Video and Sensor-Based Rope Pulling Detection in Sport Climbing

Iustina Ivanova
Free University of Bozen-Bolzano
Bozen-Bolzano, Italy
iivanova@unibz.it

Marina Andrić
Free University of Bozen-Bolzano
Bozen-Bolzano, Italy
maandric@unibz.it

Sadaf Moaveninejad
Free University of Bozen-Bolzano
Bozen-Bolzano, Italy
sadaf.moaveninejad@unibz.it

Andrea Janes
Free University of Bozen-Bolzano
Bozen-Bolzano, Italy
ajanes@unibz.it

Francesco Ricci
Free University of Bozen-Bolzano
Bozen-Bolzano, Italy
fricci@unibz.it

ABSTRACT

Sport climbing is becoming an increasingly popular competitive sport as well as a recreational activity. For this reason, indoor sport climbing operators are constantly trying to improve their services and optimally use their infrastructure. One way to support such a task is to track the climbing activities performed by visitors while climbing. This paper considers a scenario in which a sensor is attached to a piece of climbing equipment that connects the climbing rope to the bolt anchors (quickdraws) and a camera is overlooking a climbing wall. Within this scenario, this paper explores two approaches to detect when a climber finishes a climb and pulls the rope from the wall: 1) a hybrid approach in which sensors and cameras are used and 2) a video-based approach where only cameras are used. The evaluation resulted in recognition precision of 91% for the hybrid and 76% for the video-based approach, respectively. This paper also discusses advantages and disadvantages of analysed approaches and points out future research directions to allow the automatic tracking of climbing activities.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous computing**; • **Computing methodologies → Machine learning**; • **Hardware → Sensors and actuators**.

KEYWORDS

Sensors; video analysis; climbing activity recognition

ACM Reference Format:
Iustina Ivanova, Marina Andrić, Sadaf Moaveninejad, Andrea Janes, and Francesco Ricci. 2020. Video and Sensor-Based Rope Pulling Detection in Sport Climbing. In *3rd International Workshop on Multimedia Content Analysis in Sports (MMSports'20), October 16, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3422844.3423058

1 INTRODUCTION

Sport climbing is becoming an increasingly popular competitive sport as well as a recreational activity. Sport climbing can be performed outdoors, e.g., on natural rocks as well as indoors, i.e., on artificial routes in climbing gyms, and requires protection such as harness and ropes.

Today, science and engineering are used in various sports to help athletes to improve their performance [3]. A popular way to achieve this is activity tracking [4], where mobile sensing technologies gained popularity due to their ability to collect and process measurement data. For example, smart watches and fitness trackers are used by cyclists and runners to obtain and visualize statistics about their endurance and performance. In contrast, sport climbing did not receive an equal amount of attention in consumer electronics so far. Indeed, collecting quantitative climbing data would allow a number of applications: on an individual level, it could support performance assessment and training; on a gym level, anonymous usage statistics could help to understand which routes are popular and guide route builders to construct climbing gyms that people like. Moreover, anonymous sensing technology could help to enforce minimum safety standards, e.g., reporting strong falls.

Previous research efforts concerned with monitoring sport climbing activities demonstrated promising systems able to detect several activities (e.g., climbing, resting during a climb, falling) and assess climbing performance. To this end, a variety of sensing technologies has been proposed, such as body-worn sensors, video analysis and instrumented climbing walls. However, most of these systems involve wearable devices for tracking climber's movements (e.g. [7, 21, 22, 28]), and some are constrained to laboratory settings (e.g. [14, 32, 34]) due to their complex and expensive design. This paper is the result of research cooperation with Vertical-Life (https://www.vertical-life.info), a company that develops climbing gym management software, and therefore deals with an activity monitoring approach that is suitable for deployment in commercial climbing gyms i.e., does not require user interaction and modification of the existing climbing infrastructure that would require safety and compliance certifications.

In this work, we augment a standard piece of a climbing equipment present on climbing routes in every climbing gym, namely a quickdraw, with a small 3-axial sensor that measures acceleration of the quickdraw during a climb. We hypothesise that climbing

MMSports'20, October 16, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-8149-9/20/10…$15.00
https://doi.org/10.1145/3422844.3423058

activities can be inferred from recordings of movements of the augmented quickdraw. The setup envisioned for data collection is depicted in Fig. 1a. We decided to equip the second-lowest quickdraw to the ground with a sensor since, in practice, climbers often neglect to clip the rope into the first.

Our overarching goal is to detect activities throughout the day, and understand what the climber is actually doing, e.g., clipping the rope into a quickdraw or pulling the rope after the climb is finished (removing the rope). That assumes a scenario in which sensors are distributed across the climbing gym and continuously collecting data about events caused by various climbs and various climbers. For now, it is not our intention to differentiate between climbers. The problem of the current interest is to detect climbing episodes within continuous stream of data. Moreover, this paper only considers "lead climbing", which describes the type of sport climbing depicted in Fig. 1a: the rope is carried up by the climber, who routinely stops to clip it into quickdraws for safety. The other side of the rope is led through a device that—in case of a fall—stops the rope and is operated by the belayer. The belayer, when the climber reached the top, also slowly releases the belaying device to lower the climber to the ground. After one climb is finished, the rope is removed, and another climb can begin. A possible approach to identify climbing episodes is, therefore, to detect when the rope is pulled out of all the quickdraws after a climb.

