Automated Human Movement Segmentation by Means of Human Pose Estimation in RGB-D Videos for Climbing Motion Analysis

Raul Beltrán B.[a], Julia Richter[b] and Ulrich Heinkel[c]
Professorship Circuit and System Design, Chemnitz University of Technology, Reichenhainer Straße 70, 09126 Chemnitz, Germany

Keywords: Climbing Motion Analysis, Movement Segmentation, Human Pose Estimation, Video Analysis.

Abstract: The individual movement characterization of the human body parts is a fundamental task for the study of different activities executed by a person. Changes in position, speed and frequency of the different limbs reveal the kind of activity and allow us to estimate whether an action is well performed or not. Part of this characterization consists of establishing when the action begins and ends, but it is a difficult process when attempted by purely optical means since the subject's pose in the image must first be extracted before proceeding with the movement variables identification. Human motion analysis has been approached in multiple studies through methods ranging from stochastic to artificial intelligence prediction, and more recently the latest research has been extended to the sport climbing employing the centre-of-mass analysis. In this paper, we present a method to identify the beginning and end of the movements of human body parts, through the analysis of kinematic variables obtained from RGB-D videos, with the aim of motion analysis in climbing. Application tests with OpenPose, PoseNet and Vision are presented to determine the optimal framework for human pose estimation in this sports scenario, and finally, the proposed method is validated to segment the movements of a climber on the climbing wall.

1 INTRODUCTION

With the increasing accessibility to devices for recording and analysing people and objects in the 3-D space, through image processing and artificial intelligence (AI), every day more products appear that provide us with real-time information about our activities. An example of this is the video processing technology in real-time applied to sport, which makes it possible to give online feedback to the athletes by simply recording exercise sequences on their smartphones and then analysing them on the spot using an application. In this type of application, a fundamental process consists of extracting the human figure, determining the pose, and finally characterizing the movement. Each of these phases is a matter of research, which has either been approached individually (Xiaohui et al., 2018; Khuangga and Widyantoro, 2018; Zheng et al., 2020) or jointly using AI (Fan et al., 2017; Cao et al., 2017; Papandreou et al., 2018).

Human Pose Estimation (HPE) is a trending solution that AI offers to determine the position and orientation of a person's body in a given image. While there is already an acceptable level of precision in 2-D pose estimation, in many scenarios, the 3-D case still requires more work to produce accurate models with data fusion techniques, which is a challenging task. In this area, sport climbing draws attention not only because of the widespread use it has had in recent years, but also due to the challenges it implies for the recognition of human postures.

The characterization or classification of human movement according to kinematic variables such as displacement and distance, velocity and speed, acceleration, and time, requires segmenting the motion observation sequences into smaller components, called motion primitives. It is a principal task to describe or analyse the execution of human activities, to facilitate the identification, modelling and learning of movement (Lin et al., 2016). The climbing action is divided into phases, usually composed of movements, in which these can be segmented. e.g. Firstly, one hand reaches for a hold, then the feet are placed, and finally, the climber stands up to grab the next grip

[a] https://orcid.org/0000-0001-6612-3212
[b] https://orcid.org/0000-0001-7313-3013
[c] https://orcid.org/0000-0002-0729-6030

Beltrán B., R., Richter, J. and Heinkel, U.
Automated Human Movement Segmentation by Means of Human Pose Estimation in RGB-D Videos for Climbing Motion Analysis.
DOI: 10.5220/0010817300003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 5: VISAPP*, pages 366-373
ISBN: 978-989-758-555-5; ISSN: 2184-4321

with the other hand. Then, it is especially important to determine when each of these phases starts and ends to analyse individual movements, adding that specific movement errors can occur in each of these phases. After the climber has reached the new hold and replaced the feet, for example, this very arm has to be kept straight to save energy within the same phase of the climbing action. To evaluate and determine if there is error, the segments of the whole climbing sequence have to be obtained.

Our research has as its main objective the movement segmentation for the hands, feet, and waist using the 3-D skeleton joints of a climber recorded in an RGB-D video through an optical device. Thus, a mechanism is provided to model individual climbing movements using the estimated pose in 2-D images and its projection in 3-D using the point cloud delivered by the device. Our contributions are included in the first phases of the climbing analysis, namely in the selection of a suitable HPE framework and the application of techniques for segmentation of the movement primitives.

