Computer Science Expository Work
Literature Review

Asad Khan, Laura Madrid, & Lucas Noritomi-Hartwig,
Supervised by Professor Lisa Zhang

Department of Mathematical and Computational Sciences
University of Toronto

October 2023

Introduction

Indoor Rock climbing originated in the 70's. Since then, it has gained massive popularity, given it provides a physical and mental stimulus for individuals of diverse age groups and abilities. The world of indoor rock climbing is in constant growth and development, and thus there is lots of room for new methods of analysis in terms of athletic performance, feedback, and health benefits among other areas. Researchers in computer science have shown significant interest in studying various aspects of indoor rock climbing like path generation, pose estimation, hold detection, and human-object interaction of the sport. Typically, climbers rely heavily on their vision to generate their path, estimate their pose, detect the holds and create a plan for how to climb the route. However, when climbers are visually impaired they no longer rely on this sense and the approach must be different. This literature review aims to dive deeper into the adaptive technologies and exploratory approaches that have emerged to provide visually impaired individuals with the thrilling experience of rock climbing.

Pose Estimation Overview

As climbers ascend a wall, their body movements and positions play a crucial role in determining their success and safety. These positions are known as "poses", and in the context of rock climbing a pose refers to a specific position of a human subject's body, joints, and limbs at any moment in time in relation to the climbing wall that they are ascending. For the purposes of developing a system that can assist climbers scale a wall, being able to accurately assess these poses are crucial. The system must also be able to adjust in instances where certain limbs and joints are occluded behind other body parts as is often the case when climbers take up certain unorthodox poses during their ascent. To assist with this, we inquire into recent advancements in pose estimation technology. Pose estimation is the computational process by which information on a human subject's pose is captured through video data and processed using pose detection algorithms to get a digital representation of their pose that accurately represents the position of the climber's body parts in space. This is useful to us as in order to suggest to the climber what their next move should be, the computer should have an accurate assessment of the positions of the climber's body parts in relation to the wall so that the next move it suggests is possible and useful.

To assess the current state-of-the-art in pose estimation technology, we look into the main components of what comprises an accurate and efficient pose estimation system. First, we look into the type of video capture methods that are being used. Traditional capture methods mainly comprise of RGB Video capture which provides a 2D representation of the environment. However LiDAR-based video capture provides a 3D representation of the environment, by estimating the depth of each point in space to reconstruct the climber's point cloud and their environment's point cloud in 3D-space. For higher-budget setups, IMU sensors can be used to provide real-time motion capture data of the climber's position and orientation.

One issue with traditional pose estimation libraries is that they are not always trained on data that incorporates human poses during niche activities such as rock climbing. During the process of rock climbing, the climber will often place themselves in unorthodox poses that are vastly different from regular activities such as walking, running, standing, etc. Thus, it is valuable to look into pose estimation methods as well as training datasets that are specifically tailored towards rock climbing or sports in general.

Pose Estimation: Data capture devices

As mentioned, RGB video is a staple of video capture and is the most common amongst commercial recording devices. However, recently the emergence of depth-sensing technology has become prevalent amongst common recording devices such as the ones on phones, cameras, and webcams. An analysis of the use of RGB-D (Red, Green, Blue + Depth) video capture devices by Beltrán et al. (2022) examines the use of two consumer-grade RGB-D recording devices, namely the Intel RealSense D435 and the iPad Pro 12.9-inch 4th Generation Beltrán et al. (2022).

Though both devices have depth-sensing technology, they approach this feature in different ways. Devices such as the Intel RealSense D435 use an active infrared sensor to capture depth which produces point clouds that it stores in a separate file alongside the RGB video. The iPad Pro utilizes a LiDAR scanner built in to its two pro cameras alongside a motion sensor. While the upside of these video capture devices are evident due to their accessibility and ease of use, their limitations become evident as a greater distance between the camera and the subject are observed, as is the case with Beltrán et al. (2022). which saw poor results with the iPad Pro at 6 meters. The iOS documentation itself recommends a 5.5m maximum from the object, which is a cause for potential issues as climbers naturally move further away from the camera as they ascend up the wall, with shorter lead climbing walls being between 25-30 feet Crux (2020).

