Sensor fusion for multi-player activity recognition in game sports

Thomas Kautz
Digital Sports Group
Friedrich-Alexander University
Erlangen-Nürnberg (FAU)
Martensstraße 3
Erlangen, Germany
thomas.kautz@fau.de

Benjamin H. Groh
Digital Sports Group
Friedrich-Alexander University
Erlangen-Nürnberg (FAU)
Martensstraße 3
Erlangen, Germany
benjamin.groh@fau.de

Bjoern M. Eskofier
Digital Sports Group
Friedrich-Alexander University
Erlangen-Nürnberg (FAU)
Haberstraße 2
Erlangen, Germany
bjoern.eskofier@fau.de

ABSTRACT
The use of wearable sensors for automatic recognition of human activities has pervaded both professional and recreational sports. While many activities involving only a single athlete can be classified robustly, the automatic classification of complex activities involving several athletes is still in its infancy. In this paper, we present a novel approach for the recognition of such multi-player activities in the context of game sports. Our method is based on the fusion of position measurements with inertial measurements in a set of interaction features. We demonstrate the efficacy of our method in the recognition of tackles and scrums in Rugby Sevens. The results of our current work suggest that the proposed features can be leveraged to achieve classification accuracies of more than 97 %.

Categories and Subject Descriptors
I.5.4 [**Pattern Recognition**]: Applications

Keywords
activity recognition, IMU, wearable sensor, feature, sports, data fusion

1. INTRODUCTION

1.1 Motivation
Advances in pattern recognition methodology and measurement technology as well as miniaturization of sensors have propelled the use of various kinds of sensing devices in both professional and recreational sports. Positioning systems, video recordings and inertial measurements units (IMUs) are used to gather vast amounts of data during competitions and training. In order to transform this data into useful information and even knowledge, sophisticated analysis procedures are required. Research in the recognition of complex activities with multiple participants as they occur in game sports is still in its infancy [5]. Various work has been published on the analysis of activity recognition using IMUs [1, 12, 15]. In these publications, IMU measurements from each subject were analyzed separately. Many activities, as they occur for example in game sports include several persons interacting with each other. By analyzing IMU data from interacting subjects separately, information about mutual influences and interactions which could improve the activity recognition is discarded. However, player interactions are an essential part of game sports and therefore should be included in the analysis procedure.

One possible application for the automatic recognition activities that include several interacting players is Rugby Sevens. Rugby Sevens has gained increasing interest all over the world in the last years. Rugby Sevens is a variation of Rugby Union with only 7 players per team on the field. An automatic analysis system for Rugby could provide spectators, players and trainers with valuable information about a match and individual player performance. Moreover it could be used to identify risks of injuries of the players. It has been shown that tackles are the main cause of injuries in rugby [2] and that there is significant correlation between the number of tackles and different indicators for muscle damage [17]. An automatic detection of tackles during matches and training could be used to reveal players that are involved in a high number of tackles. This information can then be employed to adapt the training or the game strategy to relieve such players and thereby avoid injuries.

The combination of inertial and position measurements offers new opportunities for the recognition of activities which include multiple participants. Features, which combine information from both measurement systems as well as features which capture relations and interactions between several players can be acquired. In order to demonstrate how such features can be leveraged for activity recognition in game sports, we focus on the automatic recognition of two important elements of the Rugby game: tackles and scrums.

1.2 Definitions
For our work, we use the definitions from the World Rugby laws of the game [18]. There, a tackle is defined as follows:

Definition 1. "A tackle occurs when the ball carrier is held by one or more opponents and is brought to ground. A ball carrier who is not held is not a tackled player and a tackle has not taken place. Opposition players who hold the ball

KDD Workshop on Large-Scale Sports Analytics 2015, Sydney, Australia

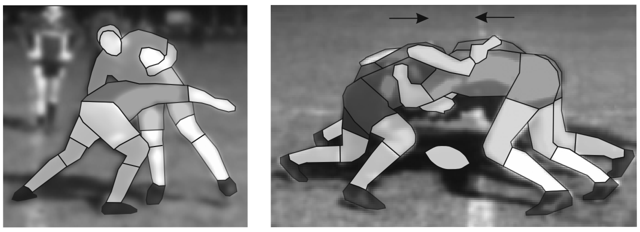

Figure 1: Left – Tackle: The ball carrier is brought to the ground by 1 or more opponents Right – Scrum: 3 players of each team interlock and compete for ball possession

carrier and bring that player to ground, and who also go to ground, are known as tacklers. Opposition players who hold the ball carrier and do not go to ground are not tacklers."

