

# Signal Processing for Haptic Surface Modeling: a Review

Antonio Luigi Stefani<sup>a,c,\*</sup>, Niccolò Bisagno<sup>a,c</sup>, Andrea Rosani<sup>b</sup>, Nicola Conci<sup>a,c</sup>, Francesco De Natale<sup>a,c</sup>

<sup>a</sup>*University of Trento, Via Sommarive, 9, Trento, 38123, Trento, Italy*

<sup>b</sup>*Free University of Bozen-Bolzano, Piazza Università, 1, Bolzano, 39100, Bolzano, Italy*

<sup>c</sup>*CNIT - Consorzio Nazionale Interuniversitario per le Telecomunicazioni, Via Sommarive, 14, Trento, 38123, Trento, Italy*

## Abstract

Haptic feedback has been integrated into Virtual and Augmented Reality, complementing acoustic and visual information and contributing to an all-round immersive experience in multiple fields, spanning from the medical domain to entertainment and gaming. Haptic technologies involve complex cross-disciplinary research that encompasses sensing, data representation, interactive rendering, perception, and quality of experience. The standard processing pipeline, consists of (I) sensing physical features in the real world using a transducer, (II) modeling and storing the collected information in some digital format, (III) communicating the information, and finally, (IV) rendering the haptic information through appropriate devices, thus producing a user experience (V) perceptually close to the original physical world. Among these areas, sensing, rendering and perception have been deeply investigated and are the subject of different comprehensive surveys available in the literature. Differently, research dealing with haptic surface modeling and data representation still lacks a comprehensive dissection. In this work, we aim at providing an overview on modeling and representation of haptic surfaces from a signal processing perspective, covering the aspects that lie in between haptic information acquisition on one side and rendering and perception on the other side. We analyze, categorize, and compare research papers that address the haptic surface modeling and data representation, pointing out existing gaps and possible research directions.

## Keywords:

Haptic Surface Understanding, Signal Processing, Virtual Reality

## 1. Introduction

The primary objective of extended-Reality (XR), in all its possible variants (virtual-, augmented-, mixed-) is to either faithfully replicate real-world experiences or enhance them [1, 2], enriching the real-world information by overlaying additional virtual contents[3, 4]. To effectively complement and/or replace their actual real-world counterparts, virtual experiences must be designed to engage as much as possible the various human senses [5].

For decades, visual and acoustic stimuli have been the focal points of XR solutions, due to the large availability of reliable and affordable technologies. For example, current AR/VR headsets are able to render realistic immersive environments with synchronized and spatialized audio-visual information. However, whenever the application requires the interaction with virtual objects, the need of engaging user's touch emerges, being the only way to provide a realistic perception of the contact with objects. Notwithstanding their importance in this respect, haptic technologies have remained relatively unexplored, lagging behind the rapid development of spatial audio and high-definition images, and with very little impact on the market, especially at a consumer level.

To behave in a more natural way when interacting with a virtual environment, users' sense of touch should also be stimulated, allowing them to grasp a virtual object, and feel the roughness of its surface. To do so, researchers focused on how to generate haptic feedback as close as possible to the real world. This implies a number of operations to be performed, which span different technological domains. Accordingly, we propose in Fig. 1 a breakdown of the haptic rendering process that defines in a clear and interpretable manner the steps that are required to effectively achieve such feedback.

In particular, the overall process can be split into five main blocks, namely: (I) sensing, (II) haptic data processing and modeling, (III) representation and communication, (IV) rendering, and (V) perception. Sensors, such as inertial or vision-based sensors, are employed to scan objects and capture their haptic properties, such as stiffness, roughness, and slipperiness. Haptic data processing and modeling focuses on the encoding of the collected physical information to some digital format (images, haptic maps, spectrograms) and its processing mechanisms. This is also referred to as Haptic Surface Modeling. The block of representation and communication of haptic data studies how to efficiently compress, encode and transmit such information. Rendering is performed by haptic devices (e.g., haptic gloves, haptic displays), which are able to generate an appropriate stimulus starting from the captured model. Finally, user's feedback (perception) and the Quality of Experience (QoE) are

\*Corresponding author

Email address: antonioluigi.stefani@unitn.it (Antonio Luigi Stefani)

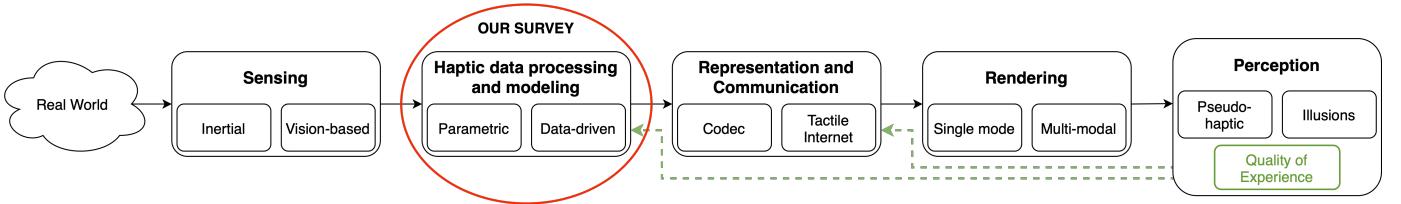


Figure 1: The haptic processing pipeline consists of 5 main steps: sensors to acquire the data, processing and modeling to store acquired data, communication to transmit information associated with the collected signals, haptic devices to render the haptic feedback to the user, and perception that studies the factor that influence the haptic experience of the user.

studied to understand how close the haptic interaction is to the real-world touch feeling, making use of pseudo-haptic and haptic illusions techniques.

In Tab. 1, we list the surveys currently available in the literature, covering the different areas of the haptic pipeline. As can be seen, the focus of such works is either on sensors, or on the later stages of rendering and perception. A few works are also classified as 'general' (dealing with cross-disciplinary research in collateral areas, such as psychology, perception, or engineering), and application-oriented (dealing with real-world setups and systems that make use of haptics). None of the listed works, however, explores in detail the available literature in the area of Haptic Surface Modeling.

Topic	Ref.	Description
<b>General</b>	[6, 7, 8, 9, 10, 11]	General topics with convergent research on haptic perception, mechanics, electronics, and material technologies
<b>(I) Sensors</b>	[12, 13, 14, 15]	Hardware-related topics to collect haptic data, such as vision-based and inertial sensors
<b>(II) Haptic data processing and modeling</b>	Our	Focusing on Haptic Surface Modeling, thus how the haptic data can be stored, processed and analysed
<b>(III) Representation and communication</b>	[16, 17, 18, 19, 20]	Focusing on allowing the transmission of haptic information over a network (Tactile Internet)
<b>(IV) Haptic devices</b>	[21, 22, 23, 24, 25, 26, 27, 28, 29]	Multimodal and single mode devices to convey the haptic sensation to the user
<b>(V) Perception</b>	[30, 31, 32]	Focusing on how realistic is the tactile sensation perceived by the user
<b>Applications</b>	[33, 34, 35, 36, 37, 38, 39, 40]	Various applications of the haptic rendering process

Table 1: Available surveys in the domain of haptics focus on different steps of the haptic rendering process, but none of them zeros in on the data representation for haptic signals.

In this survey, we aim at reviewing such missing key aspect, i.e., the representation of real-world haptic features in the digital domain. We provide a comprehensive review on the representation of haptic surfaces from an image and signal processing perspective. We examine, which data are mostly used in today's applications, looking at the existing datasets, and we indicate how they are being adopted by the research community. Drawing a parallelism with the research in the visual and acoustic domains, we highlight the shortcoming of the current haptic processing pipelines, pointing out possible research directions.

The paper is organized as follows. In Section 2, we introduce the definitions that will be used throughout the paper. Section 3 delineates the current processing pipelines for the visual and acoustic signals and compares them to the haptic one, highlighting

ing differences and similarities. In Section 4, we present the available literature on Haptic Surface Modeling and processing. Section 5 showcases available datasets for haptic processing research, while in Section 6 we highlight current research tasks, including baselines and metrics. Section 7 concludes the paper, providing a critical discussion and highlighting future trends and open research issues.

## 2. The Haptic Processing Pipeline: Definitions

According to the representation in Fig. 1, when dealing with haptic technologies we can identify five main building blocks, which contribute to the definition of the overall pipeline, namely: sensing, processing and modeling, representation and communication, rendering, and perception. In this paragraph we briefly describe each of these elements, introducing definitions and concepts that will be used throughout the paper, and providing links to the relevant surveys for each specific topic.

### 2.1. Sensing

Haptic sensors aim at mimicking the receptors that are present in our fingers and hands, to extract essential tactile information. Two main types of sensors exist: inertial and vision-based. Inertial sensors adopt various technologies (resistive, capacitive, fiber-based) to detect the roughness of the surface through the contact with it. They produce one-dimensional signals that capture the forces of interaction between sensor and object, which convey information about surface properties. Inertial sensors have been largely employed in robotic applications [13, 12], where tactile sensing refers to the point of contact between robot fingers and objects.

Vision-based sensors [15, 14] have recently emerged as the most simple and affordable way to capture tactile information at high resolution from images or video. The use of images and video appears to be very intuitive to tackle the problem. This is generally implemented using a RGB camera recording the deformation of a membrane upon its contact with a surface, thus generating a haptic map.

### 2.2. Haptic data processing and modeling

Haptic data processing and modeling, also called Haptic Surface Modeling, investigates how the sensed haptic signals are

handled and stored in digital form. Unlike other sensory domains such as visual [41] and acoustics [42], where standardized representations are commonly available, the representation of tactile stimuli remains an area without established standards. In the early stages, haptic surface modeling relied on parametric functions to model the surface shape [43]. However, this approach either introduced a high amount of noise and approximation in the model or relied on too complex mathematical expressions. Consequently, researchers have opted for a paradigm change, moving towards data-driven approaches [44], where the object's haptic information is collected as a one-dimensional signal, associated to classical representations such as spectrograms. The natural evolution of these initial approaches was then to turn to 2D signals, such as images and videos as well as haptic maps, which provide a much richer information that can be approached through well-studied computer vision techniques.

In this survey, we focus on these aspects, investigating the different options in terms of processing and highlighting advantages and shortcomings for each of them.

### 2.3. Representation and communication

Even more than its visual and audio counterparts, the haptic signal needs to rely on low latency communication to develop real-time interactive systems [19]. Potential applications include tele-surgery, industrial processes automation and augmented reality for live music concerts [16]. To enable such applications, researchers have focused on building the so-called Tactile Internet [16, 45, 19], elaborating on suitable technologies and communication protocols in the different layers of an architecture, such as the information representation using suitable codecs [45], as well as network design and customer's haptic devices.

### 2.4. Haptic devices

Haptic devices aim at providing the tactile experience to the user, effectively translating digital data into physical stimuli that hit the receptors of human haptic channels [28]. The haptic stimulus can be delivered using single mode or multi-modal devices. Among single mode devices, surface haptic displays [25] consist of an active physical planar device that deforms so as to reproduce a surface model, usually sensed by the bare finger. Differently, multimodal haptic devices, like hand-held (joystick-like, unconstrained devices), wearables (usually gloves) and encountered/grounded (surfaces/devices for the user to touch "on-demand") [28, 22, 21, 29] aim at exploiting different stimuli, such as forces, vibrations and thermal, to convey the haptic sensation to the user.

### 2.5. Perception

Human haptic perception encompasses a range of sensory modalities, which can act externally in the form of cutaneous sensing, and internally as proprioception and kinesthesia [30]. While single mode haptic devices aim at reproducing pseudo-haptic feedback [31, 32], they struggle to fulfill the whole range

of human haptic perception. Multi-modal devices could expand the spectrum of haptic feedback, reproducing the sensations produced by touching an object rather than its physical features, thus creating the so-called haptic illusions [30]. In human-centered applications, Quality of Experience (QoE) plays a crucial role in influencing the design and overall success of the whole pipeline. However, QoE in haptics is usually evaluated through subjective tests with the user-in-the-loop [45]. The desirable evaluation of QoE through objective testing aims at defining algorithmic models of human perception to calculate meaningful perceptual metrics [45, 7]. Other than human-based applications, perception is also key in robotics, for instance when dealing with object grasping [38, 46]. In such scenarios, understanding objects' properties plays a fundamental role to modulate the applied force in the grasping action and to minimize failures.

## 3. Real world objects representation in the physical and the perceptual spaces

Real-world objects are defined by their physical characteristics, such as color, shape, roughness, and produced sound. Our perception of an object depends on these characteristics as well as on a number of subjective factors, such as sensory sensitivity and cognitive processing. To develop a digital model for virtual interactions, both physical and perceptual characteristics must be considered.

Fig. 2 illustrates the modeling pipeline, which connects physical phenomena sensed by specific sensors, the physical feature space that defines the object's characteristics, and the perception feature space that identifies its properties as they are perceived by humans. While physical and perceptual spaces are well-established for images and audio signals, haptic signals still lack a standard representation able to characterize and replicate the relevant properties [47].

In the following paragraphs we highlight the existing definitions of the physical and perceptual spaces for both visual and acoustic domains, trying to draw a parallelism with the haptic domain, and highlighting gaps and possible research directions.

