Your Personal Movie Producer: Generating Highlight Videos in Soccer Using Wearables

Dominik Schuldhaus[1], Carolin Jakob[1], Constantin Zwick[2], Harald Koerger[2], Bjoern M. Eskofier[1]
[1] Digital Sports Group, Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany
[2] adidas AG, Herzogenaurach, Germany
Corresponding author: *dominik.schuldhaus@fau.de*

ABSTRACT
Manually browsing through high amount of sports videos, selecting interesting highlight scenes, and applying video effects are time-consuming and one major burden these days. Automatic approaches are preferred, but currently require high-quality TV broadcast material which is usually not available in recreational sports. Thus, the purpose of this paper was to develop a personal low-cost movie producer for highlight videos using wearables. The feasibility of the proposed approach was shown for soccer scenarios. The automatic highlight video generation included three contributions: (i) sensor-based full-instep kick detection and extraction of corresponding video segments, (ii) sensor-based ball speed estimation for provision of highlight-related metadata shown in the final video, and (iii) sensor-driven video effect generation. The proposed system was evaluated on eleven subjects which were equipped with inertial sensors in the cavity of soccer shoes. A mean sensitivity of 95.6 % and a mean absolute error of 7.7 km/h were achieved for the full-instep kick classification and the ball speed estimation, respectively. This personal movie producer based on wearables is a novel idea to provide recreational athletes with attractive automatically generated highlight videos in sports.

Author Keywords
Wearables; Inertial Sensors; Soccer; Data Mining; Highlight Videos.

ACM Classification Keywords
I.5.4. Pattern Recognition: Applications

INTRODUCTION
Sports videos are highly attractive internationally, with massive production and a huge audience base [5]. Nevertheless, the value of sports videos drops after a while. A short and compact summary (highlights) often appears to be more attractive than the whole video sequence and offers more personalized applications [14]. Since manually browsing through the high amount of captured video data and selecting interesting highlights are time-consuming, automatic approaches are mandatory.

In recent years, algorithms for the automatic generation of highlight videos were developed for TV broadcast sports videos [7, 13, 2]. TV broadcast material includes high-resolution video data, special cinematic features like replay scenes and textual overlays, and audio data of announcers. Current state-of-the-art highlight video generation techniques exploit these features in the proposed systems. Compared to high-quality TV broadcast material in professional sports, videos in recreational sports are mainly based on low-cost video equipment, e.g. camcorder or smartphone camera. The mentioned video generation techniques are therefore hard to apply to the described low-quality data. There is a major need to provide users in recreational sports with low-cost tools for the automatic highlight video generation.

Thus, the purpose of this paper was to develop a personal low-cost movie producer using wearables. Wearable-based signal processing was applied in order to detect time sequences of highlights and extract specific metadata for the highlight. Based on the derived time sequences, the video cutting is performed with additional video effect generation. As a proof of concept the proposed architecture was applied to soccer. Wearables were placed inside the apparel of soccer players. The wearables were used for (i) full-instep kick classification (highlight), (ii) ball speed estimation (metadata), and (iii) sensor-based video effects to increase the attractiveness.

RELATED WORK
In recent years, algorithms for the automatic generation of highlight videos were developed for TV broadcast sports videos, e.g. in baseball [13], football [2], and soccer [7]. Most approaches of highlight video generation for TV broadcast material are based on visual, audio, and textual features extracted from the video stream itself. In [7], low-level video-based analysis for cinematic feature extraction was performed including robust dominant color region detection, video shot boundary detection, video shot classification, and slow-motion replay detection. Goals were detected by a cinematic template that should fulfill certain requirements. The algorithm was evaluated on more than 13 hours of soccer

ISWC '16, September 12-16 2016, Heidelberg, Germany

ACM 978-1-4503-4460-9/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2971763.2971772