Hence, in this paper, our goal is **to detect the *rope pulling* activity** and correctly label with that information data obtained from the sensor augmented quickdraw. This is important by itself, but, in future work we also aim at using **only sensor data** and extract from it other information about the climbs. This requires Machine Learning techniques that use training data where the sensor data is correctly labelled with semantic information describing the performed activity. In fact, video recording is not always feasible: climbers do not like it and it requires a specific infrastructure and authorization. Hence, the research described in this paper can also help to achieve this long term goal, that is, to acquire sensor data and their correct semantic labels. We investigate here two methods that could be used to speed up a totally manual labelling of the sensor data: a hybrid method that relies on both sensor and video data and a video-based method that relies only on video data.

The rest of the paper is structured as follows: Sect. 2 reports about related work, Sect. 3 and Sect. 4 introduce the case study and two methods we adopted to detect rope pulling activity, Sect. 4.1 presents the hybrid method, Sect. 4.2 presents the video-based method. Sect. 5 evaluates the two methods and Sect. 6 concludes the paper.

2 RELATED WORK

To the best of our knowledge, this work is the first study on detecting *rope pulling* activity during climbing. Hence, there is no earlier work specifically dedicated to that problem to be compared with ours. Hence, in this section, we briefly review the broader subject of activity detection in climbing. Previous works on that subject can be divided into two main groups: **video-based** and **sensor-based** techniques. The former relies on human action recognition, which is an area of computer vision [26] while the latter includes works on time-series analysis of data obtained from sensors.

In video-based methods, athletes may be tracked by several cameras. Like in other applications such as robotics or surveillance systems, also for sports activities, one can take advantage of human activity recognition and object detection. In [13], the authors extracted temporal human 2D pose sequences from video frames, followed by automatic event detection in the athlete's motion using convolutional neural networks (CNN). Similarly [1] and [10, 20], implemented pose estimation based on the estimation of the skeleton of every person in an image in a soccer match and climbing, respectively. In [2], players in a soccer match are represented with blobs, i.e., regions segmented out from the playfield by color differentiating. Video-based activity recognition in climbing is mainly performed by using the depth map obtained from a Microsoft Kinect device, also used in computer games [18, 20]. The limitation of such device is hardware complexity: it consists of two cameras and an infrared projector, therefore, it requires adjustment between the camera and the projector.

In sensor-based methods, climbing activities are considered as temporal events that may be detectable from data measured through sensors. Event detection is an important task in the analysis of time-series data. Each event is referred to as a point in time where data that belonging to an interval around that event has specific characteristics [15]. In this regard, sliding windows are used to transform each interval of time-series data into an appropriate feature vector. Other research areas relevant to event detection are changed point detection and time-series segmentation [15]. In previous studies, different types of sensors were embedded in different places and employed to analyze sport climbing. The authors of [22] and [21] measured the acceleration of climbers through wearable sensors on wrists for performance assessment and route recognition, respectively. In the same way, the authors of [28] developed an analytical framework for assessing general climbing performance during training using a single ear-worn accelerometer-based sensor. In addition to wearable sensors, [23] presents a sensor-equipped climbing wall in which capacitive sensors are embedded into climbing holds to detect any touch by hands or feet. Similarly, in [27] force sensors are utilized to measure the load applied to the holds as a child hangs and steps up during climbing. In [37] the authors describe fence climbing recognition based on the data obtained from a three-axial accelerometer mounted on the fence. Their method has two steps: first, to discriminate activity vs. no activity and second, to classify the activity into climbing or rattle. They used signal variation to detect events with vibration, from non-event with no or little vibration. They proposed a threshold for a signal variation to distinguish between the two situations.

Differently from previous studies, we do not attach any sensor to the wall or the body, but we augment a climbing instrument. Introducing our own climbing holds might involve certification, insurance, and legal costs. For example, a patent for a "touch-sensitive, illuminated climbing hold" is already registered with the patent number US 2019 329 113A1 [11]. Regarding wearable sensors, it is not convenient nor desired by the climbers to wear a device (either on their hand, belt, or leg) while climbing a wall. Also, when the sensor is attached to the quickdraw, as we propose, there is no need to instruct climbers on how to use the tracking device each time. Moreover, we utilize video frames recorded by the camera to obtain information on the vertical movements of the climber. We employ a person detection approach, which is computationally less expensive than pose estimation.

3 CASE STUDY: *ROPE PULLING*

As stated in the introduction, we aim at detecting the activity of removing the rope from the quickdraws, which is commonly performed by climbers after they are lowered to the ground. We call this activity *rope pulling* as it involves pulling the rope down by the climber, until the rope has passed through all the quickdraws it had previously been clipped in while the climber was ascending.

3.1 Data collection

We used two types of devices for detecting the *rope pulling* activity, i.e., a video camera and a quickdraw enhanced with an accelerometer. Two groups of data, one consisting of a video recording and the other containing accelerometer readings were generated and collected during climbing sessions on the selected route.

One male and one female climber participated in the experiment. The participants' skill levels were intermediate i.e., self-estimated as 5a and 6b on-sight on the French Rating Scale of Difficulty (FRSD). One participant had a climbing experience of five years and the other of thirteen years. For the purpose of data collection, the participants were asked to climb in the leading style five times in succession on one pre-selected route. The participants climbed in their usual speed, clipping the rope into every quickdraw, and were free to take resting time between climbs. The participant who climbed pulled the rope after each ascent. The other participant was responsible for belaying. After the first session of five climbs was completed, the participants swapped the roles.

Data collected during the first climbing session were used to develop detection algorithms. During the later evaluation of the detection performance (in Section 5), data collected from the second climbing session was used for benchmarking them. In the following, we refer to the respective sets of data as training and test set.