There may also be instances where the camera would need to be significantly farther from the climber when assessing an ascent of a larger wall. Other endeavours into this area also seem to run into the same issue such as the case of Vähämäki (2016), who used the Microsoft Kinect sensor and deemed it unsuitable for rock climbing off the shelf Vähämäki (2016). Vähämäki (2016) makes the claim that standard commercial-grade depth sensors are "optimized to detect front-facing standing poses that are typical in a living room environment", which is the likely cause for issues at higher distances. However, Beltrán et al. (2022) still recommend LiDAR capture as found in the iPad Pro due to the quality of the RGB-D video capture. While the height of the wall was not specifically mentioned in their paper, it seems their pose estimation system tested by was tested on a very short wall (roughly 3 times the size of the climber) as seen in photos of the wall in comparison to the climber, meaning potential use cases for LiDAR cameras for shorter climbs. As a follow-up study, Richter et al. (2023) implements a real-time feedback system using the same iPad Pro for video capture with promising results of being able to educate beginner rock climbers using feedback from their pose data.

On top of assessing individual video capture methods, it is also important to assess these video capture methods in relation to the pose estimation methods that they employ. Considering the work done by Beltrán et al. (2022) with the iPad Pro, their work suggests using the iPad Pro LiDAR camera running the native Apple Vision pose estimation framework yields the best results for that specific

hardware. In their comparative analysis, Beltrán et al. (2022) measures the likelihood of detecting each skeletal joint correctly for each Camera/Pose Estimation Algorithm combination. Apple Vision running on the iPad Pro provided an average certainty of 82% whilst measuring 19 joints compared to the RealSense-D435 running OpenPose which yielded a certainty of 67%. The work done by Vähämäki (2016) institutes a novel approach which implements the offset joint regression (OJR) method for pose estimation. OJR works through using a body part classification system which is able to predict joint positions directly from a single input depth image. This method operates without the intermediate body part representation using a random forest regression model which estimates skeletal joint positions directly from single depth images. Results showed that OJR produces better results in situations where joints are occluded. Benefits of this approach are the speed at which the random decision forest model is able to evaluate poses, which allows for more efficient real-time pose estimation. However, approach taken by Vähämäki (2016) only uses synthetically generated data to train and test their results, which they mention generalizes to real world data reasonably well but is missing some of the variation observed in real world depth images.

Pose Estimation: Datasets and Models

The most comprehensive dataset procured specifically for the purposes of pose estimation in rock climbing activities is the CIMI4D dataset which incorporates complex movements that use a multi-modal video capture system consisting of 60 minutes of RGB videos, 179,838 LiDAR point clouds, 180 minutes of IMU poses, and an accurate global trajectory Yan et al. (2023). This dataset was trained on a few of the state-of-the-art pose estimation methods that utilize RGB, LiDAR, and scene capture. The comparative results collected by Yan et al. (2023). gives us a good baseline to see how some of the leading pose estimation methods work when trained on the most comprehensive pose estimation system. CIMI4D tests the following methods for RGB-based pose estimation: VIBE, MAED, DynaBOA. CIMI4D also tests LiDAR-based pose estimation methods: LiDARCap and P4Transformer and scene-aware pose estimation methods: PROX and LEMO.

The VIBE (Video Inference for Body Pose and Shape Estimation) model estimates the 3D pose and shape of a human body from video sequences Kocabas et al. (2020). It uses a type of machine learning called adversarial training, where one part of the model tries to generate realistic human motions, while another part tries to tell real from generated motions. It's trained using a large dataset of real human motions and applies this knowledge to analyze new videos, predicting how the body moves in 3D space across each frame of the video. Through this approach, VIBE can produce accurate and realistic 3D body pose and shape estimations from ordinary videos.

The MAED (Multi-level Attention Encoder-Decoder Network) model for pose estimation operates by paying attention to different aspects of human motion in videos Wan et al. (2021). It is composed of a Spatial-Temporal Encoder (STE) that looks at each frame of a video to understand how body parts are positioned and how they move over time, as well as a Kinematic Topology Decoder (KTD) that models a hierarchy of joints, where the estimation of a joint's pose parameters is influenced by the predicted pose parameters of its parent joints in a top-down manner. By combining these two approaches, the MAED model can more accurately estimate 3D poses and shapes of a human body, even in challenging scenarios like when there's a cluttered background or parts of the body are obscured.

