Text2Texture: Generating 3D-Printed Models with Textures based on Text and Image Prompts

Joshua Yin
University of Washington
Seattle, Washington, USA
joshjyin@uw.edu

Faraz Faruqi
MIT CSAIL
Cambridge, Massachusetts, USA
ffaruqi@mit.edu

Martin Nisser
University of Washington
Seattle, Washington, USA
nisser@uw.edu

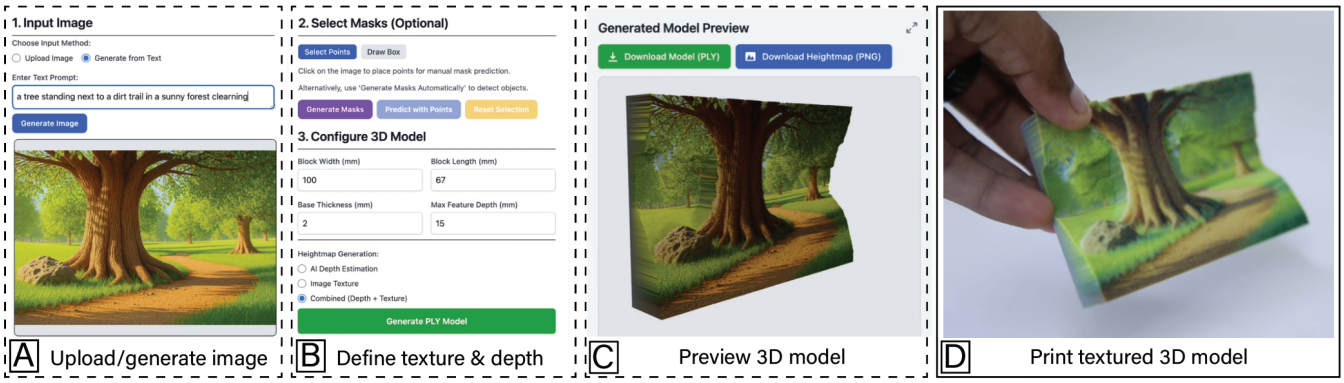

Figure 1: Text2Texture generates an image from a text prompt and converts this into a textured 3D object ready for 3D printing. (A) Users enter a text prompt in the web tool or upload an image directly. (B) Users can [2] generate or draw labeled masks such as "tree" which synthesizes a texture applied locally to the mask. [3] Users then select scaling parameters that a depth estimator uses to extrude the 2D image into a 2.5D surface. (C) Text2Texture then superimposes the texture and depth map to render a 3D-printable model in the viewport. (D) The 3D printed model captures both macroscopic depth and locally applied textures.

Abstract

To support users' understanding of physical properties in 2D images, we propose Text2Texture, a webtool that converts 2D color images into textured 3D objects ready for 3D printing. This is achieved by extracting depth information using a monocular estimator, extracting local texture information using a fine-tuned stable diffusion model, and superimposing these macro- and micro-scale geometries to produce a composite 3D model with color, depth and texture. Images can be uploaded directly or generated via text prompt, and we print a variety of objects generated using each approach to suggest applications in physicallizing virtual worlds, adding haptic cues to photographs, and conveying information about scale in images.

CCS Concepts

• **Human-centered computing → Empirical studies in collaborative and social computing**.

Keywords

Generative AI, Digital fabrication

UIST Adjunct '25, Busan, Republic of Korea

ACM ISBN 979-8-4007-2036-9/25/09
https://doi.org/10.1145/3746058.3758373

ACM Reference Format:
Joshua Yin, Faraz Faruqi, and Martin Nisser. 2025. Text2Texture: Generating 3D-Printed Models with Textures based on Text and Image Prompts. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '25), September 28–October 01, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3746058.3758373

1 Introduction

Images are a primary medium through which people experience the world, from sharing photographs to consuming visual art. While images capture visual information about the physical world, they do not transmit critical information about the size, texture, and materials of an image's contents in the way that tactile information can through touch. However, 3D printing presents the ability to transmit the physical information implicit in images by converting these to physical objects with tactile or haptic cues [4, 6, 12].

