PG23
Daejeon, Korea

Color3d: Photorealistic texture mapping for 3D mesh

Chenxi Zhao[1], Chuanmao Fan[1], Payal Mohadikar[1], Ye Duan[2]

1. University of Missouri, USA
2. Clemson University, USA

PROBLEM

3D reconstruction stands as a long-standing and fundamental topic in computer vision. Successful reconstruction requires high-quality meshes and accurate color representation. In this study, we focus on the coloration of the meshes. We proposed a Multiview texture mapping procedure for coloring. Given a mesh without color and Multiview RGBD images of the colored object with known camera poses and intrinsic, our method can generate photorealistic color for the mesh. Figure 1 illustrates the input mesh, Multiview images, ground truth, and the colored mesh from our method.

Figure 1. An illustration for the input mesh, Multiview RGBD images, ground truth and our colored mesh.

RELATED WORK

In the realm of 3D reconstruction, numerous methods have been proposed. Approaches like NeRF [1] represent scenes through neural radiance fields, utilizing Multiview images as their input. These techniques require a substantial number of images and are generally trained on a per-scene basis. In contrast to images, point clouds encompass inherent 3D information, making them a more intuitive medium for representing objects or scenes. However, rendering point clouds presents a challenge due to their lack of volume. Among the various 3D reconstruction methods from point clouds, the concept of implicit fields stands out. Notably, the Implicit Feature Network (IF-Net) [2] falls short in predicting color information, although a texture extension variant has been proposed [3]; the predictions remain deficient in capturing details.

OVERVIEW

Given a reconstructed mesh without color, our goal is to produce a mesh with photorealistic color. Our approach takes the mesh, Multiview RGBD images, and camera matrices as input. While deep learning methods can also learn to predict color, the process is time-consuming and often results in low-quality outcomes restricted by resolution. All the inputs can be acquired through an RGBD system, which is cost-effective. This texture mapping technique is capable of generating colors that are more photorealistic and intricate, such as capturing subtle facial expressions on the human body, or detailed pattern of the object surfaces. Ideally, there are no restrictions on the input mesh as long as the projection process is correct, which means our method could be applied to any single object or larger scene. In our experiment, we use the human body object as an example.

METHODOLOGY

Traditional methods like texture mapping can provide higher-quality colors compared to deep learning approaches. However, the reconstructed mesh surfaces, such as the human body object, are often more complex than the ground truth. This complexity arises because surfaces generated from deep learning methods are not always smooth and may contain more individual triangles with normals oriented in various directions. As a result, to address this issue, we replace the conventional UV map-based texture mapping with vertex colors. These colors can be directly back-projected from Multiview images. Initially, we extract all vertices in world coordinates from the provided or predicted mesh.

Subsequently, using camera matrices, we calculate the corresponding pixel coordinates in the Multiview images. Using camera matrices, we can compute the depth of vertices in each camera coordinate system. We then compare the differences between the computed depth and the values from depth images. If the difference is greater than a threshold value, the vertex will be counted as invalid. Each valid pixel coordinate corresponds to a specific color in the image. This color is then projected back onto the corresponding vertex. The colors of the faces are obtained by interpolation from the vertex colors of each triangle. The pipeline of our texture mapping method is illustrated in Figure 3.

