Signal Processing: *Image Communication* 35 (2015) 35–45

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Real-time event classification in field sport videos

Rafal Kapela [a,*], Aleksandra Świetlicka [a], Andrzej Rybarczyk [a], Krzysztof Kolanowski [a], Noel E. O'Connor [b]

[a] *Department of Computer Engineering, Poznan University of Technology, Poland*
[b] *INSIGHT: Centre for Sensor Web Technologies at Dublin City University, Dublin, Ireland*

ARTICLE INFO

Article history:
Received 3 January 2015
Received in revised form
14 April 2015
Accepted 15 April 2015
Available online 23 April 2015

Keywords:
Real-time sports event detection
Neural networks
State machines
Field sports

ABSTRACT

The paper presents a novel approach to real-time event detection in sports broadcasts. We present how the same underlying audio–visual feature extraction algorithm based on new global image descriptors is robust across a range of different sports alleviating the need to tailor it to a particular sport. In addition, we propose and evaluate three different classifiers in order to detect events using these features: a feed-forward neural network, an Elman neural network and a decision tree. Each is investigated and evaluated in terms of their usefulness for real-time event classification. We also propose a ground truth dataset together with an annotation technique for performance evaluation of each classifier useful to others interested in this problem.

1. Introduction

Sport is consistently highly rated in terms of television broadcasts [1,2] and in some countries, sports broadcast are the most watched broadcasts. This is true especially for significant sporting events like the Olympics or for the national/regional finals of the most popular sport in a given country. Across Europe, soccer usually in the center of attention. Based on publicly available statistics [3], one can observe that matches played in Germany's Bundesliga, the Premier League and Spain's La Liga are watched by over 10 million fans each year with a substantially larger audience watching at home on TV. However, soccer is not the only sport that enjoys significant popularity and large viewing figures. In Ireland, for example, soccer is considered to be in third position alongside rugby, after Gaelic football and hurling [4], the finals of which are guaranteed huge audiences both in the stadium and in front of the TV [5]. Considering other countries, we can add the following to the most popular field sports around the world: basketball, rugby, cricket, field and ice hockey or many others [6]. Depending on the country, the success of the local or national team and the time of year, sport can often be considered to be users' most desirable audio-visual information.

As a result, there has been significant interest in algorithms for automatic event detection in sports broadcasts. This is motivated by potential applications such as automatic highlight generation for summarization and second screen applications, indexing for search and retrieval in archives, mobile content delivery either off-line or as an added value in-stadium user experience. However, most event detection algorithms published thus far normally focus on a particular type of the sport (e.g., tennis, soccer, cricket, etc.) and are not robust for other types of sports, thereby limiting their applicability. Like for example event detection systems presented in [7–12] can work autonomously and some have ability to turn on themselves at specific time in order to analyze broadcasted video together with web-casting text. However, these systems suffer from the lack of flexibility that would allow it to analyze more than just one type of sport. This is a very

* Corresponding author.
E-mail address: rafal.kapela@put.poznan.pl (R. Kapela).

http://dx.doi.org/10.1016/j.image.2015.04.005
0923-5965/

good example of the state of the art in this field – although there are a plenty of examples that can be featured with high accuracy all of them work for only one type of sport. This is caused by the fact that different sports present different characteristics either in the rules for that sport of the manner in which it is captured and directed for broadcast. In addition real-time aspect is quite often neglected whereas in most application scenarios where a game is analyzed in order to provide rich content to the end users event extraction time should be one of the main parameters taken into account.

For this reason, in this paper we focus on a generic subset of all sports that can be designated as *field sports*, a term introduced in [13] to refer to any sport played on a grass pitch (soccer, rugby, field hockey, etc.) featuring two teams competing for territorial advantage. In this work, however, we extend this genre to include other sports that exhibit similar characteristics but that are not necessarily played on a grass pitch. Specifically, we extend the definition of field sports to include sports played in a playing arena that features some kind of scoring posts (e.g., goal post in soccer or basket in basketball), whereby the overall objective is territorial advancement with a view to obtaining a score.