In Rugby Sevens, a scrum is a specific formation of 6 players (3 of each team):

Definition 2. "A scrum is formed in the field of play when three players from each team, bound together in one row, close up with their opponents so that the heads of the players are interlocked. This creates a tunnel into which a scrum half throws the ball so that the players can compete for possession by hooking the ball with either of their feet"

Examples of a tackle and a scrum are depicted in Fig. 1.

1.3 Related Work

Methods for the automatic recognition of activities that include several participants have mostly been published in the context of computer vision. For example, Oliver et al. [13] and Du et al. [7] model human interactions based on video data. However, in these papers only interactions between 2 or 3 participants were considered. The described methods require detailed modeling of the analyzed processes. In game sports, where many players interact at the same time (e.g. 14 players in Rugby Sevens) a huge variety of possible interactions would have to be modeled in order to reflect all nuances of the game. The complexity of such an approach is prohibitive for the direct application of the previously mentioned methods to a game sport like Rugby. Direkoglu and O'Connor [6] and Perše et al. [14] use player trajectories to classify different team activities. While these trajectory-based approaches allow the recognition of some team activities, they are not able to capture detailed player interactions, as for example tackles in football and rugby.

Only little work about automatic event detection and classification in rugby has been published. Kelly et al. [11] presented a method for automatic detection of collisions in rugby using data from wearable acceleration sensors. Although a collision always includes at least two players, sensor data was analyzed separately for each player, thus ignoring possible interrelations between the sensor signals from different players.

The purpose of this paper is to present a novel approach for the automatic multi-player activity recognition in game sports. In our approach we analyze data from all players simultaneously. Searching for all possible player interactions would lead to unacceptable combinatorical complexity. Therefore, we also measure player positions by means of a local positioning system (LPS) and identify possibly interacting players based on their respective positions. To demonstrate our methods, we focus on Rugby Sevens as a possible area of application. First, we describe our measurement setup consisting of an LPS and IMUs. Second, the conducted study for data acquisition is briefly outlined. In the succeeding section, we point out how the advantages of the multi-modality measurement can be used for capturing player interactions within a set of novel expert features. We show how these features can be leveraged to analyze complex activities including several participants in game sports.

2. METHODS

2.1 Data Acquisition

2.1.1 Sensor Hardware

For data acquisition, we used a LPS which measured the positions of all players in a 2D local coordinate system. The LPS consisted of 12 base stations and a reference transponder with fixed and known positions as well as mobile target transponders which were attached to the shirts of the players. The LPS provided time difference of arrival measurements from which the player positions were obtained. In addition to the position measurements, the movements of each player were monitored using the miPod IMU [3]. It contained a 3-axis accelerometer (range $\pm 16\,\mathrm{g}$) and a 3-axis gyroscope (range $\pm\,2000\,^\circ/\mathrm{s}$). Inertial measurements were recorded with a sampling rate of 200 Hz. Each player was equipped with one LPS target transponder and one IMU. The sensors were put in a neoprene pouch for cushioning which was attached on the inner side of the training shirts. The sensor was positioned at the upper back and centered on mid-sagittal plane. For reference, the data acquisition was filmed with a GoPro camera (resolution 1280x960 pixels, 50 frames per second).

2.1.2 Study Design

In order to capture different situations, data was recorded during several training sessions. 7 male and 10 female amateur Rugby players participated in the study. Participating players were asked to sign an information sheet to document their informed consent. The recordings included a test game, different types of tackles (front-on, from left and right side, from behind), scrums and different running exercises. In total, approx. 10 hours of data was recorded and analyzed. This included 50 tackles and 6 scrums which were manually labeled based on the video recordings as ground truth. The tackles were performed by 7 of the 17 subjects, the scrums were performed by 6 subjects. During all sessions the trajectories of the players were captured with the LPS. In addition, accelerations and turning rates were recorded with the IMUs placed on the upper back of the players.

2.2 Preprocessing

Since the raw LPS position measurements were obtained with a non-uniform sampling rate (approx. 80 Hz on average) and contained measurement noise and outliers, the position data was filtered and resampled to a constant sampling rate. To this end, an outlier-robust Extended Kalman Filter (EKF) with fixed-lag Rauch-Tung-Striebel smoothing (100 ms lag) was used [10]. To facilitate further analyses, the measurements from the LPS and the IMUs were synchronized using pre-defined movement patterns that the subjects were asked to perform. All acquired signals were interpolated to a common time stamp (100 Hz). The IMU measurements were calibrated according to the methods proposed by Ferraris et al. [8]. The coordinate systems of the accelerometers and gyroscopes were rotated to match the body axes of the players.

