

自動ダイジェスト再生機能のためのリプレイシーン検出手法 Replay Scene Detection for Automatic Video Digesting

親松 昌幸[†]
Masayuki Oyamatsu

廣井 和重[†]
Kazushige Hiroi

森 靖英[†]
Yasuhide Mori

加藤 雅弘[†]
Masahiro Kato

木村 淳一[†]
Junichi Kimura

1. はじめに

近年大容量のHDDレコーダが普及し、ユーザが長時間にわたる複数の映像コンテンツを録画して視聴する機会が増えている。一方で録画されている映像の内容理解に多くの時間を要してしまうという課題がある。この課題を解決する手段として、映像内容を解析して映像シーンを分割・分類することにより、自動的に番組の重要シーンだけを再生する、自動ダイジェスト再生機能が必要とされている。例えば野球やサッカー番組においては、得点シーン等が重要シーンに分類され、番組中の音響的特徴を用いて検出する様々な研究が行われている[1]。

上記自動ダイジェスト再生機能における誤検出要因の一つとしてリプレイシーンの存在がある。リプレイシーンは重複したシーンなのでダイジェスト再生には不要である。しかしリプレイシーンは得点シーン等を集約したものとなっており、歓声の盛り上がり度が大きいなどの重要シーンと同じ特徴を有する。そのためリプレイシーンは重要シーンとして誤検出される可能性が高い。

このようにリプレイシーンはダイジェスト再生のシーンとしては不適切なため、検出して除去する必要がある。リプレイシーン検出手法は従来において類似画像検索による手法が考案されていた[2]。しかしこれらの手法は組み込み機器で動作させるためには更なる計算量の削減が必要である。

本研究では番組の映像及び音響の両特徴を用いた盛り上がり検出方法と、類似画像検索方法を組み合わせたリプレイシーン検出手法を提案する。リプレイシーンの検出精度が高く、かつ高速に検出する手法を検討する。

2. ダイジェスト再生機能の処理フロー

2.1. 構成

図1にダイジェスト再生機能の処理フローを示す。本研究における自動ダイジェスト再生機能のメディア処理部分をビデオダイジェストエンジンと呼ぶ。ビデオダイジェストエンジンは特徴生成部、重要シーン生成部および再生位置決定部の三つの部分から構成される。

特徴生成部は番組中の映像および音響のそれぞれについて単位時間毎の特徴データを生成する。映像の特徴データとしては長坂らの手法[2]で用いられている代表フレームの画素平均値を保存する。音響の特徴データとしては単位時間内の音響パワーを保存する。

重要シーン検出部は特徴生成部により生成された特徴データを解析し、映像や音響の変化度合いにより重要度を付与する。その重要度は番組ジャンルに依存する。

再生位置決定部は、ユーザの所望の時間に応じて、重要

度の大きいシーンから再生位置データとして選定する。

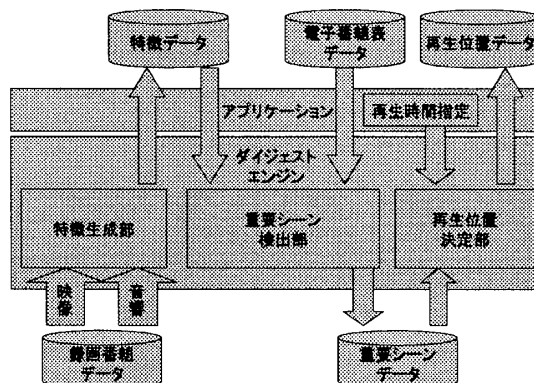


図1: ダイジェスト再生機能の処理フロー

2.2. 盛り上がりシーン検出方法

野球番組を例に挙げて重要シーンの検出方法を説明する。ホームランシーン等の野球番組における重要シーンは、単なる投球シーン等と比較して歓声の盛り上がり度が大きくなる傾向がある。そこで野球番組ではこれらの特徴を用いて重要シーンが検出可能となる。

盛り上がりシーンを検出する手順は、まず、映像特徴データにおいて前フレームとの画素毎の差分値の総和が大きいフレームをScene Change(以下SCと略す)点として検出する。続いて、2つの連続するSC点で挟まれる区間を複数集めて一つのシーンとする。次に、特徴データから得られる音響パワーの大きい順に各シーンを整列し、ある一定以上の音響パワーの大きさを持つシーンを盛り上がりシーンとして検出する。

3. リプレイシーン検出

3.1. リプレイシーン検出方法の検討

リプレイシーンは番組中に元となるシーンが含まれている。そこで従来の研究で行われている類似画像検索による手法を検討した(図2(a)) [2]。

本手法は最初に前述のSC点によりシーンを定義する。続いて3.2節で述べるシーン間の類似シーン判定を行い、番組内の類似シーン群の辞書を作成する。番組中の全てのシーンに関する辞書を作成後、類似シーン群に含まれるシーンをリプレイシーンとして決定する。以下この方式を方式(a)と呼ぶ。

方式(a)はシーンの開始点と終了点を様々に変えてシーンマッチングを行う必要がある。図2(a)に示すようにシーンAとその他のシーンとの類似シーン判定を行った後に、シーンAと一部重複するシーンBの類似シーン判定を行う。よって比較回数が多いといった欠点がある。このため、番組中の映像を一定法則で区切ったシーンで比較すること

