Towards Automatic Modeling of Volleyball Players' Behavior for Analysis, Feedback, and Hybrid Training

Fahim A. Salim

University of Twente

Fasih Haider

The University of Edinburgh

Dees Postma, Robby van Delden, and Dennis Reidsma

University of Twente

Saturnino Luz

The University of Edinburgh

Bert-Jan van Beijnum

University of Twente

Automatic tagging of video recordings of sports matches and training sessions can be helpful to coaches and players and provide access to structured data at a scale that would be unfeasible if one were to rely on manual tagging. Recognition of different actions forms an essential part of sports video tagging. In this paper, the authors employ machine learning techniques to automatically recognize specific types of volleyball actions (i.e., underhand serve, overhead pass, serve, forearm pass, one hand pass, smash, and block which are manually annotated) during matches and training sessions (uncontrolled, in the wild data) based on motion data captured by inertial measurement unit sensors strapped on the wrists of eight female volleyball players. Analysis of the results suggests that all sensors in the inertial measurement unit (i.e., magnetometer, accelerometer, barometer, and gyroscope) contribute unique information in the classification of volleyball actions types. The authors demonstrate that while the accelerometer feature set provides better results than other sensors, overall (i.e., gyroscope, magnetometer, and barometer) feature fusion of the accelerometer, magnetometer, and gyroscope provides the best results (unweighted average recall = 67.87%, unweighted average precision = 68.68%, and $\kappa = .727$), well above the chance level of 14.28%. Interestingly, it is also demonstrated that the dominant hand (unweighted average recall = 61.45%, unweighted average precision = 65.41%, and $\kappa = .652$) provides better results than the nondominant (unweighted average recall = 45.56%, unweighted average precision = 55.45, and $\kappa = .553$) hand. Apart from machine learning models, this paper also discusses a modular architecture for a system to automatically supplement video recording by detecting events of interests in volleyball matches and training sessions and to provide tailored and interactive multimodal feedback by utilizing an HTML5/JavaScript application. A proof of concept prototype developed based on this architecture is also described.

Keywords: accelerometers, activity classification, computational methods, machine learning, physical activity

Coaches and players desire and would benefit greatly from easy access to performance data of matches and training sessions (Koekoek, van der Mars, van der Kamp, Walinga, & van Hilvoorde, 2018). They use this information not only to monitor performance but also to plan training programs and game strategy. According to the assessment of volleyball coaches in The Netherlands,¹ the two areas which can substantially improve sports training are as follows:

- (a) Interactive exercises and enhanced instructions, and
- (b) Providing the trainer with information from live data on player behavior.

It is because performance in sports depends on training programs designed by team staff, with a regime of physical, technical, tactical, and perceptual–cognitive exercises. Depending

on how athletes perform, exercises are adapted or the program may be redesigned. State of the art data science methods have led to groundbreaking changes. The data are from sources such as tracking position and motion of athletes in basketball (Thomas, Gade, Moeslund, Carr, & Hilton, 2017) and baseball and football match statistics (Stensland et al., 2014).

Furthermore, new hardware platforms appear, such as LED displays integrated into a sports court (Kajastila & Hämmäläinen, 2015) or custom tangible sports interfaces (Ludvigsen, Fogtmann, & Grønbaek, 2010). These offer possibilities for hybrid training with a mix of technological and nontechnological elements (Kajastila & Hämmäläinen, 2015). This has led to novel kinds of exercises (Jensen, Rasmussen, Mueller, & Grønbaek, 2015; Ludvigsen et al., 2010) including real-time feedback, that can be tailored to the specifics of athletes in a highly controlled way.

These developments are not limited to elite sport. Interaction technologies are also used for youth sports (e.g., the widely used player development system of Dotcomsport.nl) and school sports and physical education (Koekoek et al., 2018).

Identification and classification of events of interest in sports recordings, therefore, is of interest for not only coaches and players

Salim and van Beijnum are with the BSS, University of Twente, Enschede, The Netherlands. Haider and Luz are with the Usher Institute, The University of Edinburgh, Edinburgh, United Kingdom. Postma, van Delden, and Reidsma are with the HMI, University of Twente, Enschede, The Netherlands. Salim (f.a.salim@utwente.nl) is corresponding author.

but also for sports fans who might, for example, wish to watch all home runs hit by a player during the 2013 baseball season (Matejka, Grossman, & Fitzmaurice, 2014), or a coach searching for video recordings related to the intended learning focus for a player or the whole training session (Koekoek et al., 2018). Analysis of videos, displaying different events of interest, may help in getting insightful tactical play and engagement with players (Harvey & Gittins, 2014). Video edited game analysis is a common method for postgame performance evaluation (Koekoek et al., 2018).

