

映像の大域的变化と局所的变化を用いたショット分割

村井 陽介† 村井 泰裕† 藤吉 弘亘†

† 中部大学 大学院 工学研究科

E-mail:yoo@vision.cs.chubu.ac.jp, yasu@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

あらまし 本稿では、映像の大域的变化と局所的变化の共起を用いたショット分割法を提案する。提案手法は、Space-Time Patchを用いて映像の大域的变化を、SIFTを用いて局所的变化を捉え、ショット検出を行う。画像全体を1つのST-Patchとして2つのST-patch間の動きの相関を求めることで、フレーム間の大域的な変化を捉える。一方、SIFTにより映像中の特徴点の対応から局所的な変化を捉える。これらの特徴量間の共起を表現し AdaBoost 識別器によりショット分割を行う。評価実験により、大域的な変化と局所的な変化の特徴量間の共起を表現することで、単体の特徴からの判別に比べ、ショット分割の精度を約 19% 向上させることができた。

Shot Boundary Detection using Global and Local Changes in Video Stream

Yosuke Murai† Yasuhiro Murai† Hironobu Fujiyoshi†

†Dept. of Computer Science, Chubu University

E-mail:yoo@vision.cs.chubu.ac.jp, yasu@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp

Abstract This paper presents a method of shot boundary detection based on the co-occurrence of global changes and local changes in video stream. Our approach use Space-Time patch for detecting global changes, and Scale-Invariant Feature Transform(SIFT) for detecting local changes in video stream. Our approach use the AdaBoost classifier for detecting shot boundaries using the co-occurrence of global and local changes. Experimental results show that our method has about 19% higher performance than that of the not using the co-occurrence of features.

1 はじめに

映像のショット分割とは、編集された映像から本来ばらばらに撮影された個々のショットに分割することである。映像のショット分割はメタデータの付加、動画像圧縮処理を行う上でも重要な技術であり、既に多くの手法が提案されている [1]-[4], [7]-[11]。ショットの分割フレームでは、瞬時切り替え、空間的切り替え、時間的切り替え等のショット切り替えが発生する。これらの様々なショット切り替えを自動的に検出する手法として、既にヒストグラムインターセクションとバッファリング手法を用いた検出法 [8] や、突出検出フィルタを用いたショット

検出法 [9] が提案されている。これらの手法はリアルタイムで高精度なショット検出が可能である。しかし、フレーム間の輝度値の変化からショット切り替えの検出を行うため、時間的切り替え等の特殊効果を用いたショット切り替えの検出が困難である場合がある。特殊効果を用いたショット切り替えの検出法として、高橋らはオプティカルフローを用いて画像の動きを評価する手法を提案している [11]。しかし、映像内の物体が非剛体であると、正確なオプティカルフローの推定が困難となり、正確なショット分割は難しい。

本稿では、映像の大域的变化と局所的变化の共起に着目したショット分割法を提案する。単一の特徴

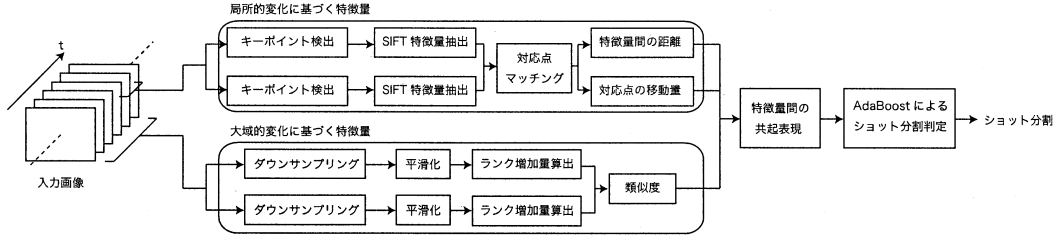


図 1: ショット分割の流れ

量では判別が困難なパターン (例えばカメラワークと特殊切り替え) に対する識別精度を向上させるために、大域的变化に基づく特徴量と局所的变化に基づく特徴量の共起性を表現し、全特徴の同時確率を用いた AdaBoost によるショット分割を実現する。本研究で対象とするショット切り替えは、瞬時切り替え、空間的切り替え (ページターン、押し出し)、時間的切り替え (ディゾルブ) とする。これらの映像中にはカメラワークが存在する。カメラワークには、左右方向の動きであるパン、上下方向の動きであるチルト、さらにズームアップ、ズームダウンが含まれている。このようなカメラワークに影響を受けずショット分割フレームを検出することを目的とする。

