Calibrationless monocular vision motion capture for drop jump

Ryo Ueno[1], Claire V. Hammond[1], Wan M. R. Rusli
[1]Department of Research and Development, ORGO Inc., Sapporo, Japan
Email: ueno@orgo.co.jp

Summary

Estimating global position of motions in multilevel floor plane is challenging for monocular vision pose estimation methods. This study introduced a technique to fit the ground plane and refine the foot-ground contacts for monocular pose estimation. Monocular vision motion capture results for a drop jump video was demonstrated with accurate foot-ground contacts.

Introduction

Computer vision-based pose estimation technologies are rapidly improving the accuracy and usefulness. Recent pose estimation only requires a monocular video for estimating human body motion in 3D global space without calibrating the camera. However, a motion task which contains multilevel floor plane such as drop jump is challenging for the monocular video pose estimation. This is because many methods assumed that the motion is on a single floor plane to reconstruct smooth motion by utilizing ground contacts probability [1]. The other pose estimation methods without this assumption often fails to have accurate foot contacts to the ground and causes foot-skating and foot floating/penetration error [2]. To address this problem, this study demonstrated a non-machine learning technique to fit the ground plane and refine the foot-ground interaction for a monocular vision pose estimation.

Methods

A previous pose estimation method [2] was used as the base method. The base method utilized SMPL model and reconstructed accurate relative global translation and rotation of the root joints in camera coordinate system but did not consider the ground plane and foot-ground interaction in global coordinate system. To fit the ground plane to the estimated motion, it was assumed that an average of all the vectors from foot contact points to the whole-body center of mass (COM) would replicate the normal vector of the ground. SMPL joint position data was used to calculate the COM position. Foot contact states were detected if the contact points attached to the feet were lower than a height threshold (< 0.1 m) and slower than a velocity threshold (< 1.0 m/s). Vectors from each contact point to COM were averaged over the time. This averaged vector was considered as the normal vector of the ground. A Rotation matrix was derived from the normal vector and applied to SMPL root translation and rotation to transform the motion data in the global coordinate system. The lowest position coordinate of SMPL vertices in all time was subtracted from the position data to eliminate the offset in vertical axis.

After fitting the ground, an optimization was performed to refine the foot-ground contacts. Pytorch LBFGS function was used to optimize SMPL root translation and poses to minimize the velocity and height error of the foot contact points when they were in contact states while discouraging the large changes from the original motion.

A drop jump video from OpenCap dataset [3] was used for the evaluation of the developed technique. Toe position error in horizontal plane (foot skating) and vertical axis (foot floating/penetration) during contact phase was evaluated before and after the optimization.

Results and Discussion

Figure 1 shows the qualitative results of estimated marker positions from SMPL vertices. The foot position error in horizontal plane before the optimization was 128.1 mm and decreased to 19.3 mm after the optimization. Foot height error was 42.2 mm before the optimization and decreased to 26.4 mm (Figure 1).

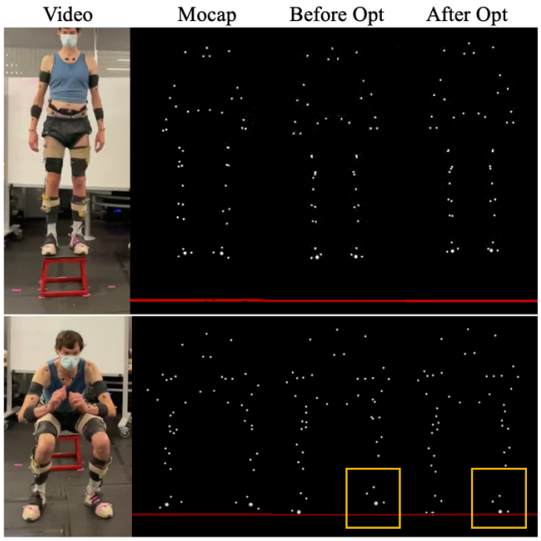

Figure 1: Optimization recovered natural foot-ground contacts.

The developed non-machine learning technique is generalizable to many other pose estimation methods and motion tasks. This encourages the development of future monocular vision motion capture system for biomechanical analysis.

Conclusions

Drop jump motion was estimated in 3D global space from a monocular video with accurate foot-ground contacts using a developed technique.

References

[1] Shin S et al. (2024). *CVPR*: 2070-2080.
[2] Wang et al. (2024). *ECCV*: *467 - 487*.
[3] Uhlrich SD et al. (2023) *PLOS Comput Biol*, 19(10)