Understanding Activity Segmentation for Multi-sport Competitions

Justin Whitlock
Old Dominion University
jscha002@odu.edu

Orkun Krand
Old Dominion University
okrand@odu.edu

Shubham Jain
Old Dominion University
jain@cs.odu.edu

ABSTRACT

Despite the advances in activity detection, their applications in the sports domain are limited. Athletic environments are fast and challenging. Athletes often perform more than one activity in a single workout, especially if they are training for a multi-sport competition, such as a triathlon. These competitions require an athlete to transition from one activity to another quickly. Current logging applications require the user to select the activity they are about to perform, and start and stop the timer for each activity. This could increase the athlete's transition time, and it would not give the athlete an estimate of how much time they spent in transition between the activities.

This paper explores activity segmentation for multi-sport scenarios. Our goal is to identify the activities and segment a user's workout trace into constituent activities, including the transition periods. We use an Apple Watch to gather inertial sensor data and validate our system in the context of a triathlon. The system was trained and tested on 3 activities (running, biking, and swimming), as well as simple actions performed in transition, from 5 different participants. Our system achieves 91% accuracy in detecting the activity, and can accurately identify the start and stop times for each. We also validate our results with data collected from a volunteer at a triathlon.

ACM Reference Format:
Justin Whitlock, Orkun Krand, and Shubham Jain. 2018. Understanding Activity Segmentation for Multi-sport Competitions. In *WearSys'18: 4th ACM Workshop on wearable systems and applications , June 10, 2018, Munich, Germany*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3211960.3211972

1 INTRODUCTION

For today's athletes, many resources and wearable devices have been created to assist with training. Numerous motion studies have covered activity recognition for running, cycling and swimming [1–6] using accelerometers and gyroscopes, but current applications still depend on the athlete to notify it of what activity they are performing. For most applications, this is sufficient, since athletes tend to do only one activity at a time. An exception to this assumption are multi-sport athletics. We consider the example of a triathlon. Triathlons are a combination of three activities performed back-to-back: running, cycling, and swimming. In addition to the three events, there is a transition period between activities which counts against the athlete's overall time. Top of the line sports watches like the Garmin Forerunner 935 [7] include triathlon tracking with a convenient interface for quickly switching between sports at a tap. Unfortunately, the switch is manual and it happens during the transition period and does not accurately reflect the event times as the transition time will be split between events. It has been shown that if athletes are able to accurately divide their activities, and include transition between the activities as part of their training, they can significantly improve their overall time [8]. Simply adding a start and stop for transition periods would just be one more thing for athletes to think about and could potentially increase their transition time. In a sport where it is common for athletes to begin the cycling portion barefoot to save time, slipping their shoes on after they are up to speed, automating event tracking so athletes can focus on the race could be an advantage.

This paper explores continuous activity segmentation in the context of multi-sport scenarios, using a wearable sensor. Our system identifies the athletic activity a user is performing, based on inertial sensor data from a single limb. We also demonstrate how the workout can be segmented to identify the start and stop times for each activity, including the transition duration in between. Such a system provides athletes with the ability to train and evaluate performance from a single workout trace.

Related Work. Activity recognition using wearable sensors has been explored previously. In the past, researchers have used accelerometers or gyroscopes to identify or evaluate swimming [6, 9, 10], running [6, 11–13], or cycling [1, 3–5], and some success using a wrist mounted sensor [3–6, 9, 14]. However, they either assume that the activity is known, seeking to evaluate the activity, or they focus on daily activities such as sitting, walking, standing etc. While some of them can determine activity duration in addition to recognition, there is not an attempt to string similar short durations together to represent an overall activity segment.

Recently, researchers have proposed solutions for multiple-activity segmentation in more controlled environments, using rich sensors, such as cameras [15–18]. This may be an effective solution for a gym or fitness center, but most triathlons are performed outdoors, in parks or other large recreational areas. As an athlete training for a triathlon, limiting oneself to indoor training or the environs of a fitness center, could be a disadvantage.