### 3.1. Physical feature space

The physical feature space contains the characteristics to be captured in order to represent the physical phenomenon of interest. The role of the sensor is to capture one or more physical characteristics of the phenomenon in a given domain. As an example, cameras provide information in the visual domain by capturing wavelengths in the visible (or infrared) spectrum, while microphones record sounds in terms of frequency, amplitude and phase. When dealing with haptics, roughness and stiffness are among the most common features to be sensed. To better explain haptic sensing, in the next paragraphs, we will draw an analogy between the haptic domain and the more familiar visual and audio domains.

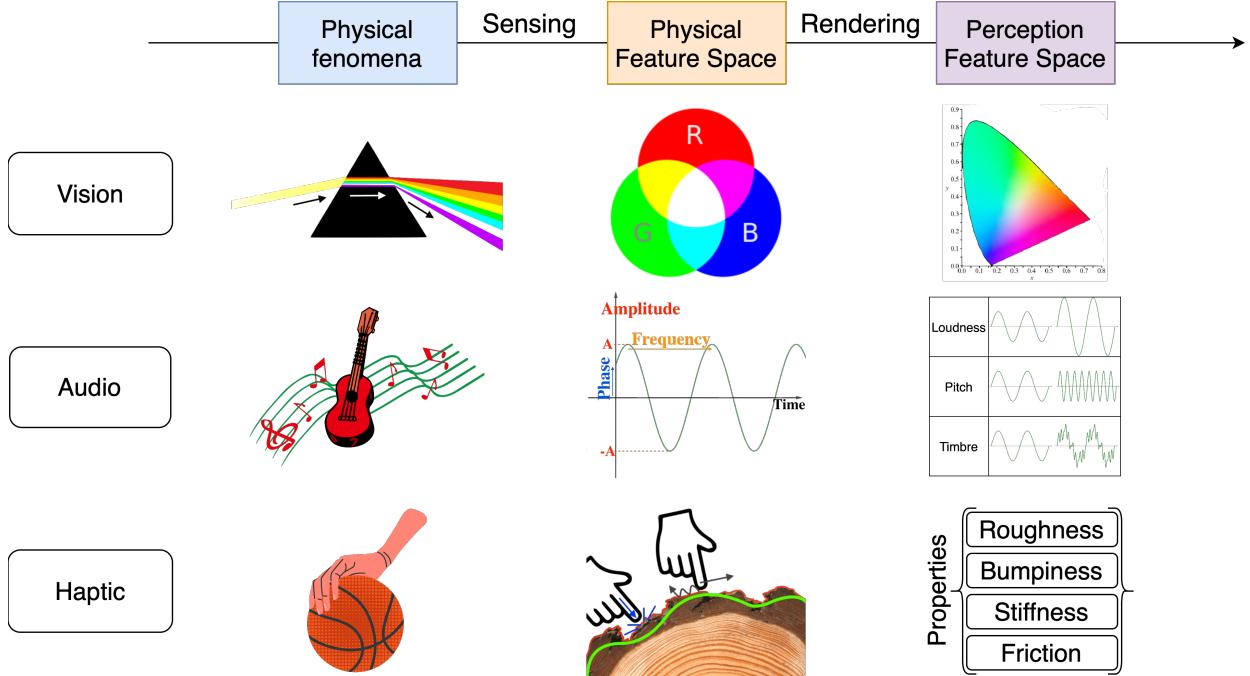


Figure 2: The classic signal processing pipeline (top row) aims at mapping a physical phenomenon to a physical feature space such that a similar experience of the phenomenon can be delivered to the user in the perception feature space (bottom row). While the pipeline is well-established in both the visual and acoustic domains, it still lacks a common standard in the haptic domain.

**Visual and acoustic domains.** The signals captured by cameras and microphones are mapped to standardised physical spaces. In the visual domain, light waves are filtered to obtain the RGB color model [41, 48], which resemble the wavelengths perceived by human eyes. In the acoustic domain, sound waves are one-dimensional signals, described by frequency, amplitude, and phase [42, 49].

**Haptic domain.** The properties describing the haptic qualities of an object in the physical space and a standard model encapsulating them did not reach a consensus among researchers yet. Nevertheless, as will be seen in the next paragraphs, researchers agree on the inclusion of some common macro categories as, for example, the concepts of roughness and friction.

The first attempt to define a haptic physical space and to correlate the physical properties of a material with their perception is described in [50]. In this work, the authors identified four physical properties that characterize a surface, namely: roughness, compliance, friction, and cooling rate. According to the authors, roughness measures the variations of the height of the surface, compliance measures its softness, friction measures the sliding resistance of the surface, and cooling rate measures the variations of temperature across it.

More recently, in [51, 52] the authors proposed a new way to characterize objects according to a set of haptic properties that can be considered orthogonal in what they called Haptic Attribute Space (HAS). In particular, the attributes that describe the four-dimensional HAS space are defined as follows: **roughness** (or micro-roughness) corresponds to the finer spaced irregularities of the surface texture that usually result from the inherent action of the production process or material condition

[53]; **bumpiness** (also called macro-roughness or waviness) is associated to the macro-behavior of the surface texture, such as ridges, bumps and valleys; **stiffness** (or elasticity), it is the term used to describe the force needed to achieve a certain deformation of a structure [54]; **friction** (or slipperiness) refers to the resistance that one surface or object encounters when moving over another. In other words, it is the force that opposes the relative motion or tendency of such motion of two surfaces in contact [55]. In fig. 3, a visual example is depicted of the above properties.

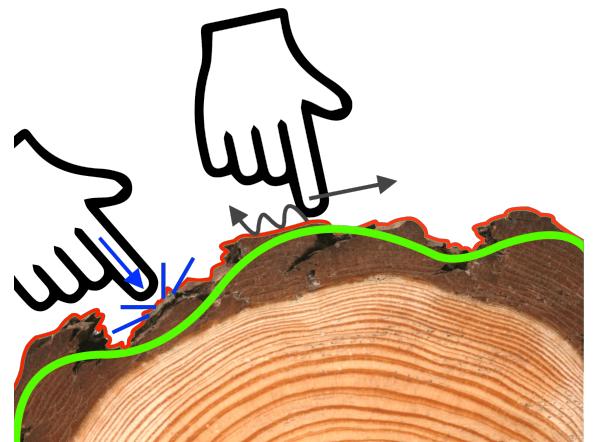


Figure 3: In this example illustrating the section of a tree the red line is the roughness, the green line is the bumpiness, the blue line represents the stiffness, and the dark gray line is the friction.

Despite some attempts to define a standardized physical fea-

ture space for haptic data, a universally accepted representation has not emerged yet. Furthermore, the physical space is often conflated with the perception space, and this frequently generates confusion and misunderstandings when describing applications related to sensing and actuating surfaces that recreate haptic perceptions.

### 3.2. Perception feature space

The perception feature space aims at describing the properties that enable us to sense and interpret information from the surrounding world. These properties have been the subject of study for decades, as they inform the design of sensors and actuators. Also in this case, to better highlight the underlying concepts, we propose a simple parallel with other known perceptual spaces, in the visual and audio domains.

*Visual and acoustic domains.* The color feature space describing the human perceptual system is the CIELAB standard [56], which defines the relation between the wavelength distributions in the electromagnetic visible spectrum and the psychologically perceived colors in human vision [57]. In the acoustic domain, perception is characterized by pitch, loudness, and timbre, corresponding to frequency, amplitude, and phase in the physical space. Pitch relates to the perception of fundamental frequency [58, 49], loudness refers to sound intensity [59, 60], while timbre represents the tonal quality that distinguishes different sounds [61].

*Haptic domain.* Also in this case, differently from visual and acoustic stimuli, haptic stimulus is quite complex to describe from a perceptual viewpoint. In fact, understanding which are the surface properties within the physical space that influence our perception when interacting with an object is still an open problem [25]. Therefore, many studies have been conducted to identify an orthogonal perceptual feature space that optimizes the representation of the properties to transmit to our skin, allowing us to recognize and categorize the surfaces.

In early studies, such as [62], the authors suggested a possible two-dimensional space that describes the haptic perception in terms of roughness and hardness. Subsequently, slipperiness was added as a third dimension [63]. In [64], which has become the most common representation, the haptic features that compose the perceptual space were extended to four, namely roughness, compliance, coldness and friction. However, in [65], the authors further refined this model, proposing a 5-dimensional haptic perception space that included macro-roughness, micro-roughness, coldness, hardness, and friction. Other studies [66, 67] tried to apply PCA on a big set of properties classified by users, to identify the components that carry more information, in an attempt to minimise the dimensional space. In the opposite direction, other studies proposed to further enlarge the dimensional space. In [68], for instance, two new elements called "volume" and "naturalness", were added to the perceptual space. According to the authors, "volume" describes a shape property of the surface and "naturalness" measures how much a surface is close to the natural world (as opposed to the artificial world). Another study analyzed the neural cortex response of primates and humans when scanning a

surface [69]. The authors claim that the perceptual representation of a texture is high-dimensional and that the four properties usually adopted are not capable of conveying all the descriptors that are needed to uniquely describe a surface.

Notwithstanding all the recent research efforts, the four-dimensional model identified in 2010 [64] remains the one the research community still agrees more. Even when agreeing on these dimensions, however, it is still unclear which features carry more perceptual information (e.g., the L component in the CIELAB model conveys more information to the human eye than the other two), and how to model the relationship among features (e.g., in CIELAB, the non-linear relation between its components try to mimic the non-linear response of the human visual system).

### 3.3. Linking physical and perceptual feature spaces in haptics

While the haptic perception feature space has received broader attention from researchers, it is still unclear how to establish a direct link between the physical feature space and the perceptual feature space [70, 71]. A first attempt to define this mapping function was made in [71], where the authors worked at the mapping between textures acquired by a sensor and relevant features, such as roughness and bumpiness, as perceived by the users involved in their experiment.

In short, despite the available literature, we are still far from establishing a universally accepted link between the two spaces (physical and perceptual), which remains an open problem.

### 3.4. About the temperature

Temperature is a fundamental dimension of the haptic space, significantly affecting our exploration of the world and our interaction with it. It is often considered an independent property, because the sensors required to measure it (both from human and machine viewpoints) are different and complementary to those needed for other tactile properties [65].

In physics, temperature is clearly defined as one of the fundamental dimensions in the International System of Units [72]. However, in the perceptual domain, human thermal sensitivity depends on the physical characteristics of the object in contact. Specifically, the perceived temperature depends on the thermal transfer between the object itself and the human skin. This thermal interaction involves complex mechanisms that include the thermal conductivity and heat capacity of the materials involved, as well as the surface geometry and the nature of the contact [73].

When two objects of different materials are touched, the perceived temperature is influenced by the rate of heat transfer. For instance, metals typically feel colder than wood at the same temperature because metals have higher thermal conductivity, allowing heat to be transferred away from the skin more rapidly. This concept, known as thermal diffusivity [74], explains why certain materials feel warmer or cooler upon contact despite being at the same physical temperature. Moreover, the perception of temperature is not only a function of the material properties but also involves cognitive processes. Our brain integrates these thermal cues with other sensory information to form a coherent

perception of the material, influencing our ability to identify and discriminate between different objects based on their thermal properties.

Given the independence and uniqueness of the temperature with respect to other haptic properties, we do not delve into the details in the scope of this survey. For further analysis, please refer to the following comprehensive review [73].

#### 4. Haptic Surface Modeling and Processing

Haptic surface modeling aims at mapping a physical object surface into the corresponding physical feature space of choice (i.e., roughness, bumpiness, stiffness and friction). Along the modeling, physical features may require some processing, filtering, and adaptation to be transmitted to an haptic device and eventually rendered in the perception space.

Being one of the main building blocks between the physical space and the perception, haptic surface modeling plays a crucial role in delivering a satisfactory experience. In this section we delve into the different methods proposed in recent research to model the haptic characteristics of a surface.

As shown in Fig. 4, existing modeling approaches can be divided in two main categories: parametric approaches [43], where the object surface is modeled as a parametric function and the more recent data-driven approaches [25], where the mapping between a surface and the relevant physical features is learnt from real data samples.

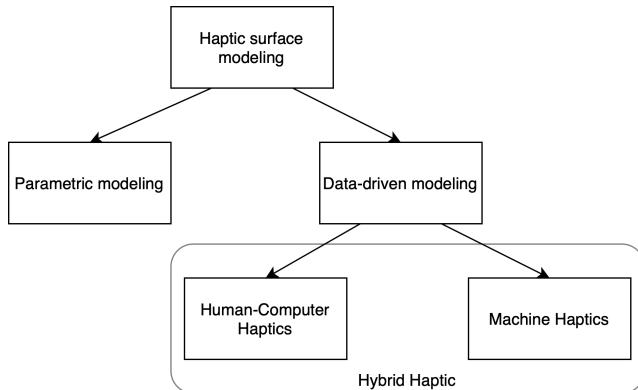


Figure 4: Haptic surface modeling approaches can be categorized into parametric and data-driven methodologies. Furthermore, data-driven methods can be subdivided into human-computer haptics, which model haptic properties with a focus on human-based applications, and machine haptics, which model haptic properties specifically for robotic applications.

##### 4.1. Parametric approaches

Parametric approaches involve using physical, mathematical, and mechanical principles to simulate the surfaces properties. Due to their simplicity, explicit models were popular in the early studies about haptic rendering [43, 104, 105]. The parameters of the modeling functions can be set through measurements taken from real objects. A glance into these methods can be found in [43]. Notwithstanding their attractive features, parametric models show two main limitations. Firstly, they require some a-priori knowledge of the physical phenomenon to

be simulated, and this is not always available. Secondly, except for some simple cases, they require considerable sophistication to make them suitable for real-time rendering, where complex and realistic scenes with non-linear behaviors lead to heavy computation [106]. Due to these problems, such methods have found limited application up to now, and researchers have moved towards data-driven approaches, which do not require any a-priori knowledge of the underlying physical world.