The paper is structured as follows: Section 2 presents extant works in the research areas related to the present document. Thereupon, Section 3 explains the methods to carry out the data collection, model construction, and information analysis. Next, in section 4, the results are exposed and discussed. Finally, conclusions of the overall work are presented and future work in the climbing analysis is outlined.

2 RELATED WORKS

This paper focuses on the field of climbing motion analysis and is linked to the segmentation of human motion primitives. Recent work in these two research areas is presented below.

2.1 Motion Primitives Segmentation

In extant studies, human motion and human posture has been commonly modelled by means of dynamic systems and approximations with stochastic methods to carry out temporal and dimensional segmentation of the body parts associated with the displacement (Kulić et al., 2011). Other works in the same direction as Meier et al. (2011) have reformulated the problem from the perspective of trajectory recognition in parameterized libraries of motion primitives, an approach that is valid for the study of specific locomotion activities. With the rise of AI in the last decade, projects of movement primitives segmentation have emerged (Lin et al., 2016; Vögele et al., 2014), but as proposed by Lin et al. (2016), from the definition of what a segment is to how the data is collected, make each solution to the problem have an application to specific requirements. These studies, like others (Jansen et al., 2019; Aoki et al., 2016), have used inertial measurement unit (IMU) sensors attached to the joints of the body to collect the position and velocity data of the limbs, obtaining consistent signals that facilitate the analysis. Cutting-edge investigations (Zago et al., 2020; Colyer et al., 2018) using marker-less sensors, have been facing problems such as body parts occlusion and the quality of the gathered data, requiring multiple sensors at different viewpoints to improve data quality, or data virtualization to predict unknown poses with predefined kinematic models. There are also investigations (McCay et al., 2020) in this sense where the posture analysis is carried out from RGB-D images, to train neural networks and translate them into a classification problem.

2.2 Climbing Motion Analysis

In the field of sport climbing, there are recent studies to analyse the trajectory of the climber's centre-of-mass (CoM) through optical devices, such as Richter et al. (2020a), where information on the fluency, force, and distance to the wall is acquired to provide the climber with information to prevent possible injuries from a therapeutic point of view. Richter et al. (2020b) moreover provide a profound survey on climbing motion analysis. Cha et al. (2015) analyse postures and movements also employing optical devices, but with an orientation to the construction of 3-D graphic animations. Others like Seifert et al. (2014), using IMUs attached to the climber body, attempt the recognition of climbing patterns using cluster analysis to process the position of the limbs and waist of the subject. Nevertheless, these studies that involve wearables are dedicated to laboratory analysis, since they have the difficulty to be transferred to applications of daily use due to the cost of the implements (Jansen et al., 2019) and how cumbersome these accessories can be for climbers.

3 METHODS

In this section we first introduce a segmentation concept for our climbing scenario, after which we proceed with the explanation of the steps taken to collect, process and analyse the information.

3.1 Segment Definition

As in other sports, climbing movements aim to reduce effort and improve performance, seeking to solve a task while saving effort. Climbing is an acyclic sport with three phases: preparation, continue-reaching, and stabilization, usually with a combination of movements in a single phase (Winter, 2012). The period in which each of these movements is executed is what we define here as a segment. Figure 1 shows an example of a sequence of climber movements, where the speed changes indicate when the climbing actions begin and end.

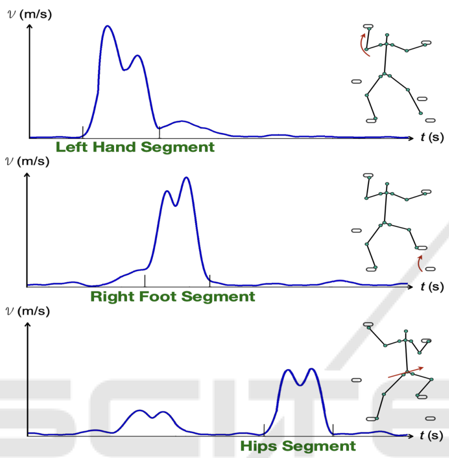


Figure 1: Schemes of the temporal segmentation of climbing movements.

