



Fast Detection of Partially Occluded Humans from Mobile Platforms: Preliminary Work

David Frank • Dr. David Feil-Seifer

Robotics Research Lab, Department of Computer Science & Engineering, University of Nevada, Reno

Introduction

The final goal of this project is to create a robust system for finding the pose, i.e. locations of arms and legs, of a human within a depth image. The system will take depth images as input and use a trained Convolutional Neural Net (CNN) to find the locations of the body segments. It will also make no assumptions about the orientation of the camera relative to the human, so the system will still function when used on cameras in any arbitrary position. In order to train the CNN, large amounts of training data must be created. This data must have depth images with each pixel labeled as one of the body parts. Creating this data manually is a time consuming task. This preliminary work involves creating a system to generate synthetic data suitable for training.

Method

Synthetic data is an effective means of training a system to recognize body parts [1]. By creating a system to quickly generate training data, immense amounts of time can be saved when compared to manually labeling training data. This allows us to create a large data set to train a CNN on. Currently, the freely available image generation application “Blender” is being used. Blender is free to use for any purpose, so the system can be used by collaborators.

The images being generated are of a person in a random pose and a random amount of occlusion. The pose of the person can be any valid position that a real person could take. The occlusion can vary from covering nothing of the person to covering up most of the person. This provides a wide range of training data for the CNN to train on.

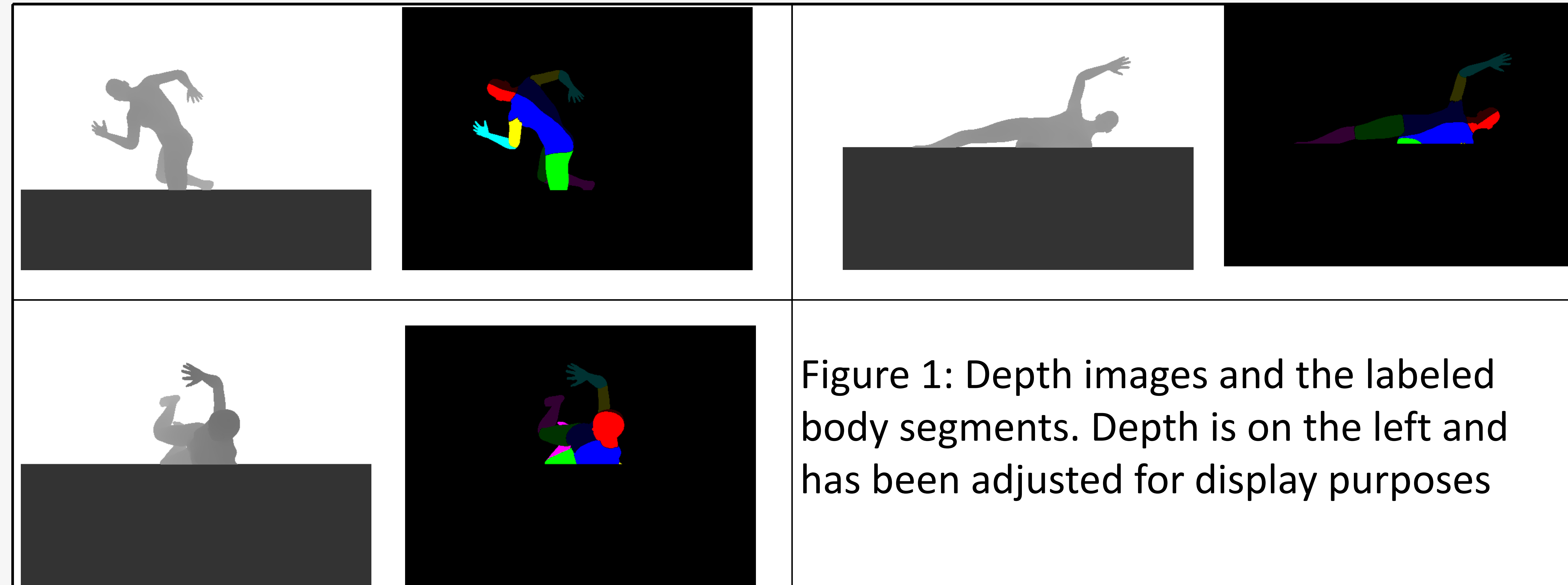


Figure 1: Depth images and the labeled body segments. Depth is on the left and has been adjusted for display purposes

Conclusion

Images generated by the system can be seen in Fig. 1. The system can generate and apply necessary post processing in approximately 1.3 seconds. This allows a large training set to be constructed in a few days.

Future Work

The data will be used to train a CNN to label each pixel in a depth image as a certain body part. Once this is done, the labeled images will be used to estimate the pose of the person. Most applications that rely on pose estimation use a skeletal model of the person, since it is more efficient. Thus the image must be reduced to only a few key points. This is the future work that will be done on the project.

Reference

- [1] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time Human Pose Recognition in Parts from Single Depth Images,” in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2011, pp. 1297–1304

Acknowledgements

This material is based upon work supported by the National Aeronautics and Space Administration under Grant No. NNX10AN23H issued through the Nevada Space Grant. We would also like to thank Dr. Richard Kelley