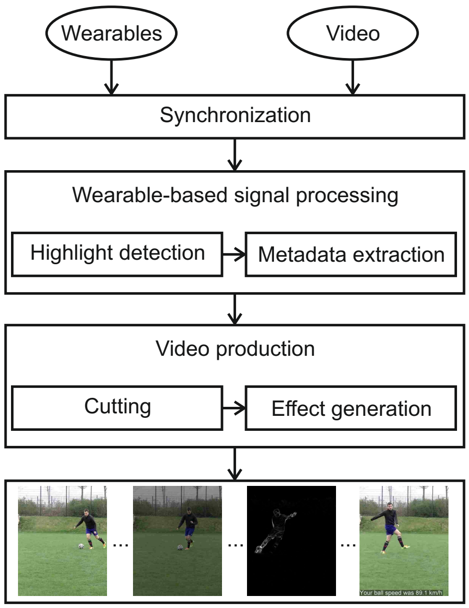


Figure 1: General architecture of movie producer.

videos and achieved a sensitivity of 90.0 %. Since cinematic features were required, the proposed approach can mainly be applied to TV broadcast material. Users in recreational sports should also be provided with automatically generated highlight videos. Thus, alternative solutions have to be developed. According to [3], a tremendous effort has been made in the field of wearable technology since the late 1990s toward closing the gap between vision and reality. Wearables like inertial sensors provide a low-cost solution for monitoring human's behavior and classifying certain events in a long period of time. Inertial sensors are more and more used in sports applications. Examples can be found e.g in team sports like soccer [6]. In [6], the general applicability of inertial sensors for event classification in soccer was shown. Nevertheless, the approach only addressed event classification not considering important metadata such as the resulting ball speed after the full-instep kick. The information of the ball speed right after a full-instep kick is an important metadata. Current solutions require expensive equipment and are usually established in a laboratory environment [11]. In this paper, the ball speed estimation was included in the processing pipeline based on inertial sensor data. The advantage is the applicability in outdoor environments with a cheap setup.

MATERIALS AND METHODS

Movie Producer Architecture

The general architecture of the proposed movie producer includes three steps, which are shown in Figure 1. In the first step, sensor data acquired by wearables and video data are synchronized. The movie producer is independent on the video source. Video data can be acquired, e.g. by a camera with fixed position or by a smartphone. In the second step, wearable-based signal processing is performed consisting of highlight detection and metadata extraction. Since full-instep and side-foot kicks are the most frequently used soccer kicks, these events were considered in the highlight detection [10]. Since the ball speed is a measure of success in soccer [9], this important value was extracted as metadata. In the third step, the wearable-driven video production is performed in order to generate the final highlight video. The three mentioned steps are performed on a local PC.

Data Acquisition

A custom-made system comprising of sensor and storage unit was used in order to acquire soccer-specific events like full-instep kicks. The sensor unit including an inertial measurement unit was placed in the cavity of a soccer shoe. The inertial measurement unit consisted of a triaxial accelerometer and a triaxial gyroscope with a range of ± 16 g and ± 2000 °/s, respectively. Soccer-specific movements were acquired with a sampling rate of 1000 Hz. The storage unit was integrated in the shin guard and was connected to the sensor unit by a cable. Each leg was equipped with one custom-made system.