For movement analysis, we took five target joints from the skeleton data: the climber's wrists and ankles, plus the hip. For each joint an independent discrete signal was constructed with the measured velocity at each frame. Thus, the movement of one hand to the next grip, the rearrangement of the feet or the hip's displacement when standing up on the supporting leg, for example, can be identified as peaks in the speed signal. Considering that in the obtained signal, detections of interest consist of several consecutive peaks, the analysis must use the signal envelope or its cumulative value to find the local maximums. To rule out small peaks due to jittering in the skeleton joints localization, we decided to use the cumulative value of the signal where sustained slopes can be seen when a long movement is executed.

3.2 Movement Segmentation Procedure

Figure 2 depicts the overall process followed to achieve the movement segmentation through a 3-D optical device, which is described in detail in the following.

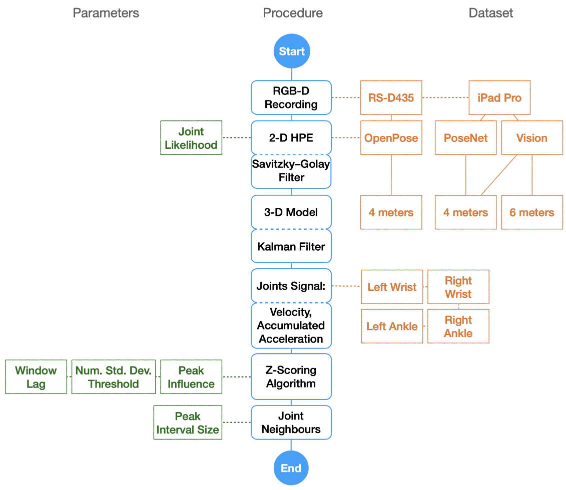


Figure 2: Block diagram of the movement segmentation algorithm.

3.2.1 RGB-D Video Recording

Our study was conducted on RGB-D videos captured with two different devices. Initially, an Intel RealSense D435 camera was used, with which video samples were taken at 4 m from the climbing wall with a resolution of 848×480 pixels at 30 fps, obtaining a density in the point cloud of one point per pixel. Subsequently, an iPad Pro 12.9-inch 4th Generation was used to record videos of 1440×1920 pixels at 60 fps, at 4 m and 6 m from the wall, with a density of one point per each 8,62 and 6,12 pixels respectively. The distances to the climbing wall depended on the sensor used, considering that the entire wall should fit within its angle of view.

3.2.2 2-D Human Pose Estimation

The pose detection in the RealSense (RS) videos was carried out by means of OpenPose (Cao et al., 2019), a real-time multi-person detection library capable of jointly detecting human body, face, and foot keypoints. For the iPad case, the PoseNet (Papandreou et al., 2018) framework was tested first with a Machine Learning (ML) model developed for iOS in TensorFlow Lite ; however, better results were obtained when using the Vision framework for HPE built into the device SDK provided by Apple Inc.

In the skeleton model obtained, fluctuations between good and bad joint detections translate into a high rate of jittering in the position of the recognized body joints. In our work, we reduced the rate of this jittering by applying an implementation of the Savitzky-Golay filter algorithm (Savitzky and Golay, 1964), whose principle is the calculation of local polynomial regression to determine the new value of each non-conforming point.

3.2.3 3-D Model Construction

The estimation of the third coordinate for the 3-D skeleton joints was performed in the post-processing phase, using the collected data from the device and the Point Cloud Library (PCL).

The RS-D435 device includes an active infrared (IR) stereo vision sensor to capture the depth of the scene, producing point clouds in a modified Rosbag file accessible through the camera's SDK . In the iPad's case, the device uses a LiDAR scanner that performs depth-sensing with the help of its pro cameras, motion sensors, and the GPU, such that the more it scans an area, the more details are resolved. The iOS SDK allows communication through Shaders with the GPU memory and thus to control the delivery density of the point cloud in each frame, to the detriment of the available RAM; for this reason, a threshold is introduced between the duration of the video and the desired density of the point cloud. We choose this threshold based on the distance to the climbing wall, carrying out tests from 4 m and 6 m as shown in Table 1.

Table 1: Values for the suitable point cloud size. Low values for iPad come from the threshold between video length, PCL persistence time, and available memory on the device; having a maximum of 49.192 depth points with an individual confidence level.

