Computer Vision 3D Reconstruction for Prosthetic Limb Design

Ben Hofflich
bhofflic@stanford.edu

Nurmohammed Patwary
npatwary@stanford.edu

Ali Yaghoubzade
aliy@stanford.edu

A. Abstract

The aesthetic appearance of prosthetics is a concern for many amputees and so there is a direct need for accurate prosthetic modeling to match the existing figure of the user. We propose a method to generate a 3D reconstruction of the patient's limb using NVIDIA's Instant-NGP NeRF generation software. We evaluate the accuracy of the model in comparison to the computer vision technique, Space Carving, and compare both models to the ground truth measurements. We found that both models performed very well with Space Carving having an accuracy of 95% and the Instant-NGP model having an exceptional accuracy of 99%. We show that computer vision based 3D modeling can provide accurate reconstructions for anatomical structures for use in prosthetic modeling.

B. Introduction

Prosthetics research and development has been largely focused on improving functionality and usability of the prosthetic for the user, however, has been lacking in accommodating the individuality of the users [1]. The absence of diversely sized prosthetic parts may impede prosthetic users from fully considering their prosthetic a true extension of themselves. 3D reconstruction from multiple views is a powerful technique that can be used to create highly accurate and detailed 3D models of the patient's body, which can be used to design and manufacture orthotic and prosthetic devices that fit perfectly and function optimally. The process of 3D reconstruction from multiple views involves taking multiple 2D images of the patient's body from different angles and using computer algorithms to reconstruct a 3D model of the body. Additionally, traditional means for measuring body parts, such as CT, MRI, and 3D scanners, are not always readily available and are costly.To address these issues, we propose a framework using Neural Radiance Fields (NeRF) to provide an accessible, cost-efficient method for modeling prosthetic limbs.

B.1. Problem Statement

Individuals who have lost a limb require prostheses, and currently, there are approximately 1.9 million people who fall under this category in the US. The cost of a prosthetic device ranges between $2,500 and $50,000, and adults usually need to replace them every one to three years. Children require replacements more frequently, approximately every three to six months, as they continue to grow. [2] Previous studies on prosthetic design have primarily relied on manual design and the use of CAD/CAM systems. Typically, when using manual design, the shape of the residual limb is determined by creating a mould of the limb itself, which is then manipulated by a trained practitioner to properly distribute the pressure on the patient. Once the physical model is created using a milling or carving machine, it is transformed into a foam or plaster shape and applied to the patient as a medical support device. [3]The amount of time and number of attempts required in the process of obtaining a well-fitted prostheses through trial and error depends on the experience of the prosthetist. [4]. Compared to all previous techniques, Computer Vision 3D reconstruction method is a low-cost and portable solution, capable of easy manipulation and of making captures simultaneously. [5]

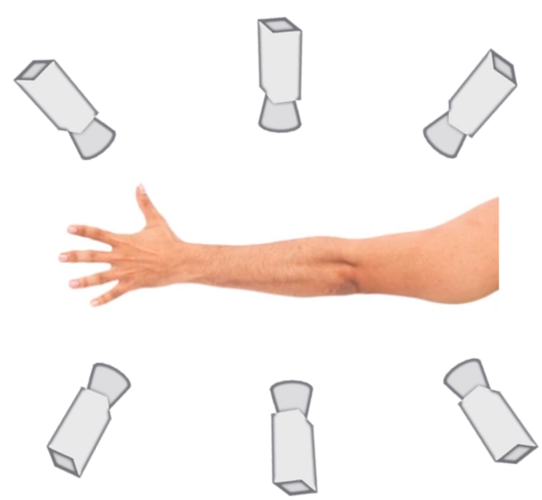


Figure 1. Project Illustration

C. Prior Work

C.1. Instant-NGP

We render the 3D model of a hand using the methods described in [6], which the authors calls Instant NGP. In this paper, authors proposed a computationally efficient way to render 3D volume from a set of 2D images using NeRF [7], which is a seminal work in this field. Ref [6] achieved 20-60X performance improvement in terms of rendering time. For example, Instant NGP can reconstruct a gigapixel image in 2.5 minutes, where the original NeRF method takes 36.9 hours on a NVIDIA RTX 3090 GPU.