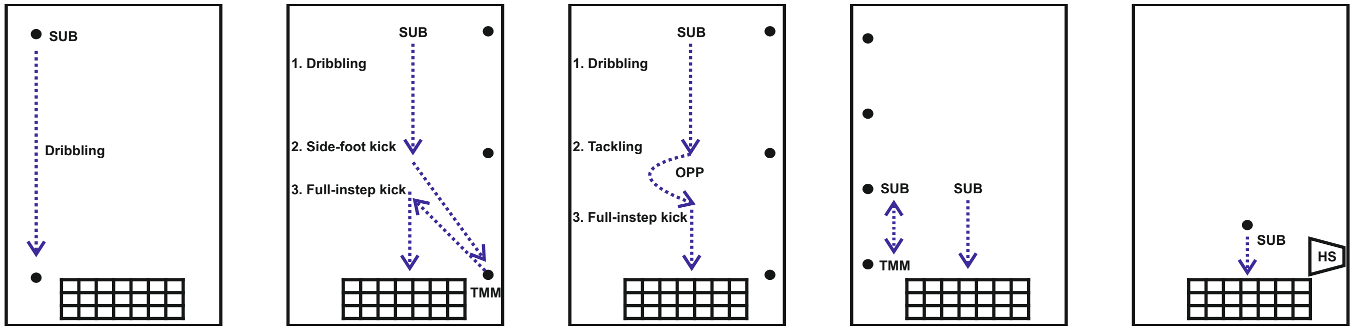
Eleven male amateur players (age 29.6 ± 9.2 years, height 182.3 ± 6.3 cm, weight 77.7 ± 9.8 kg) participated in a study and performed a protocol including six controlled exercises, which are shown in Figure 2. Exercises shown in 2a, 2b, and 2c were performed three times with the preferred leg. Figure 2d shows two exercises. Single subsequent side-foot kicks were performed eight times with each leg and three different distances. Single full-instep kicks were performed three times with each leg. Ground truth data was determined by video labeling. The mentioned exercises were used for full-instep kick classification. The exercise shown in Figure 2e was performed in a laboratory environment and used for ball speed estimation. The subject was asked to perform different full-instep kicks toward a goal with three different increasing self-selected intensities. A full-instep kick of each self-selected intensity was performed once with each leg. Ground truth ball speed was determined by simultaneously capturing high speed video data and postprocessing.

Full-Instep Kick Classification

The full-instep kick was defined as the highlight which should be included in the final video. The goal of the full-instep kick classification was to detect the event and the corresponding time points. The time points defined the borders of the video segment. The approach was divided into four major steps which are explained in the following sections.

Peak Detection

The first step was to detect peaks in the inertial sensor data. It was assumed that peaks represented candidates of full-instep kicks. A Butterworth high-pass filter was applied to the gyroscope data of left and right leg with an order of 2 and a cutoff frequency of 250 Hz. In order to reduce the complexity and to remove the direction of movement, the signal magnitude vector was computed for each high-pass filtered signal. Peaks, for which the corresponding amplitude exceeded a predefined threshold of 67.2 °/s, were assumed to be candidates for the location of full-instep kicks. The mentioned parameters were optimized by grid search in a leave-one-subject-out cross-validation (LOSO-CV). The true positive rate was used as performance measure (true class: full-instep kick, side-foot



(a) Dribbling with the ball from one cone to a second cone.

(b) Dribbling, side-foot kick to teammate, full-instep kick after receiving the ball.

(c) Dribbling, tackling against opponent, full-instep kick.

(d) Single subsequent side-foot kicks with three different distances (left) and single full-instep kicks (right).

(e) Setup for ball speed estimation.

Figure 2: Illustration of exercises (SUB: subject, TMM: teammate, OPP: opponent, HS: high speed camera); black circles indicate cones and grid pattern indicates goal; 2a, 2b, 2c, 2d were performed outdoors and used for full-instep kick classification (dimension of pitch: 16.5 m x 40 m); 2e was performed in a laboratory environment and used for ball speed estimation.

kick; false class: other non-kick instances). The following processing steps were performed for segmented sensor data around each peak.

Event Leg Classification

The second step was to classify the leg which was used to perform the full-instep kick. For the rest of the paper, this leg is denoted as event leg, the other leg is denoted as supporting leg. By the event leg classification, a distinction of the cases, in which the full-instep kick was performed with the left or right leg, was not needed. The signal magnitude area of the accelerometer signal was computed for the left and right leg in a predefined window [8]. The start and end of the window was chosen as 0.1 s before and 0.4 s after the peak motivated by [4]. The higher signal magnitude area defined the event leg.

