with C being a hyperparameter that controls the influence of how strong the condition of linear separability can be violated. This version of SVMs is denoted as soft-margin linear SVM and is applied whenever linear decision boundaries are desired, but noisy or scattered data potentially violates the linear decision boundary condition. The slack-based penalty added to the learning rule allows for the classification of overlapping samples between the decision boundary. Hence, the inclusion of C increases the generalization capability of the SVM classifier, but potentially decreases the classification performance.

Another technique changes the input feature space directly. For applications based on physical measurements, such as the material classification task in this thesis, the decision boundaries between data clusters (i.e., different classes) are not necessarily linear, but more likely to form clusters of specific materials samples. One reason can be found in the distribution of data points, and hence, the distribution of the feature values. The feature space of any linear classifier, e.g., the perceptron or SVMs, can be transformed non-linearly to solve non-linear classification tasks, which is generally referred to as the "kernel trick" [203]. Since repeated physical sensor measurements (e.g., vibrotactile scans of the same material surface) potentially follow a Gaussian probability distribution, the most relevant kernel function for this thesis is the (Gaussian) radial basis function (RBF) kernel, which is defined for the distance between the two vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ as

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = e^{\left(\frac{-||\mathbf{x}_i - \mathbf{x}_j||^2}{2 \cdot \sigma^2}\right)} \tag{4.73}$$

with σ being the bandwidth of the Gaussian kernel, which is typically redefined as second hyperparameter $\gamma = -\frac{1}{2 \cdot \sigma^2}$. The image montage in Fig. 4.37 based on two range-normalized features illustrates the reason why different linear and the non-linear SVM (RBF) need specific care considering the tuning of C, and notably γ.

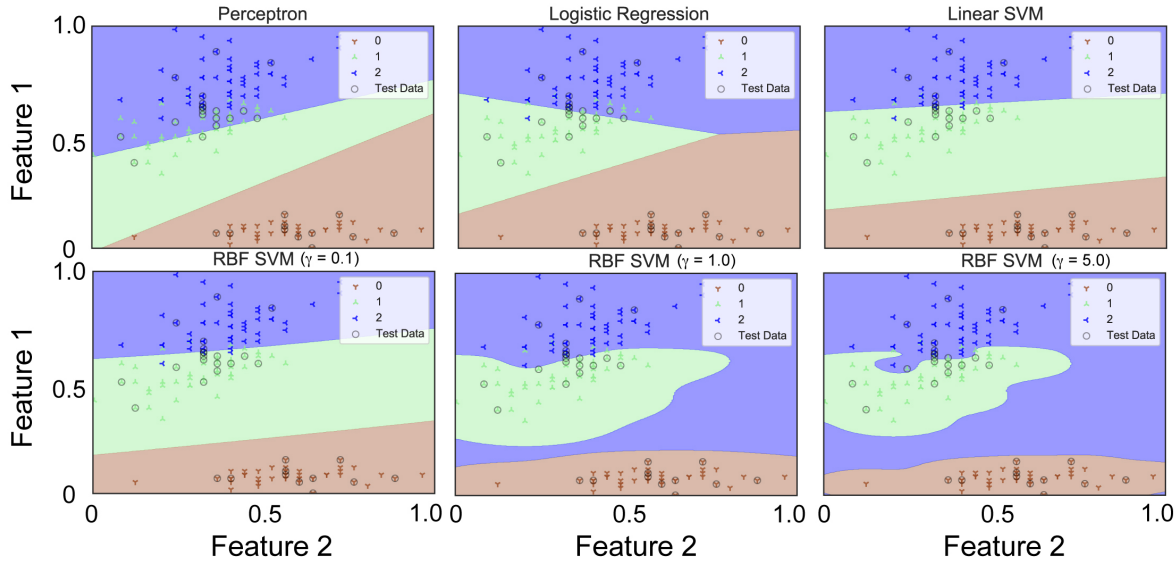


Figure 4.37: Different linear classifiers and hyperparameter configurations to visualize the different decision boundaries caused by the choice of γ. The soft margin parameter C is constant for all plots with $C = 1.0$. Large values for γ may lead to higher accuracies, but also tends to learn unreasonable decision boundaries.

Purely linear classifiers generalize better, whenever the classes are likely to be linearly separable, but the soft-margin SVM even optimizes the decision boundary. The application of the kernel trick (RBF) further improves the classification accuracy, but introduces over-fitting for specific hyperparameter choices of γ and C. For example, the lower middle and right trained SVM classifiers learn not intuitive decision boundaries for class 2 (purple) which unnaturally revolve around class 1 as γ becomes larger than 1. Since higher dimensional feature spaces cannot be visualized, unreasonable and over-fitted classifier hyperparameter settings cannot be spotted by the application designer. These observations lead to the conclusion that nonlinear SVMs should be used only with hyperparameter settings that turn their decision boundaries slightly nonlinear, e.g., γ should be between 0.1 and 1 being precisely quantified by a grid search.

4.3.2 Nearest Neighbor Classifiers

Instead of learning a model or a representation from the training dataset, a classifier can also memorize the training samples directly. More specifically, if the k closest elements of a training observation are considered and the overall class is determined from a majority voting scheme, the corresponding classifier is known as k-Nearest Neighbor (k-NN) [239]. The general concept of distance d between two n-dimensional points $\mathbf{x}$ and $\mathbf{y}$ in the feature space is described by the Minkowski distance

$$d_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{n=1}^{N} |\mathbf{x}[n] - \mathbf{y}[n]|^p \right)^{1/p} \tag{4.74}$$

with p being the order of the distance metric [203]. Although the Minkowski distance metric offers an infinite number of specific metrics by varying the order of the equation, only

three (p = 1, 2, ∞) are of practical importance and commonly used in classification tasks [203]. The 1-norm refers to the Manhattan distance and the 2-norm is known as Euclidean distance. Note that the initial assumption is that there is no cross-axis related distribution of the data, i.e., the covariance matrices $\mathbf{\Sigma}$ have negligible off-diagonal entries and the features are not correlated. In practical cases, however, the more general Mahalanobis distance metric between two points $\mathbf{p}_{1,2}$ (in 2D, otherwise hyperpoints)

$$d_M(\mathbf{p}_1, \mathbf{p}_2|\mathbf{\Sigma}) = \sqrt{(\mathbf{p}_1 - \mathbf{p}_2)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{p}_1 - \mathbf{p}_2)} \tag{4.75}$$

takes such correlations into account. Figure 4.38 shows a 2D example of the two features (MiCO$^{\mathrm{TUM}}$ and MiRO$^{\mathrm{TUM}}$) based on class-level classification ($C_{1...8}$). The orientations and distances between the contour lines of each class depend on the feature standard deviation, correlation, and scattering of the training data points.

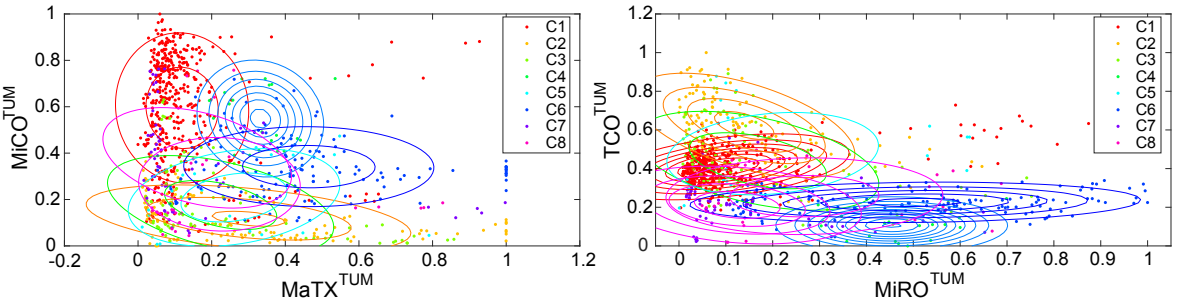


Figure 4.38: Mahalanobis distances between data points from two features. The condition for Euclidean distance is not given, since the distribution scatters more along the x-axis and is rotated for several classes.

By choosing the Mahalanobis distance as underlying distance metric for k-NN, the data scattering is inherently considered. Further advantages of k-NN are its non-parametric definition, i.e., the training procedure does not require to estimate any learning-relevant parameters [230], and its ability to learn non-linear decision boundaries like nonlinear SVMs. Additionally, new training samples can be easily integrated into previously trained k-NN models. However, k-NN is comparably susceptible to the curse of dimensionality, i.e., the more dimensions are considered, the sparser the feature space becomes and the distance between two data point looses its interpretability. Hence, feature selection and feature space reduction schemes are especially important for k-NN.

4.3.3 Probabilistic Models and Naïve Bayes Classifier

Probabilistic models are based on the idea that features and target variables can be modeled as random variables [203]. A probabilistic classifier is able to predict a probability distribution over all classes $\hat{\mathbf{y}}$ given the evidence, i.e., data $\mathbf{X}$. In the following, the hypothesis $\hat{y}$ is the material class label C prediction based on the taxonomy introduced in Section 3.1.2. Each assignment to a specific material class C_k has a particular distribution of the underlying data, which are the numeric feature values of the feature space $\mathbf{X}$ in this thesis. Given

a set of classes $\mathbf{C} = \{1, 2, ..., N\}$, the frequency that class C_k occurs is defined by the probability distribution $P(C_k)$, also called prior distribution and constitutes the class probabilities before evaluating the data (i.e., by random guessing) and determines the global occurrence of the class based on prior knowledge. In this thesis, the a-priori probabilities of the classes C_1 (woods) and C_6 (stones), for example, are larger due to class imbalance and need to be considered in the case of major class or subclass-level classification. Given the evaluation of labeled training data, i.e., $\mathbf{X}$ assigned to specific class labels, the objective is to calculate the posterior class probability $P(C_k|\mathbf{X}) = P(C_k|\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ [240], which shows that an observation falls into class C_k, given the corresponding data, i.e., n features $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, and is known as Bayes' rule

$$P(C_k|\mathbf{X}) = \frac{P(\mathbf{X}|C_k)P(C_k)}{P(\mathbf{X})} = \frac{P(\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n|C_k)P(C_k)}{P(\mathbf{X})} \tag{4.76}$$

The denominator $P(\mathbf{X})$ denotes the probability of evaluating the data independent of the class labels, and is a constant for all classifications, and hence, omitted for further calculations. The numerator $P(\mathbf{X}|C_k)P(C_k)$ is also called joint likelihood of the observation and class $P(\mathbf{X}, C_k) = P(\mathbf{X}|C_k)P(C_k)$. The naïve assumption of feature orthogonality, i.e., feature independence, decomposes the joint likelihoods to marginal likelihoods

$$P(C_k|\mathbf{X}) = P(C_k) \cdot \prod_{i=1}^{n} P(\mathbf{x}_i|C_k) \tag{4.77}$$

The decision rule used to select the class with the largest maximum a posteriori (MAP) probability is

$$\hat{y}_{\text{MAP}} = \underset{c_{k=1,...,N}}{\arg\max}\, P(\mathbf{X}|C = C_k) \cdot P(C = C_k) = \underset{C_{k=1,...,N}}{\arg\max}\, P(C_k) \cdot \prod_{i=1}^{n} P(\mathbf{x}_i|C_k) \tag{4.78}$$

and can be simplified to the maximum likelihood (MLi) decision rule

$$\hat{y}_{\text{MLi}} = \underset{c_{k=1,...,N}}{\arg\max}\, P(\mathbf{X}|C = C_k) = \underset{C_{k=1,...,N}}{\arg\max} \prod_{i=1}^{n} P(\mathbf{x}_i|C_k) \tag{4.79}$$

whenever a uniform prior probability is assumed, i.e., there is no prior knowledge about the occurrence of materials $P(C_k)$. The decision boundary is situated where the two posterior probabilities are equal, e.g. $P(C = C_k|\mathbf{X}) = P(C = C_{k+1}|\mathbf{X})$. For a new data point, i.e., a test set point, the predicted class $\hat{C}$ will be the class with the largest MAP value.

Considering supervised classification and numerical feature values in this thesis, the joint likelihood $P(\mathbf{X}|C_k)$ follows a probability function depending on the features and their corresponding class labels. A Gaussian Naïve Bayes classifier (GNB) assumes that the data is generated by a normal distribution and follows a Gaussian probability density function (PDF) [203], [241]. For example, Fishel et al. [116] also assumed a normal distribution of their scanned data based on repeated physical measurements. It is worth noting that this ideal assumption might be violated for specific features, which is one possible explanation why the Naïve Bayes classifier might achieve a lower accuracy than other classifiers.

4.3.4 Decision Trees

Decision trees (DTs) [233] can be considered as the most intuitive classifiers for supervised tasks. They do not require feature scaling, are non-parametric, and inherently capable of multiclass classification. Notably, their concept follows a similar methodology as humans would categorize a material sample, i.e., by ranking the features according to their (perceptual) relevance and the subsequent decisions according to the most dominant features. However, individual DTs are prone to data variance [203] and are generally outperformed by other classifiers [235], but they provide the base classification concept for the well-performing ensemble classifiers explained in the next Section 4.3.5.

A DT consists of a root node with N children nodes, which likewise either contain further children nodes, or the leaf nodes which are associated to a classification decision given a sample input data instance traversing through the tree. For illustration, a binary tree is considered in the following based on two classes with the labels C_1 and C_2. Given a numeric feature space as in this thesis, the DT is built (i.e., trained) beginning from the root to the leafs using an inequality split decision based on one out of F features in the feature space. The fundamental task for the building algorithm is to identify the most distinctive feature for each node to decrease the impurity of the node, i.e., its uncertainty of assigning a specific class label to the data instances. One generally applied metric is the Gini impurity

$$I_{\mathrm{G}} = 1 - \sum_{c=1}^{C} \mathbf{p}[c]^2 \tag{4.80}$$

which quantifies impurity based on the percentage **p**[c] of all samples correctly assigned to the class c of the current node. For example, if feature F_1 is applied to split the training dataset, 60 observations fall into the left node and 40 to the right node. Within the 60 observations in the left node, 50 actually belong to C_1 and 10 to C_2, i.e., 10 observations are wrongly classified as C_1. Hence, $I_{\mathrm{G,left}}$ equals

$$I_{\mathrm{G,left}} = 1 - \left(\frac{50}{50+10}\right)^2 - \left(\frac{10}{50+10}\right)^2 = 0.278 \tag{4.81}$$

In the right node, 20 out of 40 are classified as C_2, and 20 out of 40 wrongly as C_1, leading to

$$I_{\mathrm{G,right}} = 1 - \left(\frac{20}{20+20}\right)^2 - \left(\frac{20}{20+20}\right)^2 = 0.5 \tag{4.82}$$

The overall split quality is the weighted average of both Gini impurities

$$I_{\mathrm{G}} = \frac{\mathbf{w}[1] \cdot I_{\mathrm{G,left}} + \mathbf{w}[2] \cdot I_{\mathrm{G,right}}}{\mathbf{w}[1] + \mathbf{w}[2]} = 0.389 \tag{4.83}$$

and is lower than the initial impurity of 0.5 of the root node, and hence, a valid split. Note that the weight vector **w**[C] contains equal weights for the DT classifier, however, the AdaBoost classifier in Section 4.3.5 internally adjusts the weights to further influence the split decision. The same procedure is repeated for all subsequent child nodes with their respective

subset of data instances. The algorithm terminates either when the hyperparameter maximum tree depth is violated, or, whenever the I_G of the child is less then the I_G of the parent node. In this case, a child node becomes a leaf node.

A single decision tree, however, is outperformed by improved ensemble classification methods such as the random forest classifier which was evaluated on various datasets in [234], or adaptive boosting classifiers (AdaBoost [242]) explained in the following.

4.3.5 Ensemble Classifier

Ensemble methods aim to combine multiple classifiers $c_1, c_2, ...c_\mathrm{m}$ to form a superior classifier [230], [243]. The development of such combined models comprises the main steps listed in the following.

1. Choice of either different classifiers or the same basic classifier with different configurations
2. Selection of features (columns) and instances (rows) of the feature space
3. Random selection and duplication of instances or different weighting of instances
4. Majority voting of individual classifier results

In the following, the most commonly used ensemble models relevant for this thesis are presented.

4.3.5.1 Bootstrap Aggregating, Bagged Trees, and Random Forests

Bootstrap aggregating (Bagging) is a method that changes the input feature space before a classifier is trained. In detail, a subset of the instances are randomly bootstrapped (i.e., selected) to form a new feature space, which also allows duplicates [244]. If the selected classifier of each bootstrapped feature space is the decision tree from Section 4.3.4, the ensemble method is consequently called bagged trees. The fundamental concept behind this method is to average the classification output of these multiple, presumably highly-variant decision trees on the differently bootstrapped datasets for the sake of a more robust classification result. It has been shown in [244] that the addition of more trees does not over-fit the training dataset, but further increases the robustness of the classifier. Note that bagged trees still contain all columns (i.e., features) of the initial feature space. By contrast, a very popular implementation of bagging is the Random Forest (RF) classifier [234] which uses bootstrapped feature spaces, but further randomly selects f features for each k^th trained decision tree to obtain a reduced feature space $\mathbf{X}^*$ for training each classifier

$$c_\mathrm{k} = train(\mathbf{X}^*, \mathbf{y}) \tag{4.84}$$

The final prediction of a new instance $\mathbf{x}$ (e.g., a test instance) is based on a majority voting of all K tree predictions as

$$\hat{y}_\mathrm{MV} = \begin{cases} 1 & \text{if } \sum_{k=1}^{K} \hat{\mathbf{y}}_\mathbf{x}[k] > 0 \\ -1 & \text{otherwise} \end{cases} \tag{4.85}$$

with

$$\hat{\mathbf{y}}_{\mathbf{x}}[k] = predict(c_{\mathrm{k}}, \mathbf{x}) \tag{4.86}$$

for a binary classification task. Most importantly, this averaging procedure of RFs further mitigates the well-known over-fitting behavior of individual decision trees [234]. Another advantage is the lower number of hyperparameters compared to other complex classifiers like nonlinear SVM, which make RFs an ideal classifier candidate for this thesis.

4.3.5.2 Adaptive Boosting

Another commonly used ensemble method is adaptive boosting, in short, AdaBoost [245] and is considered in this thesis based on its superior performance on different machine learning tasks and its low number of hyperparameters similar to the random forest classifier, which is another striking argument for their selection as classifiers. Boosting also uses the same type of classifier, i.e., decision trees based on random subsets of the dataset. In contrast to RFs, AdaBoost imposes a strict maximum tree depth of 1, i.e., one parent node and two child nodes which is denoted as tree stump. Since decision trees made of only one feature, and hence, one split criterion, generally lead to poor classification performance, the stumps are also known as weak learners. Let $\mathbf{X}$ consist of N instances and F features, i.e., F columns, and $\mathbf{y}$ the corresponding binary label vector, and K different stumps are trained sequentially. The adaptation mechanism originates from the fact that a weight for each observation focuses on the data row in the feature space which was previously misclassified, leading to a weight vector $\mathbf{w}$ containing N entries (i.e., same amount as instances in $\mathbf{X}$). Initially, all weight entries are weighted equally ($\mathbf{w}_{1...\mathrm{N}} = 1/N$). The original AdaBoost M1 algorithm (M1 for binary classification) iterates over all weak learners from k = 1 ... K and repeats the following steps.

1. The k$^{\mathrm{th}}$ tree stump is trained using the decision tree classifier (e.g., based on the Gini impurity metric)

   $$c_{\mathrm{k}} = train(\mathbf{X}, \mathbf{y}, \mathbf{w}) \tag{4.87}$$

   and the prediction vector $\hat{\mathbf{y}}$ is obtained by

   $$\hat{\mathbf{y}} = predict(c_{\mathrm{k}}, \mathbf{X}) \tag{4.88}$$

   and contains N entries. The feature with the largest information gain is used as split criterion for the stump. Note that within the next iterations, the weight vector $\mathbf{w}$ is adjusted by the AdaBoost algorithm and influences the split decision.

2. The weighted classification error

   $$\mathbf{e}[k] = \frac{\sum_{n=1}^{N} \mathbf{w}[n] \cdot I(\mathbf{y}[n] \neq \hat{\mathbf{y}}[n])}{\sum_{n=1}^{N} \mathbf{w}[n]} \tag{4.89}$$

   is calculated in the next step. I is the indicator function leading to 1 if its argument is true, and 0 otherwise. Note that both $\hat{\mathbf{y}}$ and $\mathbf{y}$ are binary vectors of length N either containing the positive or negative class label. The more misclassifications ($\mathbf{y}[n] \neq \hat{\mathbf{y}}[n]$) occur, the more $\mathbf{e}$[k] becomes closer to 1.

3. Next, the weak learner influence factor

$$\boldsymbol{\alpha}[k] = \frac{1}{2} \cdot ln\left(\frac{1 - \mathbf{e}[k]}{\mathbf{e}[k]}\right) \tag{4.90}$$

is calculated. This value represents how much influence the k$^{\text{th}}$ weak learner receives for the later majority voting compared to all other weak learners. If the weighted error is small, than $\boldsymbol{\alpha}[k]$ becomes large indicating that the k$^{\text{th}}$ weak learner performs well. If $\boldsymbol{\alpha}[k]$ becomes negative due to a very high classification error, then the classification output of the k$^{\text{th}}$ weak learner is negated in order to predict the opposite class label.

4. Lastly, each $n \in N$ entry of the weight vector $\mathbf{w}$ is updated by

$$\mathbf{w}[n] = \mathbf{w}[n] \cdot e^{\boldsymbol{\alpha}[k] \cdot I(\mathbf{y}[n] \neq \hat{\mathbf{y}}[n])} \tag{4.91}$$

Consequently, the observations which led to misclassifications $\mathbf{y}[n] \neq \hat{\mathbf{y}}[n]$ are weighted with $\boldsymbol{\alpha}[k]$, and hence, depend on the classification performance of the current weak learner. The next weak learner (k+1) is then supposed to emphasize on these more critical (i.e., previously misclassified) instances instead of those which are already correctly classified; the changed weight vector influences the probability estimates in the Gini impurity calculation for the next weak learner. This step is the reason why AdaBoost learns different tree stumps during the K iterations.

After K tree stumps are trained, i.e. K iterations, the final (scalar) majority voting prediction for a novel observation $\mathbf{x}$ is

$$\hat{y}_{\text{MV}} = \begin{cases} 1 & \text{if } \sum_{k=1}^{K} \boldsymbol{\alpha}[k] \cdot \hat{\mathbf{y}}_{\mathbf{x},k} > 0 \\ -1 & \text{otherwise} \end{cases} \tag{4.92}$$

with

$$\hat{\mathbf{y}}_{\mathbf{x},k} = predict(c_{\text{k}}, \mathbf{x}) \tag{4.93}$$

During these training steps, no further hyperparameters are required. Since many weak learners are trained instead of complex decision rules, AdaBoost is also less likely to overfit, and hence, a very suitable classification algorithm for this thesis.

4.3.6 Classification Results

The following section evaluates the different classifiers on the features spaces used in this thesis.

4.3.6.1 Initial LMT 108 Material Database

During the creation of this thesis, the data collected with the Texplorer version 1.c [7] was used to create the LMT 108 material feature space based on the features from Section 4.1.4. Table 4.11 shows the achieved accuracies for the three best (Compliance, MiRAS, MiRSS),

Classifier	Acc (%) (3 features)	Acc (%) (5 features)	Acc (%) (8 features)
KNN (Mahalanobis)	56.1 ± 0.4	82.3 ± 0.5	84.1 ± 0.4
SVM (RBF)	47.6 ± 0.5	73.2 ± 0.4	79.1 ± 0.5
Gaussian Naïve Bayes	46.2 ± 0.4	71.7 ± 0.3	78.1 ± 0.3
Ensemble - Random Forest	**63.2** ± 0.8	**84.2** ± 0.7	**91.2** ± 0.8
Ensemble - AdaBoost	54.9 ± 0.7	78.8 ± 0.8	86.8 ± 0.8

Table 4.11: Classification results of the initial 108 LMT texture dataset.

five best (additionally: MaRS, MaRiA), and all eight tactile features from the thesis-related work in [7].

These results served as an intermediate result to show that the feature engineering on such a database seemed promising. The relevant extension of introducing the novel taxonomy (Section 3.1.2), the Texplorer2 (Section 3.3), and novel features (Section 4.1.5) led to further improvement of the overall classification task on the more extensive LMT-184-HDB database from Section 3.1.2.

4.3.6.2 LMT-184-HDB Results

The LMT-184 haptic material database is trained and tested based on five-fold partitions using cross-validation. Similarly, the classifiers introduced in Section 4.3 such as support-vector machines (SVMs), k-nearest neighbor (k-NN), neural network (NN), random forest (RF), AdaBoost (AB) and Gaussian Naïve Bayes (GNB) are used. A feed-forward neural network (Matlab Neural Network toolbox) is fed with all features as input and trained using 75% training, 5% validation, and 20% test data based on 500 hidden layers to show a comparison to the other classifiers. All classifiers are evaluated based on three different classification levels (class C, subclass S, material-based M) using the two metrics accuracy and F_1 score.

Class and Subclass-level Results Figures 4.39 and 4.40 show the accuracies (top) and F_1 scores (bottom) for the feature spaces with ten selected features $\mathbf{X}_{10}$ from different domains, e.g., purely image-based features, evaluated on the class- and subclass level. The confusion matrix in Fig. 4.41 further shows the subclass results of the RF classifier and aids to understand which subclasses are confused by the trained model. It becomes evident that the major confusions occur only within the major classes; e.g., softwood and hardwood are confused, but both belong to the major class C_1 (woods).

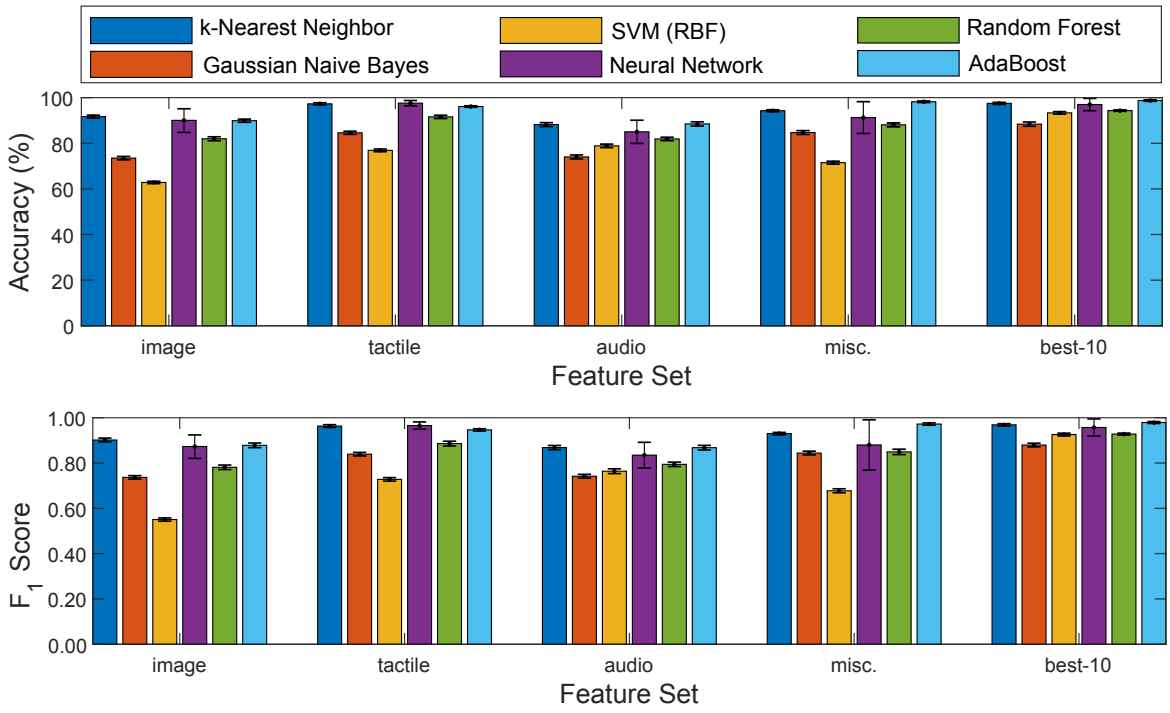


Figure 4.39: Major class C classification results (top: accuracy, bottom: F_1) for different classifiers. The x-axis shows the different sets of features, e.g., a selection of purely image-based features used to train the different classifiers. Note that misc. summarizes features both from related work (Table 4.3) and features going beyond human perception (Table 4.5). The names of the best features of each domain have been identified in Section 4.2.4 and are summarized in Table 4.6 (adapted from [2] © 2019 IEEE).

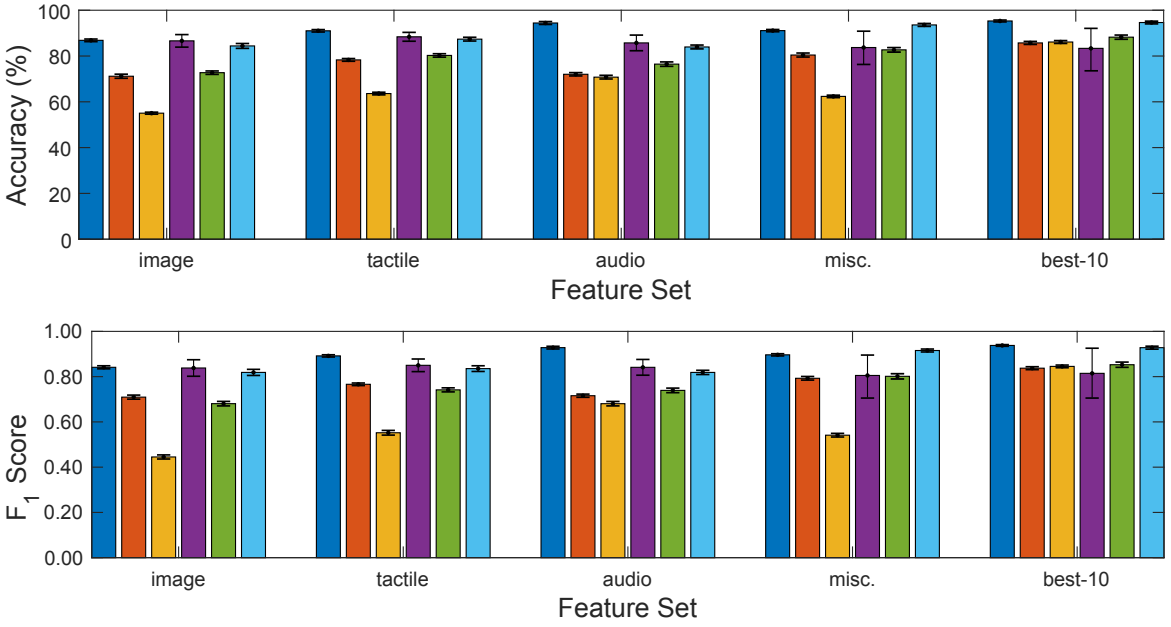


Figure 4.40: Subclass S classification results (top: accuracy, bottom: F_1) for different classifiers. The x-axis shows the different sets of features, e.g., a selection of purely image-based features used to train the different classifiers.

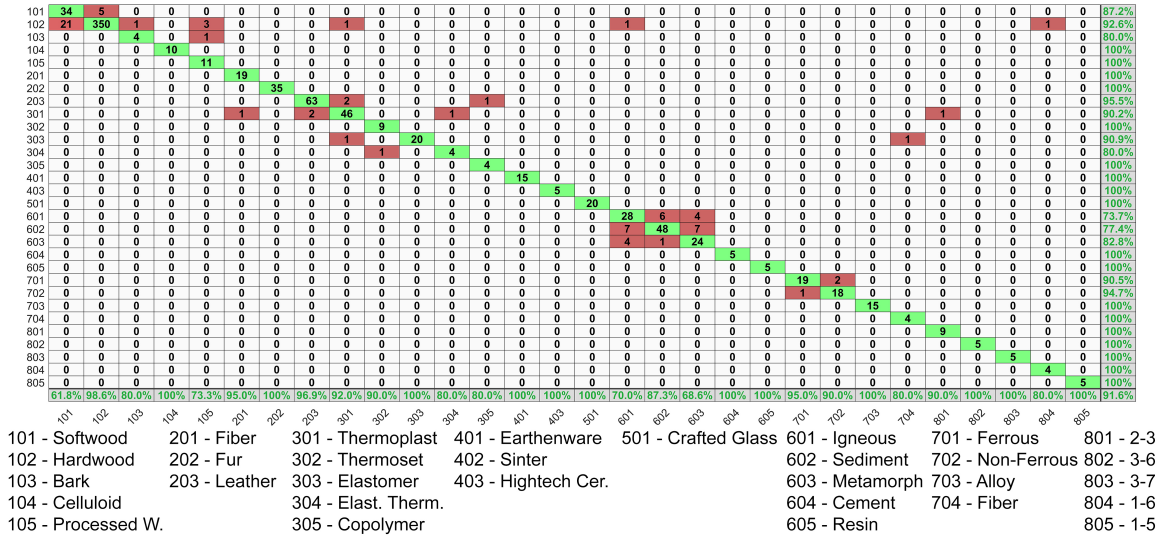


Figure 4.41: Subclass classification confusion matrix based on RF classifier and the ten best features. Note that C_8 (composites) are abbreviated by their combinations of the classes C_1 – C_7. The confusions generally occur within the major classes like C_1 or C_6, i.e., between their subclass labels S_1-S_3, and hence, leading to the conclusion that the selected features inherently are also taxonomy-related.

Material-level Results The identification of the class or subclass-level label helps to show that the proposed features are relevant, yet the core task of a material classification engine is to precisely identify the materials on their material-level base, i.e., tin, copper, iron, etc. If these material identifiers M are used as the true labels $\mathbf{y}$, the classifiers yield the results shown in Fig. 4.42. Note that no analytical proof can be made why a specific classifier exactly

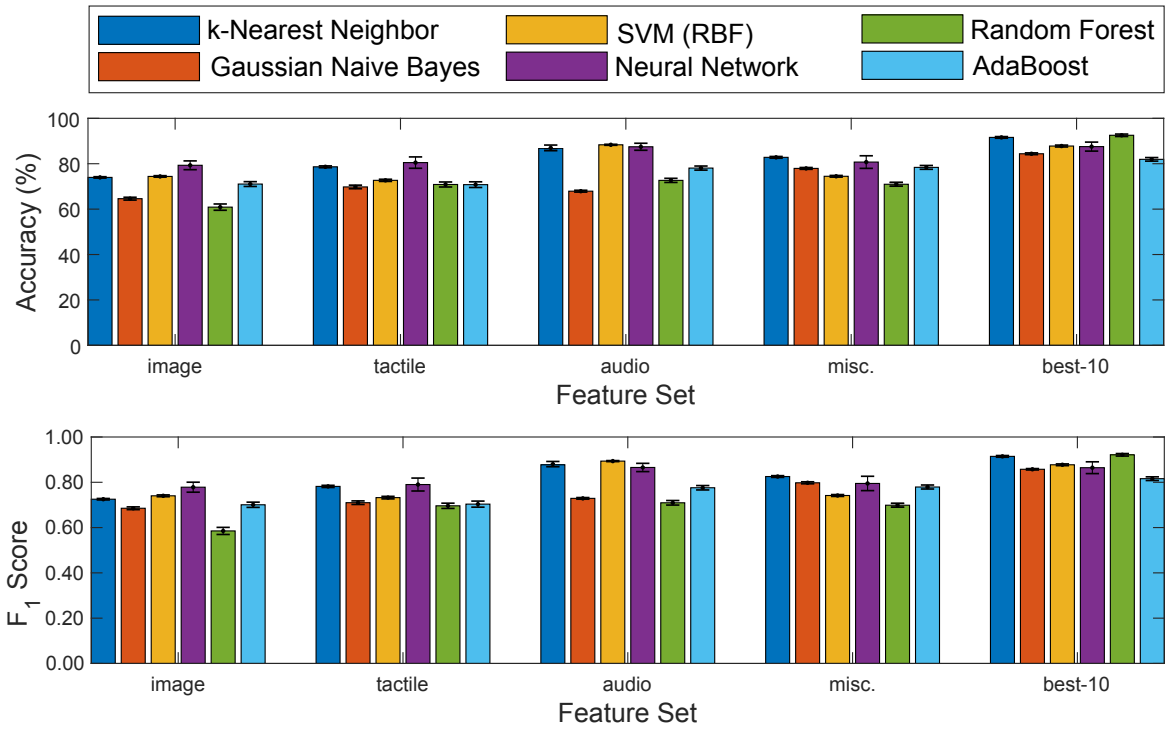


Figure 4.42: Material-level M classification results (top: accuracy, bottom: F_1 score) for different classifiers. The x-axis shows the different sets of features, e.g., a selection of purely image-based features used to train the different classifiers. Note that misc. summarizes features both from related work (Table 4.3) and features going beyond human perception (Table 4.5). The results for individual domains are comparable, however, the selection of the best features from each domain (best-10) lead to a significantly higher classification performance which is a strong indicator that the application of only audio or image features under-fit the overall classification task (adapted from [2] © 2019 IEEE).

outperforms another classifier in general, however, it can be observed that the results for the ten best features selected from all domains generally outperform the individual modalities such as the ten best audio features. However, all classifiers show predictive capabilities, and consequently, the combination of the individual classifier outputs can form a superior **majority voting**-based (MV) classifier using the posterior probabilities of the three best classifiers (RF, KNN, and AdaBoost), which votes for the most frequent class as classification output. If no consensus is found, the voting of the best individual classifier is selected. Figure 4.43 compares the KNN, the RF, and the MV classifier for different numbers of the n-best features across all domains. Comparable to the votings of experts in a real-life situation, the overall voting of the MV classifier outperforms all ordinary classifiers and achieves 91.0 ± 0.2 % accuracy on the testing set for 10 features.

