Material Attribute Estimation as Part of Telecommunication Augmented Reality, Virtual Reality, and Mixed Reality System: Systematic Review

Nicole Christoff * and Krasimir Tonchev †

Faculty of Telecommunications, Technical University of Sofia, 8 Kliment Ohridski Blvd., 1756 Sofia, Bulgaria; k_tonchev@tu-sofia.bg
* Correspondence: nicole.christoff@tu-sofia.bg
† These authors contributed equally to this work.

Abstract: The integration of material attribute estimation (MAE) within augmented reality, virtual reality, and mixed reality telecommunication systems stands as a pivotal domain, evolving rapidly with the advent of the Tactile Internet. This unifying implementation process has the potential for improvements in the realism and interactivity of immersive environments. The interaction between MAE and the haptic Internet could lead to significant advances in haptic feedback systems, enabling more accurate and responsive user experiences. This systematic review is focused on the intersection of MAE and the Tactile Internet, aiming to find an implementation path between these technologies. Motivated by the potential of the haptic Internet to advance telecommunications, we explore its potential to advance the analysis of material attributes within AR, VR, and MR applications. Through an extensive analysis of current research approaches, including machine learning methods, we explore the possibilities of integrating the TI into MAE. By exploiting haptic and visual properties stored in the materials of 3D objects and using them directly during rendering in remote access scenarios, we propose a conceptual framework that combines data capture, visual representation, processing, and communication in virtual environments.

Keywords: material attribute estimation; Tactile Internet; object representation; augmented reality; virtual reality; mixed reality

1. Introduction

The basis of augmented reality (AR), virtual reality (VR), and mixed reality (MR) technologies as part of telecommunications systems is undergoing a revolutionary transformation, fundamentally changing our interactions with digital information and the physical world. A crucial aspect of improving the user experience in these systems is the precise estimation and rendering of material attributes in the virtual environment. In this context, the evaluation of material attributes serves as a core element of telecommunication AR, VR, and MR systems. This will allow users to interact with virtual objects that display realistic properties. Accurately simulating material behavior under different lighting conditions and viewing angles allows users to seamlessly integrate virtual objects with their surroundings. This in turn enriches the overall sense of presence and immersion. The Tactile Internet (TI) is emerging as a core element of this infrastructure. It offers communication capabilities to refine the estimation of material attributes in telecom AR, VR, and MR systems. TI can be implemented to provide real-time feedback, foster collaborative environments, integrate haptic sensations, and more. By leveraging the capabilities of the Tactile Internet, these systems can provide immersive and interactive experiences that closely approximate real-world interactions with physical materials. Thus, while AR, VR, and MR technologies remain the focus, it is the integration with TI that serves as the foundation, enhancing their functionality and realism.
In researching the intersections of technology and sensory perception, various methods find application in MR, VR, and the Tactile Internet. Multimodal imaging techniques inspired by medical imaging [1] are applicable to MR and VR, aiding scene understanding and object recognition tasks. By providing a more accurate and efficient method of aligning virtual objects with real-world scenes, ref. [1] improves realism and immersion in AR, VR, and MR experiences. The proposed method effectively handles different materials and lighting conditions, providing a seamless integration of virtual and physical environments. Moreover, understanding surface properties such as gloss and mechanical attributes informs the development of realistic simulations in MR and VR environments [2]. The authors of this work evaluate intrinsic properties such as reflectance, shape, and illuminance. This is achieved through energy minimization frameworks such as the SIRFS model, which optimizes reflectance, depth, and illumination conditions. In this way, AR, VR, and MR systems can achieve a more realistic rendering of virtual objects, resulting in improved immersion, presence, and user engagement.

Similarly, understanding tissue attributes through visual and tactile data informs the creation of immersive experiences in MR, VR, and the Tactile Internet [3]. The study focused on assessing the reliability of softness judgments using visual and haptic information and assessing the contribution of each sense to the judgments. It also investigates how visual and haptic information are integrated into the perception of softness. In [4], material texture patterns are analyzed using image filtering techniques such as fast Fourier transform (FFT). Vibrations are then modeled based on these texture patterns and vibration motors are used to provide the appropriate tactile feedback. Tasks involving fabric texture recognition and interaction benefit from datasets designed for fabric-related tasks, enriching experiences in MR, VR, and the haptic Internet [5]. In this study, the attributes are defined based on the output voltage waveforms generated by the sensor array as it slides over the fabric surface. Different textures lead to different stress fluctuations, which allows the distinction between fabric types. Haptic technology, fundamental in simulating tactile sensations in VR environments [4], finds direct relevance in improving the user experience within the Tactile Internet through tactile feedback [6]. Attributes such as viscosity, elastoplasticity, and contact forces are defined by mathematical formulations and constraints within computational models [6]. Three-dimensional rendering techniques [7] and reconstruction [8] contribute to realistic rendering under various conditions in MR and VR settings. In addition, joint, material, and lighting geometry estimation methods hold promise for improved realism in MR, VR, and Tactile Internet applications [9]. The geometry is represented using a sign distance field (SDF), the material properties are estimated using a Bidirectional Reflectance Distribution Function (BRDF) field, and the illumination conditions are estimated using neural incident light fields (NeILFs) and output radiation fields. Frameworks and metrics tailored for haptic Internet applications aim to improve haptic sessions on networks, improving the user experience [10]. These achievements demonstrate the interconnectedness of technologies driving innovation in MR, VR, and the haptic Internet, enriching user experiences in immersive and interactive environments.

