Automated Fitness Tracking Using Machine Vision

Justin Kang, Andrew Dworschak[1], Jacob Budzis, Jonah Killam, Rahat Dhande and Ryan Cotsakis

Abstract— **TODO: Summarize the document**

I. INTRODUCTION

Physical exercise is one of the most important contributors to maintaining good health. Despite this, 2 out of 3 Americans fail to meet the recommended minimum amount of weekly exercise. Among reasons for this lack of exercise is the high attrition rate faced by fitness facilities. According to the International Health and Racquet Sports Association, the difficulty in receiving immediate, structured feedback on one's workout is one of the principal drivers of this attrition [1].

One way to address the problem of limited feedback in the gym is to create a workout log that allows users to easily observe trends in their exercise history. However, current products in the market are all either overly burdensome to the user or severely limited in their scope of exercises.

As an example of an overly burdensome technology, BodySpace is a mobile application that keeps track of user's exercise history provided the user enters each of their sets into the application during their workout. As a rough estimate, if the average user performs 25 sets in their workout and takes 1 minute to enter the information into their phone, then 25 minutes or half of their workout is spent typing information into their phone.

As an example of a limiting technology, eGym is a company specializing in sensor-enabled workout machines. They can only track the user's workout on several specific machines manufactured by eGym and have no potential to track free-weight exercises, the largest class of exercises. In this sense, eGym limits the fitness facility to only sourcing their machines from a single supplier, and limits the user to a small set of machine-based exercises to enjoy the automatic logging of their workout.

Outside the market, there has been many recent efforts in academic research to develop exercise classification and repetition counting capabilities [2], [3]. Most approaches in the literature focus on the analysis of accelerometer data [4], [5], [6], [7], [8]. However, given that most gym users are reluctant to wear accelerometers during their workout and the recent advances in machine vision, we anticipate that machine vision will be the technology that facilitates broad exercise monitoring capability.

We propose a system that uses machine vision to automate the classification of exercises, the weight lifted and repetitions performed, uploading the workout data in real-time to the appropriate user's fitness profile. This system gathers structured feedback without creating a burden for the user by creating a record of the user's workout history without requiring user input. Further, the system can potentially accommodate any exercise, making it more versatile than the current market alternatives.

The system is illustrated in a block diagram in Figure 1. The system is composed of three subsystems: and Exercise Classification Subsystem, a Repetition Counting Subsystem and a Weight Detection Subsystem.

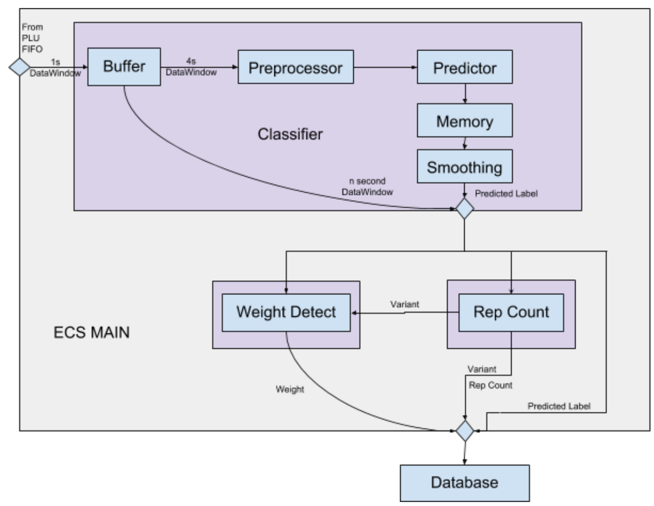

Fig. 1. Block Diagram of the System

The remainder of the paper is structured as follows. Simplifications to the problem for this iteration of the system are described in the Setup section. Each of the three subsystems are described in the Method section. The performance of the system is quantified in the Results section. Finally, we discuss our ongoing work on this project in the Conclusions section.

