

サッカー動画からのハイライトシーンの抽出

笹原 規睦 花泉 弘

法政大学情報科学部

1. まえがき

近年、マルチメディアコンテンツが充実し、誰もが手軽に利用できる環境になりつつある。そのような環境の中で、ユーザーにとって数あるサッカー動画の中から自分が見たい動画を選ぶのは容易ではない。サッカーの動画を選ぶ際に、映像データの重要なシーンをハイライトシーンとして抽出しておけば、時間をかけずに選択できて便利である。ハイライトシーンとは試合中の重要なシーンであり、ゴール、シュート（ゴールでない）、ファウルの3つに大別できる。サッカー動画の中からあらかじめハイライトシーンを自動的に抽出しておき、ユーザーが見たいハイライトシーンをいつでも素早く簡単に選択できるようなシステムを構築することをこの研究の目的としている。

従来研究では Play-Break（試合が動いている時のシーンと止まっている時のシーンの組み合わせ）でハイライトシーンの抽出を行っていた[1][2]。試合映像において、リプレイシーンは重要なハイライトシーンの直後に表示され、視聴者に何が起きたのかを知らせているので、本研究においてもリプレイシーンに着目し、そのリプレイがどのような意味を持つのかを分類することで、ハイライトシーンとして抽出する。リプレイシーンの抽出には、その直前と直後にロゴが表示されることを利用する。試合映像中からロゴを検出する手法として、コントラストを用いて検出する手法[1]が提案されているが、検出できるロゴに限られていた。本研究では、ロゴが表示される位置が常に同じであることを利用して、フレーム内の少数の領域に注目した手法を提案する。また、ゴール付近であるかどうかの判定においても従来研究[1]ではコートラインを検出する方法を用いていたが、さまざまなサッカー動画に対応するのが困難なため、本研究では従来手法[3]を改良した新しいゴールポスト検出法を提案する。

2. 原理と処理手順

入力されたサッカー動画をフレームに分解し、そこから領域探索によってロゴが表示されているフレームを検出し、ロゴの間のフレームをリプレイシーンとして検出する。そのリプレイシーンの継続時間、フレーム内にグラウンド（芝）がどれだけ映っているか、ゴールポスト映っているか、スコアテキストが変化しているかどうかの特徴を抽出する。ゴールポストはリプレイシーン内で映っていないこともあるため、リプレイシーンの直前のロゴから一定時間（450フレーム分）遡って検出を行う。ゴールとシュートの決定的な違いとしてスコアテキストの変化が挙げられる。リプレイシーンの開始の450フレーム前と終了の200フレーム後のスコアテキストを比べ、変化の有無を調べる。抽出した特徴を基に物理的指標で

リプレイシーンを分類していく。分類したリプレイシーンとその直前の試合が動いているシーンを組み合わせてハイライトシーンとして出力する。

2.1. ロゴの検出

ここでは、特定のサッカーリーグに限定しないロゴ検出を目指す。各ロゴが表示される位置は決まっていることを利用して検出を行う。各々のロゴを特徴づける領域を3点ずつ選び、テンプレートで指定した位置での画素値が同じになることでそのロゴを検出する。テンプレートは手動で作成し、ノイズの影響を避けるために3x3の領域内の画素を比較に用いる。ロゴが出現していても画素値がテンプレートと完全に一致しないことがあるため、RGBすべての画素値の平均の差が閾値以下の場合、領域が一致したと判定する。3箇所すべてで一致していると判定されたときにロゴが検出されたとする。

2.2. グラウンド領域の検出

シュートやゴールのリプレイシーンではリプレイシーンの前半部分に表示されるため、リプレイシーンの始めの20フレームでグラウンド領域の検出を行う。サッカー動画では主にグラウンドの芝が映ることに基いて検出を行う。サッカー動画全体からランダムに100枚のフレームを抜き出し、フレーム毎にランダムに10個の点を選ぶ。サッカー動画では、フレームの上部分に観客席が映ることが多いため、点を選ぶ際はフレームの下半分から選ぶ。全1000個の点の色相値を取得し、60~180の範囲で属する画素数が30個以上の値を芝の色相値とする。