NeRF model represents a continuous scene as a 5D function where input is a 3D location x = (x, y, z) and 2D view angle (θ, ϕ), and the output c = (r, g, b) and the volume density, σ. These information are feed to a Neural Network which is based on DeepSDF [8] architecture. NeRF adds one additional activation layer. It has an additional output layer that outputs the density and a feature vector. Density and feature vector is concatenated to the positional encoding of the input viewing direction. This layer is processed by a 128-channel ReLU layer that outputs the RGB radiance from the input direction. The positional encoding helps to recover high frequency information while rendering 3D shape. The positional encoding is similar to the Transformer architecture to map the input to a higher dimensional space.

The authors derive a differentiable discrete density function using the classical principal of volume rendering. This allows to use gradient-descent for parameter optimization. The volume is sampled in Hierarchical manner in two steps: (a) Coarse and (b) Fine. Coarse step creates a "coarse" network at some sampled locations and Fine step produced better sampling using coarse results.

Training NeRF requires only a dataset of captured RGB images of the scene, the corresponding camera poses and intrinsic parameters, and scene bounds. The network is trained as follows:

1. Sample a batch of camera rays from the set of all pixels in the dataset
2. Apply hierarchical sampling
3. Apply rendering
4. Compute the loss
5. Optimize the model and update the parameters.

C.2. Triangulation

Li et al have introduced a technique that utilizes photogrammetry to produce 3D models of prosthetics and orthotics for customization and fitting purposes [5]. The authors' method is based on the principle of triangulation, where the camera is calibrated and point correspondence is established before creating the 3D surface in space. They implemented the Scale Invariant Feature Transform (SIFT) [9] to transform image data into scale-invariant coordinates based on local features. Then, they used epipolar geometry to establish feature point correspondence and perform 3D reconstruction of the scene, accounting for an overall 3D projective deformation. According to the paper, the photogrammetry-based system was tested on five subjects with transtibial amputation and produced accurate and reliable 3D surface reconstructions.

D. Methods

For our project, we aimed to evaluate the accuracy of NVIDIA's Instant-NGP NeRF generating network for anatomical 3D reconstruction. As a baseline for our evaluation, we utilized the space carving method implemented in Problem Set 3. The accuracy for these models was determined using the relative measurements of various anatomical measures compared to the ground-truth.

D.1. Data Collection

The Instant-NGP model and Space Carving model require training images to perform the 3D reconstruction. We recorded a 360 degree video of a hand using an iPhone 14 Pro camera. The video was recorded horizontally with a video quality of 1080x1920 with a duration of 58 seconds.

D.2. Preprocessing

To extract image frames from the video recording, as well as to determine the intrinsic and extrinsic camera properties, we utilized the open source COLMAP software. The video recording was sampled at 2 frames per second, yielding 116 images. The model trained using COLMAP successfully estimated the properties for 73 of these images. It is important to note that the successfully chosen images comprise views primarily of the backhand, and the left and right side views, with significant occlusion of the front facing and palm views.

These camera properties were exported and formatted into a .mat file to input into the space carving algorithm. The silhouette for each image was determined manually in Adobe Photoshop. The silhouette image was then quantized and formatted into a binary image and included in the input .mat file. Figure 2 shows the comparison between a raw frame and its silhouette.

D.3. Space Carving

As a baseline for our project, we utilized the space carving method implemented in Problem Set 3. We used the silhouettes for all camera views to enforce consistency in our reconstruction. Initially, we determined bounds for the

Figure 2. Raw Image and Silhouette

object's location using the minimum and maximum camera positions as determined using COLMAP. Using those bounds, we determined the coordinates and size of the voxels from which the model would be constructed. For each view, we projected each voxel onto the image space as determined by the view's transformation matrix P. We enforced 2 criteria to determine whether the voxel belongs to the model:

1. The projection falls within the image bounds [0:1080, 0:1920]
2. The voxel projects into the space enclosed by the view's silhouette.

