

## 1. はじめに

映像のショット境界検出, シーン抽出, カメラワーク解析, 顔検出, 物体検出, イベント検出, 要約映像生成など, シーン解析技術に関する研究は幅広く行われている. これらのシーン解析において映像の変化や状態をより良く表現できる特徴量が重要だが, 年々複雑化する映像の特殊効果や高解像度化の影響により, 従来用いられてきた特徴量のみでは十分に映像の変化や状態を捉えることが困難となっている. そこで本手法では, 映像の大域的な動きに着目した新たな特徴量を提案し, シーン解析において最も基本であり重要な技術であるショット境界検出へ適用することで有効性を示す.

## 2. ショット境界検出

ショット境界検出とは, 編集された映像から本来はばらばらに撮影された個々のショットのつなぎ目を検出することである. ショット境界は, 1 フレームで場面が切り替わるカット境界と, 2 フレーム以上で徐々に場面が切り替わるグラジュアル境界に大別できる. 従来, ショット境界検出にはカラーヒストグラム特徴や動きベクトル差分特徴が用いられてきたが, これらの特徴量ではグラジュアル境界検出が困難であった. この問題に対し, 我々は映像の大域的な変化に基づく特徴量を提案し, 複数の異なる特徴量の共起を表現することで高精度なショット境界検出を実現する.

## 3. 大域的な動きに基づく特徴量

映像に発生する変化には, カメラワークや編集による人工的な変化と, 映像内に含まれる人や背景等の動きによる自然な変化に大別できる. 従来シーン解析に用いられてきたアピアランスに基づく特徴量や動きのフロー等の局所的な変化に基づく特徴量では, 人工的な変化と自然な変化の区別が困難であった. そこで我々は, これらの変化を捉える新たな特徴量として大域的な動きに基づく特徴量を提案する.

## 3.1. ST-patch 解析

Shechtman 等により提案された ST-patch[1] は, 時間間における勾配から求められる. ST-patch を用いることで, 映像から 2 種類の特徴を表現することが可能である. 1 つ目の特徴量として, ある一つの ST-patch  $P$  のランク増加量  $\Delta r$  は, パッチ内の動きの一貫性を表現することが可能である.  $\Delta r \approx 0$  のときはパッチ内の各画素の動きが一貫していることを表し,  $\Delta r \approx 1$  のときはパッチ内の各画素の動きが複数あることを表している. 2 つ目の特徴量として, 異なる 2 つの ST-patch 間の動きの類似度  $m$  を表現することが可能である. 動きの類似度  $m$  は, ST-patch  $P1$  のランク増加量  $\Delta r_1$  と ST-patch  $P2$  のランク増加量  $\Delta r_2$ , さらに  $P1, P2$  を一つにした ST-patch  $P_{12}$  のランク増加量  $\Delta r_{12}$  から次式のように算出される.

$$m = \frac{\min(\Delta r_1, \Delta r_2)}{\Delta r_{12}} \quad (1)$$

$m \approx 1$  の場合は ST-patch 間の動きの類似度は高く,  $m \approx 0$  の場合は類似度が低いことを表す. 以上のように算出される ST-patch の動きの一貫性  $\Delta r$  と動きの類似度  $m$  を映像の大域的な変化に基づく特徴量と定義する.

## 3.2. ST-patch を用いた大域的な動き特徴抽出

通常, ST-patch は時系列画像の局所領域から作成される小型のパッチである. そのため, 画像全体の変化のような大域的な特徴を表現することは難しい. そこで我々は, 画像全体を 1 つのパッチとして ST-patch を作成する (図 1). 以下に, ST-patch 算出処理の流れを示す.