The quickdraw movements were captured using a small accelerometer sensor attached to the strip in the central part of the quickdraw (see Fig. 1a). We used one Movesense[1] sensor and configured it to sample tri-axial acceleration data at 50Hz frequency, i.e., one sample every 20ms. The communication with the sensor for data download was based on Bluetooth connection with an iPhone X running the "Movesense Showcase" mobile app[2]. The sensor-enhanced quickdraw was placed on the second-lowest position to the ground. Upon placement, the sensor's x-axis was horizontal and parallel to the wall, the y-axis was vertical, and the z-axis was horizontal and orthogonal to the climbing wall. Moreover, the climbs were video recorded using a fixed camera that was placed close to the ground and facing the climbing wall with a selected route. All video recordings were taken with the same constant frame rate of 30 fps. Each climbing session was fully recorded.

4 METHODS

Based on the available data—sensor and video data—we explored two analysis techniques that detect rope pulling-related events. First, a hybrid approach, in which initially by means of person detection, we identify the belayer and the climber, see Fig. 1b. Then, once we know when the climber is back on the ground, we analyse the sensor data to find the exact point when the rope pulling ends, see Fig. 1c.

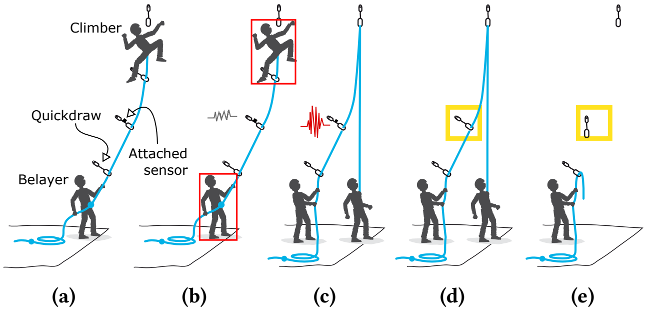

Figure 1: The setup, the two steps of the hybrid approach, and the video approach (original picture of the climbers courtesy by Petzl). The setup (a): Climbing a route in the lead climbing style: the belayer holds one end of the rope while the climber clips the rope into quickdraws for protection. Step 1 hybrid (b): Detection of position of the climber and belayer from video data. Step 2 hybrid (c): Identification of rope-pulling events from sensor data. Video (d): Detection of climbing rope within the light-coloured rectangle. Video (e) When *rope-pulling* is finished, the light-coloured rectangle does not contain a rope.

In the video-based approach, we detect the position of a quickdraw, that was marked on the climbing wall surrounding it with a light-colored adhesive strip (see Fig. 1d). We actually detect if within the mentioned rectangle the rope is present. If not, we consider the rope to be removed (or pulled, as in Fig. 1e).

An illustration of both detection approaches (also considering the timeline) is presented in Fig. 2. It is worth mentioning that the video and accelerometer data are synchronized based on timestamps assigned locally i.e., by the video camera and sensor device, respectively.

4.1 Hybrid approach

This section presents a two-step approach for detecting occurrences of the *rope pulling* activity in sensor data using a sliding window data analysis procedure [19]. At first, we use video processing to identify occurrences of a discrete event, i.e., that a climber reached the ground again and therefore is considered to be *lowered*. The *lowered* event marks the earliest time for the beginning of *rope pulling*; *rope pulling* starts after the climber had been lowered to the ground. Then, the start and the end of *rope pulling* are identified using a sliding window analysis of acceleration measurements coming from the sensor following each *lowered* event. The method is based on the observation that before the start and after the end of *rope pulling*, there is no or little movement of the quickdraw on which the accelerometer sensor is located. Although the video camera and sensor device do not share a common temporal basis, and therefore the synchronization of *lowered* events in video and acceleration data is possibly not precise at the sample level, the

[1] https://www.movesense.com
[2] https://apps.apple.com/us/app/movesense-showcase/id1439876677

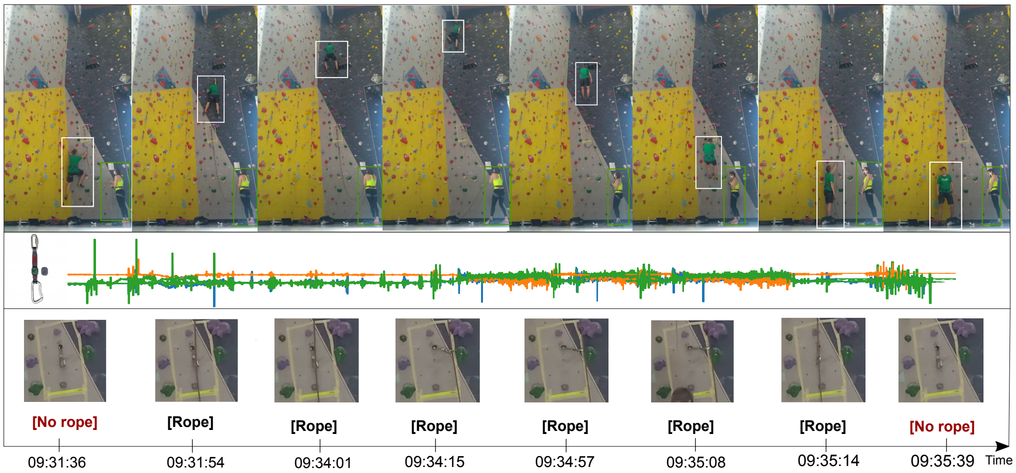

Figure 2: (Top) Sample frames from a video recording of one climb. White and green rectangles indicate the bounding boxes of the climber and the belayer, respectively. Upon reaching the top of the route the climber lowers to the ground and subsequently pulls the rope. (Middle) Illustration of a three-axial acceleration signal that is acquired by the sensor-enhanced quickdraw and synchronized with the above video frames. (Bottom) Examples of detected regions in video frames for object detection with classes *rope* (selected quickdraw has a rope), and *no rope* (selected quickdraw is without a rope).

algorithm could still accurately identify the start and end of each *rope pulling*, as it will be shown.