Motivation: With the advent of the TI and its potential in telecommunications, there is a need to explore its integration into material attribute analysis within AR, VR, and MR applications. So far, the evaluation of material attributes in these technologies has been discussed in detail, but the use of the TI has hardly been addressed. Our motivation stems from the key role that the TI can play in developing the user experience. This could be achieved by improving the accuracy and efficiency of material attribute estimation, thereby advancing the realism and interactivity of AR, VR, and MR environments.

Contributions: Our research presents a comprehensive review of the current approaches to material attribute evaluation, with a focus on the integration of the Tactile Internet (TI). This includes machine learning methods to estimate visual attributes based on 2D and 3D information. We explore key aspects such as mesh generation, texture generation, material evaluation, object manipulation, object property evaluation, object tracking, and performance evaluation in the context of the TI. Our goal is to advance the un-
derstanding and practical implementation of the TI by proposing a conceptual framework that combines data capture, visual presentation, processing, and communication within virtual environments.

2. Object Representation

Object representation refers to the process of creating digital models or descriptions of physical objects in a virtual environment. In VR, objects are usually represented as 3D models with geometry and texture, and made of some material. In a similar way, objects are also represented in AR, but superimposed on the user’s real environment. This requires a realistic resemblance to real-world surfaces and objects. MR combines elements of both VR and AR, allowing virtual and real objects to coexist and interact with the user’s physical environment. Spatial sensation and depth perception must be considered when representing objects in MR. The input data for object representation can be 2D images or 3D models. In the context of haptic imaging, these inputs are important. In VR, 3D models are integrated into the virtual environment, providing detailed geometry, texture, and material properties. In AR and MR, these models are superimposed on the real world, which requires precise alignment and realistic representation. Haptic rendering involves creating touch sensations corresponding to virtual objects, allowing users to feel as if they are interacting with real objects. Through the Tactile Internet, it would be possible to transmit and interact with virtual objects as if they were physically present. Haptic feedback devices such as tactile gloves for example are used to convey sensations of touch, pressure, texture, and force feedback. Through a haptic interface, the user can perceive the properties of the objects with which he or she interacts. Also, the Tactile Internet facilitates the transmission of representations of objects through digital models and simulations. This allows the representation of a physical object, including its shape, texture, and mechanical properties, in virtual environments. The representation of objects plays a key role in creating authentic and immersive virtual experiences, and the integration of the Tactile Internet enables real-world interaction with digital replicas of physical objects, changing the way we interact with virtual environments.

Figure 1 presents the step-by-step process involved in the analysis of objects within 2D images and 3D models and their digital representation. It begins with the segmentation and modeling of the objects; then, attributes are extracted and property evaluation is performed. Each of the stages contributes to the complete extraction of the features and properties of the object. Through this algorithm, objects can be accurately identified, categorized, and analyzed for various applications. The first step in the representation of the objects in a virtual environment is the pre-processing of the images. This includes enhancing the visual qualities of the image, such as removing noise, in order to subsequently analyze and model the image content. Examples of object detection are You Only Look Once (YOLO) [11] or Single Shot Multi-Box Detector (SSD) [12]. They are used for the identification and segmentation of objects in the image. In the context of VR, these exposed objects are represented as 3D models with detailed geometry, texture, and material properties. Analysis and segmentation are also performed in 3D, with dimensioning performed directly from the model. In 2D, object size and texture extraction can be improved by, for example, measuring the size of a window in which the region of interest is located and, for material texture, examining the textural features of the image. Both physical properties and reflective properties can then be evaluated. When rendering objects in a virtual environment, factors such as lighting conditions, occlusion, and perspective must be taken into account to create a seamless integration between the virtual and real elements. The accuracy of object representation is critical to immersion in a virtual environment. This may require models to undergo optimization to ensure smooth rendering and interaction at high frame rates. In addition, material physical properties can be applied to objects to enable realistic interactions such as gravity, collision, etc. Visual information about an object refers to the features that can be perceived through visual perception. When we observe an object,
we try to extract different information by which to characterize it, such as shape, size, color, texture, orientation, and spatial relations. To describe the shape, we use the overall outline or shape of an object, including its contours and silhouette, as seen from multiple viewpoints in 3D space. In terms of appearance, the specific hue, saturation, and brightness of an object’s surface can characterize its color, and, if it is presented in 3D, possible shading from other objects must also be taken into account. It is important to note that the texture or surface characteristics of an object, such as smoothness, roughness, or patterns, can affect its visual and tactile properties. The physical dimensions of an object relative to its surroundings provide information about its size and distance, including its distance from the observation point. For object perception in three-dimensional space, occlusion, relative size, shading, and viewing perspective must be taken into account. The perception of visual information can be difficult if the spatial orientation and location of the object in the field of view, including its angle, tilt, and position relative to other objects, are not taken into account. While many aspects of visual information are common and inherent to 2D and 3D representations, in 3D, additional perceptual elements are depth of representation and spatial relationships. Any dynamic changes in the object’s appearance, position, or orientation over time that may provide additional information about its properties or behavior can be used to facilitate the object characterization process. Consideration should also be given to the surrounding environment or context in which the object is observed, including background elements, lighting conditions, and environmental cues.

Figure 1. Graphical representation of object analysis and representation pipeline. Correct representation of the object increases its correct automatic segmentation. This helps for a correct model of the object and the extraction of its attributes, which in turn are represented in the form of a model. The two models taken together could be used as combined features that in turn serve to predict the object and extract its physical properties.

