weSport: Utilising Wrist-Band Sensing to Detect Player Activities in Basketball Games

Lu Bai
School of Engineering and Digital Arts
University of Kent
L.Bai@kent.ac.uk

Christos Efstratiou
School of Engineering and Digital Arts
University of Kent
C.Efstratiou@kent.ac.uk

Chee Siang Ang
School of Engineering and Digital Arts
University of Kent
C.S.Ang@kent.ac.uk

Abstract— **Wristbands have been traditionally designed to track the activities of a single person. However there is an opportunity to utilize the sensing capabilities of wristbands to offer activity tracking services within the domain of team-based sports games. In this paper we demonstrate the design of an activity tracking system capable of detecting the players' activities within a one-to-one basketball game. Relying on the inertial sensors of wristbands and smartphones, the system can capture the shooting attempts of each player and provide statistics about their performance. The system is based on a two-level classification architecture, combining data from both players in the game. We employ a technique for semi-automatic labeling of the ground truth that requires minimum manual input during a training game. Using a single game as a training dataset, and applying the classifier on future games we demonstrate that the system can achieve a good level of accuracy detecting the shooting attempts of both players in the game (precision 91.34%, recall 94.31%).**

Keywords— Wearable Sensors; Classification; Accelerometer; Activity monitoring

I. INTRODUCTION

Over the last decade, there has been a significant increase in the use of wristbands as companion devices for tracking physical activities. One of the primary purposes for such devices is to act as trackers for sports activities [1-3]. Relying on a combination of inertial sensors and biosensors, wristbands aim to extract higher-level information about the physical activities performed by the user. Typical usage scenarios include the tracking of sports activities the wearer performs alone (e.g. running), or the collection of generic information that is relevant to the specific activity performed (e.g. energy expenditure) [4-6].

When considering team-based sports activities, such as playing basketball, existing sensing software that is commonly available for wristbands has very little extra information to offer. We believe that in the realm of team-based sports games, wristbands can potentially be used to detect game-specific events, group activities, and games statistics. Such information can enhance the experience of the players, and offer new insights about their performance within the game. Developing such systems, however, may require the departure from the traditional "single wristband – single user" sensing approach. Instead, team-based tracking need to exploit the wrist-mounted devices of multiple users in order to extract information that may related to multiple players.

In this paper, we explore the potential of collaborative sensing using wrist-worn devices within the context of "one-to-one" basketball games. Our aim is to develop a system that can capture game statistics, such as the number of shot attempts performed by each player during a game. Although such statistics refer to a single user, the accurate detection of the shooter during a basketball game may require information from both players in the game, to resolve potential ambiguities. In this work we attempt to accurately detect such shooting events by employing a two-level activity classification approach. Firstly, individual classifiers tailored to each player are used to detect potential shooting actions. In the second stage, the classifier combines information from multiple users to accurately identify the player who is performing the shooting action at any given time. The system is designed with particular attention to reduce the overhead of training the classifiers, and in particular to make the collection of "ground truth" information easy enough to be performed by typical users in the field.

The *weSport* system is developed using the Microsoft Band device. The system is evaluated through a small-scale study, including data from 2 one-to-one basketball games (in total 20 minutes of data). We demonstrate that our approach can detect shooting statistics with 91.34% precision and 94.31% recall, relying on a single game as a training session.

II. RELATED WORK

Researchers have used wearable sensing technologies for monitoring physical activities and evaluate the performance of individuals in different sports games [7-9]. Wearable technologies have also been employed as visual augmentation aids in team sports [10]. In outdoor games, such as cricket, and soccer, GPS technologies have been employed to track the players' movement patterns and activity profiles. Players' performance during a team game is influenced by other players' speed or field positions [11]. For specific indoor sport games, such as swimming and tennis, camera based systems and accelerometers based inertial sensing systems have been used in recent studies [12, 13]. In the study by Montgomery et al. [14], multiple basketball players' movement patterns and

