ExerSense: Real-time Exercise Segmentation, Classification and Counting using IMU

Shun Ishii[1] Anna Yokokubo[1] Mika Luimula[2] Guillaume Lopez[1]

Abstract: Even though it is well known that physical exercises have numerous emotional and physical health benefits, maintaining a regular exercise routine is quite challenging. Fortunately, there exist technologies that promote us to do physical activities. Nonetheless, almost all of these technologies only target a narrow set of physical exercises (e.g., either running or physical workouts but not both) and are only applicable either in indoor or in outdoor environments, but do not work well in both environments. This paper introduces ExerSense, a real-time segmentation and classification algorithm that recognizes physical exercises, and that works well in both indoor and outdoor environments. The proposed algorithm achieves a 95% classification accuracy for five indoor and outdoor exercises, including segmentation error. This accuracy is similar or better than previous works that handled only indoor workouts, and those use a vision-based approach. Moreover, while comparable machine learning-based approaches need many training data, the proposed correlation-based method needs only one sample of motion data of each target exercise.

1. Introduction

While regular exercises provide us numerous benefits, including healthier life and emotional well-being [1], maintaining regular physical activity is challenging. Fortunately, research studies have demonstrated that automatic tracking can motivate physical activities [2]. However, almost all of the current products target only walking/running or indoor specific workouts separately. In contrast, most of the exercises and sports that have positive effects on health are combinations of running and other workouts.

This research aims to develop an algorithm that provides a very accurate real-time segmentation, classification, and counting of physical exercises, including both indoor and outdoor, using only one sensor data collected from the general usage wearable devices.

2. Related Work

Dan et al., introduced RecoFit, a system for automatically tracking repetitive exercises such as weight training and calisthenics via an arm-worn inertial sensor [3]. They identify exercise periodes with precision and recall greater than 95%, recognize with 96% accuracy 13 exercises, and count accurately to ± 1 repetition 93% of the time.

Prakash et al. introduced the advantages of an earable device, eSense [4], in counting the number of steps a user has walked [5]. While head movement can still pollute this bouncing signal, they developed methods to alleviate the problem. Results show 95% step-count accuracy even

1 Aoyama Gakuin University, Japan
2 Turku University of Applied Sciences, Finland

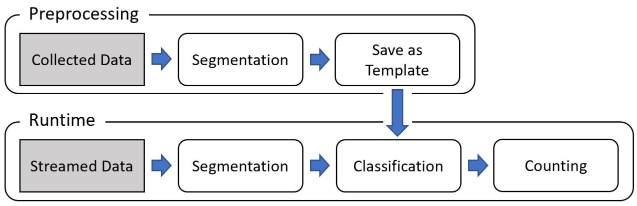

Fig. 1 Schematic of the architecture of the proposed method

in the most challenging test case, very slow walk, where smartphone and Fitbit-like systems falter. Importantly, their system is robust to changes in walking patterns and scales well across different users.

Though most of the related works achieved around 95% exercise recognition accuracy, it is under the condition of only indoor workouts or outdoor exercises. Also, machine learning models need to be re-trained if one want to add a new exercise. On the other hand, only a few samples are necessary for a correlation approach. This research aims to recognize both indoor and outdoor exercises with around 95% accuracy using a correlation-based method.

3. Proposed Method

3.1 Outline of ExerSense

The proposed method to segment, classify, and count physical exercise in real-time is separated into two stages: pre-processing and runtime (Figure 1). The pre-processing stage collects some targeted exercise data and saves a single motion as a template 3-Dimensional (3-D) acceleration for the exercise. The runtime stage first segments the acceleration signal into a single motion.Then, it uses a correlation-based algorithm to classify every segment.Finally, the count of exercise is incremented.

3.2 Details of Segmentation

Figure 2 illustrates the process of segmentation algorithm from a 3-D acceleration signal collected during push-ups activity. First, the synthetic acceleration of streamed inertial sensor data, which is the norm of the 3-D acceleration signal, is calculated. The synthetic acceleration is more appropriate than one axis acceleration, though it as the disadvantage of reducing the differences between movements that are similar but along a different axis.

Then, as shown in the second plot of Figure 2, the result of the norm includes much noise. Applying short-term energy, enables not only to emphasize significant signal variations, but also to smooth it as shown in the third plot. Smoothing is important to detect a motion start, and the end peaks easily. Then, a sliding window of 0.25[s] length is used to detect peaks. If the center value of the window is the maximum value of the window, it is determined as a peak. The fourth plot shows detected peaks plotted on the smoothed norm of acceleration signal collected during push-ups exercise. Finally, the 3-D acceleration signal is segmented by extracting the data between the period of two consecutive peaks.