One of the main characteristics of an object is its geometric shape or configuration. This encompasses the physical dimensions and contours that determine its appearance and
structure in space (see Table 1). Form representation in virtual environments includes techniques such as geometric modeling, mesh generation, and surface reconstruction [12]. This would be an easy task as long as we are dealing with basic shapes such as spheres, cubes, etc. (see Table 2). However, when it comes to objects with a different shape from this, then the use of more innovative methods based on visual information is necessary. Bargmann et al. [13] estimate shape attributes such as size, shape, spatial orientation, distribution of grains, and grain boundaries using experimental techniques like serial sectioning and imaging based on transmissive radiation (tomography), as well as computational methods including physics-based simulation and geometry-based approaches. Several studies, such as [14–16], estimate geometry attributes through joint optimization using physically based rendering techniques, differentiable rendering, and Monte Carlo integration. They optimize shape from multi-view images through 2D supervision, represent it with a signed distance field defined on a three-dimensional grid, and reduce it to a triangular surface mesh. Material properties are estimated using a physically based material model, and environment lighting is represented using a high dynamic range light probe stored as a floating-point texture. End-to-end inverse rendering pipelines employing Monte Carlo sampling-based path tracing are used to estimate geometry attributes, among others. Chen et al. [8] estimate shape attributes using a differentiable rendering framework based on deferred shading, leveraging a hybrid differentiable renderer combining rasterization and ray tracing. Monte Carlo (MC) integration and spherical Gaussians (SGs) are used to approximate outgoing radiance and lighting. Liang et al. [17] estimates shape attributes by using radiometric features extracted from different images, such as RGB, NIR, etc. They exploit different behaviors of different materials in terms of light reflection, refraction, and absorption, observed through polarization properties and NIR absorption. The object attributes can also be estimated using deep learning algorithms. Achlioptas et al. [18] estimate shape attributes using a deep AutoEncoder (AE) network and various generative models like Generative Adversarial Networks (GANs) and Gaussian Mixture Models (GMMs). Shape operations are performed via algebraic manipulations in the latent space of the AE. Authors like Sharma et al. [19] estimate shape attributes based on the similarity of materials at different pixels in the image to the material at the query pixel location, using a material-aware multi-scale encoder trained on synthetic renderings with material ground-truth labels. Lagunas et al. [20] measure the similarity in appearance between different materials based on human similarity judgments collected through crowdsourced experiments. They use a deep learning architecture with a novel loss function to learn a feature space for materials correlated with perceived appearance similarity. Other authors, such as Baars et al. [21], estimate the volume of an object by using the mesh created from the point cloud and dividing the resulting mesh into tetrahedra. The volume of each tetrahedron is calculated and summed to give an estimate of the total volume. Mass is estimated by combining information about the object’s 3D shape, density, and volume. Density and volume are estimated using neural networks, with the geometry module providing object shape information [22]. Pose estimation is an essential element of object orientation determination in various applications, whether in augmented reality, virtual reality, or mixed reality. Armeni et al. [23] entered the field of object position estimation using a combination of experimental techniques and computational methods to reveal not only object size but also other related attributes. The focus is on extracting object pose information using the Scene Graph paradigm in 3D, generating a 3D Scene Graph. Besides pose estimation, ref. [24] investigated the physical properties of materials. Through innovative approaches such as quantile loss function, machine learning for prediction intervals, and Gaussian processes, [24] aims to decipher not only object poses but also the complex details of material attributes. Meanwhile, ref. [25] uses a holistic approach, and combines object position estimation with additional attributes such as shape and reflection coefficients. Taking advantage of neural SDF-based shape reconstruction and material distillation and lighting stages, ref. [25] aims for a comprehensive understanding of objects in their environment. Size estimation is an important aspect of site analysis.
Table 1. Object representation and material attribute estimation comparison.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Geometric Shape Estimation</th>
<th>Material Attribute Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bargmann et al. [13]</td>
<td>Experimental and computational methods (serial sectioning, tomography, simulation)</td>
<td>N/A</td>
</tr>
<tr>
<td>Zeng et al. [14]</td>
<td>Joint optimization with physically based rendering</td>
<td>Physically based material model</td>
</tr>
<tr>
<td>Chen et al. [6]</td>
<td>Differentiable rendering framework (deferred shading)</td>
<td>Monte Carlo integration, spherical Gaussians</td>
</tr>
<tr>
<td>Liang et al. [17]</td>
<td>Imaging modalities (RGB, polarization, NIR)</td>
<td>Behavior analysis of materials (reflection, refraction, absorption)</td>
</tr>
<tr>
<td>Achlioptas et al. [18]</td>
<td>Deep AutoEncoder network, Generative models (GANs, GMMs)</td>
<td>N/A</td>
</tr>
<tr>
<td>Sharma et al. [19]</td>
<td>Multi-scale encoder trained on synthetic renderings</td>
<td>Similarity to the material at query pixel location</td>
</tr>
<tr>
<td>Lagunas et al. [20]</td>
<td>Deep learning architecture with novel loss function</td>
<td>Human similarity judgments</td>
</tr>
<tr>
<td>Baars et al. [21]</td>
<td>The surface meshing of a point cloud, neural network towers</td>
<td>Physically based material model</td>
</tr>
<tr>
<td>Richardt et al. [12]</td>
<td>Geometric modeling, mesh generation, surface reconstruction</td>
<td>Texture datasets, fabric attribute understanding</td>
</tr>
<tr>
<td>Zhang et al. [14]</td>
<td>Signed distance field representation, BRDF field, neural incident light fields</td>
<td>Reflectance, depth, illumination terms optimization</td>
</tr>
<tr>
<td>Corsini et al. [1]</td>
<td>Multimodal image registration, scene understanding, object recognition</td>
<td>Vibration modeling based on texture patterns</td>
</tr>
<tr>
<td>Yoon et al. [4]</td>
<td>Image filtering techniques (fast Fourier transform)</td>
<td>N/A</td>
</tr>
<tr>
<td>Standley et al. [22]</td>
<td>Image-based mass estimation</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Geometry-related attributes.