2.3. ゴールポストの検出

ゴールポストは細長くどこから撮影してもほぼ垂直で、必ず白いという特徴を基に検出を行う。フレームをグレースケールに変換し、図1のような検出マスクを用いる。

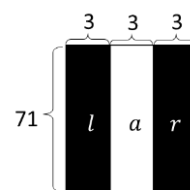


図1 検出マスク

検出マスクをフレームの左上から1ピクセルずつずらし、式(2)を用いてdの数値を調べ、マスク処理を行った。

$$d = \frac{1}{213} \left\{ \sum_{i=0}^2 \sum_{j=0}^{70} a(i,j) - 0.3 \times \sum_{i=0}^2 \sum_{j=0}^{70} l(i,j) - 0.3 \times \sum_{i=0}^2 \sum_{j=0}^{70} r(i,j) \right\} \quad (1)$$

得られたdの値のうち、第1ピークと第2ピークをゴールポストとして検出する。

2.4. スコアテキスト

ゲーム中にゴールが決まればスコアテキストが変化するため、その変化を認識する。各国のサッカーリーグではスコアテキストが表示される座標は異なる。ロゴを検出

Extraction of Highlight Scenes from Football Videos

Noritomo SASAHARA, Hosei University

Hiroshi HANAIZUMI, Hosei University

した段階で、どの国のサッカーリーグなのかを知ることができるため、それを利用してスコアテキストの変化をみていく。

前述したとおり、リプレイシーンの 450 フレーム前と 200 フレーム後のスコアテキストを比べていく。スコアテキストをグレースケールに変換し、それぞれ表 2 の閾値で二値化して処理を行う。式(2)を用いてスコアテキストの変化を観察する。

$$s = \frac{d}{m} \quad (2)$$

類似度を s とおく。 d は画素値が異なる画素数で m は画素値が一致している画素数である。閾値を T とすると、

$s > T$ の場合にスコアが変化したとする。

2.5. シーン分類の実験結果の考察

実験で使用する動画は DAZN で 2017 年 4-5 月に録画 (bandicam でキャプチャ) したものである。動画をフレームに分解し、実験を行った。フレームレートは 30fps、フレームのサイズは 1280x720 ピクセルである。128 個のリプレイシーンからゴール、シュート、ファウルの特徴と閾値を得る。図 2 従来研究[1]を基にしたそれぞれのハイライトシーンのレーダーチャートを示す。数値の単位は継続時間とゴールポストがフレーム数、芝領域がピクセル数、スコアテキストが式(1)で求めた類似度の数値である。グラウンド領域はリプレイシーン中の平均を求めている。リプレイシーンの直前の 450 フレームすべてでゴールポストを検出すると膨大な時間がかかってしまうため、フレームレートを下げて検出を行っている。ゴール、シュート、ファウルの変数を G, S, F とおく。継続時間、グラウンド領域、ゴールポスト、スコアテキストの得られた数値をそれぞれ t, g, gp, st とし、それらの数値によって変数に 1 を足していく。表 1 に抽出した各特徴の閾値を示す。シーン分類では特徴抽出に用いた動画とは別の動画を使用した。表 2 にシーン分類の結果を示す。

