3D Human Pose Analysis of Sport Climbing

Cristian-Alexandru Tanase
AI Multimedia Lab
National University of Science and Technology
POLITEHNICA Bucharest
Bucharest, Romania
cristian.tanase2601@upb.ro

Bogdan Ionescu
AI Multimedia Lab
National University of Science and Technology
POLITEHNICA Bucharest
Bucharest, Romania
bogdan.ionescu@upb.ro

***Abstract*—Sport Climbing 3D Human Pose Estimation represents a yet under-explored area of science. More recently, the number of publications regarding sport climbing started to grow and there is much more interest in studying this popular sport. A lot of information can be obtained from researching into both 2D and 3D pose estimation of climbers. While it is challenging due to the heavy occlusions caused by both the climber's body and the climbing wall, successful estimation can prove useful in injury prevention and increasing performance of climbers. Another huddle commonly observed is the back-facing position of the climber on the wall. Most current 3D Human Pose Estimation models are trained on datasets that are mainly comprised of individuals facing the camera or observing system. Besides the initial inference that was realsed using ViTPose 2D pose and MotionBERT for obtaining a 3D pose estimation we plan on firstly fine-tuning this ViTPose-MotionBERT pair before looking into other options. Further biomechanical analysis will be done after fine-tuning the model to obtain an acceptable mean per join position error, metric that represents the accuracy of 3D Human Pose Estimation systems.**

***Index Terms*—sport climbing, 3D Human Pose Estimation, Artificial Intelligence, transformer, convolutional neural networks**

I. Introduction

Following its addition to the Olympic games in Japan, sport climbing has caught the attention of the scientific community. It is broken down into three disciplines, bouldering, lead climbing and speed climbing, each of them posing a different challenge. At the Paris Olympic games, speed climbing has been separated from the other two disciplines because of how different it is. Scientific publications on the matter of Sport Climbing injuries and performance are blooming in the latest years [1]. There are initial studies such as a climbing dataset for training 3D human pose algorithms [1] and force development studies. AI is widely used for 3D human pose with models such as Motion Bert [2] that requires the 2D pose output of a model like Alpha Pose or ViTPose [3] to generate a 3D pose. These models are based on the human 3.6M or MS Coco datasets, where usually the subject is facing the camera [4] or the actions that occur do not include heavy occlusions [1]. Such occlusions frequently occur during climbing. Biomechanical analysis has been done in some isolated situations in climbing [5], but not paired with 3D human pose.

1st Workshop on Artificial Intelligence for Multimedia,
AI Multimedia Lab, CAMPUS Research Institute, POLITEHNICA Bucharest,
Bucharest, Romania, 8 November, 2024

II. 3D Human Pose Estimation

With the advancement of neural networks and transformer architectures, more accurate human pose estimation algorithms have been developed. Such algorithms are based on different architectures [6], with a common denominator, their output. The output that is expected out of 2D or 3D pose networks is an estimation of joint and body position. This estimations is useful to analyse diseses that are related to gait, performance of athletes or movements that are prone to cause injury.

Some of the most popular 2D human pose estimation networks are OpenPose, AlphaPose, ViTPose and RTWPose [7]. These models are all based on different architectures. Open Pose is based on a key 2D pose estimation architecture using a multi-stage convolutional neural network (CNN) that predicts confidence maps for keypoints and part affinity fields for limb connections [6]. It focuses on both keypoint detection and limb association. RTWPose and AlphaPose are also CNN based approaches, they use a HRNet backbone for feature extraction and realise whole-body pose estimation including hand and foot keypoints. Unlike the others, ViTPose is based on the vision transformer (ViT) architecture. It uses transformers to capture global relationships between keypoints [3].

The above networks have and can be trained on the MSCOCO and MPII datasets for human pose estimation [8]. The evaluation of such networks is done using several metrics such as average precision (AP), particularly AP@OKS (Object Keypoint Similarity). Another metric relevant to keypoint localisation methods is mean per joint position error (MPJPE) that evaluates the average euclidean distance between predicted and ground truth keypoints, with lower values predicting better precision. Some models are based on heatmaps, the relevant metric for these is normalized mean squared error (NMSE), quantifying the difference between predicted and actual heatmaps. Other general metrics are inference speed, in frames per seconds (fps) and model complexity.

