Information Sciences 294 (2015) 64–77

Contents lists available at ScienceDirect

Information Sciences

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

An efficient pitch-by-pitch extraction algorithm through multimodal information

Kai-Lung Hua [a], Chao-Ting Lai [a], Chuang-Wen You [b,*], Wen-Huang Cheng [b,*]

[a] Dept. of CSIE, National Taiwan University of Science and Technology, Taipei, Taiwan
[b] Research Center for IT Innovation (CITI), Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form 6 July 2014
Accepted 3 September 2014
Available online 16 September 2014

Keywords:
Pitch-by-pitch events
Video understanding
Baseball videos
Multi-modal analysis

ABSTRACT

Sport video analysis facilitates the discovery of semantic structures in sport broadcast videos and enables a wide spectrum of applications. For example, coaches can analyze offensive and defensive plays performed during games to assess a team's capabilities. In general, identifying interested shots, e.g. pitch shots, from broadcast baseball videos requires great human labor to browse through those videos. In this work, we proposed a novel technique that automatically extracts pitch-by-pitch shots by recognizing the reliable emergence of pitching speed displayed on the scoreboard, estimating when and where the pitcher is present, and identifying the pitch shots based on the pitcher's motion degree. To validate the performance and accuracy of the proposed technique, we collected a dataset of baseball videos broadcasted in various countries. The experimental results verify that the proposed technique successfully extracts the desired pitch-by-pitch videos. Furthermore, it outperforms the state-of-the-art approach in terms of accuracy and time complexity.

1. Introduction

Sport video analysis [23,22,33,8,24,7,5,4,18,20,17] has been an active research field for a long time. It facilitates the discovery of semantic structures in recorded sport broadcast videos and enables a wide spectrum of applications. For example, to decide suitable strategies for the future games, coaches and managers have to carefully watch specific shots, e.g., pitching or batting shots in a baseball game, to study the strength and weaknesses of rival players. In general, identifying interested shots, such as pitch shots, from broadcast baseball videos requires great human labor to browse through those videos. This study proposes a novel technique to efficiently and accurately segment complete pitch shots through multimodal information [34,35,37,38].

Previous researchers, Chen et al. [4], designed a technique to extract *pitch-by-pitch* videos by first tracking ball trajectories to locate pitch shots in *per-batter* videos. Two types of videos were analyzed or extracted by Chen et al.:

- *Pitch-by-pitch video*: A video clip encompasses frames captured in a period between the starting time when pitchers take a ready stance on the pitching rubber and the ending time when the ball reaches the home plate.
- *Per-batter video*: A video clip includes frames between two adjacent batter change events [33], and it happens when the batter is either out (flies out, grounds out, and strike out) or credited with a base (on base hit or balls).

* Corresponding authors. Tel.: +886 2 27872327; fax: +886 2 27872315 (C.-W. You). Tel.: +886 2 27872356; fax: +886 2 27872357 (W.-H. Cheng).
E-mail addresses: cwyou@citi.sinica.edu.tw (C.-W. You), whcheng@citi.sinica.edu.tw (W.-H. Cheng).

http://dx.doi.org/10.1016/j.ins.2014.09.001
0020-0255/

Based on the identified pitch shots, Chen et al. [4] extracted pitch-by-pitch videos mainly by computing baseball trajectories. This technique would fail to extract pitch-by-pitch shots for at least following three instances: (1) ball trajectory is out of a predefined search region, (2) pitch replays, and (3) white-on-white problems [4]. Furthermore, the computational complexity of [4] is high because it requires a lot of time to compute ball trajectories. To overcome the above limitations and issues, in this work, we propose an efficient algorithm that detects pitch speed emerging in the scoreboard as an indication of ending time of a pitch. Afterward, we will utilize the motion degree information of a certain pitcher to determine the starting time of a pitch.