Device	Distance from Wall	Image Size (pixels)	Point Cloud	
			Size ($\frac{\text{points/}}{\text{frame}}$)	Density($\frac{\text{pixels}^2\text{/}}{\text{point}}$)
RS-D435	4 m	848x480	407.040	1
iPad Pro	4 m	1440x1920	37.288	8,6
iPad Pro	6 m	1440x1920	74.576	6,1

Determining the depth of the body limbs presents a problem when the point cloud density is low, then the distance to the elbows, hands, knees, and feet often coincides with the climbing wall. As a solution, we use the Kalman filter (Kalman, 1960) to predict the correct distance and thus reduce the jittering produced in the z-coordinate of the skeleton joints.

3.2.4 Joint Signal Construction

For the analysis, the joint's position along the entire climbing route is recorded independently of the other joints. Hence, the velocity and acceleration at the i-th frame of the video are given by:

$$v_i = f\sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2} \quad (1)$$

$$a_i = \frac{1}{2}\sqrt{\left(\frac{v_{xi}^2 - v_{xi-1}^2}{x_i - x_{i-1}}\right)^2 + \left(\frac{v_{yi}^2 - v_{yi-1}^2}{y_i - y_{i-1}}\right)^2 + \left(\frac{v_{zi}^2 - v_{zi-1}^2}{z_i - z_{i-1}}\right)^2} \quad (2)$$

where f is the sampling frequency depending on the recording device, 30 Hz or 60 Hz. The series of n values for all frames of the video sequence constitute the discrete velocity and acceleration signals that model the movement of the climber's limbs. As shown in equation 3, to reduce the noise influence and facilitate the detection of the joints state, a low-pass moving average filter was applied on v, with a period of $f/3$ samples.

$$V[i] = \frac{1}{M}\sum_{j=1}^{M-1} v[i+j], \quad with \quad M = f/3 \qquad (3)$$

As preconditions were defined: *i)* The start and end of a motion segment is the moment when the climber is at rest, the hands gripping the holds and the feet firm on the grips. *ii)* The minimum window for a detection movement is 1/3 s ($f/3$), considering that below this limit, climbing movements occur between close grips and are too short for our analysis. *iii)* The study is conducted on the climber's extremities and their CoM, so the joints involved are the subject's wrists, ankles and hips. *iv)* In the 2-D case, the z-coordinate is zero.

3.2.5 The Z-Scoring Algorithm

The peaks in the signal were located employing an implementation of the z-score algorithm, a procedure based on the principle of dispersion, which identifies as local maxima those data points that are within a certain number of standard deviations (σ) from the mean (μ) of a moving window. This procedure uses three parameters, the window lag, the threshold or number of σ's at which the peak is marked, and the influence of the peak on the μ and σ. To illustrate the technique employed, Figure 3 shows in dark blue the velocity signal related to the left wrist of a climber, as schematized in Figure 1. The algorithm identifies the movements as peaks in the velocity signal, so there are four detections around frame numbers 160, 264, 656 and 852. This method detects significant variations in the signal, leaving out those sudden peaks considered as noise, such as occurs in the area of frame 372.

3.2.6 Joint Movement Intervals

Through testing, we determined that for long or sustained climbing movements, the cumulative acceleration works better to identify the movement peaks than the flat velocity. Figure 4 shows the cumulative acceleration of the wrist joint analysed in figure 3, there the four significant movements detected along the signal were transformed into slopes when we had long and

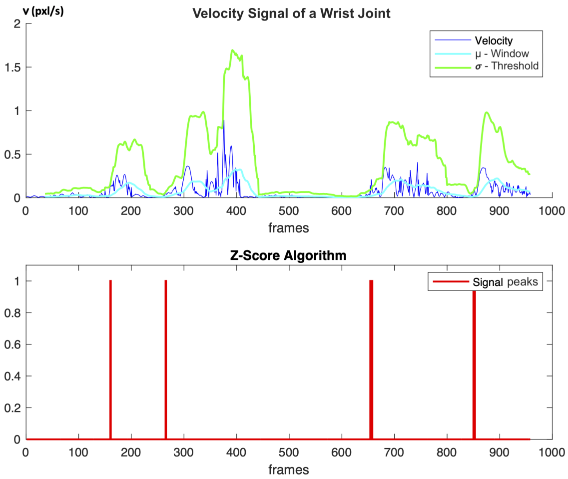


Figure 3: Application of the z-score algorithm on the wrist velocity signal with lag window 35 and σ threshold 6.

sustained movements, while the relatively short and fast ones were interpreted as steep slopes. This is the case for the limbs when they are occluded and the uncertainty in position is reflected as jittering in the signal, as both images show around frame 372. Frame intervals were created around the peaks obtained in the accumulated acceleration signal, and only those peaks whose intervals intersect adjacent ones were taken. This process provided us with a series of frame intervals where the joint showed consistent changes in velocity, i.e. a starting point with increasing acceleration and an end point with the decreasing magnitude. These local maxima indicate when the limb of the body is in motion and are therefore directly related to the segmentation of the movement sought.