However, these examples require events to be manually tagged which not only requires time and effort but would also split a trainer's attention from training to tagging the events for later viewing and analysis. A system which could automatically tag such events would help trainers avoid manual effort and has the potential to provide tailored and interactive multimodal feedback to coaches and players. The approach described in this paper precisely addresses the above issue.

The context of the current paper is the Smart Sports Exercises project in which we aim to use multimodal sensor data and machine learning techniques to not only enable players and coaches to monitor performance but also to provide interactive feedback (Postma et al., 2019). This paper extends our previous research (Haider et al., 2019, 2020; Salim et al., 2019a, 2019b) and details the architecture, components, and a comprehensive analysis of a machine learning-based system which automatically classifies volleyball actions performed by players during their regular training sessions. The presented paper demonstrates the following:

- Description of a proof of concept prototype of a real-time video supplementary system to allow coaches and players to easily search for the information or event of interest (e.g., all the serves by a particular player).
- Description of an annotated and anonymized data set of inertial measurement units (IMUs) data of players while playing volleyball in real-life training scenarios.
- A novel and comprehensive analysis to: the evaluation of each sensor data from IMUs (3D acceleration, 3D angular velocity, 3D magnetometer, and air pressure) and their fusion for automatically identifying basic volleyball actions, such as underhand serve, overhead pass, serve, forearm pass, one hand pass, smash, and block.

Evaluate the role of dominant and nondominant hand for modeling the type of volleyball action.

Related Work

There are many applications of automatically identifying actions in sport activities (Bagautdinov, Alahi, Fleuret, Fua, & Savarese, 2017; Matejka et al., 2014; Pei, Wang, Xu, Wu, & Du, 2017; Vales-Alonso et al., 2015). Due to their portability and reasonable pricing, wearable devices, such as IMUs (Bellusci, Dijkstra, & Slycke, 2018; x-io Technologies, 2019), are becoming increasingly popular for sports-related action analysis (Pei et al., 2017). Researchers have proposed different configurations in terms of number and placement of sensors (Wang, Zhao, Chan, & Li, 2018); however, it is ideal to keep the number of sensors to a minimum due to issues related to cost, setup effort, and player's comfort (Cancela et al., 2014; Ismail, Osman, Sulaiman, & Adnan, 2016; von Marcard, Rosenhahn, Black, & Pons-Moll, 2017; Wang et al., 2018).

The IMU sensors (Bellusci et al., 2018; x-io Technologies, 2019) have been utilized to automatically detect sport activities in

numerous sports, for example, soccer (Mitchell, Monaghan, & O'Connor, 2013; Schuldhaus et al., 2015), tennis (Kos, Ženko, Vlaj, & Kramberger, 2016; Pei et al., 2017), table tennis (Blank, Hoßbach, Schuldhaus, & Eskofier, 2015), hockey (Mitchell et al., 2013), basketball (Lu et al., 2017; Nguyen et al., 2015), and rugby (Kautz, Groh, & Eskofier, 2015). Many approaches have been proposed for human activity recognition. They can be categorized into two main categories: wearable sensor based and vision based.

Vision-based methods employ cameras to detect and recognize activities using computer vision technologies, while wearable sensor-based methods collect input signals from wearable sensors mounted on human bodies, such as accelerometer and gyroscope. While wearable sensor-based methods collect input signals from wearable sensors mounted on human bodies, such as accelerometer and gyroscope. For example, Liu, Nie, Liu, and Rosenblum (2016) identified temporal patterns among actions and used those patterns to represent activities for automatic action recognition. Kautz et al. presented an automatic monitoring system for beach volleyball based on wearable sensor devices which are placed on the wrist of players' dominant hand (Kautz et al., 2017). Serve recognition from beach volleyball using a wrist-worn gyroscope was proposed by Cuspinera, Uetsuji, Morales, and Roggen (2016), who placed gyroscopes on the player's forearms. Kos et al. proposed a method for tennis stroke detection (Kos et al., 2016). They used a wearable IMU device which is located on the players' wrists. A robust player segmentation algorithm and novel features are extracted from video frames, and finally, classification results for different classes of tennis strokes using Hidden Markov Model are reported (Zivkovic, van der Heijden, Petkovic, & Jonker, 2001).

