FIND: An Unsupervised Implicit 3D Model of Articulated Human Feet - Supplementary Material

Oliver Boyne
ob312@cam.ac.uk
James Charles
http://www.jjcvision.com
Roberto Cipolla
cipolla@eng.cam.ac.uk

Machine Intelligence Lab
Department of Engineering
University of Cambridge
Cambridge, U.K.

1 Dataset details

1.1 Poses and scanning

Index	Extreme 1	Extreme 2	Description
0	T-Pose	-	Neutral pose
1	Plantarflex	Dorsiflex	'Pitch' of the foot
2	Inversion	Eversion	'Roll' of the foot
3	Lateral rotation	Medial rotation	'Yaw' of the foot
4	Toe Flexion	Toe Extension	Toes clenched towards sole (flexion) or lifted upwards (extension)
5	Toe Abduction	Toe Adduction	Toes outwards (abduction) or inwards (adduction)
6	Standing on Floor	-	-
7	Tiptoes	-	-

Table 1: Pose descriptions

Poses. We ask users to form a combination of poses for each scan, from the available degrees of freedom of the human foot. The possible poses are shown in Table 1. These poses are based on foot articulation described in foot anatomy literature [2].

Figure 1: Images taken during the scanning process, showing the placement of the leg on a table, and the suspension of the foot in the air.

Pose label vectors. To form these pose descriptions into a vector for our contrastive loss, we transform each description into an 8-long vector, according to the indices in Table 1. For poses with only one extreme, the value of that vector can be 0 or 1. For poses with two extremes, the value can be -1 or 1. For example, T-Pose sets the 0th index to 1; Plantarflexion sets the 1th index to -1. As a full example, for pose *Dorsiflexion, Inversion, Toe Extension*, the corresponding vector would be $[0,1,-1,0,1,0,0,0]$.

Scanning. We show in Figure 1 an image of the position subjects would sit in while their foot was scanned. We found subjects had minimal difficulty in holding the desired pose for the approximately 2 minutes the scan took, and any slight movements were accounted for by the post-processing software [?].

1.2 Alignment process

Once meshes have been processed in Artec Studio [?], we are left with unaligned meshes, with a plane sliced approximately perpendicular to the leg, at differing heights up the leg. We use this to looseley register the meshes and standardise the cutoff, as follows:

Loose alignment. We assume that the cut-off planes used in the previous stage were precise enough to provide an initial alignment for the feet. We rotate all feet such that this plane is perpendicular to the global up (Z) direction, and that a line connecting the centroid of the cut-off plane and the centroid of the ‘footprint’ (all of the mesh within 5 mm of the bottom) lies along the YZ plane.

Plane slicing. We find the ‘heel’ point by finding the minimum X-wise vertex below the foot’s centroid. Next, we select a height 10 cm above the heel, and slice the foot along the XY-plane at that height. This provides a more consistent cut-off of the foot. To manage the texture on this plane, we simply colour all faces on the new slice plane as white, and set their UV coordinate to (0,0) so that they can be tracked in our differentiable rendering pipeline and masked out.