In most cases, one peak is detected for each motion. However, in the case of sit-ups, three peaks are detected for each motion (see Figure 3). To be able to deal with this case, one of the three peak-to-peak periods (yellow-colored in Figure 3) is defined as sit-up base motion, and when it is detected, the latter two segments are ignored. If the person starts another exercise after sit-ups, there is no problem for counting because for motions such standing up, sitting down, and other posture changes, the smoothed synthetic acceleration also includes more than two peaks.

In order to use the correlation-based approach, templates data of all targeted exercises are necessary. The segmentation algorithm above described was used to get one motion data of each exercise. For each exercise, one of the detected segments of motion data from a peak to the next peak was selected randomly as the template. Compared with the machine learning approach, what is needed is just doing the exercise for once or a few times to collect template data.

3.3 Details of Classification

After extracting the 3-D acceleration signal corresponding to a single motion through the segmentation process, the raw data are used to calculate the correlation with every exercise's templates. The Dynamic Time Warping (DTW, Algorithm 1) algorithm is applied to calculate the distance between every template signal and the extracted signals. The DTW calculates the distance between two time-series data with different lengths. This property is crucial since it enables dealing with the shape of signals issued from one identical exercise, independently of the speed of the exercise motion.

Then, multiple weights are applied to the calculated DTW score for each exercise template. The variance of

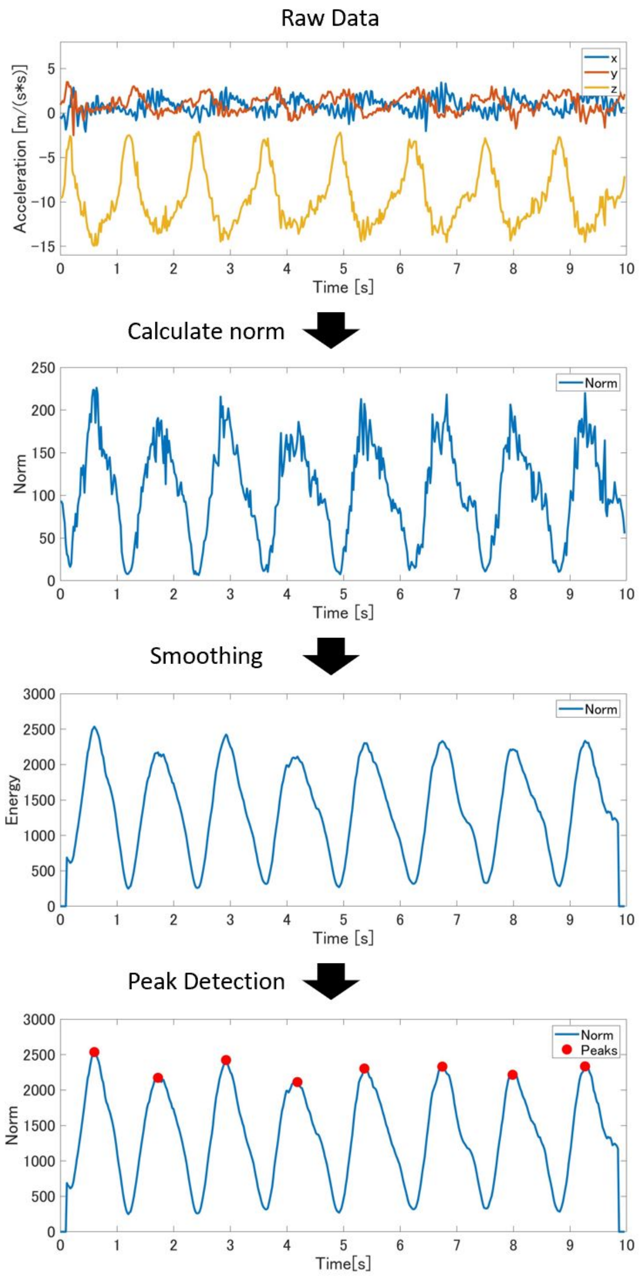

Fig. 2 Example of the processing flow of segmentation algorithm applied to 3-D acceleration signal collected during push-ups exercise