1. 入力画像を 1/10 のサイズにダウンサンプリング
2. ダウンサンプリングした入力画像を平滑化

3. 平滑化した入力画像のフレームバッファから ST-patch を算出

4. ST-patch から動きの一貫性  $\Delta r$  と動きの類似度  $m$  を算出

以上のようにして, 大域的な変化に基づく特徴量を算出する. 図 2 は映像から動きの一貫性  $\Delta r$  を算出した例である. 従来用いられてきた特徴量 (図 2 では動きベクトル (緑色)) ではカメラワークとグラジュアル境界で特徴量に大きな変化は見られないが, 動きの一貫性  $\Delta r$  (図 2 中の赤線) を用いることでカメラワークとグラジュアル境界の違いをはっきりと捉えることができる.

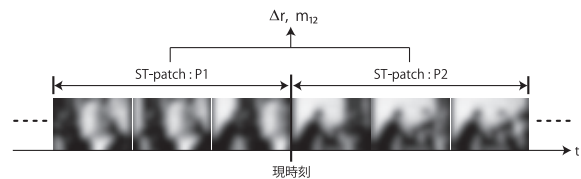


図 1: ST-patch の算出

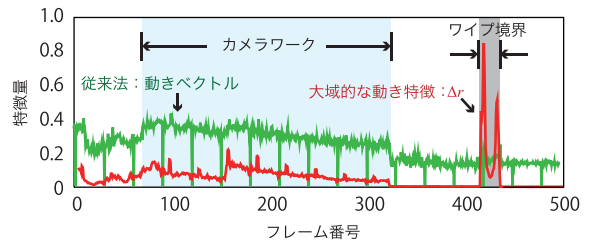


図 2: 算出した大域的特徴量

## 4. 共起表現を用いた AdaBoost 識別器の構築

提案手法では, 複数の異なる特徴量を同時に捉えることで単独の特徴量では困難なシーンの変化を捉える. 複数の異なる特徴量を同時に捉えて識別を行うために, 我々は Mita 等が提案する同時確率による共起表現を用いた AdaBoost 識別器 [2] を構築する. まず, 複数の特徴量の同時確率を求めるために特徴量を量子化する. 今回は, 2 クラス識別によるシーン解析を対象とするため 0 と 1 に 2 値化する. ある特徴量  $v_i$  について, 式 (2) に示すように 2 値化を行う. 2 値化のための符号  $s$  はベイズの定理 (式 (3)) により決定される.

$$s = \begin{cases} 1 & P(C_p|v_i) > P(C_n|v_i) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$P(C_k|v_i) = \frac{P(v_i|C_k)P(C_k)}{P(v_i)} \quad (k = p, n) \quad (3)$$

ここで,  $P(v_i|C_k)$  は確率密度分布から得られる確率,  $P(C_k)$  は事前確率,  $P(v_i)$  は出現確率,  $C$  はクラスである. 事前確率は両クラスとも等しいと仮定し, 0.5 を与える. 確率密度分布は, 全学習サンプルから得られる特徴量の出現頻度をヒストグラム化したものを用いる. このように 2 値化を行った特徴量を用いて, 新たな組み合わせ特徴量  $c$  を算出する. 例えば, 複数の特徴量から 3 種類を選択し, その組み合わせを表現する場合, 選択された 3 種類の特徴量  $f1, f2, f3$  を 2 値化した結果,  $f1 = 0, f2 = 1, f3 = 0$  が観測されたとする. このとき組み合わせ特徴量  $c$  は  $c = (010)_2 = 2$  のように算出される. ここで,  $c$  は 2 進表現された特徴の組み合わせのインデックス番号となる. AdaBoost 識別器の学習時, 組み合わせ特徴量  $c$  は図 3 に示すように, フレームバッファ ( $2N$ ) からエラー率が最小になるように自動的に選択される. 従って, 各弱識別器は複数のフレームから算出された複数の特徴量を表現することが可能となる.

次に, 得られた組み合わせ特徴量  $c$  を用いて AdaBoost の強識別器  $H(x)$  を構築する. 最終的に構築される識別器

$H(x)$  は,  $T$  個の弱識別器  $h_t(x)$  の線形結合として次式により表される.

