

局所ヒストグラムの時間変動に着目したディゾルブ検出
Dissolve Detection based on Time Variation of Local Histogram

前田 憲志†
Satoshi Maeda

田中 孝志†
Takashi Tanaka

黒木 修隆†
Nobutaka Kuroki

廣瀬 哲也†
Hirose Tetsuya

沼 昌宏†
Masahiro Numa

1. まえがき

近年の放送デジタル化による番組の増加、大容量ハードディスクの普及に伴い、視聴者が蓄積する映像は増加している。より効率的な視聴を可能にするために、番組の要約・構造化に関する研究が盛んに行われている。特にスポーツ映像において、リプレイなどの重要なシーンの前後にはディゾルブと呼ばれる映像編集技術が用いられている。本研究では、画像解析により特徴量を抽出し、得られた特徴量ベクトルからサポートベクトルマシン(SVM)によりディゾルブを検出する手法を提案する。

2. ディゾルブ検出手法

2.1 ディゾルブ

ディゾルブとは、図1のように2つの映像を徐々に切り換える映像効果である。

2.2 輝度変化区間の探索

一般的に、ディゾルブ区間では輝度変化が激しいため、前処理として輝度変化の激しい区間を探索し、その区間においてのみディゾルブ判定を行うことで処理の軽減を図る。

2.3 特徴量の抽出

2.3.1 特徴量抽出領域の分割

特徴量を算出するにあたり、画素単位でフレーム間差分を算出した場合、被写体のわずかな動きやカメラのゆれだけで変化を検出してしまふ。そこで画像を図2のように縦 $P \times$ 横 Q 個の領域 z_i ($i=1, 2, \dots, P \times Q$)に分割し、その各領域に対して以下の5つの特徴量を求める。

(1) 輝度ヒストグラム特徴量

分割した各領域 z_i に対し、輝度ヒストグラム $H_{z_i}(n, Y_x)$ を作成する。 $H_{z_i}(n, Y_x)$ は第 n フレームにおける輝度値 Y_x の出現画素数である。ここで輝度値 Y_x は F 段階に量子化されているものとする。隣接フレームにおいてその差の総和

$$YD_{z_i}(n) = \sum_{x=1}^F |H_{z_i}(n, Y_x) - H_{z_i}(n+1, Y_x)| \quad (1)$$

を輝度ヒストグラム特徴量とする。



図1 ディゾルブ画像

(2) エッジ特徴量

各フレームにおいてエッジ画像を生成し、第 n フレームの領域 z_i におけるエッジ数を $EN_{z_i}(n)$ とする。隣接フレームにおけるエッジ画素数の差

$$ED_{z_i}(n) = EN_{z_i}(n) - EN_{z_i}(n+1) \quad (2)$$

をエッジ特徴量とする。

(3) 両端フレームにおける差分特徴量

輝度変化区間 $n \sim n+t$ の両端フレームにおいて、輝度ヒストグラムの差分

$$BEYD_{z_i}(n) = \sum_{x=1}^F |H_{z_i}(n, Y_x) - H_{z_i}(n+t, Y_x)| \quad (3)$$

とエッジ画素数の差分

$$BEED_{z_i}(n) = EN_{z_i}(n) - EN_{z_i}(n+t) \quad (4)$$

を算出し、両端フレームにおける差分特徴量とする。

(4) 単調性特徴量

第 n フレームの座標 (x, y) における輝度値を $f_n(x, y)$ とする。隣接3フレームに対し前後フレームの比較から、輝度値が単調に増加

$$f_{n-1}(x, y) < f_n(x, y) < f_{n+1}(x, y) \quad (5)$$

もしくは減少

$$f_{n-1}(x, y) > f_n(x, y) > f_{n+1}(x, y) \quad (6)$$

している画素数の総和を求め、単調性特徴量とする。

(5) 理想との差分特徴量

ディゾルブ区間内における理想的な画素値 $g_n(x, y)$ は、

$$g_n(x, y) = \frac{f_{n-1}(x, y) + f_{n+1}(x, y)}{2} \quad (7)$$

と表される。そこで、式(7)の値と実際の輝度値との差の絶対値を

$$ID_{z_i}(n) = \sum_{x, y \in z_i} |g_n(x, y) - f_n(x, y)| \quad (8)$$

で求め、理想との差分特徴量とする。

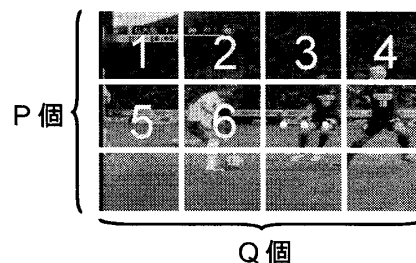


図2 特徴量抽出領域の分割

† 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University

表1 編集した映像に対するディゾルブ検出精度

解析対象映像	適合率 (%)		再現率 (%)		F 値 (%)	
	従来手法	提案手法	従来手法	提案手法	従来手法	提案手法
クロスディゾルブ(100ヶ所)	41.0	89.0	86.0	81.0	55.5	84.8
ホワイトアウト(100ヶ所)	98.0	91.7	97.0	100.0	97.5	95.2
ディゾルブ(100ヶ所)	56.5	92.6	52.0	99.0	54.2	96.2
暗転 (100ヶ所)	90.7	92.6	97.0	100.0	93.7	96.2
ランダム反転 (100ヶ所)	86.2	92.8	56.0	90.0	67.9	91.4
平均値	67.7	91.8	77.6	94.0	72.3	92.9