<table>
<thead>
<tr>
<th>Shape</th>
<th>3D Model</th>
<th>Geometry</th>
<th>Surface Area</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sphere</td>
<td>1 curved surface</td>
<td>0 edges 0 vertices</td>
<td>$S = 4\pi r^2$</td>
<td>$V = \frac{4}{3}\pi r^3$</td>
</tr>
<tr>
<td>Cube</td>
<td>6 faces 12 edges 8 vertices</td>
<td>$S = 6s^2$</td>
<td>$V = s^3$</td>
<td></td>
</tr>
<tr>
<td>Cylinder</td>
<td>2 edges 1 curved surface 2 edges 0 vertices</td>
<td>$S = 2\pi r^2 + 2\pi rh$</td>
<td>$V = \pi r^2 h$</td>
<td></td>
</tr>
<tr>
<td>Pyramid</td>
<td>4 faces 6 edges 4 vertices</td>
<td>$S = B + \frac{1}{2} l$</td>
<td>$V = \frac{1}{3} Bh$</td>
<td></td>
</tr>
<tr>
<td>Cone</td>
<td>1 face 1 curved surface 1 edge 0 vertices</td>
<td>$S = \pi r^2 + \pi \ell$</td>
<td>$V = \frac{1}{3} \pi r^2 h$</td>
<td></td>
</tr>
</tbody>
</table>

3. Visual Attribute Grouping and Estimation

The presentation of materials in MR, VR, AR, and the Tactile Internet should also include simulation of the visual and tactile properties of the materials. Material properties such as color, gloss, transparency, and reflectivity are simulated to match the appearance of physical materials. Surfaces can be rendered with textures, shaders, and lighting effects to emulate the appearance of various materials such as metal, wood, fabric, glass, and more. This requires the material properties to be adjusted to match the lighting and environmental conditions, ensuring a seamless blend between virtual and real objects. It is possible to simulate sensations such as texture, stiffness, elasticity, and surface roughness through haptic interfaces. Material properties are encoded into haptic feedback algorithms, enabling
the user to feel the physical characteristics of virtual objects with which he or she interacts remotely. The visual characteristics of objects make it possible to calculate and simulate the physical properties of the object. One of the main characteristics is the color. This represents the visual perception of the wavelength of light reflected or emitted from the surface of an object. Different materials absorb and reflect light in different ways, resulting in color variations. Features such as roughness, smoothness, graininess, or softness can be used to characterize the surface texture of an object. Texture is defined by the composition of surface features visible at a given object scale. For example, a smooth surface implies homogeneity, while a rough surface indicates the presence of bumps or protrusions. Materials with a high gloss, such as polished metal or glass, create sharp, mirror-like reflections, while matte surfaces scatter light diffusely. Luster is affected by factors such as surface smoothness, refractive index, and microstructure. Transparent materials, such as glass, allow light to pass through with minimal distortion, while opaque materials block or scatter light. Translucent materials partially transmit light while scattering it in different directions. Opaque materials, such as metal, ceramics, and others, stop light from passing through and appear solid. Materials with a high luster, such as metal, reflect light brightly and evenly, giving a metallic or glassy appearance. On the other hand, matte materials have a low gloss and diffuse reflect light. The purpose of examining the luster of an object is to infer its surface smoothness, reflectivity, and optical properties. The current section aims to make material visual attribute comparison, visually presented in Table 3.

Table 3. Material visual attribute comparison.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Estimation Techniques</th>
<th>Methods</th>
<th>Evaluation Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflectance-related</td>
<td>SVBRDF-net [26], NeRD [7], MLPs, spherical Gaussians, diffusion priors, semantic segmentation [27]</td>
<td>SVBRDF-net prediction accuracy, perceptual evaluation</td>
<td>Realistic image synthesis, BRDF parameter estimation, perceptual evaluation</td>
</tr>
<tr>
<td>Albedo</td>
<td>Cascaded network architectures [28], MLPs, spherical Gaussians, autoencoders, neural textures, renderers [29], NeRF [30]</td>
<td>SVBRDF prediction accuracy, perceptual evaluation</td>
<td></td>
</tr>
<tr>
<td>Roughness</td>
<td>Scoring method [31], differentiable rendering [8], manual annotation [32], classifier training [33], SVBRDF-net [26]</td>
<td>Roughness estimation accuracy, perceptual evaluation</td>
<td></td>
</tr>
<tr>
<td>Metallicity</td>
<td>Similar estimation methods [31], NeRD [7], StyleGAN2 [34]</td>
<td>Metallicity perception evaluation, joint optimization of shape, BRDF, and luminosity</td>
<td></td>
</tr>
<tr>
<td>Translucency</td>
<td>Behavioral tasks [35], diffusion, semantic segmentation, and material estimation models [27]</td>
<td>Translucency estimation accuracy, perceptual evaluation</td>
<td></td>
</tr>
<tr>
<td>Emissivity</td>
<td>Human labeling and measurements [36], scene division into mesostructure textures and basic shapes [30]</td>
<td>Direct measurements, perceptual evaluation</td>
<td></td>
</tr>
<tr>
<td>Texture Analysis</td>
<td>Pattern Recognition Image filtering techniques [4], semantic understanding of natural language descriptions [37]</td>
<td>Pattern analysis accuracy, semantic understanding evaluation</td>
<td></td>
</tr>
<tr>
<td>Structural Analysis</td>
<td>Structural Analysis Image processing techniques, machine learning algorithms</td>
<td>Structural analysis accuracy, machine learning model evaluation</td>
<td></td>
</tr>
</tbody>
</table>