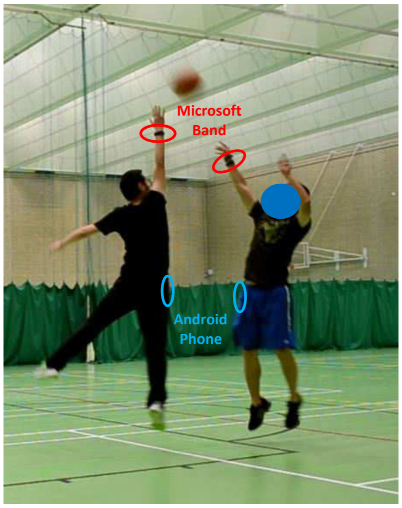

Fig. 1 A typical shooting event. The shooter on the right side is jumping and the defender on the left side is jumping with hand raised. Both participants wearing the system used for dataset collection, with a Microsoft Band worn on the wrist and an android phone put in the trouser pocket.

heart rate have been monitored but no team work has been analyzed from these monitoring results. There is also research focusing on the collaborative tracking in team sports where multiple cameras are used [15-17], which makes the motion tracking systems complex and expensive.

Overall, existing work in sports activity tracking relies primarily on expensive infrastructure that is more applicable for commercial venues. When considering casual sports games, off-the-shelf sensing technologies can offer a cheap way to enhance the experience of the players, and allow them to capture statistics about their performance. In this paper we explore the tracking of players' performance in a 'one to one' basketball game by using the inertial sensing data from typical wristbands and smartphones.

III. Data Collection

As aforementioned, we collected the players' data using the inertial sensors from a Microsoft Band mounted on the participant's dominant wrist and an Android phone placed in the players' trouser pocket (See Fig.1).

Data were collected from two sessions of one-to-one basketball games of two same users. Ethics permission was obtained from the University of Kent Research Ethics committee. Written informed consent was obtained from each participant before enrolment and participant in the study. Each session lasted for 10 minutes. The data collected included 3D acceleration and rate gyro captured from the Microsoft Band at 32Hz and from the Android phone at 100Hz. In addition, the Microsoft Band also provided physiological information e.g. heart rate and skin temperature through the built-in PPG (Photoplethysmogram) sensor and temperature sensor.

Two games were video recorded by a Nikon D3100 camera (14.2 megapixels) and the timestamp of inertial measurement data from Microsoft Band and Android phone were synchronized with the videos manually after the data capture. An android application for inertial measurement data collection from both Microsoft band and android phone was developed. The app collected data from all the devices in the field and uploads them to a back end server for analysis.

IV. Observations

Following the data collection, we performed a preliminary analysis of the sensors data, and compared them against the video footage. The aim was to get some insights on the mobility patterns that can be observed during the game. Through manual observation, the dataset was labeled, marking the incidents where players perform an "attack", "defense" or "normal play" (e.g. dribbling). **Fig. 2** shows the normalized acceleration data from the wristband, for both players with the manually generated labels.

The preliminary analysis shows that every game session consists of periods of "normal play", with relatively low levels of mobility, interspersed with high activity events where a player is either attacking (and eventually shooting) or defending against an attack. Our primarily aim is to identify "attack" instances that also contain a shot attempt. One important realization that emerged from this study is that