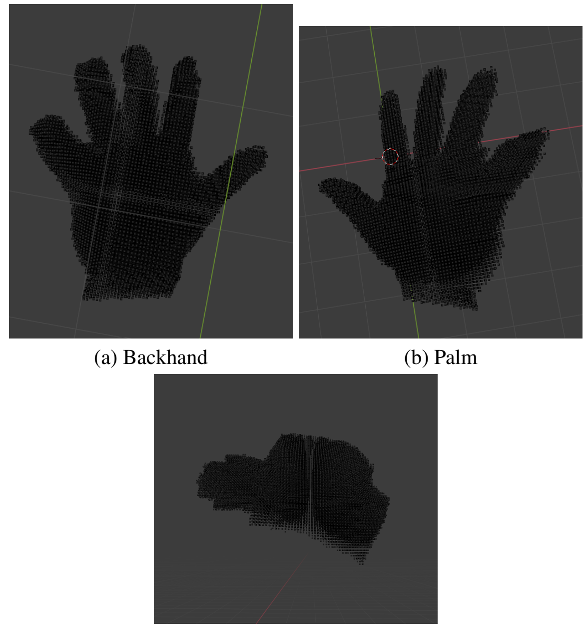
Any voxel that satisfied these criteria was retained and any that did not was removed from the reconstruction set. This process was repeated for each camera view to ensure consistency and achieve a more thorough reconstruction. This model was then exported as a point cloud to Blender for measurement as shown in Figure 3

D.4. Instant-NGP

Using the Instant-NGP library, we generated a 3D reconstruction of the scene as a NeRF. We inputted the reconstructed points as determined by COLMAP for each view which Instant-NGP trained on to create the reconstruction shown in Figure 4 To isolate the portion of interest i.e., the hand, we adjusted the bounding box and exported the 3D model as an .obj for post-processing and measurement. These models are shown in Figure 5

D.5. Post-Processing

The outputs of the Instant-NGP model included a significant number of noisy residuals, though the target object was still visible and discernable from the noise. To simplify the measurement process and clean the model, we imported the model into the Blender software and manually removed the residuals outside of the target model. Figure 6 shows



(a) Backhand (b) Palm

(c) Pinkyside

Figure 3. Space Carving Models

the comparison between the original output and the post-processed output object models.

D.6. Measurement

We utilized the integrated measurement tool in Blender to measure the following anatomical features, shown in Table 1. The reference map for these features is shown in Figure 7

E. Results

To evaluate the accuracy of the models used, we needed to compare the anatomical measurements from the models to the ground-truth measurements. Because the model measurements are in relative units, we need to standardize the measurements to enable comparison. For each model, the ratio between each anatomical feature was then calculated and plotted as a matrix in Figure 8.

To evaluate the accuracy of our models, we compared each element of the model ratio matrices to the ground truth matrix. Error was calculated as follows:

$$Error = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{x'_{i,j}-x_{i,j}}{x_{i,j}} \tag{1}$$

where $x_{i,j}$ is the ground truth ratio of feature i and feature j and $x'_{i,j}$ is the model ratio of feature i and feature j. This calculation yielded the accuracies listed in 2

Figure 4. Instant-NGP Training and Reconstruction

#	Features	Ground Truth	Instant-NGP	Space Carving
1	Thumb	6.5	0.733	0.974
2	Index	7.5	1.061	1.099
3	Middle	8.7	1.095	1.222
4	Ring	8.2	0.9598	1.120
5	Pinky	6.5	0.7282	1.009
6	Palm height	10.5	1.045	1.696
7	Palm Width	10.5	1.149	1.999
8	Pinky to Thumb Separation	20.9	2.170	3.387
9	Finger Width (Index)	1.8	0.227	0.480
10	Hand Height (Thumbside)	19.2	1.946	2.310