3.1. Material Visual Attributes: Reflectance-Related

Reflectance-related material visual attributes are essential for understanding surface properties. Various evaluation techniques are used to quantify these attributes. One of these is albedo, including both diffuse and specular components. It can be estimated using innovative solutions such as SVBRDF-net, a convolutional neural network (CNN) approach [26], or through Neural Reflectance Decomposition (NeRD), which optimizes shape, BRDF, and illumination [7]. Additionally, a combination of techniques, including Multilayer Perceptron (MLP) and spherical Gaussians, has been effective [33], along with diffusion priors and semantic segmentation [27]. Reflectance estimation methods encompass raw Time-of-Flight (ToF) measurements with processing algorithms like depth normalization and noise removal [38], as well as differential angular imaging and dataset collection [39]. Furthermore, techniques like diffusion priors, semantic segmentation, and material estimation models
have been utilized [27]. Estimation of the Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) involves advanced network designs for shape, illumination, and SVBRDF predictions, such as cascaded network architectures [28]. Moreover, combinations of methods like MLPs and spherical Gaussians have shown promise [33], along with neural networks incorporating autoencoders, neural textures, and renderers [29]. Disentangling scenes into meso-structure textures and underlying base shapes has enabled the estimation of diffuse and specular coefficients using neural radiance fields (NeRFs) [30].

Figure 2 illustrates how the combination of roughness and specularity affects the appearance of materials, particularly in terms of their gloss and surface texture.

![Figure 2. Visual representation of surface roughness and specular reflectance: the parameters of a BRDF model [40].](image)

### 3.2. Material Visual Attributes: Surface Property Estimation

Understanding the visual attributes of materials is critical in a variety of fields. Surface properties such as roughness, metallicity, translucency, and emissivity play a major role in determining the appearance and behavior of materials in real and virtual environments. The use of methods and algorithms from machine learning and computer vision has increased the accuracy in estimating these attributes.

Quantifying the roughness of a given surface is accomplished with a variety of approaches, ranging from observer estimates to neural network-based solutions. Nagai et al. [31] introduce a scoring method in which observers rate sample photographs based on nine surface features. Chen et al. [8] present a novel approach using a differentiable rendering framework combining rasterization, ray tracing, and Monte Carlo algorithm integration for roughness estimation. Schwartz et al. [32] introduce a manual annotation approach combined with classifier training. Other authors, Chen et al. [33], introduce a method using MLP and spherical Gaussians. Li et al. [26] propose a convolutional neural network-based SVBRDF-net that estimates the roughness and other characteristics of a material. Another attribute to evaluate is metallicity. Nagai et al. [31] use a similar estimation method for metallicity perception. They focus on understanding how dynamic changes in reflectivity affect the perception of metallicity. Through temporal variation analysis, their approach improves the accuracy of the representation of metal surfaces compared to other methods. Boss et al. [7] propose NeRD, a neural reflectance decomposition technique for metallicity estimation involving joint optimization of shape, BRDF and luminosity. NeRD is a method that separates the reflectance properties of individual components, optimizing these aspects simultaneously to achieve realistic metal surface simulations. The advantage of this method is an improvement in both accuracy and computational efficiency. Zhou et al. [34] present TileGen, a generative model based on StyleGAN2. TileGen creates highly detailed and realistic metal textures using the generating capabilities of StyleGAN2. This
model excels at generating complex patterns and textures that closely mimic real-world metal surfaces, making it suitable for digital content creation and design. Evaluating transparency proves to be quite challenging due to its nature. Liao et al. [35] use behavioral tasks including binary classification, semantic attribute estimation, and material categorization to assess translucency. Their study involved volunteers performing tasks to distinguish and categorize materials based on translucency. This provides detailed insight into human perception, which is essential for developing more accurate models for imaging translucent materials. Wu et al. [27] propose a method combining diffusion, semantic segmentation, and material estimation models to estimate translucency. Diffusion models capture light-scattering properties, while semantic segmentation and material evaluation increase the accuracy of identifying and classifying translucent materials. This hybrid approach greatly improves the realism and fidelity of rendering translucent material.

Another attribute of crucial importance for the understanding of the behavior of the material is the emissivity and, accordingly, its evaluation. Yuan et al. [36] propose an approach based on human labeling and direct measurements of thickness and density using a ruler and scale to estimate radiance. This empirical approach, based on physical measurements and human perception in evaluating emissivity, provides a practical framework for material research applications. Huang et al. [30] introduce a method dividing the scene into mesostructure textures and a basic shape for emissivity estimation. The authors use neural radiance fields (NeRFs), decomposing the scenes into simpler components for detailed analysis. This method allows the precise investigation of emission properties. Their algorithm performs a high-resolution analysis of textures and shapes, which contributes to the emissive behavior of the material.

