

GENERATIVE AI FOR AUTOMATED MATERIALITY SYNTHESIS IN SPATIAL 3D MODELS: STATE-OF-THE-ART

Donata Šermukšnė¹, Filippo Sanfilippo², and Algirdas Noreika³

¹Department of Software Engineering, Kaunas University of Technology, Kaunas, Lithuania

²Department of Engineering Sciences, University of Agder, Grimstad, Norway

³Department of Multimedia Engineering, Kaunas University of Technology, Kaunas, Lithuania

KEYWORDS

Generative AI, generative technologies, 3D, material, texture, haptics.

ABSTRACT

This research investigates the state-of-the-art of generative AI for the automated synthesis of physically accurate materiality in spatial 3D models. By addressing two core research questions, the study reveals that generative algorithms can effectively synthesize a wide array of digital-visual, digital-tactile, and procedural material properties. These encompass standard Physically Based Rendering (PBR) maps and Spatially Varying Bidirectional Reflectance Distribution Function (SVBRDF) models, alongside perceptual tactile attributes like softness and friction, enabling multi-sensory experiences via haptic devices. The algorithmic landscape is analyzed through Generative Adversarial Networks, Diffusion Models, and Transformers, underscoring a paradigm shift toward multimodal architectures and resolution-independent procedural node graphs. Ultimately, this state-of-the-art findings demonstrate that the convergence of generative technologies may enable the creation of high-fidelity 3D materials, addressing the growing demand across industries such as gaming, robotics, film, and product design for more efficient, controllable, and automated 3D content generation techniques.

INTRODUCTION

The evolution of 3D graphics, from early innovations like Sketchpad (Sutherland, 1963) to modern real-time rendering, has revolutionized visual representation across industries such as gaming, film, and product design (Berto et al., 2016). This widespread integration has consequently driven an exponential demand for high-quality 3D assets. However, traditional manual content creation remains labour-intensive, costly, and time-consuming, necessitating urgent advancements in automated 3D design workflows to reduce production costs, accelerate development, and foster greater creative exploration.

This need for automation is especially crucial for complex robotic systems involved in human-robot collabora-

tion (HRC), physical human-robot collaboration (PHRC), and human-robot teaming (HRT) tasks (Langås et al., 2025). In these contexts, high-quality, physically accurate visual and tactile materials of 3D models provide simulated sensor data for advanced robot training, significantly improving the control algorithms of complex robotic systems. Furthermore, materials with realistic tactile properties, such as roughness, friction, and temperature, offer robots a precise sense of virtual touch. This capability allows robots to "learn" to interact physically with objects and enables humans to issue specific commands based on perceived tactile feedback, fostering more natural and efficient collaboration. Such advanced applications necessitate the development of generative AI models capable of autonomously producing realistic and diverse material properties for 3D spatial models.

Generative AI has significantly advanced content creation by offering novel approaches for producing complex and realistic 3D assets and enabling the efficient and rapid generation of physically accurate 3D materials (Tang et al., 2025). These generative AI methods greatly assist 3D artists, enhancing their capabilities. Nevertheless, current generative AI approaches encounter considerable challenges related to the limited availability of high-quality and diverse training data, computationally intensive optimization processes, and ensuring the integrity and physical accuracy of generated materials.

While there's broad interest in generative AI for 3D content, a specific gap exists in comprehensive reviews focusing on how AI automatically generates high-quality, physically accurate materials and its unique challenges. This review aims to bridge that gap by systematically examining generative AI algorithms for automated materiality generation in spatial 3D models by answering the main research questions (RQ):

- RQ1: What are the key physically-based 3D material properties that generative AI algorithms can effectively reproduce?
- RQ2: What are the current state-of-the-art approaches of generative AI algorithms employed for the synthesis of visual and/or tactile material properties?

Table 1: Thematic Keyword Clusters

Generative AI terminology	3D graphics terminology	Haptics terminology
generative technologies, generative AI, generative artificial intelligence, generative model*.	3D material*, physically-based, PBR, materiality, texture*, texture map*, texture inpainting, material*.	haptic*, tactile, touch, wearable*.

METHODOLOGY

The research process for this review began with the formulation of two central research questions, which directly target the core challenges and opportunities in applying generative AI to reproduce physically-based material properties (visual and tactile). The literature search for this review targeted scientific publications from 2021 to 2026, a time frame selected to capture the accelerated evolution of generative AI models. This search was conducted using the IEEE Xplore and Web of Science Core Collection digital databases, which are widely recognized for their comprehensive coverage of technology research and high-quality publications.

The search strategy for this review utilized thematic keyword clusters (see Table 1), organized into three categories: generative AI, 3D graphics, and haptics terminology. Keywords within each category were combined with logical "OR" operators for broad coverage, using wildcards for variations. The generative AI and 3D graphics categories were then combined with "AND", with the resulting query "OR"-ed to the haptics category, to identify studies integrating generative AI methods into 3D graphics, especially those advancing physically accurate materials and haptics technology.

Following the initial database searches, duplicates were first removed. Next, the remaining records underwent multi-stage filtering based on predefined inclusion and exclusion criteria. This process began by retaining only English-language papers published between 2021 and 2026, followed by restricting document types to "Article", "Early Access Articles", "Review Article", and "Proceedings Paper". Subsequently, quality filters were applied using automated text filtering tools, retaining only journal articles from Q1 or Q2 quartiles, which rank publications annually by citation impact within their field, and proceedings papers from CORE-ranked A* or A conferences, renowned for their low acceptance rates and high-impact papers.

