BlendFusion: Procedural 3D Texturing Assistant with View-Consistent Generative Models

Qinchan Li*
Finley Torrens*
Kenneth Chen
Qi Sun
Tandon School of Engineering, New York University, USA

ABSTRACT

Modeling 3D assets is universal in various applications, including animation and game development. However, a key challenge lies in the labor-intensive task of 3D texturing, where creators must repeatedly update textures to align with modified geometric shapes on the fly. This iterative workflow makes 3D texturing significantly more cumbersome and less efficient than 2D image painting. To address this, we introduce BlendFusion, an interactive framework that leverages generative diffusion models to streamline 3D texturing. Unlike existing systems that generate textures from scratch, BlendFusion integrates the procedural nature of texturing by incorporating multi-view projection to guide the generation process, enhancing stylistic alignment with the creator's intent. Experimental results demonstrate the robustness and consistency of BlendFusion across both objective and subjective evaluations.

ACM Reference Format:
Qinchan Li, Finley Torrens, Kenneth Chen, and Qi Sun. 2025. BlendFusion: Procedural 3D Texturing Assistant with View-Consistent Generative Models. In *Companion Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D Companion '25), May 7–9, 2025, Jersey City, NJ, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3722564.3728376

(a) initial status (b) manual after 20 min (c) ours 2 min inference

Figure 1: (a) shows the overall and zoomed-in 3D scene for re-texturing. (b) shows the modeled texture after 20 minutes by a professional designer, and (c) shows the results using our framework with an inference time of 2 minutes with prompt "a child drawing style of a cabinet in a cartoony kitchen". Model credit to SketchFab user MiwhyMiwh.

1 INTRODUCTION

Texturing is a creative, yet labor-intensive task in creating 3D assets. This process requires navigating complex software like Blender and mastering specialized skills, such as mentally mapping 2D screen interactions onto 3D objects [3]. One of the most challenging tasks in 3D texturing is designing texture [6], a task which demands spatial ability to iteratively align 2D textures and 3D geometries with UV-mapping [2], resulting in prolonged design cycles [8].

To improve 3D texturing efficiency, various 3D interfaces have been proposed, such as sketching in VR [1]. More recently, text-to-3D generative AI has created a new paradigm [7]. However, both methods face tradeoffs in full automation vs. fine-grained control.

We present BlendFusion, a generative-AI-assisted interface for rapid 3D asset texturing and editing. BlendFusion simplifies the time-consuming process of run-time texture editing, allowing for seamless adjustments to frequent geometric edits. Our GenAI-assisted interface frees creators from the tedious manual texturing.

We adopt 3D-geometry-conditioned 2D image generation, via geometric measurements in 2D space with depth and color projections. The 2D-projected information then controls the text-to-texture generation in 2D space. Our experiments show improved 3D texturing efficiency and visual quality.

*Both authors contributed equally to this research.

I3D Companion '25, May 7–9, 2025, Jersey City, NJ, USA

ACM ISBN 979-8-4007-1833-5/2025/05.
https://doi.org/10.1145/3722564.3728376

2 METHOD

As in Figure 2, users first select camera views, and the inputs to the generative diffusion model are rendered. View-dependent textures are then generated and projected onto the 3D model. With our pipeline, modelers may only manually select viewpoints or fully automatically generate with Text2tex [4].

Selection of camera views. To leverage 2D generative models in 3D texturing, we tackle the dimensional gap by generating view-dependent textures. BlendFusion allows users to automatically or manually select camera views. We leverage the framework of Chen et al. [4] for automatic camera sampling, and modelers can also manually select the views to control the covered regions.

Deferred shading to generate view-dependent guidance. We proposed several inputs to achieve geometry-aware and view-consistent texture generation: a depth map as view-dependent geometry, a 2D mask, and the initial color map. The depth map controls the generated image's geometry. The mask specifies generated regions of the texture image, and the color map serves as the visual guidance. All inputs are rendered through Blender's EEVEE rendering engine.

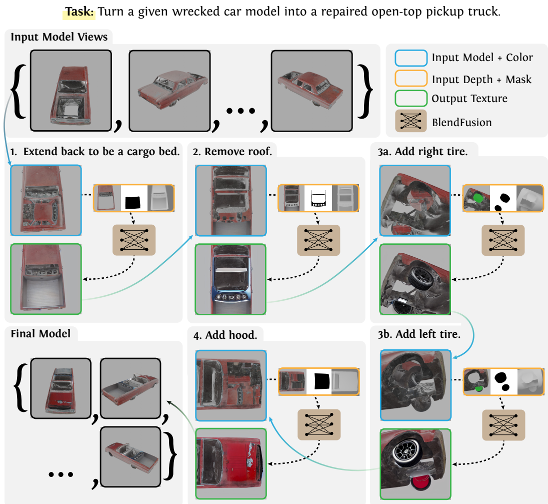

Figure 2: With the 3D assets' multi-view color and depth maps, BlendFusion leverages a controllable diffusion model [10] to generate the texture that aligns with 3D geometry and users' prompts. The workflow may be procedurally performed with users-in-the-loop. Model credit to SketchFab user Arthur_mf.