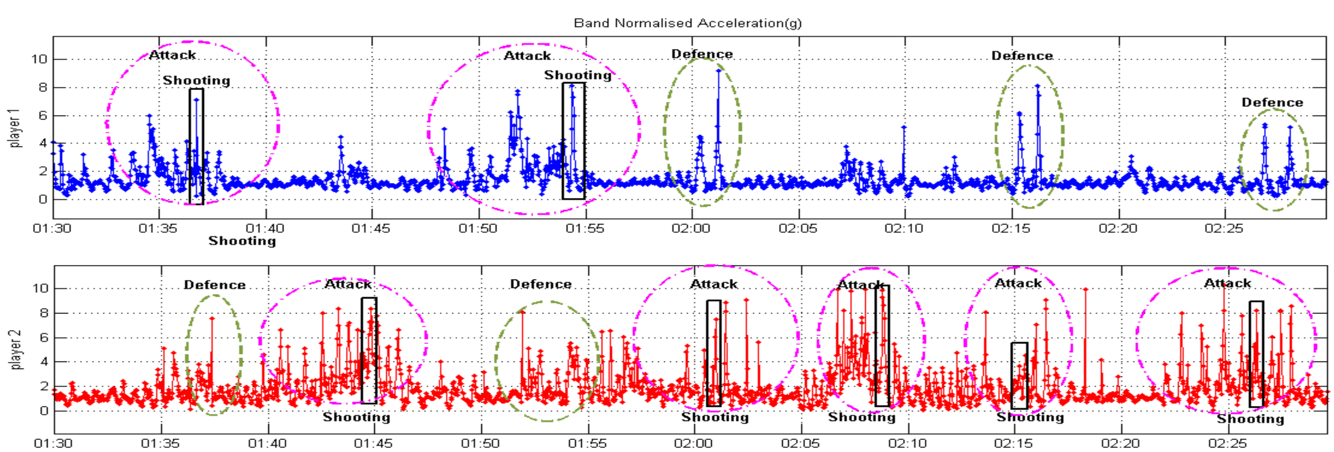

Fig. 2. An example of comparing the two players' performance through utilizing the captured inertial measurement data against video footage. We can observe different patterns during the Attack, Defense and No activity.

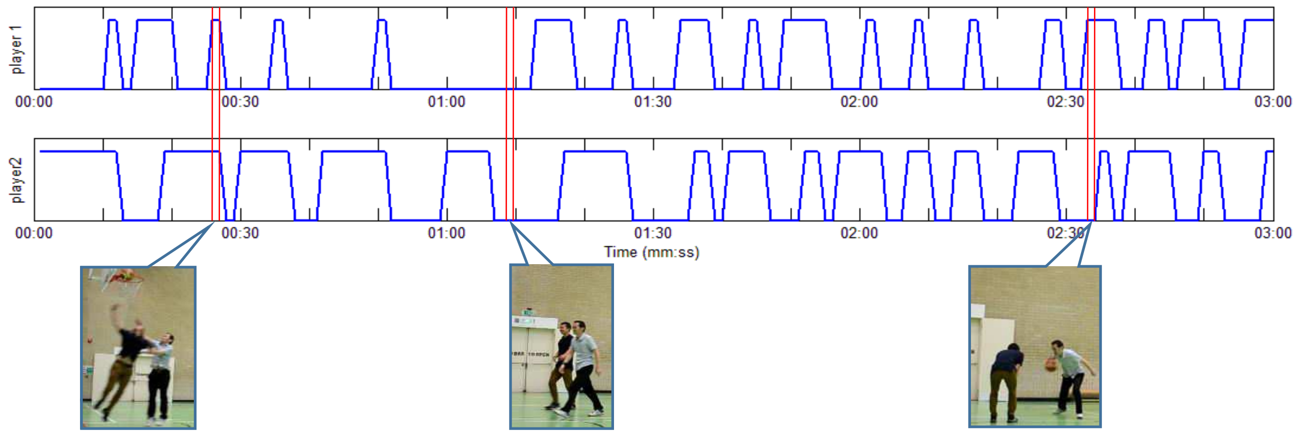

Fig. 3. Comparing the two players' activities against video footage. Case 1: both players jump during a shot action. Case 2: no action time in the game; Case 3: one player initiates and attack while the other player is inactive.

manual labelling of an attack sequence is subjective. Identifying when an attack started depends highly on the perception of the researcher/observer. This also means that capturing the ground truth of that detail would be extremely cumbersome and potentially unrealistic for a system that should be easy to deploy and use. Indeed, in the final system, we relied only on the actual shooting events (that can be objectively captured by an observer) as the ground truth.

Further observation of the dataset shows that a highly active instance of one player attacking tends to overlap with an active instance of the opposing player defending. **Fig. 3** demonstrates the active periods (either attacking or defending) for both players within a game session. It is clear that there are cases where both players can be active at the same time (case 1) or instances where only one (case 3) or none (case 2) is active. This observation implies that detecting the activities of a user through their sensor data coming from their own devices can potentially result in situations where both users are mistakenly considered as performing a shot at the same time. Therefore, collaborative sensing, combining data from both players, may be required.

V. System Design

With *weSport* we aim to develop a system able to capture the shooting actions of basketball players, generating statistics that the player can explore to learn about their performance in the game. Moreover, the system should be easy to deploy and train.

