Geometry Aware Texturing

Evgeniia Cheskidova
Ready Player Me
Tallinn, Estonia
cheskidovaeugenia@gmail.com

Aleksandr Arganaidi
Ready Player Me
Tallinn, Estonia
einstalek@gmail.com

Daniel-Ionut Rancea
Ready Player Me
Berlin, Germany
daniel-ionut.rancea@readyplayer.me

Olaf Haag
Ready Player Me
Berlin, Germany
info@olafhaag.com

Figure 1: 1. Original 3D asset with a diffuse map created by a professional artist, 2. Original 3D asset, mesh only. 3. Original asset with generated textures from prompt "Sailor Moon anime", 4. Close-up of boots and UV-atlas of generated diffuse texture.

ABSTRACT

In this work, we propose a novel approach to texture generation, making use of recent advancements in Latent Diffusion models, [Rombach et al. 2022] unlocked by [Zhang and Agrawala 2023], introducing control inputs to generation pipelines via ControlNet. We find that a special condition, where the mesh is encoded into UV space, can serve as a control input, producing textures that are geometrically and visually coherent, and of high quality. Using this approach, we are able to generate a unique look guided by text for existing meshes in a matter of seconds.

SA Posters '23, December 12–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0313-3/23/12.
https://doi.org/10.1145/3610542.3626152

CCS CONCEPTS

• **Applied computing** → *Media arts*; *Fine arts*.

KEYWORDS

Generative Texturing, 3D texture, Stable Diffusion, ControlNet

ACM Reference Format:
Evgeniia Cheskidova, Aleksandr Arganaidi, Daniel-Ionut Rancea, and Olaf Haag. 2023. Geometry Aware Texturing. In *SIGGRAPH Asia 2023 Posters (SA Posters '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3610542.3626152

1 INTRODUCTION

Mesh texturing stands as an important problem in generative 3D pipelines. With Stable Diffusion able to produce high-quality images following text guidance, text-to-image conversion has become easier than ever before. Applied to the problem of texturing, a straightforward solution was proposed by [Richardson et al. 2023], making use of Stable Diffusion's inpainting abilities paired with re-projection. Although this and other similar methods allow for

the effective creation of textures, one texture creation usually takes a significant amount of time, and the resulting texture contains excessive baked lights and shadows. The latter might not be suitable for many typical 3D pipelines in the industry that make use of physically-based rendering materials. Moreover, assets with lots of overlapping details might get a worse texture due to re-projection artifacts.

To tackle these issues, we propose a different method. Instead of iterative texturing, we attempt to generate textures directly in UV space, conditioning them only on geometry-specific inputs in image form. Our key finding is that the spatial information of a mesh can be effectively encoded into a single image. This allows us to achieve consistent colors and reduce seam problems on the edges of UV islands. The generated textures are "aware" of the assets' symmetry, back and frontal parts, and small geometrical details. Our method can generalize to unseen 3D meshes that were not included in the training set.

2 METHOD

In our experiments, primarily focused on meshes of avatar outfits, we tested various control inputs with ControlNet and their influence on the quality of textures. We were interested not only in producing coherent-looking textures that match text descriptions but also in the consistency of textures between the edges of UV islands. It is often the case that geometric parts of a 3D asset positioned closely together are split into different UV islands, leading to seams around UV edges. We discovered that the conditioning input most

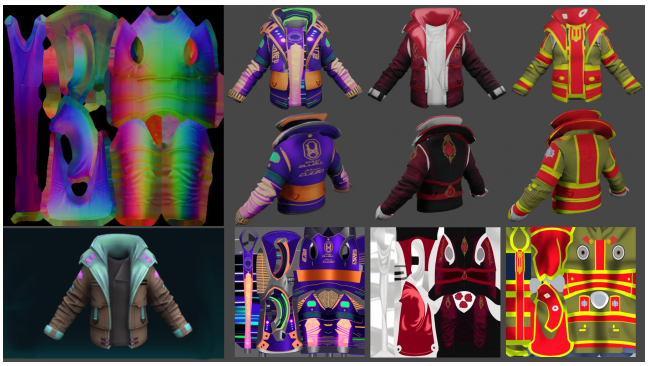

Figure 2: 1. Position encodings + Object Space Normals as a condition for ControlNet for the original 3D jacket shown below. 2. Variations of textures produced from different prompts. From left to right: "Futuristic gear, LED accents", "Graf Dracula", "Firefighter". UV textures produced by the model are also shown below each render.

effectively mitigating this issue is a linear combination of object space normals and position maps aligned with the original UV layout of an asset, as shown in Figure 2. These maps were precomputed using Blender ([Blender Online Community 2023]), for a set of ~ 1000 3D outfits, varying in shape (from shoes to upper clothing) and textures, and we trained a ControlNet on this dataset. In our experiments, we exclusively used 3D assets manually created by 3D artists at Ready Player Me. This input equips the generation pipeline with enough positional awareness to create novel details that conform to the asset's structure. Additionally, other custom normal maps could be generated to provide more visual support for these new details. We implemented this aspect as a procedural generation based on the original asset's normals.

3 LIMITATIONS AND CONCLUSIONS

Our method demonstrates the ability to quickly generate spatially coherent textures of high quality. However, seamless texture generation remains a primary challenge, and our approach represents an attempt to mitigate this issue to some extent. While positional maps encode the geometry well enough to follow the asset structure, we still observe inconsistencies between adjacent UV islands, especially if they share a continuous pattern like the straps shown in Figure 3. This issue can be addressed in various ways, such as creating an alternative set of UVs for each model to conceal the seams. Regarding future work, we plan to explore model alignment to fine-tune the model, aiming to produce consistent colors on adjacent regions and simplify the process of generating aesthetically pleasing textures while better adhering to additional aesthetic requirements.

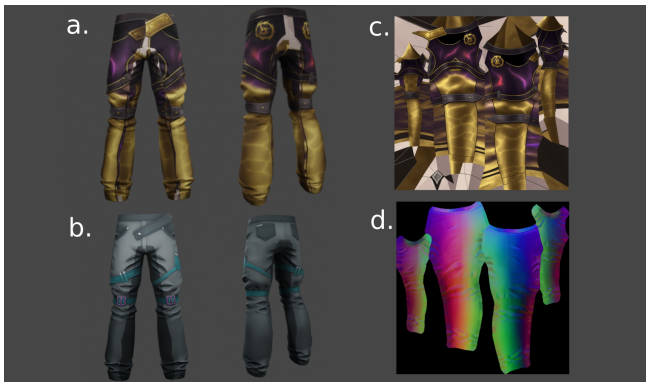

Figure 3: a.) Render of pants with generated textures demonstrating misalignment in seams region. b.) The original asset created by a 3D artist. c.) Generated UV-texture d.) Conditional image that demonstrates how UV islands are separated on the image.

ACKNOWLEDGMENTS

To Rainer Selvet, for being curious and getting hands-on tests with segmentation ControlNet APIs. Without his small experiment, we could have missed the idea of this work.

REFERENCES

Blender Online Community. 2023. *Blender - a 3D modelling and rendering package.* Blender Foundation, Blender Institute, Amsterdam. http://www.blender.org

Elad Richardson, Gal Metzer, Yuval Alaluf, Raja Giryes, and Daniel Cohen-Or. 2023. TEXTure: Text-Guided Texturing of 3D Shapes. arXiv:2302.01721 [cs.CV]

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752 [cs.CV]

Lvmin Zhang and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543 [cs.CV]