##### 4.2. Data-driven approaches

The general pipeline of data-driven approaches is to collect the signal generated by a sensor that explores the surface, and map it into a physical counterpart in the feature space. The mapping function is directly learnt from real data samples belonging to a given category. Notably, the learned mapping should be successfully applied to model the haptic features of unseen objects, thus demonstrating its generalisation ability. Data-driven approaches, reported in Tab. 2, can be classified as either **computer haptic**, which refers to human-computer interaction applications, **machine haptic**, which refers to robot-based applications, or **mixed haptic**, which mixes data usually employed by computer haptic or machine haptic, depending on the collected signal.

*Human-computer haptic signals.* Applications based on the so-called computer haptic aim at modeling the haptic properties of a surface by mimicking the interaction between the surface and the human finger [75]. The usual pipeline involves the use of sensors, which touch the surface of interest and produce one or multiple one-dimensional signals corresponding to the haptic information, as shown in Fig. 5 (a). The most widely used sensors are the inertial ones, such as accelerometers, probes and force sensors.

The collected signals depend on the motion and position of the sensor when touching the object: *position*, meaning the displacement of the sensing device, *velocity*, namely how fast the sensing device moves, *sliding velocity*, which refer to how fast the sensing device moves on the plane tangential to the surface and, similarly, *acceleration*, *force*, and *sliding force* to which the device is subject.

As shown in the upper part of Tab. 2, various studies have explored different one-dimensional signals translating them into the physical features. In [44], the authors focused on modeling visco-elastic properties of surfaces, particularly silicon-based and rubber materials. They measured position and force applied simultaneously, using interpolation methods to model transient behaviors. In [75], they utilized a Generalized Maxwell Model for elasticity, focusing on visco-elastic bodies and viscous fluids. In [76], they examined bumpiness and roughness using acceleration and force data collected with a What probe, applying an IIR filter for their modeling. The work in [77] assessed human force sensing capacity and evaluated their previous modeling method [44] by testing on two silicon cylinders. The authors used both static and dynamic models based on radial basis function interpolation. In [78], the authors reduced a 3D high-frequency acceleration signal to a 1D signal using their DFT321 transformation. This method approximates

Signal	Year	Method	Collected signal	Physical features				Exploration	Neural-based
				Bumpiness	Friction	Roughness	Stiffness		
Human-Computer Haptic	2008	[44]	F, P, SF ,V	✗	✓	✗	✓	Human	✗
	2009	[75]	F, P, SF ,V	✗	✓	✗	✓	Human	✗
		[76]	A, F	✓	✗	✓	✗	Human	✓
	2010	[77]	F, P, V	✗	✗	✗	✓	Human	✗
		[78]	A	✓	✓	✓	✓	Human	✗
		[79]	A, F	✓	✗	✓	✓	Human	✗
	2011	[80]	A, F, P	✓	✓	✓	✓	Human	✗
	2012	[81]	A	✓	✓	✓	✓	Human	✗
	2013	[82]	A, F, V	✓	✓	✓	✓	Human	✗
		[83]	F, P	✗	✓	✗	✓	Human	✗
	2014	[84]	A, F, SV	✓	✓	✓	✓	Human	✗
	2015	[85]	A, F, V	✓	✓	✓	✓	Machine	✓
	2016	[86]	A, F, P	✓	✓	✓	✓	Human	✓
		[87]	F, P, SV	✗	✓	✗	✓	Human	✓
		[88]	A, F, V	✗	✓	✓	✓	Human	✗
	2018	[89]	A, F, P	✓	✓	✓	✓	Human	✓
	2020	[90]	A, F, P	✓	✓	✓	✓	Human	✗
	2022	[91]	A, F, P	✓	✓	✓	✓	Human	✓
		[92]	F, P	✗	✗	✗	✓	Machine	✓
	2023	[93]	F, P, V	✗	✗	✗	✓	Machine	✓
Machine Haptic	2016	[94]	Img	✗	✗	✗	✓	Human/Machine	✗
	2017	[95]	Img	✓	✓	✓	✗	Human	✗
		[96]	Vid	✗	✗	✗	✓	Human/Machine	✓
	2021	[97]	Img	✓	✗	✓	✗	Human	✓
		[98]	Img	✓	✗	✓	✗	Machine	✓
		[99]	Img	✓	✗	✓	✗	Machine	✓
Mixed Haptic	2023	[100]	Img	✓	✗	✓	✗	Machine	✓
	2018	[101]	Img, F, A	✓	✓	✓	✓	Machine	✗
	2020	[102]	Img, F, A	✓	✓	✓	✓	Machine	✗
		[103]	Img, V, F	✓	✓	✓	✓	Not specified	✓

Table 2: A = Acceleration, F = Force, P = Position, SF = Sliding Force, V = Velocity, SV = Sliding Velocity

the 3D signal by aligning the module and phase through Discrete Time Fourier Transformation. In [79], they used a sensor fusion technique combining haptic data from a probe with visual data from a camera. They applied the Hertzian contact model to describe the stiffness of objects, achieving a 3D profile from the collected acceleration and force data. Building on [78], [80] used Linear Predictive Coding (LPC) to model time-domain patterns from 3D data reduced to 1D, creating a look-up table for data prediction. Subsequently, in [81], they enhanced the method from [80] using an Auto-Regressive Moving Average (ARMA) model, combining past outputs and weighted moving averages of past inputs to associate time-domain patterns with spatial domains. In [82], data were collected with users moving freely along surfaces. The authors used an Auto-Regressive model, assuming signal stationarity and segmenting signals into chunks using an Auto-PARM algorithm. The study by [83] relied on a PHANToM Premium 1.5/3DOF device and an ATI Nano 17 force/torque sensor to measure force and position, focusing on friction and stiffness. In [84], they continued with DFT321, using an AR model and Auto-PARM algorithm for signal segmentation, and employed Delaunay triangulation for interpolation. The authors of [85] modeled isotropic haptic textures with a frequency-decomposed neural network, demon-

strating its applicability to anisotropic textures. The data were collected using a Texanner and processed with a neural network model. In [86], they segmented the data using a bottom-up algorithm and interpolated it with a Radial Basis Function Network (RBFN) to estimate Line Spectrum Frequency coefficients. They analyzed homogeneous anisotropic grain textures, scanning in two directions. In [87], the authors used a force sensor and PHANToM device, developing an interpolation model based on the proxy position of an imaginary contact point. They identified the sliding yield surface through automated palpation. The authors in [88] incorporated Dahl friction model for friction, labeled stiffness with the speed of the tooltip, and used an AR model for roughness. They rendered virtual surfaces with a haptic interface augmented with a Tactile Labs Haptuator. In [89] they followed the segmentation and interpolation methods from their previous work, analyzing bumpiness, friction, roughness, and stiffness using an accelerometer, force sensor, and PHANToM device. Next, the same authors, in [90], proposed another method to collect and process data. They demonstrated that by combining their segmentation framework and a user interface, which guides the data collection process to create a human-in-loop system, the approximation quality of the model increases. In [91], they trained a neural network with contact

acceleration data collected through a manual scanning stylus. They used attention-aware 1D CNNs and encoder-decoder networks with Bi-LSTM to capture spatial and temporal dynamics. The authors from [92] used a force-feedback device and load cell to measure force and position, employing Radial Basis Function and Random Forest models for interpolation. Lastly, in [93] they utilized a CatBoost approach, which is a variant of gradient boosting. Here they train decision trees to learn the required mapping function for modeling the objects. A notable mention should be made of the work presented in [78, 80], which introduced the DFT321 method. This technique substantially contributed to the modeling of haptic properties and led to the creation of the first dataset containing data on surface haptic properties [107]. However, the DFT321 approach relies on lookup tables to model surfaces, which limits its ability to generalize effectively to new, unseen surfaces. In contrast, machine learning techniques offer enhanced generalization capabilities, enabling more robust and versatile modeling of haptic properties across a broader range of surfaces.

Concluding, human-computer haptic signals have long been the focus of Haptic Surface Modeling thanks to their simple processing techniques, which allows to immediately characterise the features of a specific point on the surface. However, the same sensed signals (e.g., acceleration, force) have been modeled into different physical features depending on the design of each sensor and the dataset being used. This makes it difficult for researchers to perform comparative studies. As an example, relating how different sensors characterise the physical features of the same surface would help in comparing the different modeling techniques. The other main limitation of human-computer approaches is the difficult in geometrically relating different points in space.

*Machine haptic signals.* While human-computer haptic applications have been developed to work with humans using one-dimensional signals, machine haptics aims at making a robotic hand deal with objects to accomplish different tasks, such as object manipulation [38]. To achieve this goal, object geometry must be made available. Therefore, machine haptic generally relies on vision-based sensors to obtain 2D high-resolution information in the form of RGB images and haptic/tactile maps, as shown in Fig. 5 (b). The sensors capturing these data rely on a RGB camera recording the deformation of a membrane upon its contact with a surface. The images of the deformed membrane can be directly used as an haptic map or they can be used to estimate the depth of an object representing its shape. In [94], a method for estimating object hardness using a touch sensor and linear regression is proposed, introducing the GelSight sensor. The development and application of the GelSight sensor are thoroughly reviewed in [95]. This paper covers the sensor's optical system design and algorithms for deriving the haptic physical features from the images. GelSight's performance in geometry and force measurement is experimentally evaluated, demonstrating its capabilities in capturing friction, roughness, and bumpiness. In [96], a neural network is employed to map an image sequence to a scalar hardness value. This approach measures the hardness of objects with different shapes under loosely

controlled contact conditions, utilizing both manual and robotic interactions. In [97] the authors introduce the GelStereo sensor and the method that converts tactile information into geometric deformation of a silicon gel layer, captured using visual sensing methods. The study employs a self-supervised GS-DepthNet estimation technique to enhance the accuracy and resolution of the contact geometry, focusing on roughness and bumpiness. The DelTact sensor, discussed in [98], uses a Gaussian density regression model to estimate roughness and bumpiness from an image. The model solves a point cloud that aligns with prior depth information, ensuring depth consistency and generalizing across sensors with minimal parameter tuning. In [99], the DenseTact model estimates depth images from RGB input images using a DenseNet. This method aims to capture roughness and bumpiness by converting tactile information into depth images. The improved DenseTact2.0 sensor is presented in [100], which includes an encoder-decoder network for shape estimation and force estimation models using transfer learning.

Differently from human-computer haptic signals, machine haptic applications leverage the larger expressiveness of images to represent the data acquired, introducing a new aspect to haptic sensing. By directly sensing the surface geometry, it is possible to achieve higher precision and richer data representation. This capability allows robotic systems to better understand and interact with their environments, facilitating more complex tasks such as detailed object manipulation, texture recognition, and adaptive control based on the tactile feedback received.

*Mixed haptic modeling methods.* Human-computer and machine haptic applications have traditionally been studied separately due to their different requirements, each facing limitations such as difficult in scaling to the large variety of real materials or inadequate modeling of skin-material interactions. Recently, researchers have proposed methods that integrate the two approaches to model the haptic properties of surfaces. These methods use vision-based sensors or standard cameras to estimate 2D geometric data and inertial sensors for stiffness and friction. In [101], the authors combined data-driven and parametric modeling to estimate the haptic properties of surfaces. They used a photometric stereo algorithm to estimate the bidirectional reflectance distribution function (BRDF), modeling surface texture through data-driven methods for roughness and bumpiness, and physical-based approaches for stiffness and friction. In [102], the model is enhanced by incorporating an LPC-based method to capture vibrations. By merging force and vibration feedback models based on their spectral characteristics, a comprehensive hybrid framework is devised, which allows for more accurate 3D surface reconstruction. In [103], the authors enhanced the data collected in [107] by employing a GelSight sensor to refine the samples of some materials with additional geometric information. Their architecture encodes the GelSight images into texture representation vectors, which are then combined with encoded action representations derived from the user's force and velocity. This combined representation is used by an acceleration predictor module to predict the desired DFT (Discrete Fourier Transform).

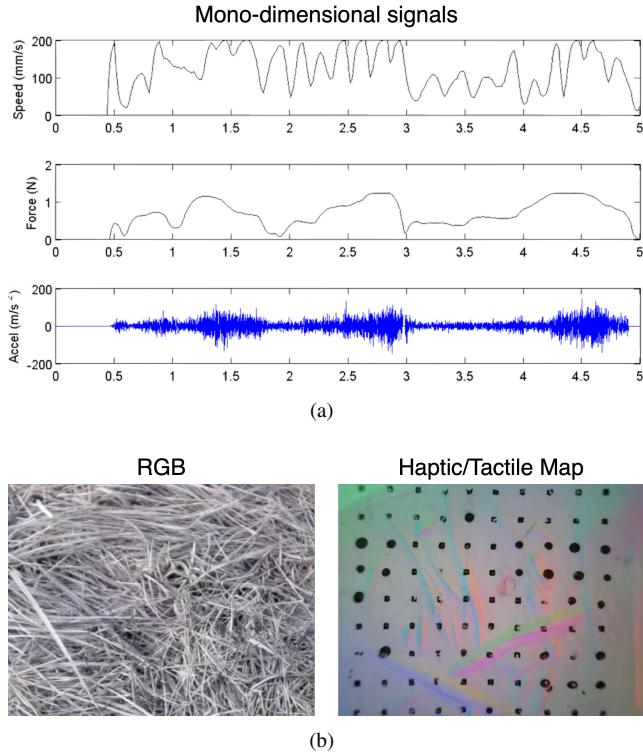


Figure 5: Data-driven approaches require the collection of haptic signals, which can be either (a) mono-dimensional (taken from [81]) or (b) vision-based (taken from [108]). Mono-dimensional signals usually describe a single feature acquired by a sensor, like speed, force, and acceleration. Vision-based signals provide spatially-related 2D high-resolution information of a surface.