2 特徴量抽出

図 1 に提案するショット分割の流れを示す。映像の大域的变化に基づく特徴量抽出には Space-Time Patch (以下、ST-patch と呼ぶ) を、局所的变化に基づく特徴量抽出には Scale-Invariant Feature Transform (SIFT) を用いる。抽出された異なる特徴量間の共起を表現し、AdaBoost によりショット・非ショットの分割を行う。以下に、各処理の詳細を示す。

2.1 ST-patch による特徴量抽出

Shechtman 等により提案された ST-patch[5] は、時空間における勾配から求められる。図 2 に示すように異なる色で示したフレームバッファをそれぞれ ST-patch P1 と ST-patch P2 とする。3 つのラインは個々の画素の動き、 $[u \ v \ w]^T$ は ST-patch 中の個々の画素の動き方向ベクトルを表し、 ∇P_i は個々の画素の勾配方向ベクトルである。

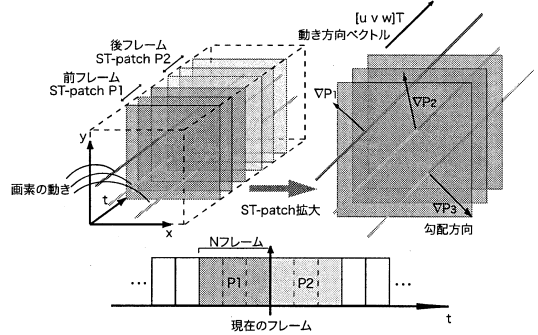


図 2: ST-patch の概念図

2.1.1 ST-patch

画像中の動きが一定の場合、それぞれの画素の勾配 $\nabla P_i = (P_{x_i}, P_{y_i}, P_{t_i})$ は、画素の動き方向ベクトル $[u \ v \ w]^T$ に対して垂直となるため、式 (1) が成り立つ。

$$\nabla P_i \begin{bmatrix} u \\ v \\ w \end{bmatrix} = 0 \quad (1)$$

画素数が n の場合、式 (1) は式 (2) となる。

$$\begin{bmatrix} P_{x_1} & P_{y_1} & P_{t_1} \\ P_{x_2} & P_{y_2} & P_{t_2} \\ \vdots & \vdots & \vdots \\ P_{x_n} & P_{y_n} & P_{t_n} \end{bmatrix}_{n \times 3} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1} \quad (2)$$

ST-patch 中の n 画素の ∇P_i からなる $n \times 3$ の行列を $\mathbf{G}$ とし、行列 $\mathbf{G}^T$ を掛けると式 (3) となる。

$$\mathbf{G}^T \mathbf{G} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3 \times 1} \quad (3)$$

行列 $\mathbf{G}^T \mathbf{G}$ は 3×3 の行列である。ここで、行列 $\mathbf{G}^T \mathbf{G}$ を $\mathbf{M}$ とすると式 (4) となる。

$$\begin{aligned} \mathbf{M} &= \mathbf{G}_0^T \mathbf{G} \\ &= \begin{bmatrix} \sum P_x^2 & \sum P_x P_y & \sum P_x P_t \\ \sum P_y P_x & \sum P_y^2 & \sum P_y P_t \\ \sum P_t P_x & \sum P_t P_y & \sum P_t^2 \end{bmatrix} \quad (4) \end{aligned}$$

2.1.2 2つのST-patchの一致

2つのST-patch P1, P2間に共通のベクトル $\mathbf{u} = [u \ v \ w]^T$ が存在する場合、P1とP2間の動きは一致し、各ST-patchにおける行列 $\mathbf{G}_1$ と $\mathbf{G}_2$ は $\mathbf{G}_1 \mathbf{u} = 0$ かつ $\mathbf{G}_2 \mathbf{u} = 0$ となる。ここで、 $\mathbf{G}_1 \mathbf{u} = 0$ と $\mathbf{G}_2 \mathbf{u} = 0$ をまとめると式 (5) となる。

$$\mathbf{G}_{12} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix}_{2n \times 3} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{2n \times 1} \quad (5)$$