Based on the above literature, we have concluded that most studies take into account the role of dominant hand particularly for volleyball action modeling and the role of the nondominant hand is less explored. It is also noted that none of the studies above evaluated the IMU sensors for volleyball action recognition. The paper extends our previous work (Haider et al., 2019, 2020; Salim et al., 2019a, 2019b), in which we evaluated the IMU sensors for two class problems (action and no action). However, this study evaluates the sensors for type of volleyball action, such as serve or block, which is a seven-class problem.

By combining machine learning models based on IMU sensors with a video tagging system, this paper opens up new opportunities for applying sensor technologies, such as IMU sensors, with interactive system to enhance the training experience.

Approach

The presented paper extends upon the ideas presented in our previous work (Haider et al., 2019, 2020; Salim et al., 2019a, 2019b). Figure 1 shows the overall system architecture. This paper focuses on Step 3 of the proposed system. However, this section provides a brief summary of all the steps to provide a full idea of the proposed approach.

Data were collected in a typical volleyball training session, in which eight female volleyball players wore an IMU on both wrists and were encouraged to play naturally. See Step 0 in Figure 1. The details of the data collection protocol and annotation procedure are presented in section "Volleyball Data Set." Time-domain features, such as mean, SD, median, mode, skewness, and kurtosis are extracted over a frame length (i.e., time window) of 0.5 s of sensor data with an overlap of 50% with the neighboring frame. See Step 1 of Figure 1.

Figure 1 — Prototype system architecture.

Classification is performed in two stages, that is, Step 2 and Step 3. In Step 2, binary classification is performed to identify whether a player is performing an action or not, using supervised machine learning with unweighted average recall (UAR) as high as 86.87%. The details of the action versus nonaction classification procedure is described in Haider et al. (2019, 2020) and Salim et al. (2019b). Next in Step 3 (Figure 1), the type of volleyball action performed by the players is classified using supervised machine learning algorithms. The details of type of action classification is described in section “Experimentation.”

Once the actions are identified, its information along with the time stamp is stored in a repository for indexing purposes. Information related to the video, players, and actions performed by the players are indexed and stored as documents in tables or cores in the Solr search platform (Velasco, 2016). An example of a smash indexed by Solr is shown in Table 1.

An interactive system is developed to allow the player and coaches access to performance data by automatically supplementing video recordings of training sessions and matches.

The interactive system is developed as a web application. The server side is written using asp.net MVC framework, while the front end is developed using HTML5/Javascript.

Figure 2 shows a screen shot of the front end of the developed system. The player list and actions list are dynamically populated by querying the repository. The viewer can filter the actions by player and action type (e.g., overhead pass by Player 3). Once a particular action item is clicked or taped, the video is automatically jumped to the time interval where the action is being performed. Currently the developed system lets a user filter types of action performed by each user. Details of the interactive system are described in previous work (Salim et al., 2019a, 2019b).

Volleyball Data Set

In order to collect data for the experimentation, eight female volleyball players wore an IMU on both wrists during their regular training session (see Figure 3). All players were amateur volleyball players and belonged to different age groups. The players were encouraged to play naturally so that the data are representative of real-life training scenarios. The video is also recorded using two video cameras. Later the IMU sensors data and video streams are

Table 1 Sample Solr Structure

```

{
  "id": "25_06_Player_1_action_2",
  "player_id": ["25_06_Player_1"],
  "action_name": ["Smash"],
  "timestamp": ["00:02:15"],
  "_version_": 1638860511128846336
}

```

Figure 2 — Interactive front-end system.

synchronized. No screenshots of the recorded session are added due to explicit requests by players not to publish their pictures or videos. It is done so that the models trained are capable of performing in the wild instead of controlled settings. It is for this reason that the collected data are highly imbalanced, for example, for the binary classification task of action versus nonaction recognition (Haider et al., 2020), there are 1,453 versus 24,412 s of data, respectively. Similar imbalance can be seen in the type of volleyball actions performed by players. Table 2 shows the frequency of each volleyball action performed by each player.

Figure 3 — Player wearing two IMUs on both wrists. IMU = inertial measurement unit.

Table 2 Number and Type of Actions Performed by Each Player

ID	No. of actions	Forearm pass	One hand pass	Overhead pass	Serve	Smash	Underhand serve	Block
1	120	40	3	16	0	29	28	4
2	125	36	2	14	32	15	0	6
3	116	50	3	3	34	25	0	1
5	124	46	2	19	21	28	4	4
6	150	30	1	70	0	12	30	7
7	106	39	4	13	0	14	34	2
8	105	34	4	16	34	17	0	0
9	144	42	1	58	33	4	1	5
Total	990	317	20	209	154	144	97	49

Three students annotated the video using Elan software (Brugman, Russel, & Nijmegen, 2004). All annotators were the participants of eINTERFACE2019, and the annotation task is not paid. Since volleyball actions performed by players are quite distinct, there is no ambiguity in terms of interannotator agreement. The quality of the annotation is evaluated by a majority vote, that is, if all annotators have annotated the same action or if an annotator might have missed or mislabeled an action.

