Human Movement Science 41 (2015) 165–178

Contents lists available at ScienceDirect

Human Movement Science

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

Team activity recognition in Association Football using a Bag-of-Words-based method

Raúl Montoliu *, Raúl Martín-Félez, Joaquín Torres-Sospedra, Adolfo Martínez-Usó

Institute of New Imaging Technologies (INIT), Jaume I University, Castellón, Spain

ARTICLE INFO

Article history:
Available online 25 March 2015

PsycINFO classification:
3720

Keywords:
Team activity recognition
Sport analysis
Association Football
Bag-of-Words
Pattern recognition

ABSTRACT

In this paper, a new methodology is used to perform team activity recognition and analysis in Association Football. It is based on pattern recognition and machine learning techniques. In particular, a strategy based on the Bag-of-Words (BoW) technique is used to characterize short Football video clips that are used to explain the team's performance and to train advanced classifiers in automatic recognition of team activities. In addition to the neural network-based classifier, three more classifier families are tested: the *k*-Nearest Neighbor, the Support Vector Machine and the Random Forest. The results obtained show that the proposed methodology is able to explain the most common movements of a team and to perform the team activity recognition task with high accuracy when classifying three Football actions: Ball Possession, Quick Attack and Set Piece. Random Forest is the classifier obtaining the best classification results.

1. Introduction

The motion of individual professional sport players across the playing field is complex and depends on some factors including the main tactic, the game's context and some previously defined plays. Players train before matches in order to prepare specific plays, so that they can be perfectly executed.

* Corresponding author at: Institute of New Imaging Technologies, Jaume I University, Av. Sos Baynat s/n, E-12071 Castellón, Spain. Tel.: +34 964 728258.

E-mail addresses: montoliu@uji.es (R. Montoliu), martinr@uji.es (R. Martín-Félez), jtorres@uji.es (J. Torres-Sospedra), auso@uji.es (A. Martínez-Usó).

http://dx.doi.org/10.1016/j.humov.2015.03.007
0167-9457/

Team activity recognition is very useful for coaches since they can analyze the opponent team's plays and weaknesses and this information may be vital for preparing the next match's tactics. Nowadays, professional coaches perform this identification relying on manual video segmentation by human experts who simply store video fragments and annotate them. However, this procedure is time-consuming, tedious, expensive and it may produce errors due to the subjective nature of manual annotation. Recently, computer vision techniques have been successfully introduced in the team activity recognition research field (Barris & Button, 2008; D'Orazio & Leo, 2010) which allows this process to be automatized and non-subjective.

In this paper, we propose a methodological procedure based on machine learning that can certainly help to understand and analyze human movements in team sports, in line with other works such as Grunz, Memmert, and Perl (2012), Vucković et al. (2014), Díaz-Pereira, Gómez-Conde, Escalona, and Olivieri (2014). Moreover, the proposed approach opens up a new research line that directly competes with the traditional methodologies in this field that use tracking procedures. The proposed system learns how to recognize similar team behaviors in future matches from the previously labeled video fragments. For instance, the two first season's matches labeled by human experts could be used to train the automatic detection system during the third week of the league. The use of this methodology could replace the current manual process, which is time-consuming and prone to errors, with an automatic one. Broadcasters could also use this methodology to identify important highlights in real-time. In particular, coaches would be able to: more quickly identify the opponent team's strengths and weaknesses before the match, obtain the opponent's main tactics during a match, observe his/her team's performance during a match, and evaluate the players as a whole or individually immediately after the match, among other possible applications.

In the context of team activity recognition, most machine learning-based methods, such as (Blunsden, Fisher, & Andrade, 2006; Chen, Chou, Fu, Lee, & Lin, 2012; Direkoglu & O'Connor, 2012; Grunz et al., 2012; Hervieu, Bouthemy, & Cadre, 2009; Perl, Grunz, & Memmert, 2013; Perse, Kristan, Kovacic, Vuckovic, & Pers, 2009), estimate the players' positions and their trajectories using a tracking method. Despite the advances in tracking multiple objects in a scene (Yilmaz, Javed, & Shah, 2006), this task is still an open problem in complex scenarios (Barris & Button, 2008; Chen et al., 2012; Khokhar, Saleemi, & Shah, 2013), such as sport videos, due to a number of reasons: (i) players have almost the same appearance, (ii) they exhibit quick and unpredictable movements and changes of direction, (iii) there are multiple and frequent occlusions with other players, (iv) relative size of players can continuously vary, and (v) important lighting changes can appear in different parts of the playing field and during a match which could be misclassified as movement. All of these characteristics violate the assumptions on which computer tracking algorithms are typically based, so a different approach is needed to face the problem of automatically analyzing team tactics on sport videos.

Recently, the use of strategies that do not rely on tracking for general automatic video analysis has gained an increasing interest from the research community (Hospedales, Gong, & Xiang, 2012; Saleemi, Hartung, & Shah, 2010; Wang, Ma, & Grimson, 2009). Most of them use Bag-of-Words based (BoW) techniques that have been mainly developed to analyze traffic and surveillance related scenes. A first implementation of BoW within the sport science field was performed in (Rodríguez-Pérez & Montoliu, 2013). The BoW model is a representation used in natural language processing, information retrieval and computer vision fields, among others. In this model, a text is represented as the bag of its words, disregarding grammar and word order but keeping multiplicity. The BoW model is commonly used in text classification problems, where the frequency of occurrence of each word is used as a feature to feed a classifier. The BoW model has been adapted to other domains such as image and video classification, where words can refer to *visual* words or *motion words* as in the method proposed in this paper.

Several team pattern recognition methods have been developed (see Barris & Button, 2008; D'Orazio & Leo, 2010 for a review), most of them using Neural Networks (NNet) based methods (such as in Grunz et al., 2012; Memmert & Perl, 2009a, 2009b; Perl et al., 2013). Although, in general, NNet-based methods tend to obtain high performance, in the literature of pattern recognition and machine learning there are other classification methodologies than can improve their recognition results, such