

野球ゲームにおける盛り上がり箇所の自動検出 —画像解析の一考察—

A Study of Image Analysis for Hot Spots Auto Detection In Baseball TV Game

赤木 信也*

Shinya Akagi

*NTTデータ先端技術株式会社, Email: akagis@intellilink.co.jp

はじめに

動画の盛り上がり箇所の検出は、ハイライト動画の作成に応用することができ、動画編集を手助けできる技術である。盛り上がりと言っても定義は様々であり、「草」や「w」といったコメントに注目したもの[1]、笑い声や相槌に注目したもの[2]が存在する。また、ハイライト動画の作成に関する研究として、音声や画像に着目したものがいくつか研究されている[3]。著者の過去研究[4]において、切り抜き動画を対象とした盛り上がり箇所の検出実験を実施したが、音声振幅やコメントを用いた方法では、盛り上がり箇所をピンポイントで推定できた結果を上位10箇所を含む場合が10件中1件程度しかなく、検出精度の低さが課題として挙げられていた。特にプロ野球の練習試合動画については、音声振幅を用いた方法では得点シーンやホームラン箇所の推定に失敗しており、個別に最適化された方法を試した方が良いという結果が得られていた。本研究では、過去研究で音声やコメントによる盛り上がり箇所の推定が難しいとされた野球に関する動画として、実際のプロ野球動画よりも形式的で解析しやすいプロ野球ゲームを対象とし、画像解析によって盛り上がり箇所（ピンチやチャンス、得点シーン、ホームラン）を自動検出することができないか分析した結果について報告する。

研究の技術領域

本研究の目的は、ハイライト動画の作成を支援または自動化することにある。本研究の技術領域は下記の通りいくつか分かれており、盛り上がり箇所の検出は1番目の技術領域である。

- (1) 盛り上がり箇所の検出
- (2) 場面の切り分け・結合
- (3) タイトル・説明文の作成

過去研究

野球などのスポーツに対するハイライト動画の作成方法としては、(1)カラーヒストグラム、カメラショットの種類、動作情報などの視覚的特徴、およびゼロクロスレート、周波数スペクトル、信号エネルギーレベルなどの音声特徴を利用する方法、(2)字幕を利用したり、スコアボックスの内容を光学文字認識を適用して利用したりする方法、(3)スローモーションリプレイを検出して利用する方法、(4)マルチモーダルのアプローチを利用する方法が研究されてきた[5]。マルチモーダルのアプローチの研究としては、GMMを用いた得点シーンに関する6種類のラベル付与について、音声だけの場合はF値が最大21.0%、動画像だけの場合はF値が最大35.8%、音声と動画像を組み合わせた場合はF値が最大45.5%となるという研究がある[6]。

本研究では、(1)の視覚的特徴を利用する方法をプロ野球ゲームに適用した結果について報告する。

プロ野球ゲームの特徴

研究対象とした実況パワフルプロ野球2022の特徴としては、投球・打撃時にストライク、ボール、アウトのカウント、ランナーの有無が固定の位置に表示されたり、得点時に固定の位置に総得点が表示されたり、ホームラン時にホームランの文字が表示されたりする。これらの視覚的特徴を利用し、画像解析を用いてピンチやチャンス、得点シーン、ホームランを自動検出する方法を検討した。

ピンチやチャンスの自動検出方法

野球におけるピンチやチャンスは、ランナーが2塁以上にいる状態を指す。ランナーの状態は、なし、1塁、2塁、3塁、1・2塁、1・3塁、2・3塁、満塁の8パターンであり、固定の位置に表示される。ただ、常に表示されている訳ではなく、打者が打ったり走者が盗塁したりした場合などには画面が切り替わり、一時的に非表示となる。パターンが少ないことおよび一時的な非表示にも対応できることを踏まえ、予めランナーの表示パターンを取得しておき、各表示パターンでテンプレートマッチングを実施してランナーの状態を分類する方法を採用した。