The workflow for detecting the *lowered* event is as follows. First, we use RetinaNet [24, 33] to extract bounding boxes of persons on a per-frame basis (Fig. 2 top). These are specified with a pair of coordinates of the corners of rectangles around the belayer and the climber. The distance of the climber from the ground is crucial to the event detection; therefore we estimate it by calculating the maximum vertical pixel distance of the bounding boxes' centers from the bottom of the frame. After having obtained the climber's distance from the ground for each frame independently, we apply temporal smoothing based on zero-phase filtering for outlier elimination. Finally, the *lowered* events are identified in a simple iterative procedure from the smoothed signal as a subset of the local minima points. Examples of temporal trajectories that illustrate *lowered* event detection are presented in Fig. 3.

The sensor located on the second quickdraw records the acceleration along three axes, denoted by A_x, A_y and A_z. These time-series cannot be used directly because the acceleration due to the movement of the rope is registered together with the gravity component. The norm of this component is well-known (9.81 m/s^2), but it cannot be removed directly due to the presence of rotational movements. A low-pass filter, as proposed in [6], was demonstrated to be effective for isolating the gravity component from the quickdraw movement-related component. The acceleration due to gravity is calculated using a recursive low-pass filter as follows: $A^{gr}[n] = a \cdot A[n] + (1-a) \cdot A^{gr}[n-1]$, where a is a constant dependent on the sensor sampling rate and A is the raw input data (A_x, A_y, A_z). The acceleration component due to the quickdraw movement A^m is then obtained by subtracting the low-pass filtered data from the original signal along the three-axes.

To detect the start and end event of *rope pulling* in acceleration signals A^m, we employ a sliding window procedure that extracts data windows of 1s (50 samples) length with a shift of 1 sample between one window and the next one. The optimal window size depends on the target activity, however, for common human activities window size of 1-2s has often been identified as optimal [5, 12]. For each window, a single scalar value is calculated, derived from the three-dimensional accelerometer's vector, that represents the movement energy over a short period of time (short-term energy). Then, the short-term energy of each window is compared to a threshold of 7.2g, and if it is greater, then the window is added to the sequence of previously found windows. This process is repeated for each *lowered* event until a consecutive sequence of windows of at least 5 seconds duration is found, which is a reasonable assumption for

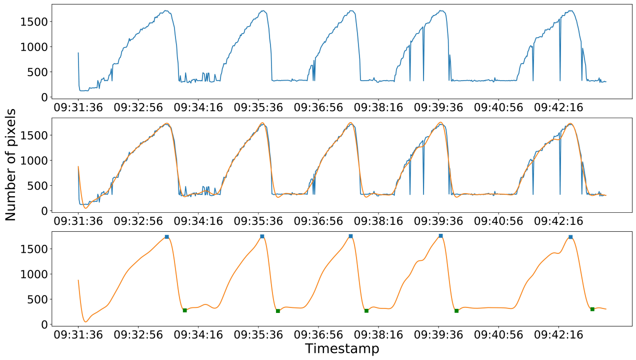

Figure 3: (Top) Height of computed bounding boxes in pixels for the training dataset of five climbs. (Middle) Blue line: original signal, orange line: zero-phase filter. (Bottom) Points indicating when climber reached the top (blue squares) and lowered to the ground (green squares).

Algorithm 1 Event detection with sliding window

```
Input: list of lowered timestamps $T_l$; energy threshold $e$; duration threshold $d$; acceleration signal $A$; window length $W$
Output: lists of timestamp pairs $T_{rp}$ that indicate the timing of the start and end event of rope pulling for given signal $A$
procedure DetectRopePulling($T_l, A, e, d$)
    $A^m = A - A^{gr}$                                  ▷ Gravity removal
    $T_{rp} = \emptyset$                                ▷ Initialize timestamp set
    for all timestamp $t$ in $T_l$ do
        $T = \emptyset$                                 ▷ Initialize auxiliary timestamp set
        for all windows $w$ after $t$ do
            Calculate short-term energy:
            $E_w = \left(\sum_{i=1}^{W} w_x(i)^2 + w_y(i)^2 + w_z(i)^2\right)^{1/2}$
            if $E_w > e$ then                           ▷ Energy thresholding
                $T = T \cup timestamp(w)$
            else if $|max(T) - min(T)| > d$ then
                $T_{rp} = T_{rp} \cup \{(min(T), max(T))\}$
                break                                   ▷ Start search for a new $t$
            else
                $T = \emptyset$                         ▷ Restart search
end procedure
```

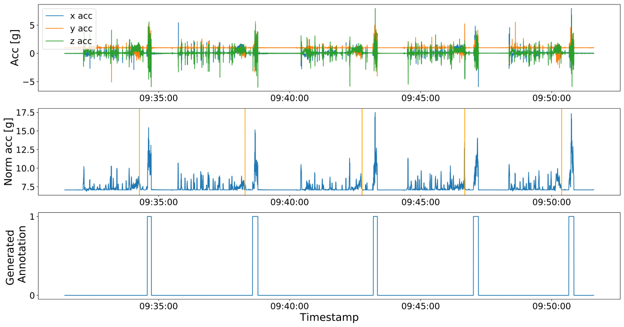

Figure 4: (Top) Acceleration signals for the session of five climbs in the training set. (Middle) Short-term energy of frames of 1s length and a shift of 1 sample. Vertical bars indicate *lowered* events. (Bottom) Generated annotation synchronized with the acceleration signal. 1: *rope pulling*, 0: not *rope pulling*.

the minimum duration of *rope pulling*. Algorithm 1 summarises the described procedure.