Instant and Phase Detection

The third step was to detect different biomechanically-driven instants and phases which occurred during the execution of a full-instep kick. The instants and phases were needed for the feature extraction in both the full-instep kick classification and the ball speed estimation as well as for special video effects in the final highlight video. For the instant and phase detection, only the event leg was considered. The ball contact was determined by computing the maximum angular acceleration in the sagittal plane [12]. The start of the leg-acceleration phase was determined by computing the last zero-crossing of the angular velocity in the sagittal plane before the ball contact [11]. The last heel strike before the kick was computed according to [1].

Event Classification

The fourth step was to classify the regions around the detected peaks regarding full-instep kicks (FULL-INSTEP) and side-foot kicks (SIDE-FOOT). For the classification task, six features were extracted. The absolute sum of the signal in the leg-acceleration phase was computed for each accelerometer and gyroscope axis of the event leg. These features reflected the different movement directions. The features were used as input for a SVM classifier with linear kernel. The sensitivity was used as performance measure for the event classification and was determined by LOSO-CV. In each LOSO trial, the optimization of the cost parameter C of the SVM was performed by grid search with $C \in \{2^N\}, N \in \{-10, ..., 10\}$ and an inner LOSO-CV.

Ball Speed Estimation

The ball speed estimation consisted of two steps. In the first step, one feature was extracted exploiting the fact that the accelerometer reached its saturation in the leg-acceleration phase. In this case, the total number of samples of the saturation phase was individually computed for each axis and summed over all axes. In the second step, the feature was used as input for Multiple Linear Regression. The mean absolute error was used as performance measure for the ball speed estimation and was determined by LOSO-CV.

Video Production

Each ball contact, which was classified as part of a full-instep kick, was considered for video production. A video clip was generated with a window that was defined by the user, e.g. 1 s before and 3 s after the ball contact. Based on the outcome of the instant and phase detection, the video clip was further divided into three segments: (A) beginning of video clip to heel strike, (B) heel strike to ball contact, and (C) ball contact to end of video clip. In each segment, configurable video effects were applied, which offered personalized video clips. In segment A, a fade-out effect was performed with a configurable fade-out factor. In segment B, different images between two subsequent frames were computed. Furthermore, a slow motion effect was performed with a configurable slow-motion factor. In segment C, a predefined sound was played starting at ball contact and a text box appeared with the estimated ball speed. The position of the text box can be defined by the user. For the video production, the FFmpeg framework was used (`https://ffmpeg.org/`).

RESULTS

In this section, the results for peak detection, event classification, and ball speed estimation are described. For the peak detection and event classification, 116 full-instep kicks, 561 side-foot kicks, and 333 NULL events of eleven subjects were available. NULL events included peak-rich instances of the dribbling exercise. For the ball speed estimation, 45 instances with ground truth ball speed values between 60.3 km/h and 101.7 km/h were available for evaluation. For the peak detection a mean true positive rate of 95.7 % was achieved. For the event classification, class-dependent sensitivities of 97.9 % and 95.6 % were achieved for SIDE-FOOT and FULL-INSTEP, respectively. For the ball speed estimation, a mean absolute error of 7.7 ± 4.1 km/h was achieved.

DISCUSSION

The achieved results in peak detection, full-instep kick classification, and ball speed estimation showed the general applicability of sensor-based approaches for event analysis in soccer. The rather generic design of the peak detection can be applied to other highlights in sports which are represented as peaks in the inertial sensor data, e.g. tackling in football, stroke in tennis, or jump shot in basketball. In this work, the highlight detection step shown in the generic architecture (Figure 1) was optimized for a limited set of highlights represented as peaks in the sensor data. To cover other events in sports like dribbling sequences, walking sequences, or tricks, the highlight detection can be replaced, e.g. by sliding window approaches which are optimized for those scenarios. The novel idea of a wearable-driven personal movie producer offers athletes in recreational sports with highlight videos. The highlight videos can be used for reporting the own performance during sports, comparing different athletes, or posting on social network sites. In future work, the set of implemented video effects will be extended, e.g. burning balls, radial effects for the player, and artificial announcers.