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right) \quad (4)$$

ここで,  $\alpha_t$  は弱識別器  $h_t(x)$  の信頼度を表す重みである. 弱識別器  $h_t(x)$  において, 入力画像から組み合わせ特徴量  $c$  を観測する関数を  $C_t(x)$  とする.  $C_t(x)$  が組み合わせ特徴量  $c$  を観測したとき, 弱識別器  $h_t(x)$  を条件付き確率に基づく識別関数として次式で表す.

$$h_t(x) = \begin{cases} +1 & P_t(y = +1|c) > P_t(y = -1|c) \\ -1 & \text{otherwise} \end{cases} \quad (5)$$

ここで,  $P_t(y = +1|c)$  及び  $P_t(y = -1|c)$  は, それぞれ  $c$  を観測した結果であり, AdaBoost の学習サンプルの重みに基づいて算出される組み合わせ特徴量の同時確率である. このようにして構築された識別器を用いることで, 対象/非対象の 2 クラス識別を行う.

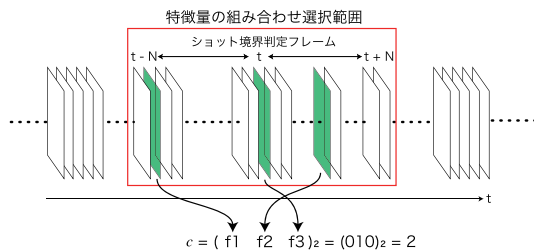


図 3: 特徴量間とフレーム間の共起を同時に表現

## 5. 評価実験

提案手法の有効性を示すため, カット・グラジュアル境界それぞれ従来法との比較を行う. 提案する特徴量の有効性と識別器の有効性を示すため, 従来用いられてきた特徴量であるカラーヒストグラム, オプティカルフロー, エッジ差分 (これらを特徴セット 1 とする) を使用した場合の従来法と提案手法の検出精度と, 提案する特徴量を追加した特徴セット (これらを特徴セット 2 とする) を使用した場合の検出精度の比較を行う. 検出精度の評価は, 次式に示す再現率 (Recall), 適合率 (Precision), F 値により行う.

$$R = \frac{C}{T} \times 100, P = \frac{C}{S} \times 100, F(R, P) = \frac{2RP}{R+P}. \quad (6)$$

$C$  は検出境界数,  $T$  は基地の正解境界数,  $S$  は検出したすべての境界数である. 検出対象はニュース, ドラマ, パラエティを含む映像約 150,000 フレームである.

### 5.1. 実験結果

表 1, 表 2 に実験結果を示す.

表 1: カット境界フレーム検出結果

|     |   | 従来法<br>(特徴セット 1) | 提案手法<br>(特徴セット 1) | 提案手法<br>(特徴セット 2) |
|-----|---|------------------|-------------------|-------------------|
| cut | R | 98.7             | 97.1              | 98.2              |
|     | P | 96.6             | 98.4              | 98.6              |
|     | F | 97.6             | 97.7              | 98.3              |

表 2: グラジュアル境界フレーム検出結果

|      |   | 従来法<br>(特徴セット 1) | 提案手法<br>(特徴セット 1) | 提案手法<br>(特徴セット 2) |
|------|---|------------------|-------------------|-------------------|
| grad | R | 74.0             | 75.4              | 80.1              |
|      | P | 92.0             | 91.7              | 92.4              |
|      | F | 82.0             | 82.7              | 85.8              |

結果を見ると, 提案手法はカットに対して特徴セット 2 で F 値 0.2% 向上している. また, グラジュアルに対して特徴セット 2 で F 値 3.8% 向上している. 以上の結果から, ショット境界検出に対する提案手法の有効性が確認できる.

### 5.2. 考察

提案手法では, 識別に有効な特徴量として AdaBoost の各学習ラウンドごとに異なるフレームから異なる特徴が自動的に選択される. 以下では, カット境界, グラジュアル境界ごとに学習時に選択された特徴量やフレームが識別能力に与える影響を調査する.