The aforementioned threshold is the one providing better system performance over a range of threshold values and using the training set of sensor recordings. For this, the actual start and end of *rope pulling* activities were transcribed from the video recording and used to manually annotate three-axial acceleration signals. A number of threshold values, ranging from relatively low to relatively high short-term energies, was tested, and the *best* threshold was determined. The *best* threshold is defined as the threshold, which resulted in the smallest difference between generated and manual annotation. During the later measurement of detection performance, the same similarity metric, known as the Jaccard index, was used. An illustration of the hybrid approach is shown in Fig. 4.

4.2 Video-based approach

In the video-only based approach, we trained a CNN for object detection via transfer learning to localize the position of a yellow rectangle that is attached around a target quickdraw and to identify whether the quickdraw has a rope clipped into [26].

This method exploits video data for detecting the *end* of the *rope pulling* activity, while the *start* is roughly estimated by assuming that all *rope pulling* events have the same duration, as the average duration of the activity in the first session of climbs. The hypothesis is that the video of a quickdraw is a valuable source of information: by the detection of rope inside a quickdraw, we can understand, whether a climber is still performing some activity. Furthermore, when the rope is not present there, then the *rope pulling* has happened, and the climber is not using the route anymore. As previously mentioned, we consider the second quickdraw, but the described system can be used for the other quickdraw positions as well. Using the second quickdraw has two advantages: it is easy to install the system and—in our experimental setting—it is almost perpendicular to a camera filming from the ground. If the camera is installed on the ceiling of a climbing gym, it might be better to observe one of the top quickdraws even though that will not record some climbing attempts where the climber did not reach the top.

Hence, we have develop an object detection procedure for the *target quickdraw* with or without a rope based on a CNN (the term *target* relates to an area that is clearly marked, e.g., with a light-colored adhesive strip forming a rectangle as in Fig. 2, bottom). Usually, object detection combines two sub-tasks: the detection of the location of an object in the image (with a confidence score), and the prediction of the class label for the object[16]. Existing CNNs are capable of localizing objects with high accuracy [17, 30], but quickdraws are too specific and not detected. Nevertheless, transfer learning allows us to retrain a network to detect new objects [31]. To implement it, we manually annotated all the images in videos so that for each image we created a text file with the location of a target quickdraw and the label for the object (where the label *rope* indicates a rope inside the region, and *no_rope*, the absence of it, see Fig. 2 (bottom)). In this way, we obtained an object detection problem with two objects to detect: *no_rope* and *rope*. We employed a CNN RetinaNet with a ResNet-50 backbone, pre-trained on the COCO2017 [25] dataset. The network is trained on a set of annotated video frames from our training video. Since it has been shown that dataset augmentation can improve the prediction accuracy of a model (e.g. [9]), we augmented the training dataset with images that were selected from the same dataset by applying transformation, such as increasing or decreasing brightness, contrast and RGB colors up to 10%, rotation by up to 10%, shifting within a frame by up to 6% of width and resizing the image by up to 20% or the original size. In that way, we created a dataset where 24989 images are annotated as *rope* and 12100 as *no_rope*. To speed up the training process, we used freezing hidden layers [8]. We trained RetinaNet for 20 epochs using a GPU.

To evaluate the model, we used images from the second filming session: 13965 of them were manually labeled as *rope* and 6607 as *no_rope*. For every test image, the model produces the list of objects present in the image along with their position and probability. For every frame in a test set, the model found at least one object. For some frames, the model produced more than one objects as a resulting prediction, but we only considered the object with maximum probability as a label for the frame.

The proposed model was able to discriminate quickdraws with or without a rope in a fairly accurate way. Fig. 5 illustrates the prediction results of the network evaluation on the testing video recording. The vertical axis represents the class prediction for the detected object (when the red line is equal to zero, there is no rope in the detected object; when it is close to one, there is a rope). The gray line is a normalized position of a climber on the wall. The frames with predicted class 0 are clearly occurring when the climber climbs the wall, i.e., when the gray dotted curve clearly moves from the bottom to the top position and from the top to bottom back, indicating the movement of a person who ascends the wall. Furthermore, it follows the activity of *rope pulling*, i.e., when the gray curve remains at the bottom part of the figure, indicating that the climber stays on the floor and removes the rope from the wall before starting the next route. Although after training the network, we were able to detect objects with high accuracy, there were some inaccuracies in the predictions (Fig. 5, top). They usually happen at the beginning of an ascent and at the end of it: at the beginning, the model can not distinguish the object when, for example, the quickdraw is hidden by the hand of a person when they clip; at the end of the ascent, the model is confused when there are two ropes in the region. To remove them, we assigned the label for a frame to be *no_rope* only if the next 90 out 100 frames are detected as *no_rope*; thus it is clear that in the next 3 seconds of video, the model is confident about *no_rope* prediction (see formula 1).

$$c'_i = \begin{cases} 0 & \text{if} \quad \sum_{j=i}^{j=i+99} c_j <= 10, \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

where c'_i is the newly obtained class label for frame i, c_j is the predicted class label for frame j, $\sum_{j=i}^{j=i+99} c_j$ is the amount of *rope* objects for a window of 100 consecutive frames after the target frame (including the frame itself).

After having obtained the sequence of the frame labels, the end of a *rope pulling* activity is detected as a change point (from 1–*rope* to 0–*no_rope*). However, the beginning of the *rope pulling* activity cannot be identified from video data in a similar direct way. In particular, such an effort would require extracting poses of a climber (e.g., as done in [10]) for which a different sensing approach (i.e., depth camera) is better suited. Hence, we take a less complex approach, which yet yields a good estimate of the activity start, as we show later. Namely, by using the training dataset, we measured that *rope pulling* activity lasts on average for 12.3 seconds. Moreover, the standard deviation of 1.3 seconds suggested that an individual activity duration tends to deviate only slightly from the mean value. Taking these facts into account, we determine the start of *rope pulling* activity by subtracting the observed mean duration of 12.3s from the detected activity end timestamp.

