I-026

多次元特徴空間解析に基づく映像のカット検出手法

A Cut Detection Method for a Video Sequence based on Multi-Dimensional Feature Space Analysis

岩元 浩太† 山田 昭雄†
Kota Iwamoto Akio Yamada

1．はじめに

映像コンテンツへの効率的なアクセスを実現するためには、映像を時間軸上で分割することが重要であり、その基本単位となるのがショットである。ショット間の境界には(1)瞬時に切り替わるカットと、(2)徐々に切り替わるトランジション(ディゾルブ、ワイプなど)があるが、本稿ではショット間の境界の大半を占めるカットを検出する手法について述べる。従来、カット検出手法には、フレーム間の比較に基づく手法が多数提案されている。この手法では、各フレームから特徴量を抽出し、隣接するフレーム間の特徴量の差分値がある閾値を超えた場合にカットを検出する。特徴量として、画素値を用いるもの[2]、画像のブロック単位で算出された平均画素値・分散値を用いるもの[3]、色ヒストグラムを用いるもの[4]、エッジ情報を用いるもの[5]などが提案されている。また、MPEG 符号化情報のマクロブロックの DC 成分を用いるもの[6]、符号化モードの情報を用いるもの[7]なども提案されている。しかしながらこれらフレーム間の比較に基づく手法は、(1)カメラのフラッシュなどの一時的な異常値(特に連続的に続く場合)、(2)フレーム間の差分値が大きくなるカメラモーションなどの動きの激しいシーン、に対して過剰にカットを検出してしまうという問題がある。本稿では、フラッシュや動きに頑健なカット検出手法を提案する。

2．提案手法

提案するカット検出手法は、フレームから抽出した特徴量が特徴空間内で連続的に移動する特性を利用し、多次元特徴空間での軌跡解析に基づく手法である。フレーム特徴量の特徴空間内における軌跡を追跡し、その非連続点をカットとして検出する(図 1)。具体的には、動きの影響を抑えるため過去の特徴量系列から予測処理を行い、実際の特徴量との予測誤差を評価する。予測誤差が大きい点が存在する場合は、フラッシュなどの一時的な異常値である可能性を考慮し、その点を除いた予測処理・予測誤差の算出を行う。予測誤差の時系列的な推移を評価し、予め定められた閾値を超える予測誤差が連続する場合にカットを検出する。

提案手法では各フレームに対して、(1)特徴量抽出、(2)予測処理に基づくカット検出、(3)ポストフィルタリング、の 3 段階の処理を行い、カットであるか否かの判定を行う。ポストフィルタリングは、テロップ重畳に起因する過剰検出を排除するために行う処理である。以下に各処理について詳細に説明する。

† NEC メディア情報研究所
Media and Information Research Laboratories, NEC Corp.

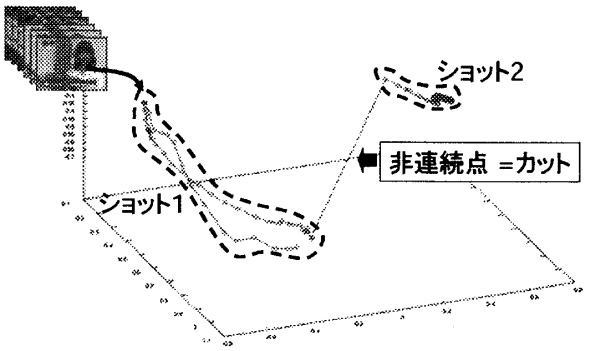

図 1: 提案手法の概念図

2.1 特徴量抽出

各フレームから特徴量を抽出する。カット判定を行ううえで十分な特徴表現力を持つ特徴量が必要になるため、複数の視覚特徴量を統合する。まず、フレーム画像から、(1)ブロック平均値、(2)ブロック分散値、(3)カラーレイアウト、(4)エッジヒストグラム、(5)HSV ヒストグラムの 5 種類の視覚特徴量(合計 550 次元)を抽出する。表 1 に、5 種類の視覚特徴量の詳細を示す。抽出された 550 次元の視覚特徴量に正規化処理を施し、主成分分析を実行する。主成分分析で得られた上位の主成分(累積寄与率が 90%以上となる約 100 主成分)を各フレームの特徴量とする。

表 1: 5 種類の視覚特徴量

視覚特徴量	次元数	説明
ﾌﾞﾛｯｸ平均値	94	ﾌﾞﾛｯｸに含まれる輝度値の平均値
ﾌﾞﾛｯｸ分散値	94	ﾌﾞﾛｯｸに含まれる輝度の分散値の平均値
ｶﾗｰﾚｲｱｳﾄ	84	8x8 画素に縮小された画像の周波数成分
ｴｯｼﾞﾋｽﾄｸﾞﾗﾑ	150	ﾌﾞﾛｯｸに含まれるエッジ方向の出現頻度分布
HSVﾋｽﾄｸﾞﾗﾑ	128	HSV 表色系による色の出現頻度分布

2.2 予測処理に基づくカット検出

特徴量の予測処理に基づきカット検出を行う。過去のフレーム列の特徴量から、現フレーム(以後、現フレームを n で表す)の特徴量の予測値を算出する。予測には、線形予測モデルである自己回帰モデル(AR モデル)を用いる。すなわち、フレーム n の特徴量の予測値 $\hat{\mathbf{x}}_n$ を、それよりも過去のフレーム列の特徴量 $\mathbf{x}_{n-1}$、$\mathbf{x}_{n-2}$、$\mathbf{x}_{n-3}$、…の重み付き線形和として求める。

$$\hat{\mathbf{x}}_n = \mathbf{A}_1\mathbf{x}_{n-1} + \mathbf{A}_2\mathbf{x}_{n-2} + \mathbf{A}_3\mathbf{x}_{n-3} + \ldots \quad (1)$$