Experimentation

The feature set for this paper is extracted from the feature set of a previous study conducted to distinguish actions from nonactions in volleyball training sessions (Haider et al., 2019). In that study, we used time domain features, such as mean, *SD*, median, mode, skewness, and kurtosis which are extracted over a frame length of 0.5 s of sensor data with an overlap of 50% with the neighboring frame. For the current study, we did not apply frequency-domain approaches or deep learning approaches due to the fact that the data set is rather small for such approaches. The second reason for not

opting to use deep learning methods is to evaluate the IMU's sensor information in resource-constrained settings, such as a mobile application.

For the current study, we calculated an average of frame-level features over the time window length of an action. It is done in this way because the current models are intended to be used on the classification performed by the previous model: first a classifier such as the one described in Haider et al. (2019, 2020) would identify the presence of an action (start and end time of an action); subsequently the model trained and reported in this paper would further classify the type of that action.

Classification Methods

The classification experiments were performed using five different methods, namely decision trees (with leaf size of 10), nearest neighbor (with $K = 5$), linear discriminant analysis, Naive Bayes (with kernel distribution assumption), and support vector machines (with a linear kernel, box constraint of 0.5, and sequential minimal optimization solver).

The classification methods are implemented in MATLAB² using the statistics and machine learning toolbox. A leave-one-subject-out cross-validation setting was adopted, where the training data does not contain any information of the validation subjects. To assess the classification results, we used the UAR as a primary measure as the data set is imbalanced but we also reported overall accuracy, unweighted average precision (UAP), and kappa (Landis & Koch, 1977) for the best results.

The UAR is the arithmetic average of recall of all classes and UAP is the arithmetic average of precision of all classes.

Results

The UAR of dominant hand and nondominant hand for all sensors are shown in Tables 3 and 4 respectively. These results indicate that the dominant hand (UAR = 61.45%, UAP = 65.45, and $\kappa = .652$) provides better results than the nondominant hand (UAR = 45.56%, UAP = 55.45%, and $\kappa = .553$). The averaged UAR across sensors indicates that the support vector machines classifier provides the best average UAR (40.34%) across sensors for dominant hand and Naive Bayes provides the best averaged UAR (34.85%) across sensors for nondominant hand for action type detection. It is also noted that the accelerometer provides the best averaged UARs across classifiers for dominant (53.92%) and nondominant (42.70%) hand. The pressure sensor provides the least UAR across classifiers, and the gyroscope provides better UAR across classifiers than the magnetometer. For further insights, confusion matrices of the best results using dominant hand and nondominant hand are shown in Figures 4 and 5 along with precision, recall of each class, overall accuracy, UAR, UAP, and kappa (Landis & Koch, 1977). From Figures 4 and 5, it is also noted that the dominant hand provides better kappa (.652) than nondominant hand (.533). It is noted that the dominant hand provides better precision for “underhand serve” (78.79%), “serve” (80.95%), “overhead pass” (74.80%), “one hand pass” (50.00%), and “forearm pass” (75.12%). However, the nondominant hand provides better recall for “smash” (76.67%) and “block” (44.44%). It is also noted that the nondominant hand (63.30%) provides better recall for “smash” action than the dominant hand (55.05). For all other actions, the dominant hand provides better recall than the nondominant hand. It suggests that both hands are important in classifying the type of volleyball actions. That is why we also experimented with combining different sensors and also with using both the dominant and nondominant hand to see if using both hands instead of only one hand would provide better results.

Table 5 shows the UAR using fusion of different sensors and using the dominant hand, the nondominant hand, and both hands. The dominant hand gives better results (UAR = 61.79%) compared

with the nondominant hand (UAR = 54.28%). However, using both hands (UAR = 67.87%) provided better results than the dominant hand. We also noted that the linear discriminant analysis provides better results than support vector machines. For further insights, confusion matrix of the best result for both hands is shown in Figure 6. It is noted that the fusion improves precision of five volleyball actions but results in a decrease of recall for “one hand pass” (35.29%) and “block” (25.00%). However, the overall accuracy (78.17%), UAR (67.87%), and kappa (.727) are improved. It is also noted that the fusion improves the recall of five volleyball actions but results in decrease of recall for “block” (from 41.67% to 37.50%) and “forearm pass” (from 85.99% to 81.64%).