5 EVALUATION AND DISCUSSION

We evaluate the proposed approaches for *rope pulling* detection by using the video and sensor recordings, on the data collected during the second climbing session consisting of five climbs. After the start and the end timestamps of *rope pulling* activities were computed

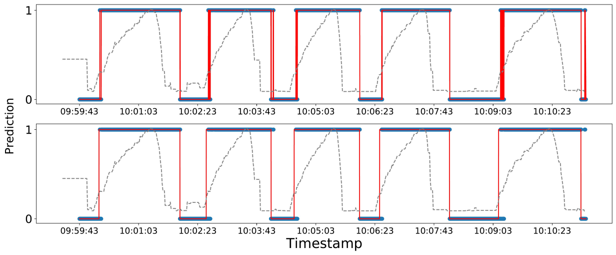

Figure 5: (Top) Predicted object labels by retrained RetinaNet (red line) and ground truth labels (blue line) for video frames in the test dataset. x-axis indicates the frame timestamp and y-axis indicates the prediction: 0: *no_rope*, 1: *rope*. Climber's trajectory is shown as a dotted grey line. When the climber is on the wall, the prediction is *rope*, in the opposite, when the climber is on the floor, the prediction is *no_rope*. Prediction errors are visible at the beginning and end of an ascent (the red line deviates from the blue line). (Bottom) Class membership after thresholding defined in Equation 1.

by the proposed methods, every sample in the sensor dataset was also manually annotated based on the acquisition timestamp. That enables to compute the actual accuracy of the proposed methods predictions.

We report detection accuracy using the Jaccard index-based measure for each *rope pulling* occurrence. Namely, the reported index represents the degree of similarity (ranging from 0 to 1) between two sets: one consists of sample timestamps automatically generated, and the other consists of sample timestamps in the ground truth. As mentioned above, the true start and end of *rope pulling* were manually transcribed from the video recordings of the climbing session. The start time is taken when the climber grasps the rope while the end of the activity is taken when the rope leaves the sensor enhanced quickdraw. If M is the set of *rope pulling* sample timestamps based on the ground truth and A is the set of sample timestamps that were automatically annotated as *rope pulling*, then the recognition accuracy for the climb is calculated as the Jaccard index of similarity for the two sets using $J(M, A) = \frac{|M \cap A|}{|M \cup A|}$.. The results are visualized for each climb independently in Figure 6. Table 1 summarizes the evaluation results for each climb in the test dataset.

The evaluation resulted in recognition rates of 91% for the hybrid and 76% for the video-based approach, respectively. This difference is statistically significant (t-test, 0.05 p). While the latter approach resulted in generally more accurate activity detection, the video-based method proved to be more precise in detecting specifically the end of a *rope pulling* event.

An important aspect to consider is the complexity and the required resources of the two proposed methods. Both utilize video-analysis: the hybrid approach is based on person detection and employs a pre-trained network, which can be run on a CPU, but for a real-time scenario GPU is required. The video-based method involves transfer learning for a quickdraw detection which needs a GPU for training. In our experiments, we have used a NVidia 2080

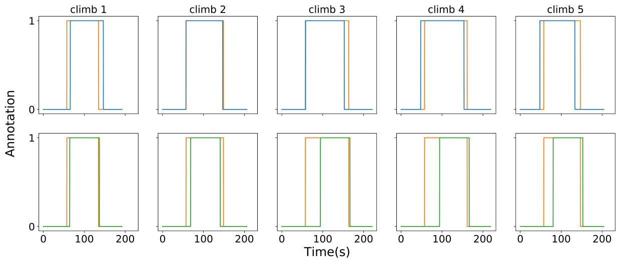

Figure 6: Performance of rope pulling detection on the test dataset. Orange: Manual annotation transcribed from the video recording, blue: annotation generated using the hybrid approach, green: annotation generated using the video-based approach.

Table 1: Performance of rope pulling detection using video-based and hybrid approaches on the test dataset.

Climb id	Hybrid	Video
1	0.89	0.91
2	0.99	0.86
3	0.89	0.65
4	0.93	0.65
5	0.85	0.74
Average	0.91	0.76

Ti GPU[3] on a machine with 126 GB of RAM, 20 TB of SSD harddrive and an 8 core CPU (Intel Xeon W-2123 3.6G). The training time was about 4 hours for object detection (the data is 37089 images, with the image size of 1920 × 1080 pixels).

Hence, we believe that the obtained results are satisfying, however, each has advantages and disadvantages that are discussed in the following:

- Using the hybrid method we obtained a higher precision, though the approach requires two data sources: a sensor and a camera and therefore requires more effort to set up and to maintain, e.g., the sensor batteries have to be changed regularly.
- In both cases, video analysis is used, and the approval from climbers has to be obtained so that they can be recorded.
- In both cases, when the climbing gym is crowded, and multiple climbers are climbing in close proximity, failures to detect *lowered* events become more likely. While the proposed methods can be extended to address tracking multiple objects within the video [35], a camera-based detection approach may still be impractical in some settings due to the large number of cameras that would need to be installed to ensure the coverage of all routes.

6 CONCLUSION AND FUTURE WORK

In this paper, we have described two data analysis techniques that can automatically detect discrete events in the sport climbing domain. Specifically, the focus of our investigation was the detection of events **associated with the activity of pulling the rope after the ascent is finished**, using video and sensor data. The long term goal of our work is to identify effective data analytics methods to correctly label sensor data, namely, to mark a sensor reading with semantic labels describing the activity performed at that point in time. Sensor data augmented with this information can be further used to train even more sophisticated Machine Learning methods, aimed at precisely extract activity and performance information from a large batch of test sensor data.