SUMMARY

Highlight videos often appear to be more attractive than the whole video sequence and offer more personalized applications. Current automatic solutions are often based on high-quality TV broadcast material. Thus, the purpose of this paper was to develop a personal low-cost movie producer for highlight videos and to show the feasibility in soccer scenarios. Sensor-based pattern recognition techniques were applied to detect full-instep kicks, estimate the corresponding ball speed, and apply sensor-driven video effects. In the full-instep kick classification and ball speed estimation, a mean sensitivity and a mean absolute error of 95.6 % and 7.7 km/h were achieved, respectively.

ACKNOWLEDGMENTS

We thank the adidas AG for financial and technical support of the study. This work was further supported by the Bavarian Ministry of Economic Affairs and Media, Energy and Technology.

REFERENCES

1. Aminian, K., Najafi, B., Büla, C., Leyvraz, P.-F., and Robert, P. Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. *J Biomech 35*, 5 (2002), 689–699.
2. Babaguchi, N., Kawai, Y., Ogura, T., and Kitahashi, T. Personalized Abstraction of Broadcasted American Football Video by Highlight Selection. *IEEE Trans Multimedia 6*, 4 (2004), 575–586.
3. Bonato, P. Wearable sensors and systems. *IEEE Eng Med Biol Mag 29*, 3 (2010), 25–36.
4. Brophy, R. H., Backus, S., Kraszewski, A. P., Steele, B. C., Ma, Y., Osei, D., and Williams, R. J. Differences Between Sexes in Lower Extremity Alignment and Muscle Activation During Soccer Kick. *J Bone Joint Surg Am 92*, 11 (2010), 2050–2058.
5. Chang, S.-F. The Holy Grail of Content-Based Media Analysis. *IEEE Multimedia 9*, 2 (2002), 6–10.
6. Dorschky, E., Schuldhaus, D., Koerger, H., and Eskofier, B. M. A Framework for Early Event Detection for Wearable Systems. In *Proc ISWC 2015*, ACM (2015), 109–112.
7. Ekin, A., Tekalp., A. M., and Mehrotra, R. Automatic Soccer Video Analysis and Summarization. *IEEE Trans Image Process 12*, 7 (2003), 796–807.
8. Khan, A. M., Lee, Y.-K., Lee, S. Y., and Kim, T.-S. A Triaxial Accelerometer-Based Physical-Activity Recognition via Augmented-Signal Features and a Hierarchical Recognizer. *IEEE Trans Inf Technol Biomed 14*, 5 (2010), 1166–1172.
9. Lees, A., and Nolan, L. The biomechanics of soccer: A review. *J Sports Sci 16*, 3 (1998), 211–234.
10. Levanon, J., and Dapena, J. Comparison of the kinematics of the full-instep and pass kicks in soccer. *Med Sci Sports Exerc 30*, 6 (1998), 917—-927.
11. Nunome, H., Asai, T., Ikegami, Y., and Sakurai, S. Three-dimensional kinetic analysis of side-foot and instep soccer kicks. *Med Sci Sports Exerc 34*, 12 (2002), 2028–2036.
12. Nunome, H., Lake, M., Georgakis, A., and Stergioulas, L. K. Impact phase kinematics of instep kicking in soccer. *J Sports Sci 24*, 1 (2006), 11—-22.
13. Rui, Y., Gupta, A., and Acero, A. Automatically Extracting Highlights for TV Baseball Programs. In *Proc MULTIMEDIA 2000*, ACM (2000), 105–115.
14. Tjondronegoro, D. W., and Chen, Y.-P. P. Knowledge-Discounted Event Detection in Sports Video. *IEEE Trans Syst, Man, Cybern A, Syst, Humans 40*, 5 (2010), 1009–1024.