While recent tools have advanced specific aspects of translating visual inputs into physical form, they extract individual physical properties from images. For example, depth estimation techniques [2, 3] can extract macroscale depth from a single image, capturing global shape and spatial layout, but does not capture local micro-features for texture. Separately, tools such as Gelsight [9] enables the extraction of microgeometry details for surface textures, producing 3D-printable tactile models. Recent methods [5] enable prediction of these micro-geometry features, but no existing system integrates macroscale geometry, microscale texture, and color appearance into a unified 3D model. Our system bridges this

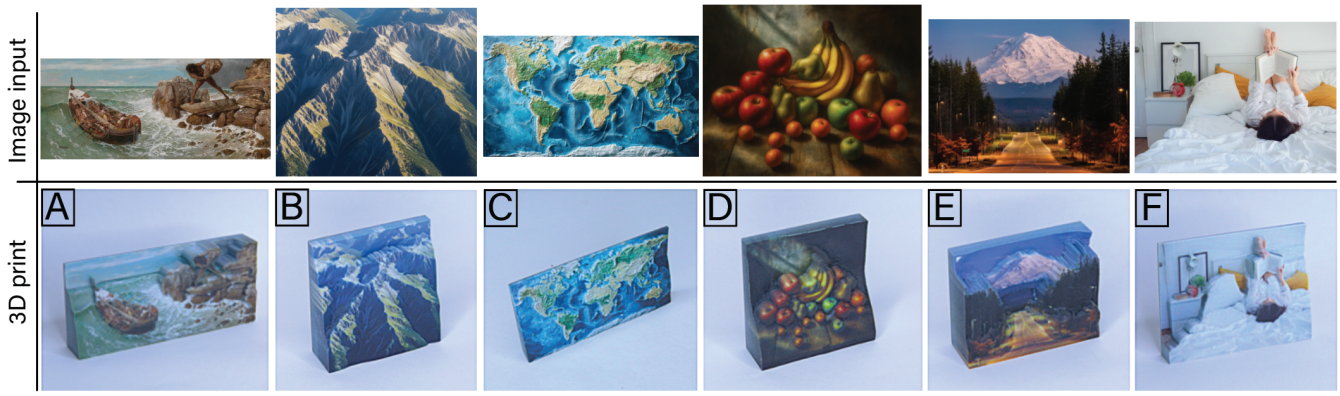

Figure 2: (Above) Image inputs to Text2Texture and (below) the resulting 3D printed objects for (A) a painting, (B) a mountain range, (C) a topological world map, (D) a still life image, (E) a vista of Mt. Rainier, and (F) a photo of a family member reading.

gap by combining monocular depth estimation, semantic texture synthesis, and visual color cues into a single end-to-end pipeline. The result is a textured, colored, and depth-aware 3D model that aims to encode both the *form* and *feel* of the input image.

Text2Texture converts a 2D color image into a textured 3D object ready for 3D printing. Implemented as a web interface, users can upload an image directly or generate one with a text prompt. A monocular estimator predicts a macroscale heightmap from this image, extruding the 2D image to form a closed 2.5D surface. In parallel, semantic segmentation is used to partition the image into local textures, such as tree, rocks, grass etc. The system extracts these textual features and uses a fine-tuned stable diffusion model to extract the microgeometry representing their textures. Text2Texture then superimposes the macroscale geometry from the heightmap with the microscale geometry representing each feature's texture. The result is a 3D-printable model of the image that captures both depth information from the scene and haptic information related to the texture of individual features within the scene.

2 Text2Texture

Text2Texture integrates an interactive web interface with an AI-driven multi-model processing pipeline. To use the tool, users either upload an image or provide a textual scene description, the latter being synthesised into an image by OpenAI's DALL-E 3 [1]. Next, the image is automatically segmented with Segment-Anything Model (SAM) [10]. These semantic regions are automatically labeled with Florence-2 [14], which the user can edit if needed. The interface allows additional user-guided edits—clicking points or drawing bounding boxes to re-run SAM on selected areas—thereby producing high-quality masks even for intricate object boundaries.