We have so far exploited state-of-the-art for object detection techniques from computer vision to obtain 2D bounding boxes as representation of climber's motion. Using timestamps of *lowered* events inferred from the motion analysis, we have considered video and acceleration data separately for detecting the start and end of *rope pulling*.

Although the results of evaluation are satisfying, there are aspects that need more attention in the future. First, a larger number of participants and climbs will have to be included in the evaluation. This will give insights into how well the methods performs in a variety of climbing scenarios. There are two factors with implications on *rope pulling* duration that were not considered. Namely, the activity will last longer for the higher routes and, secondly, different climbers may pull the rope at different speed. Current study involved two participants for which the standard deviation of rope pulling duration was 1.1 seconds; thus lower than the standard deviation for only the training dataset, which consisted of data of one participant. With increased number of climbers, the average activity duration has to be revised. This may well result in higher detection accuracy for the video-based method as the standard deviation of rope pulling duration tends to decrease. Only successfully completed climbs were used in the study, while in real conditions, climbers may lower to the ground following a fall off the wall. Partially climbed routes will presumably result in recognition failures. This might not be a flaw if the intended used of the system is counting the number of successfully completed climbs. However, this problem could be mitigated by extending methods to include *fall* detection, as done in [36], for example.

To get closer to the goal of making a *rope pulling* detection system suitable for deployment, optimization for sensor battery consumption would be desirable. Sensors attached to quickdraws require batteries, which discharge fast if data is collected and transferred at high frequency. In this study, data were recorded at the sampling rate of 50Hz, thus allowing sensor to work only for some hours. In the future we need to identify the lowest frequency (therefore the most battery saving) that we can use so that event detection methods (such as the once described in this paper) still achieve satisfying results. Business goals will determine what "satisfying" means. If the major goal is to roughly estimate route usage, a lower accuracy might be enough. Conversely, if the goal is to detect particular climbing moves or situations within a climb, a higher frequency might instead be needed.

Detection of the *rope pulling* activity is clearly useful for climbing gym management. A byproduct is that the frequency of rope pulling provides a direct measure of how often sections of the wall are climbed as an indicator of popularity. Based on these methods, a system for providing performance feedback to a climber could be developed. Namely, the *rope pulling* activity segments a climbing

[3] https://www.nvidia.com/it-it/geforce/graphics-cards/rtx-2080-ti/

session into individual climbs, thus climb duration and resting time can be derived directly. In fact, a common method in endurance training is to climb a route in quick repetitions with short rest intervals between repetitions [29]. Further segmentation would make it possible to detect climber's performance at a finer granularity, i.e., in sections between *rope clipping* events or to detect problems (e.g., a *falling* event) that occur at specific parts of the route.

Of further interest is to devise techniques for activity tracking in "bouldering" areas in climbing gyms, where climbing is done without the use of the rope and quickdraws. These areas are equipped with cameras for security reasons as climbers can get injured as a result of a fall on the bouldering mat. A system that is able to identify dangerous situations using video analysis techniques and that warns climbing gym staff could improve the safety of the sport.

In conclusion, we believe that soon automatic systems that provide usage analytics and climbing performance feedback will become a reality in every indoor sport climbing gym. The results presented in this study show a promising step towards achieving this goal.

ACKNOWLEDGMENTS

This work has been partly supported by the project "Sensors and data for the analysis of sports activities (SALSA)", funded by the EFRE-FESR programme 2014-2020 (CUP: I56C19000110009).

REFERENCES

[1] Reza Afrouzian, Hadi Seyedarabi, and Shohreh Kasaei. 2016. Pose estimation of soccer players using multiple uncalibrated cameras. *Multimedia Tools and Applications* 75, 12 (2016).

[2] Jürgen Assfalg, Marco Bertini, Carlo Colombo, Alberto Del Bimbo, and Walter Nunziati. 2003. Semantic annotation of soccer videos: automatic highlights identification. *Computer Vision and Image Understanding* 92, 2 (2003). Special Issue on Video Retrieval and Summarization.

[3] International Sports Engineering Association, Eckehard Fozzy Moritz, Steve Haake, et al. 2006. *The engineering of sport 6*. Vol. 6. Springer.

[4] Akin Avci, Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu, and Paul Havinga. 2010. Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey. In *23th International conference on architecture of computing systems 2010*. VDE.

[5] Oresti Banos, Juan-Manuel Galvez, Miguel Damas, Hector Pomares, and Ignacio Rojas. 2014. Window Size Impact in Human Activity Recognition. *Sensors* 14, 4 (April 2014).

[6] Akram Bayat, Marc Pomplun, and Duc A. Tran. 2014. A Study on Human Activity Recognition Using Accelerometer Data from Smartphones. *Procedia Computer Science* 34 (2014).

[7] Jeremie Boulanger, Ludovic Seifert, Romain Herault, and Jean-Francois Coeurjolly. 2016. Automatic Sensor-Based Detection and Classification of Climbing Activities. *IEEE Sensors Journal* 16, 3 (Feb. 2016).

[8] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. 2017. Freezeout: Accelerate training by progressively freezing layers. *arXiv preprint arXiv:1706.04983* (2017).

[9] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. 2020. Albumentations: Fast and Flexible Image Augmentations. *Information* 11, 2 (2020).

[10] Kyungsik Cha, Eun-Young Lee, Myeong-Hyeon Heo, Kyu-Cheol Shin, Jonghee Son, and Dongho Kim. 2015. Analysis of Climbing Postures and Movements in Sport Climbing for Realistic 3D Climbing Animations. *Procedia Engineering* 112 (2015).