Furthermore, titles and abstracts of the resulting records underwent manual review, excluding studies unrelated to generative AI applications in 3D material synthesis. Connected Papers was then used to identify additional relevant studies missed in the initial search, applying the same inclusion and exclusion criteria for consistency. Finally,

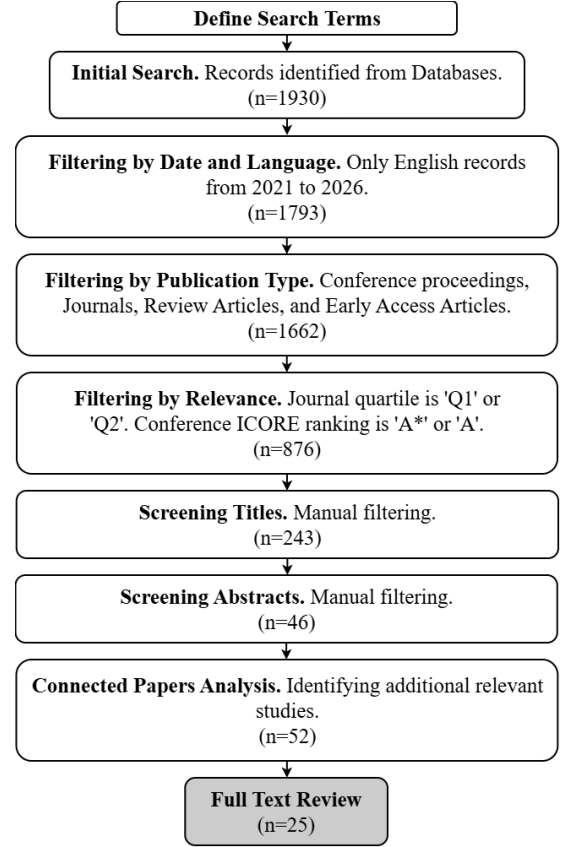


Figure 1: Flow Diagram Illustrating the Study Selection Process

full-text analysis of the remaining articles confirmed their relevance and eligibility for this review. Figure 1 provides a visual representation of the systematic review process, from initial search to final selection of 25 papers.

MATERIAL PROPERTIES

Materials are fundamental to how we perceive and interact with both the real and virtual worlds. Their properties can be broadly categorized into real-world physical attributes and their digital representations, which further break down into visual and tactile characteristics. Real-world materials possess a wide array of physical properties that govern their behavior and interaction with external forces and energy. These properties are what we intuitively sense and understand about objects through sight, touch, and even other senses. In the digital domain, these multifaceted properties must be accurately simulated to achieve realistic and immersive 3D environments.

Digital-Visual Materials

Digital-visual materials are commonly defined by a set of 2D texture maps that describe color, detail, and material properties—such as roughness, reflectivity, or transparency—without increasing geometric complexity of 3D models (see Table 2). These maps are integral to PBR workflows, enabling a realistic appearance under varying

Table 2: Most Common Types of 2D Texture Maps (Vecchio and Deschaintre, 2024)

Type	Description
Albedo Map	Represents the raw color values of the material, without any lighting or shading information.
Normal Map	Stores vector information that defines the orientation of surface normals. This simulates fine surface details like bumps and grooves.
Roughness Map	A grayscale map that defines how light scatters across a surface, controlling the sharpness or blurriness of reflections.
Metallic Map	A grayscale map that indicates whether a material is metallic or non-metallic/dielectric.
Ambient Occlusion Map	A grayscale map that simulates soft shadows in crevices, corners, and areas of the 3D model, where ambient light is obstructed.
Opacity Map	A grayscale map that controls the transparency of different parts of a material, allowing varying levels of visibility.
Diffuse Map	Represents the main colors or texture of an object and may include some baked-in lighting information.

lighting conditions (Vecchio et al., 2021; Syarawy et al., 2025). PBR workflows utilize standardized texture maps, such as albedo, metallic, roughness, and normal maps, to accurately represent how materials interact with light, ensuring visual consistency across diverse virtual environments (Syarawy et al., 2025).

These maps are often applied to form a SVBRDF representation, which is used to describe how light interacts with the surface and defines how light is reflected from every point on an object’s surface in every possible direction, taking into account that the material properties might change across the surface itself (Li et al., 2022). The primary purpose of SVBRDFs is to enable the realistic rendering of materials that have varying appearances, such as scratched metals, weathered wood, or fabrics with different weaves in different areas. Albedo, normal, roughness, metallic, ambient occlusion, opacity, and diffuse maps are the most common texture maps used to represent a SVBRDF material model. Some implementations simplify this to four core maps: normal, diffuse, roughness, and specular (Wang et al., 2023).

Digital-Tactile Materials

Tactile materials are characterized by properties that evoke a sense of touch. Simulating these properties is critical for creating immersive virtual and haptic experiences (Gao et al., 2023). Haptic technology plays a crucial role in bridging the gap between the virtual and physical worlds by allowing users to experience the sense of touch within

Table 3: Tactile Properties of Digital Materials (Gao et al., 2023; Naeem et al., 2025)

Tactile Property	Description
Friction	The resistance encountered when one surface slides over another.
Stickiness	The tactile sensation commonly related to the micro-stretching of the skin.
Smoothness / Roughness	The perception of surface texture ranging from very even to highly uneven.
Hardness	Resistance to deformation.
Temperature	The perceived warmth or coolness of a surface.

digital environments. Haptic devices can offer a multi-sensory experience, enabling users to perceive a range of tactile properties (see Table 3) such as friction, stickiness, smoothness, roughness, hardness, and temperature.