## 5. Datasets and Simulators

The field of haptic data acquisition and simulation has seen significant advancements, driven by the need for more realistic and immersive tactile interactions in robotics, virtual reality, and human-computer interaction. Various datasets and simulators have been developed to capture and model the complex sensory experiences associated with touch. These resources provide comprehensive data that combine visual, tactile, auditory, and sometimes linguistic information, facilitating the development and evaluation of algorithms for tasks such as texture classification, material recognition, and robotic manipulation. In this section, we summarize the main datasets available in the literature, describing their properties, data collection methods, features, and potential applications, distinguishing between synthetic and real datasets. The details are summarized in Tab. 3. Additionally, we review recent simulators that enhance the realism of tactile data, bridging the gap between simulated and real-world haptic interactions.

### 5.1. Datasets

In [107] (HaTT), acceleration data was collected to model vibrations from textures, position, and pressure force, along with RGB images for 100 haptic textures and friction models. The dataset, including ten-second recordings of force, speed, and high-frequency acceleration from a handheld tool, uses a lookup table where AR models are labeled by force and speed.

The paper [109] introduces a dataset (LMT v1) of acceleration data for 43 textures to aid in developing algorithms for texture recognition through tool-mediated interactions. It has been improved in [110] (LMT v2), introducing a novel dataset and methodology for classifying surfaces based on acceleration signals from human freehand movements and in [111] (LMT v3), where high-resolution images, haptic feedback, and sound recordings of various surface materials were collected. The dataset includes diverse materials like wood, metal, fabric, and plastic, with annotations for supervised learning, capturing visual features (color histograms, texture descriptors, edge detection), tactile features (vibration frequency, amplitude, force measurements), and auditory features (sound frequency, intensity, temporal patterns) to support feature fusion techniques for improved classification accuracy.

In [112] (FoS), a 7-DoF Sawyer robotic arm with a Weiss WSG-50 gripper and GelSight tactile sensors, along with a Microsoft Kinect 2 for depth data, was used to collect over 9000 grasping trials. Visuo-tactile deep neural networks trained on this data showed that incorporating tactile information significantly improved grasp prediction accuracy. The work in [113] (Fabrics) presented a dataset and methodology to bridge visual and tactile perceptions of materials using 118 fabric samples (Fabrics). Convolutional Neural Networks (CNNs) trained on this data generated embedding vectors to link visual inputs with tactile sensations, demonstrating that models incorporating both visual and tactile data outperformed those using visual data alone. The dataset from [114] (UHL) aims to create a universal library of haptic textures assignable to surfaces based on their visual features. This library enables automatic haptic feedback based on visual appearance, enhancing realism and immersion in virtual environments, such as virtual reality and remote manipulation. [115] (Texplorer) includes diverse surface materials, captured using high-resolution imaging and tactile sensors. Designed for developing and evaluating retrieval algorithms, it supports multi-modal research and is ideal for applications in digital material libraries, e-commerce, and virtual reality. In [120] (Texplorer2), a Tactile Computer Mouse (TCM) is introduced to simulate tactile sensations like hardness, friction, warmth, and roughness, enhancing virtual material perception. Equipped with actuators, the TCM provides calibrated tactile feedback matched to visual textures across 184 different surfaces, aiming to deliver realistic material interaction experiences.

The dataset in [116] (ViTac) comprises various cloth textures captured using both visual and tactile sensing methods. This multi-modal dataset supports research into integrating visual and tactile features for enhanced cloth texture recognition, aiding in the development and evaluation of algorithms for improved material identification and classification. In [117] (MTaF), the dataset combines high-resolution visual data capturing object appearance and shape with tactile sensor readings detailing contact forces and surface texture.

[118] (Haptex) aims to enhance virtual texture realism through tactile displays, featuring data from 120 fabric samples with diverse textures. In [119] (VisGel), RGB and GelSight data from 118 surfaces are collected to explore integrating vi-

Year	Name	Ref.	Data annotations								Dataset size			
			RGB image	Haptic map	Depth/ 3D data	Acceleration/ vibration	Position	Pressure Force	Friction force/ coefficient	Sound	Real/Synth	Objects	Touches	Link
2014	HaTT	[107]	✓			✓	✓	✓			Real	100	-	<a href="#">Link</a>
	LMT v1	[109]				✓					Real	43	-	<a href="#">Link</a>
2015	LMT v2	[110]				✓				✓	Real	69	-	<a href="#">Link</a>
2016	LMT v3	[111]	✓			✓			✓	✓	Real	69	-	<a href="#">Link</a>
2017	FoS	[112]	✓	✓							Real	106	9.2k	<a href="#">Link</a>
	Fabrics	[113]	✓	✓		✓					Real	118	-	<a href="#">Link</a>
	UHL	[114]	✓						✓	✓	Real	84	-	<a href="#">Link</a>
	Texplorer	[115]	✓				✓		✓	✓	Real	108	-	<a href="#">Link</a>
2018	ViTac	[116]	✓	✓							Real	100	3.0k	<a href="#">Link</a>
	MTaF	[117]	✓	✓							Real	65	6.45k	<a href="#">Link</a>
2019	Haptex	[118]					✓	✓	✓		Real	120	-	<a href="#">Link</a>
	VisGel	[119]	✓	✓							Real	195	12.0k	<a href="#">Link</a>
	Texplorer2	[120]	✓			✓			✓	✓	Real	184	-	<a href="#">Link</a>
2020	HaTT + Gelsight	[103]	✓	✓		✓	✓	✓	✓		Real	100	-	<a href="#">Link</a>
2021	ObjectFolder	[121]	✓		✓			✓			Synth	100	-	<a href="#">Link</a>
2022	Touch and Go	[108]	✓	✓							Real	3971	13.9k	<a href="#">Link</a>
	ObjectFolder2.0	[122]	✓		✓			✓			Synth	1000	-	<a href="#">Link</a>
	PoseIt	[123]	✓	✓					✓		Real	26	1.84k	<a href="#">Link</a>
	SSVTP	[124]	✓	✓							Real	10	4.5k	<a href="#">Link</a>
2023	ObjectFolder Real	[125]	✓	✓	✓						Real	100	-	<a href="#">Link</a>
	MTTD	[126]				✓					Real	12	3.6k	<a href="#">Link</a>
	YCB-Slide	[127]	✓	✓	✓						Real/Synth	10	-	<a href="#">Link</a>
2024	MPI-10	[128]				✓		✓			Real	10	-	<a href="#">Link</a>
	SSVTP + HCT*	[129]	✓	✓							Real	-	44.0k	<a href="#">Link</a>
	SENS3 <sup>T</sup>	[130]	✓				✓	✓	✓	✓	Real	50	-	<a href="#">Link</a>
	Touch100k*	[131]	✓	✓							Real	-	100k	<a href="#">Link</a>
	TARF	[132]	✓	✓	✓						Real	-	19.3k	<a href="#">Link</a>

Table 3: This table provides a comprehensive overview of datasets used in haptic sensing and interaction research, highlighting their diverse properties. Each dataset is detailed with its name, key features, data collection methods, and data size. The symbols (\*) and (<sup>T</sup>) indicate the datasets providing as data annotations text descriptions and temperature respectively.

sual and tactile sensory information for enhanced physical interactions. The study focuses on training a cross-modal prediction model using conditional Generative Adversarial Networks (GANs), enabling robots to predict tactile sensations from visual inputs and vice versa.

[103] (HaTT + GelSight) provides models for 100 materials using a sensorized pen measuring 6 degrees of freedom (DoF) in force/torque, acceleration, and positional tracking during 10-second circular motions. The dataset supports research in texture rendering and hardness estimation for enhancing haptic feedback realism in virtual reality and teleoperation applications.

The dataset presented in [121] (ObjectFolder) is a dataset of multisensory neural and real objects, containing 100 virtualized objects with visual, auditory, and tactile representations. The successor of ObjectFolder is [122] (ObjectFolder2.0). The purpose of the dataset is to facilitate sim-to-real transfer in computer vision and robotics. They synthesized 1000 common household objects, with implicit neural representations. The rendering time has been improved, it's orders of magnitude faster, with improved rendering quality for visual, acoustic, and tactile modalities. Models trained on virtual objects in ObjectFolder2.0 successfully transfer to real-world counterparts. A comprehensive suite of tasks designed for multisensory object-centric learning is provided in [125] (ObjectFolder Real). The benchmark focuses on object recognition, reconstruction, and manipulation using sight, sound, and touch. It emphasizes the importance of multisensory perception in computer vision and robotics. It contains real data. Measurements cover 3D meshes,

videos, impact sounds, and tactile readings. Systematic benchmarking reveals the roles of vision, audio, and touch in different object-centric learning tasks. Multisensory perception significantly impacts performance.

The work by Yang et al. [108] (Touch and Go) has introduced a multimodal visuo-tactile learning dataset focusing on associating touch with sight during physical interactions with objects. Human data collectors used tactile sensors to probe objects in natural environments, simultaneously recording egocentric video. In [123] the authors demonstrated how holding poses impact grasp stability. The purpose of the dataset (PoseIt) is investigating grasp stability considering different holding poses. An LSTM classifier trained on PoseIt achieves 0.85 accuracy on this task. Multi-modal models outperform using vision or tactile data alone. Moreover, it generalizes to unseen objects and poses.

The dataset (SSVTP) [124] enables robots to learn multi-task visuo-tactile representations in a self-supervised manner. It includes spatially-aligned visual and tactile image pairs, allowing the robot to perform tasks like feature classification, contact localization, anomaly detection, visual query-based feature search, and edge following along cloth edges.

In [128] (MPI-10) they capture haptic-auditory recordings during human exploration of 10 surfaces using three steel tools, collecting haptic (tool and finger accelerations), auditory (contact sounds), and force/torque applied to the surface. In [129] (SSVTP + HCT), the SSVTP dataset is expanded by incorporating HCT (Human Contact Tasks). Thus, tactile and visual data are aligned with linguistic labels. It includes training, testing,

and fine-tuning splits to facilitate the development of models that interpret tactile sensations joined with visual inputs and descriptive language. Applications include open-vocabulary classification and text generation based on multimodal cues. The dataset in [130] (SENS3) is an extensive open-access repository of multisensory data acquired from fifty surfaces during finger-surface interactions.

The work in [126], with (MMTD), focuses on dynamic tactile data for texture characterization. It was generated using a tactile-enabled finger equipped with a multi-modal tactile sensing module. This dataset incorporates data from pressure, gravity, angular rate, and magnetic field sensors, designed for texture classification. In the paper [127] they introduce (YCB-Slide). MidasTouch is a tactile perception system designed for online global localization of a vision-based touch sensor as it slides across an object’s surface. The YCB-Slide dataset, which includes real-world and simulated forceful sliding interactions between a vision-based tactile sensor and standard YCB objects, accompanies the MidasTouch framework.

The dataset in [131] (Touch100k) is a large-scale touch-language-vision dataset designed for touch-centric multimodal representation. It features 100000 paired samples with tactile sensation descriptions at multiple granularities. These descriptions include sentence-level natural expressions with rich semantics, capturing contextual and dynamic relationships, as well as phrase-level descriptions highlighting key features of tactile sensations.

In [132] (TaRF), the authors present a dataset that introduces a novel scene representation that unifies vision and touch in a shared 3D space. It estimate visual and tactile signals for specific 3D positions within a scene. Leveraging insights from multi-view geometry, they register touch signals to visual scenes and train a conditional diffusion model. This model generates tactile signals corresponding to RGB-D images rendered from neural radiance fields.

## 5.2. Simulators

Generating synthetic data to avoid the cumbersome data collection and annotation process has been commonly used in multiple domains [133]. Similarly, simulation has been used in haptic domain to collect synthetic data. In this section we report the available simulators in the state of the art.

The simulator described in [134] (Tacto) is designed to emulate the tactile sensations perceived by robotic systems. It offers a high degree of realism and versatility by using high-resolution vision-based tactile sensors to simulate the deformation and contact of objects with the sensor surface. Tacto provides a user-friendly interface and supports integration with popular robotics frameworks like PyBullet and ROS, facilitating seamless deployment in various robotic applications. The simulator is open-source.

In [135] (Taxim) they provide a sophisticated method to emulate the tactile responses of GelSight sensors by leveraging an example-based approach. Taxim generates realistic tactile images by mapping the deformation patterns of the sensor’s gel surface to pre-recorded examples of tactile data, allowing for

accurate reproduction of sensor outputs under various contact conditions. This method improves the fidelity of the simulation, making it particularly useful for machine learning and robotic manipulation tasks.