3.3. Visual Material Attributes: Texture Analysis

The presence of texture can provide clues to the material’s origin, manufacturing process, or structural properties. For example, patterns in wood show the direction of wood fibers, the recognition of which allows the calculation of its strength and flexibility, and can also be used for wood classification. Similarly, the surface texture of metals can reveal information about the machining processes or treatments they have undergone, which in turn can indicate their durability and potential applications. Analyzing textural patterns involves a variety of techniques, each offering unique insight into material characteristics. Yoon et al. [4] use image filtering techniques such as fast Fourier transform (FFT) for precise texture pattern analysis, allowing the identification of repeating patterns and frequency components of an object’s surface. This method is particularly effective for materials with regular, periodic textures, such as fabrics or certain types of composites. Deschaintre et al. [37] adopt a new approach based on the analysis of natural language descriptions, allowing the evaluation of texture patterns through semantic knowledge. This method allows the interpretation of texture characteristics to evaluate materials based on human perceptual attributes. This approach is suitable for materials where visual inspection is combined with subjective qualities, such as softness or roughness, which are often described in everyday language.

3.4. Material Datasets

To date, the datasets contain images of various materials, which can be of real and synthetically obtained materials. This facilitates the development and evaluation of algorithms for material synthesis, texture mapping, and material attribute evaluation. Each dataset varies in its nature, the types of materials included, the number of images, and the methods used to synthesize and evaluate the material (see Table 4). PhotoMat [41] consists of photos of real flash material with hidden maps of the material including albedo, roughness, and normals. It uses a conditional relightable GAN for material images in the RGB domain and a BRDF parameter estimator. In particular, training avoids focusing on specific materials to avoid inconsistencies with lighting conditions. UMat [42] consists of 2000 textile materials divided into 14 families, including different textures, colors, patterns,
fabric types, and material thicknesses. It uses a U-Net generator in a GAN framework for image-to-image translation and uses different loss functions for training and evaluation. MatSynth [43] covers a large collection of unduplicated high-quality, high-resolution realistic material obtained from online sources. The dataset is artificially enhanced, including material blending, rotation, cropping, and ambient lighting variations for visualizations. OpenIllumination [44] focuses on generating 3D models of various materials and includes detailed material properties such as reflectance, roughness, and surface normals. It uses a combination of physically based rendering and machine learning techniques to synthesize materials.

Table 4. Comprehensive comparison of four distinct datasets used in material synthesis and analysis.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>2D</td>
<td>2D</td>
<td>2D</td>
<td>3D</td>
</tr>
<tr>
<td><strong>Materials</strong></td>
<td>Real dataset of flash material photos with hidden material maps: albedo, roughness, normals, etc.</td>
<td>Textile materials such as crepe, jacquard, fleece, leather, etc.</td>
<td>Realistic materials like wood, stone, metal, fabric, etc.</td>
<td>Various materials including metals, plastics, fabrics, and ceramics.</td>
</tr>
<tr>
<td><strong>Material Categories</strong></td>
<td>N/A</td>
<td>14 families of textile materials</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Size of the Dataset</strong></td>
<td>N/A</td>
<td>2000</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Material Attributes</strong></td>
<td>Albedo, roughness, normals, etc.</td>
<td>Texture, color, pattern, fabric type, material thickness.</td>
<td>Reflectance, roughness, texture, color, surface finish.</td>
<td>Reflectance, roughness, surface normals, texture, color.</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>Conditional relightable GAN for material images in RGB domain and BRDF parameter estimator.</td>
<td>Uses U-Net generator within a GAN framework for image-to-image translation.</td>
<td>Utilizes a combination of physically based rendering and machine learning techniques for material-related tasks.</td>
<td>- Focuses on generating 3D models of various materials.</td>
</tr>
<tr>
<td><strong>Dataset Details</strong></td>
<td>- Uses a relightable generator to produce material images under conditional light source locations.</td>
<td>- Employs various loss functions including pixel-wise, adversarial, style, and frequency losses.</td>
<td>- Comprises a large collection of non-duplicate, high-quality, high-resolution realistic materials.</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Data Augmentation</strong></td>
<td>- Random cropping of real photos with flash highlights to obtain images with varied highlight locations.</td>
<td>- Patch-based training, affine transforms, random rescales, rotations, and intensity changes.</td>
<td>- Material blending, rotation, cropping.</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Uncertainty</strong></td>
<td>- Training strategy avoids baking highlights in neural materials to prevent mismatch with light conditions.</td>
<td>- Proposes an uncertainty quantification mechanism applied to individual per-map estimations.</td>
<td>- Environment illumination variations for renders.</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Evaluation</strong></td>
<td>- Evaluated based on the ability to produce realistic material images and BRDF parameter estimation.</td>
<td>- Evaluated on various criteria including categories, tags, creation methodology, and stationarity.</td>
<td>- Evaluated on various criteria including categories, tags, creation methodology, and stationarity.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

4. Conceptional Realization through Tactile Internet

4.1. Realization

Taking into account existing methods, we can present a theoretical solution to object analysis and representation through the Tactile Internet. Stuijt et al. [45] propose a methodology that aims to generate accurate yet efficient meshes for physical simulations (see Figure 3), especially for applications such as the Tactile Internet where bandwidth and computational cost are critical. The methodology described in the document consists of two main steps. The first is mesh reconstruction using screened Poisson surface reconstruction and marching cubes. Screened Poisson surface reconstruction models the problem as a Poisson equation, but focuses on surfaces that are not waterproof. Marching cubes, on the other hand, do not strictly require waterproof surfaces and offer a good compromise between quality and speed. The second step is to simplify the network. After reconstruction, the meshes are simplified to reduce redundant triangles.
Algorithms based on edge collapse are used, including techniques such as Hoppe’s. These
techniques iteratively shrink the edges while preserving the shape of the mesh. To evaluate
the accuracy, a test set of meshes ranging from simple geometric shapes to complex geome-
tries is used. A ground-truth model is created and error metrics, specifically the Hausdorff
distance, are used for evaluation. Network reconstruction with marching cubes is efficient
even with noise, although deformations increase with noise levels. Changing the sampling
distance affects detail and noise tolerance. Mesh simplification significantly reduces the
number of faces without a significant increase in errors. The maximum directional residual
is proposed as a stopping criterion for simplification. The overall results show a significant
reduction in the number of faces with a minimal increase in error.