行列 $\mathbf{G}_{12}$ はST-patchのP1とP2の両方の勾配を含む。式 (3)、式 (4) と同様に、式 (5) に行列 $\mathbf{G}_{12}$ の転置行列 $\mathbf{G}_{12}^T$ を掛け、 $\mathbf{M}_{12} = \mathbf{G}_{12}^T \mathbf{G}_{12}$ とすると式 (6) となる。

$$\mathbf{M}_{12} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}_{3 \times 1} \quad (6)$$

また、 $\mathbf{M}_{12}$ は式 (7) により求めることもできる。

$$\mathbf{M}_{12} = \mathbf{M}_1 + \mathbf{M}_2 = \mathbf{G}_1^T \mathbf{G}_1 + \mathbf{G}_2^T \mathbf{G}_2 \quad (7)$$

2.1.3 ランク制約による状態判定

画素の動き方向ベクトル $[u \ v \ w]^T$ に対して垂直となる勾配 ∇P_i は、動きが一定の場合、2次元平面上に現れる。2つのST-patchの動きが一致している場合、それぞれの ∇P_i は共に同一2次元平面上に現れ、行列 $\mathbf{M}_{12}$ のランクは $\text{rank}(\mathbf{M}_{12}) \leq 2$ となる (図3(a))。2つのST-patchの動きが不一致の場合、 ∇P_i は別々の2次元平面上に現れ、行列 $\mathbf{M}_{12}$ のランクは $\text{rank}(\mathbf{M}_{12}) = 3$ となる (図3(b))。ここで、行列 $\mathbf{M}$ の 2×2 の左上の行列 $\mathbf{M}^\diamond$ を式 (8) より求める。

$$\mathbf{M}^\diamond = \begin{bmatrix} \sum P_x^2 & \sum P_x P_y \\ \sum P_y P_x & \sum P_y^2 \end{bmatrix} \quad (8)$$

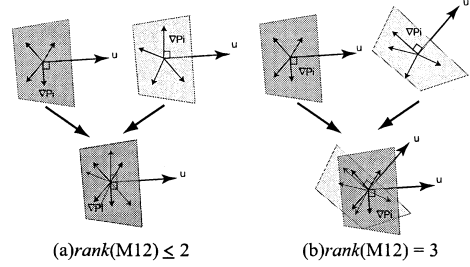


図3: 勾配とランクの関係

ST-patchの空間の特性と式 (8) で求められる行列 $\mathbf{M}^\diamond$ を用いて、ランク制約によるST-patchの動きの状態判定を行う。ST-patchの空間の特性から、単一の動きのST-patchは、 $\text{rank}(\mathbf{M}) = \text{rank}(\mathbf{M}^\diamond)$ のようなランクの状態となる。そして、複数の動きを含むST-patchでは、1つの行と列だけが $\mathbf{M}$ から $\mathbf{M}^\diamond$ へ変換されているので、ランクの違いが1となる。よって、式 (9) のような関係が成り立つ。

$$\begin{aligned} \Delta r &= \text{rank}(\mathbf{M}) - \text{rank}(\mathbf{M}^\diamond) \\ &= \begin{cases} 0 : \text{single motion} \\ 1 : \text{multiple motions} \end{cases} \quad (9) \end{aligned}$$

同様に、2つの異なるST-patchにおいても、動きが互いに一致しているかを式 (9) により求めることができる。

2.1.4 連続的なランク増加量 Δr

ランク増加量 Δr は、動きが一致か不一致かの2通りしか求めることができず、動きがどれだけ類似しているかを求めることができない。そこで、行列 $\mathbf{M}$ と行列 $\mathbf{M}^\diamond$ の固有値を用いて連続的なランク増加量 Δr を定義し、2つの異なるST-patchの類似度を求める。式 (4) に示した行列 $\mathbf{M}$ の固有値 $\lambda_1 \geq \lambda_2 \geq \lambda_3$ と、式 (8) に示した行列 $\mathbf{M}^\diamond$ の固有値 $\lambda_1^\diamond \geq \lambda_2^\diamond \geq \lambda_3^\diamond$ から、連続的なランク増加量 Δr を次式のように定義する。

$$\Delta r = \frac{\lambda_2 \cdot \lambda_3}{\lambda_1^\diamond \cdot \lambda_2^\diamond} \quad (0 \leq \Delta r \leq 1) \quad (10)$$

