

フレームベースクラスタリングを利用したショット検出手法 の一検討

梅田直樹[†] 青木輝勝^{†‡} 沼澤潤二^{†‡}

[†]東北大学 大学院情報科学研究科 [‡]東北大学 電気通信研究所

1. はじめに

近年、データの圧縮技術やネットワーク関連技術、記憶媒体の発展を背景に、映像コンテンツが一般的に広く利用されてきている。しかし、ユーザが映像コンテンツの取得、蓄積を容易に行うことができる一方、映像データが大量になることで、ユーザが閲覧したいコンテンツにたどり着くことが困難になっている。そのため、映像コンテンツに対してメタデータによる検索を行うために、自動的にメタデータを付与する研究が盛んに行われている。

メタデータを付与する前処理として、映像の構造化が必須である。本研究ではその構造レベル(フレーム、ショット、シーン、クリップ)の中の、ショットレベルの構造化を行うためにショットを検出することを目的としているが、特に筆者らは検出漏れをゼロに保ったまま、誤検出を減らすことを目指した手法の研究を進めている。そのため、フレーム単位で種々の画像特徴量を抽出しその類似度によりフレームのクラスター解析を行い、ショット検出する手法を提案してきた。

2. 従来研究の問題点

2007 年の国際的な動画検索を対象とするワークショップである TRECVID^[1]で行われたショット境界検出の成果を見ると、CUT に対しては Recall, Precision 共に 98%近い結果が発表されている。また、GT (Gradual Transition)では、最も良い結果でも Recall, Precision 共に 80%を超えるものはない。

検出漏れと誤検出の原因として従来研究でいくつか指摘されてきた。しかし、筆者らが従来研究で最も問題であると考えたことは、検出率を上げる過程で、検出漏れが存在していることである。

メタデータ付与のために、ショット境界検出で処理された結果を用いるときに、検出漏れが 1箇所でもある場合その漏れを探すコストは、誤検出を探すコストに比べて非常に高いと考えられる。なぜなら検出漏れがある場合は映像コンテンツを再び最初から最後まで見る必要があるためである。

3. 提案手法

本研究では、映像クリップからフレーム間差分の数値を分類してショット境界を分割して行く従来研究で主に使われている手法を採らない。全てのショットが、一枚のフレームという状態、つまり検出漏れがゼロの状態から同じショットであるフレーム同士を繋げていくことで、検出漏れがゼロの状態を保ったまま、ショット自体を検出する手法を提案する。

また、検出の精度を上げるためにショット境界検出に特化したベクトル空間を見つけることを目指す。そのために、フレームから抽出できる画像特徴量をできるだけ多くショット境界検出のシステムに組み込み、次元縮約を行うことを提案する。次元縮約とは、本質的な情報を保持したままより低次元のベクトル空間に変換する処理のことである。これにより、低次元のショット境界検出に特化したベクトル空間を発見することができると考えられる。

以上の提案を盛り込んだシステムを図 1 に示す。また、以下でブロックの詳細について説明する。

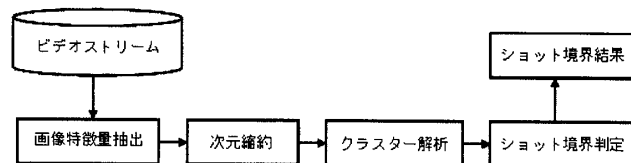


図1 提案システムの概要

3.1. 画像特徴量抽出

検出漏れを避けるためにフレーム同士の差を

“Shot boundary detection with frame-based clustering”

Naoki UMEDA[†], Terumasa AOKI^{†‡}

and Junji NUMAZAWA^{†‡}

[†]GSIS, Tohoku Univ.

[‡]RIEC, Tohoku Univ.