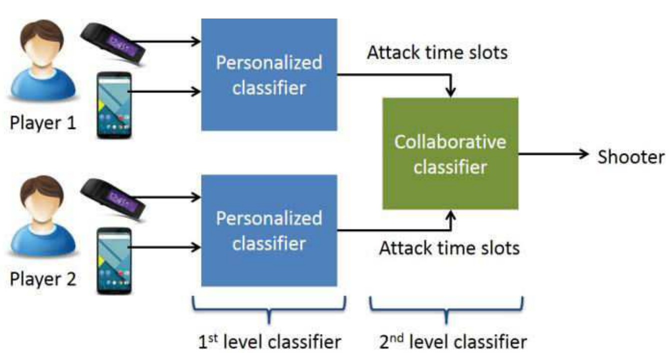

Fig. 4. System architecture. Employing two-level classification.

Specifically, we envisage a system that can be trained during a single basketball game session, and can be used to capture statistics of any subsequent sessions of the same users. As part of the training phase, we believe that ground truth should be easy to capture without the need for specialized equipment (e.g. video recording). Hence, in the design of our system, we consider the capture of ground truth through a third-party observer, who will record which player performs a shot. This means that the ground truth for our system consists only of timestamped events with the indication of the shooter during each shot. In section VI.B we describe how this ground truth is used to generate more elaborate labels for the design of the system.

Based on the observation of our dataset we noted that there were situations where both the shooter and the defender could appear to be performing similar actions: a jump and raise of hands (See **Fig. 1**). Exploring data captured through the sensing devices worn by a single user, it was clear that a shot classifier could therefore potentially generate a high level of false positives, classifying defense actions as shots. In order to address such problems, we developed our system using a two-level classification architecture. As seen in **Fig. 4**, data from the wristband and smartphone were first used to identify a specific action of each user independently. This classifier could potentially generate false positives when trying to detect a shot. The outputs of the first level classifiers are then used to distinguish between actions performed by both players at the same time. The second level classifier is designed to identify the actual shooter in the cases where the individual classifiers mark both players as performing a shooting event.

VI. Data Processing and Methods

weSport consists of two classifiers: a "personalized classifier" aimed to detect the activities of each player independently, and a game-level "collaborative classifier" that identifies the actual shooter at any given time.

A. Ground truth labels

The original ground truth that was captured in each session consists of timestamped events that indicate which player performed a shot during the game. In order to develop a classifier that can detect the actual activity performed by each player, we needed to generate labels that span across a specific duration to capture the full movement of a player when they performed the actions. Through observation of the movement data it is clear that there were distinctive patterns of movement during a shot sequence that lasted for a number of seconds, extending to some seconds before the actual shot (see **Fig. 2**). We therefore needed to create labels for these sequences, identifying the "attack" sequence and differentiate those from the "normal" play activities.

The generation of these labels in an automated way is based on an iterative approach. For each shoot instance labeled manually by an observer, we define an "attack window" that extended t seconds before the actual shot. We start the process with an attack window of t=1 second and progressively extended this by 1 second for each iteration. This window is used to label the duration of an attack using the timestamp of the shot event as the end time point.

For each potential attack window, we train and validate (through a 10-fold cross validation) a classifier, using the features described in section B. After running this process for 10 iterations (maximum attack window tested is 10 seconds), we select the window length that gives us the best performance for the classifier. The length of the attack windows is calculated for each individual player. This reflects the different styles of play for each player where they may have performed attack actions that took more or less time than the other player. Through this process, we define the typical duration of an attack for each player, and we use this as the ground truth label for our classifiers. Specifically, for our two participants, using the first game session for training, we identified an attack window W_1=4 seconds for player 1 and W_2=6 seconds for player 2.

B. Data processing and Feature Extraction

In order to identify the different activity events during the game session, we extract a set of features from the raw data captured by both the wristband and the Android phone. The features are extracted using a sliding window with a 1 second overlap. The length of the sliding window is equal to the attack window that was described in section A.