The DynaBOA (Dynamic Bilevel Online Adaptation) pose estimation model is designed to adapt to videos that come from different domains, making it capable of handling varying camera angles, bone lengths, and backgrounds which typically challenge standard models Guan et al. (2021). It introduces what's called temporal constraints which is information from previous frames, to refine predictions in current and future frames, making the most out of the sequential nature of video data. The model employs a two-stage optimization process, known as bilevel optimization, to strike a balance between making general predictions and accurately fitting to challenging or tricky frames. Through

this adaptive approach, it effectively deals with the challenges posed by out-of-domain videos and achieves high accuracy in human mesh reconstruction, even in complex and varied scenarios.

LiDAR pose estimation methods such as LiDARCap and P4Transformer use the LiDAR point clouds in combination with spatial and temporal dynamics in data to achieve accurate pose estimation. LiDARCap specifically focuses on long-range human motion capture using LiDAR point clouds, leveraging temporal encoding to aggregate features hierarchically over time for pose regression Li et al. (2022). On the other hand, P4Transformer employs a point 4D convolution to embed spatio-temporal structures in point cloud videos, further leveraging a transformer mechanism to capture appearance and motion information across the entire video Fan et al. (2021).

Through qualitative results, Yan et al. (2023) suggests that the PROX (Proximal Relationships with Object eXclusion) method works better than the other RBG and LiDAR-based methods when trained on the CIMI4D dataset. This is due to it being a scene-aware method Hassan et al. (2019). It aims to improve 3D human pose estimation by incorporating the static 3D scene structure in the estimation process from monocular images. Additionally, it also penalizes situations where the estimated human body pose interpenetrates with objects in the scene, ensuring that the body and scene objects occupy distinct spaces. By doing this, the PROX method endeavors to resolve ambiguities in 3D human pose estimation, making the estimated poses more accurate and consistent with the 3D scene structure.

Results showed pose estiamation algorithms trained on the CIMI4D dataset have a better ability to reconstruct the true climbing environment compared to individual methods like LiDAR, RGB, and IMU alone. However, benchmarks still show a considerable degree of error when CIMI4D is trained on the models mentioned, with the authors citing issues such as lack of detailed hand poses leading to a penetration of the holds. Yan et al. (2023). suggests utilizing the human-scene interaction annotation provided in CIMI4D could significantly improve results in these areas.

The outcomes from these papers highlight the enhanced accuracy and robustness achieved through specific video capture methods. A common success point was the use of depth/LiDAR cameras and joint estimation models to provide an accurate estimation of the climber's pose. For our implementation, replicating the methods used by Beltrán et al. (2022) could be promising given the simple equipment required and the robustness of the already developed Apple Vision framework.

Benchmarking and model evaluation are crucial steps in assessing the performance and robustness of machine learning models, especially in the domain of human pose estimation. A significant work in this area is the "MPII Human Pose" benchmark, a comprehensive dataset sourced from YouTube, covering over 800 human activities Andriluka et al. (2014). This dataset was designed to challenge and evaluate pose estimation models with its diverse set of images, rich annotations, and varied imaging conditions. The paper explored several leading human pose estimation approaches, including the Flexible Mixture of Parts (FMP), Pictorial Structures (PS), Multimodal Decomposable Models (MODEC), and Armlets. Among these, the PS approach stood out on the LSP dataset, while the Armlets approach excelled on its namesake dataset with numerous truncations and occlusions.

Hold Detection

In their work, Ayesha Arif and Kim (2021) were able to outperform state-of-the-art methods for detection of human-object interaction (HOI) by combining individual methods used to solve the pose estimation and object detection problems. Ayesha Arif and Kim (2021) discuss that novel methods of object detection can be helpful in better estimating the pose of the human body. Ayesha Arif and Kim (2021) combined K-Means clustering and YOLO, with the novel approaches of Fuzzy C-Means for super-pixels and Random Forests for segmentation of objects. This combination, paired with human detection using centroid, extreme points, and inscribed ellipses, allowed them to outperform state-of-the-art methods on the PAMI'09 dataset, one of static images used for action recognition, and the UIUC's ISD dataset, one used for image sense discrimination. The combination of both human

body detection as well as object detection may prove to simplify the task of having to both estimate the pose of a climber, while also detecting a specific hold.