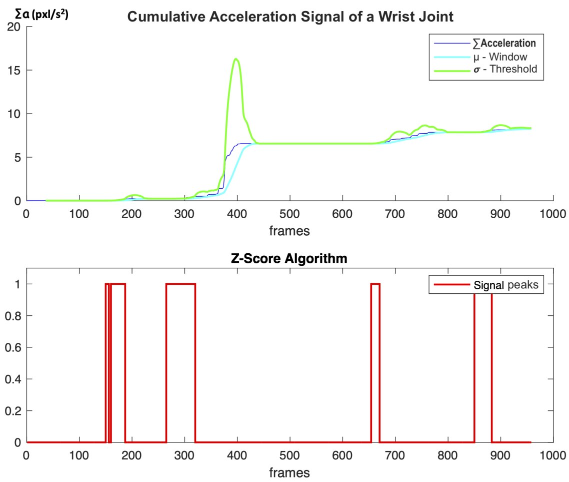


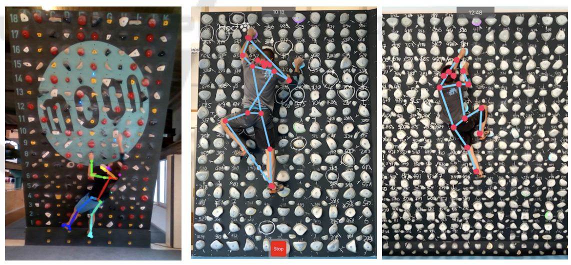
Figure 4: Signal analysis in 2-D of a wrist joint obtained from an 16-second video sequence at 60 fps, and application of the z-score algorithm to cumulative acceleration.

4 RESULTS AND DISCUSSION

This section presents an evaluation of the quality of data delivered by the skeleton extraction algorithms used, followed by a contrast of results using pure 2-D images against the complement with 3-D information.

4.1 HPE Framework Selection

In images recorded with the RS-D435 camera, OpenPose results are notably good in most positions taken by the climber. The skeleton extraction for people from behind is precise and in the climbing scenario, it is possible to detect particular climber poses when the occlusion of the extremities is not very pronounced. In addition, OpenPose offers the location of the hands and feet, which is important when it comes to analysing how effective the climber's pose is before making a move (Kosmalla et al., 2020). Nevertheless, the algorithm has problems by detecting false hands and feet positions when they are confused with shapes of the holds on the climbing wall, especially when lighting conditions or the colour of the grips make it difficult to differentiate them from body parts, see Figure 5a. It is relevant to mention here that OpenPose has a licence restriction that strictly prohibits its use for sports for commercial purposes (CMU, 2019); additionally, the annual cost of this licence makes it awkward to implement outside the research field.



(a) (b) (c)

Figure 5: Pose estimation by three different frameworks: (a) Hold misidentified as a right hand by OpenPose in an RS image. (b) Twisted skeleton by OpenPose trying to fit the subject in front. (c) Correct skeleton detection by Vision at 6 m from a wall tilted at 25°.

For the sequences recorded with iPad, PoseNet allowed the extraction of the skeleton with an accuracy of up to 53 % in each video. Index calculated with the number of false orientations of the skeleton, in 12 videos of 19,8 s on average at 60 fps (14.270 images approximately), plus the number of duplicate joints as a result of partial twists of the skeleton detected, see Figure 5b. This does not include false detections of limbs hidden by the climber's body, which are difficult to count algorithmically, hence the percentage of good detections could decrease as a function of the time the climber covers a limb. The problem with PoseNet lies in the Microsoft COCO keypoint dataset (Lin et al., 2014) used to train the algorithm , where the content does not provide enough knowledge for the algorithm to recognize climbers on a climbing wall. After correcting the body limbs' laterality through heuristic rules the result reached a 67 % of effectiveness in the detection of the poses.