Table 1. Anatomical Feature Measurements

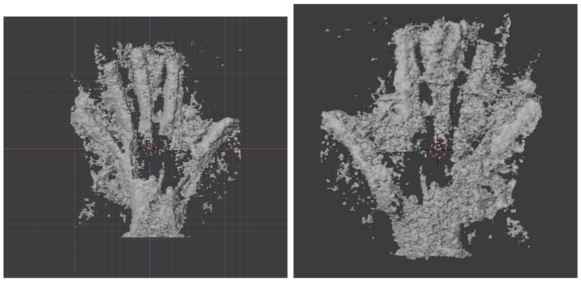
Model	Accuracy(%)
Space Carving	95.61
Instant-NGP	98.98

Table 2. Model Accuracies

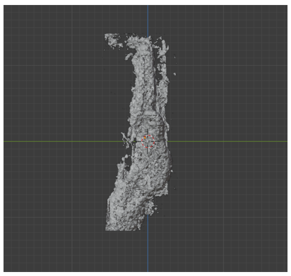
F. Conclusion

We see that both models perform very well in terms of their mean accuracy from the ground truth, with the Instant-NGP model performing exceptionally well with a mean accuracy of 99%. We have shown that 3D reconstructions of anatomical parts can be very accurate compared to the ground truth, potentially allowing for a novel method of modelling prosthetics. Our main limitation was the accuracy of the silhouette projections in the space carving method and so to expand on this project in the future, we would train the space carving method with a larger image set and acquire more precise camera matrices for that reconstruction.

Github Repository: https://github.com/bhofflich/Anatomical-Instant-NGP



(a) Backhand (b) Palm



(c) Thumbside

Figure 5. Instant-NGP Models

References

[1] Chloé Valentine Toscano. Prosthetics are built for disabled people. but they still lack inclusivity. 2022. 1

[2] Nasrul Humaimi Mahmood and Tardi Tjahjadi. 3d reconstruction from multiple views for orthotic and prosthetic design: An overview. pages 70 – 75, 07 2006. 1

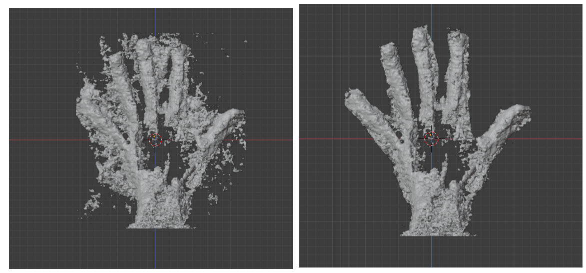
[3] Nasrul Humaimi Mahmood, Camallil Omar, and Tardi Tjahjadi. Multiviews reconstruction for prosthetic design. *Int. Arab J. Inf. Technol.*, 9:49–55, 01 2012. 1

[4] Marco A. Barreto, Jorge Perez-Gonzalez, Hugh M. Herr, and Joel C. Huegel. Aracam: A rgb-d multi-view photogrammetry system for lower limb 3d reconstruction applications. *Sensors*, 22(7):2443, Mar 2022. 1

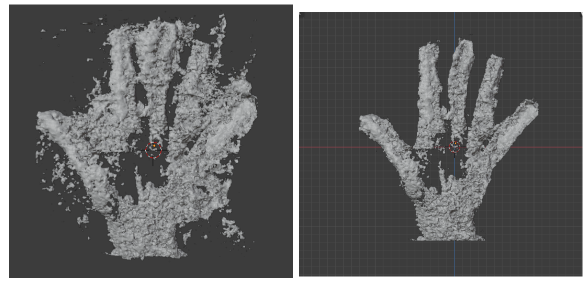
[5] Guang-kun Li, Fan Gao, and Zhi-gang Wang. A photogrammetry-based system for 3d surface reconstruction of prosthetics and orthotics. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2011:8459–62, 08 2011. 1, 2