It is a common technique to further augment the feature space with additional observa-

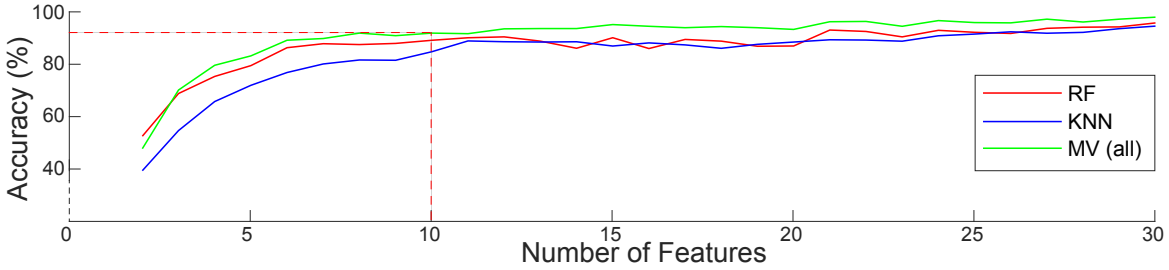


Figure 4.43: Comparison of the KNN, RF, and MV classifiers depending on the number of best features. The MV approach already achieves a high accuracy based on five features only. Note that all classifiers do not significantly increase their performance with more than ten best features; also more features increase the curse of dimensionality to which classifiers like KNN are susceptible.

tions, for example, generated from copies of the original feature values overlaid by adding noisy values drawn from the standard normal distribution [246]. Each feature value in a copied observation is multiplied by the factor $(1 + l \cdot r)$ with r being drawn from the standard normal distribution and l being a scaling factor between 0.0 and 1.0 times the normal standard deviation. Figure 4.44 shows the achieved accuracy, depending on l, if a five-fold cross-validation is applied.

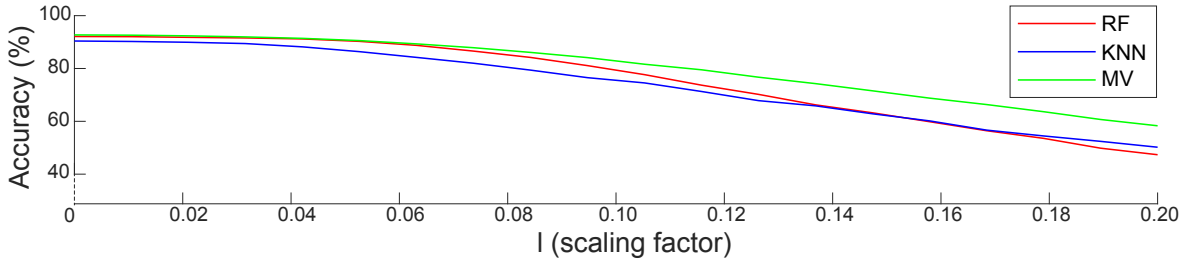


Figure 4.44: Feature space augmentation with added noisy observations. Comparably to Fig. 4.43 the majority-voting-based approach still outperforms all individual classifiers. Further note that the MV approach leads to slightly more robust accuracies in the presence of noisy feature values.

4.4 Subjective Experiment

During the creation of this thesis and the evaluation of the Texplorer system, the fundamental question arose whether features based on human ratings should be preferred to sensorized systems. The SynTouch feature space allows for the application of an n-point Likert scale [247] to quantify the tactile dimensions based on such human ratings. Consequently, a corresponding experiment with 20 subjects between the age of 18 and 62 was performed in this thesis. The design of such a subjective experiment follows the concept of the six exploratory procedures. Note that lifting and enclosure lead to the kinesthetic dimensions weight and volume, and contour following and sliding are combined into a single examination step. Consequently, the four exploratory procedures (EPs) static touch, sliding, contour following and pressure were examined for a representative subset of material samples originating from the LMT 184 database, denoted as LMT-24-HDB, which is visualized in Fig. 4.45 by its cor-

responding surface images. Three materials per major major class C based on the taxonomy from Section 3.1.2 were selected to cover a highly representative range of material samples.

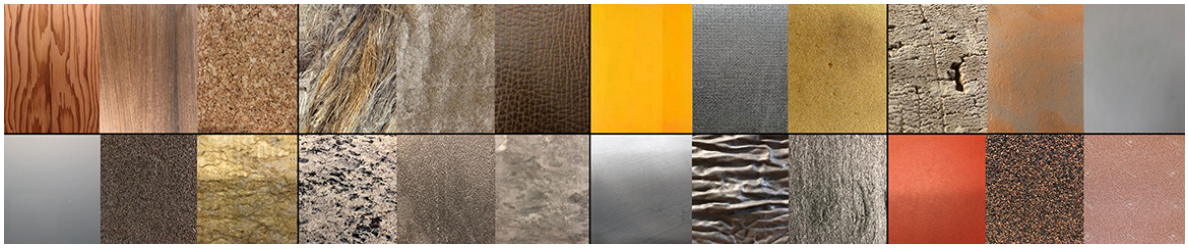


Figure 4.45: Selection of 24 materials from the LMT-184-HDB, denoted as LMT-24-HDB, which has been used for the subjective experiment. Three materials per major class C (separated by black lines) have been selected: C_1: Larch, Oak, Cork, C_2: Hemp, Sheep Wool, Bison Leather, C_3: PMMA, Isoprene, Polyurethane, C_4: Refractory Material, Terracotta, Ceramic Tile, C_5: Float Glass, Foam Glass, Glass Wool, C_6: Granite, Clay Shale, Marble, C_7: Stainless Steel, Lead, Steel Wool, C_8: Linoleum, Tar Paper, Sandpaper.

4.4.1 Experimental Setup and Procedure

Each subject sat in front of a table with the 24 sliced material samples randomly placed. The subject were not blindfolded due to prior test run observations of the experiment supervisor. The feasibility of such an experiment absent visual information would drastically influence the results since the subjects would have to memorize the tactile impressions for the extensive number tactile subdimensions in this study. It was also observed beforehand that judgments about visual properties such as regularity were reconsidered if a rating about perceived haptic regularity was asked for. This was also observed later for the subjects which mentioned that their prior knowledge about the material and its visual appearance did not influence their haptic rating.

After touching and inspecting every material sample, the 15 tactile subdimensions were explained thoroughly, which then had to be rated from 1 - 5 following a 5-point Likert scale, and the supervisor of the experiment filled the verbal ratings into an excel sheet for later processing. Note that it was identified beforehand that broader scales, e.g., 7-point-Likert scales, were too challenging to quantify consistently even for the experiment designer. The subjects were given the original explanations of the different tactile dimensions originating from [66] and listed in Section 2.1.2. Several tactile dimensions are clearly associated to specific EPs in order to facilitate the experimental procedure as shown in Table 4.12.

The thickness of material samples potentially influences the perception of material hardness. Consequently, the experiment phase Pressing contains two modes of operation. First, the material samples had to be assessed while lying on a table (no asterisk), and second, they had to be rated during free, unconstrained interaction, marked with asterisks *, leading to ten hardness-related dimension ratings, and overall, 20 tactile dimension ratings. Note that the introduction of unconstrained exploration increases the variance introduced by different subjects, but likewise extends the descriptiveness of the tactile features compared to, e.g., robotic-based scan systems. As discussed in Section 2.2.2, subjects can easily identify if a

Table 4.12: SynTouch subdimensions categorized into EPs for the subjective experiment. The samples were placed and explored on a planar table. The asterisks for the phase pressing indicate that, additionally, free interaction with the material was allowed during the phase pressing which led to two different feature vectors for this EP.

#	Phase Sliding/Contour Following	Phase Static Touch	Phase Pressing
1	**Macrotexture** (mTX)	**Macrotexture Coarseness** (mCO)	**Tactile Compliance*** (cCM*)
2	**Microtexture Roughness** (uRO)	**Macrotexture Regularity** (mRG)	**Local Deformation*** (cDF*)
3	**Microtexture Coarseness** (uCO)	**Thermal Cooling** (tCO)	**Damping*** (cDP*)
4	**Tactile Stiction** (fST)	**Thermal Persistence** (tPR)	**Relaxation*** (cRX*)
5	**Sliding Resistance** (fRS)	**Adhesive Tack** (aTK)	**Yielding*** (cYD*)

material sample is thin enough to be lifted and fold; notably these actions lead to completely different values for features like local deform (cDF vs. cDF*). Also, the feature yielding cYD is very limited when a soft material sample is placed on a table. However, if free interaction is allowed, cYD* is used to express how likely the material sample stays deformed after applying reasonable human-material interaction forces. Notably foams or pieces of fabrics will stay deformed after arbitrary interactions which cannot be captured by any other tactile feature, and hence, is another independent data source.

4.4.2 Experimental Results

The subjective ratings are range-normalized between 0 and 1. The mean opinion score (MOS) feature vector $\mathbf{m}$ across all subjects and materials is calculated and represents an averaged feature value for each tactile dimension. First of all, it is relevant to identify the outliers in the dataset, i.e., the subjects which deviated most from the average opinion. The Spearman rank correlation coefficient (SRCC) between each subjects' 20-by-1 tactile dimension feature vector (i.e., the individual ratings for each feature) and the MOS feature vector are calculated. Note that this step is performed for all 24 material classes, leading to 24 SRCC values per subject. Their mean and standard deviations are visualized as error bars in Fig. 4.46.

Figure 4.46 further shows that the average correlation to the MOS is 0.75 ± 0.08. The subjects did not consistently rated the tactile dimensions alike due to potential outliers. The z-score [248] is a first calculation step to remove outliers, notably values beyond two-times the standard normal deviation should be excluded, which is, however, not the case even for subject 15 with a z-score of about 1.96. As the next step, the two-sample Kolmogorov-Smirnov (KS) test [249] is performed on each pairwise comparison between the MOS mean vector and each individual subject's rating mean vector to check if both underlying distributions are equal. The result hypothesis h equals 1 if the test rejects the null hypothesis at the significance level α, which is set to 1%, and 0 otherwise. Table 4.13 shows the test result values of the subjects whose ratings led to KS test rejection (h = 1), and which are hence

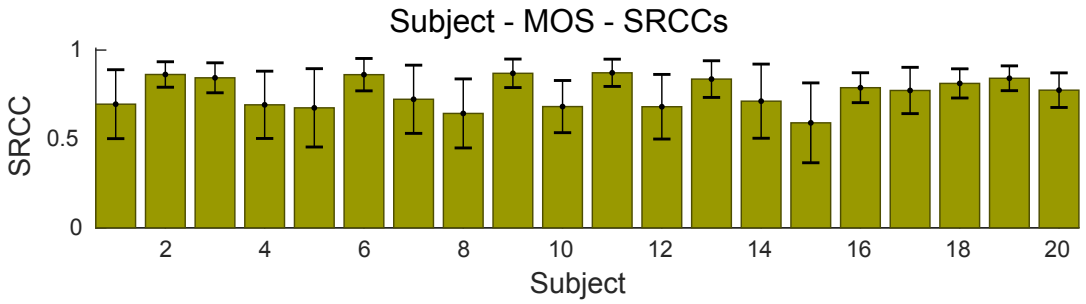


Figure 4.46: SRCCs of each individual subject's rating to the mean opinion score, i.e., overall mean rating. Note that, e.g., the ratings of subject 15 have a low correlation to the mean and also large variance across the individual tactile dimensions (black error bars).

considered as outliers.

Table 4.13: Subjects that did not pass the KS test (h = 1) on equal distribution to the MOS distribution. These seven subjects are considered as outliers and their ratings are not further used.

Subject ID	test hypothesis h	p-value	KS test value
1	0	0.4442	0.250
2	1	0.0027	0.525
3	1	0.0018	0.541
4	0	0.6291	0.216
5	0	0.0694	0.375
6	1	0.0022	0.533
7	0	0.2614	0.291
8	0	0.0158	0.450
9	1	0.0009	0.566
10	0	0.0222	0.433
11	1	0.0002	0.616
12	0	0.1602	0.325
13	1	0.0018	0.541
14	0	0.4880	0.241
15	1	0.0050	0.500
16	0	0.7257	0.200
17	0	0.5809	0.225
18	0	0.0805	0.366
19	0	0.0597	0.383
20	0	0.8167	0.183

The ten best subject ratings, i.e., these being closest to the MOS, were considered for further evaluation using ten-fold cross-validation. Comparable to the Texplorer2-based feature space made of the 15 TacTUM features (Section 4.1.5), a subjective feature space is created based on the individual ratings as instances (rows) and the tactile dimensions as features (columns), leading to a 240-by-21 feature space $\mathbf{X}_{\text{subj}}$ as shown in Fig. 4.47.

Note that the true labels are appended as last column (column 21) containing the numerical class labels ranging from 1 to 24. This step investigates if human ratings are more suitable for material classification than sensor-based features. The same classifiers from Section 4.3

	mTX	uRO	uCO	fST	fRS	mCO	mRG	. . . . .	cRX	cRX*	cYD	cYD*	Label
Subj1 ⋮ Subj10	0.25 ⋮ 0.00	⋮											1 ⋮ 1
Subj1 ⋮ Subj10	⋮												2 ⋮ 2
⋮													⋮
Subj1 ⋮ Subj10													24 ⋮ 24

Figure 4.47: Illustration of subjective feature space created from the best ten subjects ratings. The labels are consecutively labeled according to the 24 materials from Fig. 4.45.

are selected to make the accuracies of the subjective experiment comparable to the Texplorer2 approach. Note that the comparability is decreased by the fact that only 24 materials were rated by the subjects; the data collection of 184 materials during a subjective experiment is far more time-consuming, and possibly even infeasible, than the Texplorer2 data acquisition and already emphasizes one of the critical drawbacks of human voting-based data acquisition. Table 4.14 shows the accuracies based on the best three, ten, and all subjective tactile feature votings which have been identified using the same feature selection techniques from Section 4.2.4. The ten most distinctive subjective features are MiRO^{TUM}, TCO^{TUM}, $\text{CYD}^{\text{TUM}*}$, $\text{CCP}^{\text{TUM}*}$, FST^{TUM}, MaTX^{TUM}, MaRG^{TUM}, MiCO^{TUM}, FRS^{TUM}, and MaCO^{TUM}.

Table 4.14: Results of the subjective experiment with the 24 materials of the LMT-24-HDB. Compared to the Texplorer2-related features calculated and evaluated in Section 4.3, the voting-based classification performs significantly worse for a much smaller dataset (24 vs. 184 materials), supporting the initial assumption that a sensorized system outperforms human-made material votings.

ML Approach	Acc. (%) (3 best features)	Acc. (%) (10 best features)	Acc. (%) (all 20 features)
SVM	42.1 ± 0.4	71.3 ± 0.5	62.0 ± 0.4
AdaBoost	52.1 ± 0.3	72.5 ± 0.3	77.5 ± 0.4
k-NN	49.1 ± 0.4	75.7 ± 0.4	75.0 ± 0.4
Neural Network	53.8 ± 0.4	77.1 ± 0.4	81.4 ± 0.3
GNB	42.3 ± 0.8	61.3 ± 0.8	53.7 ± 0.9
Random Forest	51.3 ± 0.7	72.3 ± 0.7	75.6 ± 0.8

These results show that the subjective features for only 24 materials already lead to lower accuracies across all classifiers compared to the sensorized classification results. Note that the inclusion of all subjects, i.e., with the outliers, performs worse throughout the classifier results. Figure 4.48 further illustrates the resulting accuracies for the RF, k-NN, and MV classifiers depending on the number of used subjective features.

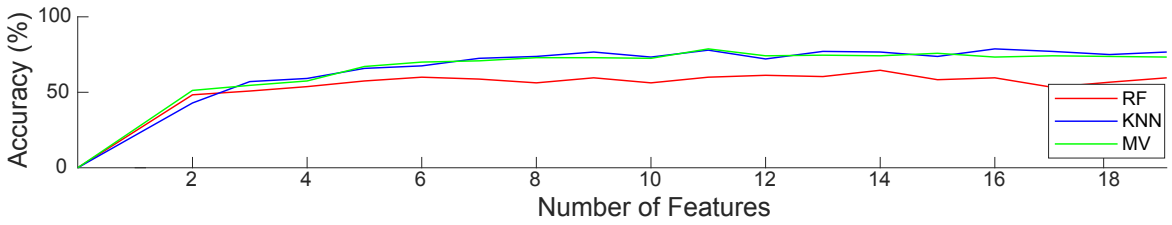


Figure 4.48: Comparison of different classifiers on the subjective ratings-based tactile feature space. Similar to Fig. 4.43 the accuracies saturate if ten or more best subjective rating features are used.

4.4.3 Discussion

The majority of the subjective tactile dimension feature values reveal noticeable variations for a small set of 24 materials with considerable experimental time effort. The feasibility of such an approach beyond 24 materials rapidly becomes impractical and also prone to subjective interpretations of the material samples and the tactile features. As final observation of this subjective experiment, the data acquisition based on human ratings has been found to be inferior to a sensor-based data collection procedure such as the proposed Texplorer2 approach.

4.5 Deep Learning-based Material Classification

An alternative way of generating features in the context of material classification is based on the training of neural networks which directly learn hidden, i.e., inherent, features from the raw input data. In the following the relevant components are briefly introduced.

4.5.1 Background

With advances in computer systems, notably in terms of storage capacities and GPU performance, the underlying concepts of the linear classifiers have been applied to define deep learning (DL) algorithms. In this context, a neural network (NN) is created by arranging multiple neurons in layers of perceptrons [250], known as multilayer perceptron (MLP) model and further introduces non-linear operations such as linear rectifying units and max-pooling, which are explained in the following sections. Combined with an input layer, intermediate (hidden, or dense) layers in between, and an output layer containing the material classes, a network of such neurons forms a feed-forward neural network (FNN). This concept can be further improved for spatial data, i.e., images, by introducing convolutional layers to form convolutional neural networks (CNNs). Similar to handcrafted image filters used in image processing (e.g., Gabor or Schmidt filters [205]), the network is trained to learn its own set of image filters defined by the input data. CNNs have been first successfully applied for handwritten number classification by LeCun et al. [251]. Krizhevsky et al. [252] further have proven the success of using CNNs for image-based object recognition, and ever since, several adaptations like VGGM by Cimpoi et al. [253] or VGGVD [254] have been proposed to also classify surface textures. Bello et al. [255] extensively compared deep networks with handcrafted image features. Deep networks outperform other approaches in terms of texture classification whenever non-stationary textures and the presence of multiple acquisition condition changes is considered. By contrast, handcrafted features are still superior in distinguishing stationary textures under steady imaging conditions and have proven to be more robust than CNN-based features for variation in the image rotation [255]. Due to the rapid progress of deep learning, a comparison of FNNs and CNNs to the handcrafted feature approach for material classification is inevitable in this thesis. In contrast to speech recognition, however, the analysis about the temporal evolution of material-related data traces is not required, and hence, recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) are not further examined.

4.5.2 Network Design Components

The overall design process of FNNs and CNNs comprises several trade-offs and cannot be determined analytically [250]. However, several techniques and heuristics have been developed which aid to impose a specific behavior on parts of the network and its individual components as well as to reduce over-fitting. Consequently, several components are considered in the following in order to understand, why and how the network is optimized.

4.5.2.1 Activation Functions

One fundamental aspect for the success of neural networks is reasoned by non-linear activation functions (AFs), which are applied on the weighted neuron input

$$y = AF(\mathbf{w}^T \cdot \mathbf{x} + w_0) \tag{4.94}$$

Various AFs have been proposed and summarized in [256]. The currently most frequently used ones are listed in the following.

1. The sigmoid AF allows for a soft transition and limit-ranges any input value between 0 and 1.
$$y = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{4.95}$$
However, the gradients of this function at very high or low input values become very small, and hence, slow down the learning process [250].

2. Rectified linear units (ReLUs) [257] avoid small gradient values, and hence, they improve the learning rate for very small or large input arguments compared to Eq. (4.95).
$$y = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{4.96}$$
Whenever the input is negative, the neuron stops learning entirely. Nevertheless, this AF has been used frequently due to its trivial derivative and computational speed [252], [258].

3. Leaky ReLUs [259] are similar to ReLUs, but introduce another parameter α to avoid the problem of vanishing ReLU gradients during training:
$$y = \begin{cases} x \text{ if } x \geq 0 \\ \alpha \cdot x \text{ otherwise} \end{cases} \tag{4.97}$$
with $\alpha = 0.01$, i.e., 1% of the original input data slope. Comparable results to ReLUs have been reported, but gradients do not stop learning as for ReLUs. Consequently, leaky ReLUs presumably are a better choice for the task at hand which was further tested during the creation of this thesis.

4. The softmax AF is generally applied on the last layer in classification tasks [250] to transform the output values $\mathbf{y}$ into class probabilities, and is calculated as:
$$\mathbf{y}[j] = \frac{e^{\mathbf{x}[j]}}{\sum_{k=1}^{K} e^{\mathbf{x}[k]}} \tag{4.98}$$
and ensures that the sum of each output neuron $\mathbf{y}[j]$ is normalized with respect to the overall layer sum of all neurons.

Despite more complex AFs having been published recently [256], the majority of all successful network architectures like AlexNet, GoogleNet, ResNet, VGGNet use ReLU layers for hidden neurons and softmax as output AF, but in order to avoid the learning stop of ReLUs due to a negative input value, leaky ReLUs are preferable in the context of this thesis.

AF Conclusion for this thesis: Leaky ReLU layers are used in hidden neurons, and softmax activation for the classification layer.

4.5.2.2 Cost Functions

A network is trained by changing the network weights $\mathbf{w}$ and biases $\mathbf{w}_0$ with the purpose that the network's predicted output $\hat{\mathbf{y}}$ approximates the true label array $\mathbf{y}$. In the following, the network connection weights and biases are summarized into the single vector $\mathbf{w}$. Different cost functions have been inferred to quantify the achievement of this objective. For example, the quadratic cost function mean squared error MSE is defined as

$$C(\mathbf{w}) = \frac{1}{2 \cdot T} \cdot \sum_{t=1}^{T} |\mathbf{y}[t] - \hat{\mathbf{y}}[t]|^2 \tag{4.99}$$

over all training inputs T. The MSE cost function, however, suffers from slower learning progress rates whenever softmax and sigmoid AFs are used due to saturation caused by very large values. Consequently, the cross-entropy cost function

$$C(\mathbf{w}) = -\frac{1}{T} \cdot \sum_{t=1}^{T} (\mathbf{y}[t] \cdot ln(\hat{\mathbf{y}}[t]) + (1 - \mathbf{y}[t]) \cdot (1 - ln(\hat{\mathbf{y}}[t])) \tag{4.100}$$

was specifically designed to remove the derivative of, e.g., the sigmoid function in the partial derivatives of C($\mathbf{w}$) with respect to the weights $\mathbf{w}$ [250]. It further penalizes wrong predictions with larger prediction certainty, is generally used in comparable classification tasks as cost function, and hence, likewise used in this thesis.

The overall goal of training a neural network is to minimize the cost function $C(\mathbf{w})$ using a gradient descend (GD) approach. Since the weights $\mathbf{w}$ and biases $\mathbf{b}$ may have millions of entries, the gradient of the cost function cannot be calculated analytically. However, since the predicted and true labels are known depending on a given $\mathbf{w}$ at each iteration step, $\mathbf{w}$ is adjusted backwardly beginning from the output layer, which is denoted as backpropagation [260]. The cost function difference of the currently predicted and true label entries also determines the corresponding direction and magnitude on how to change the gradient based on the partial derivatives

$$\nabla C(\mathbf{w}) = \left(\frac{\delta C}{\delta w_1}, \frac{\delta C}{\delta w_2}, \ldots, \frac{\delta C}{\delta w_N} \right) \tag{4.101}$$

The next iteration $i+1$ of updated weights and biases $\mathbf{w}_{i+1}$ is then obtained by a stepwise multiplication with a small value, the learning rate η, leading to

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \eta \cdot \nabla C \tag{4.102}$$

with η being set to 0.01 as default learning rate value [250], which, however, is generally adjusted during training to improve the learning procedure.

Instead of calculating the true gradient over all T training inputs in each iteration, the stochastic gradient descent (SGD) randomly picks M training samples $T_1, T_2, ..., T_M$ (generally known as hyperparameter mini-batch size M) and calculates an approximated gradient as

$$\nabla C(\mathbf{w}) \approx \frac{1}{M} \sum_{m=1}^{M} \nabla C_{T_m}(\mathbf{w}) \tag{4.103}$$

Altogether, the next iteration using SGD can be written as

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \frac{\eta}{M} \cdot \sum_{m=1}^{M} \nabla C_{T_m}(\mathbf{w}) \tag{4.104}$$

Note that too small mini-batch sizes cause significant errors in the gradient calculations, especially whenever many classes as in this thesis are considered. On the other hand, too large mini-batch sizes will cause the update frequency to decrease, and hence, perform the learning in too few steps, and potentially, not reaching a minimum [250]. Consequently, the mini-batch size is adjusted to the closest power of two value of the number of classes in the dataset, i.e., 128 or 256 in the case of the LMT-184-HDB. Thereby, the majorities of labels is present on average in a mini-batch.

To decrease possible learning instabilities of a neural network due to different feature value range and to speed up the learning process, batch normalization (BN) [261] normalizes each output element $\mathbf{x}[i]$ of a previous activation layer by subtracting the batch mean μ_M and by dividing the batch standard deviation σ_M similar to the feature space scaling in Section 4.2.3.3 using

$$\mathbf{x}[i] = \frac{\mathbf{x}[i] - \mu_M}{\sqrt{\sigma_M^2 + \epsilon}} \tag{4.105}$$

Two hyperparameters shift β and scale γ are introduced in [261] to each layer to further control the range and shift of the normalized entry with

$$\mathbf{x}[i] = \gamma \cdot \mathbf{x}[i] + \beta \tag{4.106}$$

The additional effort of saving these hyperparameters is acceptable for an offline training algorithm compared to the higher accuracies and more stable learning as reported in comparable classification tasks, and hence, is applied in this thesis for all networks.

Cost function conclusion for this thesis: The cross-entropy loss function is employed using the back-propagation algorithm based on mini-batch SGD.

Mini-batch size conclusion for this thesis: The choice of the mini-batch size is based on the number of classes to represent the majority class labels in each iteration on average, and batch normalization is applied on each layer.

4.5.2.3 Gradient Descent Learning Options and Optimizers

The learning rate η does not necessarily need to stay constant during the learning iterations. After training a network with an initial value for η which successfully was employed for comparable networks, it is intuitive to either use per-epoch-decreasing or adaptive values. Moreover, SGD can also get stuck on saddle points, i.e., by having a local minimum in one dimension, but not in an other dimension [250]. One approach to improve SGD is based on second-order derivatives (Hessian matrix) which considers the change of the gradient. This technique is computationally very expensive and currently infeasible for larger datasets like the one used in this thesis.

An alternative to the second-order derivatives are momentum-based techniques which accumulate previous gradients to increase the learning in the direction of the steepest descend as

$$\mathbf{v}_{i+1} = \mu \cdot \mathbf{v}_i - \frac{\eta}{M} \cdot \sum_{m=1}^{M} \nabla C_{T_m}(\mathbf{w}) \tag{4.107}$$

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mathbf{v}_{i+1} \tag{4.108}$$

with μ being the momentum coefficient. This hyperparameter can be considered as the amount of friction introduced into the update rule [250]. However, the minimum of the cost function potentially can overshoot the minimum using pure momentum without a deceleration term. Nesterov et al. [262] added such a deceleration condition which led to more stable convergence when approaching the local optimum and is denoted as SGD with Nesterov momentum (SGDM).

$$\mathbf{v}_{i+1} = \mu \cdot \mathbf{v}_i - \frac{\eta}{M} \cdot \sum_{m=1}^{M} \nabla C_{T_m}(\mathbf{w} - \gamma \cdot \mathbf{v}_{i-1}) \tag{4.109}$$

The challenge of choosing a learning rate remains for SGD with Nesterov momentum. Several approaches have been introduced to infer a learning rate from the gradient ever since. The adaptive momentum algorithm, denoted as Adam [263] computes individual learning rates during each iteration. Instead of requiring an initial learning rate λ, the moving mean

$$\mathbf{m}_i = \frac{\beta_1 \cdot \mathbf{m}_{i-1} + (1 - \beta_1)\nabla C(\mathbf{w})}{1 - \beta_1} \tag{4.110}$$

and the moving variance

$$\mathbf{v}_i = \frac{\beta_2 \cdot \mathbf{v}_{i-1} + (1 - \beta_2)\nabla C(\mathbf{w})^2}{1 - \beta_2} \tag{4.111}$$

of the gradient are inferred from previous calculations. β_1 and β_2 are constants with empirical values of 0.9 and 0.999 originating from [263]. Eventually, the weight is updated as

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \Delta i \cdot \frac{\mathbf{m}_i}{\sqrt{\mathbf{v}_i} + \epsilon} \tag{4.112}$$

with the step size Δi and a small number $\epsilon \approx 10^{-8}$ to avoid division by zero.

However, Wilson et al. [264] report that Adam can face convergence issues and does not always generalize as good as SGDM. It has been shown that SGDM can solve a popular classification task (image classification on CIFAR-10 dataset) with notably higher accuracy than Adam or any other adaptive algorithm [264]. Consequently, optimizers like Adam and SGD with Nesterov momentum need to be compared in this thesis.

Optimizer conclusion for this thesis: The comparison of SGDM and Adam needs to be performed. Evaluations on the MNIST dataset (which can be considered as the base line for DL approaches) done by the authors of Adam [263] also revealed that Adam only performs slightly better than SGDM. Furthermore, the Adam algorithm has shown to generally outperform other previously defined adaptive optimizers such as Adagrad or AdaDelta [263].

4.5.2.4 Reduction of Over-fitting

It is well-known that neural networks, and especially CNNs, are prone to over-fitting [252]. Several techniques have been proposed to reduce the effect of such over-adaptation to the training data and are listed in the following.

Data Augmentation One effective approach to combat over-fitting is to enhance the training set with more training instances to learn from. A set of transformations such as rotations, flips, or, intensity variations heavily increase the amount of available data and was one of the key steps in [252] to achieve remarkable results on the ImageNet dataset.

Regularization - L1 and L2 As in traditional machine learning approaches, low complexity systems tend to generalize better than complex models. For example, a linearly or quadratically fitted curve through a set of two-dimensional points is typically less prone to over-fitting than a higher order polynomial, and especially more resistant to noise in the training data. To equivalently account for this phenomenon in FNNs and CNNs [250], the cost function C from the previous section can be regularized by punishing larger weights $\mathbf{w}$ and to enforce more weights being set to zero (L1 regularization) or close to zero (L2 regularization) using

$$C_{\text{reg}} = C_0(\mathbf{w}) + \frac{\lambda_{\text{L1}}}{T} \sum_{o=1}^{W} |\mathbf{w}[o]| + \frac{\lambda_{\text{L2}}}{2 \cdot T} \sum_{o=1}^{W} \mathbf{w}[o]^2 \tag{4.113}$$

with $C_0(\mathbf{w})$ being the unregularized cost function (i.e., cross-entropy in this thesis), T the size of the training set, and W the number of overall weights in $\mathbf{w}$. Since the cost function needs to be minimized, a tradeoff between the original cost function C_0 and small weights is enforced. The hyperparameters $\lambda_{\text{L1,L2}}$ adjust the influence of both regularizers. More specifically, the L1 regularization enforces sparse weight vectors with a few important connections only, whereas the L2 regularization enforces small values in the weight vector. Note that generally only one of the two regularizers is used, namely, the L2 regularizer. Large weights are only learned when they considerably improve C_{reg}. The learning rule for gradient descent using only L2 regularization is

$$\mathbf{w}_{i+1} = \mathbf{w}_i - \mathbf{w}_i \cdot \frac{\eta \cdot \lambda_{\text{L2}}}{T} \cdot -\frac{\eta}{M} \cdot \sum_{m=1}^{M} \nabla C_{T_m}(\mathbf{w}) \tag{4.114}$$

Dropout Layer Dropout layers randomly remove neurons for each new training mini-batch to further reduce over-fitting [265]. The intuition can be compared to the random forest classifier by learning multiple, weaker networks that ultimately form a robust network. Instead of learning one model, a set of submodels with different features and an averaged result lead to less variance in the classification process. This technique has proven to actually reduce over-fitting CNNs [265].

Over-fitting reduction conclusion for this thesis: The network input images are subject to data augmentations such as rotations. Additionally, 50% dropout and L2 regularization ($\lambda_{\text{L2}} = 0.001$) are applied to further decrease potential over-fitting.

4.5.2.5 Network Depth

Another fundamental parameter is the actual number of layers, i.e., the network depth, which is determined by the classification task at hand and increases with higher task complexity. For example, Simonyan et al. [254] extended AlexNet [252] by adding more and smaller 3-by-3 convolutional layers, up to 19 in total. They follow the argumentation of decreasing the number of training parameters, i.e., less parameters per kernel, and to increase the network's capability of learning more complex image patterns by introducing more non-linear ReLU functions. However, by multiplying the layers with a number between 0.0 and 1.0 too many times, the gradient becomes too small when reaching the earlier layers after several applications of the chain rule, which also leads to the vanishing gradient problem [266]. In this context, a very relevant consideration intuitively reduces the total number of hidden layers necessary for the thesis-related DL task. Surface patterns are, in any case, less complex than images of common objects from databases like ImageNet. Each hidden layer conceptually learns a family of patterns, e.g., the first hidden layer learns edges and corners, the second layer simple geometric patterns, and so forth. Since textural structure elements are comparable to geometric elements, e.g., hexagons representing foam-like structures, it is expected that few layers are sufficient to identify such patterns.

Network depth conclusion for this thesis: Less layers lead to better generalization and are sufficient to capture the required geometrical shapes determined by the given task of material classification.

4.5.3 DL-based Texture Classification of LMT 184 Haptic Database

Liu et al. [205] summarize that the network structures AlexNet, VGGM, VGGVD, and TCNN literally solved the texture classification problem on the largest existing texture datasets such as CUReT [267], KTH-TIPS2 [268], or ALOT [269] by achieving a classification performances between 95% and 99% [253]; the actual bottleneck is the amount of correctly labeled data to learn from. Consequently, the extensive LMT image database with the new material labeling

taxonomy proposed in this thesis is highly suitable to perform image-based DL texture classification. In the following, only the material images from the LMT-184-HDB are considered as input for a CNN.