II. SETUP

A. Simplifications of the problem

Recall that the goals of the full system are to be able to classify any exercise performed with any weights, anywhere in the gym and to upload the data to the appropriate user's profile. However, this iteration of the system addresses only a subset of this functionality, making several simplifications:

- The system is limited to classifying 11 free-weight and body-weight exercises and their variants. Classified exercises include bicep curl (simultaneous and alternating variants), squat, sitting shoulder press, lateral arm raise, lunge (left and right variants), sit-up, dumbbell row (left and right variants), and sitting tricep extension.

[1]This article was written by Andrew Dworschak and Justin Kang, based on the research of all 6 authors dating July 2017 - April 2018.

- The system is limited to classifying three sets of color-coded weights.
- The system assumes that the user will take at least a 5 seconds break between their exercises.
- The system only uses one camera, and thus has a limited field of vision.
- The system does not incorporate the identification of users after they leave the frame.
- The system does not handle obstructions to the field of vision.

We feel that despite these simplifications, the system demonstrates the viability of the machine learning approach to exercise classification. By addressing free-weight exercises, the current bottleneck of the industry, we demonstrate the capacity of the system to outperform existing solutions.

B. Technology used

In this project, we use the Microsoft Kinect V2 camera. This camera generates RGB images, depth-based images and skeletal vectors recognizing of up to six people at a time. The two sets of images and skeletal vectors are fed into the system as inputs.

III. METHOD

The system consists of three subsystems. The Exercise Classification Subsystem receives a time series of skeletal vectors as input, determines which exercise is being performed, and when that exercise has been completed. The Repetition Counting Subsystem receives data from a completed exercise and determines the number of repetitions performed in that exercise. The Weight Detection Subsystem receives data from a completed exercise and determines which weight, if any, is being lifted by the user.

A. Exercise Classification

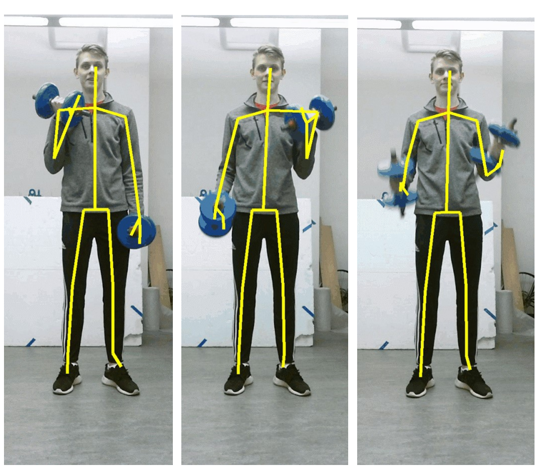

Fig. 2. Example of generated skeleton, and mapping onto color image

The Exercise Classification Subsystem consists of a preprocessing step that enforces translational and rotational invariance in the data, a predictor that takes a window of data and outputs a set of probabilities for which exercise is being performed, and a smoothing algorithm that determines at which time exercises start and end.

1) Preprocessing: Because we want our system to be robust to people performing exercises facing in different directions and standing in different parts of the frame, it is important to enforce translational and rotational invariance. To accomplish this, the preprocessing step receives the raw skeletal vectors from the kinect camera which consist of 75 points per frame representing 25 joints in 3 dimensional space. We then use 21 different skeletal angles along with the normalized (mean of 1) distance of each point to the centroid of the skeleton in each frame, for a total of 46 time series. This corresponds to a loss of data, but improves the translational and rotational invariance of the predictor.

This data is stored in a queue and delivered to the predictor to prevent the dropping of frames.

2) Predictor: The predictor employs a Recurrent Neural Network (RNN) that accepts skeletal input vectors and outputs its most likely classification of the exercise from the given list. In addition to the 11 variants, the RNN can output a 'NULL' exercise to indicate that the user is not performing any recognized exercise at that time. An RNN was the appropriate choice for this application due to the time series nature of the input data. Once trained, it exhibited an extremely quick computation time suitable for running the system in real time.