Tests with the Vision framework were successful with 93 % effectiveness, obtaining the skeleton without the need for orientation corrections as Figure 5c shows, and applying the same metric used to evaluate PoseNet as shown in Table 2. Detection difficulties occur when the subject occupies less than 1/3 of the overall image height, as recommended by the documentation. The detections are then confused with the adjacent holds just like OpenPose. A curious situation happens when a climber has hair styled like a ponytail, then the algorithm detects this kind of hair-knot as if it were the subject's nose and tries to rotate the skeleton.

Regardless of the algorithm used for the pose estimation, they all provide a likelihood in the detection of each skeleton joint. By averaging the individual probabilities of all the detected joints per frame, to assign a global certainty to the climber's skeleton, it could be observed that, in the case of OpenPose, the detections are generally made with a certainty of 67 %. While with PoseNet and Vision this value is 68 % and 82 % respectively. Although this calculation is not conclusive, it does make it possible to identify the certainty as an inherent parameter of the algorithm employed, which cannot be used as cross-sectionally indicator.

Table 2: Effectiveness comparison of the three tested pose detection frameworks. OpenPose ran only on the RealSense device without rotation problem, while PoseNet and Vision were able to run on the iPad.

Algorithm / Device	Joints	RS-D435	iPad Pro		
		Average Certainty	Average Certainty	Back Detection Raw	Back Detection Rot. Corrected
OpenPose	18	67%			
PoseNet	17		68%	53%	67%
Vision	19		82%	93%	94%

4.2 Algorithm Evaluation

The climber's movement segmentation algorithm was evaluated in six scenarios, which are derived from the use of three groups of videos analysed in 2-D and 3-D, see Table 3. The first video group made up of 24 recordings created with a RealSense D435 camera, 4 m away from the climbing wall, for which HPE was done using the OpenPose framework. The second and third video groups were recorded on an iPad Pro 4th Generation, at 4 m and 6 m from the climbing wall respectively, using the device's framework for HPE. On the other hand, the six parameters required for the algorithm execution were calculated in advance for each scenario before executing the evaluation.

Table 3: Specifications of the three datasets used to tuning the segmentation algorithm.

Device	Distance from Wall	Total Videos	Average Time	Frame Rate	Aprox. Images	Dataset
RS-D435	4 m	24	16,3	30	11.736	*RS_4m*
iPad Pro	4 m	12	19,8	60	14.270	*iP_4m*
iPad Pro	6 m	14	25,6	60	21.538	*iP_6m*

The ground truth for the evaluation of the algorithm was constructed manually by observing each of the climber's movements in the different sets of videos, taking time measurements for the actual segments where a movement of each limb was observed. To classify True Positives, an intersection of at least 70 % between the interval lengths of the detected and the expected segment was considered; otherwise, it was treated as False Negative.

In the video group of the RS-D435 camera, both in the 2-D and in the 3-D cases, the skeleton jittering could be significantly reduced using the Savitzky–Golay filter. Relying only on those skeleton nodes whose likelihood was greater than 65 % and 76 %, respectively. Given that the point cloud density was higher there, the calculation of the z-coordinate for the joints presented less abrupt variations. Thus, as shown in Figure 6, the detections of the movement of the climber's limbs, in both scenarios remained similar. However, rapid movements between adjacent holds were discarded by the algorithm. It may be attributed to the fact that they occurred within the sliding window of the z-scoring algorithm, within time lower than 1/3 s, or due to no intersections of the 10-frame interval around the detected peak were found. In general, the detections were made with a probability of 73,74 % in 2-D and 69,51 % in 3-D, where part of the fails can be attributed to the fact that many skeleton joints below the certainty threshold were discarded to avoid signals from other doubtful nodes.