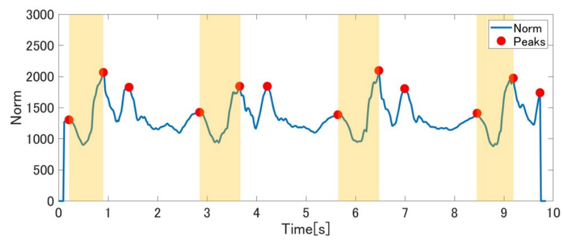

Fig. 3 The smoothed synthetic acceleration signal and detected peaks during sit-ups exercise

the three axes determines the weights. For example, if the maximum variance of axis is Z-axis, the weight of "Exercise 2" is "0.9." The direction of maximum movement and body influence the final DTW scores, which are used for classification. It is an essential point of the proposed

Algorithm 1 Dynamic Time Warping

Require: $n > 0 \wedge m > 0$
  $DTW \Leftarrow array[0...n, 0...m]$
  for $i = 0 \ldots n$ **do**
    $DTW[i, 0] \Leftarrow infinity$
  end for
  for $j = 0 \ldots m$ **do**
    $DTW[0, j] \Leftarrow infinity$
  end for
  $DTW[0, 0] \Leftarrow 0$
  for $i = 0 \ldots n$ **do**
    for $j = 0 \ldots m$ **do**
      $cost \Leftarrow ||(s[i] - t[j])||$
      $DTW[i, j] \Leftarrow cost + min(DTW[i-1, j], DTW[i, j-1], DTW[i-1, j-1])$
    end for
  end for
  return $DTW[n, m]$

method. Finally, the proposed method classifies the exercise that has the minimum weighted DTW score as the ongoing exercise. However, if the minimum score is higher than a pre-defined threshold, the segment is ignored as non-exercise.

4. Validation

4.1 Conditions

As mentioned in the introduction, this research targets both indoor and outdoor regular exercises. A circuit of five regular exercises, performed in a random order, has been created to evaluate the proposed method.

Fifteen participants have been recruited from University students. Participants varied in weight from 58kg to 80kg, and also they self assessed as performing exercise "at least once a week," with an average of four times a week. Each participant performed once all exercises according to the conditions described above. Due to the missing value of six participants, we used valid data of nine participants to validate the proposed method.

Though this research aims at being deployable with various types of commercially available general use wearable devices such as smartwatches, smart earphones, smart glasses, and chest bands, the study reported in this thesis is based on the signal of an IMU that is mounted to the chest to develop basic classification algorithm (Figure 4). One can assume that the chest movement, as the head, has less noise than other body parts. Chest sensors are also commonly used by people practicing exercise several times a week to monitor their heart rate.

The proposed method has been evaluated and validated with the five exercise. Participants run and walk more than 20 steps each without caring whether they start with the right or left foot. They also perform jumps, push-ups, and sit-ups around ten times each. The author itself performed the exercises to produce the templates for all five exercises, which are necessary for real-time classification.

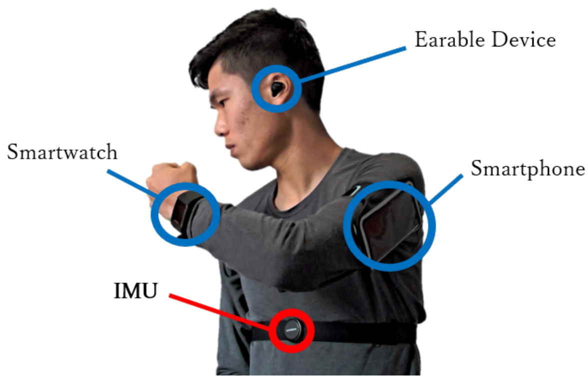

Fig. 4 Sensors used for the experiment and their positionings

4.2 Results

Table 1 shows the results of the validation of the proposed method using data collected through the experiments described in the previous section. "Truth" corresponds to the labels that were assessed watching at the video recorded during each exercise practice. "Predicted" corresponds to the labels that have been classified by the proposed method. "Overlooked" corresponds to uncounted exercise even though labeled. Contrarily, "mistook" corresponds to the number of classified labels though there was no exercise.

The precision, the recall, and the F1-score have been defined as following to evaluate the segmentation algorithm from the results of Table 1.

- The precision is the fraction of predicted exercise segments that correspond to actual exercise.
- The recall is the fraction of actual exercise segments predicted as exercise.
- The F1-score is the harmonic mean of precision and recall.

Table 2 shows the results of the segmentation algorithm and classification method. The proposed method achieved approximately 95% classification accuracy, although it includes segmentation error. Some overlooked segments are observed because of their low similarity with the templates, while some mistook exercises can be observed because of the high similarity between transition movements and template exercises.