The captured RGB images $\mathbf{I}_{\text{disp}}$ from Section 3.3.4 are sliced into 500-by-500-by-3 RGB patches with no pixel overlap between slices which is denoted as TexTUM-184 database, representing only the images extracted from the LMT-184-HDB. Each image is rotated by 90° to augment the dataset, leading to 350 images per material class, and a split of 80% training and 20% test data is applied. The image size is based on the observations from Section 4.1.1.2. Typical texture image recording distances lead to about 300 pixel-sized coarse texture elements, and hence, 500-by-500-by-3 RGB patches sufficiently capture macroscopic roughness features. Figure 4.49 shows a few randomly selected patches from the DL database to illustrate the variance of patterns to learn from.

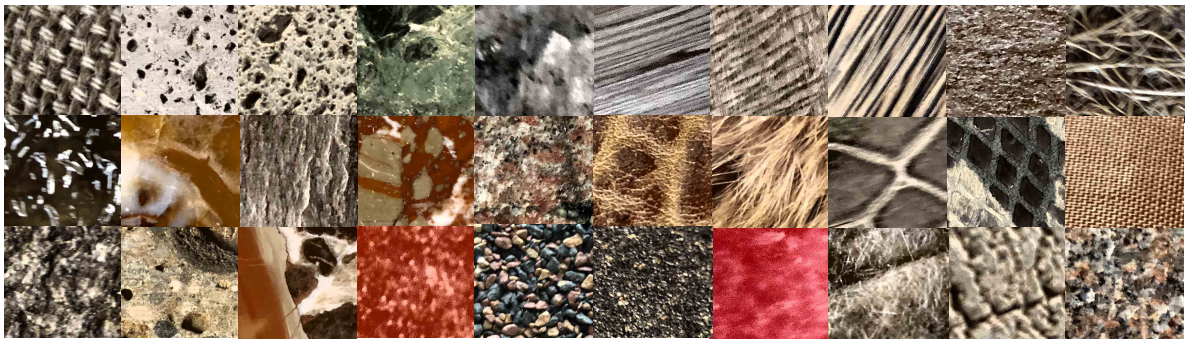


Figure 4.49: Thirty 500-by-500-by-3 RGB material image patches. Overall, 64,400 of these patches (TexTUM-184 database) are used to train (80%) and test (20%) the DL network in this thesis to perform the fine classification of the 184 materials constituting the largest taxonomy-based image texture database to date (adapted from [2] © 2019 IEEE).

As a recommended DL procedure, a smaller subset of the final material database has been tested before the entire database was evaluated in order to draw initial conclusions about the required network structure. During the subjective experiment in Section 4.4, a subset of 24 out of the 184 material representations from the LMT-184-HDB was denoted as LMT-24-HDB. From exactly these complete material recordings, the display images are extracted and denoted as TexTUM-24 image dataset in the following, also for the sake of further comparisons.

Figure 4.50 shows the DL network used in this thesis. It is adapted from AlexNet [252] which has proven to work excellently for object [252] and texture [205] recognition. The selected network settings were cross-validated with the ALOT database and achieved a similar result of 94.91 $\pm$ 0.5%. Note that the ALOT texture images were originally sliced to a 768-by-330-by-3 image patch size [269], but the TexTUM related images have the comparable dimensions of 500-by-500-by-3.

4.5.3.1 Results

The following hyperparameters were used for the training of both the TexTUM-24 and TexTUM-184 DL image databases.

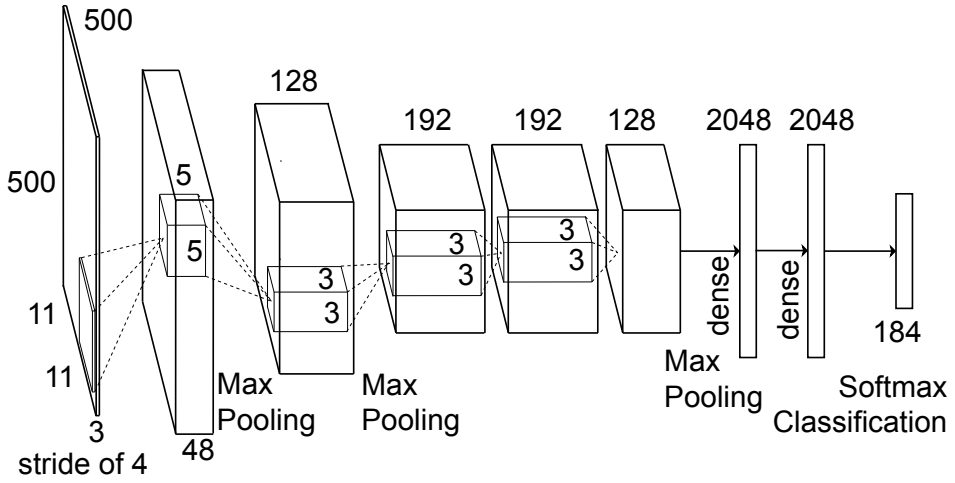


Figure 4.50: DL network based on AlexNet [252] adapted to perform a fine classification of the 184 material classes in this thesis. Each convolutional layer is followed by a batch normalization and a leaky rectified linear unit layer with $\alpha = 0.01$. If not explicitly stated, the stride was set to 2 and the padding to "same" (adapted from [2] © 2019 IEEE).

- Stochastic gradient descend with momentum of 0.9 (SGDM)
- 75% training data, 5% validation data, 20% test data
- Mini-batch size of 128
- Initial learning rate of 0.02, 1% learning rate drop after each 200 iterations
- L2 regularization of $\lambda_{\mathrm{L2}} = 0.001$
- Dropout layers (50%)

All parameters are based on the theoretical concepts explained in the previous section. Overall, the TexTUM-184 network achieves a classification accuracy of $90.5 \pm 0.3\%$ after 50 epochs of training, and hence, only slightly outperforms the handcrafted approaches. Table 4.15 shows the results for other tested texture databases and the TexTUM-24 material selection.

Figure 4.51 (top) and Fig. 4.52 show the independent implementation results of the network both implemented in Python 3.7.4. (using the framework TensorFlow 2.0.0 and Keras API 2.2.4) and Matlab 2019a, respectively.

The SGDM approach has also been compared to the Adam optimizer as shown in the top and bottom part of Fig. 4.52. Reconsidering the argumentation of Wilson et al. [264], this thesis also supports the statement that the Adam optimizer does not necessarily performs better than SGDM. By contrast, the SGDM optimizer with a linearly-decreasing learning rate (1% drop after 200 iterations) reveals a smoother convergence compared to the Adam optimizer and leads to a slightly higher accuracy (+ $1\% \pm 0.4\%$) after less iterations of training.

Further modifications of the network, such as the addition of multiple 3-by-3 CNN layers as proposed by Simonyan et al. [254] have been tested as well and only lead to barely significant improvements in accuracy, however, also include the extended risk of over-fitting.

Table 4.15: Comparison of the CNN performances on the TexTUM, ALOT and KTP-TIPS-2b datasets. Note that ALOT reports more material classes, however, these do not follow a consistent taxonomy. Consequently far less actual classes are present in the dataset with lesser number of samples compared to TexTUM-184, which is hence the largest image texture dataset currently available.

	TexTUM-24	TexTUM-184	ALOT [269]	KTH-TIPS-2b [268]
Classes	24	184	250	11
Samples	8,400	64,400	25,000	1,100
Samples/Class	350	350	100	100
Image Sizes	500-by-500-3	500-by-500-3	768-by-330-3	200-by-200-3
Accuracy	$92.18 \pm 0.6\%$	$90.50 \pm 0.3\%$	$94.91 \pm 0.5\%$	$85.55 \pm 0.3\%$
F_1 score	0.92 ± 0.01	0.90 ± 0.02	0.94 ± 0.01	0.85 ± 0.02

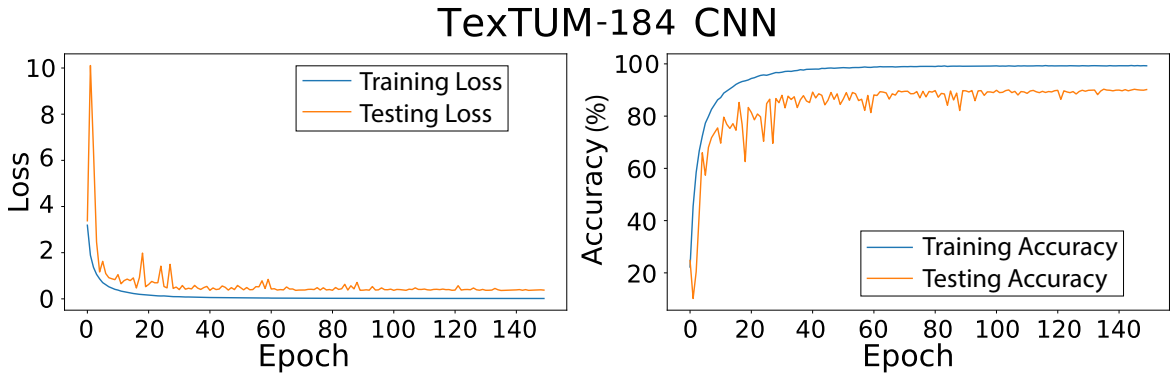


Figure 4.51: Texture image CNN python implementation results.

That is why the more generalizable model with maximal five convolutional layers from Fig. 4.50 is used in this thesis. Figure 4.53 shows the response maps for two exemplary material images of this network architecture. More complex texture patterns, such as hexagons or blob-like structures, are characteristically encoded in the deeper response maps.

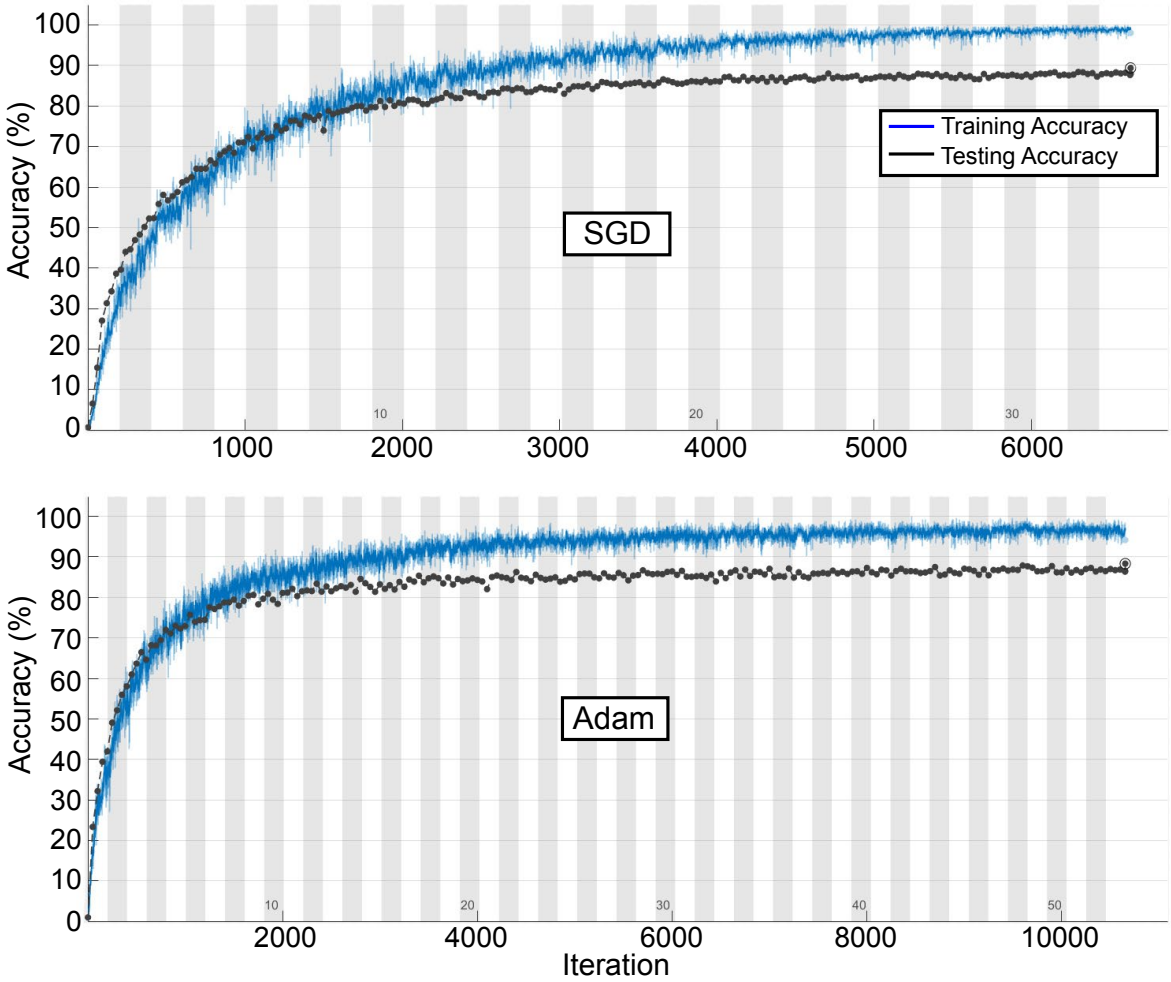


Figure 4.52: TexTUM-184 CNN results using the Matlab implementation of the SGDM (top) and Adam optimizer (bottom). Note that the SGDM approach (with learning decay rate 1% per 200 iterations) converges smoother and leads to a better classification result of the testing set than the Adam optimizer based on the same network configuration and dataset.

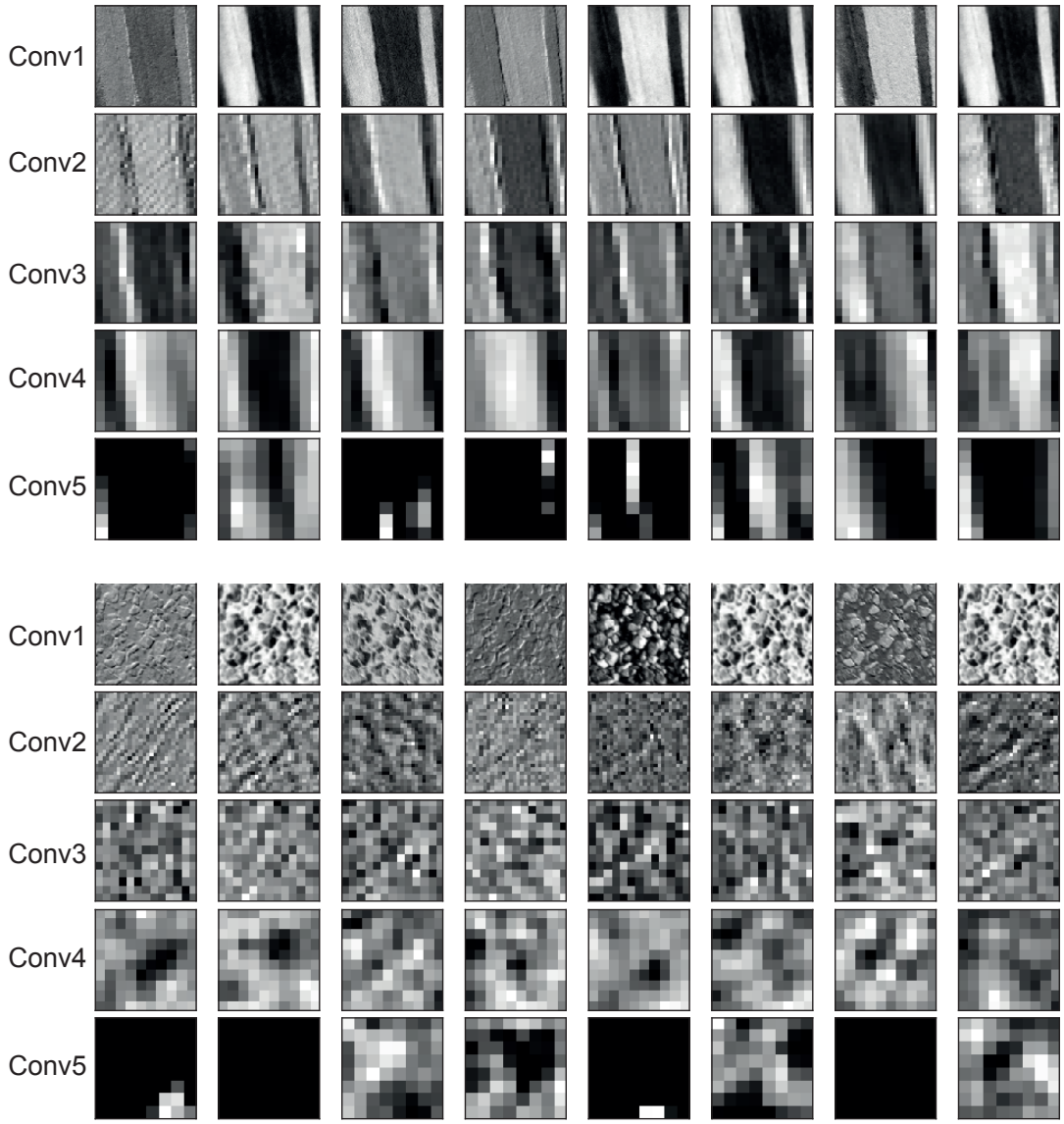


Figure 4.53: Response maps for two different materials (top: silver fir $C_1S_1M_5P_1X_1$, bottom: tar paper $C_8S_4M_1P_1X_{15}$). Relevant structure elements are differently encoded by the trained CNN filters with increasing level of abstraction.

4.5.3.2 Transform-based CNN

The channels of the RGB texture images used in the previous section and related work approaches are intuitively correlated. Consequently, it is worth considering to modify the M-by-N-by-C CNN input in terms of the insights from previous haptic image-based texture recognition approaches. Since transforms such as the DFT2 and the graylevel co-occurrence matrices have been extensively used for the handcrafted image feature engineering (see Section 4.1.1), they are potentially useful for DL approaches as well. Notably the pixel neighbor relations encoded in GLCMs used by the Haralick image features, or the 2D DFT with respect to texture regularity, provide relevant information to learn hidden features which is another relevant insight and small contribution of this thesis. A 500-by-500-by-3 texture patch is transformed and passed as input into the same CNN used in the previous sections as shown in Fig. 4.54, which is denoted as DFT2-graylevel-GLCM, or short, DGG, approach.

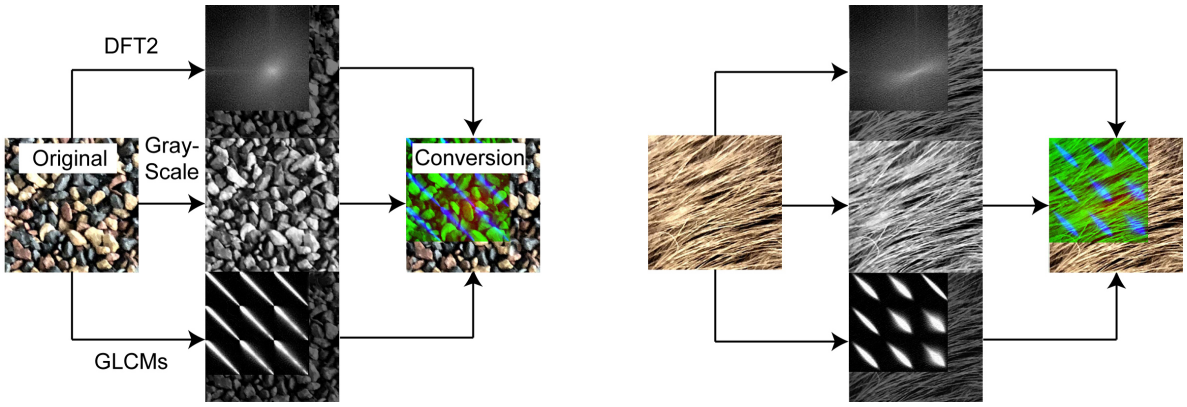


Figure 4.54: Conversion of RGB texture images to DFT2, GLCM, and gray-level image components.

The first channel (Red) is mainly replaced with the absolute value DFT2 of the image, the second channel (Green) width the grayscale converted image, and the third channel (Blue) with nine different GLCMs.

The GLCMs are calculated based on different configurations (different directions 0^o, 45^o and 90^o and the pixel distances 1,2, 3 pixels) to encode different rotations and scales of the original input images. The resulting nine GLCMs are horizontally and vertically stacked and placed within the 500-by-500-by-1 blue color image channel. The lower and right-most pixels are set with the original green channel pixels, and likewise, the lower and right-most part of the red channel is overlaid with the original red channel pixels to partly keep the color information. Using exactly the same network structure from Fig. 4.50 and data source, the accuracy improves to $92.0 \pm 0.5\%$ as shown in Fig. 4.55 compared to the same CNN network with the RGB images as input ($90.5 \pm 0.3\%$). Note that the DGG approach still contains the RGB color information (lower right part of the images in Fig. 4.55), which is useful for specific material classes such as woods or metals. Further potential improvements can result from, e.g., the inclusion of different image scales in future work.

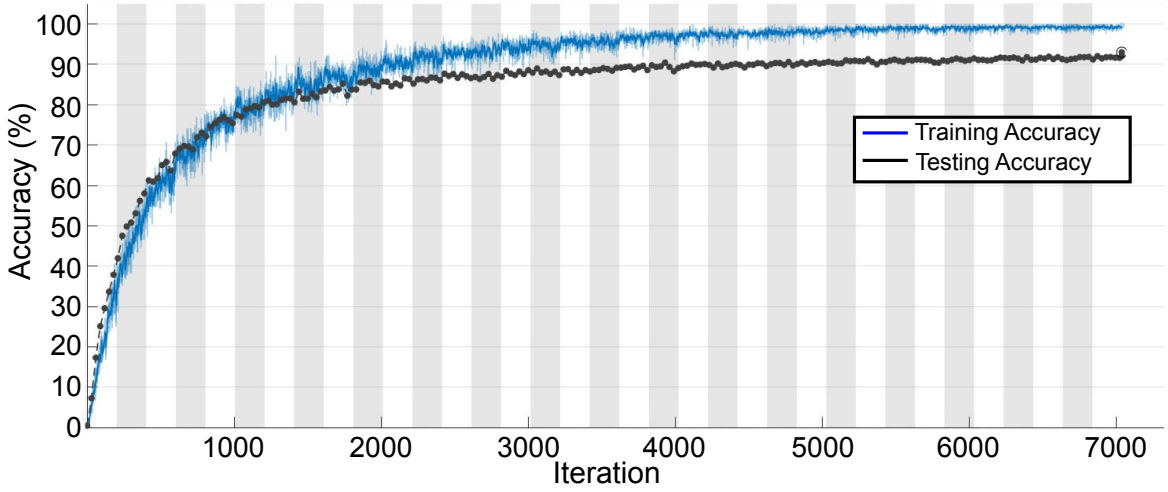


Figure 4.55: DFT-Graylevel-GLCM (DGG) approach training and testing accuracy (92.0 $\pm$ 0.5 %) which outperforms RGB approaches (90.5 $\pm$ 0.3%) based on the exactly same network and parameters.

4.5.3.3 Misclassification Comparison of Image-based Deep Learning and Handcrafted Features

The image-based DL network learns the majority of visual texture patterns from the training images and outperforms the handcrafted feature approaches in terms of classification accuracy. However, one critical point arises considering haptic material classification. Several relevant material characteristics, like the friction coefficient or the compliance, cannot be causally learned by such an image-based deep neural network, and hence, the system has a poor understanding of the haptic material properties and the man-made taxonomy, i.e., the classification into woods, ceramics, and so forth. The defined metric taxonomy similarity $Sim_{\text{tax,C}}$ from Section 4.2.3 quantifies how related the misclassifications are to the proposed taxonomy from Section 3.1.2.

The best handcrafted (HC) approach using the random forest classifier leads to $Sim^{\text{HC}}_{\text{tax,C}}$ = 0.96, and the trained image-based DL network leads to $Sim^{\text{Img-DL}}_{\text{tax,C}}$ = 0.90. Despite a better classification accuracy (i.e., less misclassifications), the labeling error of the TexTUM-184 DL network is still larger than the one of the handcrafted approach. Intuitively, a purely image-based approach faces a critical challenge while distinguishing between hard and soft materials, and has no notion of the characteristic microscopic roughness or surface friction. This fundamental insight has been first identified by Katz [64] who claimed that surface microscopic roughness can only be identified by active sliding. This claim has also been experimentally verified by Ziegler et al. [270] for different paper samples, which are challenging to classify for vision-only classification systems. Tactile features can represent such properties. Their inclusion into a material classification engine is hence inevitable, even though they introduce operator-based variance. On the contrary, image-based DL approaches show remarkable classification abilities of visual surface textures. Consequently, a fusion of multiple approaches based on handcrafted and deep learning-based images features or comparable data sources seemed promising during the creation of this thesis and is investigated in the

following.

4.5.3.4 TUM SpectraNet

Beyond purely image-based approaches, deep neural networks have also been trained on vibrotactile data as shown by Gao et al. [271] and in the thesis-related previous work [6], [10], e.g., on spectrograms of vibrotactile signals in order to learn hidden features from the tactile domain. The 1D vibrotactile and audio data spectra from Section 3.3.2 may be passed into a FNN in order to learn hidden features which has also been examined in this thesis and is explained in the following. As shown in Fig. 4.56, 1,000 spectral values from the finger-based and 1,000 spectral values from the steel tool tip-based vibrotactile recordings, and likewise, 10,000 audio spectral values from the same audio spectra were concatenated to form a 22,000-by-1 input vector and trained with the following optimizer parameters.

- SGD with momentum of 0.9 (SGDM)
- 75% training data, 5% validation data, 20% test data
- Mini-batch size of 128
- Initial learning rate of 0.01, 1% learning rate drop after each epoch
- L2 regularization of $\lambda_{\mathrm{L2}} = 0.001$
- Dropout layers (50%)

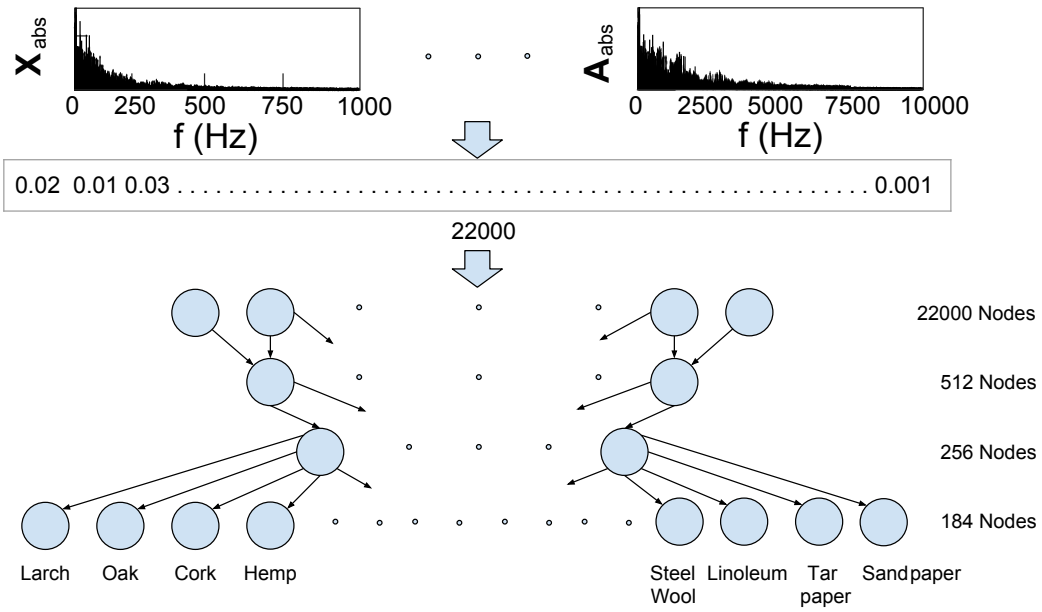


Figure 4.56: SpectraNet structure. The combination of vibrotactile and audio spectrum data during sliding over the material surfaces provides the input for a feed-forward neural network.

The training and validation accuracies over 300 epochs are shown in Fig. 4.57 and saturate at 87.0 % ± 1.9 %. Note that they introduce a larger variance, which presumably results from the operator variance in recording speed.

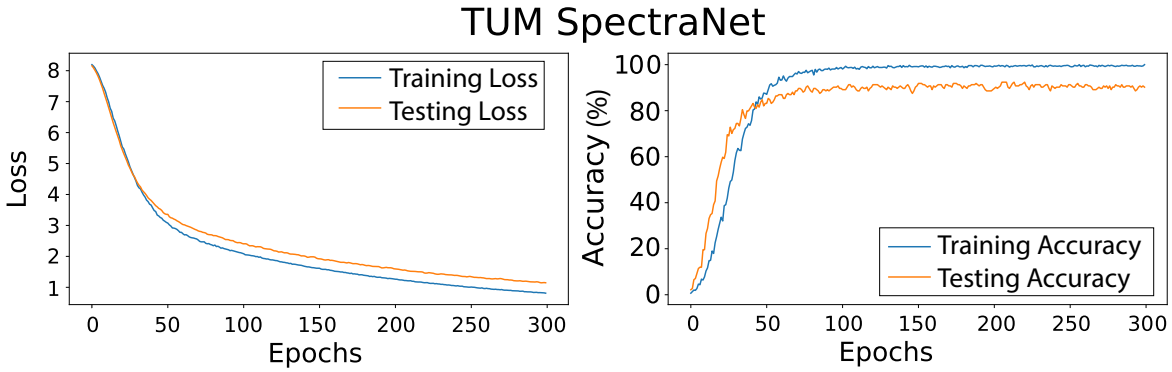


Figure 4.57: SpectraNet accuracies for one fold (80 % training - 20 % testing split).

4.5.4 Hybrid Model for Material Classification

As shown in the previous section, image-based DL approaches are very suitable for texture classification, but face partial challenges while learning haptic-related information absent physical interaction data. The SpectraNet results may depend on operator speed variations during a scan, and, the handcrafted SynTouch features also depend on these conditions in part. That is why an overall hybrid model (HM) comes into consideration which late-fuses the classification outputs from all the different HC and DL approaches mentioned in this thesis. The novel hybrid approach hence aims to connect the surface image-based approaches, the SpectraNet, and the handcrafted feature classifiers (Section 4.3.6) measured by the Texplorer2. The hypothesis is that models using tactile data can help distinguish materials that feel different but look similar, which are challenging to classify for the TexTUM-184 CNN, and vice versa. Another advantageous aspect of this method is the opportunity to let the hybrid network learn the weights of its subcomponents, i.e., the influence of the handcrafted feature-based classifiers compared to the TexTUM CNN or the SpectraNet.

4.5.4.1 HM-24

A smaller version of the hybrid network, namely HM-24, is based on the LMT-24-HDB selected material samples (see Section 4.4) and contains exactly the same 24 material classes as the TexTUM-24 dataset-based CNN. Using this configuration, all aforementioned feature domains (image-based TexTUM-24 CNN, handcrafted features, SpectraNet, and subjective feature evaluation from Section 4.4) are included and lead to four distinctive feature domains. The output probabilities, i.e., prediction probabilities of these four approaches are ultimately concatenated and passed to a shallow FNN containing only one hidden layer in order to determine how each approach needs to be weighted as shown in Fig. 4.58. The combination of the models for 24 classes of the TUM dataset is shown in Table 4.16 which gave valuable design indication for the larger TexTUM-184 dataset. Table 4.17 further shows

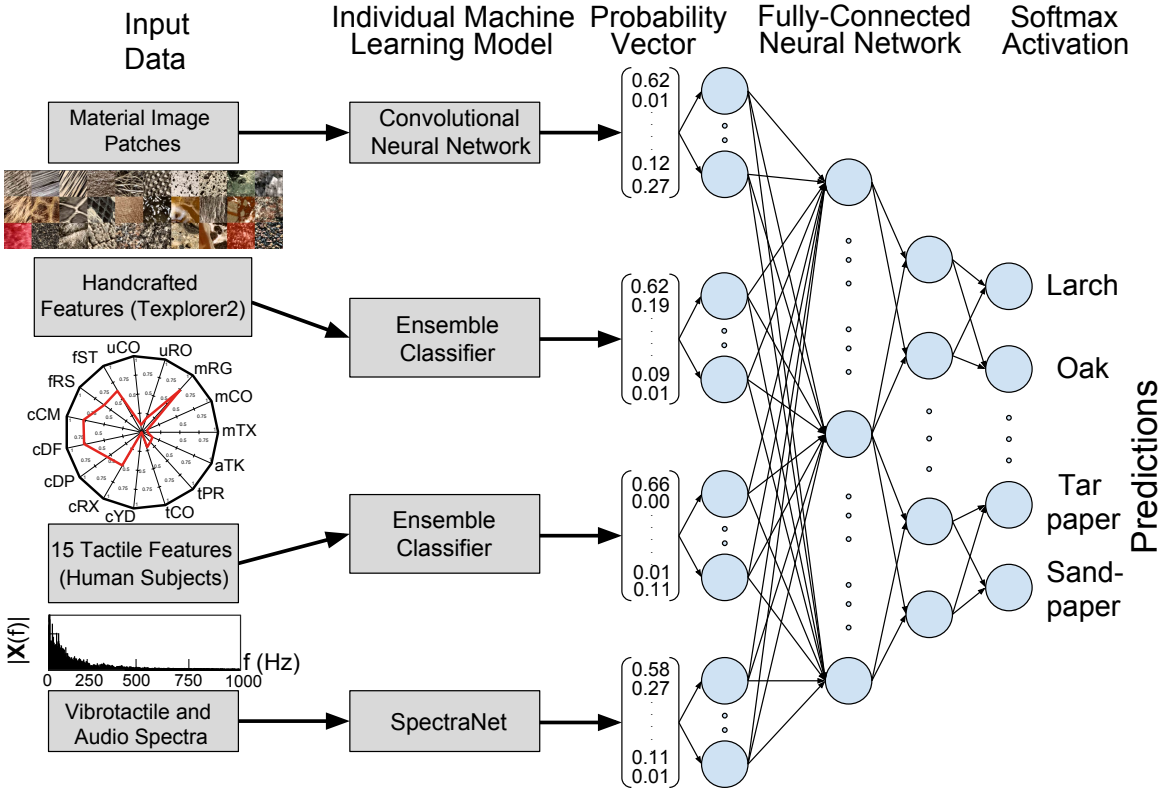


Figure 4.58: Hybrid model (HM-24) configuration based on the image-based DL approach, a trained handcrafted feature-based random forest (ensemble) classifier, the results of the subjective experiment about the 15 SynTouch subdimensions using the same ensemble classifier type, and the vibrotactile and audio data-based SpectraNet. The larger HM-184 uses the same subcomponents except the ratings from human subjects. The individual classification results (prediction probabilities) are late-fused by a shallow neural network which learns how to weight the different models to output a fused classification result.

the result depending on different depths of the fusion neural network and its shape which was tested on the HM-24. Note that only few parameters can be optimized; in this context only the shape (e.g., constant or pyramid-decreasing number of neurons) and depth of the neural network.

Table 4.16: All possible combinations of the HM-24 components, with the other components being disabled. I: TexTUM CNN, II: best 15 Texplorer2-based HC features, III: 15 tactile feature evaluations from subjective experiment, IV: SpectraNet. Note that the inclusion of the subjective ratings is decreasing the overall performance compared to the other sensor-based approaches, and also, if all four feature domains are considered. Consequently, the subjective ratings should be omitted from the fusion approach.