For the wristband, the features are computed for all three axes and 3D normalized vectors. Considering that the placement of the band was fixed on the player's wrist, we tried to maintain the directionality of movement in the data processed from the band. For the android phone, all the features were computed only for the 3D normalized vectors. Considering that the position of the phone can shift inside the trouser pocket, we only relied on the total magnitude of movement captured from that device. There are 66 features for Microsoft Band and 15 features for android phone that have been considered. These features were computed for each window, which include time domain features e.g. mean value, standard deviation, median value, maximum value, minimum value, zero crossing rates, number of peaks and correlation coefficient between different axes and frequency domain features e.g. FFT coefficients. The details of these features are described in Table I.

TABLE I. SELECTED FEATURES FROM ACCELEROMETERS AND GYROSCOPES OF MICROSOFT BAND AND ANDORID PHONE

Features	***Description***
Mean value	The average value of all the samples in one window
Standard Deviation	The variation from the average value of all the samples in one window
Median value	The median value of all the samples in one window
Maximum value	The maximun value of all the samples in one window
Minimum value	The minimum value of all the samples in one window
Zero crossing rates	The rates at which the signal changes from positive to negative or back of all the samples in one window
Number of peaks	The numbers of the local maxima of all the samples in one window
Correlation Coefficient	The coefficient that illustrates a quantitative measure of the correlation between samples from two different axes in one window
FFT	The sample vaules in the frequency domain in one window

C. Classification of each player's actions

As part of the first stage of the activity detection, we developed personalized classifiers for each player, where we attempted to detect when each player performed an attack. Specifically, the classifier is designed to classify each second of the player's activities as either "attack" or "normal play". Although the classifier was developed through traditional training and cross validation, our ultimate aim was to ensure that we captured the sequence of actions where the actual shot event is performed. Therefore, our actual validation of this classification process was to test whether a sequence of actions classified as "attack" contained the actual shot events captured through ground truth. This essentially means that even if the classifier did not detect accurately the duration of an attack sequence, it would be still considered good for our purposes if the shot event was properly captured.

For the development of the classifier, we tested a range of classification methods. Specifically Random Forest, Naïve Bayes, Decision tree, SVM and k-NN classification algorithms were explored. The best performance was achieved through Random Forest. The output of the classifier was a set of time periods that represent the attack and normal play of each player. The performance of the classifier was checked by testing whether each attack period contained a shot event as specified by the ground truth.

D. Shooter detection

The output of the first level (personalized) classifiers is a set of attack sequences performed by each player. As described earlier there were many situations where the classifiers of both players would indicate that both of them were performing an attack. The purpose of the second level (collaborative) classifier was to identify who was the actual shooter in such ambiguous situations.

For this classifier, the input consisted only of the time periods that had been identified as potential "attacks" for each player through the first level (personalized) classification. Specifically the collaborative classifier did not deal with any data related to "normal play" where no player was potentially shooting. The dataset was essentially a set of time periods where either one or both players were detected as "shooting".

For the collaborative classifier, we used the original ground truth to label each attack sequence with the actual player that performed the shot (e.g. "Player 1 shooting", "Player 2 shooting", or "No player shooting"). For this classifier we used all the features in the Table I as representatives of the movement patterns of each player. Moreover, we augmented our dataset with features that could discriminate the movements of one player against the other. In particular, a typical behavior that could be observed was that when both players were involved during a shot event, one jumping to shoot, and the other jumping to defend, there was a time delay between the actions of the attacker and those of the defender. Indeed, the defender would normally initiate a defensive jump when observing that the attacker was about to shoot. In order to capture such dependencies we introduced two extra features: the differences of peak normalized acceleration vector between two players in one window and the time differences between these two acceleration peaks. Using this feature set we developed the second level collaborative classifier using the datasets from the first game session.

VII. RESULTS AND DETECTION PERFORMANCE

The system was trained using one game session (Game 1) as the training dataset and validated on the same game session through 10-fold cross validation. Moreover, the trained classifier is also tested on a different game session (Game 2) that was captured on a different day. We report the accuracy of the system on both games.

A. *Classification Results in detecting shot event*

We evaluated each of the two players' performance by using all the obtained movement features extracted from both Microsoft Band and android phone for each game respectively. Using personalized sampling window sizes, we trained a classifier for each player. We report the results for a classifier based on Random Forest, which achieved the best results. The accuracy of correctly classified instances for player 1 is 92.83% and for player 2 is 83.17% in Game 1. Using the original ground truth (shot events), we marked each classified attack sequence as True-Positive / False-Positive whether there is a shot event within it. Validating this classifier on both games, we show that it can correctly detect 21 out of 22 shots for player 1 and 41 out of 41 shots for player 2 in Game 1; and 26 out of 27 shots for player 1 and 33 out of 33 for player 2 in Game 2 (see Table II).