ここで求められた連続的なランク増加量 Δr の値は、動きが一定のとき $\Delta r \approx 0$ となり、動きが複数のとき $\Delta r \approx 1$ となる。

2.1.5 2つのST-Patchの相互関係

式 (10) により求められた連続的なランク増加量 Δr から、2つのST-patch P1とP2の動きの類似度

を求めることができる。P1 の連続的なランク増加量を Δr_1 、P2 の連続的なランク増加量を Δr_2 、P1 と P2 をまとめた連続的なランク増加量を Δr_{12} とすると、式 (11) により動きの類似度が求められる。

$$m_{12} = \frac{\min(\Delta r_1, \Delta r_2)}{\Delta r_{12}}$$

ここで求められた類似度 m_{12} は、 $m_{12} \approx 1$ の場合は類似度が高く、 $m_{12} \approx 0$ の場合は類似度が低いことを意味する。

2.1.6 ST-patch による大域的な変化

類似度 m_{12} を映像の大域的变化に基づく特徴量とする。本研究では、画像全体を 1 つの ST-patch とし、フレーム間の類似度 m_{12} を以下の処理により算出する。

Step1 画像サイズを 64×48 ピクセルにダウンサンプリングする

Step2 ダウンサンプリング画像に対して平滑化を施す (図 4)。対象フレームの前後の 3 フレーム分をそれぞれ ST-patch P1 と P2 とする

Step3 P1 の連続的なランク増加量 Δr_1 、P2 の連続的なランク増加量 Δr_2 と、P1、P2 間の連続的なランク増加量 Δr_{12} を算出

Step4 Δr_1 、 Δr_2 、 Δr_{12} より、類似度 m_{12} を算出
類似度 m_{12} より、時系列での大域的な変化を捉えることが可能となる。図 5 に、ある映像における

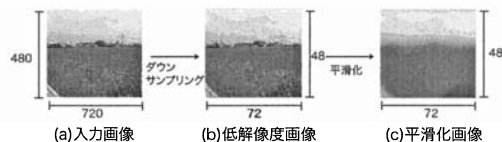


図 4: ST-patch の算出

ST-patch により算出された動きの類似度の変化を示す。図 5 から、ST-patch はカメラワークやショット切り替えがあるところで類似度 m_{12} が小さくなっており、映像の大きな変化を捉えていることがわかる。しかし、ST-patch のみでは画像全体の大きな変化すべてに反応するため、ショットの検出とカメラワークの判別が不可能であるという問題がある。

2.2 SIFT による特徴量抽出

カメラワークとショットの違いを捉えるために、局所的な変化に着目する。本研究では、SIFT を用いてフレーム間での対応点を求め、局所的な変化を求める。SIFT は、Lowe により提案された回転・スケール変化・照明変化に不変な特徴点 (以下、キーポイントと呼ぶ) の検出と特徴量の抽出を行う手法 [6] である。キーポイントの検出はスケールスペース極値検出、キーポイントのローカライズからなり、特徴量の抽出はオリエンテーション割り当てと、SIFT Descriptor による特徴量の記述からなる。

2.2.1 スケールスペース極値検出

SIFT は、スケール変化に対する不変性を得るため、対象とするキーポイントの特徴を最も良く表現することができる近傍領域の範囲を自動的に決定する。範囲を決定するパラメータをスケールといい、キーポイント検出における最適なスケールを求める。まず、スケール (σ) のガウス関数 $G(x, y, \sigma)$ と入力画像 $I(x, y)$ を畳み込んだ平滑化画像 $L(x, y, \sigma)$ を次式により得る。

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (11)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (12)$$

スケールの異なる平滑化画像の差分をとることを Difference-of-Gaussian (DOG) といい、 $D(x, y, \sigma)$ で表す。DOG 画像は式 (13) により得る。

$$D_i(x, y, \sigma) = L(x, y, \sigma_{i+1}) - L(x, y, \sigma_i) \quad (13)$$

次に、図 6 に示すように、得られた DOG 画像から極値の検出を行う。注目画素 (図中 $\times$ 印) が 3 次的に隣接する 26 近傍において極値である場合、注目画素をキーポイント候補点として検出する。