[136] (Tacchi) offers a streamlined and efficient approach to simulating the deformation of elastomer-based tactile sensors. Tacchi is designed to be easily integrated (“pluggable”) with existing systems and achieves low computational overhead by focusing on essential deformation characteristics, thereby maintaining real-time performance. It accurately replicates the tactile feedback by simulating how the elastomer deforms upon contact with objects, ensuring that the output closely mirrors actual sensor readings. This makes Tacchi a practical tool for researchers and developers without the burden of high computational demands.

[137] (Diffactile) provides a sophisticated simulation environment designed to model the intricate interactions between tactile sensors and objects during robotic manipulation tasks. This simulator stands out by incorporating a differentiable physics engine, which allows for the computation of gradients with respect to the tactile interactions. This feature enables the optimization of sensor designs and control policies through gradient-based methods. Diffactile excels in simulating complex, contact-rich scenarios with high accuracy.

The application described in [138] (Tacipc) offers an advanced simulation framework tailored for optical tactile sensors, utilizing the Finite Element Method (FEM) to model elastomer deformation. TacIPC distinguishes itself by ensuring simulations are free from geometric intersections and inversions, which enhances the stability and realism of the deformation modeling. This high-fidelity approach allows for accurate replication of how elastomer materials behave under contact, providing valuable insights for sensor design and application.

In the paper [139] (Fots) authors provide a rapid and realistic simulation environment for training tactile-motor skills in robots. By focusing on high-speed rendering of optical tactile feedback, Fots ensures that the simulated interactions are highly accurate and closely mimic real-world tactile responses. This simulator is specifically designed to facilitate the sim2real transfer, allowing robotic systems to effectively apply the skills learned in simulation to physical tasks.

The paper [140] (Tactile) offers a detailed simulation framework tailored for GelStereo sensors, which combine visual and tactile sensing capabilities. This simulator accurately replicates the tactile imprints left on the sensor’s gel surface when objects make contact, capturing the intricate details of these interactions. By focusing on the visuotactile feedback, the simulation enables researchers to study and optimize the sensor’s performance in a variety of tactile sensing tasks.

Finally [141] presents a comprehensive framework for simulating, training, and deploying vision-based tactile sensors in extensive applications. This framework supports large-scale simulations that accurately model the tactile feedback from vision-based sensors, enabling the development of robust tactile sensing algorithms. By integrating learning methodologies, the framework facilitates the training of machine learning models to interpret tactile data effectively. The scalable nature of this

simulation platform allows for extensive testing and optimization, making it ideal for broad deployment in various robotic and industrial contexts where tactile sensing is critical.

## 6. Research tasks

In this chapter we provide an overview of the main tasks currently tackled by researchers in the haptic signal processing domain. Starting from the datasets listed in Section 5, in Tab. 4 we provide a thorough classification of the available works for each task. While some of them, like classification and compression, are akin to other domains and straightforward to understand, other tasks are very specific to haptics, like contact localisation and grasp prediction. For a comprehensive overview, in the next paragraphs we delve into each task, highlighting their peculiarities as well as the common metrics used for the evaluation.

**Classification.** is one of the tasks mainly performed when dealing with haptic data. As the collected datasets usually provide ground truths in different domains (e.g. haptic maps, RGB images, audio produced by the object upon contact) for the same sample, classification can be performed according to different strategies: *single mode*, *multi-modal*, and *cross-modal*. Works adopting the *single mode* strategy aim to classify materials/objects by training and testing in the same domain. For approaches employing only haptic data, this task is usually performed on tactile maps [116, 120, 168, 175, 176, 121, 172, 169, 125, 178, 183, 163], since dealing with images enables the usage of established approaches, such as deep neural networks (DNNs). Differently, other works, such as [109, 142] proposed a material classification method that exploits acceleration data, achieving state-of-the-art results and showing that acceleration data establish a unique fingerprint for each material. *Multi-modal* approaches employ multiple data sources to classify materials and objects. These works rely mainly on visual and haptic data [150, 176, 152, 151, 124, 149, 108, 153, 182, 143, 178, 132, 129, 184, 163], since their combination allows for the exploitation of both mono-dimensional and multi-dimensional data. However, also audio cues are widely used jointly with haptic data [154, 111, 115, 171, 120, 121, 172, 125, 170, 185, 183] enabling approaches that leverage the combination of all these data types. *Cross-modal* classification is a task that leverages knowledge acquired on one data domain to classify data from another one. This allows, for example, to exploit data domains that are more easily accessible, such as images and audio domains, and apply knowledge on an unexplored domain, like haptic. Cross-modal classification has been performed with audio data in [155], and using visual data in [116, 176, 155, 129]. Furthermore, a subset task of the cross-modal classification application is *cross-modal retrieval* [156, 157, 121, 125, 180]. Cross-modal retrieval sits between classification and generation applications. The core idea consists of retrieving the data sample in one domain based on a query from another domain by finding the best correspondence within the dataset used to train the system.

To quantitatively evaluate the results obtained and compare them with other works, various metrics have been employed.

Other than the standard classification metrics, such as *Classification Accuracy*, *Classification Precision*, *Recall*, *F1-score*, mean Average Precision (mAP) and rank@, an haptic-specific metric *L2 distance* aims at measuring the distance between the feature vectors extracted by two instances from different data sources.

**Cross-modal generation.** refers to the creation of content in one sensory modality based on the information gathered from another one. This process involves translating the haptic information of a surface across different types of media. Recently, this task has gained popularity in many real-world applications, since it allows to collect data in one modality and to transfer the information to different ones. From the haptic perspective, cross-modal generation can be applied to generate vibro-tactile acceleration signals from other types of data, such as texture images [159, 158, 179, 144, 162, 160]. Another approach involves mapping acoustic and acceleration signals, exploiting their similarities [125, 161]. Furthermore, visual and acoustic domains can also be combined [173]. Another data type widely employed in literature for the cross-modal generation task is the haptic map, which enables the usage of computer vision algorithms to map RGB images into haptic data [177, 125, 181, 132, 162, 163]. The most common metrics include *Structural Similarity Index Measure* (SSIM), *PSNR*, *RMSE*, *MAE*, *Fréchet inception distance* (*FID*), and *L2-distance*, which assess the quality of the generated samples and the difference from the ground truth. In addition to these, many authors employed non-standard metrics. In [181], the authors introduce the *Contrastive Visuo-Tactile Pre-Training* (*CVTP*) metric, a cosine similarity that evaluates the distance between visual and tactile embeddings, obtained within a contrastive learning framework. *Material Classification Consistency* (*MCC*) [181, 163] is another evaluation strategy that is gaining popularity. This method consists of using a material classifier trained on ground truth samples to classify the cross-generated samples and check whether they correspond to the proper class. In [125], the *STFT-distance* and *Envelope Distance* metrics are introduced to evaluate the distance between the ground truth and the spectrograms or signals predicted. Furthermore, in [125], the authors assess the quality of the samples generated within the audio-tactile generation task by taking into account also the perception similarity. To do this, they employed the *CDPAM* metric, defined in [186].

**Stylization.** is a recent application that, given a desired haptic map, aims at "stylizing" the RGB image of an object, modifying its appearance to reflect the tactile characteristics accurately. This application has been applied mainly to the Touch And Go dataset using methods based on contrastive learning [108] or diffusion models [181, 163]. Despite this task emerging from the use of RGB and Haptic Maps data, the HaTT dataset has also been previously employed. In [145], the authors exploited a physical surface as a constraint, modifying the accelerations emitted by the haptic device to render a different surface according to the model collected in HaTT.

Dataset	Classification			Cross-modal generation	Stylization	Grasp prediction	Contact localization	Perceptual similarity	Video generation/prediction
	Single mode	Multi-modal	Cross-modal						
HaTT	[142]	[143]		[144]	[145]			[146, 147] [148]	
LMT	[109]	[111][149] [150, 151, 152] [153, 154]	[155, 156] [157]	[158, 159] [160, 161] [162]					
FoS						[112][163] [164, 165] [166, 167]			
Fabrics	[168, 169]	[149, 152]							
Texplorer	[120]	[115, 120] [170, 171] [143, 172]	[157]	[173]				[174]	
ViTac	[116] [175, 176]	[176]	[116][176]	[177]					
MTaF					[117][178]				
Haptex				[179]					
VisGel			[180]	[181]				[180]	
ObjectFolder		[125][182]	[121][125] [180, 163]	[125]		[125][163]	[125]		[180]
Touch and Go	[178, 183] [163]	[108][178] [183, 184]	[163, 180]	[181, 163] [162]	[108] [163, 181]	[108][163]			[108][180]
PoseIt					[123]				
SSVTP	[163]	[124]	[163]		[163]	[124]			
MPI-10		[185]							
SSVTP + HCT		[129]	[129]						
TaRF		[132]		[132]		[132]			

Table 4: This table illustrates the tasks performed on the most used datasets presented in Table 3, highlighting in **bold** the baselines methods.

To evaluate the results, two metrics mainly employed. For the Touch and Go dataset, the authors used a material classifier pre-trained on ground truth samples. This approach, previously presented in cross-modal generation tasks, checks if images modified to reflect a specific haptic style are classified correctly as belonging to that style. For the HaTT dataset, the authors verify whether the acceleration generated by the haptic device matched the ground truth model.

**Grasp prediction.** is a robotic application that focuses on determining the best way for a robotic hand to manipulate an object, performing tasks like moving or positioning it. While not being strictly modeling-related, it requires the understanding of the stiffness and friction properties of an object, thus modeling of haptic properties plays a fundamental role in this task. These algorithms rely mainly on vision-based sensors to gather haptic information on the part of the object occluded by robotic fingers

that cannot be seen by external cameras. To perform grasp prediction various strategies have been developed. In [112, 125], the authors rely on the difference between the initial and final grasp position to train a DNN to predict the success of the grasp. Furthermore, in [117], the authors introduced a conditioning factor that enables the DNN to adjust the grasping action dynamically if the first attempt fails. Starting from the same idea of training the classifier to predict if the grasp will be successful or not, in [164, 165] the authors designed a self-attention mechanism, while in [108, 166, 167, 178, 163] contrastive learning frameworks have been adopted. All these works employ only tactile maps and RGB images to perform grasp prediction. The work in [123] represents the only exception, where, to understand if grasping will be successful in challenging conditions, such as while shaking the object, the authors also employed mono-dimensional signals collected using force/torque sensors to train the classifier. Furthermore, grasp prediction is of particular interest for the works that aim at bridging the gap between simulation and reality, as presented in [125, 167].

For the evaluation, grasp prediction presents well-established metrics: *Accuracy*, *Precision*, *Recall*, and *F-score*.

**Contact localization.** is another key application in robotics. This application aims at identifying the exact point on an object where a sample has been taken. Initially, the primary goal of this task was to determine where an object was touched to teach robots how to grasp object correctly [124, 127]. Recently, however, research has moved towards identifying the location on an RGB image or a 3D model where other types of data, such as acoustic signals [125] or haptic signals [132], were generated.

To assess the performances of a contact localization algorithm, the metrics commonly used are *mAP*, to verify whether the system precision and recall are well balanced, *RMSE*, to assess the accuracy of both the position and the rotation measurements by comparing them to the ground truth, and *average normalized distance* (ND), to measure the distance between the predicted and the ground truth points.

**Perceptual similarity.** is an application aiming at assessing how human touch feels different textures or surfaces. The goal of this task is to identify models capable of predicting human perception based on physical interactions with surfaces. In [146] and [147], the authors propose an experiment using the HaTT dataset where users has to feel different textures by exploring randomly real surfaces made of different materials. The goal of these works is to identify the similarities that different textures present to cluster them according to the feature extracted. In [174], the authors try to cluster the textures, according to the feature perceived by user, into a feature space. To do this they employ respectively a K-means strategy and a DNN method. In [148], the authors exploit GelSight images to capture textures information and, by modeling the user interaction with them, they create virtual textures that close up with their real counterpart.

This task is evaluated by performing user study evaluation and thus its significance strongly depends on the number of participants.

**Video generation and prediction.** is a very specific application that relies on video streams from vision-based sensors like GelSight or standard RGB cameras. The goal of this task is to retrieve missing frames within videos collected from a specific data source. In [108], two approaches are explored: single-mode, which relies only on previous frames to generate missing ones, and multimodal, which leverages previous frames coming from different data sources. Differently, in [180], the authors rely on a cross-generation framework, using GelSight-captured video information to generate missing frames in RGB videos and vice-versa.

To evaluate the reconstruction capabilities of these methods, the authors used various metrics, such as *L1* and *L2* distances, and *SSIM*. Furthermore, in [108], the authors employed also the *Learned Perceptual Image Patch Similarity (LPIPS)* metric. LPIPS, firstly introduced in [187], aims at assessing the performance of a model by considering also the human perception.

This chapter has explored the different application domains in which haptic datasets have been adopted. The tasks encompassed deal with a multitude of research areas, demonstrating how the research in rapidly evolving in a multidimensional and multimodal fashion, requiring competences in signal processing and representation, machine learning, human perception, to name a few.

## 7. Conclusions

Digitalizing real objects and surfaces to enable virtual interaction requires following a rather strict pipeline, which stems from the definition of the sensing device and ends with the user perception study.

In this paper we have decided to delve into one of the mostly unexplored areas in haptics, namely the representation and modeling of surfaces. To do so we have identified the most salient properties of objects, providing an exhaustive overview of the existing datasets and the related applications.