[6] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 2

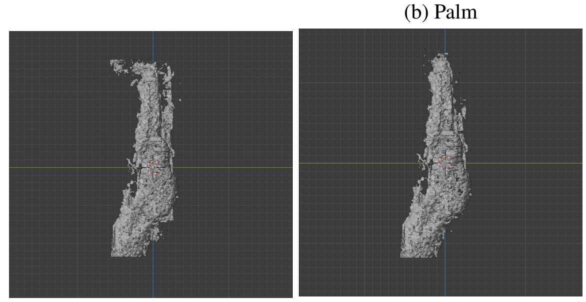
[7] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng.



(a) Backhand



(b) Palm



(c) Thumbside

Figure 6. Cleaned Models

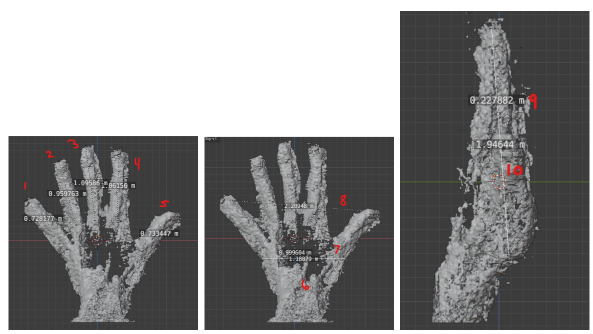
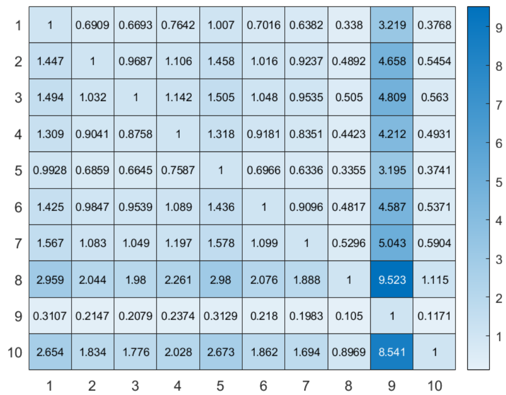


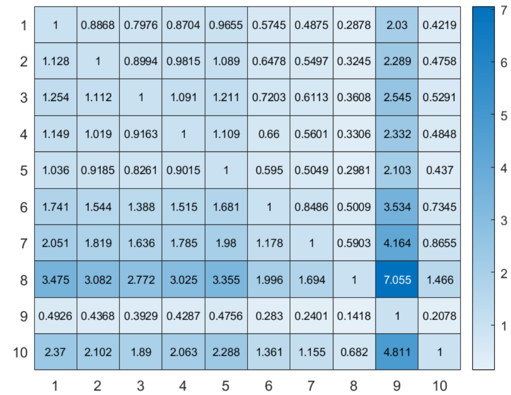
Figure 7. Anatomical Feature Map

Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2

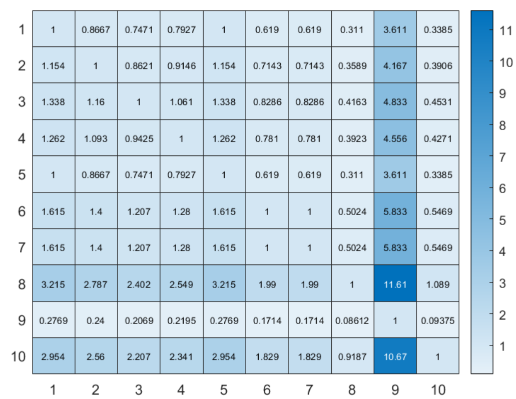
[8] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–



(a) Instant-NGP



(b) Space Carving



(c) Ground Truth

Figure 8. Feature Ratio Matrices

174, 2019. 2

[9] G LoweDavid. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004. 2