2.2.2 キーポイントのローカライズ

検出された極値は、スケール変化に不変なキーポイント候補点であるが、開口問題が発生しやすい点も含まれている。そこで、コントラストの低い点やエッジ上の点を削除し、より頑健な点をキーポイントとして選択する。

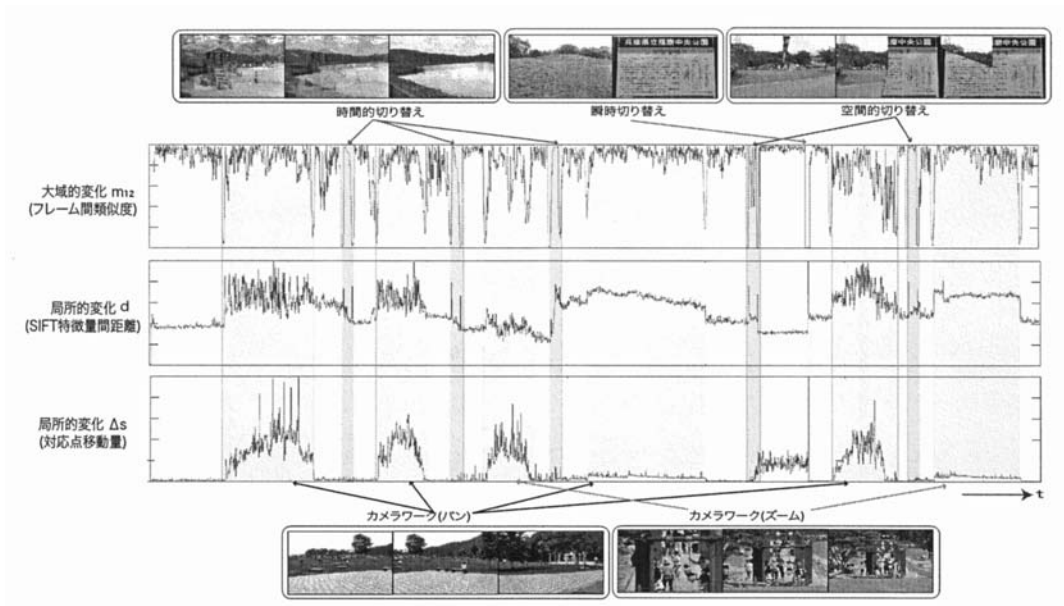


図 5: 映像の特殊効果に伴う各特徴量の変化

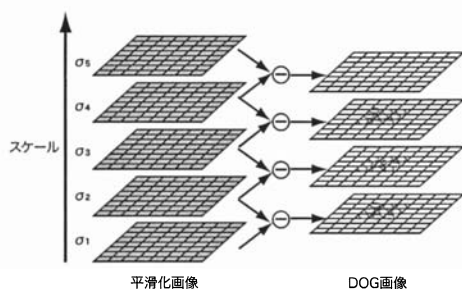


図 6: 極値検出の流れ

2.2.3 オリエンテーション割り当て

検出された各キーポイントに対してオリエンテーションを割り当てる。キーポイントのオリエンテーションを求めるために、キーポイントが検出された平滑化画像の各画素の勾配強度 $m(x, y)$ と、勾配方向 $\theta(x, y)$ を次式により求める。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (14)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (15)$$

$$\begin{cases} f_x(x, y) = L(x+1, y) - L(x-1, y) \\ f_y(x, y) = L(x, y+1) - L(x, y-1) \end{cases} \quad (16)$$

求めた勾配強度と勾配方向を用いて図 7 に示するような 36 方向の重み付方向ヒストグラムを作成する。得られたヒストグラムの最大頻度の 80% 以上になる方向をキーポイントのオリエンテーションとして割り当てる。

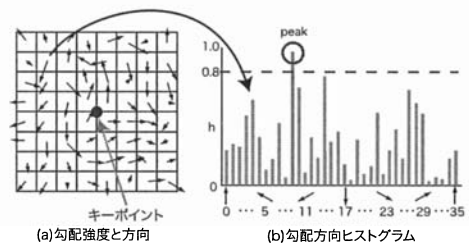


図 7: オリエンテーション割り当て

2.2.4 SIFT Descriptor による特徴量抽出

検出されたキーポイントに対し、SIFT Descriptor による特徴量の記述を行う。キーポイントの周辺領域を、割り当てられた方向を基準とした軸に回転させる (図 8)。このように、キーポイントの向きをあわせて特徴量を記述するため、回転に対する不変性が得られる。特徴量の記述の際には、ガウス窓を用