Being able to detect when a rock climber has a grip on a rock hold is useful for a live-assisting model in order to determine if the move made was effective, and whether or not the climber is ready to attempt the next move Sarah Ekaireb and Manjunath-Murkal (2020). In their paper, Sarah Ekaireb and Manjunath-Murkal (2020) discuss methods of detecting the rock holds on the wall. Sarah Ekaireb and Manjunath-Murkal (2020) attempted using Canny Edge detection in tandem with openCV's blob detection. However, one of the disadvantages of these models is that they operate most reliably under grayscale input data. Given that the rock holds are coloured specifically for each individual' climbing path, it was important to keep the colour of the rocks in the input. Sarah Ekaireb and Manjunath-Murkal (2020) then attempted using colour binning, and later colour frequency analysis. While there were some promising results, Sarah Ekaireb and Manjunath-Murkal (2020) recommended more research be done on tuning the model in order to effectively implement the method.

Next, Sarah Ekaireb and Manjunath-Murkal (2020) trained a neural network using a Roboflow model to detect and segment holds. This achieved a Mean Average Precision (mAP) of 96.3%. Following this, Sarah Ekaireb and Manjunath-Murkal (2020) then trained a neural network to perform a multi-class classification on the different colours of holds, to separate different paths from one-another. This performed very well with a mAP of 99.3% accuracy. However, the model would sometimes confuse similar colours (in the paper it is shown how sections of two adjacent paths of colours green and turquoise, respectively, were confused to be part of the same path). To mediate this, Sarah Ekaireb and Manjunath-Murkal (2020) again used colour frequency analysis, but only on the sections of the image with the bounding box segmented by the first network. An important note is that detection of holds becomes more difficult when the holds are covered in chalk, sometimes even to climbers themselves. The work of Sarah Ekaireb and Manjunath-Murkal (2020) suggests that training separate neural networks for the small pieces of the task of hold-detection can be useful in terms of fixing problems that will arise when training a model. Their work also suggests a consideration of the state of the wall i.e., whether the wall has been cleaned, or if there is still an amount of chalk layered on top of the holds, rendering them less discernible to both the model and even other climbers.

Path Generation

In their work, Duh and Chang (2021) reasoned that, by their sequential nature, path generation problems are more similar to natural language processing problems. They thus, chose to use a long short-term memory (LTSM) network, a type of recurrent neural network (RNN) to build DeepRouteSet, which was built from adapted source code from a Coursera exercise on LTSM networks. DeepRouteSet is a path generating model which has achieved a higher percentage of "quality problems" generated. A problem is considered to be of "quality" if it is reasonable; exhibits; avoiding redundant/unused holds and awkward sequences that may cause injury, as well as being easy to benchmark; having consistent difficulty between moves, and overall follows a "natural" climbing flow, allowing for climbers to hold a natural posture throughout the problem. These criteria can be useful as heuristics when creating a tool that involves generating a rock climbing route, considering we want the climbers to be able to relatively easily perform the suggested moves. While the work done is focused on MoonBoard problems, the general idea of treating the rock-climbing path generation problem as sequential can prove effective. This work, and proof of effectiveness of the LTSM network on a path generation problem, shows that a live rock-climbing assistant model can be built using LTSM networks, which generates a path for a climber and provides next move suggestions as the climber progresses, taking into account the previous moves performed throughout the problem.

In his paper, Stapel (2020) also reduced the path generation model to the MoonBoard. Since the holds on the MoonBoard are fixed, the responsibility of the model is not selecting the holds in the problem, but instead choosing which of those holds to be part of the route. This is translatable

to an indoor top-rope climbing problem, as the holds of the problems are colour-coded, and thus, the challenge would be to choose the most effective sequence of these holds which would allow the climber to solve the problem. In his paper, Stapel (2020) implements a greedy algorithm to find the easiest best move, graded on a system of heuristics taking into account factors such as the next hold's angle angle relative to the ground, the reaching distance, and the type of hold. Stapel (2020) briefly mentions the use of machine learning for route generation, though focuses mostly on the heuristics of determining what the best next move would be, given a starting position. He also notes that focusing on the MoonBoard has constraints, as the algorithm is unable to consider a problem in 3 dimensions, where volumes attached to the climbing wall serve as either holds or obstacles in a path. A flaw found in testing Stapel (2020)'s algorithm was that it was not able to differentiate between a left handhold and a right handhold, making for relatively awkward, poor flow route generations. Since we plan to implement a pose estimation model to track the climber along the route, this flaw should be addressed. Stapel (2020) categorizes a number of different moves such as "normal moves", "dyno-ing", and "crossing" which he uses to assign levels of difficulty to next moves and paths as a whole. Translating this over to a pose estimation and hold detection model. we should be able to have the model determine the relative "awkwardness" or "smoothness" of a move given the current target limb position and it's rock-hold target position.