により、比較回数を削減する手法を検討した(図 2(b))。以下この方式を方式(b)と呼ぶ。

しかし方式(b)は図 2(b)に示すように、例えばリプレイシーンがシーン B とシーン C の境界付近に存在した場合、検出漏れとなる可能性がある。そこで提案手法では、2.2節に示した盛り上がりシーンに対する類似シーン判定によるリプレイシーン検出を行う(図 2(c))。リプレイシーンは盛り上がり度が比較的大きいという特徴を有しているため、図 2(c)のようにシーン A とシーン C における盛り上がり度のピーク付近に位置することが多い。よって盛り上がり度のピークより数点前の SC 点でシーンを分割することで、リプレイシーンの途中で途切れることを防ぐことが可能である。以下この提案方式を方式(c)と呼ぶ。

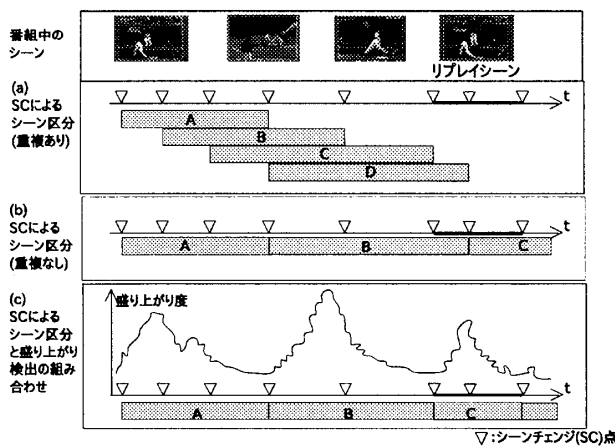


図 2: リプレイシーン検出方式の検討

3.2. 類似シーン判定方法

類似シーン判定方法については以下の方法により行う。

1. 比較するシーン A とシーン B に含まれる代表フレームについて以下の式を満たす t_A , t_B 番目のフレームを検索する。

$$\sum_{y=0}^{height} \sum_{x=0}^{width} \{ P_A(t_A, x, y) - P_B(t_B, x, y) \} \leq \alpha \quad (1)$$

α は固定値、 $P_A(t, x, y)$, $P_B(t, x, y)$ はシーン A またはシーン B に含まれるフレームの特徴データの画素値を示す。

2. 1. で検出されたフレーム以降について以下の条件を満たす区間の長さを求める。

$$\sum_k^T \sum_y^{height} \sum_x^{width} (P_A(t_A + k, x, y) - P_B(t_B + k, x, y)) \leq \beta \quad (2)$$

β は固定値、 T はリプレイと判定する最大の区間長を示す。式(2)を満たしなくなったときの k を求める区間長とする。

3. 2. で求めた区間長 k が一定以上となった場合には類似シーンとして判定する。

以上のアルゴリズムによりシーン A とシーン B が類似シーンとして判定された場合、時系列で後方に位置する方のシーンをリプレイシーンとして検出する。

4. 実験と評価

3.1節で説明した方式(a) (b) (c) のリプレイシーン検出方式を 3 試合分合計 7 時間 29 分の野球番組に適用して、比

較実験を行った。なお、方式(a),(b)におけるシーンの区切り方法は、野球のリプレイシーンの各要素が 15-20 秒で構成されていることが多いことから、前述の長さになるように SC 点を選定した。

精度及び計算量の評価を行った結果を表 1 に示す。検出率は(検出したリプレイシーン数)/(正解のリプレイシーン数)、正解率は(検出したリプレイシーン数)/(リプレイシーンとして検出したシーン数)により求めた。自動ダイジェスト再生は重要シーンを漏らしてはならない。そこで重要シーンがリプレイシーンとして誤検出されるのを防ぐため、本研究では正解率を重視する。計算量は方式(a)の処理サイクル数を 1 とした評価値で示す。

表 1: リプレイシーン検出実験結果

| | 方式(a) | 方式(b) | 方式(c) (提案方式) |
|--------|--------|--------|-----------------|
| 検出率(%) | 76.47% | 64.71% | 76.47% |
| 正解率(%) | 88.89% | 88.57% | 96.67% |
| 計算量 | 1.00 | 0.35 | 0.41 |

実験の結果、提案方式である方式(c)が最も検出率と正解率で優れた方式であった。方式(a)は方式(c)と比較して、シーンの比較回数が多い分、投球シーン等を誤検出していた。方式(b)は当初の予想通りシーンの区切り方法による検出漏れが発生した。

また一方で放送中に挿入される三振シーンのハイライトといった、比較的短い間隔で切り替わっていくリプレイシーンでは方式(a),(b),(c)で共通して検出漏れが存在する。これらは類似シーン判定における区間長の下限を下げることで検出することができるが、逆に誤検出が増え正解率が低下するため好ましくない。

計算量についてはシーンの比較回数が少ない方式(b)が最も優れていた。ただし方式(c)は方式(b)との計算量の差はわずかであり、方式(a)と比較すると約 60% 高速である。

5. おわりに

自動ダイジェスト再生機能のためのリプレイシーン検出手法を提案した。提案手法は、番組中の映像及び音声の特徴を用いた盛り上がりシーン検出と映像の類似性判別で構成されている。両手法を統合することにより、従来の類似性判別のみによる手法と比較して、精度向上と高速な検出を両立できることを確認した。今後は更なる精度の向上と提案アルゴリズムを野球以外の様々なジャンルの番組への適用方法を検討する。

参考文献

- [1] I.Otsuka, K.Nakae, A.Divakaran, K.Hatanaka, and M.Ogawa: "A Hightlight Scene Detection and Video Summarization System using Audio Feature for a Personal Video Recorder", IEEE Transactions on Consumer Electronics, Vol. 51, No. 1, 2005-2
- [2] 長坂 晃朗, 宮武 孝文, 上田 博唯: "カットの時系列コーディングに基づく映像シーンの実時間識別法", 電子情報通信学会論文誌, D-II Vol. J79-D-II No.4, pp.531 - 537, 1996-4