Neurocomputing 150 (2015) 189–199

Contents lists available at ScienceDirect

Neurocomputing

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

ATM-based analysis and recognition of handball team activities

Raúl Montoliu*, Raúl Martín-Félez, Joaquín Torres-Sospedra, Sergio Rodríguez-Pérez

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

ARTICLE INFO

Article history:
Received 31 January 2014
Received in revised form
4 July 2014
Accepted 18 September 2014
Available online 23 October 2014

Keywords:
Team activity recognition
Team activity analysis
Topic modeling
Author Topic Model

ABSTRACT

In this paper, a new methodology based on the Author Topic Model (ATM) method is presented to perform team activity recognition and analysis in handball videos. Instead of using players' trajectories we just rely on low level features related to local motion, the evolution of which is then modeled over time by the ATM. The proposed methodology is applied to the task of recognizing four kinds of team activities in handball videos from the CVBASE'06 dataset and to analyze which are the most important elements of the activities. Our method is compared with two other ways of characterizing videos based on Bag-of-Words (BoW) and Latent Dirichlet Allocation (LDA) techniques. Our proposal obtains competitive results in terms of accuracy, computing time and interpretation of the results.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Analyzing sport videos for the purpose of team activity recognition using computer vision techniques is becoming an important task. Although individual players can behave differently depending on the game's context, their motion across the playing field is not random, and it is reasonable to expect that some common coordinated team activities can be perceived, especially when considering that they are usually trained in advance. Nowadays, professional coaches use simple and manual video segmentation applications to detect and store video fragments showing the tactics of their teams and those from opponent teams to be studied. However, relying on observation and manual annotation by sport experts is risky because they tend to be subjective when annotating data, and in addition, this is a time-consuming and tedious task that not all teams can afford because of the cost of qualified experts.

The methodology that we propose in this paper could be used, in the near future, to replace this manual process for an automatic one, where coaches would only have to label some initial video fragments (e.g. one or two matches at the beginning of the season) and the system would automatically learn information to automatically recognize similar team behaviors in future matches. One of the most important benefits of such an application would be the opportunity for coaches to automatically get a broader and faster analysis of a team's performance, either from their own team or from other opponents, being thus able to identify team's strengths, weaknesses, and mistakes related to tactics.

Most team activity analysis methods such as [1–4] use a fixed multi-camera system covering the playing field. This allows to obtain the position of all players and their trajectories by player detection and tracking. Despite the advances in tracking multiple objects in a scene [5], this task is still an open problem in complex scenarios [6], such as sport videos, where multiple occlusions co-occur, players have almost the same appearance and they could perform some unpredictable motion in some instant of the match. Recently, the use of video analysis strategies that do not rely on tracking is gaining an increasing interest from research community. One of the most important works in this line is the one proposed by Wang et al. in [7], which uses low level motion features and topic modeling methods to analyze traffic and surveillance related scenes. Topic models are powerful tools initially developed to characterize text documents but their use has been recently extended to other collection of discrete data in a number of different problems, e.g. to discover human and social routines in [8], to perform trajectory analysis and semantic region modeling on video scenes in [7], and to recognize human actions in [9], among others.

In this paper, a method to perform team activity recognition and analysis in sport videos is presented. In particular, we focus on handball videos taken from fixed cameras. Instead of using players' trajectories (such as [1–4]), we just rely on low level features related to local motion, the evolution of which is then modeled over time by the Author Topic Model (ATM) [10]. ATM is a generative model for documents that extends Latent Dirichlet Allocation (LDA) [11] to include authorship information. In ATM, each author is associated with a multinomial distribution over topics and each topic, like in LDA, is associated with a multinomial

* Corresponding author.
E-mail addresses: montoliu@uji.es (R. Montoliu), martinr@uji.es (R. Martín-Félez), jtorres@uji.es (J. Torres-Sospedra), serodrig@uji.es (S. Rodríguez-Pérez).

http://dx.doi.org/10.1016/j.neucom.2014.09.053
0925-2312/© 2014 Elsevier B.V. All rights reserved.

distribution over words. Whereas in LDA, topics are obtained by means of an unsupervised method, ATM uses the class information (i.e. authorship of the documents/type of team activity) when inferring the topics. Therefore ATM obtains the topics by means of a supervised method. As far as we know, this is the first paper that uses a supervised topic modeling technique to perform team activity recognition and analysis in sport videos.

In our method, for each video clip, local motion features (position and direction of the motion) are firstly obtained for the most prominent moving pixels. We call them *motion words*. Then, we count how many *motion words* belonging to a vocabulary are repeated in a video fragment. Using this information as input, the ATM method is applied to obtain a description of each author (type of team activity in our problem) in the topics feature space, that can be used for recognition purposes and for analyzing how the team activities are.