Obtaining a good 2D human pose estimation is important when feeding it into a 3D Human Pose Estimation network that will take as inputs both the original RGB image or video and the corresponding keypoints obtained from the 2D pose. A commonly used 3D pose system is MotionBERT, the primary metric is represented by MPJPE. Anothe metric commonly used for this is procustes aligned MPJPE (PA-MPJPE), which

Fig. 1: ViTPose 2D HPE Inference Result

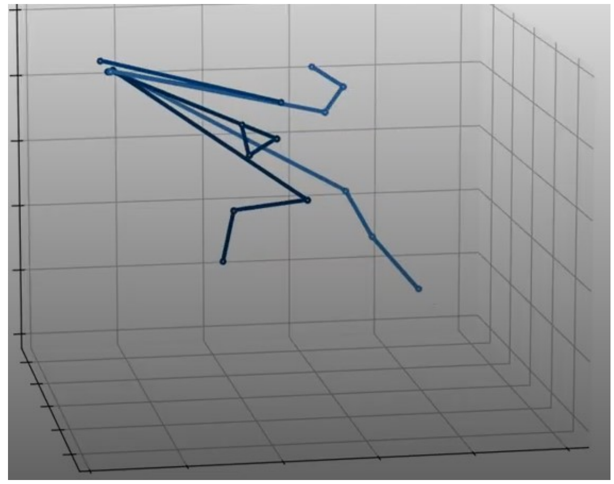

Fig. 2: MotionBERT 3D HPE Inference Result

aligns predictions to ground truth.

III. Biomechanical Analysis

Biomechanical analysis is well known for identifying lots of issues of the human body. Many of the studies on this topic are usually done using a simulation model that also has moments and forces built into it. These models can very well break down how forces affect the body during gait or movement. Gait requires such models to include ground contact and forces [9], for climbing this would be much more complex since climbing would not only require ground forces but the wall contact would need to be designed. In this study we will just focus on the join angles and lengths that are provided from the 3D pose system.

IV. Model architecture and development

The proposed model architecture for this initial study would comprise of a 2D human pose estimation model,selected depending on its performance from the ones listed above, paired with a 3D human pose estimation model. Both models require fine tuning. For this fine tuning we will use the CIMI4D dataset that is a multimedia dataset containing labeled data from both RGB monocular view and RGB multi-camera view. The dataset also contains LIDAR data that has been transformed to 2D RGB that could be used for training. After fine tuning we will be using the network to properly identify injury prone movement and if it is possible to increase climbing performance by using 3D Human Pose Estimation.

V. Initial Results

The 2D ViTPose network created a decent result on the short climbing RGB video. While most of the time the joint locations are correct, when heavy occlusions occur, hand joints tend to heavily lose accuracy this inference can be observed in Figure 1. The 3D result from inferencing MotionBERT together with the 2D pose realised by ViTPose proved faulty because of the lack of face key points. To be able to inference MotionBERT the face points values were replaced with [0,0,0] values leading to having them shown in the origin point of the figure, as observed in Figure 2.

VI. Conclusion

While an architecture based on a 2D human pose estimation resulted from ViTPose or Alpha pose and input into Motion-Bert can be quite accurate in normal conditions, it's proven difficult to obtain a qualitative inference in a sport climbing scenario without fine tuning the respective networks using a proper dataset. Future work will start with fine tuning both the 2D pose model and the 3D pose model to fit the aplication at hand.

There are multiple possible directions for studying different areas of sport climbing and for choosing what the optimal model would be for each task. Some models only limit to the body and general limbs while others do take in account the fingers also. The latter may be useful in studies that go much more in depth for specific muscle and joint groups.

References

[1] M. Yan, X. Wang, Y. Dai, S. Shen, C. Wen, L. Xu, Y. Ma, and C. Wang, "Cimi4d: A large multimodal climbing motion dataset under human-scene interactions," *arXiv preprint arXiv:2303.17948*, 2023, [Online]. Available: https://arxiv.org/abs/2303.17948.

[2] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, and Y. Wang, "Motionbert: A unified perspective on learning human motion representations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[3] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, "Vitpose: Simple vision transformer baselines for human pose estimation," in *Advances in Neural Information Processing Systems*, 2022.

[4] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1325–1339, 2014.

[5] S. Kozin, D. Safronov, Z. Kozina, H. Kniaz, O. Proskurnia, K. Prontenko, O. Lahno, V. Goncharenko, and A. Kholodniy, "Comparative biomechanical characteristics of one-arm hang in climbing for beginners and qualified athletes," *Acta of Bioengineering and Biomechanics*, vol. 22, no. 1, pp. 57–66, 2020.

[6] R. J. Cotton, A. DeLillo, A. Cimorelli, K. Shah, J. D. Peiffer, S. Anarwala, K. Abdou, and T. Karakostas, "Markerless motion capture and biomechanical analysis pipeline," *arXiv preprint arXiv:2303.10654v1 [cs.CV]*, 2023.

[7] I. Masmoudi, “Tracking and feedback in surfing: A 3d pose estimation approach,” Master’s thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2024.
[8] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, “Microsoft coco: Common objects in context,” *arXiv preprint arXiv:1405.0312*, 2015.
[9] M. Gambietz, M. Nitschke, J. Miehling, and A. D. Koelewijn, “Contributing components of metabolic energy models to metabolic cost estimations in gait,” *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 4, pp. 1228–1236, 2024.