Combination	Accuracy	F1-Score
1 Model		
I	90.27% $\pm$ 0.5%	0.92 $\pm$ 0.006
II	90.00% $\pm$ 0.9%	0.90 $\pm$ 0.010
III	50.00% $\pm$ 1.1%	0.49 $\pm$ 0.011
IV	91.23% $\pm$ 1.4%	0.91 $\pm$ 0.015
2 Models		
I&II	93.33% $\pm$ 0.5%	0.93 $\pm$ 0.006
I&III	94.53% $\pm$ 0.9%	0.94 $\pm$ 0.009
I&IV	96.14% $\pm$ 0.7%	0.96 $\pm$ 0.008
II&III	66.67% $\pm$ 0.9%	0.64 $\pm$ 0.010
II&IV	95.83% $\pm$ 0.5%	0.92 $\pm$ 0.006
III&IV	95.83% $\pm$ 0.9%	0.92 $\pm$ 0.010
3 Models		
I&II&III	95.19% $\pm$ 0.8%	0.95 $\pm$ 0.009
I&II&IV	**98.45% $\pm$ 0.4%**	0.98 $\pm$ 0.004
I&III&IV	97.82% $\pm$ 0.7%	0.96 $\pm$ 0.008
II&III&IV	95.83% $\pm$ 0.7%	0.92 $\pm$ 0.008
4 Models		
I&II&III&IV	98.24% $\pm$ 0.3%	0.98 $\pm$ 0.003

Table 4.17: HM-24 performance for different number and shape of hidden layers. The best configuration resulting from the smaller HM evaluation is used to identify the required depth and shape of the final HM-184 model.

hidden layers	0	1	2	3	4	5	10
constant	97.92%	**98.45%**	97.31%	97.22%	96.90%	96.25%	95.00%
pyramid	97.36%	98.34%	98.10%	97.69%	96.71%	96.48%	84.03%

4.5.4.2 HM-184

The most relevant insight of training the HM-24 is that the dimension ratings resulting from the subjective experiment perform worst among all other domains included, which has been similarly identified during the handcrafted feature comparison in Section 4.3.6, leading to the conclusion that the larger hybrid network model **HM-184** most likely does not require data based on subjective ratings as well. Such a vast subjective experiment would be overly time-consuming, and as shown in Section 4.4.2, highly depending on different subjective interpretations of the tactile features. Ultimately, the superiority of sensorized approaches based on different evaluation models (i.e., handcrafted or deep learning-based) becomes evident. Table 4.18 shows the result from the shallow NN for each fold (cross-validation) for each individual approach and the HM-184.

Table 4.18: Final HM-184 results of the different models performance using 184 material classes. Note that the individual approaches have a much larger variance compared to the hybrid model, which also outperforms the individual approaches.

fold	1	2	3	4	5	average
CNN	90.0%	90.9%	90.5%	90.5%	90.6%	90.5% $\pm$ 0.3%
Majority Voting (HC)	91.2%	90.8%	91.1%	91.0%	90.9%	91.0% $\pm$ 0.2%
SpectraNet	84.0%	87.4%	89.5%	86.8%	87.0%	87.0% $\pm$ 1.9%
Hybrid Model	**94.9**%	**95.1**%	**95.5**%	**94.9**%	**95.1**%	**95.1**% $\pm$ 0.2%

Figure 4.59 shows the resulting accuracies for one example fold of the cross-validation. Note that the hybrid model, which is a linear combination of its components, successfully outperforms all individual approaches and proves that the fundamental concept of majority voting is superior to individual approaches in material classification. Figure 4.60 further

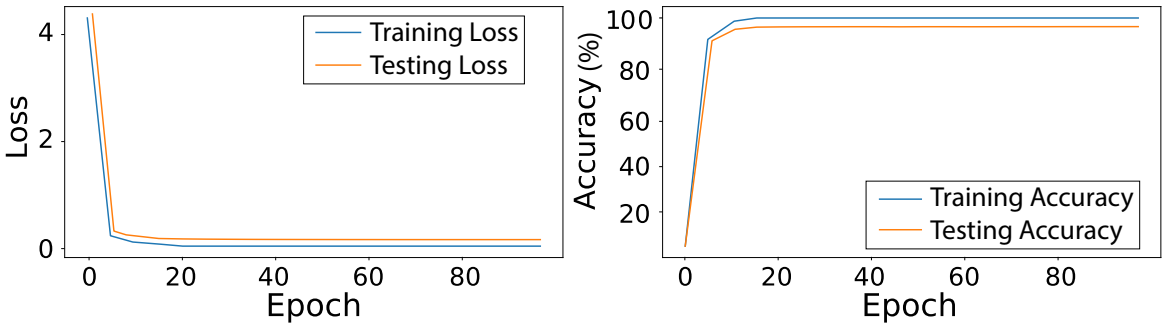


Figure 4.59: Training and validation accuracy plot for one fold of the hybrid model which quickly learns the weights of the individual components.

shows the remaining confusions for the hybrid model. Most importantly, the confusions occur within the major classes itself (e.g., only within C_1 (woods) and the resulting taxonomy similarity $Sim^{HM}_{tax,C} = 0.99$ from Section 4.2.3 is higher than all other similarities, which further proves the superiority of the hybrid model approach. Table 4.19 summarizes all average ac-

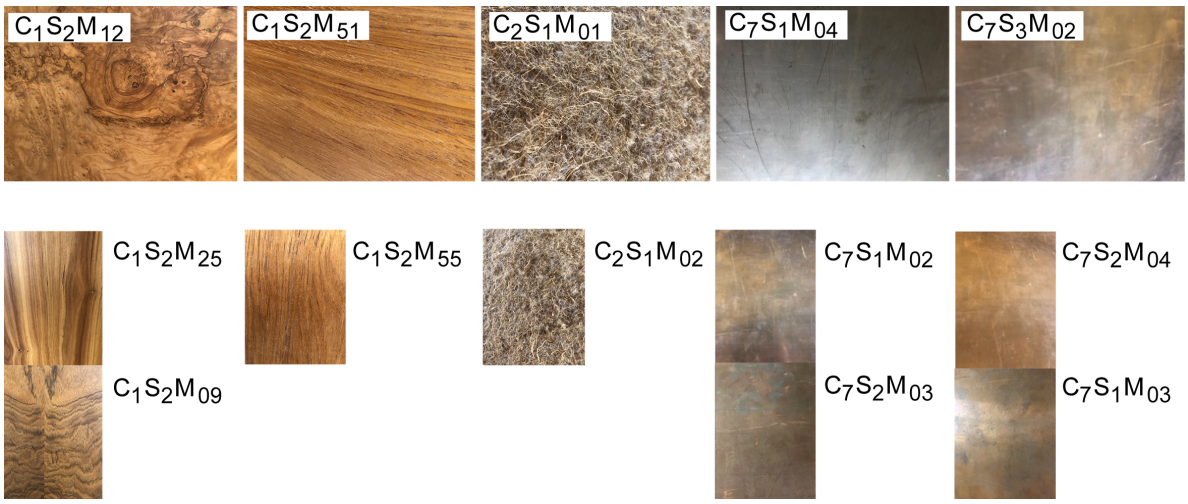


Figure 4.60: Confusions of the final hybrid model on the 184 classes. The top row shows the actual material and the bottom entries the resulting confusions. Note that not only visual information may lead to confusion, but also tactile feature values or vibrotactile input signals from the SpectraNet.

curacies and taxonomy similarities of all used approaches based on the 184 material classes in this thesis for a final overview.

Table 4.19: Accuracy and taxonomy similarity overview of all classifiers. DL_{img} is the purely image-based TexTUM-184 DL approach, and SN is the SpectraNet. Note that ordinary classifiers based on handcrafted features lead to better results in total, and the fusion into the HM combines high accuracies as well as a high taxonomy similarity for the misclassified samples.

	k-NN	NB	SVM	NN	RF	Ada.	MV	DL_{img}	SN	HM-184
Acc (%)	88.1	81.0	82.5	86.2	90.2	87.4	91.0	90.5	87.0	**95.1**
$Sim_{\text{tax,C}}$ (%)	95.1	85.5	93.2	96.3	96.7	96.6	98.0	87.9	88.1	**99.2**

The HM-184 overall achieves the highest accuracy and highest taxonomy similarity which proves the effectiveness of the presented methodology for material classification on the recorded database.

4.6 Chapter Summary

The variety of Texplorer2 data traces required the application of the typical supervised classification workflow to match real material labels to the identified feature space. Most importantly, the introduction of a hybrid material classification approach in Section 4.5.4 successfully fused different classification techniques and is another major contribution of this thesis. The results showed that the intuitive assumption of combining different schemes into one model is valid and can be compared to the concept of ensemble classifiers (see Section 4.3.5), i.e., the combination of different modality results based on different content-related data is superior to individual classifier decisions.

Chapter 5

Haptic Display

The last major contribution of this thesis addresses the haptic display of the relevant tactile dimensions introduced in Section 2.1.2.

5.1 Haptic Display Devices

The following section covers the different devices used for the display of kinesthetic and tactile dimensions. The major trend for commercial application targets wearable devices which are used in VR and gaming applications to grasp virtual objects. This thesis, however, focuses on grounded devices. Up to this point, no wearable device presented in related work accomplished to render all tactile dimensions simultaneously. The most dominant constraints in this context are based on weight and form factor limitations. The design of wearables generally requires free movement in a large workspace, and hence, imposes further constraints on the setup's power supply and cable routing. Consequently, the number of display channels as well as physical signal intensities are limited. The only way to account for this point is to rely on complex and generally expensive setups, e.g., by externally providing a ceiling mount which compensates the setup weight and allows for free motion in 3D. If the display of tactile dimensions is of major interest, similar capabilities can be achieved with notably less effort in device design using grounded haptic devices. During the creation of this thesis, two grounded approaches have been considered which are discussed in the following.

5.1.1 Augmentation of Traditional Force Feedback Haptic Devices

The first single-point of interaction haptic devices were able to render kinesthetic force feedback, and hence, the impressions of friction, hardness, weight, and shape. With the inclusion of force perturbation techniques based on height maps [16], object surfaces are enhanced with spatial textural, i.e., macroscopic roughness features. Recent work from Culbertson et al. [106] extended a Phantom Omni haptic device with a TL-002-14R haptuator (Tactile Labs, USA) to further render microscopic roughness during unconstrained surface material exploration. Similarly, Khurshid et al. [272] augmented a robotic gripper with vibrotactile feedback for teleoperation tasks. The thesis-related work in [8] extended this concept by the inclusion of a thermal unit to the grasped stylus holding a voice coil actuator. The setup in

Fig. 5.1 shows such an augmented setup for teleoperation which was created during this thesis and presented at the Mobile World Congress 2018 in Barcelona.

Figure 5.1: Mobile World Congress teleoperation setup based on augmented force feedback haptic devices. The operator is able to explore the remote environment and gets thermal, force, and tactile feedback.

Tool-mediated setups are suited for tactile dimension display, however, they face the severe challenge of different operator arm and hand weight as well as varying forces which can have significant influence on the tactile display of, e.g., vibrotactile signals. As a different approach in this thesis, another type of grounded device can compensate such applied forces while being able to sufficiently display the relevant tactile dimensions.

5.1.2 Tactile Computer Mouse

The second approach conceptualized and implemented in this thesis presents a planar-wielded haptic device resembling a computer mouse, denoted as tactile computer mouse (TCM), which provides multidimensional tactile feedback to a user and is well-suited to address the exploratory procedures mentioned in Section 2.1.2. It was also filed as patent (TUM170301PDE-48/SR: Input / Output Device and Method for the Computer-based Display and Exploration of Real or Virtual Object Surfaces) during the creation of this thesis. Comparably, the authors in [5], [273], [274] present various computer mouse-resembling devices that addressed single tactile dimensions. Golledge et al. [275] use force feedback and tactile vibrations based on haptic maps to render virtual walls and texture information using the Logitech Wingman force feedback mouse. Terry et al. [276] employ a piezoelectric transducer to generate tactile vibrations on the tip of a human finger. The device in [277] uses a solenoid for tactile vibration generation. In [273], a steered wheel displays two-dimensional paths or boundaries. The work in [278] gives directional cues using braille displays. The presented wireless device in [274] let users experience various shapes. None of these previous devices, however, were conceptualized to specifically recreate multimodal tactile dimensions of touch. Consequently, this thesis stepped toward this ambitious objective by introducing the TCM concept, which was constantly improved as visualized in Fig. 5.2.

The major objective of the first TCM version (TCM1) is the display of the five major tactile

Figure 5.2: Tactile computer mouse evolution. The first version (left) and two improved design versions [5] (middle-left and middle right) are able to display the five relevant tactile dimension from [63] either using an embedded VCA or a C2 tactor, respectively. The current version TCM2 (right) is further able to recreate the fifteen subdimensions from Fishel et al. [66].

dimensions from Okamoto et al. [63] in a monolithic approach, and has been published in [5]. The improvements implemented in the TCM2 allow for the display of all fifteen subdimensions introduced by Fishel et al. [66] directly at the human index finger. The following sections discuss the system overviews of both devices in detail as well as the display of the tactile dimensions.

5.1.2.1 TCM1

The initial TCM1 version, shown in three subsequent prototypes in Fig. 5.2 (first three images from the left) and rendered in Fig. 5.3, supports the display of friction force, macroscopic and microscopic roughness, thermal, and binary hardness (rigid or soft) feedback. A modified version (third image in Fig. 5.2) was conceptualized to display the vibrations using a C2 tactor in a localized approach to the human finger directly. Note that in the following the version in the second-left image in Fig. 5.2 is used to represent the features of the TCM1.

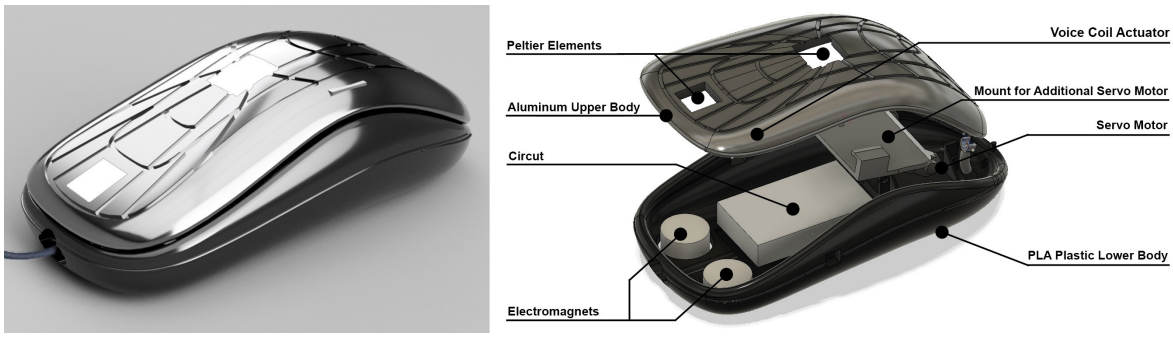


Figure 5.3: TCM1 actuator placement. All actuation components are embedded inside the TCM1. The left and right click buttons of a common computer mouse are moved to the left and right side of the TCM1, respectively. The voice coil actuator is not visible in this image, since is located beneath the upper TCM1 body in between the Peltier elements (adapted from [5] © 2018 IEEE).

All TCM versions are designed in Autodesk Fusion 360 and are made of commonly available PLA material. A controller board from a common computer mouse (Logilink ID0011, 800 dpi resolution) obtains the on-screen position $\mathbf{x}_\mathrm{p}$ in pixel coordinates. The total mass of the TCM1 is about $m_\mathrm{TCM1} = 0.25$ kg measured with an electronic scale. Figure 5.4 shows the association of the five major tactile dimensions from Okamoto et al. [63] and the used actuators of the TCM1. All components can be driven independently.

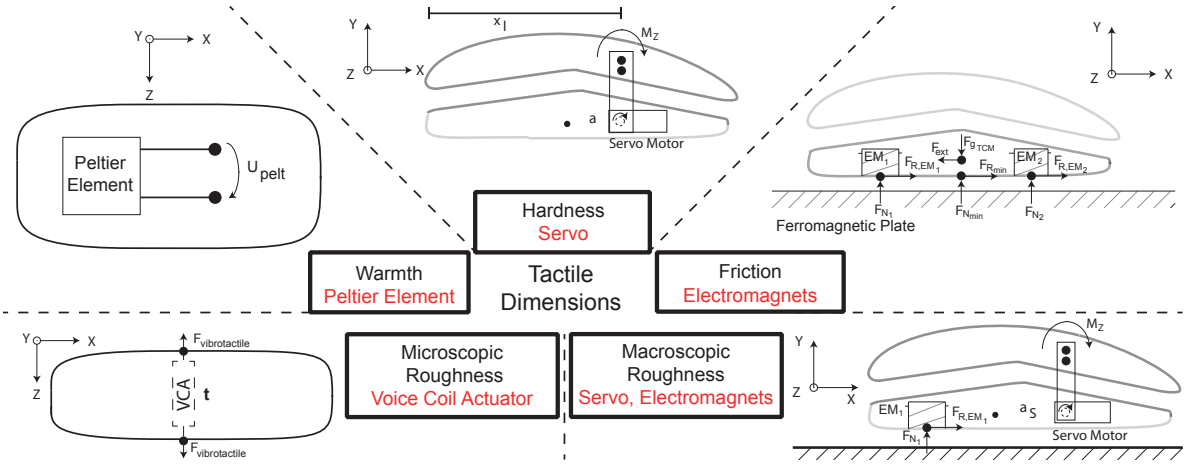


Figure 5.4: Each tactile dimension from [63] requires a different actuator. For example, the tactile dimension of friction is reproduced by adjusting the voltage of the built-in electromagnets to change the attraction forces on an iron plate, or fine roughness by displaying recorded vibrotactile signals on VCAs. The TCM actuation principle is highlighted in red font below each tactile dimension (adapted from [5] © 2018 IEEE).

The data for the display of the individual tactile dimensions is based on different sources, and Table 5.1 summarizes all values for the ten materials used in [5].

1. **Friction** The static and dynamic friction coefficients reported by Fishel et al. [116] are used.

2. **Hardness** An Optoforce sensor (OMD-10-SA-10N) is mounted on a uArm Swift Pro robot arm (Ufactory, USA) and ensured a measurable indentation depth alongside the measured force value. Values above 1,000 N/m were clipped and the corresponding materials are marked as hard, i.e., not deformable.

3. **Macroscopic Roughness** Image- and reflectance based recordings determine the intensity of gratings and surface asperities.

4. **Microscopic Roughness** Finger-based accelerometer recordings are displayed through the VCA.

5. **Warmth** Tabular thermal conductivity values [187] are used to cool down the Peltier elements based on heuristically determined cooling times.

Figure 5.5 further shows the abstracted hardware components in a block diagram, and Fig. 5.6 a flowchart of the corresponding material display application implemented in Visual Studio 2013 and the Chai3D framework (Version 3.0) based on a Windows machine (Intel Core i7-4770 at 3.4 GHz, 16 GB RAM, Windows 7 Enterprise Edition). Note that this application is also used for the most recent TCM2 (see Section 5.1.2.2) with several adaptations. The virtual material display application allows the user to change between different surface representations using the keyboard.

Table 5.1: Surface images of material samples, denoted as $S_{<ID>}$, used for the TCM1 display application and calculated model parameters to represent the five tactile dimensions of the ten material samples. Note that the term real material (RM) is also used as synonym. The spring stiffness of the thin sandpaper sheet is identical to the stiffness of the underlying wooden table. MaRS is the macroscopic roughness strength from Section 4.1.4 which determines the existence and intensity of coarse structures on a surface. MSE is the mean signal energy of a vibrotactile signal recorded at 60 mm/s (adapted from [5] © 2018 IEEE).

Property	S_0	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
μ_k	0.72	0.63	0.55	0.40	0.49	0.59	0.51	0.71	0.71	0.80
y_{ind} (mm)	5	0	0	0	0	0	0	3	9	0
$MaRS(\%)$	5	3	12	0	4	25	2	10	20	0
MSE (g)	0.08	0.1	0.11	0.01	0.09	0.21	0.06	0.05	0.03	0.20
k (N/m)	300	3,000	3,000	3,000	3,000	3,000	3,000	500	20	3,000
λ (W/(m· K))	0.04	1.31	0.75	16	0.10	0.10	0.17	0.13	0.19	0.05

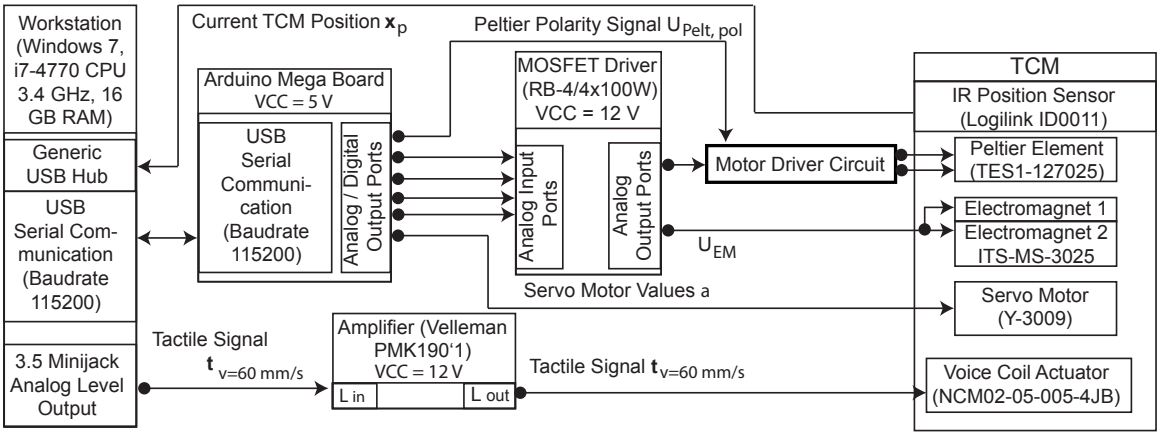


Figure 5.5: Driver circuit flowchart for control signal mapping to the actuators (adapted from [5] © 2018 IEEE).

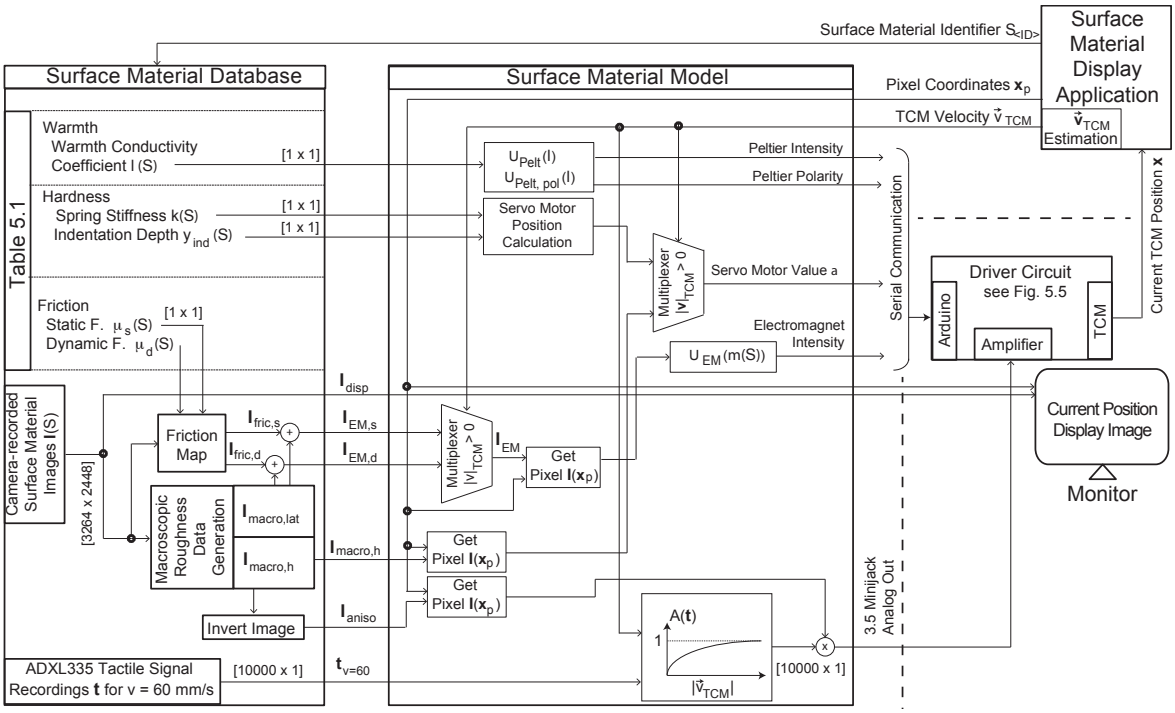


Figure 5.6: TCM1 virtual material display application overview. The user selects a material sample $S_{\text{<ID>}}$ and the corresponding model representation is fetched from the database. The physical properties are converted into signal representations and sent to the driver circuit from Fig. 5.5 using serial communication. The vibrotactile signals are sent via the sound card to the amplifier. The TCM1 actuators receive the signals and display tactile impressions to the user. The according on-screen pixel-coordinates $\mathbf{x}_\text{p}$ are used to calculate the required signals when the user moves the TCM1. The images $\mathbf{I}_\text{disp}$ are shown on a monitor and the current position in the application is shown at $\mathbf{x}_\text{p}$ as a 15-mm-radius sphere representing the human index finger. Note that the on-screen motion of this sphere was closely matched to the real mouse controller movement to avoid any visual-haptic speed differences (adapted from [5] © 2018 IEEE).

5.1.2.2 TCM2

The TCM was completely redesigned [3] to further include the more detailed SynTouch feature space [67] during the creation of this thesis, and is shown in Fig. 5.7. The novel TCM2 displays all dimensions on the index finger to fully mimic the static touch, pressing, and sliding EPs during material surface exploration.

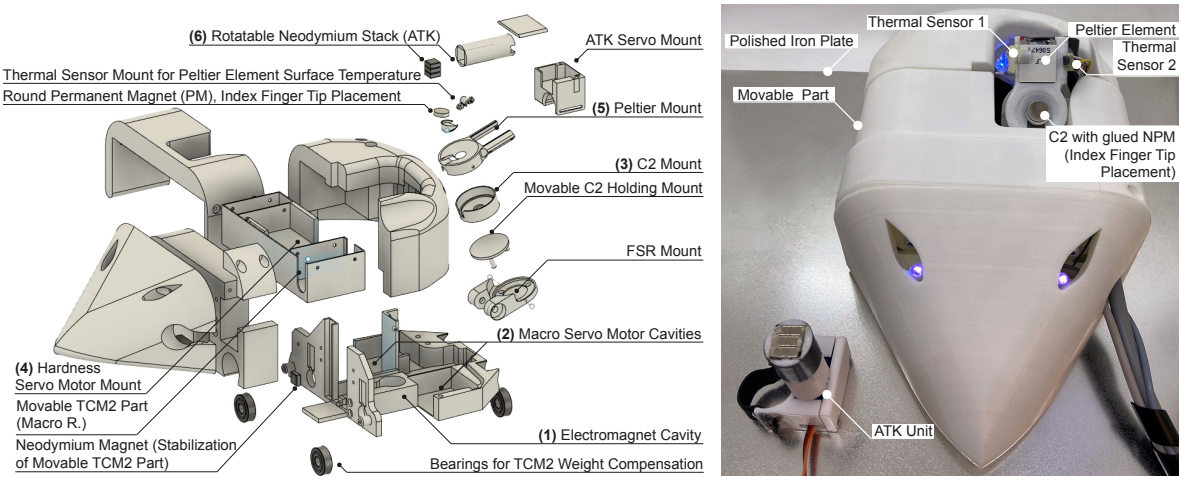


Figure 5.7: Final TCM2 mechanical parts schematic (left) and prototype (right). The human user places the index finger on the neodymium permanent magnet (NPM), or PM in short, which is firmly attached to the C2 tactor and the movable FSR mount for hardness rendering. A small wearable component allows the display of adhesion forces. The remaining part of the index finger skin surface receives controlled thermal feedback. The middle component (movable part) which holds the hardness and microscopic roughness actuators can adjust its height at two points to represent macroscopic roughness. The bottom part contains an electromagnet to adjust the friction between the TCM2 and the underlying iron plate (adapted from [3] © 2021 IEEE).

The display of the 15 SynTouch subdimensions puts challenging constraints on the design of such a novel haptic display. For example, superimposed macroscopic roughness and hardness display requires independently working servo motors and the corresponding mechanical construction. Consequently, the TCM2 prototype is relatively large and heavier compared to other haptic devices. However, four bearings have been attached to the bottom side to strongly compensate the device and user hand weight. The TCM2 can hence only move in one dimension. This is an acceptable trade-off in achieved rendering capabilities and imposed constraints, since the TCM2 movement range sufficiently resembles the tangential sliding motion of a finger over a surface.

The following list summarizes the actuators used for the display of the tactile dimensions.

1. **Friction**: An electromagnet (ITS-MS-2015-12VDC, Intertec, Germany) is embedded into the bottom part of the TCM2 (highlighted as (**1**) in the left image of Fig. 5.7) to render friction effects.

2. **Macroscopic Roughness**: Two servo motors (Reely RS-610WP MG, Conrad Electronic SE, Germany) embedded into **(2)**, see Fig. 5.7) can lift a movable part of the TCM2 to display macroscopic roughness sensations. Note that the movable part is vertically aligned to the TCM2 body using neodymium permanent magnets (NPMs).

3. **Microscopic Roughness**: The vibrotactile actuator design guidelines from Choi et al. [134] are followed by choosing the linearly moving magnet C2 tactor (Engineering Acoustics, USA) due to the ease of integration, broad vibrotactile rendering range, cost, response time, and flexible coupling to present localized vibrotactile feedback to the human index finger tip (see **(3)** in Fig. 5.7). Note that its main operating frequency range is within 200 Hz – 300 Hz [279] to mainly match the Pacinian corpuscle frequency range as determined by Israr et al. [280]. Consequently, other frequencies relevant for vibrotactile rendering need to be amplified, notably the low-frequency intensities sensed by the Meissner corpuscles. Since the user's finger tip is in direct contact with the vibrating part of the C2 via a glued smooth neodymium permanent magnet (15 mm diameter, see Fig. 5.7), a completely smooth surface is represented if the C2 is turned off. Additionally, the PM works together with another wearable attachment to the user index finger to simulate adhesive tack **(6)** (see Fig. 5.7).

4. **Hardness**: Various hardness properties like compliance or relaxation are conveyed by an index finger pressing mechanism (4 cm long lever) attached to another servo motor (Reely RS-610WP MG, Conrad Electronic SE, Germany), which is placed into **(4)** in Fig. 5.7. The maximum torque of the servo motor at 6V is 85 Nm, and hence, the resulting maximal force is 21.25 N which is sufficient [28] to exceed human finger pressing capabilities. Note that the pressing mechanism is stabilized by another bearing, and overall, the display of hardness is also exceeding the capabilities of a PhantomOmni or other comparable grounded devices. Notably the modular design allows to embed even stronger servo motors into the TCM2 to further increase hardness display if required, and also, to reduce undesired servo motor noise.

5. **Warmth**: Closed-loop thermal rendering based on three thermal sensors and a Peltier element (QuickCool, QC-127-1.4-3.7AS, Quick-Ohm Kuepper & Co. Limited) is introduced (see **(5)** in Fig. 5.7). Consequently, the initial cooling gradient and the ongoing thermal persistence can be rendered by adjusting the Peltier surface temperature.

5.2 Tactile Dimension Rendering and Display

The following section describes the rendering of the individual tactile dimensions. Since the TCM2 concepts are partly based on those of the TCM1, the following sections summarize the common rendering techniques, but also highlight the improvements achieved by the TCM2.

5.2.1 Friction

Electromagnets, which are slid over iron plates, allow for a direction-independent increase of the effective normal force, and consequently, adjustable friction forces. Note that a Teflon coating optionally can be attached to further decrease the friction during sliding. The differences between TCM1 and TCM2 result from their different weights and how the devices are held; the TCM1 is entirely grasped by the human hand, whereas the TCM2 is only moved by

the index finger to mimic bare finger sliding. Consequently, different maximal force requirements determine the sliding conditions and the number of EMs. Note that the number of electromagnets, and hence the overall friction force, is scalable, which is another advantage of the TCM in general compared to other haptic display devices.

5.2.1.1 Friction - TCM1

The total TCM1 friction force $F_{\text{Fr,TCM1}}$ that can be generated is the sum of both electromagnetic forces and the initial TCM friction force $F_{\text{Fr,min}}$ of the underlying plate, i.e., without enabling the electromagnets

$$F_{\text{Fr,TCM1}} = F_{\text{Fr,EM}_1} + F_{\text{Fr,EM}_2} + F_{\text{Fr,min}} \tag{5.1}$$

with $F_{\text{Fr,min}} = m_{\text{TCM1}} \cdot g \cdot \mu_{\text{TCM,min}}$. The required dynamic friction without powering the electromagnets is measured by using a digital spring force meter (Quantum Abacus, model A05) during the transition from sliding friction to static friction, leading to $F_{\text{Fr,min}}$ =0.5 N. Hence, $\mu_{\text{TCM1, min}} = \frac{F_{\text{Fr,min}}}{g \cdot m_{\text{TCM1}}} \approx 0.20$ is the value for the kinetic friction coefficient between the iron plate and the TCM1.

Pulse width modulation (PWM) controls the effectively applied voltage U_{EM} on both EMs. For the maximum voltage $U_{\text{EM}} = 12$ V, $F_{\text{Fr, max}} = 6.6$ N is measured, leading to a maximum friction coefficient $\mu_{\text{TCM,max}} \approx 2.69$ in the TCM1 implementation using $\mu_{\text{TCM, max}} = \frac{F_{\text{Fr,max}}}{g \cdot m_{\text{TCM1}}}$. For U_{EM} in within 0.0 V - 12.0 V, the overall electromagnetic force $F_{\text{R,TCM}}$ is proportional to the current square I_{EM}^2 through both electromagnets [281]

$$F_{Fr,TCM} \propto I_{EM}^2 = F_{R,min} + c_1 \cdot U_{EM} + c_2 \cdot U_{EM}^2 \tag{5.2}$$

with c_1 and c_2 being characteristic parameters of the electromagnets. I_{EM} denotes the electric current through both EMs. The parameters c_1 and c_2 depend on internal coil attributes like the coil winding number and the coil diameter. A spring force meter is used to determine the resulting friction force $F_{\text{Fr,TCM1}}$ depending on the input voltages U_{EM}, shown in Fig. 5.8.

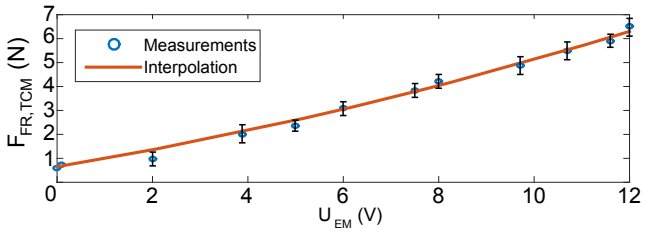


Figure 5.8: Applied EM voltage and resulting friction force for the TCM1. Similar to the approach proposed by Culbertson et al. [130], the relation between electric input and achieved force output was determined for the TCM1. A polynomial regression using a quadratic model interpolates the measured data. Without the electromagnetic force, the minimum traction is about 0.5 N, associated with $\mu_{\text{TCM,min}} \approx 0.20$. The maximum $\mu_{\text{TCM,max}}$ for 12.0 V supply voltage is hence $\mu_{\text{TCM,max}} \approx 2.69$ (adapted from [5] © 2018 IEEE).

The Matlab function polyfit interpolates the data points for $F_{\text{Fr,TCM1}}$ depending on U_{EM} as

$$F_{\text{Fr,TCM1}} = \frac{0.015\ \text{N}}{\text{V}^2} \cdot U_{\text{EM}}^2 + \frac{0.34\ \text{N}}{\text{V}} \cdot U_{\text{EM}} + 0.5\ \text{N} \tag{5.3}$$

while assuming constant, linear and quadratic components as inferred by [130]. Using this configuration, the Coulomb friction model from Section 2.3.2 was used for the ten different materials from Table 5.1.