TABLE II. RESULTS OF CLASSIFICATION ACCRUARY FOR SHOOTING DETECTION

Ground truth No. of Shootings (total duration of "attack" periods)		No. of Shootings (Correctly classified "Attacks" periods %)	TP attack sequence that contains a shooting	FP attack sequence that does not contain a shooting
Game 1				
Player 1	22 (132 secs)	21 (92.83%)	21 (95 secs)	2 (4 secs)
Player 2	41 (322 secs)	41 (83.17%)	41 (350 secs)	3 (7 secs)
Game 2 (classifier trained on Game 1)				
Player 1	27 (162 secs)	26 (84.67%)	26 (122 secs)	2 (7 secs)
Player 2	33 (260 secs)	33 (78.83%)	33 (311 secs)	5 (18 secs)

B. *Classification Results in detecting shooter*

After detecting the attack sequences by the first classifier, we use the ground truth to label them as "Player 1 Shooting", "Player 2 Shooting" or "No Shooting" for each game session. We then evaluated the results on shooter detection. We explored a range of classifiers including Random Forest, SVM, and k-NN and select the SVM (SMO) as the second level classifier to detect the shooter. We used Game 1 as the training set (and performed a 10-fold cross validation) and also validated the classifier on Game 2. In game 1, we report 21 out of 22 shots for player 1 and 42 out of 41 shots for player 2. In Game 2, we report 25 out of 27 shots for player 1 and 37 out of 33 shots for player 2.

Combining the results from both games the classifier demonstrates a precision of 91.34% and recall of 94.31%.

TABLE III. RESULTS OF CLASSIFICATION ACCRUARY FOR SHOOTER DETECTION

Ground truth No. of Shootings		No. of Shootings Detected TP	FP	FN
Player 1 Game 1	22	19	2	3
Player 2 Game 1	41	40	3	1
Player 1 Game 2	27	24	1	3
Player 2 Game 2	33	33	5	0
Overall	123	116	11	7
Precision	**91.34%**			
Recall	**94.31%**			

VIII. DISCUSSION

The overall result of this system demonstrates relatively high levels of accuracy. One of the significant limitations of this work though, is the small number of participants used in the study. However, the design of the classifier has successfully been demonstrated through an additional game session not used for the training of the classifier. We consider the positive results of this work as a good starting point. We intend to expand our work with the involvement of more users. A key challenge that we may face in future work is the potential variation of the game play of individual players when placed against different opponents.

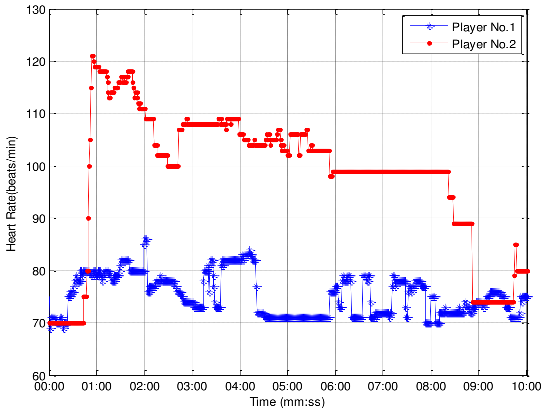

Fig. 5. Heart rate of the two players during game 1

The ease of training was a key requirement for our system. We envisage that an application that is easy to train can be used by casual basketball players to gain further insights about their performance. The datasets collected in this work can potentially be used to extract additional information, such as ball possession, and correlation between effort and results for each player. Anecdotally during the study, we observed a correlation between the heart rate levels of the two players and their performance in the field. As can be seen in **Fig. 5**, the player with the highest heart rate during the game (player 2) was the one with the most shot attempts (41 vs 22), and the highest score at the end. This could indicate a possible link between the amount of effort each player made and the outcome of the match.