[11] John Friar Chris Parsons. 2019. Modular interactive climbing wall system using touch-sensitive, illuminated climbing holds, and controller. Patent no. US 2019 329 113A1, https://patentswarm.com/patents/US20190329113A1.

[12] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook. 2012. Simple and Complex Activity Recognition through Smart Phones. In *2012 Eighth International Conference on Intelligent Environments*.

[13] Moritz Einfalt, Charles Dampeyrou, Dan Zecha, and Rainer Lienhart. 2019. Frame-Level Event Detection in Athletics Videos with Pose-Based Convolutional Sequence Networks. In *Proc. of the 2nd International Workshop on Multimedia Content Analysis in Sports*. Association for Computing Machinery.

[14] Franz Konstantin Fuss and Günther Niegl. 2008. Instrumented climbing holds and performance analysis in sport climbing. *Sports Technology* 1 (Jan. 2008).

[15] André Gensler and Bernhard Sick. 2018. Performing event detection in time series with SwiftEvent: an algorithm with supervised learning of detection criteria. *Pattern Analysis and Applications* 21, 2 (2018).

[16] David Hall, Feras Dayoub, John Skinner, Haoyang Zhang, Dimity Miller, Peter Corke, Gustavo Carneiro, Anelia Angelova, and Niko Sünderhauf. 2018. Probabilistic Object Detection: Definition and Evaluation. arXiv:1811.10800 [cs.CV]

[17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. *CoRR* abs/1703.06870 (2017). arXiv:1703.06870 http://arxiv.org/abs/1703.06870

[18] Raine Kajastila and Perttu Hämäläinen. 2014. Augmented Climbing: Interacting with Projected Graphics on a Climbing Wall. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA.

[19] Eamonn Keogh, Selina Chu, David Hart, and Michael Pazzani. 2004. Segmenting Time Series: a Survey and Novel Approach. In *Series in Machine Perception and Artificial Intelligence*. World Scientific.

[20] Jungsoo Kim, Daniel Chung, and Ilju Ko. 2017. A climbing motion recognition method using anatomical information for screen climbing games. *Human-centric Computing and Information Sciences* 7, 1 (2017).

[21] Felix Kosmalla, Florian Daiber, and Antonio Krüger. 2015. ClimbSense: Automatic Climbing Route Recognition Using Wrist-Worn Inertia Measurement Units. In *Proc. of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (Seoul, Republic of Korea). ACM Press, New York, NY, USA.

[22] Cassim Ladha, Nils Y. Hammerla, Patrick Olivier, and Thomas Plötz. 2013. ClimbAX. In *Proc. of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*. ACM Press.

[23] Mats Liljedahl, Stefan Lindberg, and Jan Berg. 2005. Digiwall: An Interactive Climbing Wall. In *Proc. of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*. ACM Press, New York, NY, USA.

[24] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. 2017. Focal Loss for Dense Object Detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*.

[25] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer.

[26] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. 2019. Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision* 128, 2 (Oct. 2019).

[27] Hisakazu Ouchi, Yoshifumi Nishida, Ilwoong Kim, Yoichi Motomura, and Hiroshi Mizoguchi. 2010. Detecting and Modeling Play Behavior Using Sensor-Embedded Rock-Climbing Equipment. In *Proc. of the 9th International Conference on Interaction Design and Children* (Barcelona, Spain). Association for Computing Machinery, New York, NY, USA.

[28] Julien Pansiot, Rachel C. King, Douglas G. McIlwraith, Benny P. L. Lo, and Guang-Zhong Yang. 2008. ClimBSN: Climber performance monitoring with BSN. In *2008 5th International Summer School and Symposium on Medical Devices and Biosensors*. IEEE.

[29] Kevin C. Phillips, Joseph M. Sassaman, and James M. Smoliga. 2012. Optimizing Rock Climbing Performance Through Sport-Specific Strength and Conditioning. *Strength and Conditioning Journal* 34, 3 (June 2012).

[30] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. *CoRR* abs/1804.02767 (2018). arXiv:1804.02767 http://arxiv.org/abs/1804.02767

[31] Dipanjan Sarkar, Raghav Bali, and Tamoghna Ghosh. 2018. *Hands-On Transfer Learning with Python: Implement Advanced Deep Learning and Neural Network Models Using TensorFlow and Keras*. Packt Publishing.

[32] Thomas Schmid, Roy Shea, Jonathan Friedman, and Mani B. Srivastava. 2007. p565-smith.pdf. In *Proc. of the 6th international conference on Information processing in sensor networks - IPSN '07*. ACM Press.

[33] Sergey Evgenievich Shtekhin, Denis Konstantinovich Karachev, and Iustina Alekseevna Ivanova. 2020. Computer vision system for working time estimation by human activities detection in video frames. *Proc. of the Institute for System Programming of the RAS* 32, 1 (2020).

[34] F. Sibella, I. Frosio, F. Schena, and N.A. Borghese. 2007. 3D analysis of the body center of mass in rock climbing. *Human Movement Science* 26 (Dec. 2007).

[35] J. C. Silveira Jacques Junior, S. R. Musse, and C. R. Jung. 2010. Crowd Analysis Using Computer Vision Techniques. *IEEE Signal Processing Magazine* 27, 5 (2010).

[36] Vinay Vishwakarma, Chittaranjan Mandal, and Shamik Sural. [n.d.]. Automatic Detection of Human Fall in Video. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg.

[37] Ali Yousefi, Alireza A Dibazar, and Theodore W Berger. 2008. Intelligent fence intrusion detection system: detection of intentional fence breaching and recognition of fence climbing. In *2008 IEEE Conference on Technologies for Homeland Security*. IEEE.