5.2.1.2 Friction - TCM2

The TCM2-based friction assembly is partly comparable to the one used in the TCM1. The index finger alone is supposed to move the entire device, and hence, less overall forces are required for friction display. Consequently, another evaluation approach is necessary to measure the friction forces only at the index finger contact point of the TCM2. As a solution to this challenge, an Optoforce sensor (OMD-20-SA-40N) is embedded into a PLA construction as shown in Fig. 5.9, which is magnetically coupled to the location where the index finger touches the TCM2 (index finger tip placement from Fig. 5.7). During several recordings, a range between 0.3 and 0.9 was determined as currently displayable friction coefficients. Note that further device improvements (e.g., Teflon coating and a second EM) can extend this range. However, the current coefficients are sufficient to represent a broad range of natural material sample friction coefficients.

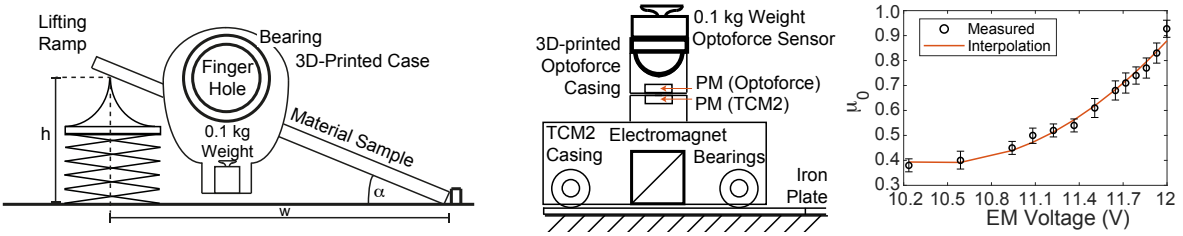


Figure 5.9: Left: Mechanical friction coefficient acquisition setup. Since the width w and height h are known, the angle α leads to the finger-surface friction during the incipient slip event. The finger is put through a bearing which guarantees that the constant weight always points toward the Earth's gravitational center. Middle and right: Electromagnet voltage to friction coefficient estimation. An Optoforce (OMD-20-SA-40N) sensor embedded into a tightly fitting PLA part is magnetically coupled to measure normal and friction forces for different electromagnet voltages. The right plot shows which voltage is required to simulate a specific friction coefficient. Note that larger coefficients can be achieved by using a supply voltage larger than 12 V if required (adapted from [3] © 2021 IEEE).

The required friction coefficients for the bare finger can be determined by using a mechanical setup like the one shown in the left part of Fig. 5.9, which has the advantages of a constant weight that always point toward the Earth's center of gravity.

As introduced in Section 2.3.2, the Dahl friction model [172] is selected as friction rendering approach as shown in Fig. 5.10 for the three material samples lead, rubber (isoprene), and silicone (from left to right). Natural material samples exhibit only dry friction, and hence, viscous friction components and the Stribeck effect are redundant. Notable stick-slip chatter can be observed for material samples like silicone (right) which have notably different static and kinetic friction coefficients.

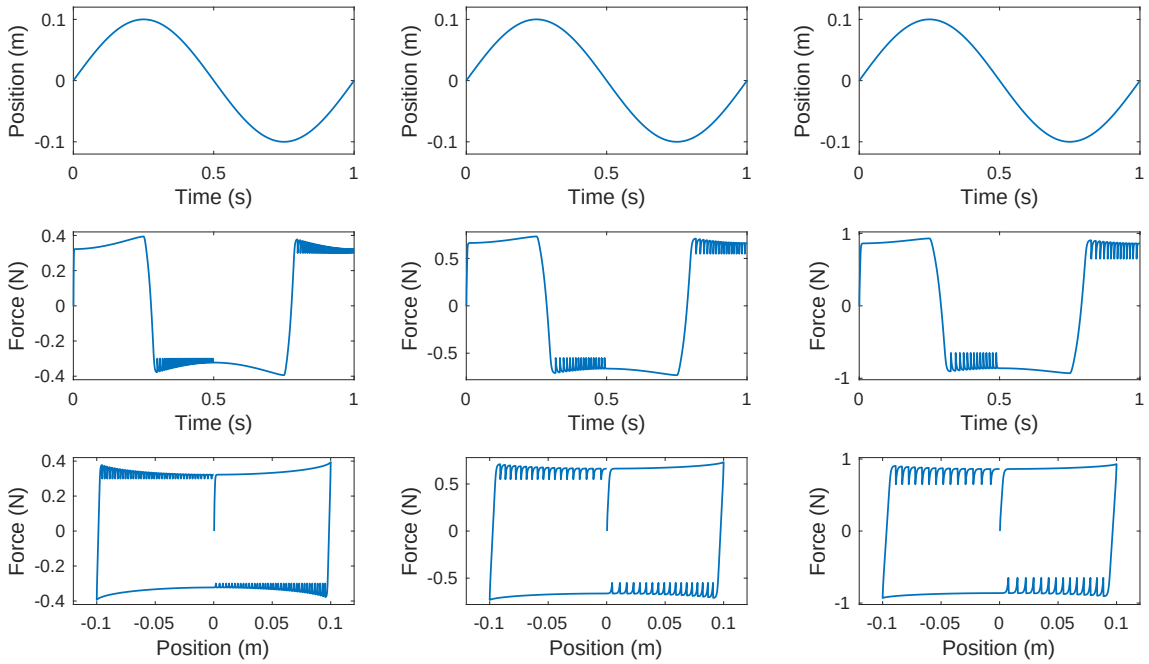


Figure 5.10: Generalized Dahl friction model for three specific materials, from left to right: lead, rubber, silicone. The first row shows the same 1 Hz input periodical sliding motion for all three cases displayed. The second row shows the temporal evolution of the calculated friction force. The force-displacement curves in the third row summarize the first and second row and visualize the frictional displacement effect, i.e., the force depending on a specific position. Note that the third row y-axes are not likewise scaled in order to better visualize the stick-slip chatter, which becomes observably strong if the material's static and kinetic friction coefficients differ significantly.

5.2.2 Macroscopic Roughness Rendering

The display of macroscopic roughness and contour information heavily depends on the device design, but is comparable for both TCM versions. Both versions follow the same concept of splitting the display of macroscopic roughness into height patterns using servo motors, and tangentially-perceivable surface obstacles using electromagnets.

5.2.2.1 Macroscopic Roughness Rendering - Surface Profile Rendering

Minsky et al. [98], [99] introduced the sandpaper model to render spatial gradients of texture height map values, and hence macroscopic roughness, which has been extensively examined and extended ever since. For example, the bump mapping approach from computer graphics has been validated by Srinivasan [16] and Basdogan [100], [131] for typical force feedback devices. However, a purely image-based approach for determining the actual height information of, e.g., a flat surface is prone to high-contrast painted patterns because of the missing depth information. If real material surfaces are considered, the reproducible transform of such images to real height maps is a challenging task. No approach has been reported that relies on a combination of painting the materials in a unitary gray tone and then

use infrared (IR) distance recordings in order to infer the actual height profile. This thesis hence introduces a novel extension to the image-based approaches that leads to actual height maps including the three macroscopic roughness subdimensions MaTX$^{\text{TUM}}$, MaCO$^{\text{TUM}}$, and MaRG$^{\text{TUM}}$. Note that the terms height and depth are interchangeable in this context and are, in the following, referred to as depth values. The IR sensor (AF2349 470) embedded into the Texplorer2 [2] from Section 3.3 is used in combination with gray-sprayed surface images $\mathbf{I}_{\text{macro}}$ to specifically design macroscopic depth maps. Since the sensor values need to be precisely matched to the real depth pattern of the material samples, the core idea is to deliberately change the visual appearance first which reveals the following advantages:

1. Man-made painted structures and metal surface gloss are completely removed and do not distort the measured depth value.

2. If the same color spray (hexadecimal color code: $\#D9D9D9$) and IR sensor mounted in a constant distance is used (as in the Texplorer2 unit 1), this approach is highly reproducible for other material samples.

3. The images of the gray-sprayed surfaces lead to $\mathbf{I}_{\text{macro}}$ applicable for haptic display which is underlying to the visual (display) rendering of $\mathbf{I}_{\text{disp}}$. Note that the corresponding display images presented to the user need to be captured first and aligned to match the positions of the underlying $\mathbf{I}_{\text{macro}}$.

A well-defined ramp (25 mm depth, 100 mm length) has been 3D-printed to serve as a depth reference as shown in Fig. 5.11. Most importantly, the ramp and all materials are sprayed with the same gray tone in order to remove all painted patterns. The Texplorer2 unit 1 is slid over the ramp and the IR sensor measures different depth levels ranging from d_0 = 7 mm (initial sensor distance) to d_0 + d_{max} = 32 mm over a length of 100 mm. Using

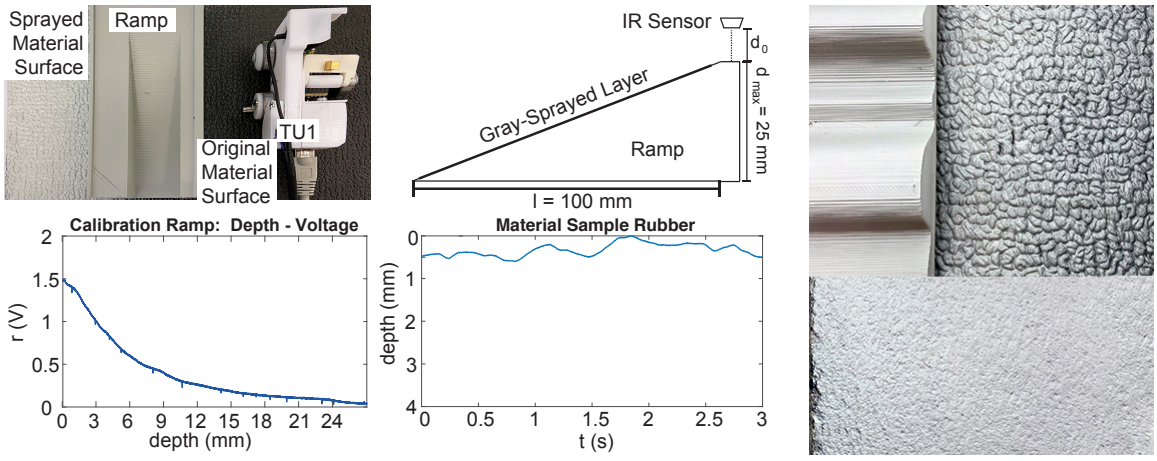


Figure 5.11: Well-defined ramp (upper left and right) to determine the mapping from IR sensor readings to the depth in mm. The collected IR values for the ramp (bottom left) are used to map the IR values of a material scan (bottom right) into the true depth value (adapted from [3] © 2021 IEEE).

linear regression with $a = -0.19$ and $b = 1.6$ (Matlab fitlm), an IR sensor reading $\mathbf{r}[n]$ (bottom

left plot in Fig. 5.11) is mapped to a depth $\mathbf{d}[n]$ using

$$\mathbf{d}[n] = \frac{ln(\mathbf{r}[n]) - b}{a} - d_0 \tag{5.4}$$

which is applied to adjust the material samples recordings for haptic display. The resulting range of the depth values then represents the macrotexture (MaTX$^{\text{TUM}}$). Based on this approach, the macroscopic roughness images $\mathbf{I}_{\text{macro}}$ are inferred from the grayscale-converted surface material display images $\mathbf{I}_{\text{disp}}$ which inherently consider the subdimensions of macro coarseness (MaCO$^{\text{TUM}}$) and macro regularity (MaRG$^{\text{TUM}}$).

As discussed in Section 2.1.1, perceivable height patterns, (or depth patterns, respectively) are mainly perceived by the slowly-adapting Merkel disks in the index finger [13], [34]. A haptic display device hence requires physical displacement between the frequency range of 0 Hz – 5 Hz. Servo motors are one option to create such physical displacements. Two servo motors (SMs) are embedded into the TCM2 as shown in Fig. 5.12. The movable

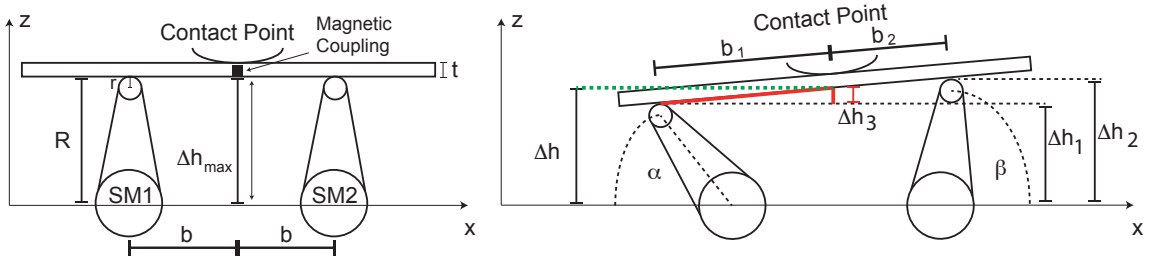


Figure 5.12: TCM2 macroscopic roughness display using two servo motors. Note that the center of the movable part (with the contact point) is vertically aligned using neodymium permanent magnets (NPMs) (adapted from [3] © 2021 IEEE).

part holding the contact point in its center is loosely placed on two 3D-printed components (length R) that are directly attached to the servo motors. However, a magnetic coupling, consisting of a 5-by-5 mm NPM inside the movable part and a 40 - by - 5 mm NPM inside the lower TCM2 body part ensures that the index finger contact point is constrained to move along a single dimension, i.e., in height direction z. The difference between the angles α and β of SM1 and SM2, respectively, determine the height Δh of the contact point. The movement radius R and baseline b are determined by the TCM2 assembly, e.g., R = 5 cm and b = 5 cm. Note that the smaller radius r of the sliding points for both sides requires $r << R$.

Two adjacent height map values Δh_1 and Δh_2 are extracted from $\mathbf{I}_{\text{macro}}$ and set the servo motor angles using

$$\alpha = \frac{2}{\pi} \cdot asin\left(\frac{min(R, \Delta h_1)}{R}\right) \tag{5.5}$$

and

$$\beta = \frac{2}{\pi} \cdot asin\left(\frac{min(R, \Delta h_2)}{R}\right) \tag{5.6}$$

based on the maximum height difference of R. The sliding points of the servo motor end-effectors approximately shift as

$$b_1 \approx b + R \cdot cos(\alpha) \tag{5.7}$$

and

$$b_2 \approx b + R \cdot cos(\beta) \tag{5.8}$$

if $\Delta h << b$. The resulting height difference for the contact point follows from the rule of proportion (see red lines in Fig. 5.12, right) as

$$\Delta h_3 = \frac{|\Delta h_2 - \Delta h_1|}{b_1 + b_2} \cdot b_1 = \frac{R \cdot sin(\beta) - R \cdot sin(\alpha)}{2 \cdot b + R \cdot cos(\beta) + R \cdot cos(\alpha)} \cdot (b + R \cdot cos(\alpha)) \tag{5.9}$$

and the overall change in height

$$\Delta h = min(\Delta h_1, \Delta h_2) + |\Delta h_3| \tag{5.10}$$

which can be approximated as

$$\Delta h \approx min(\Delta h_1, \Delta h_2) + \frac{|\Delta h_2 - \Delta h_1|}{2} \tag{5.11}$$

for smaller differences between Δh_1 and Δh_2. Note that both servo motor angles can be changed simultaneously between 0 – 90^o to display the full range of height values at the cost of a narrower range of inclinations (height gradient) at the contact point for coarsely-structured material samples.

5.2.2.2 Macroscopic Roughness Rendering - Lateral Surface Obstacles

If a human slides the bare finger over the surface of a coarse material, the finger does not only feel changes in the contour height, but also tangentially caused by coarse surface obstacles such as gratings or holes. The macroscopic roughness images do not only indicate the height profile of a material sample, but also contain the edge positions of such coarse texture elements. A Canny edge detector ($\sigma = 4$, $\tau = 0.33$) implemented in the Matlab image processing toolbox extracts the most significant edges in $\mathbf{I}_{\text{macro}}$. The resulting images are dilated

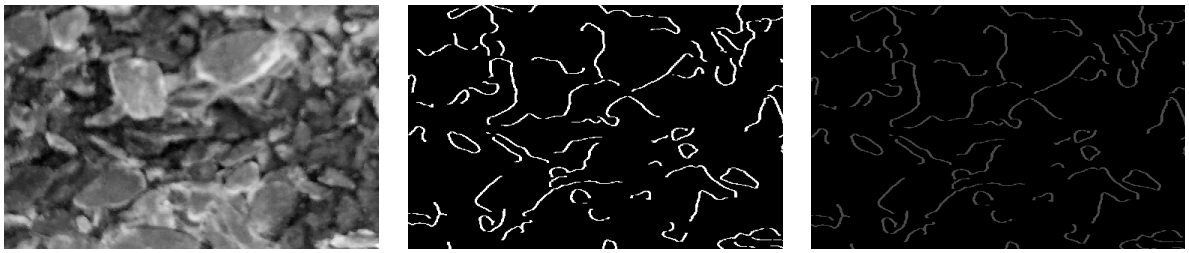


Figure 5.13: Tangential macroscopic roughness image generation $\mathbf{I}_{macro,tang}$. The grayscale converted surface images provide the positions of laterally-perceivable edges (left). A Canny edge detector (threshold = 0.33, sigma = 4) identifies relevant edges and a dilation filter (middle) broadens the edge width. The value range of the IR sensor data, normalized between 0 and 1, is then used to set the intensity of the dilated edges (right) and thereby determines the intensity of the tangential obstacle (adapted from [5] © 2018 IEEE).

using the Matlab function imdilate to broaden the edges of noticeable surface structures, leading to tangential macro images $\mathbf{I}_{\text{macro, tang}}$. Without dilation, the edges typically have the width of one pixel which is too small for a perceivable tangential obstacle effect. Figure 5.13

shows the tangential component of macroscopic roughness image generation. Note that the designer of the display application can further improve the resulting binary edge images and manually remove wrongly identified edges.

Both versions of the TCM2 superimpose the display of such surface obstacles using the friction display-related electromagnets from Section 5.2.1 underlining the specific correlation between friction and macroscopic roughness in this specific case, i.e., the pixel value $\mathbf{I}_{\text{macro, tang}}$ of the current's position height value is overlaid to the friction effect. It is worth highlighting that the designer of the application disables the tangential display of macroscopic roughness for clearly smooth and flat material samples.

5.2.3 Microscopic Roughness Rendering

Both versions of the TCM follow the same concept of displaying vibrotactile signals to the users hand based on VCAs. In the TCM1, the entire device is set into vibration by firmly mounting the VCA (NCM05-06-008-5JBL, H2W Technologies, USA) at the upper part of the movable body, whereas the TCM2 uses a C2 tactor (Engineering Acoustics, USA) to locally display the signals at the user's index finger. Thereby, the moving part of the C2 is decoupled from the device shape and only influenced by the user's finger weight. As introduced in Section 2.3.4, the two major methodologies direct playback as well as data-driven rendering are applied to display vibrotactile signals using such actuators. Data-driven approaches which were examined for pen-like tools are demanding in terms of data acquisition, and still do not capture the exact spatial properties of surface asperities, notably when anisotropic patterns like lines or gratings are present. Moreover, such examinations have not been conducted for bare-finger-based vibrotactile recordings, which are relevant for direct display at the human fingertip. Note that for the TCM2 several specific differences exist compared to tool-mediated setups, leading to the following general conditions for vibrotactile data acquisition guidelines.

1. The Texplorer2 vibrotactile data acquisition is robust to operator-induced force variations thanks to the bearings. By contrast, tool-mediated setups allow the operator to exert the weight of the entire arm on the material sample through the tool, and hence, the force variations are much larger and need to be considered. As a great advantage of a grounded device like the TCM1 and TCM2, the weight of the users arm can be completely compensated. Only the index finger may induce force variations which are well below the range of the human arm or hand applicable forces. Considering the scan speed, a grounded device further naturally constrains the movement due to the friction between the device and ground compared to freely wielding a tool, pen, or haptic stylus as in a tool-mediated setup [91].

2. Since the application designer can optimize the data acquisition procedure, it is advisable to separate the microscopic and macroscopic data acquisition as much as possible. Purely accelerometer-based setups for vibrotactile data acquisition cannot capture the exact position of coarse or large texture elements, and hence, such macroscopic patterns

are better captured with a fix-distanced camera and a distance sensor; if possible, vibrotactile data should be scanned on surface spots without such coarse particles or large structure elements. During display, both tactile information can be fused using different actuators, e.g., servo motors for macroscopic and VCAs for microscopic roughness rendering as achieved in the TCM2.

3. Weber et al. [34] examined bare-finger temporal coding of vibrotactile signals and identified the validity of time-warping the signals according to speed within the range of 40 mm/s – 120 mm/s. They clearly distinguish between spatial surface patterns which belong to macroscopic roughness, and temporal high-frequency spiking patterns which dilate or contract in time according to scanning speed belong to microscopic roughness. Hence, only a few vibrotactile scans per material are required to cover the speed dependency.

These acquisition considerations lead to further simplifications in the general conditions for vibrotactile signal modeling and display. The following sections describe these steps with respect to the **transfer function**, the **vibrotactile signal amplitude**, **vibrotactile signal characteristics**, and the microscopic roughness-based consideration of **anisotropic material surfaces**.

5.2.3.1 Voice Coil Actuator Transfer Function

The vibrotactile signals inherently contain MiRO^{TUM} and MiCO^{TUM}, and hence, do not require any additional rendering parameter. MiRO^{TUM} represents the perception frequency range of the Pacinian corpuscles (40 Hz – 1,000 Hz) [13] and MiCO^{TUM} is related to to the frequency range of the Meissner mechanoreceptors (8 Hz – 64 Hz) [13]. In order to display these frequency ranges properly, the transfer characteristics of vibrotactile actuators need to be taken into account. The C2 transfer function (TF), for example, has significant impact on the display of vibrotactile signals since low-frequency components are heavily damped which makes the display of MiCO^{TUM} impossible without adjustment.

This thesis examines the TF of the in-contact combination of the C2 and the human index finger. The transfer function is determined both using a 20-s-long chirp signal ranging from 1 Hz – 1,000 Hz using the configuration shown in Fig. 5.14. Note that the absolute part of the spectrum shown in Fig. 5.14 matches the Pacinian sensitivity curve determined by Israr et al. [280] which proves that the C2 actuator was designed to represent the Pacinian perception range. However, the range of the Meissner mechanoreceptors (8 Hz – 64 Hz) is heavily damped and needs to be enhanced if both channels are intended to be addressed. The inverse TF (ITF) is used to design a filter that adjust the signal which is only possible by highly emphasizing lower components and by damping larger frequency values. The ITF is clamped at five-times amplification of lower frequencies which reflects the average difference of the Meissner and Pacinian intensity values. All recorded material signals were adjusted to the ITF. Hence, vibrotactile sensations both related to MiRO^{TUM} and MiCO^{TUM} are transmitted via the C2 tactor during display. Figure 5.15 shows the difference without (b) and with (d) the adjustment (c) for band-limited white noise signal recording. Note that the

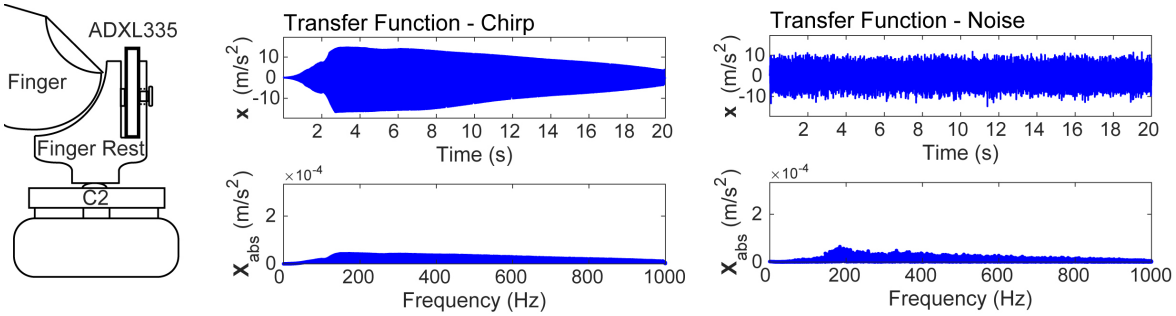


Figure 5.14: Transfer function setup which uses another ADXL335 to determine the C2 transfer function. The human index finger is placed on top of the C2 PLA component (glued to the end-effector), and hence, part of the system behavior. A chirp (envelope shown in middle plot) from 1 Hz – 1,000 Hz is sent to the C2, and the evaluation ADXL335 measures the resulting vibration, and its spectral response. The same step is repeated with band-limited Gaussian white noise, shown in the right plot (adapted from [3] © 2021 IEEE).

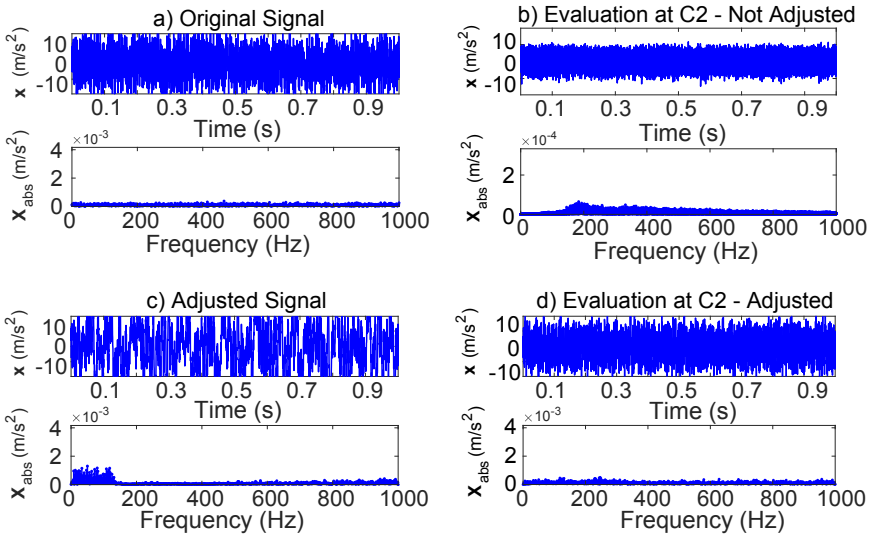


Figure 5.15: Noise input signal (a) measured at C2 tactor without (b) and with (d) ITF adjustment (c). Without adjustment, the C2 measurement reflects the ideal response for the Pacinian corpuscles only, but lower frequencies are almost completely damped, and hence, the Meissner corpuscles weakly addressed. Consequently, the signal is adjusted with the lower left correction, and hence leads to a roughly equal distribution of all frequencies in the lower right plot (adapted from [3] © 2021 IEEE).

adjusted noise signal leads to a desired noise-like spectrum (d, lower plot) when measured at the C2 tactor.

5.2.3.2 Vibrotactile Signal Amplitude

For tool-mediated devices like the PhantomOmni, the speed may range within 0 mm/s up to 240 mm/s and is generally clamped to 240 mm/s for higher input speed [183]. It is a valid expectation that a grounded device, notably the heavier version TCM2, is generally moved slower than a wearable haptic device or the light-weight end-effector of the PhantomOmni device, and hence, it has been observed that the user speed scatters around the speed range

of 0 mm/s – 120 mm/s. In order to infer a relation between input speed and signal amplitude, the Texplorer2 has been slid over several material samples and the vibrotactile signal energy was evaluated. Texplorer2 bare-finger scans are recorded at approximately equally-distanced speed levels in that range, and the mean of the absolute intensities (MSI) in the vibrotactile signal are calculated. The resulting MSI vector is min-max-normalized between 0 and 1 as shown in Fig. 5.16. Note that the MSIs of vibrotactile signals increase almost

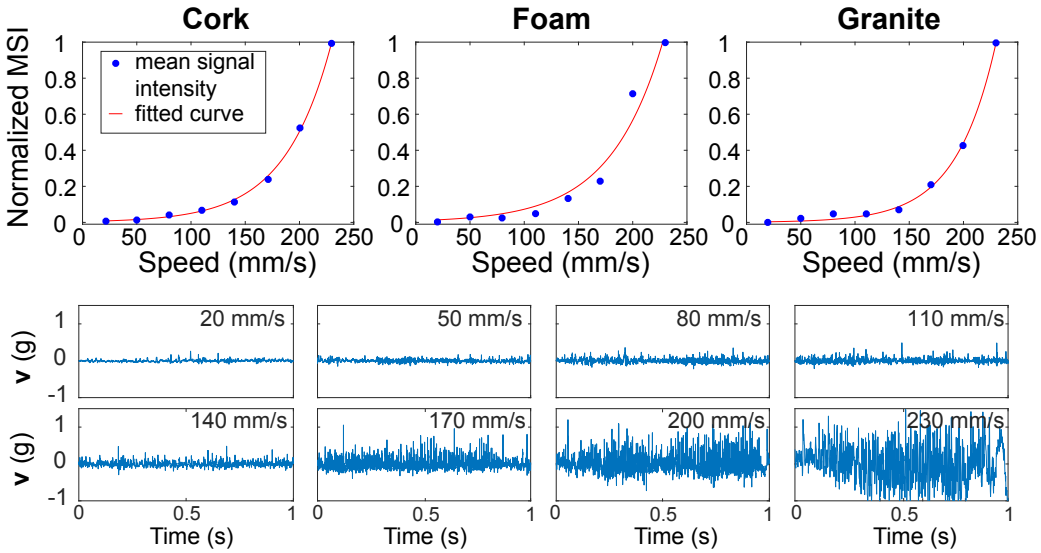


Figure 5.16: Mean signal intensities for different material samples from bare-finger Texplorer2 vibrotactile recordings. Note that the increase is linear for typical speed ranges (less than 150 mm/s) of humans, but becomes exponential for higher speed values. The lower eight plots show the bare-finger vibrotactile signals recorded for the granite material sample (adapted from [3] © 2021 IEEE).

linearly for lower speeds, but exponentially for speeds faster than 150 mm/s. Consequently, the gain of the C2 tactor is linearly scaled according to the generally lower TCM2 movement speed.

5.2.3.3 Vibrotactile Signal Characteristics

The frequency content of vibrotactile signals depends on the scan speed. Data-driven tool-mediated approaches [14], [91], [104] need to collect many vibrotactile signals recorded at different pairs of scan speed and force to account for this observation. However, the aforementioned reasons for grounded devices, notably the much smaller force dependency, lead to model simplifications. Further note that Culbertson et al. [201] identified that perceived microscopic roughness depends less on the exerted force. Additionally, the separation of x and y speed as conducted in [104] is not required since anisotropic material surface patterns (e.g., wood gratings) are better captured using images containing the exact positions of the pattern elements as will be explained in the next Section 5.2.3.4.

The perception of vibrotactile spike patterns depends linearly on the speed according to Weber et al. [34]. Their proposed speed scaling approach basically consists of collecting vibrotactile data at the speed of 80 mm/s and to linearly change the pitch to achieve invariant

speed perception. A pitch change of 0.5 then resembles a data trace of 40 mm/s, for example. This concept of vibrotactile signal warping is applied for the TCM2 which resembles a reasonable trade-off between scanning effort and perceptual relevance. Note that data-driven approaches are used to interpolate the vibrotactile signal content likewise, but only because the force also may change during the action. Thanks to the compensation of user force in the TCM2 such approaches are not required.

5.2.3.4 Anisotropic Microscopic Roughness Considerations

Anisotropic patterns, such as visible line patterns on a wooden plate, need further consideration regarding microscopic roughness display. The data-driven approach in [105] considers the scan force and both the x - and y scan speed to take anisotropic tactile properties into account. However, the exact positions of, e.g., perceivable gratings or holes, is not captured using such an approach and the amount of data required to model a material based on three dependencies heavily increase the storage conditions and scan time of such data. Consequently, the display of anisotropic patterns on a material surface is rendered differently.

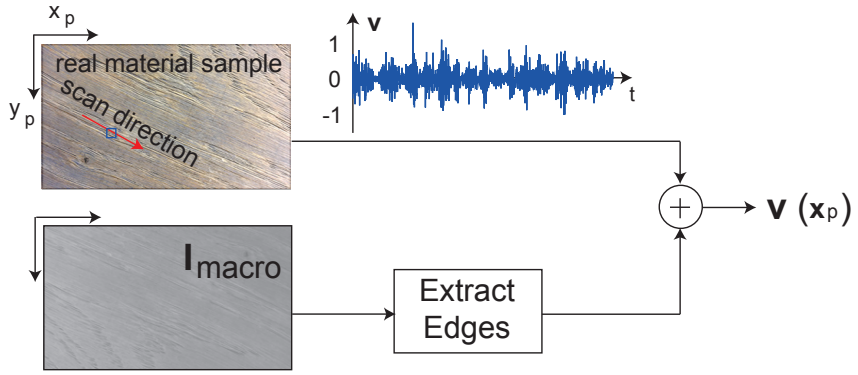


Figure 5.17: Consideration of anisotropic texture patterns. The real anisotropic material sample is scanned in its major direction and the resulting signal is added to the extracted edges of the macroscopic images during haptic display (adapted from [3] © 2021 IEEE).

Comparable to the approaches for variable-friction devices [282], directional microscopic-related sensations can be extracted from predefined haptic images. However, clearly isotropic materials, which constitute the majority of all materials, do not require the subsequent step. For anisotropic material samples, the real material sample is scanned in the major direction of surface patterns to collect a vibrotactile signal. Additionally, characteristic edges from the macroscopic roughness height profile images $\mathbf{I}_{\text{macro}}$ are extracted, and their pixel values are added elementwise to the vibrotactile signal at a specific on-screen position $\mathbf{x}_{\text{p}}$ during display as shown in Fig. 5.17. Following this approach, the vibrotactile signal amplitude is not changed for isotropic materials, however, fine gratings and lines in anisotropic surface images influence the vibrotactile signal amplitude depending on the current position of the TCM2. The inclusion of such spatial patterns is a notable improvement compared to data-driven approaches which do not consider any spatial information.

5.2.4 Hardness Rendering

The TCM1 only allows for a binary display of hardness sensation. On the one hand, completely stiff material surfaces can be displayed by using a blocking mechanism achieved by a servo motor. In this case, any stiff materials with a comparable stiffness to rigid PLA can be rendered. On the other hand, the TCM1 has a spring (k_s = 100 N/m) between the upper and lower body embedded to represent the majority of soft materials.