Feedback Systems

The essence of real-time feedback in climbing motion analysis lies in its ability to offer climbers instantaneous insights into their movements, techniques, and interactions with the environment. Such feedback, when optimized, can significantly enhance the climbing experience, ensuring safety, improving performance, and aiding in technique refinement. Delving into the research landscape, we uncover methods such as auditory feedback and tactile feedback via a tongue interface that have been explored in this domain and the potential pathways for their implementation in our project.

Feedback systems: Auditory coach

One of the first approaches established was by *MetaHolds: A Rock Climbing Interface for the Visually Impaired* Ilich (2008) paper which focused on auditory aids for those with varying degrees of vision loss. Speakers were placed on each hold to give the climber information about the type of hold, the positioning of the hold and the route it is located on. were placed beneath each hold. The climbing holds were categorized by 5 different types (pinch, crimp, sloper, pocket, jug) and each type had a specific sound effect. The climber would wear a sensor on top of their helmet that would check the positioning of the head and once the head was placed in front of the hold, the speaker for that hold would be activated. After the experiment, various participants suggested that the sound of the tone should vary based on distance and it was noted that sometimes participants forgot which holds the jugs meant. Ilich (2008) also found that verbal affirmation should only be given when the climber is on the correct hold and has correctly handled the hold based on its type. Overall, the participants did not feel distracted by the audio and understood its clear correlation the the 5 categories of holds.

Feedback systems: Tongue Interface

Another feedback approach was to help the climber see through their tongue. The *Climb-o-Vision: A Computer Vision Driven Sensory Substitution Device for Rock Climbing* paper by Richardson et al. (2022). used an Arduino compatible tongue interface called the Cthulhu Shield to convey climbing hold locations, which were detected through a YOLOv5 model in python. This interface uses an 18-electrode grid to stimulate different parts of the tongue depending on the location of the hold instead of trying to visualize the whole environment. The paper mentioned that this approach was simpler,

more accessible and affordable compared to the BrainPort device, which has 400 electrodes and can overwhelm the climber despite removing the colour and reducing pixel density of the environment as the tongue has varying degrees of sensitivity in different areas. Although this approach had an mAP of 0.913 by the 500^{th} epoch, the researchers realized that lighting conditions affected how the model detected various coloured holds and that perhaps having multi-toned holds would make the detection process easier and that higher resolution images would improve the model. The researchers also realized that mobile processing could improve the feedback system since the software of the tongue interface must be on a laptop that the climber carries on their rucksack, making the climb more strenuous than it needs to be.

Conclusion

Our goal is to build a model that can estimate the position of a climber on a climbing path, detect the rock holds along the path, and, in real time, assist the climber to make the next move. This will involve both the detection of the climber and holds, as well as providing feedback to the climber. For human body pose estimation, we plan to use Apple Vision as it has proven to have a simple development process when incorporating other models, as well as achieving high results. We have found that one of the most effective methods of implementing object, and thus rock-hold, detection is using the YOLO model. We plan to use the YOLO framework included in Apple's Core ML models alongside Apple Vision for pose estimation. Our implementation will also involve live feedback to the climber. Using the position information of both the climber and the holds on climbing path, the system must convey to the climber which limb to move (left hand, right hand, left foot, right foot) as well as the proximity of the specified limb to the target rock hold. To do this, we first plan to have the model communicate with the climber solely through audio using speech, for example: on a route, the climber may hear as the next move suggestion: "Right arm up and slightly to left" as a brief instruction. This will mainly be done to ensure that the model is guiding the climber in the correct general direction. The ultimate goal is for the model to communicate by haptic feedback via motors attached to each limb, as well as audio to indicate more specifically how to perform the next suggested move. The vibration will start at a slow frequency when the limb is far from the target hold. As the limb approaches the target hold, the vibration frequency is increased up to a solid vibration "tone" when the model detects that the limb has a hold on the target rock-hold. In order to communicate to the climber how high or low to reach for the next hold, we plan to use earbuds with a tone whose pitch represents the relative height between the limb and the target hold i.e., if the target is above the climber's limb, the pitch will be higher, and lower as the limb approaches the target. We plan to calibrate the pitch of the tone to a C natural.