Current applications for tracking athletic data such as STRAVA [19] or RunKeeper [20] allow a user to track different activities but not all in the same session. In the case of STRAVA, a user can manually divide a workout after the fact. In order for this to be effective in the context of training for a triathlon, they would need to keep track

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
WearSys'18, June 10, 2018, Munich, Germany
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5842-2/18/06…$15.00
https://doi.org/10.1145/3211960.3211972

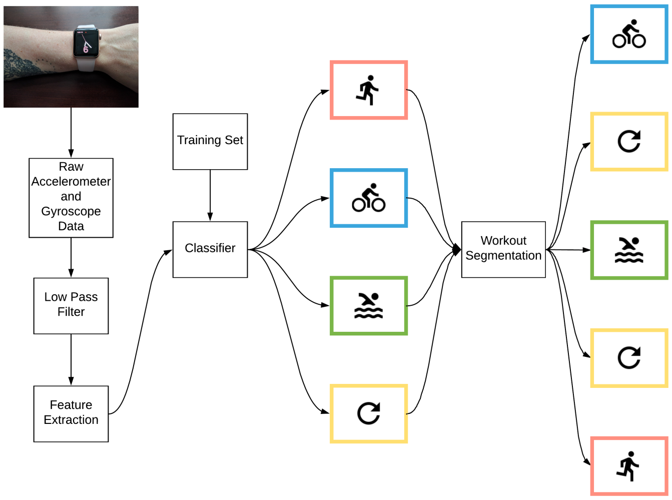

Figure 1: System Overview: A single workout is segmented into constituent activities, including the transitions between them.

of start and stop times for each activity, as well as the duration for which they were changing equipment or taking a break.

Challenges. Determining transitions between sequential activities is challenging compared to simply detecting an activity. In a multi-sport scenario, the transitions between activities do not have a well defined composition. An athlete could practically do anything in-between activities, including stretching, jogging in-place etc. On the other hand, even simple actions such as drinking water, walking, and tying their shoes can occur during an athletic activity, such as running. A major challenge is distinguishing when these actions appear during activities, and when they mark a change in activity. Moreover, for an athlete, knowing the exact start and stop times for each activity (including transitions) is crucial, and must be determined accurately. Transition times count against the overall time, and lack of techniques to measure these times offers serious disadvantages to athletes training for such events. Our proposed system addresses these challenges by not only identifying the activities in the context of a triathlon, but also determining the transitions between activities. Predicted labels from a classifier are often computed independently for small durations, and can result in outliers. Therefore, we process classifier outputs to discard outliers caused by transition-like actions within an activity.

Applications. Other multi-sport competitions, such as duathlon, swimrun, and rowathlon can also benefit from workout segmentation. Endurance races, for example, Spartan Race or Mud Run are also some of many potential applications of such a system. Additionally, the techniques proposed here can also be used in team sports, to monitor the total amount of time a player was running, walking, or sitting on the bench.

Our goal is to allow a user to capture a single workout trace with multiple activities, which is then segmented into individual activities automatically, as opposed to manual definitions.

In summary, the contributions of this work are:

- Segmenting a single workout trace by detecting and classifying 3 athletic activities and the transition period between them.

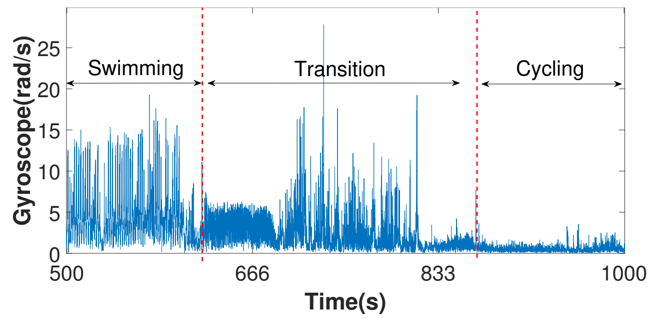

Figure 2: The magnitude of the Gyroscope data while transitioning from one activity (swimming) to a second activity (cycling).