The simplistic TCM1 rendering of hardness was one of the major reason for the TCM2 development. The improved hardness rendering approach is based on the standard linear solid model (SLSM) from Section 2.3.1 which requires Young's modulus E and damping coefficient η to render virtual material hardness. The SLSM fully includes the hardness subdimensions compliance CCP^{TUM}, damping CDP^{TUM}, relaxation CRX^{TUM}, and yielding CYD^{TUM}. The binary feature yielding CYD^{TUM} originates from subjective opinions [2] describing whether a material sample stays deformed after interaction or not. If a material certainly stays deformed, the SLSM simplifies to the Maxwell model, consisting of a spring and dashpot in series, and hence, only requires CCP^{TUM} and CDP^{TUM} for hardness rendering. If a material is viscoelastic and recovers to its original shape, the stress-relaxation test of the SLSM provides all parameters which can be estimated from the Texplorer2 TU2 recordings during the pressing EP. Stress σ is expressed as force $F_{\text{TU2}} = \sigma \cdot A_{\text{TU2}}$ using the known contact area A_{TU2} of TU2. The device is swiftly pressed on various material samples, and hence, a constant ϵ_0 is imposed to measure the subsequent stress curves shown in Fig. 5.18. The material-depending viscosity constant η quantifies the damping behavior, but is gener-

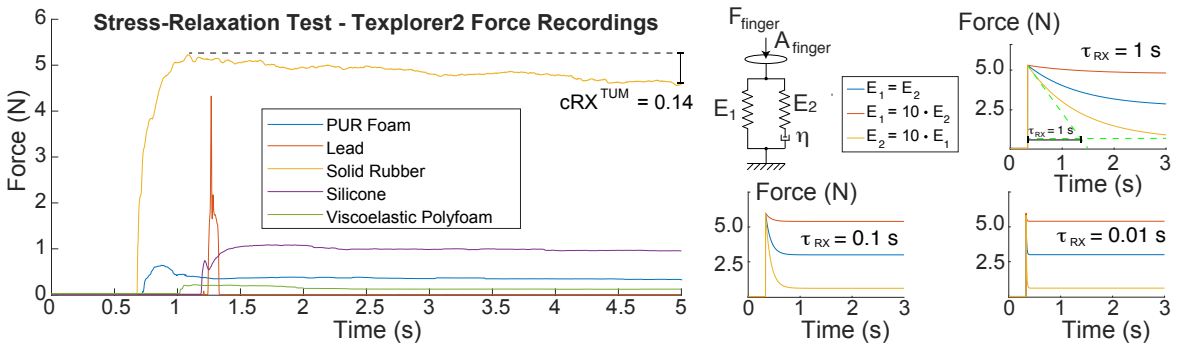


Figure 5.18: Recorded Texplorer2 stress-relaxation test data (top). Note that the SLSM can approximate any of these curves with E_1, E_2, and η, as shown on the top right plot for different relaxation times (adapted from [3] © 2021 IEEE).

ally reported for fluids only. However, it can be approximated from the relaxation time τ_{RX} [165] with

$$\eta = E_2 \cdot \tau_{\text{RX}} \tag{5.12}$$

The relaxation time is defined by the intersection of the initial force gradient and the sustained force [165], which is shown in the top right plot in Fig. 5.18 using dotted green lines.

Also note that the overall E is a constant tabular material value, but the values of the parallel E_1 and E_2 can differ and lead to different model behavior. Here, the CRX^{TUM} feature can indicate the ratio of E_1 and E_2. The lower right plots in Fig. 5.18 show the influence of

different linear combinations E_1 and E_2 for different relaxation times τ_{RX}. As an example, a rubber material sample has a relaxation feature value of CRX^{TUM} = 0.14, and consequently, E_1 needs to be about six-times larger than E_2 (red line) so that the sum E_1 + E_2 matches Young's modulus of the material. Additionally, τ is about 1 second resulting from the intersection of the green dotted lines indicating notable damping behavior.

With respect to the design of the TCM2, hardness feedback is overlaid to macroscopic roughness display by implementing an admittance-based control scheme which uses force sensor readings to control the angle, and hence the position, of servo motors. The C2 tactor is mounted on a flexible mechanical system which presses on the FSR. This unit is attached to a 4 centimeter-long lever arm connected to a servo motor (Reely RS-610WP MG, Conrad Electronic SE, Germany) and an additional bearing as shown in the left part of Fig. 5.19. The FSR continuously measures the index finger force to update the position of the servo

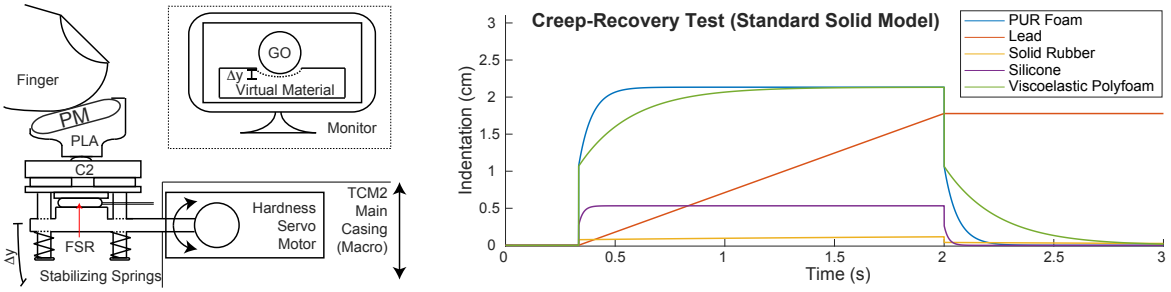


Figure 5.19: Admittance-based hardness display on the TCM2. The input force determines the position of the movable index finger placement. Local deform CDF^{TUM} is displayed visually on the computer monitor during indentation represented by the commonly used god-object (GO)[164] interaction point. Note that the material sample lead is fixed at one end, and hence, deformable (adapted from [3] © 2021 IEEE).

motor. Different to other setups, the TCM2 is able to overlay hardness sensations as well as macroscopic roughness patterns, since individual servo motors are used for each dimension.

As first simplification to display rigid, i.e., stiff materials, the servo motor actually moves against the human thumb if E is larger than 10^9 N/m^2 and a small user input force is detected. In this case, the touch point is held at a small angle of $\approx 5^o$ against the pressing direction to compensate for potential servo backlash. For viscoelastic materials, however, equation

$$\sigma + \frac{\eta}{E_2} \cdot \dot{\sigma} = E_1 \cdot \epsilon + \frac{\eta \cdot (E_1 + E_2)}{E_2} \dot{\epsilon} \tag{5.13}$$

from Section 2.3.1 leads to the required time-discrete (sample time T_s) equation

$$\boldsymbol{\epsilon}[n] = \frac{\boldsymbol{\sigma}[n] \cdot (1 + A) - A \cdot \boldsymbol{\sigma}[n-1] + B \cdot \boldsymbol{\epsilon}[n-1]}{E_1 + B} \tag{5.14}$$

with

$$A = \frac{\eta}{E_2 \cdot (n \cdot T_s)}, \quad B = \frac{\eta \cdot (E_1 + E_2)}{E_2 \cdot (n \cdot T_s)} \tag{5.15}$$

and is converted into the servo motor angle corresponding to the indentation. If the user presses on the hardness mechanism during haptic display, the pressing motion resembles the **creep-recovery test**, i.e., an input force leads to a specific indentation. The right part of

Fig. 5.19 shows the model-related indentations for the deformable materials in this thesis during about two seconds of pressing for a constant force of 1 N applied at a typical human index finger area of 75 mm^2. Note that the servo motor rotation speed and latency in the system low-pass filters the ideal model behavior during display.

The rendering of local deformation CDF$^{\text{TUM}}$ is achieved visually only. A single-point interaction display does not allow haptic rendering of multiple local forces to display a force gradient wrapping around the finger as conceptualized by Fishel et al. [66]. However, visual indentation and wrapping are displayed on a computer screen to represent the illusion of wrapping for soft material samples such as foam. Note that ongoing work in this field may further include smaller servo motors or comparable actuators that allow for a spatial wrapping mechanism around the finger tip.

5.2.5 Thermal Rendering

Peltier elements are embedded into the TCMs where a user generally places the index finger. For both TCM versions, the corresponding Peltier element is driven with a low-frequency PWM signal of 0.5 Hz, leading to an effective U_{Pelt} as the average value of the voltage on the Peltier element. Further note that the polarity can be switched ($U_{\text{Pelt,pol}}$) in order to enable a constant heating of the Peltier element surface. This effect is specifically relevant for thermally isolating materials, i.e., to simulate that the thermal energy of the finger accumulates between the skin and the Peltier element, or, to rapidly heat up the Peltier ceramic if the ambient temperature is significantly lower than the finger temperature. The thermal rendering models, however, differ significantly between TCM1 and TCM2, and are explained in the following.

5.2.5.1 Thermal Rendering - TCM1

The TCM1 implements a heuristic thermal rendering model without explicit temperature measurements based on the Peltier (TES1-127025) cooling/heating behavior when a constant voltage is set. The main reason is the form factor which does not allow for an additional sensor placement to obtain closed-loop rendering. Figure 5.20 show the measurements for different supply voltages on the TCM1 Peltier element based on the assumption that the Peltier element was in contact with the human finger and achieved approximately 32^{o}C surface temperature. Consequently, the time of applying the voltage U_{Pelt} generates different qualitative intensities of thermal conductivity $\lambda(S)$ (see Table 5.1) for the TCM1. Instead of using a timing approach, the effective Peltier voltage can be controlled using a PWM with

$$U_{\text{Pelt}}(\lambda(S)) = 3.3\text{V} \cdot max(1.0, \lambda(S)/\lambda_{\text{max}}) \tag{5.16}$$

by setting the ratio of the on–off duty cycle. This expression scales the Peltier voltage linearly up to a maximum cooling voltage of constant 3.3 V, representing the most strongly cooling material silver, which has a thermal conductivity of about $\lambda_{\text{silver}} = 500$ W/(m $\cdot$ K). The cooling effect is kept up for 10 seconds which is typical for static touch. Any other material sample thermal conductivity leads to a linearly scaled Peltier voltage in between 0.0 V – 3.3

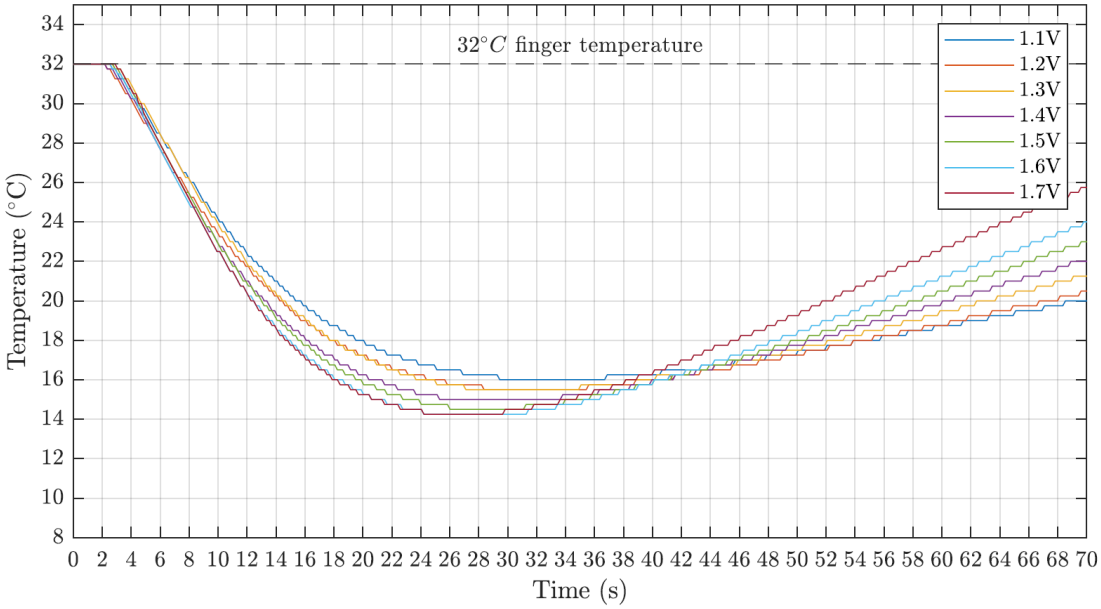


Figure 5.20: Different supply voltages set on the Peltier element of TCM1 and resulting cooling curves. Note that humans generally touch materials 5 – 10 seconds during the EP static touch which is guaranteed providing an appropriate heat sink.

V. An absolute thermal control can be achieved only by sensing the surface temperature of the Peltier element. The next subsection about the TCM2 thermal rendering introduces such a thermal control loop for the TCM2.

5.2.5.2 Thermal Rendering - TCM2

The TCM2 embodies the thermal contact model from Ho and Jones [29] which was introduced in Section 2.3.5. Their thermal semi-infinite contact model already includes the concepts to render the two thermal subdimensions TCO^{TUM} and TPR^{TUM}. The material constants thermal conductivity k, the warmth capacity c, and the density ρ have been collected and reported in corresponding tables [188] for various materials and used in this thesis for the TCM2 thermal rendering. As relevant base assumption, all material samples are selected to satisfy the Fourier number constraint from Section 2.3.5 in terms of material thickness and mass.

The left part of Fig. 5.21 shows the three thermal sensor (DS18b20, Maxim Integrated, USA) placements to monitor the Peltier element (QuickCool, QC-127-1.4-3.7AS, Quick-Ohm Kuepper & Co. GmbH, Germany). An appropriate heat sink (cooling body) and a 12 V-driven ventilator are added to provide sufficient cooling for common thermal modeling touch times of about 5 to 10 seconds [29]. When the TCM2 is powered on, the first measurement of the ambient temperature sensor (TS2) is used as ambient temperature, and consequently, as initial object temperature. When the human user starts touching the TCM2, and thereby, the thermal sensor, the initial skin temperature is obtained from TS2 as well. Note that a lower threshold of about 30°C is set [29] for all users if sensor values are missing or corrupted.

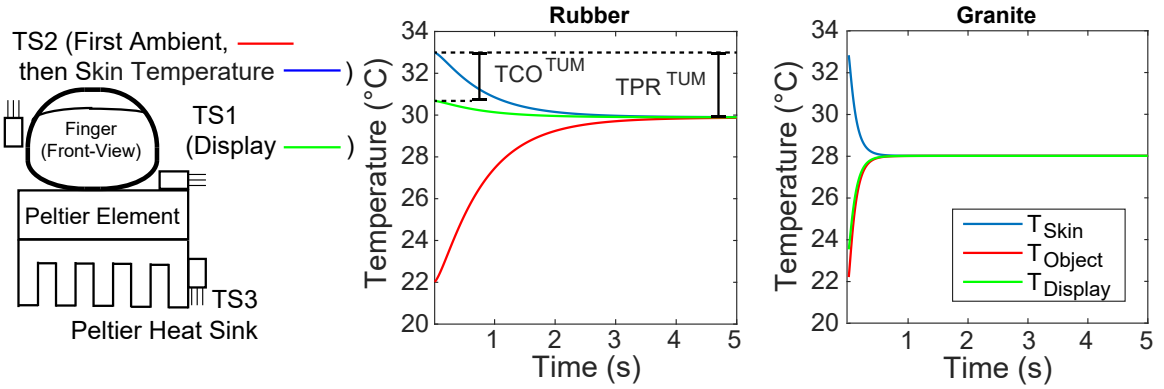


Figure 5.21: Warmth display setup (left) based on Ho and Jones [29] thermal model (right, shown for the first five seconds of the material rubber, middle plot, and granite, right plot). Note that TS2 first collects the ambient, and hence, object temperature when powered on, and afterwards the human index finger is in contact with it. During the rendering application, it hence obtains approximately the initial human skin temperature. TCO^{TUM} is quantified by the initial temperature gap between the display and the index finger, and TPR^{TUM} the persistence of this temperature gradient (adapted from [3] © 2021 IEEE).

A third sensor TS3 is used to monitor the critical temperature increase of the Peltier heat sink to avoid overheating and subsequent undesired heat flow to the other side. Notably, Gabardi et al. [158] have shown the tremendous impact if the cooling is malfunctioning, and hence, additional sensing is required to prevent overheating of the Peltier element. Thanks to the form factor provided by a grounded device, the cooling capabilities are sufficient to sustain the cooling effect robustly up to 10 seconds, which are typical for human interaction with material surfaces [29], and can be extended if required.

The middle and right part of Fig. 5.21 further show two example thermal model graphs of rubber and granite, respectively. Depending on the current ambient, and hence initial object temperature as well as the initial skin temperature, the different thermal properties of both materials lead to characteristic display temperatures. The Peltier element ultimately needs to be controlled to match the theoretical display temperature (green curve in Fig. 5.21) as closely as possible, which is achieved by a two-point thermal controller. TCO^{TUM} is interpreted as the temperature difference of the skin and the initial T_{disp} that the Peltier element is supposed to display initially, and TPR^{TUM} can be understood as the ongoing difference in between the two temperature curves. Note that both features are inherently determined by the tabular coefficients k, the warmth capacity c, and the density ρ, and hence, characteristic for each material. However, the thickness of the rendered virtual material sample should be chosen sufficiently large to satisfy the Fourier number constraint from Section 2.3.5, and surface asperities such as gratings and gaps are currently not covered by this approach and are subject to future work in thermal rendering.

5.2.6 Adhesive Tack

The C2 finger assembly contains one neodymium permanent magnet as mentioned in Section 5.1.2.2, which constitutes one component of ATK^{TUM} rendering. The other component

is implemented using a separate small device which is worn by the human user and contains a servo motor with a rotating PLA construction as shown in Fig. 5.22. A stack of three

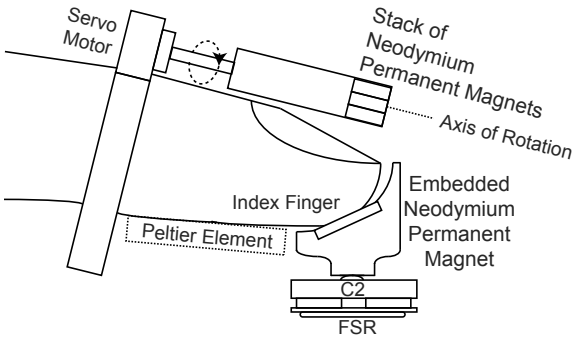


Figure 5.22: ATK$^{\text{TUM}}$ display based on currently exerted force. If a material is adhesive, the exerted pressure on the FSR is translated into a rotation of the PM stack. In its maximum, the neodymium magnets are aligned in their maximum magnetic adhesion force direction. When the finger is lifted, the servo motor swiftly returns in its initial position, and hence, stops to display adhesion forces (adapted from [3] © 2021 IEEE).

neodymium permanent magnets is embedded into this rotatable component. Note that this device is attached to the point of the index finger so that the Peltier elements still has the entire contact area for thermal rendering. The angle of the servo motor determines the adhesive forces between the rotating construction and the neodymium stack. The force sensor (see hardness Section 5.2.4) is used to further control the adhesive forces depending on the user pressure. As input source, subjective ratings of adhesive tack ATK$^{\text{TUM}}$ [2], normalized between 0 and 1, are used to parameterize adhesive tack. It is worth highlighting that the majority of naturally occurring materials are not noticeably adhesive at all. However, if a surface is adhesive, the adhesion forces are related to the effective interaction area between the fingertip and surface, and the interaction force. Note that the contact area for bare finger-material surface interaction likewise depends on the interaction force and has been empirically identified by Hiramatsu et al. [283] for the human index finger to be

$$A_{\text{finger}} = a \cdot F^b \tag{5.17}$$

with a = 60 mm^2/N and $b = 0.4$. For example, the contact area of the index finger ranges within 0 mm^2 – 105 mm^2 corresponding to 0 N – 4 N. The current force **f**[n] is hence mapped to the servo motor angle ranging between 0$^{\text{o}}$ – 90$^{\text{o}}$ based on the ATK$^{\text{TUM}}$ feature as

$$\boldsymbol{\alpha}_{\text{ATK}}[n] = ATK^{\text{TUM}} \cdot 90^{\text{o}} \cdot \frac{A_{\text{finger}}(\mathbf{f}[n])}{105\ \text{mm}^2} \tag{5.18}$$

The maximum rotation of 90$^{\text{o}}$ corresponds to the servo motor angle where the NPMs are best aligned, and hence, the magnetically-induced adhesion is strongest. The maximum displayable adhesion force was measured as 2N $\pm$ 0.3N during several measurements with a force meter. Note that this additionally worn device does not simulate adhesion locally at the finger-surface interface point. Such realistic adhesion could only be recreated by embedding actuators at or even inside the index finger pad, and is subject to future work and development in (miniaturized) actuator design.

5.3 Subjective TCM1 Experiment

A subjective experiment was conducted in [5] to evaluate the physical rendering capabilities of the TCM1.

5.3.1 Experimental Setup and Procedure

Twenty subjects, 6 females and 14 males aged between 18 and 40, participated in this study. Each subject sat at a desk in front of a computer connected to the TCM1. On the left-hand side the ten real material (RM) samples from Table 5.1 were placed. This arrangement allowed the subjects to freely examine the real and rendered virtual materials simultaneously for the purpose of comparison. The virtual material (VM) display application was used to randomly present virtual materials as black images, i.e., absent visual surface texture information, on the computer monitor. An introduction of the experiment and training for TCM1 handling was given to each participant. The subjects were asked to rate the similarity between the displayed virtual material and each real material on a scale ranging from dissimilar (1) to similar (5) while all virtual materials were displayed in a random order to avoid order effect biases. The subjects were instructed to consider all five tactile dimensions [63] for the displayed virtual materials using the TCM1 and the real materials before providing the rating.

5.3.2 Experimental Results

Figure 5.23 shows the averaged confusion matrix $\mathbf{C}_{\text{all}}$ resulting from the 20 individual subject confusion matrices. Note that the similarity ratings (between 1 - 5) were normalized (per row) as probabilities. Even if the virtual material was marked as 5 (similar) to the real material counterpart, but other virtual materials received a rating greater than 1, the overall probability for the main diagonal in the $\mathbf{C}_{\text{all}}$ was lower than 1.0, i.e., lower than 100%.

Based on the averaged probabilities in Fig. 5.23, the accuracy (mean confidence) and per-class precision (true positive rates) indicate the tactile rendering performance of the TCM1. Given the ten materials and the 20 subjects, an average rendering accuracy of 74.10% is achieved. Table 5.2 lists the per-material class precisions from the confusion matrix (Fig. 5.23).

Table 5.2: TCM1 per-class precisions (recognition probability) from confusion matrix. Virtual materials like stainless steel were easily identified and matched to the real material counterpart, whereas wooden materials such as compressed wood and oak were partly confused.

	Scouring Foam	Granite	Crushed Rock	Stainless Steel	Compr. Wood
$Prec\,(\%)$	73.70	75.80	66.90	100.00	56.10
	Profiled Wood	Oak	Profiled Rubber	Profiled Foam	Sandpaper
$Prec\,(\%)$	76.70	61.30	73.60	81.20	75.70

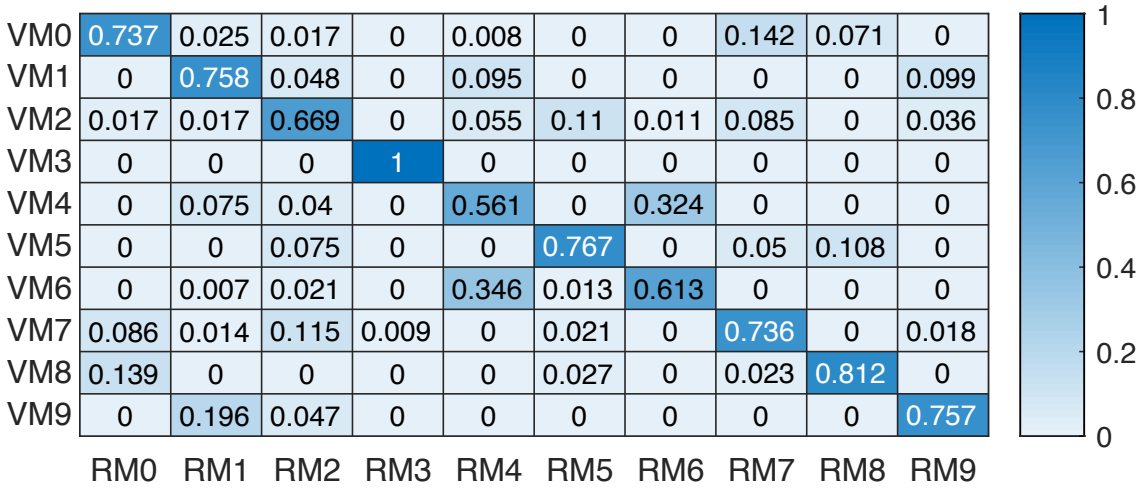


Figure 5.23: Averaged virtual material (VM) – real material (RM) confusion matrix based on 20 subjects' confusion matrices. The value 1.0 indicates 100% certainty for all subjects in identifying a virtual material sample as similar to the real material using the TCM1. The material sample names are RM0 scouring pad, RM1 granite, RM2 crushed rock, RM3 stainless steel, RM4 compressed wood, RM5 profiled wood, RM6 oak, RM7 profiled rubber, RM8 profiled foam, RM9 sandpaper (adapted from [5] © 2018 IEEE).

5.3.3 Discussion

As a conclusion of this TCM1-based study, the subjects were generally able to match the virtual material representations to the real material samples, which was considered as promising step toward tactile rendering based on the five major dimensions. However, the chosen set of materials was mainly characteristic in terms of roughness sensation. Most importantly, more complex hardness and warmth properties are not displayable with the TCM1, and hence, the selection of the study was deliberately made to include eight rigid materials, and two similarly soft pieces of foam. The subsequent experiment section about the TCM2 examined partly the same material samples, but also includes material samples that are more diverse in their haptic properties, e.g., viscoelasticity or adhesiveness. Also, the limitations of the TCM1 needed to be overcome, which was the major reason for the development of the TCM2.

5.4 Subjective TCM2 Experiment

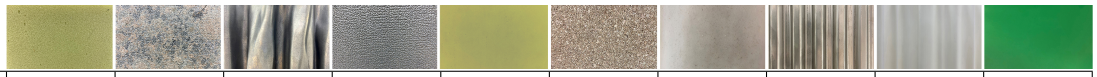
As conducted for the preceding TCM1 in [5], a comparable subjective experiment was performed in [3] to show that users can match virtual to real material samples absent visual information using the TCM2.

5.4.1 Experimental Setup and Procedure

Fifteen subjects were asked to participate in the experiment. A set of ten different material samples, shown in Table 5.3, were selected which are representative for the fifteen SynTouch

subdimensions. The model parameters from Table 5.3 provide the input data space to drive the actuators of the TCM2. Young's modulus and the thermal properties originate from corresponding tables [188], the other parameters were determined using the Texplorer2 units. The overall objective of the experiment was to identify, which currently displayed virtual

Table 5.3: Rendering parameters [186], [284] for ten selected materials (RM0 – RM9) used in the subjective experiment (reproduced from [3] © 2021 IEEE).



	Foam	Granite	Lead	Rubber	Visco.F.	Cork	Silicon	Steel	PLA	PMMA
μ_0	0.81	0.62	0.77	0.82	0.86	0.59	0.97	0.90	0.78	0.94
μ_k	0.77	0.61	0.70	0.65	0.66	0.56	0.88	0.75	0.75	0.84
ATK	0	0	0	0	0	0	1.0	0	0	0.75
E	0.003	6.9	12.5	0.04	0.002	1.0	0.005	189	1.1	2.24
τ_{RX}	0.03	0	0.001	1.20	0.75	0	0.01	0	0	0
CRX	0.4	0	0.98	0.14	0	0	0.10	0	0	0
CYD	0	0	1	0	0	0	0	0	0	0
ρ	38	2,640	10,000	920	70	120	1,300	7,600	1,010	1,160
k	0.03	1.6	35.0	0.05	0.04	0.04	0.2	14.0	0.14	0.17
c	1,210	820	130	1,880	1,300	1,800	1,200	400	1,400	1,470

material (VM) representation is most similar to each of the ten real material (RM) samples. In order to avoid any assumptions or user material knowledge bias, the real material samples were hidden in a special assembly. However, visual schematics and tangible landmarks indicated the position and the number of each real material sample, and the supervisor of the experiment ensured that the subjects always correctly located the hidden material samples.

The fifteen SynTouch subdimensions were first introduced in their original definition from Fishel et al. [66]. Afterwards, the hidden real material samples were freely explored by the users to obtain a perceptual understanding of the subdimensions. The participants were then introduced to the TCM2 rendering capabilities, and likewise asked to test the virtual material samples in the TCM2 rendering application. The virtual material samples were visually represented by a uniformly black texture image at a computer screen.

During the experiment, the virtual material samples were randomly permuted. Each user was asked which of the real material samples corresponds to which virtual material representation. The results were noted down by the supervisor, leading to a confusion matrix $\mathbf{C}_{\text{subj}}$ for each subject comparable as to the TCM1 experiment. Likewise, fractions were allowed to consider uncertainty, i.e., the subject was allowed to name more than one matching material sample and the corresponding probabilities.

After the purely tangible part of the experiment, the real material samples were uncovered and the virtual material representations regained their true visual display image on the computer screen. The subjects were asked to rate the RM – VM overall similarity (including visual and haptic properties) from 1 (not similar at all) to 7 (very similar/identical) to obtain a visual-haptic similarity vector $\mathbf{s}$, which further describes the rendering performance of the TCM2 using also visual cues.

5.4.2 Experimental Results

Figure 5.24 shows the averaged confusion matrix $\mathbf{C}_{\text{all}}$ resulting from the 15 individual subject confusion matrices. Based on the averaged probabilities in Fig. 5.24, the accuracy (mean

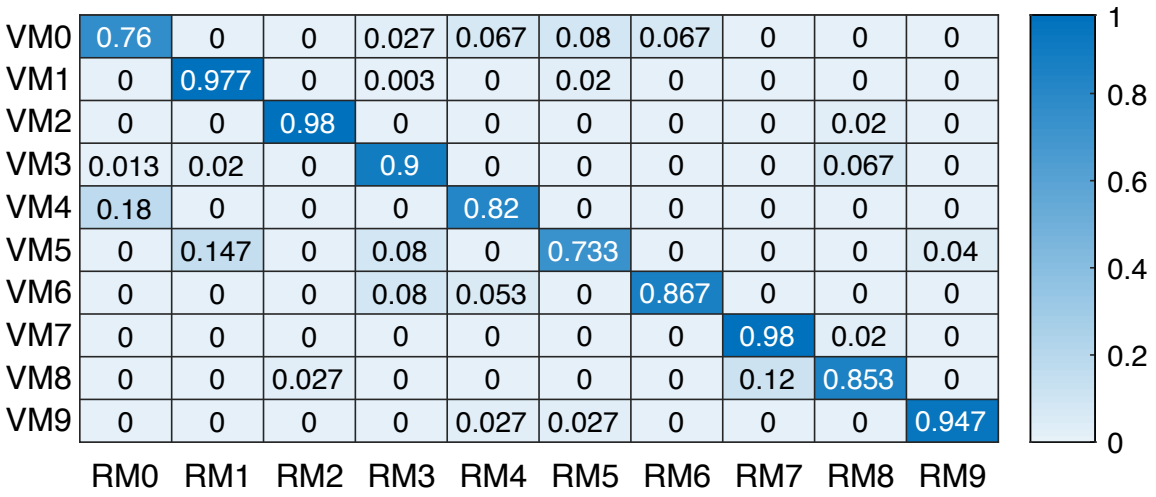


Figure 5.24: Averaged virtual material (VM) – real material (RM) confusion matrix based on 15 subjects' confusion matrices. The value 1.0 indicates 100% certainty for all subjects in identifying a material sample displayed at the TCM2. The material sample names are RM0 foam, RM1 granite, RM2 lead, RM3 rubber, RM4 viscoelastic foam (memoryfoam), RM5 cork, RM6 silicon, RM7 stainless steel, RM8 PLA, RM9 PMMA (reproduced from [3] © 2021 IEEE).

confidence) and per-class precision (true positive rates) indicate the tactile rendering performance of the TCM2. Given the ten materials and the 15 subjects, an average accuracy of 88.17% and visual-haptic similarity (VH-SIM) score of 6.2 are achieved. Table 5.4 shows the per-material class precisions from the confusion matrix (Fig. 5.24) and the visual-haptic similarity scores averaged across all subjects and their 95% two-sided confidence intervals

$$x_{\text{l,u}} = \overline{x} \pm z_{\text{l,u}}(\alpha) \cdot \frac{\sigma}{\sqrt{N}} \tag{5.19}$$

with N = 15, α = 0.05, and the corresponding standard normal distribution values $z_{\text{l,u}}(\alpha)$ = $\pm$ 1.96.

Table 5.4: Experimental results averaged across subjects. Top row: per-class precisions from confusion matrix. Bottom row: per-class virtual–real visual-haptic similarity ratings.

	Foam	Granite	Lead	Rubber	Visco.F.
$Prec\ (\%)$	76.00	97.70	98.00	90.00	82.00
$VH-Sim\ (1-7)$	5.8 ± 0.44	6.7 ± 0.26	5.9 ± 0.48	6.7 ± 0.38	6.4 ± 0.29
	Cork	Silicon	Steel	PLA	PMMA
$Prec\ (\%)$	73.30	86.70	98.00	85.30	94.61
$VH-Sim\ (1-7)$	5.8 ± 0.51	6.3 ± 0.53	6.5 ± 0.47	6.0 ± 0.52	6.2 ± 0.51

Since 15 subjects performed the experiment, each virtual – real comparison (VM – RM) is considered as vector containing 15 entries representing one subject's rating. For example, the ratings of the pair VM0 – RM0 lead to the average value 76%, or 0.76, from the confusion

matrix in Fig. 5.24. One-way Anova in a multiple comparison procedure is applied on all relevant pairs of vectors VM_i – RM_j ($i \neq j$ and $i, j \in \{0...9\}$) to check whether the subject ratings are statistically significant, leading to 45 comparisons, and hence, 45 p values. If α = 0.01, the critical F-score is (F(1,28) = 7.636, $p < 0.01$) which leads to the rejection of the null hypothesis that the ratings are not different. Bonferroni correction is applied on α. For all relevant comparisons in this context the null hypothesis can be rejected according to the pairwise comparisons, which shows that the results are statistically significant.

As overall conclusion, the subjects can distinguish between the more diverse material representations generated by the TCM2 due to the high per-class confusion matrix results, and moreover, the TCM2 appears to appropriately display material properties when visual information is provided additionally to the touch experience. The few confusions are discussed in the following.

5.4.3 Discussion

The confusion matrix in Fig. 5.24 shows a per-class precision of 98% for the steel grating in the best case, but a significantly lower result for foam (76%), which can be explained by the rigid interface between the user and the TCM2. Currently the user touches the neodymium permanent magnet, which is well-suited for the representation of rigid or slightly compressible material representations. If the application focus is set on the display of softer materials such as foams or fabrics, however, an intermediate soft layer potentially can increase the rendering realism.

Cork was partly misclassified as granite tile, which presumably results from the comparable roughness sensation and the fact that the thermal cooling and persistence was neglected by several subjects. Another confusion stems from the irregularity pattern of the material samples lead and PLA. It seems difficult for subjects to memorize an exact surface profile, however, other tactile subdimensions such as the different thermal characteristics acted as tiebreakers for the majority of the subjects who correctly distinguished both samples. Notably, specific subdimensions seem to be dominant depending on the material sample; e.g., the hardness subdimensions are highly characteristic for memory foam (viscoelastic foam), or, tactile stiction, adhesion, and sliding resistance for silicone and PMMA. It is important to mention that fifteen subdimensions cover a large display space; the 10 selected materials represent distinctive points in this space, yet several more material samples should be investigated in future work.

5.5 Chapter Summary

This chapter introduced the novel tactile computer mouse TCM1 and its final version TCM2. The data traces recorded with the Texplorer2 device and tabular values were adjusted to drive their actuators which simulate various perceptual tactile dimensions. Rendering models adapted from related work and adaptations have been presented and subjective experiments have shown that the virtual material representations can be matched to their real counterparts based on the display capabilities of both TCMs. Notably, the interpretations

of the fifteen tactile subdimensions proposed by Fishel et al. [66] can be independently displayed using the TCM2, and it was observed that specific subdimensions significantly helped in distinguishing between the material samples.

Chapter 6

Conclusion and Future Work

The presented tactile data acquisition and rendering approaches enable the parametrization and display of haptic material properties to a human user. Possible future applications range from realistic online-shopping applications to quality control and immersive virtual reality applications. This dissertation presented a complete pipeline comprising the labeling, scanning, modeling, and virtual displaying steps for a large set of real material classes toward these objectives.

6.1 Summary of the Results

Chapter 3 introduced the novel material taxonomy and the Texplorer data acquisition framework. The taxonomy allows for an unambiguous labeling of the materials in the database as well as to categorize novel material samples. Each multimodal material recording consists of 20 different data traces, which underline the complexity of material analysis. These data traces were then passed to a supervised machine learning classification pipeline in Chapter 4, and the most successful features were identified. This thesis further showed that a late fusion of the handcrafted features, a texture image-based CNN, and a neural network of the vibrotactile signal spectra obtained a classification performance of 95% accuracy, and also achieved a high similarity in terms of the proposed taxonomy of the current LMT 184 database.

