

フレームベースベクトル空間を利用したショット検出手法 の一検討

梅田直樹¹⁾ 青木輝勝²⁾ 山田洋³⁾ 沼澤潤二⁴⁾

¹⁾東北大学 工学部 ²⁾³⁾⁴⁾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、動画コンテンツが一般的に広く利用されてきている。しかし、ユーザがデジタル映像データの取得、蓄積を容易に行うことができる一方、デジタル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、動画コンテンツに対して自動的に索引付けを行う研究が盛んに行われている。その際、索引付けを行う前処理として、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、シーンなど)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。

本研究では、これは、同じショット内のフレームであれば類似しているという性質に着目し、類似画像でクラスタリングしていくことで、ショットごとのクラスタを作り、ショットを検出する手法を提案する。この際、あらかじめ固定化された特徴量を用いるよりも可能な限り多くの画像特徴を用いることが望ましいことは自明であるが、そうすると扱う画像特徴量が膨大になり計算が困難になると予想される。そこで、画像特徴量から作られる特徴量ベクトルの次元数を減らすために主成分分析を適用する。

2. 既存研究

既存の一般的なショット検出手法は、3段階の構成で説明することができる。それは、特徴量抽出、連続信号の構成、連続数値の分類の3段階である[1]。このうち連続信号の構成では、主にフレーム毎にとられた特徴量から、2フレームまたは3フレーム以上の連続フレームより、その差分を計算している。それより得られた連続数値を分類することで、ショットを検出している。近年では、その分類の手法に Support

Vector Machine(SVM)[2,3] や Finite State Machine(FSM)[4], クラスタ解析[5]を用いた手法があり、TREC Video Retrieval Evaluation (TRECVID)[6]と呼ばれる動画検索を研究対象とするワークショップで効果的なショット検出手法として提案されている。

ショット検出の評価尺度として、適合率(Precision)と再現率(Recall)が用いられている。それぞれの計算方法は以下である。

$$P = \frac{D}{D + D_F}, R = \frac{D}{D + D_M}$$

D は正しく検出されたショット境界の数、D_F は誤検出の数、D_M は検出漏れの数である。一般に、Precision と Recall は Trade-off な関係である。2007 年の TRECVID では瞬時切り替え(Cut)では、最大で Precision と Recall 共に 98%であった。一方、随時切り替え(Gradual Transition)では Precision と Recall 共に 80%を超える結果は見られなかった。Gradual Transition には様々な種類(フェード、ディゾルブ、ワイプなど)があり、フレーム毎の変化が少ないために検出しにくい。誤検出の主な原因として、激しい照明の変化や画面中のオブジェクトの激しい動き、カメラモーションが挙げられる。検出漏れの主な原因としては、ショット境界の前後フレームの構図や色の変化が少ない場合や Gradual Transition ではフレーム間での変化が少ないことが挙げられる。

3. 提案手法

既存の研究では、Precision と Recall のバランスをとりながら検出率を上げることがよいと考えられている。しかし、検索のためのメタデータをつけることを考えたとき、誤検出よりも検出漏れがある場合にその誤りを人の手で修正する際にかかる負担の方が大きいと考えられる。それは、誤検出の場合は修正箇所だけを見て判断すればよいのに対し、検出漏れがある場合、最悪また動画コンテンツを初めから見てショット境界を捜さなくてはならないからである。そこで、ある程度の誤検出を許容し、検出漏れ

“Shot boundary detection with frame-based vector space”