References

Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2014. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Mohammed Alarfaj Ahmad Jalal Shaharyar Kamal Ayesha Arif, Yazeed Yasin Ghadi and Dong-Seong Kim. 2021. Human Pose Estimation and Object Interaction for Sports Behaviour. (2021), 1–18. `https://digitallibrary.aau.ac.ae/bitstream/handle/123456789/693/Human%20Pose%20Estimation%20and%20Object%20Interaction%20for%20Sports%20Behaviour.pdf?sequence=1&isAllowed=y`

Raul Beltrán Beltrán, Julia Richter, and Ulrich Heinkel. 2022. Automated Human Movement Seg-

mentation by Means of Human Pose Estimation in RGB-D Videos for Climbing Motion Analysis. In *VISIGRAPP*. https://api.semanticscholar.org/CorpusID:246859460

Conquer Your Crux. 2020. How high are climbing walls on average? https://www.conqueryourcrux.com/how-high-are-climbing-walls-on-average/

Yi-Shiou Duh and Ray Chang. 2021. Recurrent Neural Network for MoonBoard Climbing Route Classification and Generation. https://arxiv.org/pdf/2102.01788.pdf. (2021), 1–9.

Hehe Fan, Yi Yang, and Mohan Kankanhalli. 2021. Point 4D Transformer Networks for Spatio-Temporal Modeling in Point Cloud Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 14204–14213.

Shanyan Guan, Jingwei Xu, Michelle Z. He, Yunbo Wang, Bingbing Ni, and Xiaokang Yang. 2021. Out-of-Domain Human Mesh Reconstruction via Dynamic Bilevel Online Adaptation. arXiv:2111.04017 [cs.CV]

Mohamed Hassan, Vasileios Choutas, Dimitrios Tzionas, and Michael J. Black. 2019. Resolving 3D Human Pose Ambiguities with 3D Scene Constraints. arXiv:1908.06963 [cs.CV]

Michael Ilich. 2008. MetaHolds: A rock climbing interface for the visually impaired. *Ilich M* (2008).

Muhammed Kocabas, Nikos Athanasiou, and Michael J. Black. 2020. VIBE: Video Inference for Human Body Pose and Shape Estimation. arXiv:1912.05656 [cs.CV]

Jialian Li, Jingyi Zhang, Zhiyong Wang, Siqi Shen, Chenglu Wen, Yuexin Ma, Lan Xu, Jingyi Yu, and Cheng Wang. 2022. LiDARCap: Long-range Marker-less 3D Human Motion Capture with LiDAR Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20502–20512.

Mike Richardson, Karin Petrini, and Michael Proulx. 2022. Climb-o-Vision: A Computer Vision Driven Sensory Substitution Device for Rock Climbing. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.

Julia Richter, Raul Beltrán, Guido Köstermeyer, and Ulrich Heinkel. 2023. Climbing with virtual mentor by means of video-based motion analysis. *Proceedings of the 3rd International Conference on Image Processing and Vision Engineering* (2023). https://doi.org/10.5220/0011959300003497

Prem Pathuri Priyanka Haresh Bhatia Ripunjay Sharma Sarah Ekaireb, Mohammad Ali Khan and Neha Manjunath-Murkal. 2020. Computer Vision Based Indoor Rock Climbing Analysis. 1 (2020), 1–17. https://kastner.ucsd.edu/ryan/wp-content/uploads/sites/5/2022/06/admin/rock-climbing-coach.pdf

Frank Stapel. 2020. A Heuristic Approach to Indoor Rock Climbing Route Generation. *32nd Twente Student Conference on IT* (2020), 1–16. https://essay.utwente.nl/80579/1/stapel_BA_eemcs.pdf

Joni Vähämäki. 2016. Real-time climbing pose estimation using a depth sensor. https://api.semanticscholar.org/CorpusID:40046181

Ziniu Wan, Zhengjia Li, Maoqing Tian, Jianbo Liu, Shuai Yi, and Hongsheng Li. 2021. Encoder-decoder with Multi-level Attention for 3D Human Shape and Pose Estimation. arXiv:2109.02303 [cs.CV]

Ming Yan, Xin Wang, Yudi Dai, Siqi Shen, Chenglu Wen, Lan Xu, Yuexin Ma, and Cheng Wang. 2023. CIMI4D: A Large Multimodal Climbing Motion Dataset under Human-scene Interactions. arXiv:2303.17948 [cs.CV]