Among various haptic devices, haptic gloves are particularly significant for enabling intuitive and immersive interaction with digital materials. These devices provide vibrotactile feedback, force feedback, and thermal cues, allowing users to “feel” the virtual objects and their material properties (Zhang et al., 2024b). For instance, if a user touches a “rough” digital surface, the glove might generate specific vibrotactile patterns. If a “sticky” surface is encountered, the glove might modulate friction or apply slight resistance during movement (Gao et al., 2023).

Procedural Materials

Procedural materials offer a powerful, flexible, and algorithmic way to define material properties, distinguishing them from static texture maps. These materials are defined as sets of mathematical functions or “procedures” and are typically organized and visualized as node graphs (Hu et al., 2023). In these graphs, each node represents an operation or a function with specific inputs and outputs. Nodes have input slots for arguments and output slots for providing the result of the node’s operation. Edges in node graphs represent connections between nodes, directing the flow of data or computation from one node’s output to another node’s input. Many nodes also have internal parameters that can be adjusted to precisely control the behavior of the operation. This modular structure of procedural materials allows for the creation of complex and infinitely varied material properties, providing significant benefits in terms of flexibility, resolution independence, reduced storage requirements, and the ability for dynamic generation and real-time adjustments (Guerrero et al., 2022).

GENERATIVE AI ALGORITHMS

In recent years, generative AI approaches have significantly improved the automation of creating physically based material properties, surpassing traditional manual

workflows. This section explores prominent generative AI algorithms and architectural methods used for synthesizing materiality in spatial 3D models.

Diffusion Models

Diffusion models represent a state-of-the-art class of generative models that have achieved remarkable success in image generation tasks. These models operate by gradually adding noise to training data (a forward diffusion process) and then learning to reverse this process (a reverse denoising) to reconstruct original data or generate new samples. In the context of 3D materiality, diffusion models can offer exceptional control and fidelity in generating detailed and diverse material properties.

MatFuse utilizes diffusion processes to generate high-fidelity 3D materials from diverse conditioning sources such as text, sketches, and images (Vecchio et al., 2024). Specifically, MatFuse employs a multi-encoder VQ-GAN to learn a bidirectional mapping between pixel and latent spaces, where the latent space captures essential features and variations of materials in a lower-dimensional form. The model is then trained to generate these latent representations via diffusion (Vecchio et al., 2024).

DiffMat is another latent diffusion model, which utilizes latent diffusion within StyleGAN’s latent space to efficiently generate high-resolution SVBRDF textures, overcoming challenges related to data scarcity and high computational costs (Yuan et al., 2024). The model employs a CLIP encoder to extract features from input images as conditional inputs. In particular, its overall objective is to model a conditional distribution by progressively denoising a variable that initially follows a standard normal distribution through a fixed Markov chain, in which noise is recursively added to an input. After the latent vectors are generated, they are converted into final textures via the PhotoMat generator (Yuan et al., 2024).

MatFusion formulates SVBRDF estimation as a diffusion task, utilizing an unconditional backbone diffusion model to synthesize SVBRDF textures. Following its initial training, this unconditional model can be further refined into a conditional diffusion model, enabling the estimation of material properties directly from photographs under both controlled and uncontrolled lighting conditions (Sartor and Peers, 2023). This methodology enables integration with advanced rendering for refinement without requiring backpropagation during training (Sartor and Peers, 2023).

DreamPBR is a text-driven framework for generating high-resolution, tileable SVBRDFs using a material Latent Diffusion Model (m-LDM) and a rendering-aware PBR decoder (Xin et al., 2025). By utilizing a denoising U-Net architecture to map albedo maps into a latent space and reverse the diffusion process, the model successfully integrates pretrained diffusion knowledge with material priors to produce diverse and seamless material maps (Xin et al., 2025).

The Naeem et al. (2025) use a sequential pipeline for multimodal visual and haptic texture synthesis. The pro-

cess begins by employing a Stable Diffusion model to generate high-quality visual textures from user text prompts. Furthermore, these images are processed by AttributeNet, a deep learning architecture that predicts tactile attributes such as roughness and softness (Naeem et al., 2025). Finally, an interpolation-based texture authoring algorithm synthesizes vibrotactile signals from these attributes, leveraging a pre-built texture library to enable perceptually accurate haptic feedback (Naeem et al., 2025).

Other notable diffusion-based generative AI methods have been adapted for textured 3D object generation tasks. For instance, Ouroboros3D (Wen et al., 2025) employs 3D-aware recursive diffusion to produce high-fidelity 3D assets from input images, while Sharp-It (Edelstein et al., 2024) is a multi-view to multi-view diffusion model enabling versatile 3D synthesis and manipulation. Similarly, HexaGen3D (Mercier et al., 2025) adapts Stable Diffusion for rapid and diverse text-to-3D generation.

Generative Adversarial Networks

Generative Adversarial Networks (GANs) consist of two competing neural networks: a generator that produces synthetic data and a discriminator that evaluates its authenticity. This adversarial training process pushes the generator to create increasingly realistic and convincing material outputs. While effective, GANs are constrained by their reliance on pre-existing material inputs and often lack the advanced generative capabilities seen in diffusion models (Vecchio et al., 2024).

SurfaceNet is a patch-based GAN designed to estimate the properties of SVBRDF from a single image (Vecchio et al., 2021). The encoder of the generator utilizes a ResNet-101 with Atrous Spatial Pyramid Pooling and an upsampling cascade with skip connections to capture multi-scale features and recover fine details (Vecchio et al., 2021). By employing a patch discriminator, this adversarial approach minimizes domain shift between synthetic and real data distributions while refining features in the reconstructed material maps (Vecchio et al., 2021).