IX. CONCLUSIONS

In this work, we demonstrated *weSport*, a system that is able to detect the number of shots that each player performs in a one-to-one basketball game. The system relies on motion data from wristbands and a smartphone carried by the players. *weSport* relies on a combination of personalized "attack" detection classifiers, and a collaborative shooter detection classifier, which combines data from both players. The system was validated through two basketball games, and it demonstrates a precision of 91.34% and recall of 94.31%.

REFERENCES

[1] A. Ahmadi, D. Rowlands, and D. A. James, "Towards a wearable device for skill assessment and skill acquisition of a tennis player during the first serve," *Sports Technology,* vol. 2, pp. 129-136, 2009.

[2] M. Lapinski, M. Feldmeier, and J. A. Paradiso, "Wearable Wireless Sensing for Sports and Ubiquitous Interactivity," *2011 Ieee Sensors,* pp. 1425-1428, 2011.

[3] R. Adelsberger and G. Troster, "Experts Lift Differently: Classification of Weight-Lifting Athletes," *2013 Ieee International Conference on Body Sensor Networks (Bsn),* 2013.

[4] R. Gusmer, T. Bosch, A. Watkins, J. Ostrem, and D. Dengel, "Comparison of FitBit® Ultra to ActiGraph™ GT1M for Assessment of Physical Activity in Young Adults During Treadmill Walking," *The Open Sports Medicine Journal,* vol. 8, 2014.

[5] J.-M. Lee, Y. Kim, and G. J. Welk, "Validity of consumer-based physical activity monitors," *Medicine and science in sports and exercise,* 2014.

[6] C. Strohrmann, H. Harms, C. Kappeler-Setz, and G. Troster, "Monitoring Kinematic Changes With Fatigue in Running Using Body-Worn Sensors," *Ieee Transactions on Information Technology in Biomedicine,* vol. 16, pp. 983-990, Sep 2012.

[7] A. Wixted, D. James, and M. Portus, "Inertial sensor orientation for cricket bowling monitoring," *2011 Ieee Sensors,* pp. 1835-1838, 2011.

[8] N. B. Nordsborg, H. G. Espinosa, and D. V. Thiel, "Estimating energy expenditure during front crawl swimming using accelerometers," *Engineering of Sport 10,* vol. 72, pp. 132-137, 2014.

[9] C. Z. Shan, E. S. L. Ming, H. A. Rahman, and Y. C. Fai, "Investigation of upper limb movement during badminton smash," in *Control Conference (ASCC), 2015 10th Asian*, 2015, pp. 1-6.

[10] M. Page and A. V. Moere, "Evaluating a wearable display jersey for augmenting team sports awareness," in *Pervasive Computing*, ed: Springer, 2007, pp. 91-108.

[11] M. Waldron, C. Twist, J. Highton, P. Worsfold, and M. Daniels, "Movement and physiological match demands of elite rugby league using portable global positioning systems," *Journal of Sports Sciences,* vol. 29, pp. 1223-1230, 2011.

[12] A. Callaway, J. Cobb, and I. Jones, "A comparison of video and accelerometer based approaches applied to performance monitoring in swimming," *International Journal of Sports Science and Coaching,* vol. 4, pp. 139-153, 2009.

[13] D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, M. Walsh, and C. O'Mathuna, "Multi-Sensor Classification of Tennis Strokes," *2011 Ieee Sensors,* pp. 1437-1440, 2011.

[14] P. G. Montgomery, D. B. Pyne, and C. L. Minahan, "The Physical and Physiological Demands of Basketball Training and Competition," *International Journal of Sports Physiology and Performance,* vol. 5, pp. 75-86, Mar 2010.

[15] W. Du, J.-B. Hayet, J. Piater, and J. Verly, "Collaborative multi-camera tracking of athletes in team sports," in *Workshop on Computer Vision Based Analysis in Sport Environments (CVBASE)*, 2006, pp. 2-13.

[16] F. Chen and C. De Vleeschouwer, "Automatic production of personalized basketball video summaries from multi-sensored data," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 565-568.

[17] M. Perše, M. Kristan, S. Kovačič, G. Vučkovič, and J. Perš, "A trajectory-based analysis of coordinated team activity in a basketball game," *Computer Vision and Image Understanding,* vol. 113, pp. 612-621, 2009.