The predictor outputs one list of probabilities per second representing the various exercises the user could be performing. This data is read by the smoothing algorithm to prevent erroneous classifications.

3) Smoothing: Reading the classification probabilities from the predictor, the smoothing algorithm applies a low pass filter that ignores exercises done for only several seconds. Then, by searching for transitions from the 'NULL' classification to a valid exercise, the algorithm ascertains a start and end time for the exercise window, determines the most probable exercise over the span of the window, and labels the window with an appropriate exercise label. This procedure would be problematic if a user transitioned quickly from one exercise to another, as temporal filters may prevent the 'NULL' classification from being asserted. However, this method was observed to improve the overall accuracy of the Exercise Classification Subsystem significantly.

B. Repetition Counting

When given a labelled exercise, the Repetition Counting Subsystem selects the skeletal data associated with that exercise. This skeletal data consists of a 3-dimensional set of coordinates for each of the 25 joints recorded. First, the non-vertical components of the data are thrown away, reducing the vectors to contain only the vertical component of each joint. This simplification was appropriate because all of the exercises in our sample set were limited to free-weight and body-weight exercises, which rely on the force of gravity to provide resistance.

Next, our goal is to project the remaining 25 dimensional joint data $X(t)$ defined in the interval $[t_{start}, t_{end}]$ into a 1

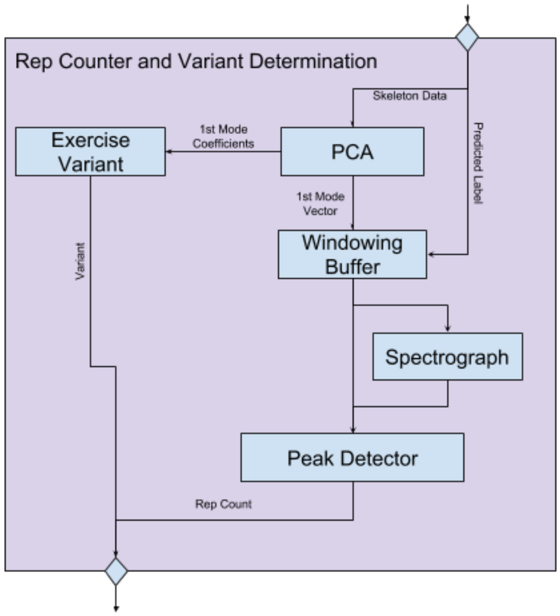

Fig. 3. Block diagram of the subsystem responsible for Repetition Counting and Variant determination

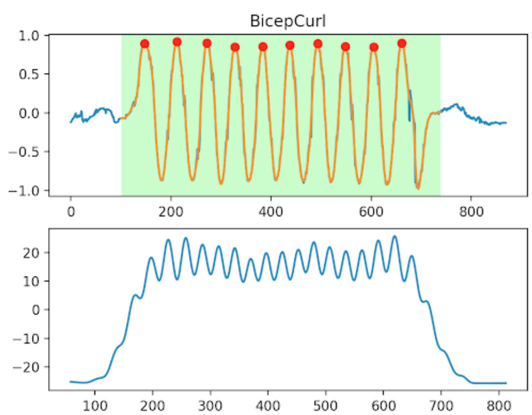

Fig. 4. Upper plot depicts the principle component of motion for the captured skeleton. Lower plot is part of a spectrogram, showing the intensity of the dominant frequency in the upper plot

dimensional signal space $y(t)$. To accomplish this, we first select the middle subset of the data by constraining $t_{start} + \tau < t < t_{end} - \tau$ for a fixed constant τ. This ignores the most uncertain portion of the data at the beginning and end of the exercise where large movements that are not part of the exercise can occur.