TileGen, based on StyleGAN2, is designed for the controllable synthesis of tileable SVBRDFs through architectural modifications that ensure seamless texture maps (Zhou et al., 2022). The TileGen model allows for the precise manipulation of materials by integrating latent style codes with optional structural patterns to define the placement of features (Zhou et al., 2022). High fidelity and realism are achieved by employing an adversarial loss and a discriminator network during the training process (Zhou et al., 2022).

The Gao et al. (2023) uses a conditional GAN to generate visual and tactile material properties from a single sketch input. The framework features a U-Net generator that branches into separate visual and tactile generators to share global structural encoding while maintaining modality-specific details (Gao et al., 2023). By utilizing multi-scale PatchGAN discriminators to enhance fine details, this controllable approach effectively bridges the gap

between abstract user inputs and rendered experiences via haptic devices (Gao et al., 2023).

Additionally, several other prominent generative AI methods based on GANs have been employed for synthesizing textured 3D objects. For instance, the work in Pavllo et al. (2021) proposes learning generative models of textured triangle meshes from real-world images using GANs trained with 2D supervision, enabling disentangled control over shape, appearance, and pose without requiring expensive key point annotations. Similarly, Dündar et al. (2022) develops a GAN pipeline with learnable position encodings and attention mechanisms to synthesize high-quality textured 3D models from multi-view images, producing coherent geometry and realistic fine-detailed textures. Zheng et al. (2024) further advances this with SDF-based 3D GANs for joint image and geometry generation.

Transformers

Transformers, originally developed for natural language processing, have found applications in material generation. Their strength lies in their ability to effectively model long-range dependencies and complex sequential data. In material generation, transformers model structural relationships and coherently synthesize visual and tactile properties.

MatFormer employs a transformer-based architecture to automatically generate procedural material graphs, which can be further adjusted by manipulating their parameters. This involves modeling a probability distribution over procedural materials and generating them in a sequential manner, handling the sparse and long-distance connections within material graphs (Guerrero et al., 2022). Building on this, the Li et al. (2024) approach uses three distinct transformers to generate sequences for procedural materials.

3DTopia-XL utilizes a 28-layer transformer architecture that integrates cross-attention layers for conditional signals, self-attention layers for modeling primitive correlations, and adaptive layer normalization for injecting timestep conditions (Chen et al., 2025). This model leverages a latent primitive diffusion process to iteratively refine geometric and textural attributes, employing conditional sampling to ensure consistency with input signals (Chen et al., 2025).

The Hu et al. (2023) work builds on the MatFormer approach by using three transformer-based models to sequentially generate nodes, edges, and parameters for material graphs. A distinguishing feature is its support for multi-modal conditioning via CLIP embeddings, allowing the generative process to be guided by text or image prompts to create diverse, user-specified material structures (Hu et al., 2023).

CLAY is a 1.5-billion-parameter generative model that utilizes an advanced latent Diffusion Transformer to synthesize high-quality 3D geometry and materials (Zhang et al., 2024a). This framework enables precise, controllable generation from diverse inputs, including text

prompts, images, and 3D primitives (Zhang et al., 2024a).

Transformers enable the synthesis of complex material structures by capturing long-distance dependencies between components with equal fidelity (Guerrero et al., 2022). This approach, as demonstrated by MatFormer, produces coherent and realistic material properties while overcoming the consistency challenges inherent in traditional generation methods (Guerrero et al., 2022).

PRACTICAL IMPACT IN APPLICATION FIELDS

Generative AI methods for the automated generation of material properties in 3D assets have progressed from conceptual prototypes to widespread adoption in industrial contexts. These approaches target domains where creating high-fidelity 3D environments has traditionally required extensive resources due to reliance on labor-intensive manual workflows.

In the gaming and film industries, generative AI models such as CLAY enable creators to produce diverse, high-fidelity 3D assets ideal for games and virtual simulations (Zhang et al., 2024a). These models can be conditioned on various inputs—including text, hand-drawn sketches, images, real-world photos, or bounding boxes—to generate complex geometries like buildings or vehicles. Moreover, they output PBR textures alongside textured 3D meshes that are directly compatible with popular rendering engines such as Unreal Engine, Unity, and Blender (Chen et al., 2025; Zhang et al., 2024a). This capability facilitates the rapid creation of diverse virtual environments for the metaverse and interactive applications.

In robotics, generative AI plays a pivotal role in synthesizing large-scale visual and tactile datasets essential for training autonomous systems in simulated environments. Since collecting real-world tactile data is labor-intensive and time-consuming, these models reduce reliance on physical sensors by generating haptic metadata and vibrotactile signals (Naeem et al., 2025). Moreover, generative AI-powered platforms enable the creation of interactive 3D scenes with precise annotations, tailored for robot training across diverse industrial scenarios. Particularly in robotic manipulation, accurate modeling of material properties is crucial for effective sim-to-real transfer (Zhang et al., 2025).

ANSWERS TO RESEARCH QUESTIONS

This section addresses the research questions outlined in this paper, which focus on systematically evaluating the current state-of-the-art for material generation, identifying gaps, and providing future research directions to continue the advancement of generating realistic and interactive 3D content. Table 4 provides an overview of the studies included in this systematic review, highlighting key characteristics of generative models for materiality synthesis.