The multimodal data and a subset of the tactile dimension features were then used as underlying model for haptic display as addressed in Chapter 5. The novel display devices TCM1 and TCM2 were conceptualized and engineered to display the tactile (sub)dimensions to a human user. These devices constitute valid attempts to fully display all of the tactile (sub)dimensions simultaneously, and subjective experiments have shown that users can identify virtual material representations absent visual information.

6.2 Limitations and Future Work

This thesis does not answer all questions related to the research field of haptic material property acquisition, modeling, and display. A few suggestions for ongoing work as well as open questions are discussed in the following.

The number of currently scanned materials is larger than any comparable haptic database to date, but still constitutes a fraction of all possible materials and their different processing shapes. This fact is, for example, demonstrated by the sheer number of types of trees (i.e., woods), which are roughly estimated to go beyond 10,000, or similarly, the number of man-made plastics and technical ceramics is likewise large and still growing. Nevertheless, by using the proposed setup and methodology in this thesis, further material samples can be scanned in order to extend the database accordingly. It is also worth mentioning and a very relevant research question how many of these material representations can be reduced to the same haptic representation that at least feel similar, and hence, would simplify this challenge for the second part addressed in this thesis, i.e., the tactile rendering of haptic impressions.

The TCM2 is currently able to display the fifteen tactile subdimensions at the index finger. It would be desirable to extend this approach to the entire human hand to further improve the touch experience. In this context, and as an even more significant limitation at this moment, the achieved rendering realism is strongly depending on the used actuators. Notably the vibrotactile rendering and the interface between the human skin and vibrating element is presumably improved by future actuator research and miniaturization. Comparable to other domains, e.g., force rendering in kinesthetic teleoperation, the relevant models generally precede their device-related implementations and mechanical solutions. Consequently, chances are high that upcoming approaches can take advantage of miniaturized actuators with improved rendering capabilities that also combine multiple tactile subdimensions. Lastly, the realistic display of adhesiveness still remains a challenging task with currently available actuators.

Bibliography

Publications by the author

Journal Publications

[1] M. Strese, C. Schuwerk, A. Iepure, and E. Steinbach, "Multimodal feature-based surface material classification", *IEEE Transactions on Haptics*, vol. 10, no. 2, pp. 226–239, 2016. DOI: 10.1109/TOH.2016.2625787.

[2] M. Strese, L. Brudermueller, J. Kirsch, and E. Steinbach, "Haptic material analysis and classification inspired by human exploratory patterns", *IEEE Transactions on Haptics*, vol. 13, no. 2, pp. 404–424, 2019. DOI: 10.1109/TOH.2019.2952118.

[3] M. Strese, X. Xu, and E. Steinbach, "TCM2: A surface material display device using a fifteen dimensional tactile feature space", *IEEE Transactions on Human-Machine Systems (under submission)*, 2021.

[4] E. Steinbach, M. Strese, M. Eid, X. Liu, A. Bhardwaj, Q. Liu, M. Al-Ja'afreh, T. Mahmoodi, R. Hassen, A. El Saddik, *et al.*, "Haptic codecs for the tactile internet", *Proceedings of the IEEE*, vol. 107, no. 2, pp. 447–470, 2018. DOI: 10.1109/JPROC.2018.2867835.

[5] M. Strese, R. Hassen, A. Noll, and E. Steinbach, "A tactile computer mouse for the display of surface material properties", *IEEE Transactions on Haptics*, vol. 12, no. 1, pp. 18–33, 2018. DOI: 10.1109/TOH.2018.2864751.

[6] H. Zheng, L. Fang, M. Ji, M. Strese, Y. Özer, and E. Steinbach, "Deep learning for surface material classification using haptic and visual information", *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2407–2416, 2016. DOI: 10.1109/TMM.2016.2598140.

Conference Publications

[7] M. Strese, Y. Boeck, and E. Steinbach, "Content-based surface material retrieval", in *IEEE World Haptics Conference*, 2017.

[8] M. Strese and E. Steinbach, "Toward high-fidelity haptic interaction with virtual materials: A robotic material scanning, modelling, and display system", in *IEEE Haptics Symposium*, 2018.

[9] M. Strese, C. Schuwerk, and E. Steinbach, "Surface classification using acceleration signals recorded during human freehand movement", in *IEEE World Haptics Conference*, 2015.

[10] M. Ji, L. Fang, H. Zheng, M. Strese, and E. Steinbach, "Preprocessing-free surface material classification using convolutional neural networks pretrained by sparse autoencoder", in *IEEE 25th International Workshop on Machine Learning for Signal Processing*, 2015.

[11] M. Strese, C. Schuwerk, and E. Steinbach, "On the retrieval of perceptually similar haptic surfaces", in *Seventh International Workshop on Quality of Multimedia Experience*, 2015.

[12] M. Strese, J.-Y. Lee, C. Schuwerk, Q. Han, H.-G. Kim, and E. Steinbach, "A haptic texture database for tool-mediated texture recognition and classification", in *IEEE International Workshop on Haptic Audio-Visual Environments and Games*, 2014.

General Publications

[13] S. J. Lederman and R. L. Klatzky, "Haptic perception: A tutorial", *Attention, Perception, & Psychophysics*, vol. 71, no. 7, pp. 1439–1459, 2009.

[14] K. J. Kuchenbecker, J. Romano, and W. McMahan, "Haptography: Capturing and recreating the rich feel of real surfaces", in *Robotics Research*, Springer, 2011, pp. 245–260.

[15] M. Strese, C. Schuwerk, R. Chaudhari, and E. Steinbach. (2020). Haptic texture database, [Online]. Available: `http://zeus3.lmt.ei.tum.de/downloads/texture/`.

[16] M. A. Srinivasan and C. Basdogan, "Haptics in virtual environments: Taxonomy, research status, and challenges", *Computers & Graphics*, vol. 21, no. 4, pp. 393–404, 1997.

[17] M. O. Ernst and H. H. Buelthoff, "Merging the senses into a robust percept", *Trends in Cognitive Sciences*, vol. 8, no. 4, pp. 162–169, 2004.

[18] E. H. Weber, *Tastsinn und Gemeingefühl*, 149. W. Engelmann, 1905.

[19] L. A. Jones, "Matching forces: Constant errors and differential thresholds", *Perception*, vol. 18, no. 5, pp. 681–687, 1989.

[20] ——, "Kinesthetic sensing", in *Human and Machine Haptics*, Citeseer, 2000.

[21] L. A. Jones and I. W. Hunter, "A perceptual analysis of stiffness", *Experimental Brain Research*, vol. 79, no. 1, pp. 150–156, 1990.

[22] ——, "A perceptual analysis of viscosity", *Experimental Brain Research*, vol. 94, no. 2, pp. 343–351, 1993.

[23] J. Morley, A. Goodwin, and I Darian-Smith, "Tactile discrimination of gratings", *Experimental Brain Research*, vol. 49, no. 2, pp. 291–299, 1983.

[24] L. Kruger, M. P. Friedman, and E. C. Carterette, *Pain and touch*. Elsevier, 1996.

[25] J. M. Wolfe, K. R. Kluender, D. M. Levi, L. M. Bartoshuk, R. S. Herz, R. L. Klatzky, S. J. Lederman, and D. M. Merfeld, *Sensation & perception*. Sinauer Sunderland, MA, 2006.

[26] S. J. Lederman, "Tactile roughness of grooved surfaces: The touching process and effects of macro-and microsurface structure", *Perception & Psychophysics*, vol. 16, no. 2, pp. 385–395, 1974.

[27] S. J. Lederman, G. Thorne, and B. Jones, "Perception of texture by vision and touch: Multidimensionality and intersensory integration.", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 12, no. 2, p. 169, 1986.

[28] L. A. Jones and S. J. Lederman, *Human hand function*. Oxford University Press, 2006.

[29] L. A. Jones and H.-N. Ho, "Warm or cool, large or small? The challenge of thermal displays", *IEEE Transactions on Haptics*, vol. 1, no. 1, pp. 53–70, 2008.

[30] A. Goodwin, V. Macefield, and J. Bisley, "Encoding of object curvature by tactile afferents from human fingers", *Journal of Neurophysiology*, vol. 78, no. 6, pp. 2881–2888, 1997.

[31] J. R. Phillips and K. O. Johnson, "Tactile spatial resolution II. Neural representation of bars, edges, and gratings in monkey primary afferents", *Journal of Neurophysiology*, vol. 46, no. 6, pp. 1192–1203, 1981.

[32] K. Sathian, A. Goodwin, K. John, and I. Darian-Smith, "Perceived roughness of a grating: Correlation with responses of mechanoreceptive afferents innervating the monkey's fingerpad", *Journal of Neuroscience*, vol. 9, no. 4, pp. 1273–1279, 1989.

[33] C. E. Connor and K. O. Johnson, "Neural coding of tactile texture: Comparison of spatial and temporal mechanisms for roughness perception", *Journal of Neuroscience*, vol. 12, no. 9, pp. 3414–3426, 1992.

[34] A. I. Weber, H. P. Saal, J. D. Lieber, J.-W. Cheng, L. R. Manfredi, J. F. Dammann, and S. J. Bensmaïa, "Spatial and temporal codes mediate the tactile perception of natural textures", *Proceedings of the National Academy of Sciences*, vol. 110, no. 42, pp. 17 107–17 112, 2013.

[35] R. J. Lundström, "Responses of mechanoreceptive afferent units in the glabrous skin of the human hand to vibration", *Scandinavian Journal of Work, Environment & Health*, pp. 413–416, 1986.

[36] H. T. Nefs, A. M. Kappers, and J. J. Koenderink, "Amplitude and spatial-period discrimination in sinusoidal gratings by dynamic touch", *Perception*, vol. 30, no. 10, pp. 1263–1274, 2001.

[37] K. O. Johnson, "The roles and functions of cutaneous mechanoreceptors", *Current Opinion in Neurobiology*, vol. 11, no. 4, pp. 455–461, 2001.

[38] I. Birznieks, V. G. Macefield, G. Westling, and R. S. Johansson, "Slowly adapting mechanoreceptors in the borders of the human fingernail encode fingertip forces", *Journal of Neuroscience*, vol. 29, no. 29, pp. 9370–9379, 2009.

[39] H. E. Wheat, L. M. Salo, and A. W. Goodwin, "Human ability to scale and discriminate forces typical of those occurring during grasp and manipulation", *Journal of Neuroscience*, vol. 24, no. 13, pp. 3394–3401, 2004.

[40] R. S. Johansson, U. Landstro, and R. Lundstro, "Responses of mechanoreceptive afferent units in the glabrous skin of the human hand to sinusoidal skin displacements", *Brain Research*, vol. 244, no. 1, pp. 17–25, 1982.

[41] J. Bell, S. Bolanowski, and M. H. Holmes, "The structure and function of pacinian corpuscles: A review", *Progress in Neurobiology*, vol. 42, no. 1, pp. 79–128, 1994.

[42] R. S. Johansson and A. Vallbo, "Tactile sensibility in the human hand: Relative and absolute densities of four types of mechanoreceptive units in glabrous skin.", *The Journal of Physiology*, vol. 286, no. 1, pp. 283–300, 1979.

[43] R. D. Howe and M. R. Cutkosky, "Sensing skin acceleration for slip and texture perception", in *IEEE International Conference on Robotics and Automation*, 1989, pp. 145–150.

[44] F. De Boissieu, C. Godin, B. Guilhamat, D. David, C. Serviere, and D. Baudois, "Tactile texture recognition with a 3-axial force MEMS integrated artificial finger", in *Robotics: Science and Systems*, Seattle, WA, 2009, pp. 49–56.

[45] W. McMahan and K. J. Kuchenbecker, "Haptic display of realistic tool contact via dynamically compensated control of a dedicated actuator", in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 3170–3177.

[46] H. Culbertson, J. M. Romano, P. Castillo, M. Mintz, and K. J. Kuchenbecker, "Refined methods for creating realistic haptic virtual textures from tool-mediated contact acceleration data", in *IEEE Haptics Symposium*, 2012, pp. 385–391.

[47] D. A. Mahns, N. Perkins, V. Sahai, L Robinson, and M. Rowe, "Vibrotactile frequency discrimination in human hairy skin", *Journal of Neurophysiology*, vol. 95, no. 3, pp. 1442–1450, 2006.

[48] S. J. Bensmaïa and M. Hollins, "Complex tactile waveform discrimination", *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1236–1245, 2000.

[49] E Francisco, V Tannan, Z Zhang, J Holden, and M Tommerdahl, "Vibrotactile amplitude discrimination capacity parallels magnitude changes in somatosensory cortex and follows weber's law", *Experimental Brain Research*, vol. 191, no. 1, p. 49, 2008.

[50] G. A. Gescheider, "Effects of signal probability on vibrotactile signal recognition", *Perceptual and Motor Skills*, vol. 38, no. 1, pp. 15–23, 1974.

[51] D. C. Spray, "Cutaneous temperature receptors", *Annual Review of Physiology*, vol. 48, no. 1, pp. 625–638, 1986.

[52] A. E. Dubin and A. Patapoutian, "Nociceptors: The sensors of the pain pathway", *The Journal of Clinical Investigation*, vol. 120, no. 11, pp. 3760–3772, 2010.

[53] R. Verillo, S. Bolanowski, C. Francis, and F. McGlone, "Effects of hydration on tactile sensation", *Somatosensory & Motor Research*, vol. 15, no. 2, pp. 93–108, 1998.

[54] K. O. Johnson, I. Darian-Smith, and C. LaMotte, "Peripheral neural determinants of temperature discrimination in man: A correlative study of responses to cooling skin.", *Journal of Neurophysiology*, vol. 36, no. 2, pp. 347–370, 1973.

[55] M. J. Caterina, M. A. Schumacher, M. Tominaga, T. A. Rosen, J. D. Levine, and D. Julius, "The capsaicin receptor: A heat-activated ion channel in the pain pathway", *Nature*, vol. 389, no. 6653, pp. 816–824, 1997.

[56] Y. Nishi, M. Osumi, S. Nobusako, K. Takeda, and S. Morioka, "Avoidance behavioral difference in acquisition and extinction of pain-related fear", *Frontiers in Behavioral Neuroscience*, vol. 13, p. 236, 2019.

[57] R. Clark and O. G. Edholm, *Man and his thermal environment*. Arnold London, 1985.

[58] D. Filingeri and G. Havenith, "Human skin wetness perception: Psychophysical and neurophysiological bases", *Temperature*, vol. 2, no. 1, pp. 86–104, 2015.

[59] M. Shibahara and K. Sato, "Illusion of wetness by dynamic touch", *IEEE Transactions on Haptics*, vol. 12, no. 4, pp. 533–541, 2019.

[60] F. A. Geldard and C. E. Sherrick, "The cutaneous rabbit: A perceptual illusion", *Science*, vol. 178, no. 4057, pp. 178–179, 1972.

[61] M. Hollins, R. Faldowski, S. Rao, and F. Young, "Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis", *Perception & Psychophysics*, vol. 54, no. 6, pp. 697–705, 1993.

[62] W. M. Bergmann-Tiest and A. M. Kappers, "Analysis of haptic perception of materials by multidimensional scaling and physical measurements of roughness and compressibility", *Acta Psychologica*, vol. 121, no. 1, pp. 1–20, 2006.

[63] S. Okamoto, H. Nagano, and Y. Yamada, "Psychophysical dimensions of tactile perception of textures", *IEEE Transactions on Haptics*, vol. 6, no. 1, pp. 81–93, 2013.

[64] K. Eilers, "D. Katz, Der Aufbau der Tastwelt (book review)", *Kant-Studien*, vol. 31, 1925.

[65] Haptics Symposium Committee. (2018). Cross-Cutting Challenge: Haptic dimensions of surfaces, [Online]. Available: `http://2018.hapticssymposium.org/ccc2`.

[66] J. A. Fishel, G. E. Loeb, B. Matulevich, and R. Davoodi, "Method and applications for measurement of object tactile properties based on how they likely feel to humans", 2016, US Patent App. 14/796,647.

[67] SynTouch. (2020). BioTac - biomimetic process, [Online]. Available: `https://www.syntouchinc.com/en/biomimetic-process/`.

[68] S. Lacey, C. Campbell, and K Sathian, "Vision and touch: Multiple or multisensory representations of objects?", *Perception*, vol. 36, no. 10, pp. 1513–1521, 2007.

[69] E. E. Brodie and H. E. Ross, "Sensorimotor mechanisms in weight discrimination", *Perception & Psychophysics*, vol. 36, no. 5, pp. 477–481, 1984.

[70] H.-N. Ho and L. A. Jones, "Material identification using real and simulated thermal cues", in *International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2004, pp. 2462–2465.

[71] W. M. Bergmann-Tiest and A. M. Kappers, "Tactile perception of thermal diffusivity", *Attention, Perception, & Psychophysics*, vol. 71, no. 3, pp. 481–489, 2009.

[72] M. Spinner, A. B. Wiechert, and S. N. Gorb, "Sticky fingers: Adhesive properties of human fingertips", *Journal of Biomechanics*, vol. 49, no. 4, pp. 606–610, 2016.

[73] R. H. LaMotte, "Softness discrimination with a tool", *Journal of Neurophysiology*, vol. 83, no. 4, pp. 1777–1786, 2000.

[74] L. Kaim and K. Drewing, "Exploratory strategies in haptic softness discrimination are tuned to achieve high levels of task performance", *IEEE Transactions on Haptics*, vol. 4, no. 4, pp. 242–252, 2011.

[75] G. A. Gescheider, *Psychophysics: the Fundamentals*. Psychology Press, 2013.

[76] S. Bensmaïa, M. Hollins, and J. Yau, "Vibrotactile intensity and frequency information in the pacinian system: A psychophysical model", *Perception & Psychophysics*, vol. 67, no. 5, pp. 828–841, 2005.

[77] M. Hollins and S. R. Risner, "Evidence for the duplex theory of tactile texture perception", *Perception & Psychophysics*, vol. 62, no. 4, pp. 695–705, 2000.

[78] S. J. Bensmaïa and M. Hollins, "The vibrations of texture", *Somatosensory & Motor Research*, vol. 20, no. 1, pp. 33–43, 2003.

[79] S. Bensmaïa and M. Hollins, "Pacinian representations of fine surface texture", *Perception & Psychophysics*, vol. 67, no. 5, pp. 842–854, 2005.

[80] R. L. Klatzky and S. J. Lederman, "Tactile roughness perception with a rigid link interposed between skin and surface", *Perception & Psychophysics*, vol. 61, no. 4, pp. 591–607, 1999.

[81] W. M. Bergmann-Tiest and A. M. Kappers, "Haptic and visual perception of roughness", *Acta Psychologica*, vol. 124, no. 2, pp. 177–189, 2007.

[82] T. Yoshioka, S. J. Bensmaïa, J. C. Craig, and S. S. Hsiao, "Texture perception through direct and indirect touch: An analysis of perceptual space for tactile textures in two modes of exploration", *Somatosensory & Motor Research*, vol. 24, no. 1-2, pp. 53–70, 2007.

[83] A. Singhal and L. A. Jones, "Perceptual interactions in thermo-tactile displays", in *IEEE World Haptics Conference*, 2017, pp. 90–95.

[84] L. Skedung, K. Danerlöv, U. Olofsson, M. Aikala, K. Niemi, J. Kettle, and M. W. Rutland, "Finger friction measurements on coated and uncoated printing papers", *Tribology letters*, vol. 37, no. 2, pp. 389–399, 2010.

[85] M. A. Srinivasan, J. Whitehouse, and R. H. LaMotte, "Tactile detection of slip: Surface microgeometry and peripheral neural codes", *Journal of Neurophysiology*, vol. 63, no. 6, pp. 1323–1332, 1990.

[86] S. M. Pasumarty, S. A. Johnson, S. A. Watson, and M. J. Adams, "Friction of the human finger pad: Influence of moisture, occlusion and velocity", *Tribology Letters*, vol. 44, no. 2, pp. 117–137, 2011.

[87] J. Drelich, E. Chibowski, D. Meng, and K. Terpilowski, "Hydrophilic and superhydrophilic surfaces and materials", *Soft Matter*, vol. 7, pp. 9804–9828, 2011.

[88] R. D. Howe and M. R. Cutkosky, "Dynamic tactile sensing: Perception of fine surface features with stress rate sensing", *IEEE Transactions on Robotics and Automation*, vol. 9, no. 2, pp. 140–151, 1993.

[89] P. Dallaire, P. Giguère, D. Émond, and B. Chaib-Draa, "Autonomous tactile perception: A combined improved sensing and bayesian nonparametric approach", *Robotics and autonomous systems*, vol. 62, no. 4, pp. 422–435, 2014.

[90] A. Burka, S. Hu, S. Helgeson, S. Krishnan, Y. Gao, L. A. Hendricks, T. Darrell, and K. J. Kuchenbecker, "Proton: A visuo-haptic data acquisition system for robotic learning of surface properties", in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2016, pp. 58–65.

[91] H. Culbertson, J. Unwin, and K. J. Kuchenbecker, "Modeling and rendering realistic textures from unconstrained tool-surface interactions", *IEEE Transactions on Haptics*, vol. 7, no. 3, pp. 381–393, 2014.

[92] R. Grigorii, M. Peshkin, and J. E. Colgate, "High-bandwidth tribometry as a means of recording natural textures", in *IEEE World Haptics Conference*, 2017.

[93] R. Hill, B. Storåkers, and A. Zdunek, "A theoretical study of the brinell hardness test", *Proceedings of the Royal Society of London. Mathematical and Physical Sciences*, vol. 423, no. 1865, pp. 301–330, 1989.

[94] T. Aujeszky, G. Korres, and M. Eid, "Measurement-based thermal modeling using laser thermography", *IEEE Transactions on Instrumentation and Measurement*, 2018.

[95] A. M. Okamura, J. T. Dennerlein, and R. D. Howe, "Vibration feedback models for virtual environments", in *IEEE International Conference on Robotics and Automation*, vol. 1, 1998, pp. 674–679.

[96] K. J. Kuchenbecker, J. Fiene, and G. Niemeyer, "Improving contact realism through event-based haptic feedback", *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 2, pp. 219–230, 2006.

[97] M. Minsky, O. Y. Ming, O. Steele, F. P. Brooks Jr, and M. Behensky, "Feeling and seeing: Issues in force display", in *Interactive 3D Graphics*, 1990, pp. 235–241.

[98] M. Minsky, "Computational haptics: The sandpaper system for synthesizing texture for a force-feedback display", PhD thesis, Massachusetts Institute of Technology, 1995.

[99] M. Minsky and S. J. Lederman, "Simulated haptic textures: Roughness", in *Proceedings of the ASME Dynamic Systems and Control Division*, vol. 58, 1996, pp. 421–426.

[100] C. Basdogan, C Ho, and M. A. Srinivasan, "A raybased haptic rendering technique for displaying shape and texture of 3D objects in virtual environments", in *ASME Winter Annual Meeting*, vol. 61, 1997, pp. 77–84.

[101] S. Shin and S. Choi, "Geometry-based haptic texture modeling and rendering using photometric stereo", in *IEEE Haptics Symposium*, 2018.

[102] M. Dulik and L. Ladanyi, "Surface detection and recognition using infrared light", in *IEEE Elektro*, 2014, pp. 159–164.

[103] J. M. Romano, T. Yoshioka, and K. J. Kuchenbecker, "Automatic filter design for synthesis of haptic textures from recorded acceleration data", in *IEEE International Conference on Robotics and Automation*, 2010, pp. 1815–1821.

[104] A. Abdulali and S. Jeon, "Data-driven modeling of anisotropic haptic textures: Data segmentation and interpolation", in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, Springer, 2016, pp. 228–239.

[105] A. Abdulali, W. Hassan, and S. Jeon, "Sample selection of multi-trial data for data-driven haptic texture modeling", in *IEEE World Haptics Conference*, 2017, pp. 66–71.

[106] H. Culbertson and K. J. Kuchenbecker, "Importance of matching physical friction, hardness, and texture in creating realistic haptic virtual surfaces", *IEEE Transactions on Haptics*, vol. 10, no. 1, pp. 63–74, 2017.

[107] Y. Tanaka, Y. Horita, and A. Sano, "Finger-mounted skin vibration sensor for active touch", in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, Springer, 2012, pp. 169–174.

[108] S. Sato, S. Okamoto, Y. Matsuura, and Y. Yamada, "Wearable finger pad sensor for tactile textures using propagated deformation on a side of a finger: Assessment of accuracy", in *IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 892–896.

[109] Y. Visell and Y. Shao, "Learning constituent parts of touch stimuli from whole hand vibrations", in *IEEE Haptics Symposium*, 2016, pp. 253–258.

[110] Y. Vardar, B. Güçlü, and C. Basdogan, "Effect of waveform on tactile perception by electrovibration displayed on touch screens", *IEEE Transactions on Haptics*, vol. 10, no. 4, pp. 488–499, 2017.

[111] H. Choi, S. Cho, S. Shin, H. Lee, and S. Choi, "Data-driven thermal rendering: An initial study", in *IEEE Haptics Symposium*, 2018.

[112] Y. Tada, K. Hosoda, Y. Yamasaki, and M. Asada, "Sensing the texture of surfaces by anthropomorphic soft fingertips with multi-modal sensors", in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, 2003, pp. 31–35.

[113] C. M. Oddo, L. Beccai, M. Felder, F. Giovacchini, and M. C. Carrozza, "Artificial roughness encoding with a bio-inspired MEMS-based tactile sensor array", *Sensors*, vol. 9, no. 5, pp. 3161–3183, 2009.

[114] J. Sinapov and V. Sukhoy, "Vibrotactile recognition and categorization of surfaces by a humanoid robot", *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 488–497, 2011.

[115] N. Jamali and C. Sammut, "Majority voting: Material classification by tactile sensing using surface texture", *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 508–521, 2011.

[116] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures", *Frontiers in Neurorobotics*, vol. 6, no. 4, pp. 1–20, 2012.

[117] J. M. Romano and K. J. Kuchenbecker, "Methods for Robotic Tool-Mediated Haptic Surface Recognition", in *IEEE Haptics Symposium*, 2014, pp. 49–56.

[118] C. H. Lin, T. W. Erickson, J. A. Fishel, N. Wettels, and G. E. Loeb, "Signal processing and fabrication of a biomimetic tactile sensor array with thermal, force and microvibration modalities", in *IEEE International Conference on Robotics and Biomimetics*, 2009, pp. 129–134.

[119] J. Hoelscher, J. Peters, and T. Hermans, "Evaluation of tactile feature extraction for interactive object recognition", in *IEEE-RAS International Conference on Humanoid Robots*, 2015, pp. 310–317.

[120] M. Kaboli and G. Cheng, "Robust tactile descriptors for discriminating objects from textural properties via artificial robotic skin", *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 985–1003, 2018.

[121] A. M. Smith, G. Gosselin, and B. Houde, "Deployment of fingertip forces in tactile exploration", *Experimental Brain Research*, vol. 147, no. 2, pp. 209–218, 2002.

[122] D. Prattichizzo, F. Chinello, C. Pacchierotti, and M. Malvezzi, "Towards wearability in fingertip haptics: A 3-dof wearable device for cutaneous force feedback", *IEEE Transactions on Haptics*, vol. 6, no. 4, pp. 506–516, 2013.

[123] (2020). Dexta robotics at http://www.dextarobotics.com/, [Online]. Available: `http://www.dextarobotics.com/`.

[124] I. Choi, E. W. Hawkes, D. L. Christensen, C. J. Ploch, and S. Follmer, "Wolverine: A wearable haptic interface for grasping in virtual reality", in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016, pp. 986–993.

[125] I. Choi, H. Culbertson, M. R. Miller, A. Olwal, and S. Follmer, "Grabity: A wearable haptic interface for simulating weight and grasping in virtual reality", in *ACM Symposium on User Interface Software and Technology*, 2017, pp. 119–130.

[126] R. Hinchet, V. Vechev, H. Shea, and O. Hilliges, "DextrEs: Wearable haptic feedback for grasping in VR via a thin form-factor electrostatic brake", in *ACM Symposium on User Interface Software and Technology*, 2018, pp. 901–912.

[127] T. Carter, S. A. Seah, B. Long, B. Drinkwater, and S. Subramanian, "Ultrahaptics: Multi-point mid-air haptic feedback for touch surfaces", in *ACM Symposium on User Interface Software and Technology*, 2013, pp. 505–514.

[128] C. Pacchierotti, S. Sinclair, M. Solazzi, A. Frisoli, V. Hayward, and D. Prattichizzo, "Wearable haptic systems for the fingertip and the hand: Taxonomy, review, and perspectives", *IEEE Transactions on Haptics*, vol. 10, no. 4, pp. 580–600, 2017.

[129] W. S. Harwin and N. Melder, "Improved haptic rendering for multi-finger manipulation using friction cone based god-objects", in *Eurohaptics Conference*, Citeseer, 2002, pp. 82–85.

[130] H. Culbertson and K. J. Kuchenbecker, "Ungrounded haptic augmented reality system for displaying roughness and friction", *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 4, pp. 1839–1849, 2017.

[131] C. Basdogan and M. A. Srinivasan, "Haptic rendering in virtual environments", *Handbook of Virtual Environments*, vol. 1, pp. 117–134, 2002.

[132] M. Bianchi, M. Poggiani, A. Serio, and A. Bicchi, "A novel tactile display for softness and texture rendering in tele-operation tasks", in *IEEE World Haptics Conference*, 2015, pp. 49–56.

[133] E. M. Young and K. J. Kuchenbecker, "Design of a parallel continuum manipulator for 6-dof fingertip haptic display", in *IEEE World Haptics Conference*, 2017, pp. 599–604.

[134] S. Choi and K. J. Kuchenbecker, "Vibrotactile display: Perception, technology, and applications", *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2093–2104, 2013.

[135] G. Campion and V. Hayward, "Fundamental limits in the rendering of virtual haptic textures", in *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2005, pp. 263–270.

[136] G. Westling and R. S. Johansson, "Responses in glabrous skin mechanoreceptors during precision grip in humans", *Experimental Brain Research*, vol. 66, no. 1, pp. 128–140, 1987.

[137] D. B. Hiemstra, "The design of moving magnet actuators for large-range flexure-based nanopositioning", PhD thesis, University of Michigan Department of Mechanical Engineering Ann Arbour, 2014.

[138] V. Yem, R. Okazaki, and H. Kajimoto, "Vibrotactile and pseudo force presentation using motor rotational acceleration", in *IEEE Haptics Symposium*, 2016, pp. 47–51.

[139] G. Park, H. Cha, and S. Choi, "Attachable and detachable vibrotactile feedback modules and their information capacity for spatiotemporal patterns", in *World Haptics Conference*, 2017, pp. 78–83.

[140] W. Huang, "Shape memory alloys and their application to actuators for deployable structures", PhD thesis, University of Cambridge Department of Engineering (Peterhouse), 1998.

[141] E. Mallinckrodt, A. Hughes, and W. Sleator Jr, "Perception by the skin of electrically induced vibrations", *Science*, 1953.

[142] O. Bau, I. Poupyrev, A. Israr, and C. Harrison, "Teslatouch: Electrovibration for touch surfaces", in *ACM Symposium on User Interface Software and Technology*, 2010, pp. 283–292.

[143] S. Wu, X. Sun, Q. Wang, and J. Chen, "Tactile modeling and rendering image-textures based on electrovibration", *The Visual Computer*, vol. 33, no. 5, pp. 637–646, 2017.

[144] G. Ilkhani, M. Aziziaghdam, and E. Samur, "Data-driven texture rendering on an electrostatic tactile display", *International Journal of Human–Computer Interaction*, pp. 1–15, 2017.

[145] C. D. Shultz, M. A. Peshkin, and J. E. Colgate, "Surface haptics via electroadhesion: Expanding electrovibration with Johnsen and Rahbek", in *IEEE World Haptics Conference*, 2015, pp. 57–62.

[146] ——, "The application of tactile, audible, and ultrasonic forces to human fingertips using broadband electroadhesion", *IEEE Transactions on Haptics*, vol. 11, no. 2, pp. 279–290, 2018.

[147] ——, "On the electrical characterization of electroadhesive displays and the prominent interfacial gap impedance associated with sliding fingertips", in *IEEE Haptics Symposium*, 2018, pp. 151–157.

[148] (2017). Tanvas corporation at https://tanvas.co/technology/, [Online]. Available: `https://tanvas.co/technology/`.

[149] O. Bau and I. Poupyrev, "REVEL: Tactile feedback technology for augmented reality", *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 89–100, 2012.

[150] T. Watanabe and S. Fukui, "A method for controlling tactile sensation of surface roughness using ultrasonic vibration", in *IEEE International Conference on Robotics and Automation*, vol. 1, 1995, pp. 1134–1139.

[151] M. Biet, F. Giraud, and B. Lemaire-Semail, "Squeeze film effect for the design of an ultrasonic tactile plate", *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 54, no. 12, pp. 2678–2688, 2007.

[152] L. Winfield, J. Glassmire, J. E. Colgate, and M. Peshkin, "T-Pad: Tactile pattern display through variable friction reduction", in *Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2007, pp. 421–426.

[153] (2017). The tpad phone at http://www.thetpadphone.com/, [Online]. Available: `http://www.thetpadphone.com/`.

[154] H.-N. Ho and L. A. Jones, "Thermal model for hand-object interactions", in *14th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2005, pp. 461–467.

[155] A. Singhal and L. A. Jones, "Space-time interactions and the perceived location of cold stimuli", in *IEEE Haptics Symposium*, 2016, pp. 92–97.

[156] T. Murakami, T. Person, C. L. Fernando, and K. Minamizawa, "Altered touch: Miniature haptic display with force, thermal and tactile feedback for augmented haptics", in *ACM SIGGRAPH 2017 Emerging Technologies*, 2017.

[157] H.-N. Ho and L. A. Jones, "Modeling the thermal responses of the skin surface during hand-object interactions", *Journal of Biomechanical Engineering*, vol. 130, no. 2, pp. 1–8, 2008.

[158] M. Gabardi, D. Leonardis, M. Solazzi, and A. Frisoli, "Development of a miniaturized thermal module designed for integration in a wearable haptic device", in *IEEE Haptics Symposium*, 2018, pp. 100–105.

[159] A. El Saddik, M. Orozco, M. Eid, and J. Cha, *Haptics technologies: Bringing touch to multimedia*. Springer Science & Business Media, 2011.

[160] P Hinterseer, E Steinbach, and S Chaudhuri, "Model based data compression for 3D virtual haptic teleinteraction", in *International Conference on Consumer Electronics*, 2006, pp. 23–24.

[161] P. Hinterseer, S. Hirche, S. Chaudhuri, E. Steinbach, and M. Buss, "Perception-based data reduction and transmission of haptic data in telepresence and teleaction systems", *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 588–597, 2008.

[162] E. Steinbach, S. Hirche, M. Ernst, F. Brandi, R. Chaudhari, J. Kammerl, and I. Vittorias, "Haptic communications", *Proceedings of the IEEE*, vol. 100, no. 4, pp. 937–956, 2012.

[163] H. Vogel *et al.*, *Gerthsen Physik*. Springer, 1995, vol. 19.

[164] C. B. Zilles and J. K. Salisbury, "A constraint-based god-object method for haptic display", in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3, 1995, pp. 146–151.

[165] D. Roylance, "Engineering viscoelasticity", *Department of Materials Science and Engineering Massachusetts Institute of Technology, Cambridge MA*, vol. 2139, pp. 1–37, 2001.

[166] P. Kelly, "Solid mechanics", *Part II, Lecture notes, The University of Auckland*, 2013.

[167] O. Caldiran, H. Z. Tan, and C. Basdogan, "An investigation of haptic perception of viscoelastic materials in the frequency domain", in *Haptics Symposium*, 2018, pp. 222–228.

[168] V. L. Popov, *Kontaktmechanik und Reibung: Ein Lehr-und Anwendungsbuch von der Nanotribologie bis zur numerischen Simulation*. Springer-Verlag, 2009.