To display the game status to audiences on sport video, the scoreboard in general appears frequently in a fixed position. Usually, the important information, e.g., scores, current inning, and the number of balls, strikes and outs, will be updated when certain pitching or batting events happen. For example, the pitch speed emerges in the scoreboard once the radar gun detects the moving speed of the ball in the end of a pitch. Therefore, we propose a novel technique, which infers ending pitching events associated with pitch shots by recognizing pitch speeds appearing in the scoreboard and estimating when and where the pitcher is present. Based on the identified pitch shots and estimated pitcher presence information, the proposed technique searches starting and releasing events by measuring the motion degree of the pitcher, and generates a list of potential pitch shots that satisfy time-based pitching constraints [3] to filter out noisy non-pitch events. By incorporating rules observed from best practices of efficient pitching [14], the proposed technique finally groups nearby events as final pitch shots and then extracts pitch-by-pitch video clips based on the event distribution in each shot from the original broadcast video footage.

The major contribution of this study is to design, prototype, and evaluate a novel pitch-by-pitch video extraction technique by incorporating a pitch speed recognition scheme, a pitcher localization technique, and a motion degree measurement scheme to accurately identify pitch shots. To validate the performance and accuracy of this algorithm, we collected a dataset of per-batter videos broadcasted in various countries by randomly selecting segmented videos generated by the per-batter video segmentation module (described later in Section 3). The experimental results show that the proposed algorithm accurately extracts all pitch-by-pitch videos from testing data and keeps a small amount of pitching frames by dropping 84% of non-pitching frames of original videos. When comparing to the existing state-of-the-art approach [4], the proposed method not only extracts better pitch-by-pitch videos, but also significantly reduces the time complexity of 92% in average.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 gives an overview of the proposed algorithm that extracts pitch-by-pitch segments from baseball broadcast videos. Section 4 infers pitching events by detecting when the pitch speed appears, estimating when and where the pitcher is present, and analyzing the pitcher's motion degree. Section 5 filters out noisy non-pitching events and extracts final pitch-by-pitch videos based on inferred potential pitching events. Section 6 summarizes the experimental results and verifies the promising performance of this technique. Section 7 concludes and gives future research directions of this study.

2. Related work

Many existing studies have designed techniques to analyze broadcast sport videos, such as baseball [23,22,33,8,24,7,5,4], tennis [25,9], soccer [16,30], or basketball [32,31,12] videos, to extract interested sport events for coaches and managers to plan future game strategies. Among these studies, techniques for identifying significant baseball events or scenes from original broadcast baseball videos involve various topics, such as scoreboard localization [13,23,22], significant baseball event identification [33,8,15], pitch performance analysis [24,5], and pitch scene detection and extraction [4,7], which are summarized as follows.

To detect superimposed captions showing in the scoreboard, researchers have developed various techniques to accurately locate the scoreboard. Shih and Huang [22] detected the superimposed caption box by using the color and motion information in video contents and then explicitly constructed the corresponding caption templates. Guo et al. [13] proposed a matching based framework, in which matched pairs of SIFT points between successive video frames are first detected for finding candidate locations of a target scoreboard, and then the text recognition is performed to help determine the true one from the candidates. Su and Hsieh [23] located a scoreboard by detecting its captions in video frames directly. Based on the observations that the location of a scoreboard is often fixed and the intensity values of the scoreboard regions are stable, the temporal averaging techniques are applied to identify the caption and scoreboard regions.

Once the scoreboard is localized, significant baseball events can be identified. Zhang and Chang [33] retrieved the baseball game status, such as score, ball count, and out number, from the scoreboard. Then they applied the transition graph model on the retrieved information to detect two types of baseball events: last pitch and scoring. Chu et al. [8,19] proposed both rule-based and model-based decision methods to detect thirteen baseball events. They transformed baseball rules into a decision tree for the rule-based decision module, and utilized the context of shots for the model-based decision module. Hung and Hsieh [15] developed a baseball event detection system which integrates mid-level cues of scoreboard information and shot transition types to classify ten different events. The event decision rules are derived from inferred results of the Bayesian belief network.

Some techniques focus on analyzing pitch performance, e.g., detecting pitch type [24] or strike zone shaping and visualization [5]. Takahashi et al. [24] first computed the pitched ball candidates based on the features of color, size, and shape.