- Evaluating the system on real world multi-sport workout data.

2 DESIGN AND METHODOLGY

Figure 1 shows an overview of our system. We train a classifier to detect three activities and the transitions between them as a fourth class. The outputs from the classifier are used to segment the workout trace, and determine the duration of the activities and the transitions. In this section, we discuss the details of our algorithm.

2.1 Data Preprocessing

Leveraging Swift 4.0 and HealthKit, we designed an application that collects accelerometer and gyroscope data from the Apple Watch 3 continuously. HealthKit workout keeps all background processes alive as it needs to keep collecting data even when the screen is off. Our WatchOS application offers a button to start and stop the workout. There is also a timer to show how long it has been since the workout started. The raw accelerometer and gyroscope readings along all three axes were captured at $30Hz$, and logged into a file on the watch. The data collection end of our system was tested thoroughly with sessions of long duration to ensure the data would be saved. We used a lowess regression filter to smooth the accelerometer and gyroscope data along all three axes. We then computed the magnitude for accelerometer and gyroscope using the filtered readings. The gyroscope magnitude is given by: $\mathcal{G} = \sqrt{G_x^2 + G_y^2 + G_z^2}$. Figure 2 shows a small trace of gyroscope data, with swimming, transition, and cycling labeled. The transition period measures the time from the finish line for the swim event, to the start line for the bike event, in this case, 3 minutes 17 seconds. During this period, the volunteer runs from the water to the transition area, changes out of wetsuit into running gear and shoes, pulls their bike off the rack, then runs, pushing their bike, to the cycling start. The distance from start and stop lines to the transition area can vary broadly between races, from a few yards, to several hundred feet.

2.2 Feature Selection

In order to classify the user's activity, we used a sliding window technique [1, 3, 4, 6, 9, 11, 13, 14, 21] and extracted features from each window. We tested with window sizes of 3 seconds, 5 seconds, and 10 seconds. The window was shifted by 1 second at a time.

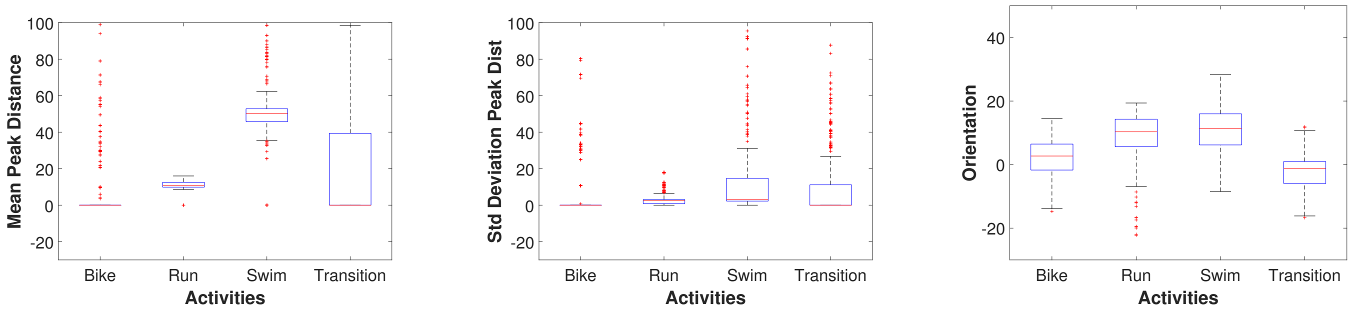

Figure 3: 11 features were extracted from the accelerometer and gyroscope data. The mean distance between peaks, the standard deviation of the distance between peaks for the accelerometer, and watch orientation showed the greatest differences across two or more activities.