Texture generation and projection. We use pre-trained text-to-image Stable Diffusion [9] controlled by ControlNet [10] for text-hint geometry-guided image inpainting. Specifically, we inject depth map and text as the conditions and use the user-specified boolean mask for seamless generation. Finally, we utilize the Blender UV Camera Project modifier to project the generated texture onto the 3D model. This method UV-unwraps the vertices for the camera view and applies the image to this UV map.

3 EVALUATION

We compare BlendFusion's interface with existing texturing frameworks from two aspects: visual quality (compared with manual way) and multi-view consistency (compared with DreamTexture[1]).

3.1 Visual Quality

We recruited 12 participants (8 male and 4 female, 22-30 years old) for 9 random-shuffled trials. During the study, the participants remained seated, viewed the images, and indicated their decision with the task detailed below. We ran a seven-point Likert scale study to measure users' visual preference for BlendFusion-made models. In addition, we investigate whether users can distinguish BlendFusion-made and expert-made models. During each trial, users were presented with three different models. The first was the ready-to-texture model, the second was the reference model, which is pegged at a rating of 4, and the third was the target model for users to rate. In addition, we randomly select **MANUAL** or **OURS** as reference and ask users to choose whether the reference model was made by AI (our pipeline). To calibrate **MANUAL** to 4, we add (4 - **MANUAL**'S rating) to **OURS** and **MANUAL**.

[1]https://github.com/carson-katri/dream-textures

Results. The user ratings, grouped by geometry-type, are in Figure 3. The average rating of **OURS** is 4.1 and the paired t-test has a $p << .01$ with an alternative hypothesis of "the ratings of **OURS** are greater than **MANUAL** ". Users only achieve a near-random guess of 54.63% in discriminating AI-made textures. It shows that our UI bridges the gap between novices and experts in texturing.

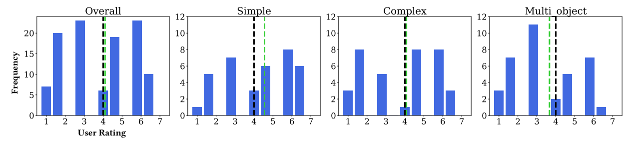

Figure 3: *User study results and statistics.* Black line is at rating 4 (reference rating), and green line is mean of user ratings. $x-$axis on Likert scale, and $y-$axis is # responses.

3.2 Multi-view Consistency

To validate the multi-view consistency, we compare **OURS** with existing GenAI-assisted UI: **DreamTexture**. We partially untextured 8 existing models, re-textured with BlendFusion and **DreamTexture**, and measured the consistency between the generated and original textures. For each object, we render two images from two randomly selected views for **OURS** and **DreamTexture**, and we calculate their DreamSim [5] similarities, which reflects perceptual similarity, with the rendering from the original texture.

We got the DreamSim score for **OURS** with 0.176±0.039 and for **DreamTexture** with 0.205±0.043 (a lower score means more visual similarity in DreamSim). We conclude that the texture created by BlendFusion is more consistent with the original texture.

REFERENCES

[1] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. Symbiosissketch: Combining 2d & 3d sketching for designing detailed 3d objects in situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–15.
[2] Ali Ihsan Benzer and Bunyamin Yildiz. 2019. The Effect of Computeraided 3D Modeling Activities on Pre-Service Teachers' Spatial Abilities and Attitudes towards 3D Modeling. *Journal of Baltic Science Education* 18, 3 (2019), 335–348.
[3] Srinjita Bhaduri, Katie Van Horne, and Tamara Sumner. 2019. Designing an informal learning curriculum to develop 3D modeling knowledge and improve spatial thinking skills. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.
[4] Dave Zhenyu Chen, Yawar Siddiqui, Hsin-Ying Lee, Sergey Tulyakov, and Matthias Nießner. 2023. Text2tex: Text-driven texture synthesis via diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 18558–18568.
[5] Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. 2024. DreamSim: Learning New Dimensions of Human Visual Similarity using Synthetic Data. *Advances in Neural Information Processing Systems* 36 (2024).
[6] Yotam Gingold, Takeo Igarashi, and Denis Zorin. 2009. Structured annotations for 2D-to-3D modeling. In *ACM SIGGRAPH Asia 2009 papers*. 1–9.
[7] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. 2022. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988* (2022).
[8] Alec Rivers, Frédo Durand, and Takeo Igarashi. 2010. 3d modeling with silhouettes. In *Acm siggraph 2010 papers*. 1–8.
[9] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.
[10] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.