To better understand the relationship between the dominant, nondominant, and both hands, we also drew the Venn diagram shown in Figure 7. In that figure, the blue area (labeled “Target”) represents the annotated labels (i.e., ground truth), the green area represents the predicted labels when the *nondominant hand* information was used, the red area represents the predicted labels when the *dominant hand* information was used, and finally the yellow area represents the prediction obtained with the *fusion* of both hands.

Table 4 Nondominant Hand: UAR

Sensor	DT	KNN	NB	SVM	LDA	Average
Acc.	39.85	37.67	45.06	45.38	45.56	42.70
Mag.	35.70	32.40	38.65	29.37	31.36	33.50
Gyr.	33.50	32.83	36.85	32.40	31.95	33.51
Baro.	16.32	12.77	18.83	14.29	15.42	15.53
Average	31.34	28.92	34.85	30.36	31.07	—

Note. LDA = linear discriminant analysis; DT = decision trees; KNN = nearest neighbor; NB = Naive Bayes; UAR = unweighted average recall; SVM = support vector machines; Acc. = accelerometer; Mag. = magnetometer; Gyr. = gyroscope; Baro. = barometer.

Figure 4 — Confusion matrix for best result using dominant hand accelerometer and barometer and SVM method. SVM = support vector machines; UAR = unweighted average recall; UAP = unweighted average precision.

Table 3 Dominant Hand: UAR

Sensor	DT	KNN	NB	SVM	LDA	Average
Acc.	46.26	54.09	50.29	61.45	57.53	53.92
Mag.	35.67	34.98	37.72	36.31	40.88	37.11
Gyr.	41.61	36.07	35.77	42.09	38.89	38.89
Baro.	24.90	15.89	14.39	21.51	22.60	19.86
Average	37.11	35.26	34.54	40.34	39.40	—

Note. SVM = support vector machines; LDA = linear discriminant analysis; DT = decision trees; KNN = nearest neighbor; NB = Naive Bayes; UAR = unweighted average recall; Acc. = accelerometer; Mag. = magnetometer; Gyr. = gyroscope; Baro. = barometer.

Figure 5 — Confusion matrix for best result using nondominant hand accelerometer, gyroscope, and magnetometer and LDA method. UAR = unweighted average recall; UAP = unweighted average precision; LDA = linear discriminant analysis.

Table 5 Sensor Fusion: UAR (in Percentage) for DH, NDH, and BH

Sensor	SVM			LDA		
	DH	NDH	BH	DH	NDH	BH
Acc.	61.45	45.38	57.61	57.53	45.56	62.96
Mag.	36.31	29.37	44.50	40.88	31.36	50.12
Gyr.	42.09	32.40	42.50	38.89	31.95	47.54
Baro.	21.51	14.29	17.40	22.60	15.42	25.76
Acc. + Mag.	59.08	45.58	60.14	61.28	50.79	65.87
Acc. + Gyr.	55.71	45.20	44.99	61.19	49.67	64.14
Acc. + Baro.	61.79	45.37	54.99	58.34	49.12	63.47
Gyr. + Mag.	47.36	36.93	43.41	50.71	40.24	61.24
Acc. + Mag. + Gyr.	55.50	43.76	44.06	60.95	54.28	67.87
Acc. + Gyr. + Baro.	55.92	44.54	44.47	61.06	50.54	64.72
All	55.43	43.59	44.22	59.76	53.87	67.78

Note. LDA = linear discriminant analysis; DH = dominant hand; NDH = nondominant hand; BH = both hands; UAR = unweighted average recall; SVM = support vector machines; Acc. = accelerometer; Mag. = magnetometer; Gyr. = gyroscope; Baro. = barometer.

The Venn diagram suggests that the information captured by dominant and nondominant hand is not similar, as only 320 out of 646 instances are detected by all the methods (i.e., dominant, nondominant, and fusion) and there are 74 out of 646 instances which have not been captured by any of methods. Those 74 instances contain eight of “block,” 16 of smash, one of “underhand serve,” 12 of “serve,” nine of “overhead pass,” 18 of “one hand pass” and 10 of “forearm pass.”

Discussion

The results reported above show that the dominant hand plays an important role in classifying the type of action, compared with the

Figure 6 — Confusion matrix for both hands and using accelerometer, gyroscope, and magnetometer and LDA method. UAR = unweighted average recall; UAP = unweighted average precision; LDA = linear discriminant analysis.