From each window we compute temporal and frequency-based features from both the accelerometer magnitude and the gyroscope magnitude. For each window these features were: mean distance between peaks, deviation of distance between peaks, deviation of peaks amplitude, mean of peaks amplitude, root-mean-square level of the window, watch orientation computed from accelerometer readings, and the maximum power spectral density. Figure 3 shows three features for all four classes in a triathlon. We can see that the mean peak distance is most distinctive for swim, but exhibits a much larger range for transitions.

2.3 Activity Segmentation

To classify each window, we use supervised learning. In the context of a triathlon, we define four classes: cycling, running, swimming, and transition. Transitions are referred to as time intervals right after an activity has been completed and before the next one is started. We divide our data into training and test subsets. The feature vectors are computed from the training set in 5 second and 10 second windows. For our classifier, we use bootstrap aggregating, also known as bagging. The classifier was chosen based on empirical performance. It is an ensemble classifier that melds several weak learners into one high-quality learner. The classifier is constructed using five fold cross validation on the training set. The training set was collected by capturing different activities from volunteers and constitutes 20% of the collected data. We obtain 98% accuracy on our training set for the 10 second window and 95% accuracy for the 5 second window.

We then divide our data set into similar sized windows and extract features for each window. Each window is classified by our classifier described earlier. We perform two-stage post processing of the predicted labels. In the first stage, we analyze the probability scores returned by the classifier. If the score for the predicted class is lower than 60%, we split the window into two and classify the first sub-window to obtain the predicted label for that window. These predictions are then used for determining the start and stop times of each activity. Since these activities are performed sequentially, we process them with respect to predicted classes in the windows before the current window. Transition-like activities, such as drinking water or walking, are common when a user is performing another activity and thus must be discarded when they appear between two segments of the same class. We make this decision based on the duration of this transition window. A transition prediction must appear multiple times in quick succession before it is determined that a transition may have begun. If the preceding activity is again detected with high confidence over a short period, it is considered to be a part of that same activity and the short period of transition-like activity is discarded. If a transition lasts longer without sustained re-occurrences of the previous activity, it is considered to mark a switch in activities. Since the transition classification was trained on simple actions like drinking water or tying shoes and doesn't cover everything an athlete may be doing during transition, we also viewed spurious appearances of activities other than the current activity as possible transition events. These are also discarded if the original activity reappears with high confidence. The predicted labels are processed using these practical constraints to output the start and stop times of each activity. The proposed activity segmentation is performed offline.

3 EXPERIMENTAL SETUP

We collected data from four volunteers for running and cycling, three male and one female. All volunteers have trained for and completed a triathlon in the past. The data we collected consisted of the x, y and z axes from both the accelerometer and gyroscope on the Apple Watch, at a rate of $30Hz$. Each session was recorded using a GoPro Hero 4 Silver at 1080p - 60 frames per second. These recordings were used to establish ground truth. We manually labeled the exact times when the user pushed the start button on the watch and their subsequent activities and transitions.

Each volunteer completed a combined workout, featuring two or more activities, in any order, to be used as test data. Each activity was approximately 10 minutes long, with a short transition period between activities. The transition periods varied in length, for a maximum total time of 25 minutes. The workouts were performed in an empty car park next to a busy public path to simulate the conditions of a triathlon, with the entrance to the path as the start line and stop line for each activity.

During the transition periods, volunteers performed actions such as changing and tying their shoes, stretching, walking around to catch their breath, sitting down, and drinking water. We also collected single traces for running, cycling and transition, totaling 45

