

Supplementary Material

A human mesh-centered approach to action recognition in the operating room

Benjamin Liu¹, Gilles Soenens², Joshua Villarreal³, Jeffrey Jopling⁴, Isabelle Van Herzeele², Anita Rau^{5,#}, Serena Yeung-Levy^{5,#}

¹Department of Computer Science, Stanford University, Stanford, CA 94305, USA.

²Department of Thoracic and Vascular Surgery, Ghent University Hospital, Gent 9000, Belgium.

³Department of Surgery, Stanford University, Stanford, CA 94305, USA.

⁴Department of Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

⁵Department of Biomedical Data Science, Stanford University, Stanford, CA 94305, USA.

#Authors contributed equally.

Correspondence to: Benjamin Liu, Department of Computer Science, Stanford University, 353 Serra Mall, Stanford, CA 94305, USA. E-mail: benliu@cs.stanford.edu

Definitions of Joint Categories

Using individual joints defined in SMPL-X^[1], we created joint categories to represent distinct groups of joints that we deemed useful for recognizing surgical actions. In the “pelvic” joint category, we included the pelvis, left hip, and right hip. In the “arm” joint category, we included the left elbow, right elbow, left wrist, and right wrist. Unlike SMPL, SMPL-X does not explicitly include individual joints for the left and right hands, and we observed that estimations of finger joints were error-prone. Hence, we assumed that the positions of the wrist joints were reasonable approximations of the positions of the hand joints. In the “cranial” joint category, we included the head, jaw, nose, right eye, left eye, right ear, and left ear. In the “thorax” joint category, we included the left shoulder, right shoulder, left sternoclavicular, right sternoclavicular, and neck. In the “leg” joint category, we included the left knee, right knee, left ankle, right ankle, left foot, and right foot. In the “spine” joint category, we included all spine joints defined in the SMPL-X model.

1. Pavlakos G, Choutas V, Ghorbani N, et al. Expressive body capture: 3D hands, face, and body from a single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019. pp. 10975-85. Available from:

https://openaccess.thecvf.com/content_CVPR_2019/html/Pavlakos_Expressive_Body_Capture_3D_Hands_Face_and_Body_From_a_CVPR_2019_paper.html. [Last accessed on 21 Jun 2024]