いて中心付近の特徴がより考慮されるように重み付けをする。このガウス窓の大きさは、キーポイントが検出された DOG 画像の平滑化時のスケールによって決定される。画像サイズが2倍になると、比例してスケールも2倍となるため、構造的に同一の範囲における特徴量の記述が行われる。これにより、スケール変化に対する不変性が得られる。次に、図8に示すように、キーポイントの周辺領域を分割した小領域ごとに方向ヒストグラムを作成する。領域を 4×4 の16に分割し、それぞれに8方向のヒストグラムを作成するため、キーポイントは128次元のベクトルの特徴量を持つことになる。

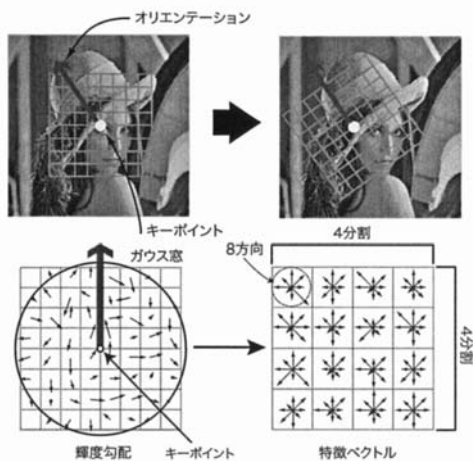


図 8: 特徴量記述

2.2.5 対応点の計算

SIFTにより検出した前後フレームのキーポイントから、ショット分割に用いる特徴量として、特徴量間距離と移動量を求める。フレーム t のあるキーポイントの128次元特徴ベクトルを $\mathbf{v}^t = (v_1^t, v_2^t, \dots, v_{128}^t)$, $t-1$ におけるキーポイントの特徴ベクトルを $\mathbf{v}^{t-1} = (v_1^{t-1}, v_2^{t-1}, \dots, v_{128}^{t-1})$ とした場合、キーポイント間のユークリッド距離を次式により求める。

$$d(\mathbf{v}^t, \mathbf{v}^{t-1}) = \sum_{i=1}^{128} \sqrt{(v_i^t - v_i^{t-1})^2} \quad (17)$$

このような距離計算を図9に示すように、全てのキーポイントの組み合わせに対して行い、最小距離がしきい値以下であればそのキーポイント同士を対応点とする。全対応点の特徴量間距離の平均値を一つ

目の局所的变化に基づく特徴量とする。このとき、対応点 (x_t, y_t) と (x_{t-1}, y_{t-1}) のフレーム間における移動量 Δs を算出する。

$$\Delta s = \sqrt{(x_{t-1} - x_t)^2 + (y_{t-1} - y_t)^2} \quad (18)$$

特徴量間距離と同様に、1フレーム間でとれた全対応点の移動量の平均を二つ目の局所的变化に基づく特徴量とする。

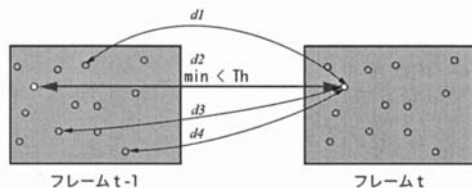


図 9: 全探索による対応点マッチング

このようにして算出した SIFT に基づく2つの特徴量を図5に示す。カメラワーク発生時には、SIFT 特徴量間距離、対応点の移動量ともに同じような特徴量の変化であるが、特殊切り替えにおいては変化に違いが見られる。この違いに着目することで、特殊切り替えとカメラワークの判別が可能となる。しかし、変化の激しいカメラワークや撮影対象の影響で、特徴量が激しく変動するため、正確にショットを検出できない可能性がある。そこで、ST-patch に基づく大域的变化と SIFT に基づく局所的变化の2種類の同時変化に着目する。

3 特徴の共起表現と AdaBoost

特殊ショットを正確に検出するために、2つの特徴を同時に捉える。2章で求めた大域的变化に基づく特徴量であるフレーム間の動きの類似度 m_{12} と、局所的变化に基づく特徴量である対応点の特徴量間距離 d 、フレーム間の移動量 Δs の3つの特徴からショット分割を行う。提案手法では、異なる3つの特徴量間の共起を表現するために、三田らが提案した特徴の共起表現法 [7] を用い、全特徴の同時確率によりショット分割判定を行う。