RQ1: What are the key physically-based 3D material properties that generative AI algorithms can effectively reproduce? Answer: Generative AI effectively reproduces

Table 4: Main Characteristics of the Reviewed Studies

Reference	Core Architecture	Primary Inputs	Target Outputs	Technical Contribution	Evaluation Metrics
Naeem et al. (2025)	Stable Diffusion v2.1.	Text.	Material image and tactile properties.	Integrates a regression-based deep learning architecture "AttributeNet" to extract tactile properties from diffusion-generated visual textures.	MAE on roughness and softness.
Xin et al. (2025)	Material Latent Diffusion (m-LDM).	Text, binary, and RGB images, geometry.	SVBRDF maps.	Integrates large-scale VLMs trained on text-image pairs with material priors from PBR samples. Employs the m-LDM model to map albedo maps to latent space and decode SVBRDF maps.	CLIP, LPIPS, and RMSE on individual textures.
Vecchio et al. (2024)	Latent Diffusion.	Text, binary, and RGB images.	SVBRDF maps.	Integrates a VQGAN multi-encoder to learn bidirectional mapping and a diffusion model to generate a latent representation of a material.	CLIP-IQA, FID, and RMSE on individual textures.
Yuan et al. (2024)	Latent Diffusion.	RGB images.	SVBRDF maps.	Integrates CLIP encoder and a multi-layer, cross-attention denoising backbone to generate latent materials and converts them into texture maps with a pre-trained StyleGAN-based generator.	FID, and inference time.
Sartor and Peers (2023)	ConvNeXt-based k-Diffusion.	RGB images.	SVBRDF maps.	Introduces an unconditional backbone k-diffusion model to synthesize texture maps and leverages ConvNeXt blocks to increase the number of activations without increasing the parameter count.	LPIPS, RMSE on individual textures, real-world validation.
Wen et al. (2025)	Recursive Video Diffusion.	RGB images.	Multiview images, 3D model, diffuse texture.	Proposes an image-to-3D framework that integrates multi-view generation with 3D reconstruction within a recursive diffusion process. Introduces a 3D-aware feedback loop to iteratively refine 2D views.	LPIPS, PSNR, SSIM, training time, inference time.
Mercier et al. (2025)	Latent Diffusion (SD-v1.5 / SD-XL).	Text.	3D mesh with diffuse texture.	Fine-tunes a pretrained text-to-image model to predict 6 orthographic projections and the corresponding 3D latent, then decodes the latent to generate a textured mesh.	CLIP, FID, CD-L1, rFID, and latency.
Edelstein et al. (2024)	Multiview to Multiview Diffusion.	Text, low-quality multiview set.	High-quality refined multiview images.	Modifies the Zero123++ architecture to take an existing 3D consistent multi-view images and enhance geometric details, and texture.	FID, CLIP, DINO, and runtime.
Vecchio et al. (2021)	GAN.	RGB images.	SVBRDF maps.	Leverages GAN properties for unsupervised texture generation and employs a patch-based discriminator to recover high-frequency details.	RMSE on individual textures, qualitative comparison.
Zhou et al. (2022)	StyleGAN2 variant.	Binary and RGB images.	Tileable SVBRDF maps	Ensures tileable outputs and enables greater control over generated materials by optionally conditioning on input structure patterns that specify the locations of dominant features.	Visual variety sampling, material capture optimization, qualitative layout consistency.
Gao et al. (2023)	GAN.	Text, binary images.	RGB image, 2-channel tactile image.	Proposes GAN model to generate visual and tactile outputs from a single sketch image, along with a pipeline to render tactile signals on haptic devices.	LPIPS, SIFID, human preference study.
Pavlo et al. (2021)	2D Convolutional GAN.	RGB images.	Textured 3D mesh, semantics.	Learns 3D mesh generation without keypoint annotations using a sphere template deformed by 2D displacement maps in UV space.	GDs, FIDs, pose estimation performance.
Dündar et al. (2022)	GAN.	Multi-view or Single-view images.	High-quality textured 3D mesh.	Proposes a progressive training strategy (geometry first, then texture), incorporating attention mechanisms and learnable position encodings.	FID, qualitative 3D, and texture visuals.
Zhang et al. (2024b)	DSS and SRResNet-based GAN.	Latent vector with a 1-D slider	Spectrogram.	Generates class-controllable, realistic, and distinguishable haptic vibrations with conditional information and encoded features.	Subjective user evaluation, classification accuracy.
Zheng et al. (2024)	3D GAN.	Noise vector, 2D images, depth maps.	Textured 3D mesh.	Utilizes a hybrid representation model for the generation of detailed features and efficient training under the supervision of low-resolution 2D images.	MMD-CD, COV-CD, and FID.
Guerrero et al. (2022)	Transformer-based.	Partial material graph.	Procedural material graph.	Introduces an autoregressive model that leverages a transformer-based architecture to model a probability distribution over the space of procedural materials.	Graph statistics, FID, normalized nearest neighbor style, and edit distance.
Li et al. (2024)	RL-trained Transformer.	RGB images.	Procedural material graph parameters.	Pre-trains a parameter generator using supervised training, and then fine-tunes it with reinforcement learning.	Reward, LPIPS, visual realism.
Chen et al. (2025)	Diffusion Transformer-based (DiT).	Text, RGB images.	Textured 3D mesh.	Primitive Patch Compression and Latent Primitive Diffusion leveraged with DiTs to model global correlation of latent primitive tokens for generative modeling.	Runtime, approximation error, CD, PSNR, user study.
Hu et al. (2023)	Transformer-based.	Text, RGB images, partial node graphs.	Procedural material graphs.	Transformer model conditioned by a CLIP embedding mapped by learnable MLPs for material graphs generation.	Style loss, conditioning visual evaluation.
Zhang et al. (2024a)	Variational Autoencoder (VAE) and latent DiT.	Text, RGB images, voxels, bounding boxes, point clouds.	Textured 3D mesh.	Extracts rich 3D priors directly from a diverse range of 3D geometries, depicts continuous and complete surfaces along with a tailored multi-resolution geometry VAE.	Render-FID, render-KID, P-FID, P-KID, CLIP, ULIP-T/I, CD, EMD, Voxel-IoU, F-score.
Syarawy et al. (2025)	Stable Diffusion XL.	Text, RGB images, box or point prompts.	PBR maps.	Implements a segmentation-guided texture transfer framework to apply context-specific textures to targeted mesh areas seamlessly.	CLIP, LPIPS, and ResNet50 cosine similarity.
Vecchio and Deschaintre (2024)	None (creates a new dataset).	Materials from online sources (CC0, CC-BY license).	Tileable 4K PBR maps.	Provides a new material dataset with 4069 unique and 683592 augmented materials with metadata and 3417960 renderings of the augmented materials.	RMSE, SSIM, LPIPS, FID for Deschaintre et al. (2018), SurfaceNet, and MatFuse methods.