Yang et al. [46] presents research on material estimation from segmented point clouds,
with a focus on material recognition, physical property determination, and tactile texture
simulation (see Figure 4). Initially, they tried to obtain color textures from segmented
point clouds and recognize materials from them using feature matching methods such as
SIFT. However, the method is inefficient due to UV distortions in the generated textures.
This necessitates the calculation of material properties for each point in the point cloud
using intensity information. The method relies heavily on point intensity, which may
not always be available, especially for indoor objects. That is why they also use a false
intensity simulated by grayscale values. The disadvantage is the recognition of different
materials with different colors. Yang et al. [46] also investigated the acquisition of physical
properties such as friction from the evaluated material properties. However, no existing
friction model based on material properties has been found, and friction coefficients are
usually measured empirically. The authors also explore the possibility of simulating tactile
textures based on the evaluated material properties. Although the method remains untested,
hypotheses suggest that users’ tactile sensations can be satisfied by aligning them with
visual perceptions of texture.

Baars et al. [21] explores the distance limitation facing the TI by implementing the
solution depicted in Figure 5. The focus is on creating a demonstration involving common
household objects on a table that can be interacted with by a remote robotic arm controlled
by a human. The feasibility of the solution is tested by assuming that an RGB-D camera captures the scene. Challenges include segmenting objects in the scene and handling missing information due to occlusion. The authors investigate object mass estimation from sensor data assuming object point cloud segmentation and no noise or missing information. The proposed method involves estimating the volume and density to calculate mass. Various approaches to estimating the volume from a point cloud are explored, with a preference for mesh creation due to its flexibility in downstream processes such as collision detection. The authors use the Point Cloud Library (PCL) for point cloud processing, including downsampling and surface reconstruction. The evaluation is performed on synthetic point clouds, comparing the volume estimation using an oriented bounding box (OBB). The results show that the surface approach generally outperforms the OBB approach, although with room for improvement, possibly by adjusting the algorithm parameters or adopting a machine learning approach.

The goal of [48] is to synthesize realistic 3D motion sequences involving the interaction of two hands with an object, where the mass of the object affects both its trajectory and the way the hands grasp it (see Figure 6). The method takes as input a scalar mass value, optionally an action label, and/or a manually drawn object trajectory. The result is a series of 3D movements represented as pairs of 3D hand and object poses. The authors use diffusion denoising models (DDMs) to synthesize 3D hand motion and object trajectory. These models have shown promising results in various tasks, including motion generation, due to their ability to generate high-quality and diverse motions without suffering from mode collapse. The authors describe the mathematical modeling and assumptions, and detail the hand motion synthesis network (HandDiff) and the trajectory synthesis algorithm (TrajDiff). The authors present 3D hands using a parametric hand model learned from large-scale 3D human scans. The pose of the object is represented by its 3D translation and rotation. The method synthesizes N consecutive 3D hand movements and optionally object poses, where global hand translations are relative to the center position of the object. To generate the object trajectory, we introduce TrajDiff, which synthesizes an object trajectory, a given mass value, and an action label. The trajectory is represented as the reference positions of the vertices on the object surface and the Procrustes alignment is used to obtain the object rotations. The method outperforms baseline methods in terms of physical plausibility, showing fewer collisions, lower collision distance, and fewer floating object artifacts, as well as stable synthesis even at unseen mass values.
Figure 6. Mass conditioned 3D hand and object motion synthesis approach [48].

The authors of [49] offer a virtual testbed tailored for model-mediated teleoperation or TI applications (see Figure 7). The authors introduce the concept of a virtual depth camera that replicates the capabilities of a real Kinect sensor using Unity. This virtual camera uses beam-casting techniques to collect depth information, allowing flexibility in adjusting the number of beams and the frame rate to suit different scenarios and devices. The point cloud is then tracked through the particle filter algorithm. Particle filtering, described as a method for estimating the dynamical states of a system from noisy or incomplete observations, overcomes the limitations of Kalman filtering, especially when dealing with nonlinear and non-Gaussian scenarios. The KLD-adaptive filter [49] dynamically adjusts the number of particles based on the approximation error and allows the tracking of different types of motion through parameter variations. The authors develop a network interface module to facilitate seamless communication between the tracking program and virtual or real devices. Using User Datagram Protocol (UDP) communication channels, this module [49] enables the transmission of point clouds and positional data, providing independent speed and efficiency between programs, similar to real-world TI applications. One of the challenges is packet loss due to tracking asynchrony.

Figure 7. A virtual platform for 3D object tracking [49].