Figure 7 — Venn diagram of the best results.

nondominant hand which provided better results for action versus no action classification (Haider et al., 2019). However, the nondominant hand certainly plays a useful role in action type classification as the results improved to 67.87% UAR compared with 61.79% using only the dominant hand. The results are highly applicable as they demonstrate the added value of using sensors on both arms for type of action classification compared with using only one arm.

The results are highly encouraging and show the viability of the trained model to be used in a real-time system (Salim et al., 2019a). While the 67.87% UAR does leave room for improvement, it is our contention that it can be easily achieved by collecting data from a couple of additional training sessions, as the models are currently trained over a single training session in which players were encouraged to play naturally resulting in an unbalanced data set.

This article focuses on the type of volleyball action recognition. The overall approach works using two steps classification method (see Figure 1). First the system classifies start and end times of an action and nonaction event (Haider et al., 2019, 2020)

(i.e., binary class problem, see Step 2 in Figure 1) and then upon detection of an action event, it further classifies the type of action (the focus of this article). In a real-life scenario, the system will use the machine learning models for both classification steps, that is, action versus nonaction classification (Haider et al., 2019, 2020) and type of action classification (see “Experimentation” section).

Concluding Remarks

This paper has proposed and described an approach to model volleyball player behavior for analysis and feedback. The described system and machine learning models automatically identify volleyball-specific actions and automatically tag video footage to enable easy access to relevant information for players and coaches. In addition to saving time and effort on the coach's behalf, the proposed approach opens up new possibilities for coaches to analyze player performance and provide quick and adaptive feedback during the training session.

The presented experiment also demonstrated the role of the dominant and the nondominant hand in the classification of volleyball action type, and presented evaluation results of different sensors and machine learning methods. The results on the relatively small and unbalanced data set are highly encouraging and applicable.

Future Directions

The outcome of the presented paper has the potential to be extended in multiple ways. In terms of machine learning models, we plan to use frequency domain features, such as scalogram and spectrogram instead of the time-domain features currently used to train the models. Apart from extending the machine learning models, the aim is to further develop the video tagging system from a proof of concept prototype to a more functional and integrated system.

The following list summarizes possible ways to extend the project:

- Further classify the actions
- Using frequency-domain approaches for feature extraction, such as scalogram, spectrogram.
- Using transfer learning approaches, such as ResNet, AlexNet, and VGGNet.
- Classification based on the above feature set.
- Further integration of Demo system and models.

In terms of further development and testing of the proposed system, we plan to conduct user studies with coaches and participants to understand the ways in which it can enhance their experience while performing their regular tasks. The user studies will be conducted using user-centric design approaches and with systematic feedback from the participants to not only understand how the system is being used by them, but also what functionalities can be added to the system to further enhance its usability for coaches and player alike.

Notes

1. <https://www.volleybal.nl/eredivisie/dames>—last accessed (June 2020).
2. <http://uk.mathworks.com/products/matlab/> (December 2018).