Li et al. (2022)	Deep Inverse Rendering.	Two RGB images with flash on and off.	SVBRDF maps.	A reconstruction pipeline that combines deep inverse rendering and reflectance sample transport to reconstruct SVBRDF maps, incorporating a jump rearrangement approach to reduce artifacts.	RMSE, LPIPS, GPU consumption, qualitative visual evaluation.
Wang et al. (2023)	Two-Level Basis Joint Prediction Network.	Hand-held captured image.	SVBRDF maps.	Introduces a two-level basis material model that leverages data priors without ground-truth material weights while optimizing lighting direction estimation.	RMSE, LPIPS, angular error, qualitative visual evaluation.
Tchapmi et al. (2022)	CNN Encoder, MLP branches.	Single RGB image.	Procedural material category and parameters.	Proposes architecture of a CNN encoder and a single-layer perceptron modules for procedural material category and parameter estimation.	Classification accuracy, average parameter regression error.

3D material properties across visual, tactile, and procedural domains. It synthesizes standard PBR texture maps, such as Albedo, Normal, Roughness, Metallic, Ambient Occlusion, Opacity, Height, and Specular, which are often integrated into SVBRDF models to ensure realistic surface rendering (Li et al., 2022; Wang et al., 2023). Beyond visuals, newer frameworks generate tactile data such as softness or friction, translating them into vibrotactile signals for immersive experiences via haptic devices (Gao et al., 2023; Naeem et al., 2025; Zhang et al., 2024b). Finally, these algorithms recreate procedural materials by generating underlying node graphs instead of static pixels (Hu et al., 2023; Guerrero et al., 2022), allowing for resolution-independent, fully editable 3D materials through automated parameter prediction (Hu et al., 2023; Tchapmi et al., 2022).

RQ2: What are the current state-of-the-art approaches of generative AI algorithms employed for the synthesis of visual and/or tactile material properties? Answer: The current state-of-the-art for synthesizing visual and tactile material properties is primarily driven by Diffusion Models, GANs, and Transformers frameworks that are increasingly moving toward multi-modal architectures capable of simultaneously generating high-fidelity texture maps and corresponding tactile properties. Diffusion models represent the pinnacle of high-fidelity material synthesis due to their ability to model complex data distributions without the stability issues of earlier models (Vecchio et al., 2024; Yuan et al., 2024). GANs remain highly effective for refining fine-grained surface details and ensuring structural consistency. Finally, Transformers are the state-of-the-art choice for procedural material generation, where models like MatFormer and the work by Hu et al. (2023) use autoregressive transformers to generate node sequences, edges, and parameters for procedural material graphs (Hu et al., 2023; Guerrero et al., 2022).

CONCLUSIONS

This systematic review demonstrates that generative AI has reached a pivotal stage in the automated synthesis of 3D materiality, successfully bridging the gap between abstract user inputs and physically-grounded digital representations. The research identifies that current algorithms can effectively reproduce a comprehensive set of material properties, ranging from Physically Based Rendering texture maps (albedo, normal, roughness, etc.) (Syarawy

et al., 2025; Vecchio and Deschaintre, 2024) for complex Spatially Varying Bidirectional Reflectance Distribution Function models (Li et al., 2022; Wang et al., 2023) to immersive tactile properties such as friction and hardness (Gao et al., 2023; Naeem et al., 2025). Furthermore, the emergence of procedural material generation marks a significant shift toward resolution-independent and fully editable assets, moving beyond the limitations of pixel-based textures (Hu et al., 2023; Guerrero et al., 2022).

The algorithmic landscape is currently dominated by a transition from Generative Adversarial Networks, which remain valued for high-frequency detail recovery (Vecchio et al., 2021; Zhou et al., 2022), toward Diffusion Models and Transformers. Diffusion frameworks like MatFuse and DreamPBR now represent the state-of-the-art for high-fidelity, controllable synthesis from multimodal inputs (Vecchio et al., 2024; Xin et al., 2025). Simultaneously, transformer-based architectures such as MatFormer, 3DTopia-XL, and CLAY have proven uniquely capable of modeling the long-range dependencies required for procedural logic and large-scale 3D asset generation (Guerrero et al., 2022; Chen et al., 2025; Zhang et al., 2024a). This convergence of architectures is increasingly facilitating multi-sensory synthesis, where visual and tactile properties are generated in a single, coherent pipeline (Gao et al., 2023; Naeem et al., 2025).