¹⁾Naoki UMEDA • Tohoku Univ. School of Engineering

²⁾Terumasa AOKI • Tohoku Univ. RIEC

³⁾Hiroshi YAMADA • Tohoku Univ. RIEC

⁴⁾Junji NUMAZAWA • Tohoku Univ. RIEC

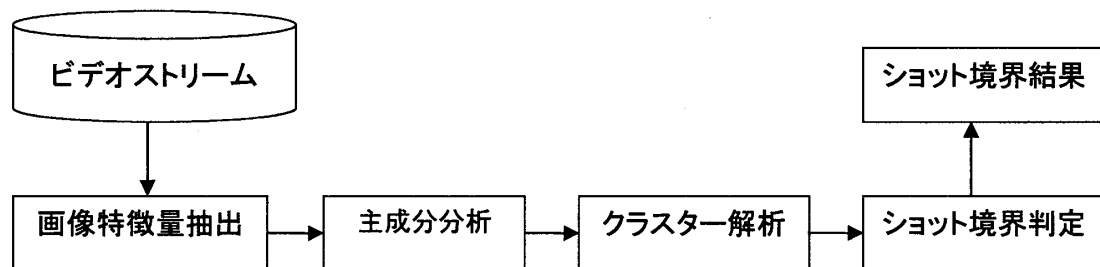


図1 ショット検出システムの概要

をゼロとすることを目指す。もちろん最終的には、誤検出もゼロとすることが望ましい。

既存研究で紹介したように、既存のショット検出では連続するフレームの差分を計算し、その数値によってフレーム間がショット境界であるかどうかの分類をしているが、本研究ではショット内のフレームは類似フレームであるという性質に着目して、クラスター解析により類似フレーム同士でクラスターを作り、フレームをショット毎に分ける手法を提案する。模式図を図2に示す。しかし、ショット内でカメラワークまたはオブジェクトの動きによりショットの最初のフレームと最後のフレームが全く違う場合はどのように解決するのかなどの問題がある。

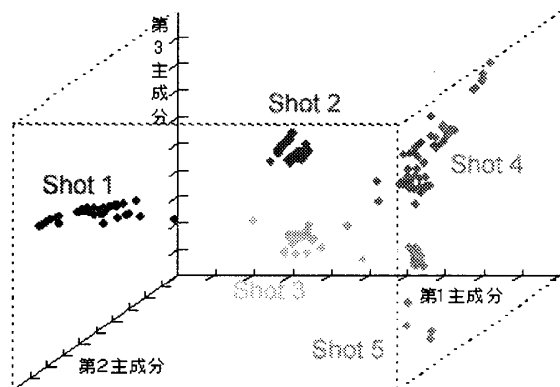


図2 クラスターの形成

クラスター解析の際に利用する特徴量ベクトルはフレーム毎の特徴量ベクトルとする。そのときにフレームから抽出され利用される画像特徴量は、検出漏れを避けるためにフレーム同士の差を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。これは、検出漏れの主な原因としてフレーム間の変化が少ないことが挙げられるため、フレーム間の差異を十分に考慮するために多くの画像特徴量を用いた方が検出漏れを減らすことができると考えられる

からである。またその際、主成分分析により次元削減を行うことにより、クラスター解析を行いやすい次元数にする。

提案するショット検出手法の流れを図1に示す。クラスター解析、主成分分析という処理では、先に記述した処理を行う。また、ショット境界判定では得られたクラスターから、連続するフレームが同じクラスター内に存在しない場合はショットの境界であるとしてショット境界フレームとして検出し出力する。このとき、フラッシュライトなどのノイズによる誤検出が起こる可能性があるため、タイムスタンプ等の情報を活用して正しいショット境界であるかの判定も行う。

参考文献

- [1] Jungui Yuan et al. "A Formal Study of Shot Boundary Detection" IEEE Transactions on circuits and system for video technology, vol. 17, No. 2, February 2007
- [2] 宮村祐一, 中村太一, 篠田浩一, 古井貞熙 "多段SVMを用いた頑健な動画ショット境界検出" 画像の認識・理解シンポジウム(MIRU 2007) IS-2-19, pp.815-820 (2007-7)
- [3] C. Liu, H. Liu, S. Jiang, Q. H. Y. Zheng and W. Zhang: "Jdl at trecvid 2006 shot boundary detection", TRECVID 2006 Workshop (2006).
- [4] Zhu Liu, E. Zavesky, D. Gibbon, B. Shahraray, P. Haffner "AT&T RESERCH AT TRECVID 2007", TRECVID 2007 Workshop(2007)
- [5] Chi Chun Lo, Shuenn Jvi Wang "Video segmentation using a histogram-based fuzzy c-means clustering algorithm", Computer Standards & Interfaces 23 (2001) 429-438
- [6] TRECVID
<http://www-nlpir.nist.gov/project/trecvid/>