2.4 特徴量ベクトルの作成

第 n フレームから第 $n+t$ フレームにおける各領域の特徴量を結合し、特徴量ベクトル $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ を作成する。それらの時間軸上の関係を図3に示す。 \mathbf{v}_1 は輝度ヒストグラム特徴量・エッジ特徴量によって、 \mathbf{v}_2 は両端フレームにおける差分特徴量によって、 \mathbf{v}_3 は単調性特徴量・理想との差分特徴量によって構成し、さらに総合特徴量ベクトル \mathbf{V} を

$$\mathbf{V}(n) = \{ \mathbf{v}_1(n), \mathbf{v}_1(n+1), \dots, \mathbf{v}_1(n+t-1), \mathbf{v}_2(n), \mathbf{v}_2(n+1), \mathbf{v}_2(n+2), \dots, \mathbf{v}_2(n+t-1) \} \quad (9)$$

のように構成する。

2.5 SVMによるディゾルブ判定

本手法では、2.4節で作成した特徴量ベクトルを用いてSVMによりディゾルブであるか否かを判定する。SVMではあらかじめディゾルブ区間が既知の映像を用いてその区間の特徴量を学習しておく。

3. 実験と考察

3.1 評価実験

実験対象とする映像は、PC上の映像編集ソフトウェアによって5種類のディゾルブを100ヶ所挿入した映像(約14分)と、実際に放送されたサッカー映像3試合(約283分)である。従来手法として、2次差分画像を用いた手法 [1] と比較・評価する。評価項目は、正確性を表す適合率、網羅性を表す再現率と、それらの調和平均であるF値を用いる。

3.2 実験結果及び考察

編集した映像と実際の放送映像の実験結果をそれぞれ表1、図4に示す。表1より、提案手法では全ての種類のディゾルブに関して高い結果が得られていることがわかる。特にクロスディゾルブ・ディゾルブ・ランダム反転において、従来手法に比べ精度が大きく向上した。

また実際のサッカー映像に対して、提案手法では適合率、再現率、F値がそれぞれ51.9%、87.8%、65.2%となった。従来手法の適合率が低い要因は主に選手のアップを過検出するためである。提案手法では、特徴量を抽出する領域を分割し、複数の特徴量からSVMによって総合的に判断することで、従来手法に比べ過検出を削減し、より高精度なディゾルブ検出が可能となった。

4. まとめ

本研究では録画したスポーツ番組の要約に役立つと思われるディゾルブ検出の高精度化を目標として、局所ヒストグラムの時間変動に着目した手法を紹介した。今後は多彩なディゾルブを含む実際の放送映像に対してSVMによる学習を強化し、検出精度を改善する予定である。

参考文献

- [1] 如澤 祐尚, 上倉 一人, 渡辺 裕, 小寺 博, “ディゾルブ検出方法に関する一検討”, 電子情報通信学会ソサイエティ大会講演論文集, pp-231 (1996).

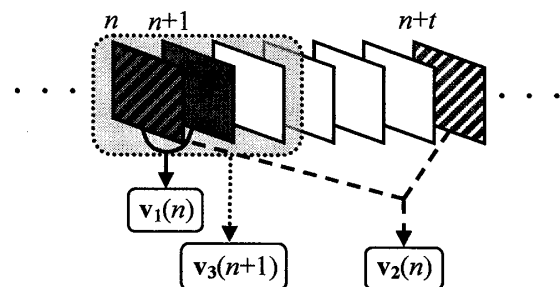


図3 特徴量ベクトルの作成

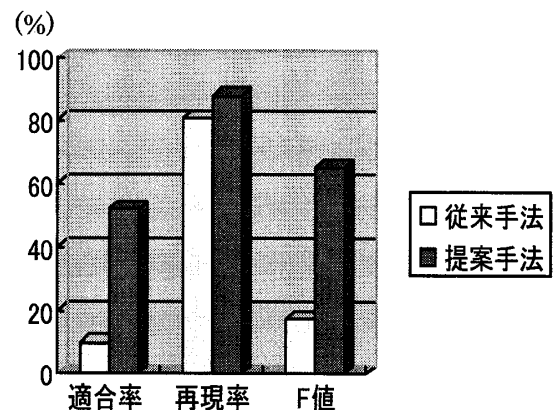


図4 実際の放送映像に対するディゾルブ検出精度