The proposed method based on ATM is applied to recognize four kinds of team activities in handball videos from the CVBASE'06 dataset[1] (*Fast Transition* (FT), *Slow Transition* (ST), *Passive Attack/Defense* (PA) and *Active Attack/Defense* (AA)). Other ways of characterizing sport videos are also applied to this recognition task, in particular, the Bag-of-Words (BoW) and LDA-based techniques. The effectiveness of these three characterization methods is assessed with a number of well-known classifiers: *k*-Nearest Neighbor (*k*NN) [12], Support Vector Machine (SVM) [13], Multilayer Perceptron (MLP) [14], Random Forest (RF) [15] and Logistic Regression (LR) [16] and their results are afterwards compared.

The main contributions of this work can be summarized as follows:

1. We propose the use of an ATM-based method to recognize team activities in handball videos from the CVBASE'06 dataset.
2. We present a comparison of the recognition results from our method with respect to those obtained when using the BoW and LDA-based strategies.
3. We analyze the semantic meaning of the results obtained by ATM when recognizing team activities in handball.
4. We study the role of some configuration parameters on the recognition process. In particular, the cell size γ when defining *motion words* and the number of latent topics T to be discovered in the ATM inference process (see Section 4).

A preliminary version of this work was published (as a conference paper) in [17]. The main difference is that, in this paper, an ATM-based methodology is used instead of the BoW-based and LDA-based ones as in the previous work. In addition, all these characterization methods have been assessed with a greater number of classifiers.

The rest of this paper is organized as follows. Related work is introduced in Section 2. Section 3 presents a background on BoW, LDA and ATM techniques. How features are extracted from videos is outlined in Section 4. We present and discuss the experimental results of the recognition task in Section 5. A semantic analysis of the results obtained by using ATM is presented in Section 6. Finally, Section 7 draws the main conclusions that have arisen from this work.

2. Related work

Sport video analysis has become an important research topic in the recent years because this might help to reveal critical information about the kind of sport, teams' and players' performance, etc. in an automatic fashion. The main issues faced by some of the related research works are as follows:

- *Categorizing videos according to the sport*: Some works (such as [18]) distinguish between video shots of a number of sports (running, basketball, football, hockey, swimming, tennis, etc.) based on the visual features of their key frames. Dynamic information is not used to this end.
- *Automatic generation of commentaries*: Another interesting application is mapping interesting game actions to previously recorded commentary concepts as proposed in [19], so that any virtual or real match could be automatically commented.
- *Highlight summarization*: Many researchers have devoted their efforts on automatically detecting highlights of a match, which refer to high-level semantic actions such as goals, saves, attacks, or fouls. Commonly, in broadcasting, highlights are repeated on the fly preceded by a transition showing an animated TV or tournament-belonging logo, so some methods [20,21] focus on recognizing those transitions as a means of detecting highlights. Alternatively, another method [22] detects highlights from audio clips with a big level of excitement. Independently of the case, detected shots are then matched with specific highlights depending on their content. Other works [23–28] do not rely on those clues to detect highlights but on visual information that is modeled over time.
- *Analysis of team tactics*: On the contrary of highlight summarization, which is mainly oriented to audience and broadcasters, professional coaches are more interested in identifying a team's strengths, weaknesses, and mistakes when executing tactics. This task requires annotating video fragments according to the tactic action or team activity that it shows, a time-consuming and tedious process that has been manually performed to this date. However, some methods [1–3,29,30] have recently appeared to help sport experts in such an ambitious objective, and this is the target application for our proposal.
- *Player identification and tracking*: This is often a core step to many of the aforementioned applications so that many works [1–3,29–33] address it as a mean to fulfill other objectives.

Our work can be classified in the Analysis of team tactics group.

Focusing on works aiming at analyzing team tactics, most methods are based on players' tracking as already mentioned before. To this end, they usually use a fixed multi-camera system [1–4] that covers the whole playing field so that they can roughly estimate the position of all players and their trajectories at any time. Some more complex methods [29–31,33] try to perform player's tracking in broadcast videos, a much more difficult task given that the camera is continuously moving and zooming only capturing part of the playing field. Tracking multiple players in a scene is a really challenging task as stated in [6] because of a number of reasons: (i) players have almost the same appearance, (ii) they exhibit quick and unpredictable movements and changes in direction, (iii) there exist frequent multiple 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 along a match. All 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 do not relying on tracking for general automatic video analysis is gaining an increasing interest from research community. One of the most important works in this line is the one proposed by Wang et al. in [7], which uses low level motion features and topic modeling methods to analyze traffic and surveillance related scenes. Similar approaches have

[1] CVBASE'06 Dataset - http://vision.fe.uni-lj.si/cvbase06/downloads.html

Download English Version:

https://daneshyari.com/en/article/412132

Download Persian Version:

https://daneshyari.com/article/412132

Daneshyari.com