Our analysis has shown how the non-standardisation of haptic data modeling provides an open research field, where a common standard has yet to emerge. Consequently, the literature is still rapidly evolving opening up to new research challenges in the domain of haptics.

## CRediT authorship contribution statement

**Antonio Luigi Stefani:** Conceptualization, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Niccolò Bisagno:** Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Andrea Rosani:** Investigation, Writing – original draft, Writing – review & editing. **Nicola Conci:** Conceptualization, Project administration, Supervision, Writing – review & editing. **Francesco De Natale:** Funding acquisition, Supervision, Writing – review & editing.

## Data availability

No data have been used to this project.

## Acknowledgments

We acknowledge the support of the MUR PNRR project iNEST-Interconnected Nord-Est Innovation Ecosystem (ECS00000043) funded by the European Union under NextGenerationEU. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or The European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

## References

- [1] N. Xi, J. Chen, F. Gama, M. Riar, J. Hamari, The challenges of entering the metaverse: An experiment on the effect of extended reality on workload, *Information Systems Frontiers* 25 (2) (2023) 659–680.
- [2] S. Alizadehsalehi, A. Hadavi, J. C. Huang, From bim to extended reality in aec industry, *Automation in construction* 116 (2020) 103254.
- [3] E. A. Keshner, Virtual reality and physical rehabilitation: a new toy or a new research and rehabilitation tool? (2004).
- [4] A. Jung herr, D. B. Schlarb, The extended reach of game engine companies: How companies like epic games and unity technologies provide platforms for extended reality applications and the metaverse, *Social Media+ Society* 8 (2) (2022) 20563051221107641.
- [5] P. Kral, K. Janoskova, A.-M. Potcovaru, Digital consumer engagement on blockchain-based metaverse platforms: extended reality technologies, spatial analytics, and immersive multisensory virtual spaces, *Linguistic and Philosophical Investigations* 21 (2022) 252–267.
- [6] S. Biswas, Y. Visell, Haptic perception, mechanics, and material technologies for virtual reality, *Advanced Functional Materials* 31 (39) (2021) 2008186.
- [7] H. Culbertson, S. B. Schorr, A. M. Okamura, Haptics: The present and future of artificial touch sensation, *Annual Review of Control, Robotics, and Autonomous Systems* 1 (2018) 385–409.
- [8] M. Fleury, G. Lioi, C. Barillot, A. Lécuyer, A survey on the use of haptic feedback for brain-computer interfaces and neurofeedback, *Frontiers in Neuroscience* 14 (2020) 524967.
- [9] S. D. Laycock, A. Day, A survey of haptic rendering techniques, in: *Computer graphics forum*, Vol. 26, Wiley Online Library, 2007, pp. 50–65.
- [10] K. Salisbury, F. Conti, F. Barbagli, Haptic rendering: introductory concepts, *IEEE computer graphics and applications* 24 (2) (2004) 24–32.
- [11] B. Varalakshmi, J. Thriveni, K. Venugopal, L. Patnaik, Haptics: state of the art survey, *International Journal of Computer Science Issues (IJCSI)* 9 (5) (2012) 234.
- [12] W. Chen, H. Khamis, I. Birznieks, N. F. Lepora, S. J. Redmond, Tactile sensors for friction estimation and incipient slip detection—toward dexterous robotic manipulation: A review, *IEEE Sensors Journal* 18 (22) (2018) 9049–9064.
- [13] Z. Kappassov, J.-A. Corrales, V. Perdereau, [Tactile sensing in dexterous robot hands — review](#), *Robotics and Autonomous Systems* 74 (2015) 195–220. doi:<https://doi.org/10.1016/j.robot.2015.07.015>. URL <https://www.sciencedirect.com/science/article/pii/S0921889015001621>
- [14] A. Yamaguchi, C. G. Atkeson, Recent progress in tactile sensing and sensors for robotic manipulation: can we turn tactile sensing into vision?, *Advanced Robotics* 33 (14) (2019) 661–673.
- [15] S. Zhang, Z. Chen, Y. Gao, W. Wan, J. Shan, H. Xue, F. Sun, Y. Yang, B. Fang, Hardware technology of vision-based tactile sensor: A review, *IEEE Sensors Journal* 22 (22) (2022) 21410–21427.
- [16] N. Promwongsa, A. Ebrahimzadeh, D. Naboulsi, S. Kianpisheh, F. Belqasmi, R. Glitho, N. Crespi, O. Alfandi, A comprehensive survey of the tactile internet: State-of-the-art and research directions, *IEEE Communications Surveys & Tutorials* 23 (1) (2020) 472–523.
- [17] G. Minopoulos, G. Kokkonis, K. E. Psannis, Y. Ishibashi, A survey on haptic data over 5g networks (2019).
- [18] G. Kokkonis, K. Psannis, M. Roumeliotis, S. Kontogiannis, A survey of transport protocols for haptic applications, in: *2012 16th Panhellenic Conference on Informatics*, IEEE, 2012, pp. 192–197.
- [19] M. Awais, F. Ullah Khan, M. Zafar, M. Mudassar, M. Zaigham Zaheer, K. Mehmood Cheema, M. Kamran, W.-S. Jung, Towards enabling haptic communications over 6g: Issues and challenges, *Electronics* 12 (13) (2023) 2955.
- [20] K. Antonakoglou, X. Xu, E. Steinbach, T. Mahmoodi, M. Dohler, Toward haptic communications over the 5g tactile internet, *IEEE Communications Surveys & Tutorials* 20 (4) (2018) 3034–3059.
- [21] A. Adilkhanov, M. Rubagotti, Z. Kappassov, Haptic devices: Wearability-based taxonomy and literature review, *IEEE Access* 10 (2022) 91923–91947.
- [22] C. Basdogan, F. Giraud, V. Levesque, S. Choi, A review of surface haptics: Enabling tactile effects on touch surfaces, *IEEE transactions on haptics* 13 (3) (2020) 450–470.
- [23] S. J. Breitschaft, S. Heijboer, D. Shor, E. Tempelman, P. Vink, C.-C. Carbon, The haptic fidelity framework: A qualitative overview and categorization of cutaneous-based haptic technologies through fidelity, *IEEE transactions on haptics* 15 (2) (2022) 232–245.
- [24] S. Choi, K. J. Kuchenbecker, Vibrotactile display: Perception, technology, and applications, *Proceedings of the IEEE* 101 (9) (2012) 2093–2104.
- [25] A. Costes, F. Danieau, F. Argelaguet, P. Guillotel, A. Lécuyer, Towards haptic images: A survey on touchscreen-based surface haptics, *IEEE Transactions on Haptics* 13 (3) (2020) 530–541.
- [26] V. R. Mercado, M. Marchal, A. Lécuyer, “haptics on-demand”: A survey on encountered-type haptic displays, *IEEE Transactions on Haptics* 14 (3) (2021) 449–464.
- [27] H. Z. Tan, S. Choi, F. W. Lau, F. Abnousi, Methodology for maximizing information transmission of haptic devices: A survey, *Proceedings of the IEEE* 108 (6) (2020) 945–965.
- [28] D. Wang, K. Ohnishi, W. Xu, Multimodal haptic display for virtual reality: A survey, *IEEE Transactions on Industrial Electronics* 67 (1) (2019) 610–623.
- [29] C. Wee, K. M. Yap, W. N. Lim, Haptic interfaces for virtual reality: Challenges and research directions, *IEEE access* 9 (2021) 112145–112162.
- [30] M. Kurzweg, Y. Weiss, M. O. Ernst, A. Schmidt, K. Wolf, A survey on haptic feedback through sensory illusions in interactive systems, *ACM Computing Surveys* (2024).
- [31] A. Lécuyer, Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback, *Presence: Teleoperators and Virtual Environments* 18 (1) (2009) 39–53.
- [32] Y. Ujitoko, Y. Ban, Survey of pseudo-haptics: Haptic feedback design and application proposals, *IEEE Transactions on Haptics* 14 (4) (2021) 699–711.
- [33] T. R. Coles, D. Meglan, N. W. John, The role of haptics in medical training simulators: A survey of the state of the art, *IEEE Transactions on haptics* 4 (1) (2010) 51–66.
- [34] D. Escobar-Castillejos, J. Noguez, L. Neri, A. Magana, B. Benes, A review of simulators with haptic devices for medical training, *Journal of medical systems* 40 (2016) 1–22.
- [35] Y. Gaffary, A. Lécuyer, The use of haptic and tactile information in the car to improve driving safety: A review of current technologies, *Frontiers in ICT* 5 (2018) 5.
- [36] F. G. Hamza-Lup, K. Bergeron, D. Newton, Haptic systems in user interfaces: state of the art survey, in: *Proceedings of the 2019 ACM Southeast Conference*, 2019, pp. 141–148.
- [37] E. Imran, N. Adanir, Z. Khurshid, Significance of haptic and virtual reality simulation (vrs) in the dental education: a review of literature, *Applied Sciences* 11 (21) (2021) 10196.
- [38] S. Luo, J. Jimbo, R. Dahiya, H. Liu, Robotic tactile perception of object properties: A review, *Mechatronics* 48 (2017) 54–67.
- [39] S. M. Petermeijer, D. A. Abbink, M. Mulder, J. C. De Winter, The effect