[169] S Derler, L.-C. Gerhardt, A Lenz, E Bertaux, and M Hadad, "Friction of human skin against smooth and rough glass as a function of the contact pressure", *Tribology International*, vol. 42, no. 11-12, pp. 1565–1574, 2009.

[170] D. C. Ruspini, K. Kolarov, and O. Khatib, "Haptic interaction in virtual environments", in *IEEE International Conference on Intelligent Robot and Systems*, vol. 1, 1997, pp. 128–133.

[171] C. Richard, M. R. Cutkosky, and K. MacLean, "Friction identification for haptic display", *Haptic Interfaces for Virtual Environments and Teleoperator Systems*, pp. 14–19, 1999.

[172] P. R. Dahl, "Solid friction damping of mechanical vibrations", *AIAA Journal*, vol. 14, no. 12, pp. 1675–1682, 1976.

[173] V. Hayward and B. Armstrong, "A new computational model of friction applied to haptic rendering", in *Experimental Robotics VI*, Springer, 2000, pp. 403–412.

[174] C. C. De Wit, H. Olsson, K. J. Astrom, and P. Lischinsky, "A new model for control of systems with friction", *IEEE Transactions on Automatic Control*, vol. 40, no. 3, pp. 419–425, 1995.

[175] N. L. Max and B. G. Becker, "Bump shading for volume textures", *IEEE Computer Graphics and Applications*, vol. 14, no. 4, pp. 18–20, 1994.

[176] T. Hausberger, M. Terzer, F. Enneking, Z. Jonas, and Y. Kim, "Surftics – kinesthetic and tactile feedback on a touchscreen device", in *IEEE World Haptics Conference*, 2017, pp. 472–477.

[177] R. C. Gonzalez and P. Wintz, *Digital image processing*. Addison-Wesley Publishing Co., 1977.

[178] E. Schrüfer, *Signalverarbeitung: Numerische Verarbeitung digitaler Signale*. Hanser, 1990.

[179] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[180] L. Tokeheim. (2020). Wavelet analysis of musical signals, [Online]. Available: `http://pages.cs.wisc.edu/~ltokheim/cs514/report/report.pdf`.

[181] M. A. Otaduy and M. C. Lin, "Rendering of textured objects", *Haptic Rendering: Foundations, Algorithms, and Applications*, pp. 371–393, 2008.

[182] Y. Takeuchi, S. Kamuro, K. Minamizawa, and S. Tachi, "Haptic duplicator", in *ACM Virtual Reality International Conference*, 2012, pp. 30–31.

[183] H. Culbertson, J. Unwin, B. E. Goodman, and K. J. Kuchenbecker, "Generating haptic texture models from unconstrained tool-surface interactions", in *IEEE World Haptics Conference*, 2013, pp. 295–300.

[184] H. H. Pennes, "Analysis of tissue and arterial blood temperatures in the resting human forearm", *Journal of Applied Physiology*, vol. 1, no. 2, pp. 93–122, 1948.

[185] A. Sarda, R. Deterre, and C. Vergneault, "Heat perception measurements of the different parts found in a car passenger compartment", *Measurement*, vol. 35, no. 1, pp. 65–75, 2004.

[186] J. H. Lienhard, *A heat transfer textbook*. Courier Dover Publications, 2019.

[187] Y. S. Touloukian, R. Powell, C. Ho, and M. Nicolaou, "Thermophysical properties of matter – the TPRC data series. Volume 10: thermal diffusivity", 1974.

[188] Y. S. Touloukian, R. Powell, C. Ho, and P. Klemens, "Thermophysical properties of matter – the TPRC data series. Volume 1: thermal conductivity-metallic elements and alloys", 1970.

[189] A. F. Mills, *Basic heat and mass transfer*. Prentice hall, 1999.

[190] E. Steinbach, S. Hirche, J. Kammerl, I. Vittorias, and R. Chaudhari, "Haptic data compression and communication", *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 87–96, 2010.

[191] H. Culbertson, J. J. López Delgado, and K. J. Kuchenbecker, "One hundred data-driven haptic texture models and open-source methods for rendering on 3D objects", in *IEEE Haptics Symposium*, 2014, pp. 319–325.

[192] W. Hassan, A. Abdulali, M. Abdullah, S. C. Ahn, and S. Jeon, "Towards universal haptic library: Library-based haptic texture assignment using image texture and perceptual space", *IEEE Transactions on Haptics*, vol. 11, no. 2, pp. 291–303, 2018.

[193] F. Musso and J. Weber. (2020). TUM collection of building materials, [Online]. Available: `https://www.ar.tum.de/ebb/startseite/`.

[194] Y. Radecker. (2020). Internationales Material-Archiv Zuerich, [Online]. Available: `http://www.materialarchiv.ch/`.

[195] H.-J. Bargel and G. Schulze, *Werkstoffkunde*. Springer-Verlag, 2008.

[196] B. Delhaye, V. Hayward, P. Lefevre, and J. L. Thonnard, "Textural vibrations in the forearm during tactile exploration", in *Annual Meeting of the Society for Neuroscience, Poster*, vol. 782, 2010.

[197] A. Akay, "Acoustics of friction", *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1525–1548, 2002.

[198] O. Kroemer, C. H. Lampert, and J. Peters, "Learning dynamic tactile sensing with robust vision-based training", *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 545–557, 2011.

[199] T. Aytaç and B. Barshan, "Surface recognition by parametric modeling of infrared intensity signals", in *European Signal Processing Conference*, 2004, pp. 1107–1110.

[200] N. Landin, J. M. Romano, W. McMahan, and K. J. Kuchenbecker, "Dimensional reduction of high-frequency accelerations for haptic rendering", in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, Springer, 2010, pp. 79–86.

[201] H. Culbertson and K. J. Kuchenbecker, "Should haptic texture vibrations respond to user force and speed?", in *IEEE World Haptics Conference*, 2015, pp. 106–112.

[202] E. Meier. (2020). The wood database, [Online]. Available: `https://www.wood-database.com/`.

[203] P. Flach, *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, 2012.

[204] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[205] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From BoW to CNN: Two decades of texture representation for texture classification", *International Journal of Computer Vision*, vol. 127, no. 1, pp. 74–109, 2019.

[206] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification", *IEEE Transactions on Systems, Man and Cybernetics*, no. 6, pp. 610–621, 1973.

[207] R. Monzel. (2020). Haralick features, [Online]. Available: `https://de.mathworks.com/matlabcentral/fileexchange/58769-haralicktexturefeatures`.

[208] S. Sornapudi. (2020). Tamura features, [Online]. Available: `https://github.com/Sdhir/TamuraFeatures`.

[209] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[210] P. V. Hough, "Method and means for recognizing complex patterns", 1962, US Patent 3,069,654.

[211] H.-G. Kim, N. Moreau, and T. Sikora, *MPEG-7 audio and beyond: Audio content indexing and retrieval*. John Wiley & Sons, 2006.

[212] T. Giannakopoulos and A. Pikrakis. (2019). Introduction to audio analysis matlab implementation, [Online]. Available: `https://de.mathworks.com/matlabcentral/fileexchange/45831-matlab-audio-analysis-library`.

[213] ——, *Introduction to audio analysis: a MATLAB approach*. Academic Press, 2014.

[214] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals", *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[215] G. H. Wakefield, "Mathematical representation of joint time-chroma distributions", in *Advanced Signal Processing Algorithms, Architectures, and Implementations IX*, International Society for Optics and Photonics, vol. 3807, 1999, pp. 637–645.

[216] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction to pattern recognition: a matlab approach*. Academic Press, 2010.

[217] M. Slaney, "Auditory toolbox", *Interval Research Corporation*, vol. 10, no. 1998, 1998.

[218] K. Wojcicki. (2020). MFCC matlab, [Online]. Available: `http://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab/content/mfcc/mfcc.m`.

[219] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, T. Darrell, and K. J. Kuchenbecker, "Robotic learning of haptic adjectives through physical interaction", *Robotics and Autonomous Systems*, vol. 63, pp. 279–292, 2015.

[220] A. Burka and K. J. Kuchenbecker, "Handling scan-time parameters in haptic surface classification", in *IEEE World Haptics Conference*, 2017, pp. 424–429.

[221] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded-up robust features", in *European Conference on Computer Vision*, Springer, 2006, pp. 404–417.

[222] J. A. Fishel and G. E. Loeb, "Sensing tactile microvibrations with the BioTac - comparison with human sensitivity", in *IEEE RAS & EMBS 4th International Conference on Biomedical Robotics and Biomechatronics*, 2012, pp. 1122–1127.

[223] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems", *Machine Learning*, vol. 45, no. 2, pp. 171–186, 2001.

[224] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection", *Journal of Machine Learning Research*, vol. 3, no. 1, pp. 1157–1182, 2003.

[225] M. G. Kendall, "Rank correlation methods.", *American Psychological Association*, 1948.

[226] J. Hochberg and A Tamhane, *Multiple comparison procedures*. John Wiley and Sons, 1987.

[227] M. H. Kutner, C. J. Nachtsheim, J. Neter, W. Li, *et al.*, *Applied linear statistical models*. McGraw-Hill Irwin New York, 2005, vol. 5.

[228] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis", *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[229] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", in *Ijcai*, vol. 14, 1995, pp. 1137–1145.

[230] S. Raschka and V. Mirjalili, *Machine Learning mit Python und Scikit-Learn und TensorFlow: Das umfassende Praxis-Handbuch für Data Science, Predictive Analytics und Deep Learning*. MITP-Verlags GmbH & Co. KG, 2017.

[231] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning", *Artificial intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.

[232] P. Domingos, "A few useful things to know about machine learning", *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[233] J. R. Quinlan, "Induction of decision trees", *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[234] L. Breiman, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[235] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?", *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3133–3181, 2014.

[236] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[237] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.", *Psychological Review*, vol. 65, no. 6, p. 386, 1958.

[238] V. N. Vapnik, "An overview of statistical learning theory", *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[239] P. Cunningham and S. J. Delany, "K-nearest neighbour classifiers", *Multiple Classifier Systems*, vol. 34, no. 8, pp. 1–17, 2007.

[240] L. Papula, *Mathematik fuer Ingenieure und Naturwissenschaftler Band 3: Vektoranalysis, Wahrscheinlichkeitsrechnung, Mathematische Statistik, Fehler-und Ausgleichsrechnung*. Springer-Verlag, 2016.

[241] K. P. Murphy, *Machine Learning: A Probabilistic Perspective. Adaptive Computation and Machine Learning*. MIT press, 2012.

[242] R. E. Schapire, "The strength of weak learnability", *Machine Learning*, vol. 5, no. 2, pp. 197–227, 1990.

[243] P. Harrington, *Machine Learning in Action*. Shelter Island, NY: Manning Publications Co, 2012.

[244] L. Breiman, "Bagging predictors", *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.

[245] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", in *European Conference on Computational Learning Theory*, Springer, 1995, pp. 23–37.

[246] T. DeVries and G. W. Taylor, "Dataset augmentation in feature space", *arXiv preprint arXiv:1702.05538*, 2017.

[247] R. Likert, "A technique for the measurement of attitudes.", *Archives of Psychology*, 1932.

[248] M. V. Selst and P. Jolicoeur, "A solution to the effect of sample size on outlier elimination", *Journal of Experimental Psychology*, vol. 47, no. 3, pp. 631–650, 1994.

[249] F. J. Massey Jr, "The Kolmogorov–Smirnov test for goodness of fit", *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[250] M. A. Nielsen, *Neural networks and deep learning*. Determination press San Francisco, 2015, vol. 25.

[251] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *et al.*, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[252] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[253] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3828–3836.

[254] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409.1556*, 2014.

[255] R. Bello-Cerezo, F. Bianconi, F. Di Maria, P. Napoletano, and F. Smeraldi, "Comparative evaluation of hand-crafted image descriptors vs. off-the-shelf CNN-based features for colour texture classification under ideal and realistic conditions", *Applied Sciences*, vol. 9, no. 4, p. 738, 2019.

[256] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning", *arXiv prepr. 1811.03378*, 2018.

[257] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines", in *Proceedings of the 27th International Conference on Machine Learning*, 2010, pp. 807–814.

[258] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions", *arXiv preprint arXiv:1710.05941*, 2017.

[259] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models", in *International Conference on Machine Learning*, vol. 30, 2013, pp. 3–9.

[260] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[261] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *arXiv preprint arXiv:1502.03167*, 2015.

[262] Y. Nesterov, "A method of solving a convex programming problem with convergence rate", in *Soviet Math. Dokl*, vol. 27.

[263] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

[264] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, "The marginal value of adaptive gradient methods in machine learning", in *Advances in Neural Information Processing Systems*, 2017, pp. 4148–4158.

[265] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[266] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[267] K. J. Dana, B. Van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces", *ACM Transactions On Graphics*, vol. 18, no. 1, pp. 1–34, 1999.

[268] P Mallikarjuna, A. T. Targhi, M. Fritz, E. Hayman, B. Caputo, and J.-O. Eklundh, "The KTH-TIPS2 database", *Computational Vision and Active Perception Laboratory*, 2006.

[269] J. M. Geusebbroek. (2020). Amsterdam library of textures (ALOT), [Online]. Available: `http://color.univ-lille.fr/datasets/alot`.

[270] M Meenes and M. Zigler, "An experimental study of the perceptions roughness and smoothness", *The American Journal of Psychology*, pp. 542–549, 1923.

[271] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, and T. Darrell, "Deep learning for tactile understanding from visual and haptic data", in *IEEE International Conference on Robotics and Automation*, 2016, pp. 536–543.

[272] R. P. Khurshid, N. T. Fitter, E. A. Fedalei, and K. J. Kuchenbecker, "Effects of grip-force, contact, and acceleration feedback on a teleoperated pick-and-place task", *IEEE Transactions on Haptics*, vol. 10, no. 1, pp. 40–53, 2017.

[273] M. Price and F. C. Sup, "A robotic touchscreen totem for two-dimensional haptic force display", in *IEEE Haptics Symposium*, 2016, pp. 72–77.

[274] L. Brayda, C. Campus, M. Memeo, and L. Lucagrossi, "The importance of visual experience, gender, and emotion in the assessment of an assistive tactile mouse", *IEEE Transactions on Haptics*, vol. 8, no. 3, pp. 279–286, 2015.

[275] R. G. Golledge, M. Rice, and R. D. Jacobson, "A commentary on the use of touch for accessing on-screen spatial representations: The process of experiencing haptic maps and graphics", *The Professional Geographer*, vol. 57, no. 3, pp. 339–349, 2005.

[276] J. A. Terry and H Hsiao, "Tactile feedback in a computer mouse", in *IEEE Northeast Bioengineering Conference*, 1988, pp. 146–149.

[277] M. Akamatsu and I. S. MacKenzie, "Movement characteristics using a mouse with tactile and force feedback", *International Journal of Human-Computer Studies*, vol. 45, no. 4, pp. 483–493, 1996.

[278] T. Pietrzak, I. Pecci, and B. Martin, "Static and dynamic tactile directional cues experiments with VTPlayer mouse", in *Eurohaptics*, 2006, pp. 63–68.

[279] Engineering Acoustics, Inc. (2020). C2 technical specifications., [Online]. Available: `https://www.eaiinfo.com/tactor-info/`.

[280] A. Israr, S. Choi, and H. Z. Tan, "Mechanical impedance of the hand holding a spherical tool at threshold and suprathreshold stimulation levels", in *EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2007, pp. 56–60.

[281] D. Dugdale, *Essentials of Electromagnetism*. Springer Science & Business Media, 1997.

[282] J. Mullenbach, C. D. Shultz, J. E. Colgate, and A. M. Piper, "Exploring affective communication through variable-friction surface haptics", in *ACM SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 3963–3972.

[283] Y. Hiramatsu, D. Kimura, K. Kadota, T. Ito, and H. Kinoshita, "Control of precision grip force in lifting and holding of low-mass objects", *Prehension on Lightweight Objects One*, vol. 10, pp. 1–19, 2015.

[284] S Coble, "Materials data book", *Cambridge University Engineering Department*, 2003.

List of Figures

1.1 Haptic material acquisition, modeling, and display addresses the research questions of how to scan various material samples using novel recording devices (upper left), how to parameterize and store virtual material representations in a computer system (lower left and right), and how to recreate the touch sensations using novel haptic display devices (right). 2
1.2 Overview of the studied haptic material acquisition, modeling, and display pipeline. Except the waveform-based compression, transmission and reduction part, all research areas are addressed within this thesis. 3

2.2 The general haptic rendering loop needs to run at least at 1 kHz to ensure realistic hardness display, i.e., kinesthetic force feedback. Further haptic effects such as Coulomb friction or macro-texture models have been included to this concept. However, thermal and high-fidelity microscopic roughness feedback were not conceptualized in the traditional haptic rendering pipeline. 21
2.3 Haptic rendering is derived from 3D computer graphic algorithms and reveals dependencies between kinesthetic and tactile dimensions. For example, shapes are represented by polygon meshes visually and can be made tangible using the GO approach [164]. Based on these meshes, the HIP history and calculated GO are used to calculate interaction forces for haptic display such as hardness, friction, or macroscopic roughness height profiles. Shape is hence connected to the local perception of hardness, which is modeled by viscoelastic systems made of springs and dashpots. In this thesis, the shown configuration (standard linear solid model) of two springs and one dashpot is notably relevant for hardness rendering. 22
2.5 High-frequency hardness information during tapping can be modeled linearly according to the impact speed while leading to a sufficient perceptual similarity to real tapping on a material surface (adapted from [106] ©2017 IEEE). 25
2.6 Basic friction models. (a) Viscous friction model for ideal fluids. (b) Basic Coulomb friction. (c) Combination of (a) and (b). (d) Coulomb friction with static friction and the commonly used implementation of the friction cone, shown in the right part of this figure. (e) Coulomb friction with deadband-based speed zone. (f) Combination of (d) and Stribeck friction for lubricant modeling (adapted from [171] ©1999 IEEE and [129] ©2002 IEEE). 26

2.7 Generalized Dahl friction model. The first row shows the input periodical sliding (from left to right: 1 Hz, 2 Hz, 3 Hz) and the second and third row the resulting force-time and force-displacement curves. Note that the typical stick-slip chatter (1) occurs whenever speed becomes faster and is notably strong for lower frequencies whenever the static and kinetic friction coefficients are different, and creeping (2) occurs whenever the motion becomes slower, i.e., before the finger gets stuck again. 28
2.8 Macroscopic structure patterns can be added to completely flat virtual material surfaces. Note that macroscopic roughness is linked to hardness modeling, represented by the two springs and the dashpot from the aforementioned SLSM. 29

3.1 Initial LMT 108 haptic database based on a loose subjective taxonomy and made publicly available on the website [15] (adapted from [7] © 2017 IEEE). 40
3.3 Final haptic material database image montage. The different classes C have been separated by black lines. The class names are C_1 woods, C_2 biodegradables, C_3 plastics, C_4 ceramics, C_5 glasses, C_6 stones, C_7 metals, C_8 composites. The class imbalance toward woods actually reflects the true circumstance that this class has the most real world representatives, i.e., types of trees (adapted from [2] © 2019 IEEE). 42
3.4 Design evolution of the Texplorer device. From left to right: Phantom-based Texplorer version 1.a, Texplorer version 1.b, Texplorer version 1.c, and two units of the final version Texplorer2. All models were designed by using the CAD software Autodesk Fusion 360. Note that the setups also gradually decreased in their component costs and increased in their flexibility of usage. The very first version requires, e.g., a Phantom Omni device, whereas the final version comes along with a low-cost USB data acquisition card NI USB-6002 (National Instruments, USA) attached to a laptop, and is mobile like the Proton Pack presented by Burka et al. [90]. 46
3.5 Texplorer (version 1.c) system overview. Six sensors provide the input data traces that are used for surface classification based on the five major perceptual dimensions from [63]. The resulting signals (right part of the figure) originate from parallel recordings while tapping on and sliding over the material surface. Tapping on and sliding over the material surfaces leads to acceleration and audio tapping and sliding signals, reflectance scans reveal surface height information, FSR scans are used to infer frictional forces, camera-captured images to visual features, and a metal detector robustly identifies metal surfaces (adapted from [7] © 2017 IEEE). 47

4.2 Custom image anisotropy rotation-invariant feature based on Hough transform (adapted from [7] © 2017 IEEE). 65
4.3 Graphical illustration of the adjacent windows with respect to the current pixel in the horizontal (left) and right-diagonal direction (right) for image coarseness calculation (adapted from [7] © 2017 IEEE). 67
4.4 Well-defined squared patterns for the identification of maximum coarse texture element sizes. A structural element of 10 mm^2 is 240 – 300 pixel large given the usual recording distance of 100 mm – 140 mm and 4K resolution. Magnified images (right) are about 1700 pixel at a fix recording distance of 25 mm. These values are relevant to determine appropriate sizes of image feature parameters, notably, the coarseness feature. Also note that these parameters help in determining the relevant image size for deep learning-based image classification in Section 4.5. If coarse materials occupy 300 pixels in width and height, and if an overlap of 100 pixels is considered for data generation, an overall width/height of 500 pixels is chosen as image patch input size. 68

4.5 Montage of surface images (hemp) captured with the TU3 extension under subsequent rotations of approximately 8°. Note that the illumination and distance conditions do not noticeably change within the images. 69
4.6 Subset of image features (image contrast, edginess, and 6th RLBP) calculated for the three material images under stepwise rotation. The image features have been min-max-normalized according to the range of values of all material samples from the LMT 184 haptic database. As expected, these features lead to very robust and sufficiently distinctive feature values and can be considered as robust to image rotation thanks to the TU3 extension. 71
4.11 Macrotexture (MaTX^{TUM}) feature components for compressed coconut fibers material sample. A spike detection algorithm based on vibrotactile signal statistics filters significant spikes which are most likely occurring due to coarse surface obstacles (top). Likewise, these surface obstacles can lead to tangentially recorded friction changes (middle) around the smoothed friction signal (red line). Independently, the IR sensor measures the distances while sliding over the material sample (below) and is likewise compared to its moving average (red line). 82
4.12 MaCO^{TUM} visual feature component, shown for makrolon material sample. The average of the n largest green circles (scale property) is related to the visual coarseness of the surface image. 83
4.14 Microtexture roughness feature visualization (top: vibrotactile, bottom: audio) for volcanic tuff material sample. Note that both the detail levels (red and blue) have similar signal energy on average, but reveal differences in their spikiness. The average of all elementwise-calculated differences is characteristic for a specific material recording (adapted from [7] © 2017 IEEE). 86
4.15 Microtexture coarseness MiCO^{TUM} feature data source (volcanic tuff material sample scan). The spectral centroid is located left to the center of the spectrum, and hence, the material structure can be considered as microscopically coarser than fine. 87
4.19 Damping-related data for material sample hemp. The current force at the maximum angular speed (dotted red line) divided by the maximum angular speed value is used to define the damping feature. 91
4.20 Tactile relaxation feature visualization. The red dotted line shows the maximum achieved pressure force during the EP pressure. It can be seen that the material sample relaxes until the end of the recording (i.e., slightly deforms) during the application of constant pressure of the TU2 unit 2. Consequently, CRX^{TUM} obtains a value larger than zero. 92
4.21 Thermal cooling curve for material sample hemp. The ambient temperature of the room is shown as dashed green line. The initial increase of temperature and ongoing decrease indicate thermal properties of the material sample. Note that the thermal cooling time τ is also characteristic for each material. 93
4.37 Different linear classifiers and hyperparameter configurations to visualize the different decision boundaries caused by the choice of γ. The soft margin parameter C is constant for all plots with $C = 1.0$. Large values for γ may lead to higher accuracies, but also tends to learn unreasonable decision boundaries. 115
4.38 Mahalanobis distances between data points from two features. The condition for Euclidean distance is not given, since the distribution scatters more along the x-axis and is rotated for several classes. 116

4.45 Selection of 24 materials from the LMT-184-HDB, denoted as LMT-24-HDB, which has been used for the subjective experiment. Three materials per major class C (separated by black lines) have been selected: C_1: Larch, Oak, Cork, C_2: Hemp, Sheep Wool, Bison Leather, C_3: PMMA, Isoprene, Polyurethane, C_4: Refractory Material, Terracotta, Ceramic Tile, C_5: Float Glass, Foam Glass, Glass Wool, C_6: Granite, Clay Shale, Marble, C_7: Stainless Steel, Lead, Steel Wool, C_8: Linoleum, Tar Paper, Sandpaper. 127
4.46 SRCCs of each individual subject's rating to the mean opinion score, i.e., overall mean rating. Note that, e.g., the ratings of subject 15 have a low correlation to the mean and also large variance across the individual tactile dimensions (black error bars). 129
4.47 Illustration of subjective feature space created from the best ten subjects ratings. The labels are consecutively labeled according to the 24 materials from Fig. 4.45. 130

5.1 Mobile World Congress teleoperation setup based on augmented force feedback haptic devices. The operator is able to explore the remote environment and gets thermal, force, and tactile feedback. 154
5.2 Tactile computer mouse evolution. The first version (left) and two improved design versions [5] (middle-left and middle right) are able to display the five relevant tactile dimension from [63] either using an embedded VCA or a C2 tactor, respectively. The current version TCM2 (right) is further able to recreate the fifteen subdimensions from Fishel et al. [66]. 155
5.3 TCM1 actuator placement. All actuation components are embedded inside the TCM1. The left and right click buttons of a common computer mouse are moved to the left and right side of the TCM1, respectively. The voice coil actuator is not visible in this image, since is located beneath the upper TCM1 body in between the Peltier elements (adapted from [5] © 2018 IEEE). 155
5.4 Each tactile dimension from [63] requires a different actuator. For example, the tactile dimension of friction is reproduced by adjusting the voltage of the built-in electromagnets to change the attraction forces on an iron plate, or fine roughness by displaying recorded vibrotactile signals on VCAs. The TCM actuation principle is highlighted in red font below each tactile dimension (adapted from [5] © 2018 IEEE). 156
5.5 Driver circuit flowchart for control signal mapping to the actuators (adapted from [5] © 2018 IEEE). 157
5.6 TCM1 virtual material display application overview. The user selects a material sample $S_{\text{<ID>}}$ and the corresponding model representation is fetched from the database. The physical properties are converted into signal representations and sent to the driver circuit from Fig. 5.5 using serial communication. The vibrotactile signals are sent via the sound card to the amplifier. The TCM1 actuators receive the signals and display tactile impressions to the user. The according on-screen pixel-coordinates $\mathbf{x}_{\text{p}}$ are used to calculate the required signals when the user moves the TCM1. The images $\mathbf{I}_{\text{disp}}$ are shown on a monitor and the current position in the application is shown at $\mathbf{x}_{\text{p}}$ as a 15-mm-radius sphere representing the human index finger. Note that the on-screen motion of this sphere was closely matched to the real mouse controller movement to avoid any visual-haptic speed differences (adapted from [5] © 2018 IEEE). 158

5.7 Final TCM2 mechanical parts schematic (left) and prototype (right). The human user places the index finger on the neodymium permanent magnet (NPM), or PM in short, which is firmly attached to the C2 tactor and the movable FSR mount for hardness rendering. A small wearable component allows the display of adhesion forces. The remaining part of the index finger skin surface receives controlled thermal feedback. The middle component (movable part) which holds the hardness and microscopic roughness actuators can adjust its height at two points to represent macroscopic roughness. The bottom part contains an electromagnet to adjust the friction between the TCM2 and the underlying iron plate (adapted from [3] © 2021 IEEE). 159
5.8 Applied EM voltage and resulting friction force for the TCM1. Similar to the approach proposed by Culbertson et al. [130], the relation between electric input and achieved force output was determined for the TCM1. A polynomial regression using a quadratic model interpolates the measured data. Without the electromagnetic force, the minimum traction is about 0.5 N, associated with $\mu_{\text{TCM,min}} \approx 0.20$. The maximum $\mu_{\text{TCM,max}}$ for 12.0 V supply voltage is hence $\mu_{\text{TCM,max}} \approx 2.69$ (adapted from [5] © 2018 IEEE). 161
5.9 Left: Mechanical friction coefficient acquisition setup. Since the width w and height h are known, the angle α leads to the finger-surface friction during the incipient slip event. The finger is put through a bearing which guarantees that the constant weight always points toward the Earth's gravitational center. Middle and right: Electromagnet voltage to friction coefficient estimation. An Optoforce (OMD-20-SA-40N) sensor embedded into a tightly fitting PLA part is magnetically coupled to measure normal and friction forces for different electromagnet voltages. The right plot shows which voltage is required to simulate a specific friction coefficient. Note that larger coefficients can be achieved by using a supply voltage larger than 12 V if required (adapted from [3] © 2021 IEEE). 162
5.10 Generalized Dahl friction model for three specific materials, from left to right: lead, rubber, silicone. The first row shows the same 1 Hz input periodical sliding motion for all three cases displayed. The second row shows the temporal evolution of the calculated friction force. The force-displacement curves in the third row summarize the first and second row and visualize the frictional displacement effect, i.e., the force depending on a specific position. Note that the third row y-axes are not likewise scaled in order to better visualize the stick-slip chatter, which becomes observably strong if the material's static and kinetic friction coefficients differ significantly. 163
5.13 Tangential macroscopic roughness image generation $\mathbf{I}_{macro,tang}$. The grayscale converted surface images provide the positions of laterally-perceivable edges (left). A Canny edge detector (threshold = 0.33, sigma = 4) identifies relevant edges and a dilation filter (middle) broadens the edge width. The value range of the IR sensor data, normalized between 0 and 1, is then used to set the intensity of the dilated edges (right) and thereby determines the intensity of the tangential obstacle (adapted from [5] © 2018 IEEE). 166
5.17 Consideration of anisotropic texture patterns. The real anisotropic material sample is scanned in its major direction and the resulting signal is added to the extracted edges of the macroscopic images during haptic display (adapted from [3] © 2021 IEEE). 171
5.20 Different supply voltages set on the Peltier element of TCM1 and resulting cooling curves. Note that humans generally touch materials 5 – 10 seconds during the EP static touch which is guaranteed providing an appropriate heat sink. 175

List of Tables

2.1 Cutaneous receptors in the glabrous (hairless) skin. Table adapted from [13], [24], [25]. . . 6

3.1 Material taxonomy table of the LMT-184-HDB [15] (adapted from [2] © 2019 IEEE). 43

4.1 Image feature dependencies with respect to rotation of images captured with the TU3 extension. The first three entries represent the slope, and the last three the standard deviations of three different material feature vectors consisting of 20 feature values. The closer an entry is to zero, the more invariant the feature is to arbitrary rotations. Features like blob-likeness, edginess, the RLBPs, color, or brightness are least affected by rotations. . . . 70

4.2 Speed dependency result table for audio features under increasing recording speed conditions. The material samples are M_1 PLA 8mm, M_2 PLA 2mm, and M_3 Granite. The first three entries denote the slope of the feature vector consisting of five entries (five different speeds), and the last three the standard deviations of the material feature vectors. Ideally, all values are close to zero to indicate independence of different recording speeds. Note that the audio signal energy (ASE) heavily depends on the recording speed, but other features like the zero-crossing rate (ZCR) are more robust to different speed levels. Nevertheless, these plots further show that any scanning system should avoid unnecessarily high scan speeds. 79

4.3 Commonly used vibrotactile features from related work which have also been examined in this thesis. 80

4.4 Speed dependency result table for tactile features under increasing recording speed conditions. The material samples are M_1 PLA 8mm, M_2 PLA 2mm, and M_3 Granite. The first three entries denote the slope of the feature vector consisting of five entries (five different speeds), and the last three the standard deviations of the material feature vectors. Ideally, all values are close to zero to indicate independence of different recording speeds. For example, MaCO^{TUM} reveals higher dependency of scan speed, since its visual component is multiplied by the vibrotactile signal energy, which itself heavily depends on the scan speed. Other features, such as FRS^{TUM} are more robust to different scan speeds. 94

4.5 Overview of features exceeding human touch discrimination capabilities, which are independent of the aforementioned audio, visual, and tactile data domains. 98

4.6 Best selected features from different domains. The last column aggregates the best features from the individual domains such as image or audio features. All rows are ordered based on the feature importance for each domain. Note that the column Misc* consists of all features that were either used in related work or do not belong to the other domains such as image or perceptually tactile features (adapted from [2] © 2019 IEEE). 109

4.7 Contingency table adapted from [203]. Note that each multiclass scenario, i.e., with more than two classes, can be reduced to a per-class binary confusion matrix (one-vs-all or one-vs-rest strategy [203]) allowing for the application of the same metrics. 111

4.8 Classification metrics [203] based on contingency matrix entries. . . . 111
4.9 Multi-class contingency table for specific material classification example based on the overall class labels C. . . . 112
4.10 Converted Multi-class contingency table for class C_3 plastics. . . . 112
4.11 Classification results of the initial 108 LMT texture dataset. . . . 122
4.12 SynTouch subdimensions categorized into EPs for the subjective experiment. The samples were placed and explored on a planar table. The asterisks for the phase pressing indicate that, additionally, free interaction with the material was allowed during the phase pressing which led to two different feature vectors for this EP. . . . 128
4.13 Subjects that did not pass the KS test (h = 1) on equal distribution to the MOS distribution. These seven subjects are considered as outliers and their ratings are not further used. . . . 129
4.14 Results of the subjective experiment with the 24 materials of the LMT-24-HDB. Compared to the Texplorer2-related features calculated and evaluated in Section 4.3, the voting-based classification performs significantly worse for a much smaller dataset (24 vs. 184 materials), supporting the initial assumption that a sensorized system outperforms human-made material votings. . . . 130
4.15 Comparison of the CNN performances on the TexTUM, ALOT and KTP-TIPS-2b datasets. Note that ALOT reports more material classes, however, these do not follow a consistent taxonomy. Consequently far less actual classes are present in the dataset with lesser number of samples compared to TexTUM-184, which is hence the largest image texture dataset currently available. . . . 141
4.16 All possible combinations of the HM-24 components, with the other components being disabled. I: TexTUM CNN, II: best 15 Texplorer2-based HC features, III: 15 tactile feature evaluations from subjective experiment, IV: SpectraNet. Note that the inclusion of the subjective ratings is decreasing the overall performance compared to the other sensor-based approaches, and also, if all four feature domains are considered. Consequently, the subjective ratings should be omitted from the fusion approach. . . . 149
4.17 HM-24 performance for different number and shape of hidden layers. The best configuration resulting from the smaller HM evaluation is used to identify the required depth and shape of the final HM-184 model. . . . 149
4.18 Final HM-184 results of the different models performance using 184 material classes. Note that the individual approaches have a much larger variance compared to the hybrid model, which also outperforms the individual approaches. . . . 150
4.19 Accuracy and taxonomy similarity overview of all classifiers. DL_{img} is the purely image-based TexTUM-184 DL approach, and SN is the SpectraNet. Note that ordinary classifiers based on handcrafted features lead to better results in total, and the fusion into the HM combines high accuracies as well as a high taxonomy similarity for the misclassified samples. . . . 151

5.1 Surface images of material samples, denoted as $S_{<ID>}$, used for the TCM1 display application and calculated model parameters to represent the five tactile dimensions of the ten material samples. Note that the term real material (RM) is also used as synonym. The spring stiffness of the thin sandpaper sheet is identical to the stiffness of the underlying wooden table. MaRS is the macroscopic roughness strength from Section 4.1.4 which determines the existence and intensity of coarse structures on a surface. MSE is the mean signal energy of a vibrotactile signal recorded at 60 mm/s (adapted from [5] © 2018 IEEE). 157

5.2 TCM1 per-class precisions (recognition probability) from confusion matrix. Virtual materials like stainless steel were easily identified and matched to the real material counterpart, whereas wooden materials such as compressed wood and oak were partly confused. . . . 178
5.3 Rendering parameters [186], [284] for ten selected materials (RM0 – RM9) used in the subjective experiment (reproduced from [3] © 2021 IEEE). 180
5.4 Experimental results averaged across subjects. Top row: per-class precisions from confusion matrix. Bottom row: per-class virtual–real visual-haptic similarity ratings. 181