画像と音声の特徴を用いたサッカー中継の自動要約生成手法

2019SC010 長谷川真士
指導教員：河野浩之

1　はじめに

NHK 放送技術研究所 [3] によると，近年映像の視聴スタイルの多様化が進んでおり，はじめから映像の全編を見るのではなく，要約映像を見て映像の大まかな内容を把握するといったスタイルが増えてきている．このような流れ中で，ソーシャルメディアなどを利用して要約映像を配信するサービスの必要性が高まっており，さまざまなジャンルの放送映像を対象とした自動要約技術の確立が求められている．自動要約技術が必要となる場面はいくつかあるが，スポーツ，特にサッカーにおいては，世界各地にプロリーグがあり，1 つのリーグに 20 チーム近くが所属しているため，様々なチームの試合内容を簡単に把握するためにはハイライト映像の利用が必須であると考えられる．

本研究では，サッカー中継の画像と音声を用いたハイライト生成手法を提案する．

2　サッカー中継映像自動要約の先行研究

サッカー中継のハイライトを生成するための主な先行研究の内容と，その結果をまとめたものを表 1 に示す．

Chakradhar ら [1] の研究では，スコアボードをオブジェクト検出アルゴリズムの YOLO を用いて検出し，得点の変化を読み取ることで得点シーンを特定することでスポーツ映像の自動キーイベント抽出を行っている．この研究ではゴールシーンにおいては様々なリーグの試合に対し，平均 0.979 の F 値を獲得することができた．しかし，得点シーン以外の検出については検討が行われていないため，十分な要約映像が作成できるとは言い難い．

Muhammad ら [2] の研究では，経験的モード分解 (EMD) を用いて中継映像から実況者の声，審判の笛の音，観客の歓声などを抽出し，抽出された音声特徴量を用いてゴールイベントの検出を行っている．この研究では 3 試合においてゴールイベントを 96% の精度で検出することができ，レッドカードや PK においては全て検出することができた．しかしシュートの検出率は 40 %にとどまっており，決定機を高い精度で検出できるとは言い難い．

3　ハイライト生成のための提案手法

本研究では，映像中の画像と音声を分析することによって重要なシーンや盛り上がったシーンを特定し，サッカー中継からハイライト映像を自動生成する手法を提案する．画像の分析では重要なシーンの後にはリプレイ映像が流れるという性質を利用し，リプレイ前後に現れる一定のロゴを検出することでリプレイシーンの特定を行う．音声の分析では音声強度を取得することで盛り上がったシーンのみを抽出する．提案手法の概要を図 1 に示す．

表 1　サッカー中継自動要約に関する研究

	概要	結果
Chakradhar ら [1]	スコアボードを YOLO を用いて検出し，スコアボードの変化からゴールシーンを特定する	ゴールシーンの検出で平均 0.979 の F 値を獲得
Muhammad ら [2]	経験的モード分解を用いてサッカーの試合映像から音声特徴を抽出し，重要イベントの検出を行う	ゴールイベントを 96% の精度で検出

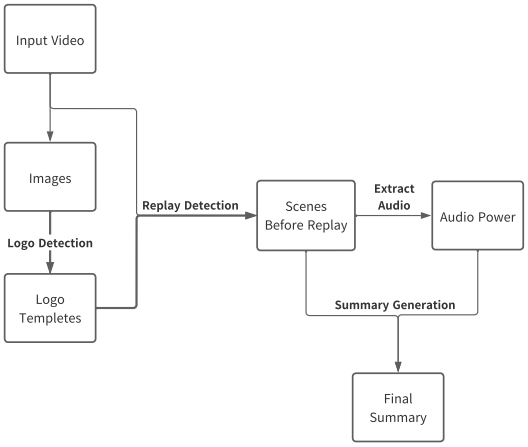

図 1　提案手法の概要

3.1　リプレイシーンの取得方法

本研究ではリプレイシーンの特定を，リプレイの前後に出現するロゴを検出することで実現する．検出には，Python で動作し，画像処理に必要な機能をまとめたライブラリである OpenCV を用いる．OpenCV には画像の類似度を計算するための機能がいくつか備わっており，動画の中から用意したロゴ画像と類似したフレームを特定することでリプレイシーンの特定を行う．今回検出するものは形やサイズ，角度などが一定のロゴであるため，テンプレートマッチングを利用して類似度を算出する．

テンプレートマッチングは入力画像にテンプレート画像を少しずつずらしながら重ね合わせ，類似度を計算していく手法で，類似度の評価指標として，今回類似度が-1 から 1 で表される NCC（Normalized Cross-Correlation）を用いる．NCC はテンプレート画像と探索対象画像における画素の濃度の相関係数を求めることで類似度を計算するも