For the iPad videos, the results differed notably between the 2-D and 3-D cases. On the one hand, in the 2-D detections, the jittering presented in the skeleton joints positions was not reduced as much as expected. It is due to the Vision algorithm producing high certainty values for the detected joints so that even invalid positions cannot be discarded. Despite

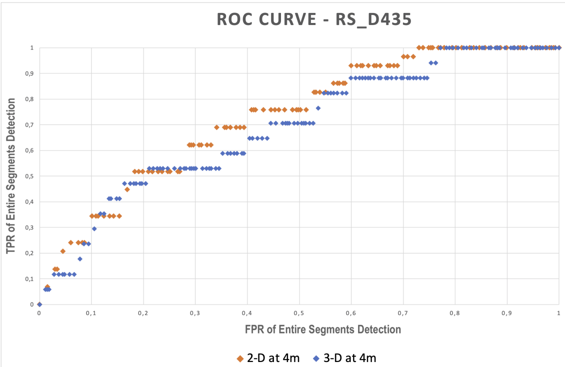


Figure 6: ROC curve comparison of the RS-D435 videos segmented in 2-D and 3-D.

this, we obtained a 74,16 % of good detections with a medium rate of false positives for videos recorded at 4 m from the wall, see Figure 7. That was not the case when it was recorded 6 m from the wall, where good detections decreased to 67,59 % and false positives increased with it. On the other hand, 3-D detections showed considerable variations in the expected results. The z-coordinate calculated for the joints exhibited a high rate of jittering, which produced many spikes in the speed signal. For now, this presents a technical restriction in our research, since the density of the point cloud depends on the physical memory of the device and the duration of the recorded sequence. However, the results for 4 m and 6 m were detections with 65,5 % and 59,56 % effectiveness.

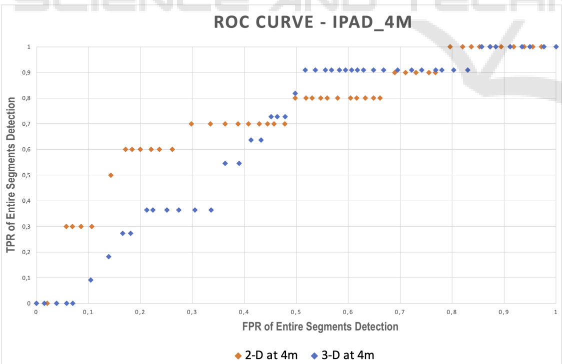


Figure 7: ROC curve for iPad Pro data sets recorded at 4 m from the wall, showing a comparison between 2-D and 3-D.

5 CONCLUSIONS AND FUTURE WORK

The objective of the climber's movement segmentation by analysing changes in speed and acceleration of their limbs was met according to the expectations. The study showed that currently there are algorithms skilled enough to detect various poses of climbers in action, such as OpenPose and Vision. It was possible to prove that the cumulative acceleration metric is valid for detecting the peaks of the climber's limbs movement. Although, there is still a significant problem to be solved, which is the sudden change of position of the hidden limbs by the climber's body, where a viable solution is to retrain one of the evaluated HPE frameworks by including a proper climbing image dataset. The latter is feasible using PoseNet, considering that Vision is a private framework from Apple Inc.

Even through the results of the automatic movement segmentation, in both 2-D and 3-D scenarios, were more consistent with the observations in the videos recorded with RS-435 using OpenPose, the quality of the video recorded with iPad and its Vision framework can not be discarded. Poor results at 6 m could be expected as the iOS documentation recommends 5,5 m maximum from the object, but it helped us in our project to test the reach of the device's technology. Considering the OpenPose licensing restrictions on the one hand, and the versatility of the iPad hardware and software on the other, this device is a suitable tool to continue our research. However, we are aware that Vision can only be used within the Apple Inc. devices environment, so its utilization limits the use of the applications.

The results presented above allow us to continue our research to evaluate the execution of technique in sport climbing, considering the relationships between posture, momentum and the effectiveness of the climber's movements.

ACKNOWLEDGEMENTS

This research was funded by the "Zentrales Innovationsprogramm Mittelstand (ZIM)" by the Federal Ministry for Economic Affairs and Energy (BMWi) with the project ID ZF4095809DH9.

REFERENCES

Aoki, T., Lin, J. F.-S., Kulić, D., and Venture, G. (2016). Segmentation of human upper body movement using multiple imu sensors. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3163–3166. IEEE.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2019). OpenPose: realtime multi-person 2d pose estimation using part affinity fields. In *IEEE trans-*

actions on pattern analysis and machine intelligence, volume 43, pages 172–186. IEEE.

Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299. IEEE.

Cha, K., Lee, E.-Y., Heo, M.-H., Shin, K.-C., Son, J., and Kim, D. (2015). Analysis of climbing postures and movements in sport climbing for realistic 3d climbing animations. In *Procedia Engineering*, volume 112, pages 52–57. Elsevier.