References

- Bagautdinov, T., Alahi, A., Fleuret, F., Fua, P., & Savarese, S. (2017). Social scene understanding: End-to-end multi-person action localization and collective activity recognition. *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 3425–3434. doi:10.1109/CVPR.2017.365
- Bellusci, G., Dijkstra, F., & Slycke, P. (2018). *Xsens MTw: Miniature wireless inertial motion tracker for highly accurate 3D kinematic applications* (pp. 1–9). Enschede: Xsens Technologies. doi:10.13140/RG.2.2.23576.49929
- Blank, P., Hoßbach, J., Schuldhaus, D., & Eskofier, B.M. (2015). Sensor-based stroke detection and stroke type classification in table tennis. *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, 93–100. doi:10.1145/2802083.2802087
- Brugman, H., Russel, A., & Nijmegen, X. (2004). Annotating multimedia/multi-modal resources with ELAN. In M. Lino, M. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)* (pp. 2065–2068). Paris, France: European Language Resources Association.
- Cancela, J., Pastorino, M., Tzallas, A.T., Tsipouras, M.G., Rigas, G., Arredondo, M.T., & Fotiadis, D.I. (2014). Wearability assessment of a wearable system for Parkinson's disease remote monitoring based on a body area network of sensors. *Sensors*, 14(9), 17235–17255. doi:10.3390/s140917235
- Cuspinera, L.P., Uetsuji, S., Morales, F.J.O., & Roggen, D. (2016). Beach volleyball serve type recognition. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (pp. 44–45). New York, NY: ACM. doi:10.1145/2971763.2971781
- Haider, F., Salim, F.A., Naghashi, V., Tasdemir, S.B.Y., Tengiz, I., Cengiz, K., . . . Luz, S. (2019). Evaluation of dominant and non-dominant hand movements for volleyball action modelling. In *ICMI '19: Adjunct of the 2019 International Conference on Multimodal Interaction*. Suzhou, China: ACM.
- Haider, F., Salim, F.A., Postma, D.B.W., van Delden, R., Reidsma, D., van Beijnum, B.-J., & Luz, S. (2020). A super-bagging method for volleyball action recognition using wearable sensors. *Multimodal Technologies and Interaction*, 4(2), 33. doi:10.3390/mti4020033
- Harvey, S., & Gittins, C. (2014). Effects of integrating video-based feedback into a Teaching Games for Understanding soccer unit. *Agora Para La Educación Física y El Deporte*, 16(3), 271–290.
- Ismail, S.I., Osman, E., Sulaiman, N., & Adnan, R. (2016). Comparison between marker-less kinect-based and conventional 2D motion analysis system on vertical jump kinematic properties measured from sagittal view. *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*, 392(2007), 11–17. doi:10.1007/978-3-319-24560-7
- Jensen, M.M., Rasmussen, M.K., Mueller, F.F., & Grønbæk, K. (2015). Keepin' it real. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems—CHI'15*, 2003–2012. doi:10.1145/2702123.2702243
- Kajastila, R., & Hämmäläinen, P. (2015). Motion games in real sports environments. *Interactions*, 22(2), 44–47. doi:10.1145/2731182
- Kautz, T., Groh, B.H., & Eskofier, B.M. (2015, August). *Sensor fusion for multi-player activity recognition in game sports*. Paper presented at the KDD Workshop on Large-Scale Sports Analytics, Sydney, Australia.
- Kautz, T., Groh, B.H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B.M. (2017). Activity recognition in beach volleyball using a Deep Convolutional Neural Network: Leveraging the potential of Deep Learning in sports. *Data Mining and Knowledge Discovery*, 31(6), 1678–1705. doi:10.1007/s10618-017-0495-0