Taking into account the diversity of the different field sports a range of event detection algorithms were presented in recent years. Even for one kind of sport the research can be conducted from different points of view. In [14,15] researchers pay their attention to the fact that low-level simple audio–visual features are often not rich enough to represent semantically complex information on the level appropriate to human perception. As a solution they propose multi-level multimodal descriptors related to the position of the camera in relation to the players and the field. The results presented by them are impressive (recall and precision on the level of about 90%) however they do not assume that their system to analyze video content in the real-time. It has been shown in [13] that about 97% of interesting moments during a game are followed by a close-up shot presenting a player who scored or who caused some interesting action. In addition, features like end of a pitch, audio activity or crowd shot detection has been shown to be very useful in event detection [13]. The presented system is proven to work with different field sports such as soccer, rugby, field hockey, hurling and Gaelic football. In this work a Support Vector Machine (SVM) was used as an event classifier. However, mainly because of the use of the Hough transform the implementation is very time consuming and inapplicable in real-time systems. A very similar approach is presented in [16]. In order to detect an event the authors declare the so-called “plays” where mainly a color histogram is calculated plus some heuristics are applied about the regions of histogram detection. An event is categorized using Bayesian Network based on the sequence of camera shots. In this work, events were detected in baseball, American football and Japanese sumo wrestling. Another example of work that belongs to this group is presented in [17] where, based on simple visual features like pitch orientation and close-up detection, the authors achieve good accuracy. However, again no time performance is given in the paper and there is a big drop in accuracy when the SVM is trained on the samples that do not belong to the same game. It is worth noting that the three approaches described above [13,16,17] are capable of extracting not only goals but also other exciting moments like penalties or close misses. In [18], very simple features like pixel/histogram change ratio between two consecutive frames, grass ratio and background mean and variation in addition to time and frequency domain audio features were used in order to detect events in soccer games. Although reporting high accuracy of the system using simple features the authors do not mention its time performance. Although the acceptance of the MPEG-7 standard in the community has been rather low, there are still approaches based on MPEG-7 descriptors. In [48] an event detection and tactics analysis is proposed. This kind of approach could be really useful for coaches and trainers for soccer game analysis after the game but from real-time analysis perspective it is not significantly interesting.

Taking the real-time approach for a given task into consideration the amount of the work is significantly lower. However, there are works worth recommending. In [19] authors use audio–visual features (Scale Invariant Feature Transform, Spatial-Temporal Interest Points, Mel frequency cepstrum coefficients, color moments, etc.) to detect events in Internet videos. The system is capable of working in real-time under an assumption that the interval between the frames for calculation is greater than 2 s. The drawback of the approach is in the precision which is on the level of about 50% for all the videos. A very interesting work is presented in [20] where authors present real-time video classification based on dense Histograms of Oriented Gradients/Optical Flow. Based on the results presented there the proposed system is capable of working at speed of almost 13 fps. The results however are presented only for 320×240 resolution short (70–200 frames) videos presenting only human actions. These assumptions are quite unrealistic for a wide range of different shots of the sport field, poses and numbers of the players in the shot.

Finally, there have been approaches significantly different from the ”standard” low level feature-based systems. In [21,22] very different approaches are taken. Both utilize the information produced by people during a game and tweeted by the popular Twitter website to detect events in different games (soccer and rugby were tested). They are, at first sight, universal approaches, however, they can suffer from quite large false positive detection rates, need constant connection to the Internet and introduce some ambiguity in the form of delay between detected and real events making the detection of event boundaries more difficult. The [23] approach uses knowledge-discounted approach to detect events. By introducing a hybrid approach which integrates statistics into logical rule-based models during event detection. It seems to be applicable for not only one type of sport but time performance of the system is not given in the paper.

Our contribution in this paper is to present a novel pseudo-generic real-time system for event detection that addresses many of the limitations of the techniques