Masked regions may be renamed, enabled, or deleted, and each accepted mask triggers the generation of a fine-grained heightfield encoding surface micro-texture. These heightfields are produced by TactStyle, a system that generates surface microgeometry as heightfields using a fine-tuned stable-diffusion model, enabling 3D-printed objects that have both the appearance and feel of real-world textures [5]. We fine-tuned TactStyle with the MatSynth [13] dataset to enable a larger set of textures to be generated, ensuring photorealistic, material-aware detail that aligns with the semantic label of each region. In parallel, the complete image is passed to ZoeDepth [3], a monocular estimator that predicts a dense global depth field. This depth map is normalized to a user-specified range and additively merged, within mask boundaries, with all regional texture height-maps. Thus, at the end of this process, there are two heightfields, one representing the texture (microgeometry), and the other representing the depth map (macrogeometry). These two heightfields are then composed together. Empirically, we found that a 90%-10% ratio of macro-micro geometry combination provides for an accurate replication of both depth and texture in the 3D model. Next, this composite heightfield is sampled at the native image resolution to displace the top surface of a uniform-thickness rectangular block; the algorithm triangulates both the modulated top face and a flat bottom face on this grid, then stitches corresponding perimeter vertices to form continuous side walls, producing a watertight mesh ready for 3D printing. Throughout the workflow, the web interface provides real-time previews together with a concise parameter set—depth scale, base thickness, and a toggle between protrude (raising height-mapped features above the block surface) and carve (engraving them below it) modes—supporting iterative refinement without exposing low-level model orchestration. On completion, the user may export the model for fabrication. We showcase fabrication by 3D printing all models on a Stratasys J55 printer, with Vero family materials that allow full-color 3D printing.

Text2Texture bridges a gap between vision and fabrication, making it possible to not just see but touch the content of images. We believe it opens the door to rich, cross-modal design experiences that are tactile, expressive, and physical.

3 Applications & Future Work

We use Text2Texture to generate and print a variety of 3D models from both image and text prompts (Fig 2). These suggest how users can interact physically with objects within art (Fig 2A,D); impart a sense of scale to landscape images (Fig 2B,C); and embed haptic cues in photographed memories (Fig 2E,F). Future work will evaluate how to superimpose macro-scale geometry and micro-scale texture to maximize perceptual attributes via psychophysical perception experiments [7, 8, 11], and how the tool can be used by blind and low-vision participants to create tactile story-boards.

References

[1] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. 2023. Improving image generation with better captions. *Computer Science. https://cdn. openai. com/papers/dall-e-3. pdf* 2, 3 (2023), 8.

[2] Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. 2022. Localbins: Improving depth estimation by learning local distributions. In *European Conference on Computer Vision*. Springer, 480–496.

[3] Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. 2023. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288* (2023).

[4] Donald Degraen, Michal Piovarči, Bernd Bickel, and Antonio Krüger. 2021. Capturing tactile properties of real surfaces for haptic reproduction. In *The 34th annual ACM symposium on user interface software and technology*. 954–971.

[5] Faraz Faruqi, Maxine Perroni-Scharf, Jaskaran Singh Walia, Yunyi Zhu, Shuyue Feng, Donald Degraen, and Stefanie Mueller. 2025. TactStyle: Generating Tactile Textures with Generative AI for Digital Fabrication. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.

[6] Martin Feick, Donald Degraen, Fabian Hupperich, and Antonio Krüger. 2023. Metareality: enhancing tactile experiences using actuated 3D-printed metamaterials in virtual reality. *Frontiers in Virtual Reality* 4 (2023), 1172381.

[7] Roland W Fleming. 2014. Visual perception of materials and their properties. *Vision research* 94 (2014), 62–75.

[8] Roland W Fleming. 2017. Material perception. *Annual review of vision science* 3, 1 (2017), 365–388.

[9] Micah K Johnson and Edward H Adelson. 2009. Retrographic sensing for the measurement of surface texture and shape. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1070–1077.

[10] Lei Ke, Mingqiao Ye, Martin Danelljan, Yu-Wing Tai, Chi-Keung Tang, Fisher Yu, et al. 2023. Segment anything in high quality. *Advances in Neural Information Processing Systems* 36 (2023), 29914–29934.

[11] Shogo Okamoto, Hikaru Nagano, and Yoji Yamada. 2012. Psychophysical dimensions of tactile perception of textures. *IEEE Transactions on Haptics* 6, 1 (2012), 81–93.

[12] Cesar Torres, Tim Campbell, Neil Kumar, and Eric Paulos. 2015. HapticPrint: Designing feel aesthetics for digital fabrication. In *Proceedings of the 28th annual ACM symposium on user interface software & technology*. 583–591.

[13] Giuseppe Vecchio and Valentin Deschaintre. 2024. Matsynth: A modern pbr materials dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22109–22118.

[14] Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, and Lu Yuan. 2024. Florence-2: Advancing a unified representation for a variety of vision tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4818–4829.