十分に表すことができるように多くの画像特徴量を使うことにする。そのために利用する画像特徴量は RGB カラーヒストグラム (HIST), HSV カラーレイアウト (CL), エッジヒストグラムである。

RGB カラーヒストグラムではフレームを 4×4 に分割し、それぞれのブロックに対して各ピクセルを各ビンに振り分ける。HSV カラーレイアウトでは 16×16 pixels のブロックごとに平均色を計算する。また、エッジヒストグラムの抽出手法として、Park らによる手法^[2]を用いた。

3.2. 次元縮約

画像特徴量抽出で得られた画像特徴量ベクトルの要素は膨大となるので、それをそのまま扱うと計算時間と計算領域が増えてしまう。そこで、何種類もの要素を、できるだけ情報の損失なしに、1 つ、または少ない数にまとめて表現する分析方法を用いて効率のよい検出ができる次元まで次元縮約を行う。また、同じショット内のフレーム同士は近づけ、異なるショットのフレーム同士は遠ざけるように次元縮約を行う。

3.3. クラスタ解析

フレーム毎の特徴量ベクトルを用いて、類似フレーム同士を繋ぎ合わせ、クラスターと呼ばれる集合を作る。このような処理をすることで、同じショット内のフレームは同じクラスターに分けられ、ショット自体を検出することができる。

クラスター解析では階層的クラスター解析の手法であるウォード法を利用する。また、再生時間の長い映像クリップを一度にクラスター解析すると計算時間が長くなるため、適当な長さに分割してクラスター解析を行う。

また、連結値の増加分が閾値を超えた箇所でもクラスター数を決定する。

3.4. ショット境界判定

クラスター解析によって、クラスター毎に分けられたフレーム群を、フレーム毎のタイムスタンプ等を考慮してショットに分割する。ショット境界は異なるショットの境目なので、ショット毎に分割されたフレームからショット境界を検出することができる。

4. 提案手法の評価

提案システムを用いて、3.1 節で説明した画像特徴量の全ての組み合わせに対して比較実験を行い、提案手法の評価を行った。また、それぞ

れの画像特徴量の重みが等しくなるように正規化を行った。今回は用いる画像特徴量が少ないため評価用システムでは次元縮約は行わなかった。

映像ソースとして 14 本の映像クリップ (総フレーム数 40,518 フレーム, CUT 数 177 ヶ所, GT 数 23 ヶ所) を用いて、それぞれの画像特徴量に対して Recall=1 としたときの最大となる Precision を求めた。TRECVID2007 参加者の CUT での結果との比較を図 2 に示す。

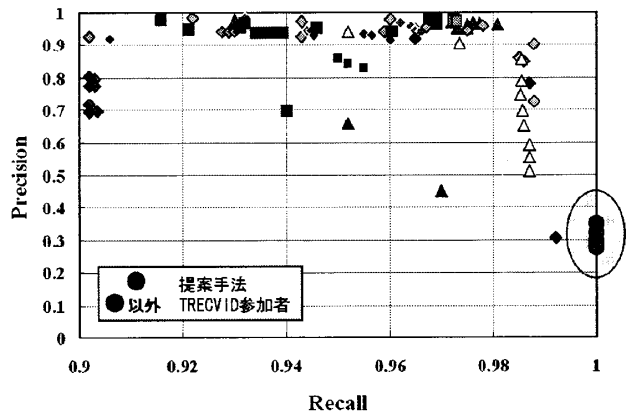


図 2 TRECVID2007 参加者の CUT の認識結果との比較

TRECVID の参加者の結果では、Recall=1 を達成している結果はなかったが、提案手法では Recall=1 としたとき、どの画像特徴量でも Precision が 0.3 付近の結果となった。この値は既存方式と比較して低い値ではあるが、Recall=1 を実現したことは実用上極めて意義が高く、今後は Recall=1 を保ちながら Precision 値を高めることにより理想的な手法の提案が行えるものと期待できる。

5. まとめ

検出漏れをゼロに保ったまま誤検出を減らすことを目指した手法として、クラスター解析によりショット内のフレームをつなぎ合わせ、ショット自体を検出することを提案してきた。

今後は、画像特徴量を増やし適切な特徴量を検討し重み付けを変えることにより、Precision の更なる改善を目指す。

文献

- [1] TRECVID
<http://www-nlpir.nist.gov/project/trecvid>
- [2] D. K. Park, etc. "Efficient Use of Local Edge Histogram Descriptor", Proceedings of ACM International Workshop, 2000