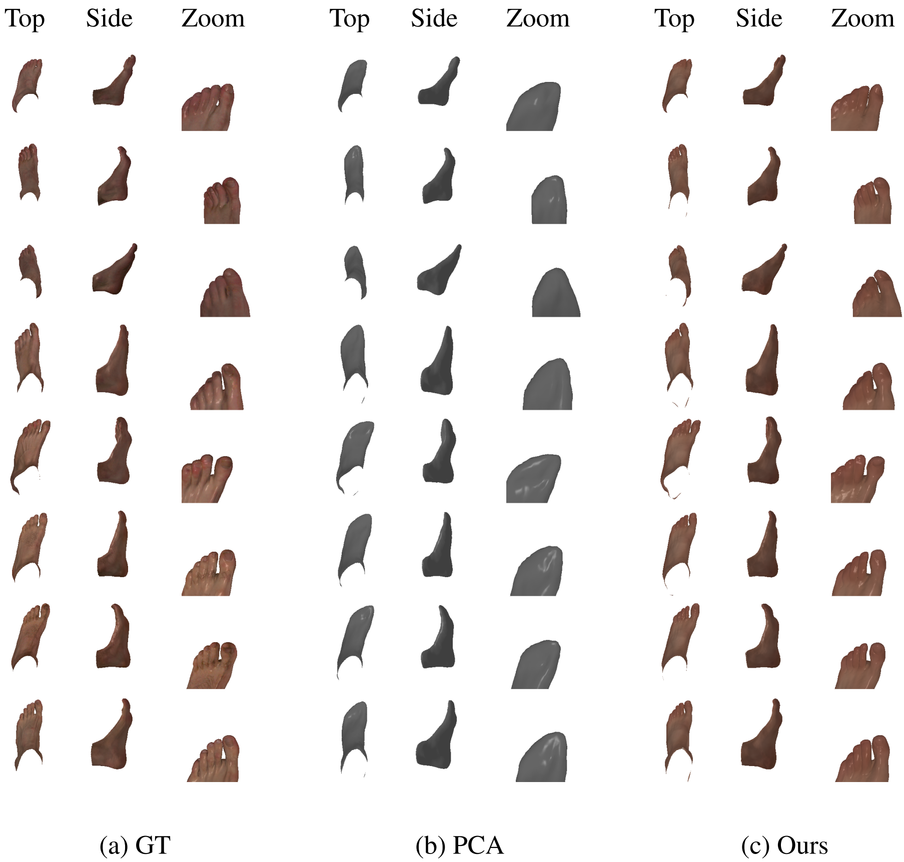
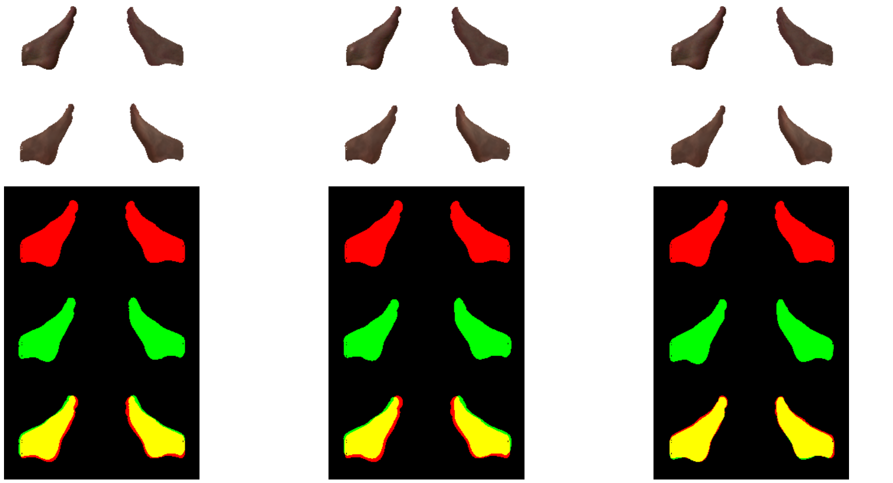


Figure 2: Qualitative results of our 3D fits to 8 validation feet, rendered from 3 views each.



(a) Silhouette only (b) Silhouette + VGG (c) Silhouette + part-based loss

Figure 3: Fits of our FIND model to 2 images of a validation foot. Only 2D losses are used in the optimisation loop. Rows show, from top to bottom, (i) GT images, (ii) Predicted renders, (iii) GT silhouettes, (iv) Predicted silhouettes, (v) Silhouette overlaps.

2 Qualitative results

3D optimisation. We show fitting of the FIND model to 8 validation feet, using a 3D chamfer loss, in Figure 2. We show significant improvement in the fidelity of shape reconstruction, especially noticeable around the toes, as the PCA model is unable to capture the finer details, such as the separation of the big toe. Furthermore, our model is capable of parameterising texture.

2D optimisation. We show fitting of the FIND model to rendered images of our validation feet in Figure 3. We optimise only with 2D losses, and show that the quality of the fit improves with the use of our novel unsupervised part-based loss over silhouette only, or silhouette and an off-the-shelf VGG-16 [3] perceptual loss.

References

[1] Artec Studio. https://www.artec3d.com/3d-software/artec-studio. Accessed: 2022-07-25.

[2] Peggy A Houglum and Dolores B Bertoti. *Brunnstrom's clinical kinesiology*. FA Davis, 2011.

[3] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.