In [10], the authors present a new objective metric, called the Tactile Internet Metric (TIM), designed to evaluate the performance of Tactile Internet (TI) sessions in real time (see Figure 8). The TIM is formulated by analyzing the components of a TI system at a detailed level and aims to provide quantifiable and reproducible measurements. TIM design goals include objectivity, evaluation based on short-term response, low complexity, easy setup, monotonic behavior, and real-time measurements. The metric takes into account both latency indicators and undetected signals to accurately capture TI system performance. In addition, theoretical modeling of tactile interaction is discussed and the use of a virtual channel compensation spring is proposed to mitigate the negative
effects of channel disturbances. In general, TIM can be used to evaluate and optimize the performance of TI systems in various use cases.

4.2. Limitations

The assessment of material performance in telecommunication AR, VR, and MR systems, and the Tactile Internet presents a myriad of challenges that require careful consideration. One of the biggest challenges is the quality and availability of data, which play a key role in the accuracy of attribute estimation. Bassetti et al. [50] highlight the problem of poor correlations between material properties, which complicates the task of deriving accurate estimates. Furthermore, Corsini et al. [1] highlight the sensitivity of these properties to small changes, emphasizing the need for precision in data collection and processing. Fleming [40] highlights the influence of noise in the data, which introduces uncertainty and reduces the reliability of estimates. In addition to data issues, the computational requirements for attribute estimation create significant obstacles. Davis [51] sheds light on the complex process of encoding and processing point clouds, which requires significant computational resources and efficient algorithms. Papastamatiou et al. [52] delved into the computational intensity of the equations involved in attribute estimation, highlighting the need for optimization and parallel processing to increase efficiency.

In addition to computational challenges, model limitations complicate the estimation process. Danaci and others [53] pay attention to model retraining, stressing the importance of model architecture and parameter tuning techniques. Farkas et al. [54] discuss the challenges of defining perceptual categories reflecting the complex nature of human perception and integrating it into attribute estimation models. Furthermore, Takahashi et al. [55] draw attention to the complexity of decoupling material and light parameters, highlighting the need for advanced modeling approaches.

Human perception is characterized by subjectivity and variability. Yuan et al. [36] and Luo et al. [56] highlight the subjective nature of human perception, which adds elements of variability to attribute estimation tasks. Fleming [40] highlights the challenges of generalizing across different scenes and materials, reflecting the inherent complexity of real-world scenarios. Optimization challenges, physical limitations, and experimental design issues further complicate the complexity of attribute assessment. Choi et al. [57], Nagai et al. [31], and Li et al. [26] address optimization challenges and experimental design considerations, emphasizing the importance of careful planning and execution. Additionally, limitations in training and evaluation, as well as limitations in sensor resolution and data collection, have hindered progress in attribute estimation [5,58–60]. Real-time estimation and network performance pose additional challenges, as highlighted by Kroep et al. [10] and Chen et al. [49], which necessitates efficient algorithms and optimized architectures. Ambiguities in trajectory and rendering quality further highlight the complexity of attribute estimation in dynamic environments [42,43,61]. In addressing these multifaceted challenges, researchers must adopt a holistic approach that includes data quality assurance, computational optimization, model refinement, and considerations of human perception and real-world complexity. Only through a concerted effort can the field of material performance evaluation in telecommunications systems truly advance and realize its full potential.
5. Conclusions

The assessment of material performance in telecommunication systems, including AR, VR, MR, and the Tactile Internet, presents a myriad of challenges that require careful consideration and innovative solutions, from data quality and availability issues to computational requirements, model limitations, complexity of human perception, optimization challenges, and real-time estimation limitations. Despite these challenges, the technological process continues to evolve. We explored various methods, from point cloud processing and feature matching to machine learning algorithms and simulation techniques, to improve the estimation of material attributes. Such advances have the potential to improve user experience, enable innovative applications, and lead to the evolution of telecommunications technology. The integration of the TI into telecommunication systems has enormous potential to improve the evaluation of material attributes, improve user experience, and stimulate innovation in various fields.

Author Contributions: Conceptualization, N.C. and K.T.; methodology, N.C. and K.T.; formal analysis, N.C. and K.T.; investigation, N.C.; resources, N.C.; data curation, N.C. and K.T.; writing—original draft preparation, N.C. and K.T.; writing—review and editing, N.C. and K.T.; visualization, N.C.; supervision, K.T.; project administration, K.T.; funding acquisition, K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financed by the European Union-Next Generation EU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0005: “Improving the research capacity and quality to achieve international recognition and resilience of TU-Sofia” (IDEAS).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

2D  Two-dimensional
3D  Three-dimensional
AE  AutoEncoder
AR  Augmented reality
BRDF  Bidirectional Reflectance Distribution Function
CNN  Convolutional neural network
DDM  Denoising diffusion model
FFT  Fast Fourier transform
GAN  Generative Adversarial Network
GMM  Gaussian Mixture Model
HandDiff  Hand motion synthesis network
KLD-adaptive filter  Kalman adaptive filter
MAE  Material attribute estimation
MC  Monte Carlo
MLP  Multilayer Perceptron
MR  Mixed reality
NeILF  Neural incident light field
NeRF  Neural radiance field
OBB  Oriented bounding box
PCL  Point Cloud Library
SDF  Signed-distance field
SG  Spherical Gaussian
SSD  Single Shot Multi-Box Detector
SVBRDF  Spatially Varying Bidirectional Reflectance Distribution Function
TI  Tactile Internet
TIM  Tactile Internet Metric


References


61. Han, X.; Wang, Q.; Wang, Y. Ball Tracking Based on Multiscale Feature Enhancement and Cooperative Trajectory Matching. Appl. Sci. 2024, 14, 1376.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.