図 4 は構築した識別器から計測した組み合わせ特徴量の識別に対する貢献度である. カット境界ではアピランス特徴とローカル特徴の組み合わせ特徴の貢献度が高いのに対し, グラジュアル境界では提案するグローバル特徴との組み合わせ特徴の貢献度が高いことがわかる. 以上から, 提案する大域的な動きに基づく特徴量のグラジュアル境界検出への有効性が確認できる.

図 5 は, 構築した識別器から計測した組み合わせるフレームの識別に対する貢献度である. ここで 0 フレームは識別対象フレームを表す. カット境界は瞬間的に変化が発生するため, 識別対象付近のフレームの識別に対する貢献度が高いことがわかる. 一方, グラジュアル境界は徐々に変化が発生するため, 識別対象フレームから広範囲のフレームが識別に貢献していることがわかる. 以上から, フレーム間の共起を表現することで, 識別対象ごとに最適な範囲の情報が選択可能されることが確認できる.

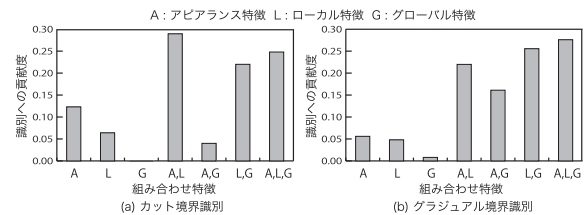


図 4: 各共起特徴の識別に対する貢献度

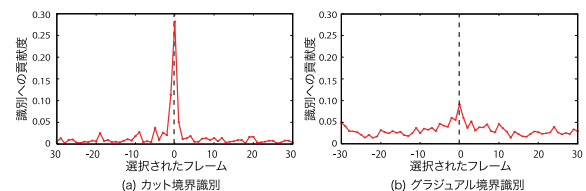


図 5: フレーム間の組み合わせとその貢献度

## 6. 他のシーン解析への応用例

本手法は他のシーン解析への応用が可能である. 表 3 は扉の開閉イベント検出への応用結果である. 検出対象イベントとそれ以外の 2 クラス識別問題として本手法を適用することで, F 値でそれぞれ 98.7%, 96.4% のイベント検出を実現している. 図 6 は扉の開閉イベント検出結果の例である. このように, 本手法は他のシーン解析への応用が期待できる.

表 3: 扉の開閉イベント検出結果

|                 | 領域 1 (右端の扉) | 領域 2 (中央の扉) |
|-----------------|-------------|-------------|
| 再現率 (Recall)    | 98.9        | 94.6        |
| 適合率 (Precision) | 98.7        | 98.3        |
| F 値 (F-measure) | 98.7        | 96.4        |

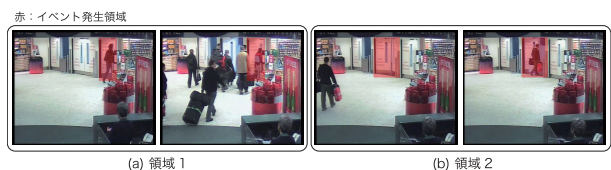


図 6: 検出結果例

## 7. おわりに

本稿では, 映像の大域的な動きに着目したシーン解析手法を提案した. 提案手法を用いたショット境界検出実験により提案手法の有効性が確認できた. また, 提案手法の他のシーン解析への応用例を示した.

### 参考文献

- [1] E. Shechtman and M. Irani, "Space-Time Behavior Based Correlation", CVPR2005, vol1, pp. 405–412.
- [2] T. Mita, T. Kaneko and O. Hori, "Joint haar-like features for face detection", ICCV2005, pp. 1619–1626.

### 研究業績

- [1] 村井陽介, 村井泰裕, 藤吉弘巨, "映像の大域的変化と局所的変化を用いたショット分割", CVIM159, pp. 217–224, 2007.
- [2] Y. Murai and H. Fujiyoshi, "Shot Boundary Detection Using Co-occurrence of Global Motion in Video Stream", ICPR, 2008.