2.3 Activity Recognition

2.3.1 Tackle Recognition

For the automatic recognition of tackles, we first applied an event detection procedure to identify possible candidates for a tackle. Tackles include a collision of the ball carrier with the opposing tackling player and a second impact, when the ball carrier (usually also the tackling player) hits the ground. These events are accompanied by high accelerations, experienced both by the tackled and the tackling player. To detect these events, we used a threshold detection based on the low-passed norm of the acceleration signal (cutoff-frequency 1 Hz). The optimal cutoff-frequency was determined with a grid search. The threshold detection results in a number of time points for each player where a tackle potentially took place. The threshold was defined such that all labeled tackles were detected (sensitivity 100 %) in a grid search. With this approach, not only tackles were detected, but also other events with a high impact, e.g. non-tackle collisions of players were marked as possible tackles.

In order to distinguish between tackles and other detected events, a second classification step was used. Therefore, signal windows with a duration of 4 seconds centered around the detected event times were extracted. For these signal windows, different generic and expert features were calculated. While in previous work signals were analyzed independently for each player [11], we defined features which are able to capture interactions between players. For each detected event, we searched for possible interaction candidates. As possible candidates, we determined the 3 closest opposing players at the time of the event based on the position data. These can be regarded as possible interaction candidates. Due to the often close proximity of multiple players during a game, this approach is more robust than selecting only the closest opposing player. This mapping allows for the calculation of features that include the player for whom an event was detected *and* his or her possibly interacting opponents. In our preliminary tests we calculated the following features (amongst others) for the detected events on the signals of the current player and the interaction candidates:

- torso pitch angle (with sensor alignment correction)
- speed (estimated from position data in the EKF)
- distance to interaction candidates
- statistical features of the acceleration and gyroscope signals (mean, standard deviation, skewness, kurtosis)

To identify the most important features, feature selection was performed with a wrapper-based 10-fold cross-validated Best First feature selection. For the subsequent classification only features that were selected in at least two folds were used. Detected events were classified using a Bayes Network, Random Forest, Multilayer Perceptron and a Naïve Bayes classifier. To account for the class imbalance of the detected events (over-representation of non-tackle events), SMOTE was applied to increase the number of tackle events in the training sets [4]. The feature selection and classification were performed using Weka [9] and the Embedded Classification Software Toolbox (ECST) [16]. The evaluation of the classification was conducted using a Leave-One-Subject-Out cross-validation.

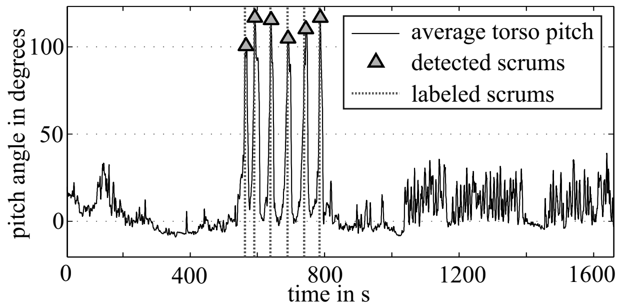

Figure 2: Scrum detection based on average torso pitch of 6 players with highest pitch

2.3.2 Scrum Recognition

In a scrum in Rugby Sevens, three players of each team interlock as exemplified in Fig. 1. In this position, they are bent forward considerably when compared to the natural upright position during standing or running. This results in a high torso pitch angle. Considering this, only a single expert feature was calculated for the scrum detection: for each time window the torso pitch was estimated for all players and sorted in descending order. The torso pitch angles of the 6 players with the highest pitch were then averaged for each window. This signal exhibits pronounced peaks for the episodes during which a scrum took place (see Fig. 2). These peaks were then detected with a threshold-based method. The threshold was determined in a grid search such that none of the performed scrums was missed in the detection.

3. RESULTS

The threshold-based event detection for the identification of possible tackles achieved a sensitivity of 100 % and a precision of 11.6 %. The feature selection for the second classification that was used to distinguish between tackles and other detected events led to different results for the different classifiers. The three features that were selected most often and which were used in conjunction with all classifiers were: the torso pitch of the player for whom an event was detected, the torso pitch of the closest player and the distance to the closest player. Using the selected classifier-specific features for the second stage of the tackle classification led to the sensitivities and specificities summarized in Tab. 1. The highest balanced accuracy was obtained with the Naïve Bayes (NB) classifier (97.2 %).

With the threshold method based on the described expert feature, scrums could be detected with 100 % sensitivity and specificity.

4. DISCUSSION

The threshold based event detection allowed the reliable detection of all labeled tackles. However, also many other

	RF	MP	BN	NB
sensitivity	95.7 %	97.9 %	95.7 %	96.8 %
specificity	97.8 %	96.0 %	96.9 %	97.5 %

Table 1: Classification results using Random Forest (RF), Multilayer Perceptron (MP), Bayes Network (BN) and Naïve Bayes (NB) classifiers

events were detected, which led to a poor precision (11.6 %). This made the second classification step necessary. The feature selection for this second classification step showed that the novel interaction features (e.g. torso pitch of the closest player) are amongst the most important features. Capturing interactions within the features allows the use of off-the-shelf classifiers without modifications. This facilitates the implementation and application of the proposed methods.