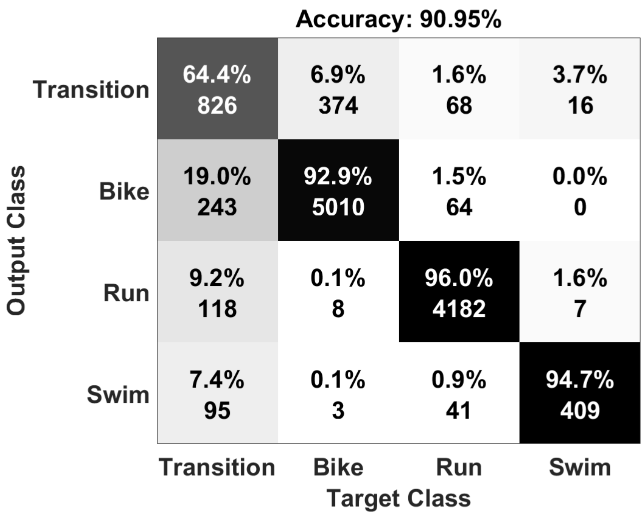

Figure 4: Classifier accuracy on unseen test data ~91%. The test data included 5 different traces that combined two or more activities.

minutes, in a controlled setting. This data was primarily used to train the classifier.

Collecting swimming data was challenging due to the lower temperatures and availability of volunteers. Swimming data was collected from 2 volunteers, both male. Data collection was done in a semi-Olympic swimming pool. Overall, we collected a total of 45 minutes of swimming data. We also collected data from one volunteer at a sprint triathlon, used as test data for our evaluation. The volunteer completed the race in 1 hour 18 minutes with an average transition time of 2.5 minutes. 20% of the collected data was used as training data for the classifier. 65% of the collected data was unseen by the classifier and used as test data.

4 EVALUATION

In order to evaluate our system, we captured a real triathlon for one of our volunteers. We used this triathlon data in addition to other data collected by our volunteers, as described in the earlier section. We selected five traces to constitute our test set. All the results presented here are derived from the 5 trace test data set. In this section we discuss the performance of our system.

Activity Classification. We measured the accuracy of our system in identifying the classes. The results are shown in the form of a confusion matrix in Figure 4. The accuracy was measured by comparing the ground truth labels and predicted labels for corresponding time windows in the test trace. The overall accuracy of the classifications was ~91%. We can see that core activities - running, swimming, and cycling, are identified more accurately than transitions. This was expected since the transition period can include other activities such as running. The transition class was trained on simple tasks like drinking water and tying shoes, while the ground truth labels more generally recognize when the transition period begins and ends. This is the reason why we are looking for spurious classifications of any class outside the current activity class as a possible indication of the start of a transition, in addition

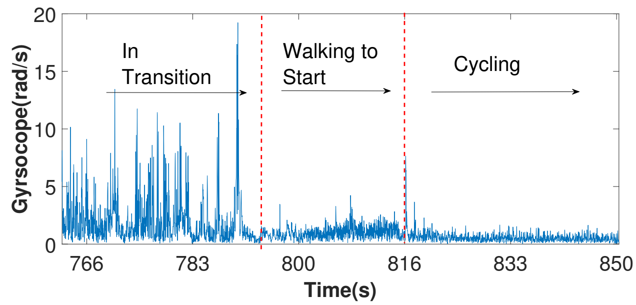

Figure 5: Walking with the bike to start vs riding the bike. The traces show a great deal of similarity due to the position and stability of the wrist while the athlete holds the handlebars.

to looking for predictions of the transition class. According to our labeling convention, cycling only starts when the athletes start riding their bicycles. However, the time duration when the athlete is walking their bicycle to the start line is also classified as cycling by the algorithm, while it's defined as transition in our ground truth labels. The similarity between these traces can be seen in Figure 5, and it does lead to an early detection of the cycling segment.

Detection Delay. In addition to classifying an activity, athletes are also interested in learning the time they've spent performing each activity, and the time spent transitioning from one activity to another. We calculate the detected start and stop times for each activity. The mean delay of the start and stop times for each activity are shown in Figure 6. Negative values denote that the activity was detected earlier than it started, while positive delay represents that an activity was detected by our system after it started/stopped.