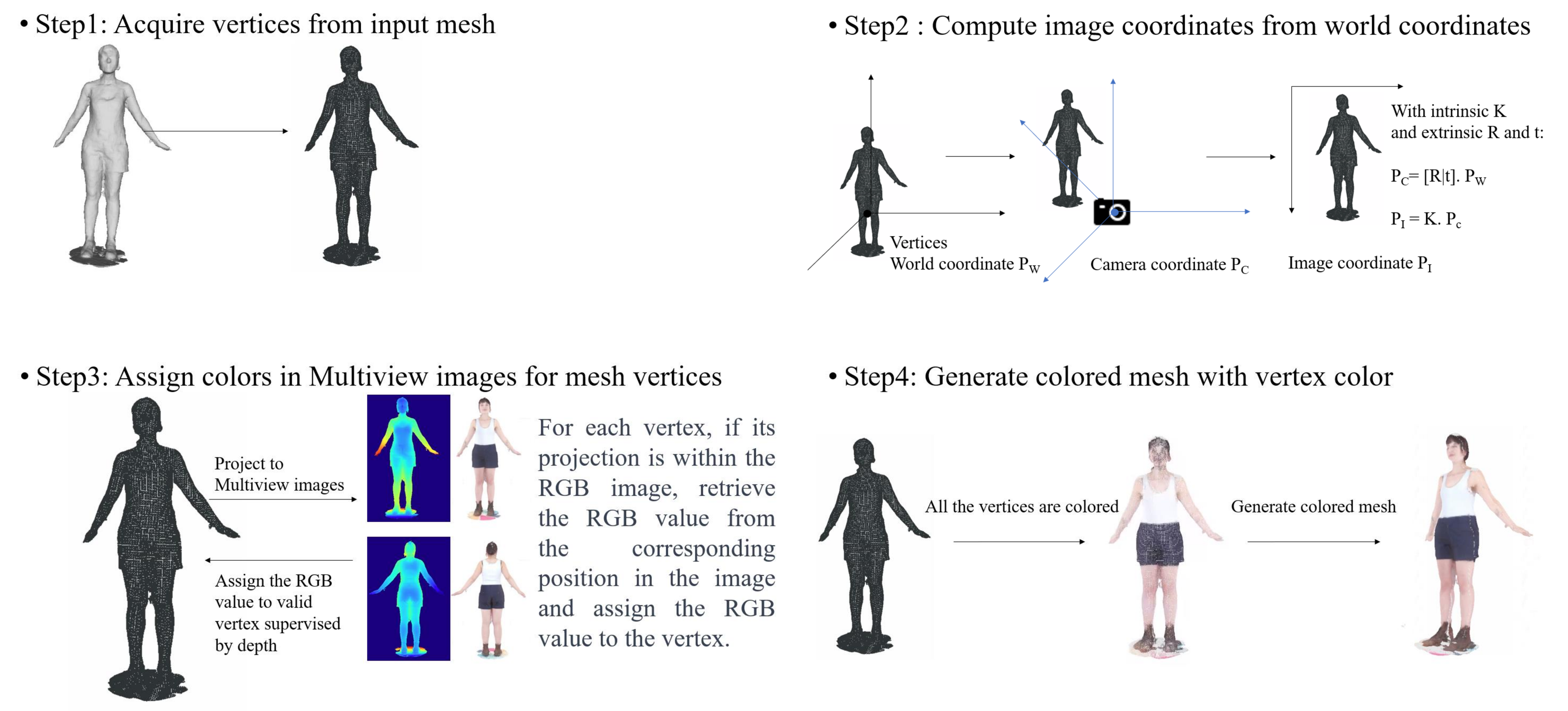

Figure 3. The pipeline of our proposed texture mapping method.

RESULTS

Fist we apply our method on some human body objects in some simple pose. The input mesh is predicted by some deep learning mesh generation network. To simulate a real RGBD (Red, Green, Blue, Depth) system, we establish a configuration with nine virtual cameras, all sharing the same intrinsic parameters but oriented in different directions. These camera poses and intrinsic are known. Some examples are shown in Figure 4. Noted that there are some differences between the input mesh and ground truth. These differences lead to some blank areas could not be projected correctly and not colored. Furthermore, we assess the performance of our approach on an animation file which contains a sequence of continuous poses. Preparing for further real-time applications, we evaluate our method on this new data, generating an animation with each colored mesh serving as an individual frame. Selected examples are shown in Figure 5.