- of haptic support systems on driver performance: A literature survey, *IEEE transactions on haptics* 8 (4) (2015) 467–479.
- [40] C. Saint-Louis, A. Hamam, Survey of haptic technology and entertainment applications, in: *SoutheastCon 2021*, IEEE, 2021, pp. 01–07.
- [41] S. Süssstrunk, R. Buckley, S. Swen, Standard rgb color spaces, in: *Color and imaging conference*, Vol. 7, Society of Imaging Science and Technology, 1999, pp. 127–134.
- [42] N. H. Fletcher, T. D. Rossing, *The physics of musical instruments*, Springer Science & Business Media, 2012.
- [43] C. Basdogan, M. A. Srinivasan, Haptic rendering in virtual environments, in: *Handbook of virtual environments*, CRC Press, 2002, pp. 157–174.
- [44] R. Hover, M. Harders, G. Székely, Data-driven haptic rendering of visco-elastic effects, in: *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, IEEE, 2008, pp. 201–208.
- [45] 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* 107 (2) (2018) 447–470.
- [46] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, H. J. Ritter, A review of tactile information: Perception and action through touch, *IEEE Transactions on Robotics* 36 (6) (2020) 1619–1634.
- [47] D. Degraen, M. Piovarči, B. Bickel, A. Krüger, Capturing tactile properties of real surfaces for haptic reproduction, in: *The 34th Annual ACM Symposium on User Interface Software and Technology*, 2021, pp. 954–971.
- [48] N. A. Ibraheem, M. M. Hasan, R. Z. Khan, P. K. Mishra, Understanding color models: a review, *ARPJ Journal of science and technology* 2 (3) (2012) 265–275.
- [49] M. Müller, *Fundamentals of music processing: Audio, analysis, algorithms, applications*, Vol. 5, Springer, 2015.
- [50] X. Chen, F. Shao, C. Barnes, T. Childs, B. Henson, Exploring relationships between touch perception and surface physical properties, *International Journal of Design* 3 (2) (2009) 67–76.
- [51] W. Hassan, A. Abdulali, S. Jeon, Authoring new haptic textures based on interpolation of real textures in affective space, *IEEE transactions on industrial electronics* 67 (1) (2019) 667–676.
- [52] W. Hassan, J. B. Joolie, S. Jeon, Establishing haptic texture attribute space and predicting haptic attributes from image features using 1d-cnn, *Scientific Reports* 13 (1) (2023) 11684.
- [53] A. S. of Mechanical Engineers. Standards Committee B46. Classification, D. of Surface Qualities, A. N. S. Institute, *Surface Texture: Surface Roughness, Waviness and Lay*, American Society of Mechanical Engineers, 2003.
- [54] E. Baumgart, Stiffness—an unknown world of mechanical science, *Injury* 31 (Suppl 2) (2000) B14–23.
- [55] F. P. Bowden, D. Tabor, *The friction and lubrication of solids*, Vol. 1, Oxford university press, 2001.
- [56] C. Standard, et al., Colorimetry-part 4: Cie 1976 l\* a\* b\* colour space, *International Standard* (2007) 2019–06.
- [57] W. S. Stiles, J. M. Burch, Npl colour-matching investigation: final report (1958), *Optica Acta: International Journal of Optics* 6 (1) (1959) 1–26.
- [58] E. Zwicker, H. Fastl, *Psychoacoustics: Facts and models*, Vol. 22, Springer Science & Business Media, 2013.
- [59] H. Fletcher, W. A. Munson, Loudness, its definition, measurement and calculation, *Bell System Technical Journal* 12 (4) (1933) 377–430.
- [60] B. C. Moore, *An introduction to the psychology of hearing*, Brill, 2012.
- [61] W. A. Sethares, *Tuning, timbre, spectrum, scale*, Springer Science & Business Media, 2005.
- [62] M. Hollins, R. Faldowski, S. Rao, F. Young, Perceptual dimensions of tactile surface texture: A multidimensional scaling analysis, *Perception & psychophysics* 54 (1993) 697–705.
- [63] M. Hollins, S. Bensmaïa, K. Karlof, F. Young, Individual differences in perceptual space for tactile textures: Evidence from multidimensional scaling, *Perception & psychophysics* 62 (2000) 1534–1544.
- [64] W. M. B. Tiest, Tactile perception of material properties, *Vision research* 50 (24) (2010) 2775–2782.
- [65] S. Okamoto, H. Nagano, Y. Yamada, Psychophysical dimensions of tactile perception of textures, *IEEE Transactions on Haptics* 6 (1) (2012) 81–93.
- [66] E. Baumgartner, C. B. Wiebel, K. R. Gegenfurtner, Visual and haptic representations of material properties, *Multisensory research* 26 (5) (2013) 429–455.
- [67] A. Metzger, M. Toscani, Unsupervised learning of haptic material properties, *Elife* 11 (2022) e64876.
- [68] M. Sakamoto, J. Watanabe, Exploring tactile perceptual dimensions using materials associated with sensory vocabulary, *Frontiers in Psychology* 8 (2017) 250102.
- [69] J. D. Lieber, S. J. Bensmaia, High-dimensional representation of texture in somatosensory cortex of primates, *Proceedings of the National Academy of Sciences* 116 (8) (2019) 3268–3277.
- [70] S. Lu, M. Zheng, M. C. Fontaine, S. Nikolaidis, H. Culbertson, Preference-driven texture modeling through interactive generation and search, *IEEE transactions on haptics* 15 (3) (2022) 508–520.
- [71] M. I. Awan, W. Hassan, S. Jeon, Predicting perceptual haptic attributes of textured surface from tactile data based on deep cnn-lstm network, in: *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, 2023, pp. 1–9.
- [72] I. B. of Weights, Measures, B. N. Taylor, A. Thompson, *The international system of units (SI)*, US Department of Commerce, Technology Administration, National Institute of . . ., 2001.
- [73] H.-N. Ho, Material recognition based on thermal cues: Mechanisms and applications, *Temperature* 5 (1) (2018) 36–55.
- [74] A. Salazar, On thermal diffusivity, *European journal of physics* 24 (4) (2003) 351.
- [75] R. Hover, G. Kosa, G. Székely, M. Harders, Data-driven haptic rendering—from viscous fluids to visco-elastic solids, *IEEE Transactions on Haptics* 2 (1) (2009) 15–27.
- [76] V. L. Guruswamy, J. Lang, W.-S. Lee, Modelling of haptic vibration textures with infinite-impulse-response filters, in: *2009 IEEE International Workshop on Haptic Audio Visual Environments and Games*, IEEE, 2009, pp. 105–110.
- [77] R. Höver, M. D. Luca, M. Harders, User-based evaluation of data-driven haptic rendering, *ACM Transactions on Applied Perception (TAP)* 8 (1) (2010) 1–23.
- [78] N. Landin, J. M. Romano, W. McMahan, K. J. Kuchenbecker, Dimensional reduction of high-frequency accelerations for haptic rendering, in: *Haptics: Generating and Perceiving Tangible Sensations: International Conference, EuroHaptics 2010, Amsterdam, July 8–10, 2010. Proceedings*, Springer, 2010, pp. 79–86.
- [79] J. Lang, S. Andrews, Measurement-based modeling of contact forces and textures for haptic rendering, *IEEE Transactions on Visualization and Computer Graphics* 17 (3) (2010) 380–391.
- [80] J. M. Romano, K. J. Kuchenbecker, Creating realistic virtual textures from contact acceleration data, *IEEE Transactions on haptics* 5 (2) (2011) 109–119.
- [81] H. Culbertson, J. M. Romano, P. Castillo, M. Mintz, K. J. Kuchenbecker, Refined methods for creating realistic haptic virtual textures from tool-mediated contact acceleration data, in: *2012 IEEE Haptics Symposium (HAPTICS)*, IEEE, 2012, pp. 385–391.
- [82] H. Culbertson, J. Unwin, B. E. Goodman, K. J. Kuchenbecker, Generating haptic texture models from unconstrained tool-surface interactions, in: *2013 World Haptics Conference (WHC)*, IEEE, 2013, pp. 295–300.
- [83] A. Sianov, M. Harders, Data-driven haptics: Addressing inhomogeneities and computational formulation, in: *2013 World Haptics Conference (WHC)*, IEEE, 2013, pp. 301–306.
- [84] H. Culbertson, J. Unwin, K. J. Kuchenbecker, Modeling and rendering realistic textures from unconstrained tool-surface interactions, *IEEE transactions on haptics* 7 (3) (2014) 381–393.
- [85] S. Shin, R. H. Osgouei, K.-D. Kim, S. Choi, Data-driven modeling of isotropic haptic textures using frequency-decomposed neural networks, in: *2015 IEEE World Haptics Conference (WHC)*, IEEE, 2015, pp. 131–138.
- [86] A. Abdulali, S. Jeon, Data-driven modeling of anisotropic haptic textures: Data segmentation and interpolation, in: *Haptics: Perception, Devices, Control, and Applications: 10th International Conference, EuroHaptics 2016, London, UK, July 4–7, 2016, Proceedings, Part II 10*, Springer, 2016, pp. 228–239.
- [87] S. Yim, S. Jeon, S. Choi, Data-driven haptic modeling and rendering of viscoelastic and frictional responses of deformable objects, *IEEE transactions on haptics* 9 (4) (2016) 548–559.
- [88] H. Culbertson, K. J. Kuchenbecker, Importance of matching physical friction, hardness, and texture in creating realistic haptic virtual surfaces,

- IEEE transactions on haptics 10 (1) (2016) 63–74.
- [89] A. Abdulali, S. Jeon, Data-driven rendering of anisotropic haptic textures, in: Haptic Interaction: Science, Engineering and Design 2, Springer, 2018, pp. 401–407.
- [90] A. Abdulali, I. R. Atadjanov, S. Jeon, Visually guided acquisition of contact dynamics and case study in data-driven haptic texture modeling, *IEEE Transactions on Haptics* 13 (3) (2020) 611–627.
- [91] J. B. Joolee, M. A. Uddin, S. Jeon, Deep multi-model fusion network based real object tactile understanding from haptic data, *Applied Intelligence* 52 (14) (2022) 16605–16620.
- [92] H. Cha, A. Bhardwaj, S. Choi, Data-driven haptic modeling and rendering of viscoelastic behavior using fractional derivatives, *IEEE Access* 10 (2022) 130894–130907.
- [93] G. Kumar, S. Prakash, A. Bhardwaj, Catboost for haptic modeling of homogeneous viscoelastic deformable objects, in: 2023 IEEE World Haptics Conference (WHC), IEEE, 2023, pp. 273–278.
- [94] W. Yuan, M. A. Srinivasan, E. H. Adelson, Estimating object hardness with a gelsight touch sensor, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2016, pp. 208–215.
- [95] W. Yuan, S. Dong, E. H. Adelson, Gelsight: High-resolution robot tactile sensors for estimating geometry and force, *Sensors* 17 (12) (2017) 2762.
- [96] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, E. H. Adelson, Shape-independent hardness estimation using deep learning and a gelsight tactile sensor, in: 2017 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2017, pp. 951–958.
- [97] S. Cui, R. Wang, J. Hu, C. Zhang, L. Chen, S. Wang, Self-supervised contact geometry learning by gelstereo visuotactile sensing, *IEEE Transactions on Instrumentation and Measurement* 71 (2021) 1–9.
- [98] Y. Du, G. Zhang, M. Y. Wang, 3d contact point cloud reconstruction from vision-based tactile flow, *IEEE Robotics and Automation Letters* 7 (4) (2022) 12177–12184.
- [99] W. K. Do, M. Kennedy, Densetact: Optical tactile sensor for dense shape reconstruction, in: 2022 International Conference on Robotics and Automation (ICRA), IEEE, 2022, pp. 6188–6194.
- [100] W. K. Do, B. Jurewicz, M. Kennedy, Densetact 2.0: Optical tactile sensor for shape and force reconstruction, in: 2023 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2023, pp. 12549–12555.
- [101] S. Shin, S. Choi, Geometry-based haptic texture modeling and rendering using photometric stereo, in: 2018 IEEE Haptics Symposium (HAPTICS), IEEE, 2018, pp. 262–269.
- [102] S. Shin, S. Choi, Hybrid framework for haptic texture modeling and rendering, *IEEE Access* 8 (2020) 149825–149840.
- [103] N. Heravi, W. Yuan, A. M. Okamura, J. Bohg, Learning an action-conditional model for haptic texture generation, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020, pp. 11088–11095.
- [104] C. Pacchierotti, D. Prattichizzo, K. J. Kuchenbecker, Cutaneous feedback of fingertip deformation and vibration for palpation in robotic surgery, *IEEE Transactions on Biomedical Engineering* 63 (2) (2015) 278–287.
- [105] A. M. Okamura, M. R. Cutkosky, J. T. Dennerlein, Reality-based models for vibration feedback in virtual environments, *IEEE/ASME transactions on mechatronics* 6 (3) (2001) 245–252.
- [106] I. Peterlik, L. Matyska, An algorithm of state-space precomputation allowing non-linear haptic deformation modelling using finite element method, in: Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07), IEEE, 2007, pp. 231–236.
- [107] H. Culbertson, J. J. L. Delgado, K. J. Kuchenbecker, One hundred data-driven haptic texture models and open-source methods for rendering on 3d objects, in: 2014 IEEE haptics symposium (HAPTICS), IEEE, 2014, pp. 319–325.
- [108] F. Yang, C. Ma, J. Zhang, J. Zhu, W. Yuan, A. Owens, Touch and go: Learning from human-collected vision and touch, arXiv preprint arXiv:2211.12498 (2022).
- [109] M. Strese, J.-Y. Lee, C. Schuwerk, Q. Han, H.-G. Kim, E. Steinbach, A haptic texture database for tool-mediated texture recognition and classification, in: 2014 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE) Proceedings, IEEE, 2014, pp. 118–123.
- [110] M. Strese, C. Schuwerk, E. Steinbach, Surface classification using acceleration signals recorded during human freehand movement, in: 2015 IEEE World Haptics Conference (WHC), IEEE, 2015, pp. 214–219.
- [111] M. Strese, C. Schuwerk, A. Iepure, E. Steinbach, Multimodal feature-based surface material classification, *IEEE transactions on haptics* 10 (2) (2016) 226–239.
- [112] R. Calandra, A. Owens, M. Upadhyaya, W. Yuan, J. Lin, E. H. Adelson, S. Levine, The feeling of success: Does touch sensing help predict grasp outcomes?, arXiv preprint arXiv:1710.05512 (2017).
- [113] W. Yuan, S. Wang, S. Dong, E. Adelson, Connecting look and feel: Associating the visual and tactile properties of physical materials, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [114] W. Hassan, A. Abdulali, M. Abdullah, S. C. Ahn, S. Jeon, Towards universal haptic library: Library-based haptic texture assignment using image texture and perceptual space, *IEEE Transactions on Haptics* 11 (2) (2018) 291–303. doi:10.1109/TOH.2017.2782279.
- [115] M. Strese, Y. Boeck, E. Steinbach, Content-based surface material retrieval, in: 2017 IEEE World Haptics Conference (WHC), IEEE, 2017, pp. 352–357.
- [116] S. Luo, W. Yuan, E. Adelson, A. G. Cohn, R. Fuentes, Vitac: Feature sharing between vision and tactile sensing for cloth texture recognition, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 2722–2727.
- [117] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, E. H. Adelson, S. Levine, More than a feeling: Learning to grasp and regrasp using vision and touch, *IEEE Robotics and Automation Letters* 3 (4) (2018) 3300–3307.
- [118] J. Jiao, Y. Zhang, D. Wang, X. Guo, X. Sun, Haptex: A database of fabric textures for surface tactile display, in: 2019 IEEE World Haptics Conference (WHC), IEEE, 2019, pp. 331–336.
- [119] Y. Li, J.-Y. Zhu, R. Tedrake, A. Torralba, Connecting touch and vision via cross-modal prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10609–10618.
- [120] M. Strese, L. Brudermueller, J. Kirsch, E. Steinbach, Haptic material analysis and classification inspired by human exploratory procedures, *IEEE transactions on haptics* 13 (2) (2019) 404–424.
- [121] R. Gao, Y.-Y. Chang, S. Mall, L. Fei-Fei, J. Wu, Objectfolder: A dataset of objects with implicit visual, auditory, and tactile representations, arXiv preprint arXiv:2109.07991 (2021).
- [122] R. Gao, Z. Si, Y.-Y. Chang, S. Clarke, J. Bohg, L. Fei-Fei, W. Yuan, J. Wu, Objectfolder 2.0: A multisensory object dataset for sim2real transfer, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10598–10608.
- [123] S. Kanitkar, H. Jiang, W. Yuan, Poseit: A visual-tactile dataset of holding poses for grasp stability analysis, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022, pp. 71–78. doi:10.1109/IROS47612.2022.9981562.
- [124] J. Kerr, H. Huang, A. Wilcox, R. Hoque, J. Ichnowski, R. Calandra, K. Goldberg, Self-supervised visuo-tactile pretraining to locate and follow garment features, arXiv preprint arXiv:2209.13042 (2022).
- [125] R. Gao, Y. Dou, H. Li, T. Agarwal, J. Bohg, Y. Li, L. Fei-Fei, J. Wu, The objectfolder benchmark: Multisensory learning with neural and real objects, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 17276–17286.
- [126] B. M. R. Lima, V. N. S. S. Danyamraju, T. E. A. de Oliveira, V. P. da Fonseca, A multimodal tactile dataset for dynamic texture classification, *Data in Brief* 50 (2023) 109590.
- [127] S. Suresh, Z. Si, S. Anderson, M. Kaess, M. Mukadam, Midastouch: Monte-carlo inference over distributions across sliding touch, in: Conference on Robot Learning, PMLR, 2023, pp. 319–331.
- [128] B. Khojasteh, Y. Shao, K. J. Kuchenbecker, **MPI-10: Haptic-Auditory Measurements from Tool-Surface Interactions** (2024). doi:10.17617/3.PM8R94. URL <https://doi.org/10.17617/3.PM8R94>
- [129] L. Fu, G. Datta, H. Huang, W. C.-H. Panitch, J. Drake, J. Ortiz, M. Mukadam, M. Lambeta, R. Calandra, K. Goldberg, A touch, vision, and language dataset for multimodal alignment (2024). arXiv: 2402.13232.