Looking forward, the integration of these physically accurate materials is essential for advancing human-robot collaboration and immersive virtual environments (Langås et al., 2025). While significant progress has been made, challenges remain regarding the limited availability of diverse, high-quality training data and the computational cost of high-resolution optimization. Future research should focus on refining these multimodal models to ensure even greater physical integrity and exploring self-supervised learning techniques to mitigate data scarcity, ultimately fostering a more seamless interaction between humans, robots, and digital 3D worlds (Langås et al., 2025; Tang et al., 2025).

References

- Filippo Berto, Relly Victoria Petrescu, and Florian Ion Tiberiu Petrescu. A review of recent results on 3d effects, 04 2016. ISSN 1941-7020, 1941-7039.
- Zhaoxi Chen, Jiayang Tang, Yu Dong, Ziang Cao, Fangzhou Hong, Yushi Lan, Tengfei Wang,

- Hongxia Xie, Tong Wu, Shunsuke Saito, Liang Pan, Dahua Lin, and Ziwei Liu. 3dtopia-xl: Scaling high-quality 3d asset generation via primitive diffusion. pages 26576–26586, 06 2025. doi: 10.1109/cvpr52734.2025.02475.
- Valentin Deschaintre, Miika Aittala, Fr’edo Durand, George Drettakis, and Adrien Bousseau. Single-image svbrdf capture with a rendering-aware deep network. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, 37(128):15, aug 2018. URL <http://www-sop.inria.fr/revues/Basilic/2018/DADDB18>.
- Ayşegül Dünder, Jun Gao, Andrew Tao, and Bryan Catanzaro. Fine detailed texture learning for 3d meshes with generative models. *arXiv (Cornell University)*, 03 2022. doi: 10.48550/arxiv.2203.09362.
- Yiftach Edelstein, Or Patashnik, Dana Cohen-Bar, and Lihi Zelnik-Manor. Sharp-it: A multi-view to multi-view diffusion model for 3d synthesis and manipulation. *arXiv (Cornell University)*, 12 2024. doi: 10.48550/arxiv.2412.02631.
- Ruihan Gao, Wenzhen Yuan, and Jun-Yan Zhu. Controllable visual-tactile synthesis. pages 7017–7029, 10 2023. doi: 10.1109/icc51070.2023.00648.
- Paul Guerrero, Miloš Hašan, Kalyan Sunkavalli, Radomír Měch, Tamy Boubekur, and Niloy J. Mitra. Matformer. 07 2022. ISSN 0730-0301, 1557-7368. doi: 10.1145/3528223.3530173.
- Yiwei Hu, Paul Guerrero, Miloš Hašan, Holly Rushmeier, and Valentin Deschaintre. Generating procedural materials from text or image prompts. 07 2023. doi: 10.1145/3588432.3591520.
- Even Falkenberg Langås, Muhammad Hamza Zafar, and Filippo Sanfilippo. Exploring the synergy of human-robot teaming, digital twins, and machine learning in industry 5.0: a step towards sustainable manufacturing. *Journal of Intelligent Manufacturing*, 37: 999–1022, 02 2025. ISSN 0956-5515, 1572-8145. doi: 10.1007/s10845-025-02580-x.
- Beichen Li, Yiwei Hu, Paul Guerrero, Miloš Hašan, Liang Shi, Valentin Deschaintre, and Wojciech Matusik. Procedural material generation with reinforcement learning. 11 2024. ISSN 0730-0301, 1557-7368. doi: 10.1145/3687979.
- Zhiqiang Li, Xukun Shen, Yong Hu, and Xueyang Zhou. High-resolution svbrdf estimation based on deep inverse rendering from two-shot images. *The Visual Computer*, 39:4609–4622, 08 2022. ISSN 0178-2789, 1432-2315. doi: 10.1007/s00371-022-02612-0.
- Antoine Mercier, Ramin Nakhli, Mahesh Reddy, Rajeev Yasarla, Hong Cai, Fatih Porikli, and Guillaume O. Berger. Hexagen3d: Stablediffusion is one step away from fast and diverse text-to-3d generation. pages 1247–1257, 02 2025. doi: 10.1109/wacv61041.2025.00129.
- M. Asif Naeem, Josiah Carberry, and Seokhee Jeon. Text-driven generative framework for multimodal visual and haptic texture synthesis. 07 2025. doi: 10.1109/whc64065.2025.11123227.
- Dario Pavlo, Jonas Köhler, Thomas Hofmann, and Aurélien Lucchi. Learning generative models of textured 3d meshes from real-world images. pages 13859–13869, 10 2021. doi: 10.1109/iccv48922.2021.01362.
- Sam Sartor and Pieter Peers. Matfusion: A generative diffusion model for svbrdf capture. pages 1–10, 12 2023. doi: 10.1145/3610548.3618194.
- Ivan E. Sutherland. Sketchpad. 01 1963. doi: 10.1145/1461551.1461591.
- Mutawally Syarawy, Rio Mukhtarom Paweroi, and Mario Köppen. From image to immersion: Ai-assisted texture transfer for 3d environment creation in the metaverse. In *2025 3rd international conference on intelligent metaverse technologies & applications (iMETA)*, pages 1–8, 01 2025. doi: 10.1109/iMETA66706.2025.11306452.
- X. Y. Tang, R. Li, and Xiaopeng Fan. Recent advances in 3d object and scene generation: A survey. *arXiv (Cornell University)*, 04 2025. doi: 10.48550/arxiv.2504.11734.
- Lyne P. Tchapmi, Trishiet Ray, Micael Tchapmi, Bokui Shen, Roberto Martín-Martín, and Silvio Savarese. Generating procedural 3d materials from images using neural networks. pages 32–40, 03 2022. doi: 10.1145/3531232.3531237.
- Giuseppe Vecchio and Valentin Deschaintre. Matsynth: A modern pbr materials dataset. pages 22109–22118, 06 2024. doi: 10.1109/cvpr52733.2024.02087.
- Giuseppe Vecchio, Simone Palazzo, and Concetto Spampinato. Surfacenet: Adversarial svbrdf estimation from a single image. pages 12820–12828, 10 2021. doi: 10.1109/iccv48922.2021.01260.
- Giuseppe Vecchio, Renato Sortino, Simone Palazzo, and Concetto Spampinato. Matfuse: Controllable material generation with diffusion models. pages 4429–4438, 06 2024. doi: 10.1109/cvpr52733.2024.00424.
- Li Wang, L Zhang, Fangzhou Gao, and Jiawan Zhang. Deepbasis: Hand-held single-image svbrdf capture via two-level basis material model. pages 1–11, 12 2023. doi: 10.1145/3610548.3618239.

