5ZK-04

音響特徴に注目したゲームを対象とする切り抜き動画生成手法

吉田　大毅,　小河　誠巳
東京電機大学理工学部情報システムデザイン学系

1　はじめに

近年、動画共有サービスにおいて長時間の配信活動が人気を集めている。配信活動では投稿者が視聴者とコメント機能を介して会話をしたり、投稿者がリアルタイムでゲームのプレイ画面を配信することが多々ある。中でも配信の動画要約として、切り抜き動画が注目を集めている。ここで言う切り抜き動画とは、長時間の動画の一部を切り取る編集を行い再投稿した動画作品であり、視聴者が気軽に本編動画の要点を把握することができるため人気を集めている。動画要約はスポーツやニュース、ドラマなど様々な映像コンテンツを対象として、ジャンルに応じた音響特徴や画像特徴に基づいて研究が多くされてきた [1]。しかしゲームを対象とした動画要約技術の研究は十分に行われておらず、ゲームの動画に適した動画要約ができない。そこで本稿ではゲーム動画、特に近年人気を集めているゲーム実況動画を対象とした動画要約を生成することを目的として、ゲーム動画の音響特徴に注目した手法を提案しその結果を述べる。

2　関連研究

動画要約に関する研究は従来から行われており、様々な動画に対しそれぞれの特徴を利用した手法が提案されてきた。スポーツ動画に関する動画要約では Tjondronegoro らはホイッスル音や観客の盛り上がりからハイライトシーンを検出する事で動画を要約する手法を提案している [2]。また歌謡番組を対象として番組の特徴をとらえた構造を定義し、それに基づいたシーンの分割を行うことで様々な歌謡番組に汎用な要約手法が提案されている [3]。

3　動画データ

使用する動画データは、任天堂の「大乱闘スマッシュブラザーズ」の対戦実況動画である。8〜10 人のプレイヤーが順番に 1 対 1 で対戦を行い、それを観戦する 2 人の実況音声とゲーム画面が収録されいている。対戦形式はストック制を採用している。ストック制とは、ストック（残機）を先に全て失ったプレイヤーが負けとなる形式である。なお開発元の任天堂と動画データの提供者からは「研究発表の範囲内ではゲーム音声・画像等の利用は問題ない」と回答を得ている。動画は全部で 8 本あり、1 本あたり 2 時間 30 分程度の長さである。

4　提案手法による切り抜き動画の生成

4.1　提案手法概要

ゲーム実況の切り抜き動画では、動画内で何かしらのイベントが発生した場面を切り抜くことが多い。本研究で使用した動画の場合、勝敗に直結するストックの変動がイベント発生時であると考えた。そしてゲーム実況では、実況者が盛り上がりをみせた前後にプレイヤーの巧みな動きや勝敗に関与する場面が多く存在する。そこでイベント発生時の前後を切り抜くために、動画に対して 1 分単位の分割を行いそれぞれの区間の音圧レベルの最大値を比較することで、切り抜き動画を生成する手法を提案する。

4.2　音響解析

本研究では音響特徴をもとに切り抜き動画の生成を行うため、”FFmpeg”を用いて動画データを WAVE オーディオファイルへ変換した。オーディオの処理と解析をするため Python パッケージの”librosa”を使用した。1 秒間に 100 回サンプルの取得を行い、RMS（Root Mean Square）から音圧レベル（dB）を求めた。以下の（1)、(2）はそれぞれ、データから取得した音圧 x のサンプル x_t（t=1,2,...n）に対する x の RMS の定義式と、取得した音圧 P_1 に対して基準音圧 P_0 を用いた音圧レベル L_p を求める定義式である。

$$RMS = \sqrt{\frac{1}{n}\sum_{t=0}^{n-1} x_t^2} \tag{1}$$

$$L_p = 20\log_{10}\frac{P_1}{P_0} \tag{2}$$

Copyright ©2023 Information Processing Society of Japan.
All Rights Reserved.

1本の動画データで求めた音圧レベルを縦軸、時間を横軸とした図を以下に示す（図1）。縦軸が大きいほど音圧レベルが高く動画内でより大きな音が発生していることを示している。

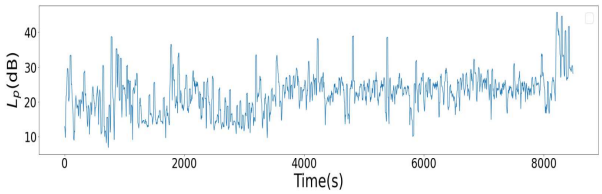

図 1: 動画の音圧レベルの推移

また、この動画データを約1分ごとに分割し、各区間で得られた音圧レベルの最大値の抽出をおこなった。これらの音圧レベルの上位10件の動画を以降ハイライトと呼ぶ。ハイライトの音圧レベル（dB）と観測した時間をまとめた表を以下に示す（表1）。

表 1: 音圧得 r ベルと観測時間（10 件）

観測時間（min）	音圧レベル（dB）
11.9	38.754692
13.1	38.526985
69.3	38.150497
70.5	37.831417
80.0	38.964779
89.6	38.520908
136.2	45.746655
137.3	44.599380
138.6	40.457054
139.8	41.692120

得られたハイライト10件を結合することで、1本の動画に対し10分の切り抜き動画を生成した。

5　評価

4.1で述べたよにうに、イベント発生時の前後の場面をハイライトとして切り抜くことを目指して実験を行った。1本の動画に対し10件あるハイライトの中で、何件が目標のハイライトであるかを評価基準とする。8本の動画データそれぞれの評価をまとめた表を示す（表2）。

6　考察

8本の内、7本の動画データでは適切なハイライト件数が6件以上となった。しかし、その内の3本の動画ではハイライトの生成箇所が極端に元動画の1か所に集中した。適切なハイライトが出来なかった要因としてマイクと話者の距離があげられる。今回使用した動画ではマイクと実況者の距離がかなり不安定であった。また、特定の人の声が他の人に比べて大きな音で入力されていることがあった。そのため特に盛り上がりのない会話や、特定の人が話しているときなど偏った切り抜き動画が生成されたと考える。

表 2: 各動画データの評価

動画ナンバー	目標ハイライト件数（10 件中）
No.1	1 件
No.2	9 件
No.3	10 件
No.4	6 件
No.5	6 件
No.6	7 件
No.7	7 件
No.8	7 件

7　まとめ

ゲーム実況動画を対象とし勝敗に関与する場面などを切り抜くため、実況者の音声に注目した切り抜き動画の生成を行った。結果として目標の場面を切り抜くことができたが、精度が不安定であることや切り抜く箇所に偏りができてしまった。本研究では使用しなかった音響特徴量や、映像情報の視点も導入することで安定性を図ることができるかもしれない。また、ゲームにも数多くのジャンルがあり全てのゲーム実況を対象とすることが困難である。より汎用性を高めることが今後の課題であると考える。

参考文献

[1] 滝嶋 康弘: 知っておきたいキーワード～映像の自動要約技術～, 映像情報メディア学会誌, Vol.62, No.5, pp.714 716(2008).

[2] Tjondronegoro, D., Chen, Y.-P. P. and Pham, B.: Integrating highlights for more complete sports video summarization, IEEE multimedia, Vol.11, No.4, pp.22–37(2004).

[3] 吉田 壮, 小川 貴弘, 長谷山美紀: 歌謡番組における映像の構造に注目したシーン分割手法, 電子情報通信学会論文誌, Vol.J97-D, No.7(2014/7)

Copyright ©2023 Information Processing Society of Japan.
All Rights Reserved.