ここで $\mathbf{A}$ は自己回帰係数行列であり、予めサンプル映像を学習して求める。

算出されたフレーム n の特徴量の予測値 $\hat{\mathbf{x}}_n$ を、実際の特徴量 $\mathbf{x}_n$ と比較し、予測誤差を算出する。予測誤差の算出には、予測誤差ベクトルの確率分布を考慮したマハラノビス距離を用いる。すなわち予測誤差 E は、

$$E = [\mathbf{x}_n - \hat{\mathbf{x}}_n]^t \sum^{-1} [\mathbf{x}_n - \hat{\mathbf{x}}_n] \quad (2)$$

と算出する。ここで $\sum^{-1}$ は同一ショット内における予測誤差ベクトルの共分散行列の逆行列であり、予めサンプル映像を学習して求める。

予測誤差 E が予め定められた閾値[1]以上である場合、フレーム n がフラッシュなどの一時的な異常値である可能性を考慮し、フレーム n の特徴量を除外して(欠損値とする)それ以降のフレーム $n+t$ (t は 1～最大 T まで)に対する予測誤差を同様に算出する。フレーム n から $n+T$ までの全てのフレームに対する予測誤差が予め定められた閾値以上である場合に、フレーム n をカットと判定する。

2.3 ポストフィルタリング

本手法は画像全体から抽出した特徴量を用いているため、テロップが重畳されるなど画像が部分的に変化した場合に特徴量が大きく変化し、誤ってカットを検出してしまう場合がある。このような画像の部分的な変化に対応するために、ブロック間の照合に基づくポストフィルタリングを行い、過剰検出を排除する。検出されたカットの前後のフレームに対して、画像を 8x8=64 のブロックに分割し、各ブロックの色特徴を求める。前後のフレームにおける対応するブロック間で特徴量を比較し、その差分値がある閾値以下である類似ブロックの数を求める。類似ブロック数がある別の閾値以上である場合に、検出されたカットが画像の部分的な変化に起因する過剰検出であると判断し、カットを排除する。

3. 評価実験

提案手法の有効性を確認するため、評価実験を行った。評価実験に使用した映像は合計約 210 分のニュース映像であり、含まれるカット数は 1276 である。なお、自己回帰係数行列、予測誤差ベクトルの共分散行列の逆行列は予め約 6 時間のサンプル映像を学習して求めた。評価実験では、従来手法と提案手法の精度の比較を行う。従来手法は、Gargi ら[1]が行った様々なカット検出手法の精度比較実験において最も高精度であったヒストグラム特徴のフレーム間比較に基づく手法を用いる。評価の方法としては、再現率と適合率を用いた。それぞれの算出式を以下に示す。

$$\text{再現率}(\%) = \frac{\text{検出された正解カット数}}{\text{全正解カット数}} \times 100 \quad (3)$$

$$\text{適合率}(\%) = \frac{\text{検出された正解カット数}}{\text{全検索結果数}} \times 100 \quad (4)$$

評価実験の結果を表 2 に示す。提案手法は、再現率 96.8%、適合率 93.4%を達成し、従来手法に比べて適合率が 18.3%改善している。精度改善の主要因は、フラッシュに起因する過剰検出が抑制されたことにある。

表 2: 実験結果

手法	再現率	適合率
従来手法	95.4%	75.1%
提案手法	96.8%	93.4%

4. まとめ

多次元特徴空間解析に基づいて映像中のカットを高精度に検出する手法を提案した。提案手法では、多次元特徴空間内においてフレーム特徴量が描く軌跡の非連続点をカットと判定する。予めサンプル映像を学習して得られた予測関数を用いて、過去のフレーム列の特徴量から現フレームの特徴量の予測値を求め、実際の特徴量との予測誤差を算出する。予測誤差の時間的な推移を評価することにより、カメラのフラッシュなどの一時的な異常値にも頑健にカットを検出する。評価実験の結果、提案手法は、再現率 96.8%、適合率 93.4%を達成し、従来手法と比較して適合率が 18.3%改善された。

参考文献

[1] Ullas Gargi, Rangachar Kasturi, and Susan H. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods", IEEE Transaction on Circuits and System for Video Technology, Vol.10, No.1, February 2000.

[2] H.J. Zhang, A. Kankanhalli, S.W. Smoliar, "Automatic Partitioning of Full-Motion Video", Multimedia Systems 1, pp.119-128, 1993.

[3] R. Kasturi, R. Jain, "Dynamic Vision", in Computer Vision: Principles, pp.469-480, IEEE Computer Society Press, 1991.

[4] A.Nagasaka, Y.Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances", in Visual Database Systems II (E. Knuth and L.M. Wegner eds.), pp.113-127, Elsevier, 1992.

[5] Ramin Zabih, Justin Miller, Kevin Mai, "A Feature-Based Algorithm for Detecting and Classifying Production Effects", Multimedia Systems 7, pp.119-128, 1999.

[6] B.L. Yeo and B. Liu, "Rapid Scene Analysis on Compressed Video", IEEE Transactions on Circuit and System for Video Technology, Vol.5, No.6, pp.533-544, December 1995.

[7] Hee-Chul Hwang and Duk-Gyoo Kim, "Shot Detection from MPEG Compressed Video", IEICE Trans. Fundamentals, Vol.E87-A, No.6, June 2004.

[1] 閾値は、予めサンプル映像を学習して得られた同一ショット内における予測誤差の確率分布に基づいて定める。