3.1 特徴量間の共起

まず、全特徴の同時確率を用いるために、特徴量の2値化を行う。ある特徴量 v_i について、式 (19)

に示すように1と0により2値化を行う。2値化のための符号 s はベイズの定理 (式 (20)) により決定される。

$$s = \begin{cases} 1 & P(C_p|v_i) > P(C_n|v_i) \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

$$P(C_k|v_i) = \frac{P(v_i|C_k)P(C_k)}{P(v_i)} \quad (k = p, n) \quad (20)$$

ここで、 $P(C_k|v_i)$ 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。

次に、2値化された各特徴量を組み合わせた新たな特徴量 c を算出することで、特徴量間の共起を表現する。提案手法では、大域的变化の特徴としてある時刻 t での前後 ST-patch 間の動きの類似度 m_{12} 、局所的变化の特徴として対応点の特徴量間距離 d と、対応点の1フレーム間の移動量 Δs の3つの特徴量から1つずつ2値化したものを用いる。例えば $m_{12} = 0$ 、 $d = 1$ 、 $\Delta s = 1$ が観測されたとき、組み合わせ特徴量 c は次式のように算出される。

$$c = (011)_2 = 3 \quad (21)$$

このように複数の特徴量を2値の組み合わせによる新たな1つの特徴量として表現することで、異なる特徴量間の共起性を表現することができる。

3.2 フレーム間の共起

ディゾルプ等のショット切り替え効果は、徐々に変化するため、判定フレームの前後フレーム間における特徴量の共起性に着目する。図10に示すように、判定フレーム $t(001)$ と時刻 t より前 N フレーム間から1フレーム (101) と後 N フレーム内から1フレーム (010) の計3フレーム分の特徴量を c' とする。従って、特徴量 $c' = (010001101)_2$ は9個の2値の組み合わせにより表現されるため、 $2^9 = 512$ パターンの特徴量となる。さらに、特徴量 c の選択は、例えば $N=15$ のとき ${}_{31}C_3 = 4495$ パターンから AdaBoost の学習時に識別エラーが最小となるように、自動的に選択される。

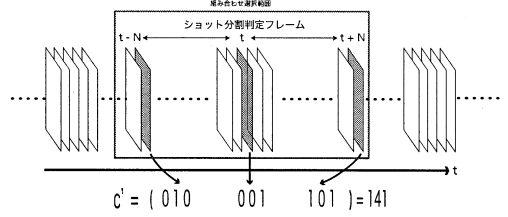


図 10: 前後フレームとの共起表現

3.3 AdaBoost 識別器の構築

特徴量 c' を用いて、AdaBoost によりショット分割判定を行うための識別器を構築する。AdaBoost によって最終的に構築される識別器は、 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。

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (22)$$

ここで、 α_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。

$$h_t(x) = \begin{cases} +1 & P_t(y = +1|c') > P_t(y = -1|c') \\ -1 & \text{otherwise} \end{cases} \quad (23)$$

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、学習サンプルの重み $D_t(i)$ 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。

4 評価実験

瞬時切り替え、特殊切り替え、カメラワークが発生する TV 番組の映像を用いた評価実験を行う。

4.1 実験概要

表1に学習用データと評価実験用データの概要を示す。TV 映像は約 180 秒の映像が 10 種類である。学習用には、ニュース映像 (スタジオのみ)、自然の多い風景映像、移動体の多い風景映像、アニメーション (CG)、ドラマの 5 種類を用いる。評価用には、ニュース映像 (中継映像)、車両が多く横切る映

像、自然の多い風景映像、ドラマ映像(字幕あり)、CMの5種類を用いる。表1に、学習用、評価用に用いるデータの内訳を示す。ショット分割の精度の

表 1: 映像データベース [フレーム]

	ショット			非ショット		合計
	瞬時	時間的	空間的	カメラ	固定	
学習用	62	414	243	802	1289	2810
評価用	123	971	834	1158	2290	5376