- [130] J. K. Balasubramanian, B. L. Kodak, Y. Vardar, Sens3: Multisensory database of finger-surface interactions and corresponding sensations, arXiv preprint arXiv:2401.01818 (2024).
- [131] N. Cheng, C. Guan, J. Gao, W. Wang, Y. Li, F. Meng, J. Zhou, B. Fang, J. Xu, W. Han, Touch100k: A large-scale touch-language-vision dataset for touch-centric multimodal representation, arXiv e-prints (2024) arXiv–2406.
- [132] Y. Dou, F. Yang, Y. Liu, A. Loquercio, A. Owens, Tactile-augmented radiance fields, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 26529–26539.
- [133] A. Figueira, B. Vaz, Survey on synthetic data generation, evaluation methods and gans, Mathematics 10 (15) (2022) 2733.
- [134] S. Wang, M. Lambeta, P.-W. Chou, R. Calandra, Tacto: A fast, flexible, and open-source simulator for high-resolution vision-based tactile sensors, IEEE Robotics and Automation Letters 7 (2) (2022) 3930–3937.
- [135] Z. Si, W. Yuan, Taxim: An example-based simulation model for gel-sight tactile sensors, IEEE Robotics and Automation Letters 7 (2) (2022) 2361–2368.
- [136] Z. Chen, S. Zhang, S. Luo, F. Sun, B. Fang, Tacchi: A pluggable and low computational cost elastomer deformation simulator for optical tactile sensors, IEEE Robotics and Automation Letters 8 (3) (2023) 1239–1246.
- [137] Z. Si, G. Zhang, Q. Ben, B. Romero, Z. Xian, C. Liu, C. Gan, Diffactile: A physics-based differentiable tactile simulator for contact-rich robotic manipulation, arXiv preprint arXiv:2403.08716 (2024).
- [138] W. Du, W. Xu, J. Ren, Z. Yu, C. Lu, Tacipc: Intersection-and inversion-free fem-based elastomer simulation for optical tactile sensors, IEEE Robotics and Automation Letters (2024).
- [139] Y. Zhao, K. Qian, B. Duan, S. Luo, Fots: A fast optical tactile simulator for sim2real learning of tactile-motor robot manipulation skills, IEEE Robotics and Automation Letters (2024).
- [140] S. Cui, Y. Wang, S. Wang, Q. Li, R. Wang, C. Zhang, Tactile imprint simulation of gelstereo visuotactile sensors, in: 2023 IEEE International Conference on Mechatronics and Automation (ICMA), IEEE, 2023, pp. 650–656.
- [141] Q. K. Luu, N. H. Nguyen, et al., Simulation, learning, and application of vision-based tactile sensing at large scale, IEEE Transactions on Robotics 39 (3) (2023) 2003–2019.
- [142] D. Nie, J. Liu, X. Sun, Influence of surface tactile data quantity on material classification in unstructured environments, IEEE Transactions on Instrumentation and Measurement 70 (2021) 1–12.
- [143] G. Liu, S. Lv, C. Wang, X. Li, W. Nai, Surface material classification based on unbalanced visual and haptic data: A double-dqn method, IEEE Transactions on Instrumentation and Measurement 72 (2023) 1–12.
- [144] S. Cai, K. Zhu, Multi-modal transformer-based tactile signal generation for haptic texture simulation of materials in virtual and augmented reality, in: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), IEEE, 2022, pp. 810–811.
- [145] H. Culbertson, K. J. Kuchenbecker, Ungrounded haptic augmented reality system for displaying roughness and friction, IEEE/ASME Transactions on Mechatronics 22 (4) (2017) 1839–1849.
- [146] Y. Vardar, C. Wallraven, K. J. Kuchenbecker, Fingertip interaction metrics correlate with visual and haptic perception of real surfaces, in: 2019 IEEE World Haptics Conference (WHC), IEEE, 2019, pp. 395–400.
- [147] B. A. Richardson, Y. Vardar, C. Wallraven, K. J. Kuchenbecker, Learning to feel textures: Predicting perceptual similarities from unconstrained finger-surface interactions, IEEE Transactions on Haptics 15 (4) (2022) 705–717.
- [148] N. Heravi, H. Culbertson, A. M. Okamura, J. Bohg, Development and evaluation of a learning-based model for real-time haptic texture rendering, IEEE Transactions on Haptics (2024).
- [149] F. Wei, J. Zhao, C. Shan, Z. Yuan, Alignment and multi-scale fusion for visual-tactile object recognition, in: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE, 2022, pp. 1–8.
- [150] H. Zheng, L. Fang, M. Ji, M. Strese, Y. Özer, E. Steinbach, Deep learning for surface material classification using haptic and visual information, IEEE Transactions on Multimedia 18 (12) (2016) 2407–2416.
- [151] T. Zhang, Y. Cong, J. Dong, D. Hou, Partial visual-tactile fused learning for robotic object recognition, IEEE Transactions on Systems, Man, and Cybernetics: Systems 52 (7) (2021) 4349–4361.
- [152] T. Zhang, Y. Cong, G. Sun, J. Dong, Y. Liu, Z. Ding, Generative partial visual-tactile fused object clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, 2021, pp. 6156–6164.
- [153] T. Zhou, X. Zhang, B. Kang, M. Chen, Multimodal fusion recognition for digital twin, Digital Communications and Networks (2022).
- [154] M. Ji, L. Fang, H. Zheng, M. Strese, E. Steinbach, Preprocessing-free surface material classification using convolutional neural networks pre-trained by sparse autoencoder, in: 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2015, pp. 1–6.
- [155] Y. Liu, S. Lu, H. Culbertson, Texture classification by audio-tactile crossmodal congruence, in: 2022 IEEE Haptics Symposium (HAP-TICS), IEEE, 2022, pp. 1–7.
- [156] W. Zheng, H. Liu, B. Wang, F. Sun, Cross-modal surface material retrieval using discriminant adversarial learning, IEEE transactions on industrial informatics 15 (9) (2019) 4978–4987.
- [157] W. Zheng, H. Liu, F. Sun, Lifelong visual-tactile cross-modal learning for robotic material perception, IEEE transactions on neural networks and learning systems 32 (3) (2020) 1192–1203.
- [158] Y. Li, H. Zhao, H. Liu, S. Lu, Y. Hou, Research on visual-tactile cross-modality based on generative adversarial network, Cognitive Computation and Systems 3 (2) (2021) 131–141.
- [159] S. Cai, K. Zhu, Y. Ban, T. Narumi, Visual-tactile cross-modal data generation using residue-fusion gan with feature-matching and perceptual losses, IEEE Robotics and Automation Letters 6 (4) (2021) 7525–7532.
- [160] Q. Xi, F. Wang, L. Tao, H. Zhang, X. Jiang, J. Wu, Cm-avae: Cross-modal adversarial variational autoencoder for visual-to-tactile data generation, IEEE Robotics and Automation Letters (2024).
- [161] Y. Zhan, X. Sun, Q. Wang, W. Nai, Method for audio-to-tactile cross-modality generation based on residual u-net, IEEE Transactions on Instrumentation and Measurement (2023).
- [162] Y. Fang, X. Zhang, W. Xu, G. Liu, J. Zhao, Bidirectional visual-tactile cross-modal generation using latent feature space flow model, Neural Networks 172 (2024) 106088.
- [163] F. Yang, C. Feng, Z. Chen, H. Park, D. Wang, Y. Dou, Z. Zeng, X. Chen, R. Gangopadhyay, A. Owens, et al., Binding touch to everything: Learning unified multimodal tactile representations, arXiv preprint arXiv:2401.18084 (2024).
- [164] S. Cui, R. Wang, J. Wei, J. Hu, S. Wang, Self-attention based visual-tactile fusion learning for predicting grasp outcomes, IEEE Robotics and Automation Letters 5 (4) (2020) 5827–5834.
- [165] C. Liu, Z. Yi, B. Huang, Z. Zhou, S. Fang, X. Li, Y. Zhang, X. Wu, A deep learning method based on triplet network using self-attention for tactile grasp outcomes prediction, IEEE Transactions on Instrumentation and Measurement (2023).
- [166] C. Liu, B. Huang, Y. Liu, Y. Su, K. Mai, Y. Zhang, Z. Yi, X. Wu, A self-supervised contrastive learning method for grasp outcomes prediction, in: 2023 IEEE International Conference on Real-time Computing and Robotics (RCAR), IEEE, 2023, pp. 737–742.
- [167] Z. Zhang, Z. Zhou, H. Wang, Z. Zhang, H. Huang, Q. Cao, Grasp stability assessment through attention-guided cross-modality fusion and transfer learning, in: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2023, pp. 9472–9479.
- [168] F. Wang, H. Liu, F. Sun, H. Pan, Fabric recognition using zero-shot learning, Tsinghua Science and Technology 24 (6) (2019) 645–653.
- [169] G. Cao, J. Jiang, D. Bollegala, S. Luo, Learn from incomplete tactile data: Tactile representation learning with masked autoencoders, in: 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2023, pp. 10800–10805.
- [170] B. Khojasteh, F. Solowjow, S. Trimpe, K. J. Kuchenbecker, Multimodal multi-user surface recognition with the kernel two-sample test, IEEE Transactions on Automation Science and Engineering (2023).
- [171] H. Liu, F. Wang, F. Sun, B. Fang, Surface material retrieval using weakly paired cross-modal learning, IEEE Transactions on Automation Science and Engineering 16 (2) (2018) 781–791.
- [172] J. Wei, S. Cui, J. Hu, P. Hao, S. Wang, Z. Lou, Multimodal unknown surface material classification and its application to physical reasoning, IEEE Transactions on Industrial Informatics 18 (7) (2021) 4406–4416.
- [173] R. Song, X. Sun, G. Liu, Cross-modal generation of tactile friction coefficient from audio and visual measurements by transformer, IEEE Transactions on Instrumentation and Measurement (2023).

- [174] K. Priyadarshini, S. Chaudhuri, S. Chaudhuri, Perceptnet: Learning perceptual similarity of haptic textures in presence of unorderable triplets, in: 2019 IEEE World Haptics Conference (WHC), IEEE, 2019, pp. 163–168.
- [175] G. Cao, Y. Zhou, D. Bollegala, S. Luo, Spatio-temporal attention model for tactile texture recognition, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, pp. 9896–9902.
- [176] G. Rouhafzay, A.-M. Cretu, P. Payeur, Transfer of learning from vision to touch: A hybrid deep convolutional neural network for visuo-tactile 3d object recognition, Sensors 21 (1) (2020) 113.
- [177] J.-T. Lee, D. Bollegala, S. Luo, “touching to see” and “seeing to feel”: Robotic cross-modal sensory data generation for visual-tactile perception, in: 2019 International Conference on Robotics and Automation (ICRA), IEEE, 2019, pp. 4276–4282.
- [178] V. Dave, F. Lygerakis, E. Rueckert, Multimodal visual-tactile representation learning through self-supervised contrastive pre-training, arXiv preprint arXiv:2401.12024 (2024).
- [179] S. Cai, L. Zhao, Y. Ban, T. Narumi, Y. Liu, K. Zhu, Gan-based image-to-friction generation for tactile simulation of fabric material, Computers & Graphics 102 (2022) 460–473.
- [180] Y. Li, Y. Du, C. Liu, F. Williams, M. Foshey, B. Eckart, J. Kautz, J. B. Tenenbaum, A. Torralba, W. Matusik, Learning to jointly understand visual and tactile signals, in: The Twelfth International Conference on Learning Representations, 2023.
- [181] F. Yang, J. Zhang, A. Owens, Generating visual scenes from touch, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 22070–22080.
- [182] B. Li, J. Bai, S. Qiu, H. Wang, Y. Guo, Vito-transformer: A visual-tactile fusion network for object recognition, IEEE Transactions on Instrumentation and Measurement (2023).
- [183] W. Lei, Y. Ge, K. Yi, J. Zhang, D. Gao, D. Sun, Y. Ge, Y. Shan, M. Z. Shou, Vit-lens: Towards omni-modal representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 26647–26657.
- [184] Y. Lyu, X. Zheng, D. Kim, L. Wang, Omnidbind: Teach to build unequal-scale modality interaction for omni-bind of all, arXiv preprint arXiv:2405.16108 (2024).
- [185] B. Khojasteh, Y. Shao, K. J. Kuchenbecker, Robust surface recognition with the maximum mean discrepancy: Degrading haptic-auditory signals through bandwidth and noise, IEEE Transactions on Haptics (2024).
- [186] P. Manocha, Z. Jin, R. Zhang, A. Finkelstein, Cdpm: Contrastive learning for perceptual audio similarity, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 196–200.
- [187] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 586–595.