The start of the bicycling activity was detected earlier as a result of the similarity between walking with a bike and riding a bike. Athletes must walk their bike from the transition area to the start line before they start riding. Since their hands are in the same position as if they were riding the bike, the algorithm classifies this as a bike activity due to the similarity in trace (shown in Figure 5). On the other hand, our manual labels only classify an activity as biking if the user was riding it. The stop times of biking activity are detected later for the same reason; athletes walk their bikes after they cross the dismount line.

For swimming, athletes often have to walk/run a certain distance to the finish line after they get out of the water. We detect an early stop to the swimming event because while the person did stop swimming, the time they spend getting to the finish line is still counted against their swimming times, which is reflected in the ground truth labels. The start line for swim can also be a distance away from the water, causing the actual swimming activity to happen several seconds after crossing the start line. We expected to see an early detection for running, as the athletes ran from the transition area to the starting point for the run, but the early detection was rare. The delayed detection was caused by the athletes looking at the watch at the start of the run, possibly checking their time. This changed the orientation of the watch, as well as the signal profile as they held their wrists steady. The resulting classifications fell under other classes often enough to make it seem the runner was still in transition.

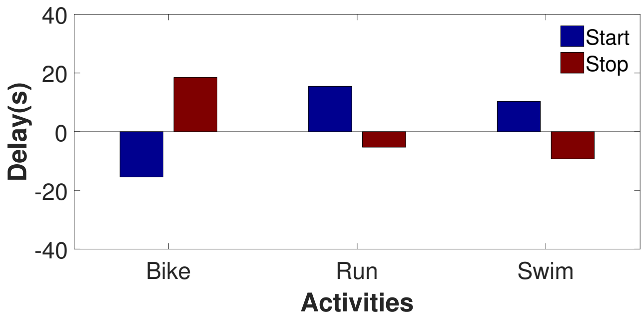

Figure 6: The recognition delay on test data. Negative values denote early detection by our system, while positive values represent a delayed detection.

5 DISCUSSION AND CONCLUSION

We proposed a system that processes time-series data for a multi-sport scenario to identify the constituent activities, and determine the duration of each activity. Our classifier achieves greater than 90% accuracy for distinguishing between swimming, biking, running, and transitions. We include transitions in our analysis because athletes who are training for such competitions are interested in knowing the time spent performing each activity and the transitions from one to another.

Sports scenarios are challenging due to their dynamic nature and less defined temporal boundaries between events. Our system can accurately detect the start and stop times for each activity. These times may not match up to the official triathlon timings. This is because, like our ground truth labels, official timings are measured from start line to finish line, while our system detects start and stop for the activity. We believe that computing the time spent performing each activity is more useful for trainees looking to measure or improve their performance. Current triathlon watches do allow activity segmentation, but the switch from one sport to the next is manual, and it would occur during transition. This means that when the user taps the watch, the preceding portion of the transition will be included in the tail end of one segment, and the rest of the transition will be part of the new activity. In a transition lasting longer than 40 seconds, that's finish line to start, the delay would always be higher than what our system is currently capable of. Current systems also do not attempt to evaluate an athlete's performance during transitions. We observed some discrepancies, such as, a time-window being classified as cycling if the volunteer's hand is on bicycle when he is walking the bike or holding the left arm up to look at the watch while running is classified as transition.

There are several ways of addressing the discrepancy in timings and improving the system accuracy. Since limbs are more involved in transitional activities, wearing an inertial sensor on the ankle instead of the wrist may perform better. For example, it might be able to distinguish between walking a bicycle and riding it. This alternate sensor position will be agnostic to wrist orientation and stability. Therefore, errors caused by athletes checking the time will not affect the system performance. We intend to explore this alternate sensor mounting position in the future.

We found that even after collecting 2 hours of data via our app, the Apple Watch was able to keep running for the rest of the day without requiring a recharge. This implies that our data collection system is not harsh on the battery of the watch. We will perform a quantitative analysis of power consumption in the near future.

This paper explores the possibility of applying activity detection to sports. By automatically segmenting a workout trace, it can measure the fraction of time spent in each activity as well as the transition periods. Such analysis has been found useful in the past [8]. We believe that such a system has immense potential in athletic training and performance evaluation.