CMU (2019). Carnegie Mellon University: OpenPose - non-exclusive commercial license. https://cmu.flintbox.com/technologies/b820c21d-8443-4aa2-a49f-8919d93a8740, visisted on 24/09/2021.

Colyer, S. L., Evans, M., Cosker, D. P., and Salo, A. I. (2018). A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. In *Sports medicine-open*, volume 4, pages 1–15. SpringerOpen.

Fan, K., Wang, P., Hu, Y., and Dou, B. (2017). Fall detection via human posture representation and support vector machine. In *International journal of distributed sensor networks*, volume 13. SAGE Publications Sage UK: London, England.

Jansen, W., Laurijssen, D., Daems, W., and Steckel, J. (2019). Automatic calibration of a six-degrees-of-freedom pose estimation system. In *IEEE Sensors Journal*, volume 19, pages 8824–8831. IEEE.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. In *Transactions of the ASME–Journal of Basic Engineering*, volume 82, pages 35–45.

Khuangga, M. C. and Widyantoro, D. H. (2018). Human identification using human body features extraction. In *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 397–402. IEEE.

Kosmalla, F., Zenner, A., Tasch, C., Daiber, F., and Krüger, A. (2020). The importance of virtual hands and feet for virtual reality climbing. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–8. Association for Computing Machinery.

Kulić, D., Kragic, D., and Krüger, V. (2011). *Learning action primitives. Visual analysis of humans*, pages 333–353. Springer.

Lin, J. F.-S., Karg, M., and Kulić, D. (2016). Movement primitive segmentation for human motion modeling: A framework for analysis. In *IEEE Transactions on Human-Machine Systems*, volume 46, pages 325–339. IEEE.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.

McCay, K. D., Ho, E. S., Shum, H. P., Fehringer, G., Marcroft, C., and Embleton, N. D. (2020). Abnormal infant movements classification with deep learning on pose-based features. In *IEEE Access*, volume 8, pages 51582–51592. IEEE.

Meier, F., Theodorou, E., Stulp, F., and Schaal, S. (2011). Movement segmentation using a primitive library. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3407–3412. IEEE.

Papandreou, G., Zhu, T., Chen, L.-c., Gidaris, S., Tompson, J., and Murphy, K. (2018). PersonLab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model.

Richter, J., Beltrán B., R., and Heinkel, U. (2020a). Camera-based climbing analysis for a therapeutic training system. In *Current Directions in Biomedical Engineering*, volume 6. De Gruyter.

Richter, J., Beltrán B., R., Köstermeyer, G., and Heinkel, U. (2020b). Human climbing and bouldering motion analysis: A survey on sensors, motion capture, analysis algorithms, recent advances and applications. In *VISIGRAPP (5: VISAPP)*, pages 751–758.

Savitzky, A. and Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. In *Analytical chemistry*, volume 36, pages 1627–1639. ACS Publications.

Seifert, L., Dovgalecs, V., Boulanger, J., Orth, D., Hérault, R., and Davids, K. (2014). Full-body movement pattern recognition in climbing. In *Sports Technology*, volume 7, pages 166–173. Taylor & Francis.

Vögele, A., Krüger, B., and Klein, R. (2014). Efficient unsupervised temporal segmentation of human motion. acm siggraph. In *Eurographics Symposium on Computer Animation*.

Winter, S. (2012). *Klettern & Bouldern: Kletter- und Sicherungstechnik für Einsteiger*, pages 90–91. Rother Bergverlag.

Xiaohui, T., Xiaoyu, P., Liwen, L., and Qing, X. (2018). Automatic human body feature extraction and personal size measurement. In *Journal of Visual Languages & Computing*, volume 47, pages 9–18. Elsevier.

Zago, M., Luzzago, M., Marangoni, T., De Cecco, M., Tarabini, M., and Galli, M. (2020). 3d tracking of human motion using visual skeletonization and stereoscopic vision. In *Frontiers in bioengineering and biotechnology*, volume 8, page 181. Frontiers.

Zheng, C., Wu, W., Yang, T., Zhu, S., Chen, C., Liu, R., Shen, J., Kehtarnavaz, N., and Shah, M. (2020). Deep learning-based human pose estimation: A survey.