ので，以下の式で求められる．

$$R(x,y) = \frac{\sum_{x',y'}(T(x',y') - I(x+x',y+y'))^2}{\sqrt{\sum_{x',y'} T(x',y')^2 \cdot \sum_{x',y'} I(x+x',y+y')^2}}$$

T はテンプレート画像を，I は入力画像を，R は結果をそれぞれ表している．また，入力画像の座標を x, y で表し，テンプレート画像の座標を x', y' で表している．画像が類似しているほど値は 1 に近づく．

テンプレートマッチングを行うためにはロゴのテンプレート画像を用意する必要があるが，手動で用意していては手間がかかってしまうため，本研究では映像の中から自動でロゴを特定し，テンプレート画像として用いる．ロゴ出現シーンは人間が映りこまないという性質を利用し，物体検出アルゴリズムを用いて人間の検出を行い，人間が映りこまなかったシーンをロゴ出現シーンであると推測する．

3.2 重要シーンの抽出方法

リプレイシーンからより重要なシーンを抽出するために，リプレイ直前の映像の音声強度を取得する．音声強度の取得方法については，リプレイ直前のシーンの音声波形から振幅の最大値を取得することにより盛り上がりを測定する．リプレイ直前のシーンごとに音声強度を取得出来たら，それらの値の平均値を求め,，音声強度が平均を上回るシーンのみを重要シーンとして抽出する．また，サッカーのリプレイ映像は得点シーンにおいてはリプレイ時間が顕著に長くなるといる特徴があり，それとは反対に，ファウルのリプレイ時間は短くなるといった特徴がある．このことから重要度が高いシーンほどリプレイ時間は長くなると推測し，リプレイ時間が平均を上回るものも重要シーンとして抽出する．

4 重要シーン抽出結果

DAZN や SPOTVNOW などのストリーミングサービスや，FIFA や各クラブがアップロードしているハイライト映像では，以下の 3 つのシーンがハイライトシーンとして用いられている．

- 得点シーン
- チャンス，決定機
- 重大なファウル (PK，レッドカード)

よってこれらのシーンを重要シーンとして定義し，これらのシーンの検出精度を検証する．

実験環境を表 2 に示す．人物検出のために YOLO の中でも比較的新しく動作の安定が保障されている YOLOv5 を用い，学習モデルは GitHub 上で公開されている人物検出に特化したモデルを用いる．

上記の環境で重要シーンの抽出を行った結果，表 3 のようなシーンが得られた．ゴールと決定機においては，実際の回数と比較を行い，検出精度を算出している．決定機の精度算出方法については，FIFA やクラブの公式チャンネルがアップロードしているハイライト動画に載っている，ゴール以外のチャンスシーンを決定機と定義し，抽出された割合を算出している．ゴールは 100%，決定機についても 80% の精度で検出できたが，今回実験対象の試合ではレッドカードや PK のシーンは見られず，精度は未知数となった．

表 2 実験環境

OS	Windows10
実行環境	Google Colaboratry
プログラミング言語	Python 3.8.8
ライブラリ	YOLOv5
学習モデル	crowd-human[4]

表 3 提案手法を用いて抽出されたシーン

	シーン数	ゴール	決定機	ファウル
試合 1	14	3/3	3/3	0
試合 2	9	2/2	2/3	1
試合 3	12	4/4	3/4	0
精度	-	100%	80%	-

5 まとめ

先行研究ではゴールイベントを高い精度で検出することができるが，それ以外のシュートやチャンスのシーンについては課題が残っていた．しかし本研究の手法を用いれば，得点シーンはもちろん，実際にハイライト映像で用いられるような決定機のシーンにおいても高い精度で検出をすることができた．今後の課題としては，レッドカードや PK といったシーンに対する精度も確認することと，サッカー以外のスポーツにおいても本研究の手法を利用できるか検討することなどが挙げられる．

参考文献

[1] Chakradhar Guntuboina, Aditya Porwal, Preet Jai, Hansa Shingrakhia,“Deep learning based automated sports video summarization using YOLO,” Electronic Letters on Computer Vision and Image Analysis, pp99-116, 2021.

[2] Muhammad Rafiqul Islam, Manoranjan Paul, Michael Antolovich, Ashad Kabir, “Sports Highlights Generation using Decomposed Audio Informatio,” 2019 IEEE International Conference on Multimedia & Expo Workshops, pp.579-584, 2019.

[3] 望月 貴裕, “映像自動要約技術の最新動向,” NHK 放送技術研究所, NHK 技研 R&D 2020 年 夏号, https://www.nhk.or.jp/strl/publica/rd/182/2.html

[4] https://github.com/deepakcrk/yolov5-crowdhuman