Using prepackaged software, we decompose $X(t) = U\Sigma V*$ using Singular Value Decomposition and select the first column of U corresponding to the eigenvector with the most variance in the data denoted w. Then, considering the full range of t:

$$y(t) = w^T X(t), t \in [t_{start}, t_{end}]$$

The upper plot of Figure 4 depicts a signal $y(t)$ generated from skeletal data of a user performing the bicep curl exercise. Due to the limitation of the classification system's ability to determine exactly when an exercises has finished, small windows of skeletal data from before and after the exercises are included.

Once a signal $y(t)$ is successfully generated, it is important to determine the exact window in which the exercises took place. First, the signal is filtered using a third order Butterworth low pass filter. The orange data set on upper plot of Figure 4 shows the output of that filter. The window is then identified by finding the dominant frequency in the entire signal using a Fourier Transform.

$$Y(\omega) = DFT(y(t))$$

$$\omega_D = argmax\,(Y(\omega))$$

A spectrogram is taken on the signal $y(t)$ using a window length that corresponds to two periods of the dominant frequency. The lower plot on 4 shows the output of the spectrogram corresponding to ω_D. By observing the first time at which the power of the dominant frequency exceeds 50% of its maximum value and the last time at which the power returns below 50%, we identify time-window when the exercise is being performed.

With this window identified, the number of reps can be counted using a simple peak detector. This is implemented by subtracting the mean from the signal and counting the peaks above zero, ignoring small oscillations around the troughs.

1) Determining the Exercise Variant: Principal Component Analysis serves as a useful tool for determining which variant of an exercise is being performed. The dumbbell row, for instance, is often performed with either the left or the right hand. Looking at the weighting of the first principal component, we determine which joints contribute most the the primary variational mode in the signal $y(t)$. If it is determined, for instance, that the component of w corresponding to the left hand has the greatest magnitude, it can be inferred that the left hand variant is being performed.

C. Weight Detection

IV. RESULTS

TODO: We classified things with xx% accuracy etc.

V. CONCLUSIONS

This project has demonstrated that the methods proposed for exercise classification and repetition counting are viable provided an accurate skeletal time series can be generated. Before this system can be implemented in a larger scale, it is paramount that systems for skeletal generation be effective over larger areas and generate a more accurate skeleton.

The next steps are to look at superior methods for skeletal generation, starting with improved sensing. Using multiple cameras, spread across a roo

ACKNOWLEDGMENT

Part of this project was completed in accordance with the Capstone course requirements for Engineering Physics at the University of British Columbia, an accredited engineering program. We thank the program's staff for their feedback and their role in shaping our work.

We also thank the Bycast Award for Entrepreneurship committee for their generosity in funding this research.

REFERENCES

[1] John McCarthy. *IHRSA's Guide to Membership Retention*. International Health and Racquet Sports Association, 2007.

[2] Mathias Sundholm, Jingyuan Cheng, Bo Zhou, Akash Sethi, and Paul Lukowicz. Smart-mat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix. pages 373–382, 09 2014.

[3] Natalia A Daz Rodrguez. *Semantic and Fuzzy Modelling for Human Behaviour Recognition in Smart Spaces*. IOS Press, AKA Verlag, 2016.

[4] Igor Pernek, Karin Anna Hummel, and Peter Kokol. Exercise repetition detection for resistance training based on smartphones. *Personal Ubiquitous Comput.*, 17(4):771–782, April 2013.

[5] Chuanjiang Li, Minrui Fei, Huosheng Hu, and Ziming Qi. Free weight exercises recognition based on dynamic time warping of acceleration data. pages 178–185, 2013.

[6] Keng-hao Chang, Mike Y. Chen, and John Canny. Tracking free-weight exercises. pages 19–37, 2007.

[7] Hristo Novatchkov and Arnold Baca. Machine learning methods for the automatic evaluation of exercises on sensor-equipped weight training machines. *Procedia Engineering*, 34:562 – 567, 2012. ENGINEERING OF SPORT CONFERENCE 2012.

[8] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, and Hugo Fuks. Qualitative activity recognition of weight lifting exercises. 2013.