アウトカウントの自動検出方法

アウトカウントはヒッティングの結果を分析することに役立つ。場面の切り分け・結合に利用できると考えられる。アウトの状態は、0アウト、1アウト、2アウトの3パターンであり、固定の位置に表示される。こちらも常に表示されている訳ではなく、打者が打ったり走者が盗塁したりした場合などには画面が切り替わり、一時的に非表示となる。こちらもパターンが少ないことおよび一時的な非表示にも対応できることを踏まえ、予めアウトの表示パターンを取得しておき、各表示パターンでテンプレートマッチングを実施してアウトの状態を分類する方法を採用した。

得点シーンの自動検出方法

ホームランを除く得点シーンでは、得点時に固定の位置に総得点が表示される。具体的には「C 0 3回表 1 M」のような総得点表示がなされる。初めに、光学文字認識を用いる方法として文字認識ツールTesseractを試したが、簡易実験の結果、標準設定では認識精度が低く、処理時間も多すぎたため、不採用とした。次に、カラーヒストグラムを用いて得点シーンを検出できないか試したところ、閾値0.65を設定すると、再現率が0.5、適合率が1.00、F値が0.667となり、篠田らの過去研究の0.455を上回

る結果となったが、2値分類にもかかわらず再現率が高くなかったため、不採用とした。最後に、得点シーンについても、画像の共通部分として「回」の部分があり、テンプレートマッチングを適用できそうだったため、「回」の部分の画像を取得しておき、テンプレートマッチングを実施して得点シーンの検出を行う方法を採用した。

ホームランの自動検出方法

ホームラン時にはホームランという文字が表示される。注意点としては、ランナーの有無で表示内容と位置が多少変化し、ランナーがいなければホームランだけ表示されるが、ランナーが1人いれば2ランホームランと表示される。光学文字認識を用いる方法も考えたが、前述の通り、認識精度の低さなどがあったため、今回は予めホームランの文字画像を取得しておき、テンプレートマッチングを実施してホームランの検出を行う方法を採用した。

実験方法

実況パワフルプロ野球2022の対戦動画（1時間36秒）において、テンプレートマッチングの閾値調整によって、どの程度まで分類精度を高めることができるか分析した。本来は別動画を用意して、検証結果まで分析することが望ましいが、今回は学習段階までの分析としている。テンプレートマッチングの値がブラックアウト時に1になる問題については、未然に判定式を用意して正例として検出されないように対処した。

実験結果

各処理は並列で実行し、処理時間は7回平均で54分55秒かかった。特にホームランのテンプレートマッチングだけ最後まで終わらず処理に時間がかかっていた。各テンプレートマッチングについて、設定した閾値ごとの正解数、正解率、再現率、適合率の結果を表に示す。

閾値	正解数	正解率	再現率	適合率	F値
ランナー0: 0.80	841	0.996	0.986	0.996	0.991
ランナー0: 0.85	841	0.854	0.369	1.000	0.539
ランナー1: 0.80	357	0.999	1.000	0.994	0.997
ランナー1: 0.85	357	0.999	1.000	0.994	0.997
ランナー2: 0.80	140	1.000	1.000	1.000	1.000
ランナー2: 0.85	140	0.999	0.986	1.000	0.993
ランナー3: 0.80	9	1.000	1.000	1.000	1.000
ランナー3: 0.85	9	1.000	1.000	1.000	1.000
ランナー12: 0.80	239	0.999	1.000	0.988	0.994
ランナー12: 0.85	239	0.999	1.000	0.996	0.998
ランナー13: 0.80	101	0.999	1.000	0.990	0.995
ランナー13: 0.85	101	0.999	1.000	0.990	0.995
ランナー23: 0.80	164	0.999	1.000	0.994	0.997
ランナー23: 0.85	164	0.999	1.000	0.994	0.997
ランナー123: 0.80	0	1.000	0.000	0.000	
ランナー123: 0.85	0	1.000	0.000	0.000	
アウト0: 0.85	760	0.698	0.986	0.408	0.577
アウト0: 0.90	760	0.805	0.986	0.518	0.679