Using the novel features, all tested classifiers achieved high sensitivities and specificities (>95 %) for the second stage of the tackle classification. Since the sensitivity of the first stage (event detection) was 100 %, the results in Tab. 1 reflect the overall performance of the presented tackle recognition. The results showed only small differences in the performance of the tested classifiers.

For the scrum detection, a simple threshold was sufficient to detect all recorded scrums based on one single interaction feature.

These results show that multi-player activities can be recognized effectively with the introduced approach. With the proposed methods, different events during the game can be detected and classified in a fully automatic manner, even without using complex interaction models. However, it needs to be considered that the presented results are based on the recordings from rugby training sessions which included only a relatively small number of tackles and scrums. Therefore, the results have to be regarded critically and further data acquisition and evaluation is necessary to affirm the presented methods on a wider scale.

5. CONCLUSION AND FUTURE WORK

Our preliminary results regarding the automatic recognition of tackles and scrums in Rugby suggest that activities with several participants can be detected reliably with the help of the proposed interaction features. These enable the implicit modelling of complex activities including multiple participants within the feature space. This approach circumvents the necessity of creating extensive explicit models. The simultaneous use and subsequent fusion of both inertial measurements and absolute position measurements offers new possibilities for the automatic analysis of Rugby and other game sports. Our goal is to combine both measurement systems to objectively and reliably analyze game sports both on the level of single events such as tackles and at the level of team strategy and tactics.

6. ACKNOWLEDGMENTS

We would like to thank the Rugby Club Leipzig for their enthusiastic cooperation and kind support.

7. REFERENCES

[1] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *2nd Int. Conf. on Pervasive Computing (PerCom)*, pages 1 – 17, 2004.

[2] A. Bathgate, J. P. Best, G. Craig, and M. Jamieson. A prospective study of injuries to elite australian rugby union players. *British J. Sports Medicine*, 36(4):265–269, 2002.

[3] P. Blank, P. Kugler, H. Schlarb, and B. Eskofier. A wearable sensor system for sports and fitness applications. In *19th Annu. Congr. of the European College of Sport Science*, 2014.

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *J. of Artificial Intelligence Research*, 16(1):321–357, 2002.

[5] L. Chen et al. Sensor-based activity recognition. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, 42(6):790–808, 2012.

[6] C. Direkoglu and N. E. O'Connor. Team activity recognition in sports. In *12th European Conf. on Computer Vision (ECCV)*, pages 69–83, 2012.

[7] Y. Du, F. Chen, W. Xu, and Y. Li. Recognizing interaction activities using dynamic bayesian network. In *18th Int. Conf. on Pattern Recognition (ICPR)*, pages 618–621, 2006.

[8] F. Ferraris, U. Grimaldi, and M. Parvis. Procedure for effortless in-field calibration of three-axial rate gyro and accelerometers. *Sensors and Materials*, 7(5):311–330, 1995.

[9] M. Hall et al. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[10] T. Kautz and B. M. Eskofier. A Robust Kalman Framework with Resampling and Optimal Smoothing. *Sensors*, 15(3):4975–4995, 2015.

[11] D. Kelly, G. F. Coughlan, B. S. Green, and B. Caulfield. Automatic detection of collisions in elite level rugby union using a wearable sensing device. *Sports Engineering*, 15:81–92, 2012.

[12] H. Leutheuser, D. Schuldhaus, and B. M. Eskofier. Hierarchical, multi-sensor based classification of daily life activities: Comparison with state-of-the-art algorithms using a benchmark dataset. *PLoS ONE*, 8(10), 2013.

[13] N. M. Oliver, B. Rosario, and A. P. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):831–843, 2000.

[14] M. Perše et al. Analysis of multi-agent activity using petri nets. *Pattern Recognition*, 43(4):1491–1501, 2010.

[15] N. Ravi, N. Dandekar, P. Mysore, and M. M. L. Littman. Activity recognition from accelerometer data. *AAAI J.*, 5:1541–1546, 2005.

[16] M. Ring, U. Jensen, P. Kugler, and B. Eskofier. Software-based performance and complexity analysis for the design of embedded classification systems. In *21st Int. Conf. on Pattern Recognition (ICPR)*, pages 2266–2269, 2012.

[17] Y. Takarada. Evaluation of muscle damage after a rugby match with special reference to tackle plays. *British J. Sports Medicine*, 37(5):416–419, 2003.

[18] World Rugby. *Laws of the Game Rugby Union*. World Rugby, Dublin, Ireland, 2015.