- Koekoek, J., van der Mars, H., van der Kamp, J., Walinga, W., & van Hilvoorde, I. (2018). Aligning digital video technology with game pedagogy in physical education. *Journal of Physical Education, Recreation & Dance*, 89(1), 12–22. doi:[10.1080/07303084.2017.1390504](https://doi.org/10.1080/07303084.2017.1390504)
- Kos, M., Ženko, J., Vlaj, D., & Kramberger, I. (2016). *Tennis stroke detection and classification using miniature wearable IMU device*. Paper presented at 2016 International Conference on Systems, Signals and Image Processing (IWSSIP), Bratislava, Slovakia. doi:[10.1109/IWSSIP.2016.7502764](https://doi.org/10.1109/IWSSIP.2016.7502764)
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi:[10.2307/2529310](https://doi.org/10.2307/2529310)
- Liu, Y., Nie, L., Liu, L., & Rosenblum, D.S. (2016). From action to activity. *Neurocom*, 181(C), 108–115. doi:[10.3390/s130405317](https://doi.org/10.3390/s130405317)
- Lu, Y., Wei, Y., Liu, L., Zhong, J., Sun, L., & Liu, Y. (2017). Towards unsupervised physical activity recognition using smartphone accelerometers. *Multimedia Tools and Applications*, 76(8), 10701–10719. doi:[10.1007/s11042-015-3188-y](https://doi.org/10.1007/s11042-015-3188-y)
- Ludvigsen, M., Fogtmann, M.H., & Grønbaek, K. (2010). TacTowers: An interactive training equipment for elite athletes. *Proceedings of the 6th Conference on Designing Interactive Systems (DIS'10)*, 412–415. doi:[10.1145/1858171.1858250](https://doi.org/10.1145/1858171.1858250)
- Matejka, J., Grossman, T., & Fitzmaurice, G. (2014). Video lens: Rapid playback and exploration of large video collections and associated metadata. *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST'14)*, 541–550. doi:[10.1145/2642918.2647366](https://doi.org/10.1145/2642918.2647366)
- Mitchell, E., Monaghan, D., & O'Connor, N.E. (2013). Classification of sporting activities using smartphone accelerometers. *Sensors*, 13(4), 5317–5337. doi:[10.3390/s130405317](https://doi.org/10.3390/s130405317)
- Nguyen, L., Nguyen, N., Rodríguez-martín, D., Pérez-lópez, C., Samà, A., & Cavallaro, A. (2015). Basketball activity recognition using wearable inertial measurement units. *Proceedings of the XVI International Conference on Human Computer Interaction*, 1–6. doi:[10.1145/2829875.2829930](https://doi.org/10.1145/2829875.2829930)
- Pei, W., Wang, J., Xu, X., Wu, Z., & Du, X. (2017). An embedded 6-axis sensor based recognition for tennis stroke. *Proceedings of the 2017 IEEE International Conference on Consumer Electronics (ICCE)*, 55–58. doi:[10.1109/ICCE.2017.7889228](https://doi.org/10.1109/ICCE.2017.7889228)
- Postma, D., Van Delden, R., Walinga, W., Koekoek, J., Van Beijnum, B.J., Salim, F.A., ... Reidsma, D. (2019). Towards smart sports exercises: First designs. *Proceedings of the Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY' 2019)*, 619–630. doi:[10.1145/3341215.3356306](https://doi.org/10.1145/3341215.3356306)
- Salim, F., Haider, F., Tasdemir, S.B.Y., Naghashi, V., Tengiz, I., Cengiz, K., ... van Beijnum, B.-J. (2019a). A searching and automatic video tagging tool for events of interest during volleyball training sessions. In *2019 International Conference on Multimodal Interaction* (pp. 501–503). New York, NY: ACM. doi:[10.1145/3340555.3358660](https://doi.org/10.1145/3340555.3358660)
- Salim, F.A., Haider, F., Tasdemir, S., Naghashi, V., Tengiz, I., Cengiz, K., ... van Beijnum, B.J.F. (2019b). *Volleyball action modelling for behavior analysis and interactive multi-modal feedback*. Paper presented at 15th International Summer Workshop on Multimodal Interface (eINTERFACE'19), Ankara, Turkey.
- Schuldhaus, D., Zwick, C., Körger, H., Dorschky, E., Kirk, R., & Eskofier, B.M. (2015). Inertial sensor-based approach for shot/pass classification during a soccer match. *Proceedings of the 21st ACM KDD Workshop on Large-Scale Sports Analytics*, 27, 1–4. Retrieved from <https://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2015/Schuldhaus15-ISA.pdf>
- Stensland, H.K., Landsverk, Ø., Griwodz, C., Halvorsen, P., Stenhaug, M., Johansen, D., ... Ljødal, S. (2014). Bagadus: An integrated real time system for soccer analytics. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 10(1s), 1–21. doi:[10.1145/2541011](https://doi.org/10.1145/2541011)
- Thomas, G., Gade, R., Moeslund, T.B., Carr, P., & Hilton, A. (2017). Computer vision for sports: Current applications and research topics. *Computer Vision and Image Understanding*, 159, 3–18. doi:[10.1016/j.cviu.2017.04.011](https://doi.org/10.1016/j.cviu.2017.04.011)
- Vales-Alonso, J., Chaves-Dieguez, D., Lopez-Matencio, P., Alcaraz, J.J., Parrado-García, F.J., & Gonzalez-Castano, F.J. (2015). SAETA: A smart coaching assistant for professional volleyball training. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(8), 1138–1150. doi:[10.1109/TSMC.2015.2391258](https://doi.org/10.1109/TSMC.2015.2391258)
- Velasco, R. (2016). *Apache Solr: For starters*. Scotts Valley, CA: CreateSpace Independent Publishing Platform.
- von Marcard, T., Rosenhahn, B., Black, M.J., & Pons-Moll, G. (2017). Sparse inertial poser: Automatic 3D human pose estimation from sparse IMUs. *Computer Graphics Forum*, 36(2), 349–360. doi:[10.1111/cgf.13131](https://doi.org/10.1111/cgf.13131)
- Wang, Y., Zhao, Y., Chan, R.H.M., & Li, W.J. (2018). Volleyball skill assessment using a single wearable micro inertial measurement unit at wrist. *IEEE Access*, 6, 13758–13765. doi:[10.1109/ACCESS.2018.2792220](https://doi.org/10.1109/ACCESS.2018.2792220)
- x-io Technologies. (2019). NG-IMU. Retrieved from <http://x-io.co.uk/ngimu/>
- Zivkovic, Z., van der Heijden, F., Petkovic, M., & Jonker, W. (2001). Image segmentation and feature extraction for recognizing strokes in tennis game videos. In R.L. Langendijk, J.W.J. Heijnsdijk, A.D. Pimentel, & M.H.F. Wilkinson (Eds.), *Proceedings 7th Annual Conference on the Advanced School for Computing and Imaging (ASCI 2001)* (pp. 262–266). Delft, the Netherlands: Advanced School for Computing and Imaging (ASCI).