REFERENCES

[1] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen. Activity classification using realistic data from wearable sensors. *Trans. Info. Tech. Biomed.*, 10(1):119–128, January 2006.

[2] Seon-Woo Lee and Kenji Mase. Activity and location recognition using wearable sensors. *IEEE Pervasive Computing*, 1(3):24–32, July 2002.

[3] Stephen J Preece, John Y Goulermas, Laurence P J Kenney, Dave Howard, Kenneth Meijer, and Robin Crompton. Activity identification using body-mounted sensors-a review of classification techniques. *Physiological Measurement*, 30(4):R1, 2009.

[4] Andrea Mannini, Stephen S. Intille, Mary Rosenberger, Angelo M. Sabatini, and William Haskell. Activity recognition using a single accelerometer placed at the wrist or ankle. *Med Sci Sports Exerc.*, 45(11):2193–2203, November 2014.

[5] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Trans. Info. Tech. Biomed.*, 12(1):20–26, January 2008.

[6] P. Siirtola, P. Laurinen, J. Röning, and H. Kinnunen. Efficient accelerometer-based swimming exercise tracking. In *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 156–161, April 2011.

[7] Garmin: Forerunner 935, 2018. https://buy.garmin.com/en-US/US/p/564291.

[8] Roberto Cejuela, Antonio Cala, José A Pérez-Turpin, José G Villa, Juan M Cortell, and Juan J Chinchilla. Temporal activity in particular segments and transitions in the olympic triathlon. *Journal of human kinetics*, 36:87–95, March 2013.

[9] Marc Bächlin and Gerhard Tröster. Swimming performance and technique evaluation with wearable acceleration sensors. *Pervasive and Mobile Computing*, 8(1):68 – 81, 2012.

[10] Neil Davey, Megan Anderson, and Daniel A. James. Validation trial of an accelerometer-based sensor platform for swimming. *Sports Technology*, 1(4-5):202–207.

[11] J. Yang, Shuangquan Wang, Ningjiang Chen, Xin Chen, and P. Shi. Wearable accelerometer based extendable activity recognition system. In *2010 IEEE International Conference on Robotics and Automation*, pages 3641–3647, May 2010.

[12] O. Thomas, P. Sunehag, G. Dror, S. Yun, S. Kim, M. Robards, A. Smola, D. Green, and P. Saunders. Wearable sensor activity analysis using semi-markov models with a grammar. *Pervasive and Mobile Computing*, 6(3):342 – 350, 2010.

[13] J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. *ACM Comput. Surv.*, 43(3):16:1–16:43, April 2011.

[14] Andrea Mannini and Angelo Maria Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. *Sensors*, 10(2):1154–1175, 2010.

[15] Guangming Zhu, Liang Zhang, Peiyi Shen, and Juan Song. An online continuous human action recognition algorithm based on the kinect sensor. *Sensors*, 16(2), 2016.

[16] Feng Niu and M. Abdel-Mottaleb. Hmm-based segmentation and recognition of human activities from video sequences. In *2005 IEEE International Conference on Multimedia and Expo*, pages 804–807, July 2005.

[17] A. Ali and J. K. Aggarwal. Segmentation and recognition of continuous human activity. In *Proceedings IEEE Workshop on Detection and Recognition of Events in Video*, pages 28–35, 2001.

[18] H. Zhang, W. Zhou, and L. E. Parker. Fuzzy segmentation and recognition of continuous human activities. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6305–6312, May 2014.

[19] STRAVA: About us, 2018. https://www.strava.com/about.

[20] RunKeeper: Everyone. every run., 2018. https://runkeeper.com/#where-to-start-module.

[21] Xing Su, Hanghang Tong, and Ping Ji. Activity recognition with smartphone sensors. *Tsinghua Science and Technology*, 19(3):235–249, 2014.