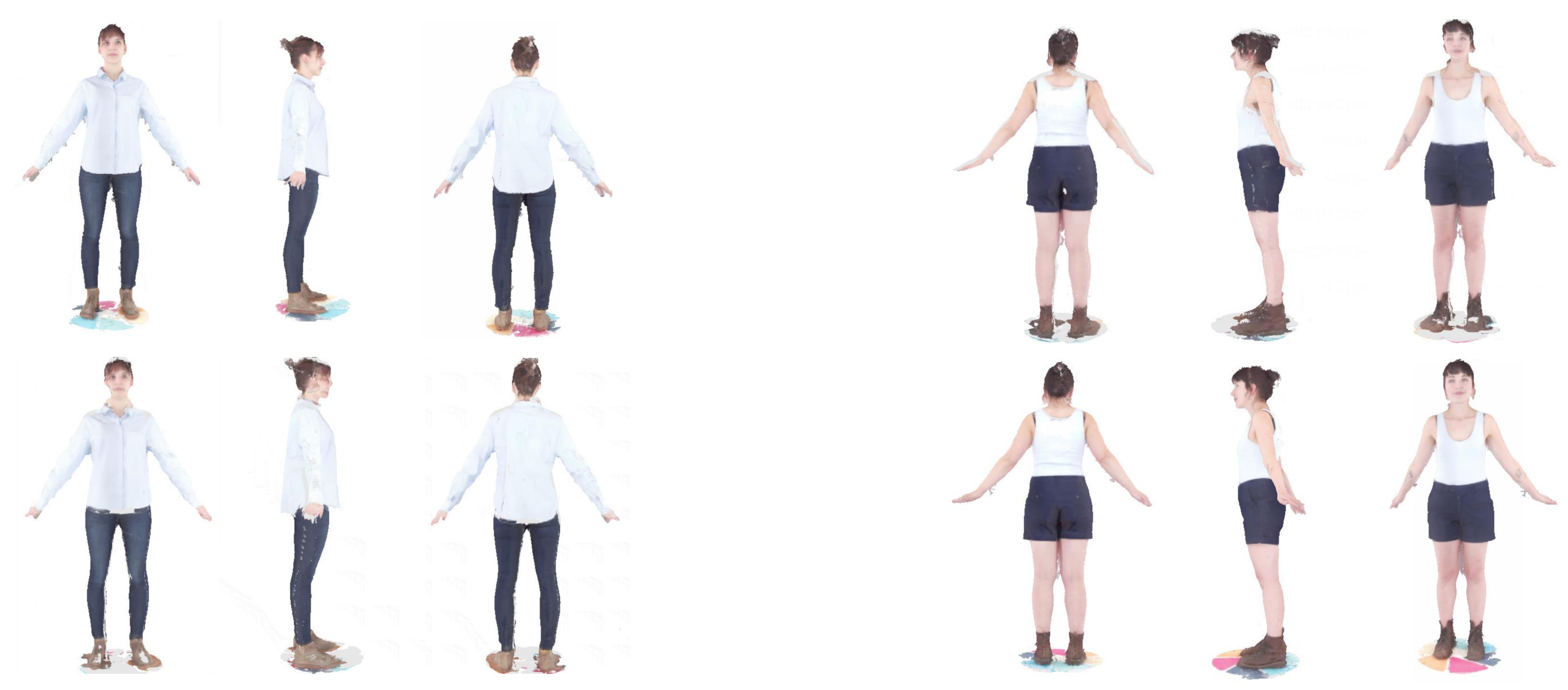

Figure 4. Comparisons between our results and ground truth. In each subfigure, the upper 3 small figures are our results, the lower 3 small figures are the ground truth.

Figure 5. Texture mapping results from animation data.

REFERENCES

[1] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65.1 (2021): 99-106.
[2] Chibane, Julian, Thiemo Alldieck, and Gerard Pons-Moll. "Implicit functions in feature space for 3d shape reconstruction and completion." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
[3] Chibane, Julian, and Gerard Pons-Moll. "Implicit feature networks for texture completion from partial 3d data." Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020.

Electrical Engineering & Computer Science
University of Missouri