評価には、フレーム単位で正誤を判定し、正解検出率、誤検出率を次式から求める。

$$\text{正解検出率} = \frac{\text{検出した正解数}}{\text{検出されるべき正解数}} \quad (24)$$

$$\text{誤検出率} = \frac{\text{誤検出した数}}{\text{実際に検出した数}} \quad (25)$$

4.2 実験結果

前後フレーム幅 N を変化させた時の実験結果を図 11 に示す。共起を用いない AdaBoost に対して、 N を 10 としたとき、共起を用いることにより最大で約 19% 向上させることができた。また、ショット

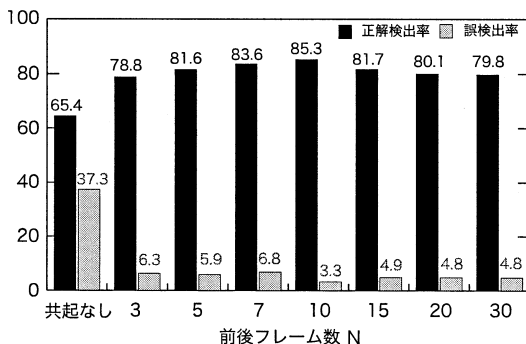


図 11: 前後フレームと識別結果の関係

分割点となるフレームは、前後フレームとの相関があるため、前後 10 フレームからフレーム間の共起を表現することが良いことがわかった。以上より、ショット検出対象となるフレームとその前後フレームの共起と、大域的变化と局所的变化の共起を表現する本手法の有効性が確認できた。

5 まとめ

本稿では、Space-Time Patch を用いて抽出した映像の大域的变化に基づく特徴量と、SIFT を用いて抽出した局所的变化に基づく特徴量の共起性に着目した映像のショット分割の手法を提案した。提案手法は大域的な変化と局所的な変化を同時に観測することで、高精度なショット分割点の検出が可能となった。今後は、より多くの映像の変化に適応した識別器の構築を行う予定である。

参考文献

- [1] Nguyen Huu Bach, 篠田浩一, 古井貞熙, “隠れマルコフモデルを用いた野球放送の自動的インデキシング”, 画像の認識・理解シンポジウム (MIRU2005), pp. 1113-1120, 2005.
- [2] 高野求, 三浦宏一, 浜田玲子, 井手一郎, 坂井修一, 田中英彦, “テキスト情報からの制約に基づく料理画像中の物体検出”, 第 17 回人工知能学会全国大会, no.2C1-02, June, 2003.
- [3] 長坂晃朗, 田中譲, “カラービデオ映像における自動索引付け法と物体探索法”, 情報学会論文誌, vol. 33, no. 4, pp. 543-550, 1992.
- [4] 谷口行信, 外村佳伸, 浜田洋, “映像ショット切替え検出法とその映像アクセスインターフェイスへの応用”, 電子情報通信学会論文誌 (D-II), Vol. J79-D-II, No. 4, pp. 538-546, 1996.
- [5] E. Shechtman, M. Irani, “Space-Time Behavior Based Correlation”, in Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, June 2005, pp. 405-412.
- [6] David G. Lowe, “Distinctive image feature from scale-invariant keypoints”, Journal of Computer Vision, 60, 2, pp. 91-110, 2004.
- [7] 三田雄志, 金子敏充, 堀修, “顔検出に適した共起に基づく Joint Haar-like 特徴”, 電子情報通信学会論文誌, Vol. J89-D, No. 8, pp. 1791-1801, 2006.
- [8] 坂江伸悟, 林義文, 熊野雅仁, 有木康雄, 春藤憲司, 塚田清志, “素材映像中のカット点検出と色調によるシーン検出”, 電気関係学会関西支部連合大会, p373, 2001.
- [9] 大辻清太, 外村佳伸, 大庭有二, “突出検出フィルタを用いた映像カット点検出法”, 電子情報通信学会論文誌, Vol. J77-D, No. 3, pp. 519-528, 1994.
- [10] 熊野雅仁, 天野美紀, 有木康雄, 上原邦昭, “映像文法に基づいた実時間使用可能ショット識別による撮影ナビゲーション”, 電子情報通信学会技術研究報告, パターン認識・メディア理解 (PRMU), Vol. 104, No. 369, pp. 1-6, 2004.
- [11] 高橋正樹, 西山裕之, 溝口文雄, “カメラロボットの協調動作による人物追跡システムの設計”, The 17th Annual Conference of the Japanese Society for Artificial Intelligence, 3D1-09, 2003.