# GymCam Critique

James Peralta
Department of Computer Science
University of Calgary
Calgary, Canada
james.peralta@ucalgary.ca

## I. Overview

GymCam [1] is a new technology that uses computer vision to detect, recognize, and track simultaneous exercises. This paper was published by the ACM (Association for Computing Machinery) in December 2018. They identified that people were failing to exercise due to a lack of motivation and feedback, so they introduced GymCam which provides feedback and workout tracking automatically without any user intervention. The goal for this technology is to gamify the experience of progression in the gym to tackle the lack of motivation. There have been previous commercial attempts at automatically tracking workouts such as the Fitbit, Apple Watch, Gymaholic, and Gymatic. All of the current solutions either use wearables or add sensors to equipment but are limited depending on where the sensor is attached. For example, a sensor placed on the arm can not detect a leg press. GymCam attempts to overcome the weaknesses of previous technology by using a Computer Vision based system. They were able to segment exercises from other activities with an accuracy of 84.6%, recognize the type of workout with an accuracy of 93.6%, and count repetition with a +/- of 1.7 repetitions.

To collect data, the Carnegie Mellon University allowed them to record the fitness center during a 5 day period. They were allowed to record anonymously as long as they blurred the faces of all users immediately. They also placed the camera 13 feet high so it was out of sight from users so they wouldn't know they are being recorded. All of this was done to maintain an unconstrained environment to reduce observer effects where the participants may change the way they perform their routines if they knew they were being studied. After the 5 days, they cleaned all of the footage collected by dropping all of the footage where the gym was empty which left them with 52 hours of footage. From the 52 hours of footage they decided to choose the best 15 hours of footage for the annotation phase. Two students were recruited and trained to annotate the footage with a start and end frame for the exercise, the repetitions of the exercise, and the exercise type. After all of the footage was annotated, they we're left with a dataset of 528 sets of exercises from 18 different exercise types. The goal of the algorithm was to label a given clip with a start/end time, repetition count, and exercise type without any user intervention.

Their solution had four steps, identifying movements and classifying them as repetitive or not, combining all movements into exercise clusters, deriving a combined trajectory that represents the exercise cluster, and then running a classifier on the combined trajectory to determine the repetition count and exercise type. To detect movements, they inputted a frame into Wang et al's [2] optical flow to extract motion trajectories from the frame. Each motion trajectory was converted into a feature vector consisting of 27 features and then a multilayer perceptron was used to classify each vector as an exercise or not. The next step was to combine all movements into exercise clusters, this was needed because exercises are often captured by many key point motion trajectories. To accurately determine which trajectory belonged in each cluster, they clustered using the spatiotemporal distribution of motion trajectories and also the phase-differences. Next to collect the combined trajectory, they simply averaged all points in the cluster. With the combined trajectory they regenerated the 27 features mentioned earlier. From here they passed these features into two separate multilayer perceptrons, one for counting repetitions and the other for exercise recognition. This approach achieved 84.6% when detecting exercises with a false positive rate of 13.5%, +-1.7 for counting reps with a with a standard deviation of 2.64, and 80.6% on 18 different exercises for exercise recognition. They mentioned that the biggest limitation for their approach was they only relied on repetitive motion for identifying and counting exercises. This caused some issues when identifying exercises. For example, when two or more exercises were happening close to each other, it would cluster these two exercises into one. Also when a user is strength training, they show high variance between repetitions due to fatigue or the loss of pace.

## II. Strengths

The GymCam technology is being introduced at a good time where many people are becoming health conscious and are looking for tools to help motivate them in the gym. The goal of the study was clear and succinct. Their paper had many technical terms but they were able to abstract away the details and still provide informative points. GymCam falls into the area of persuasive technology that performs exercise tracking which is beginning to really take off with many commercial offerings. Most of the commercial products are currently motion-sensor wearables which have pitfalls depending on their placement so there is a need for innovation in a system that detects workouts externally. The authors have a solid understanding of the related work around them in both commercial products and academia. They mentioned many of the top players such as the Fitbit, AppleWatch, Gymaholic, and Gymatic. They were able to thoroughly address the other ways this problem can be solved through wearable sensors and instrumentation of equipment. Along with acknowledging the key players, they also provided the appropriate research papers.

Their data collection phase was done carefully and produced minimal biases because they did not interact with participants and they anonymously collected the data. This allowed them to capture data when users were fidgeting, stretching, and physically challenging themselves with heavy weights. These help illustrate the robustness of their approach. When explaining the algorithm, they introduced good reasons for their design choices. For example, they provided rationale for choosing Wang et al's [2] optical flow instead of Lucas-Kanade optical flow [4]. They provided direct evidence of how they derived their conclusions by explaining how they used cross validation and they even showed the confusion matrices. This paper has the potential of changing the way we develop exercise trackers. They showed that it was possible to use this approach in a truly unconstrained environment.

## III. Weaknesses

The authors draw conclusions and evaluate their results in isolation. Many of their results were to prove that computer vision was capable of this task instead of showing it was better than alternative techniques. To propel this technology as the state of the art for this problem, it should be placed in context with all alternative techniques instead of just mentioning that they exist. For example, they don't compare this approach against RecoFit [3]. This comparison should be established because they argued that their solution does not fall into the same pitfalls as all the alternative approaches such as wearables. So it would be insightful to mention how this technology pars up against RecoFit which also recognizes, and counts repetitions.

They also mentioned that their solution handles occlusion very well but do not quantify how much of that data included occlusion or how they tested this claim. If they are to make claims that this approach advances the state of the art in exercise tracking by handling occlusion, they should provide what accuracy their approach achieved in situations with high occlusion. Furthermore, they left out many facts about their dataset. For example, their dataset only contained one viewpoint. This can be problematic because depending on the viewpoint, the accuracy of their classifier can vary greatly. There are also many challenges that they did not address in their paper such as how the algorithm would handle intra-class variation, scale variation, and illumination variation. These are very common problems with computer vision techniques so they should be addressed to further validate this approaches feasibility.

## References


[1] Khurana, K. Ahuja, Z. Yu, J. Mankoff, C. Harrison, and M. Goel, "Gymcam," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 2, pp. 1–17, 12 2018.

[2] Heng Wang, Alexander Kläser, Cordelia Schmid, and Cheng-Lin Liu. 2013. Dense trajectories and motion boundary descriptors foraction recognition.International journal of computer vision103, 1 (2013), 60–79.

[3] D. Morris, T. S. Saponas, A. Guillory, and I. Kelner, "Recofit: Using a wearable sensor to find, recognize, and count repetitive exercises," pp. 3225–3234, Association for Computing Machinery, 2014.

[4] Jean-Yves Bouguet. 2001. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm.IntelCorporation5, 1-10 (2001)