Hao Wen, Zehuan Huang, Yaohui Wang, Xinyuan Chen, and Sheng Lu. Ouroboros3d: Image-to-3d generation via 3d-aware recursive diffusion. pages 21631–21641, 06 2025. doi: 10.1109/cvpr52734.2025.02015.

Linxuan Xin, Zheng Zhang, Zhiyi Pan, Jinfu Wei, Duan Gao, and Wei Gao. Dreampbr: Text-driven high-resolution svbrdf generation with multimodal guidance. 06 2025. doi: 10.1109/icme59968.2025.11210202.

Liang Yuan, Dingkun Yan, Suguru Saito, and Issei Fujishiro. Diffmat: Latent diffusion models for image-guided material generation. *Visual Informatics*, 8: 6–14, 01 2024. ISSN 2468-502X, 2543-2656. doi: 10.1016/j.visinf.2023.12.001.

Kun Zhang, Peng Yun, Jun Cen, Junhao Cai, Didi Zhu, Hangjie Yuan, Chao Zhao, Tao Feng, Michael Yu Wang, Qifeng Chen, Jia Pan, Wei Zhang, Bo Yang, and Hua Chen. Generative artificial intelligence in robotic manipulation: A survey. 2025. doi: 10.48550/arxiv.2503.03464.

Longwen Zhang, Ziyu Wang, Qixuan Zhang, Qiwei Qiu, Anqi Pang, Haoran Jiang, Wei Yang, Lan Xu, and Jingyi Yu. Clay: A controllable large-scale generative model for creating high-quality 3d assets. 07 2024a. ISSN 0730-0301, 1557-7368. doi: 10.1145/3658146.

Mingxin Zhang, Shun Terui, Yasutoshi Makino, and Hiroyuki Shinoda. Textsensegan: A user-guided system for optimizing texture-related vibrotactile feedback using generative adversarial network. *arXiv (Cornell University)*, 07 2024b. doi: 10.48550/arxiv.2407.11467.

Han Zheng, Rong Xiong, Yue Wang, and Jun Wu. Sdf-based 3d generative adversarial networks for image and geometry generation. pages 8576–8581, 07 2024. doi: 10.23919/ccc63176.2024.10662644.

Xilong Zhou, Miloš Hašan, Valentin Deschaintre, Paul Guerrero, Kalyan Sunkavalli, and Nima Khademi Kalantari. Tilegen: Tileable, controllable material generation and capture. pages 1–9, 11 2022. doi: 10.1145/3550469.3555403.

assistant at the KTU Faculty of Informatics. Her research focuses on artificial intelligence, automation, virtual reality, 3D graphics, and multimodal human-computer interaction. Her work explores the intersection of these domains to create more intuitive and engaging user experiences in digital environments.



FILIPPO SANFILIPPO holds a PhD in Engineering Cybernetics from the Norwegian University of Science and Technology (NTNU), Norway. His research focuses on Human-Robot Teaming (HRT). He is a Professor at the Faculty of Engineering and Science, University of Agder (UiA), Norway, and the Head of the Artificial Intelligence, Biomechatronics, and Collaborative Robotics research group. He is an IEEE Senior Member, former Chair of the IEEE Norway Section, and currently Chair of the IEEE Robotics and Automation, Control Systems, and Intelligent Transportation Systems Joint Chapter. He also leads the Norway Section Life Members Affinity Group and serves on several IEEE Region 8 committees, including Chapter Coordination, Conference Coordination Committee, Public Visibility, Awards & Recognitions, and Professional & Educational Activities. He has published numerous technical papers in leading journals and conferences and actively serves as a reviewer for international publications.



ALGIRDAS NOREIKA is a professor-practitioner with a PhD in Informatics Engineering and Technology Science from Kaunas University of Technology (KTU), Lithuania. His doctoral research focused on methods of information technologies, artificial intelligence, and optimization methods. He is the head of the IT and digital technologies company “Indeform”, which is made up of highly skilled professional teams focusing on hi-tech IT, interactive solutions, and digital content production. He has more than 25 years of experience in advanced IT engineering, specializing in business process automation, AI integrations, SaaS platforms, real-time interactive systems, and high-fidelity 3D graphics and animations, using cutting-edge technologies.

AUTHOR BIOGRAPHIES



DONATA ŠERMUKŠNĖ is a PhD student in Informatics Engineering at the Kaunas University of Technology (KTU) in Lithuania. She is a software engineer at the KTU Faculty of Informatics Laboratory Center, developing technological solutions from immersive educational VR games to innovative augmented reality applications that tackle complex real-world problems. She is a researcher and a junior