閾値	正解数	正解率	再現率	適合率	F値
アウト1: 0.85	652	0.877	0.991	0.595	0.744
アウト1: 0.90	652	0.998	0.991	0.998	0.994
アウト2: 0.85	439	0.824	1.000	0.407	0.579
アウト2: 0.90	439	1.000	1.000	1.000	1.000
得点: 0.95	60	0.999	0.983	1.000	0.991
得点: 0.80	60	0.999	0.983	1.000	0.991
得点: 0.50	60	1.000	1.000	1.000	1.000
ホームラン: 0.35	6	0.999	0.333	1.000	0.500
ホームラン: 0.30	6	0.999	0.5	1.000	0.667
ホームラン: 0.25	6	0.999	0.667	0.8	0.727

考察

ランナーのテンプレートマッチングでは、閾値0.80を設定すると、再現率が0.98~1.00、適合率が0.98~1.00と非常に高い値を示したため、十分な分類精度を確保できていると言える。アウトカウントのテンプレートマッチングでは、閾値0.90を設定すると、再現率が0.98~1.00と非常に高い値を示し、適合率も0.50~1.00と高い値を示したため、十分な分類精度を確保できていると言える。False Positiveについて、他のテンプレートマッチングでTrue Positiveとなっている場合があり、テンプレートマッチングの値が高い方を採用することで適合率に関する精度の向上が期待できると考える。

得点シーンのテンプレートマッチングでは、閾値0.50を設定すると、再現率1.00、適合率1.00となり、完璧に分類できており、十分な分類精度を確保できていると言える。

ホームランのテンプレートマッチングでは、閾値0.30を設定すると、動画内に2箇所あるホームラン場面を両方検出することができており、適合率も1.00なので、十分な分類精度を確保できていると言える。

まとめ

テンプレートマッチングを利用することで、ランナー、アウトカウント、得点シーン、ホームランを高い精度で分類できることが示された。

おわりに

本研究で利用したプログラムなどはGithub[7]およびGoogleドライブ[8]に公開している。

参考文献

- [1] tf0101: YouTube Liveの生放送から盛り上がった箇所を自動抽出するCLIを作った話, <https://qiita.com/tf0101/items/efb2484a0b5b1cdc8291>, 2019-07-08
- [2] 河原達也, 須見康平, 緒方淳, 後藤真孝: 音声会話コンテンツにおける聴衆の反応に基づく音響イベントとホットスポットの検出, 情報処理学会論文誌 52 (12), 3363-3373, 2011-12-15
- [3] 望月貴裕: 映像自動要約技術の最新動向, NHK技研R&D2020年夏号, 4-15, 2020-08-15
- [4] 赤木信也: 盛り上がり検出のための音声解析の一考察, 情報処理学会FIT2023, 2023-09-07
- [5] Po-Chyi Su, Chi-Heng Lan, Chin-Song Wu, Zi-Xin Zeng, Wei-Yu Chen: Transition effect detection for extracting highlights in baseball videos, EURASIP Journal on Image and Video Processing volume 2013, 27, 2013-05-04
- [6] Koichi Shinoda, Kazuki Ishihara, Sadaaki Furui, Takahiro Mochizuki: Automatic Score Scene Detection for Baseball Video, Large-scale knowledge resources: construction and application, LKR 2008, pp. 226-240, 2008-03.
- [7] Github: create_baseball_highlightリポジトリ, https://github.com/ShinyaAkagil/create_baseball_highlight
- [8] Googleドライブ: output.mp4, <https://drive.google.com/file/d/1sJ0ygXSMOhBk9mCsJXWAqBgZ7Zq-A-j/view?usp=sharing>