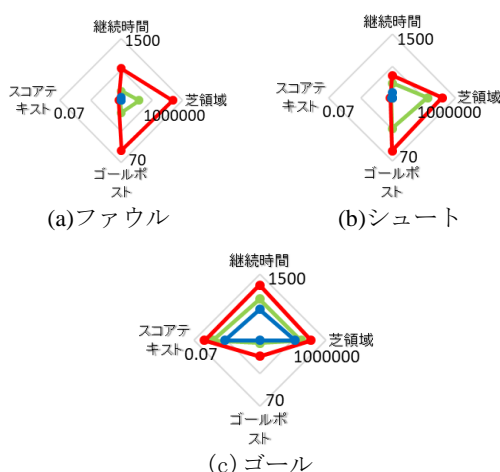


図 2 特徴のレーダーチャート

ゴールの分類において適合率 100%と非常に高い精度で分類することに成功した。

シュートをファウルと誤分類してしまった原因として、芝領域を正しく検出できていなかったことが挙げられる。リプレイシーン内でカメラが切り替わると芝の色相値も

変化するため、設定した芝色相と異なる色相値になってしまっていた。さらに柔軟性を意識した検出方法を検討する必要がある。

ファウルをシュートと誤分類してしまった 2 つのリプレイシーンはオフサイドのリプレイシーンである。オフサイドのリプレイシーンは非常にシュートシーンに近く、オフサイドのリプレイシーンをファウルと定義したため、誤分類が起きたと考えられる。さらなる特徴の追加を行うことで解決できると考える。

表 1 抽出した特徴の閾値

| シーン | 継続時間 : t | グラウンド領域 : g | ゴールポスト : gp | スコアテキスト : st |
|----------|--------------------|-----------------|---------------|----------------|
| ゴール : G | $t \geq 600$ | $g \geq 500000$ | — | $st \geq 0.01$ |
| シュート : S | $250 \leq t < 600$ | $g \geq 500000$ | $gp \geq 30$ | $st < 0.01$ |
| ファウル : F | $t < 250$ | $g < 500000$ | $gp < 30$ | $st < 0.01$ |

表 2 シーン分類の結果

| 分類結果 | ゴール | シュート | ファウル | 適合率 (%) |
|------|-----|------|------|---------|
| 正解 | | | | |
| ゴール | 6 | 0 | 0 | 100 |
| シュート | 0 | 6 | 2 | 75 |
| ファウル | 0 | 2 | 12 | 85.7 |

4. むすび

本研究では特定の領域に注目する方法でロゴを検出し、そこから作成したリプレイシーンを利用し、継続時間やグラウンド領域、ゴールポスト、スコアテキストの特徴を用いたサッカー動画のハイライトシーン分類手法を提案した。実験よりゴールシーンはスコアテキストに大きく影響を与えるため、正確に分類することができた。この手法は他のスポーツでも利用可能と考える。他のシーンもここで示した 4 つの特徴でハイライトシーンを分類できると考える。しかし、いくつかの問題点も見つかった。ロゴの未検出を防ぐために、フレームレートと画質の安定させる工夫が必要である。ゴールポストの検出をする際、処理が複雑なため時間がかかりすぎてしまった。処理をさらに効率化し、時間を短縮することが今後の課題の一つとして考えられる。シュートやファウルの誤分類においては、分類数や特徴の不足が問題点として挙げられる。今後、より詳細な分類と増やすのと同時に音などのさらなる特徴の追加でより正確にハイライトシーンを分類できると考える。

文献

- [1] Dian Tjondronegoro, Yi-Ping Phoebe Chan, and Binh Pham, The Power of Play-Break for Automatic Detection and Browsing of Self-consumable Sport Video Highlights, Mir 04 Proceedings of Acm Sigmm International Workshop on Multimedia Information Retrieval, pp. 267 – 274, 2004.
- [2] Tjondronegoro, D., Y.-P.P. Chen, and B. Pham. Sports video summarization using highlights and play-breaks. in The fifth ACM SIGMM International Workshop on Workshop on Multimedia Information Retrieval. 2003. Berkeley, USA: ACM.
- [3] 佐藤寛晃, 堀口文男: “サッカー映像におけるゴールポスト検出”, 情報処理学会第 71 回全国大会, PP. 2-397 ~ 2-398, (2009-03) .