4.3 Discussion

Nowadays, most classification methods use machine learning because it is incredibly accurate. However, in this case, it is complicated to collect enough data to train the model adequate for each exercise and each person. On the other hand, proposed correlation-based approach needs only one motion sample of each exercise, and does not require individual calibration. So, it is straightforward to deploy.

The proposed method classifies exercises after segmentation. If the segment does not match any of the available exercise templates, it ignores it. So, the overlooked segments were ignored because of their low similarity with

Table 1 Confusion matrix of Truth and Predicted labels

		Predicted					Overlooked
		Running	Walking	Jumping	Push-ups	Sit-ups	
Truth	Running	231	6	0	0	0	3
	Walking	0	272	0	0	2	22
	Jumping	0	0	80	0	0	11
	Push-ups	0	0	0	87	0	7
	Sit-ups	0	0	0	0	82	6
Mistook		3	8	0	3	2	

Table 2 Accuracy of segmentation and classification

Phase	Precision	Recall	F1-score
Segmentation	97.9%	93.9%	95.9%
Classification	96.9%	93.0%	94.9%

the templates. So, while we need to improve these templates of defined exercises, we can say that "some exercises were not performed correctly." To solve the problem, we need to define the exercise in more detail.

On the other hand, the experiments included non-exercise scenes such as the transition from an exercise to another. During these transition periods, the segmentation algorithm detects peaks from the acceleration signal. Most of these transition segments are ignored at the classification phase because they do not correlate with any template. Still, few may correlate. From the results shown in Table 1, eight transition movements were classified into walking, three into running, another three into a push-up, and two into a sit-up. Miss classification of some transitions into walking and running is due to the footsteps included in the transition movements. Besides, the reason why some transition movements were classified into a push-up or a sit-up is that the acceleration intensity of these two exercises is low. If the intensity of acceleration is low, the distance score is going to be small even though there is no correlation. Oppositely, there was no misclassification into jumping because it has high intensity.

According to the results of classification, some running steps were classified into walking. One of the reasons is because running starts from walking and ends to walking unconsciously. The first step of running is weaker than the others. Also, when stopping running, one cannot stop instantly without slowing down that the proposed method classifies into walking.

5. Conclusion

We have developed a segmentation and classification method to count some exercises. Unlike most existing approaches, the proposed approach works both in indoor and outdoor environments. Conventional exercises such as running, walking, as well as some typical workouts, have been tackled. A correlation-based approach, rather than an approach using machine-leaning techniques, has been chosen developed so that the proposed method enables easy deployment to numerous different exercises. Additionally, the proposed approach requires less training data because only one sample of motion data of each target exercise is needed.

The proposed segmentation algorithm achieved 98% precision, 94% recall, and 96% F1-score, while the proposed classification method achieved 97% precision, 93% recall, and 95% F1-score despite being affected by the errors issued from the segmentation phase. The accuracy is approximately similar or better than previous works that handled only indoor workouts and those that use a vision-based approach. These results demonstrated that our method is accurate enough to be used in actual applications.

There is still room for improvement of the proposed segmentation algorithm and classification method. By creating templates data from the mean of multiple data, the classification accuracy may increase.

References

[1] Mark Hamer, Emmanual Stamatakis, and Andrew Steptoe. Dose-response relationship between physical activity and mental health: the Scottish Health Survey. *British journal of sports medicine*, 43(14):1111–1114, 2009.

[2] Dena M Bravata, Crystal Smith-Spangler, Vandana Sundaram, Allison L Gienger, Nancy Lin, Robyn Lewis, Christopher D Stave, Ingram Olkin, and John R Sirard. Using pedometers to increase physical activity and improve health: a systematic review. *Jama*, 298(19):2296–2304, 2007.

[3] Dan Morris, T Scott Saponas, Andrew Guillory, and Ilya Kelner. Recofit: using a wearable sensor to find, recognize, and count repetitive exercises. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3225–3234, 2014.

[4] Fahim Kawsar, Chulhong Min, Akhil Mathur, and Allesandro Montanari. Earables for personal-scale behavior analytics. *IEEE Pervasive Computing*, 17(3):83–89, 2018.

[5] Jay Prakash, Zhijian Yang, Yu-Lin Wei, and Romit Roy Choudhury. STEAR: Robust Step Counting from Earables. In *Proceedings of the 1st International Workshop on Earable Computing*, EarComp'19, page 36–41, New York, NY, USA, 2019. Association for Computing Machinery.