

ワイプ・DVEのパターン非依存な検出手法

A Pattern-Independent Detection of Wipes and DVE's in Video Contents

岩元 浩太†
Kota Iwamoto

平田 恭二†
Kyoji Hirata

1. はじめに

放送番組などの映像に対する効率的なアクセスを実現するためには、映像を意味的に重要な区切りで分割し、時間軸上で構造化することが有効である。放送番組などでは、意味的に重要な区切りなどの特に注目を引きつけたい箇所にはワイプや Digital Video Effect(DVE)などの特殊効果による漸次的な映像切り替えが使用されることが多い。例えばワイプは、映画やドラマにおける場面の転換点に多用される。また DVE は、ニュース番組におけるニュース項目やバラエティ番組におけるコーナーなどのトピックの変化点に多用される。このため、ワイプや DVE を検出することにより、映像の意味的な構造化を行ううえで重要な情報を抽出することができる。そこで本稿では映像におけるワイプ・DVE を検出する手法について扱う。

ワイプ・DVE を検出する手法として、ワイプ・DVE が漸次的な映像変化であることを利用し、フレーム間の特徴量の比較に基づいて映像が緩やかに変化する区間を検出する手法が提案されている[1][2]。Zhang ら[1]は、隣接フレーム間の特徴量の差分値に対して大小 2 つの閾値を用いて、小さな差分値が連続し、かつ蓄積された差分値が大きくなる区間を検出する Twin-Comparison 法を提案している。また中島ら[2]はワイプを検出する手法として、同様に隣接フレーム間の特徴量の差分値を用い、小さな差分値が連続し、かつ前後の区間でより小さな差分値が連続する区間を検出する手法を提案している。これらの手法はワイプ・DVE のパターンに依存せず汎用的に適用できるものの、単純な特徴量の比較に基づくため、パン・ズームなどのカメラモーションや映像中のオブジェクトの動きなどのワイプ・DVE 以外の映像変化との区別ができない。このため、過剰検出が大幅に生じるという問題がある。過剰検出を低減するために、動きベクトルを用いてカメラモーション(パン・ズーム)などのパターンをモデル化し、モデルを用いてワイプ・DVE 以外の映像変化を識別する手法も検討されている[1][3]。しかしモデル化が可能な映像変化のパターンは限られ、ワイプ・DVE 以外の全ての映像変化を網羅するモデルを生成するのは不可能である。

一方、ワイプ・DVE のパターンごとに個別のモデルやルールを用いて検出する手法も提案されている[4][5][6]。河合ら[4]は DVE を検出する手法として、予めパターンごとに色や動き情報を用いたモデルを生成し、モデルとの類似度に基づき DVE を検出する手法を提案している。谷口ら[5]はワイプを検出する手法として、予めパターンごとに複数フレーム間隔の特徴量差分値のテンプレートを用意し、テンプレートとの類似度に基づきワイプを検出する手法を提案している。また Zabih ら[6]はワイプを検出する手法として、予めパターンごとに出現・消失エッジの空間分布の

時間変化をルール化し、そのルールを用いてワイプを検出する手法を提案している。これらの手法では、ワイプ・DVE のパターンごとにモデルやルールを生成する必要がある。このため放送番組などに含まれる無数のパターンのワイプ・DVE に対応することができないという問題がある。

本稿では、ワイプ・DVE のフレーム区間に共存する切り替え前後の映像の表示領域を分ける境界線に着目し、ワイプ・DVE をパターンに依存せず汎用的に、かつ他の映像変化の過剰検出を抑制して検出する手法を提案する。

2. ワイプと DVE

ワイプや DVE は切り替え前後の映像が空間的に共存しながら、その占有割合を徐々に変化させて切り替わる映像切り替えである。ワイプでは切り替え前後の映像の位置は固定で、それぞれの映像の表示領域を徐々に変化させて切り替わる(図 1 参照)。左右または上下に切り替わるパターンが多いが、斜めに切り替わるパターンなどもある。DVE では一方の映像の位置は固定で、その上に重畳される形で他方の映像が射影変換(平行移動・拡大縮小・回転)や捻りなどの画像変換を伴いながら出現または消失して切り替わる(図 2 参照)。射影変換や捻りなどの画像変換のパラメータの設定により無数のパターンが生成される。

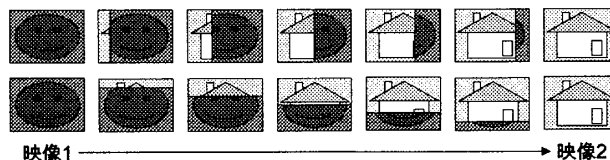


図 1: ワイプの例

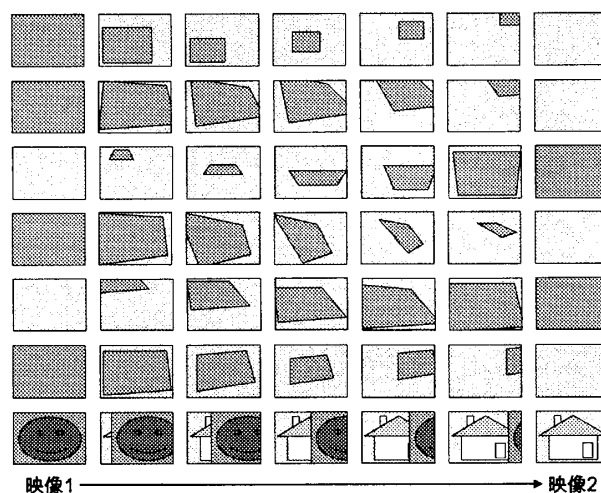


図 2: DVE の例

† NEC メディア情報研究所

Media and Information Research Laboratories, NEC Corp.

3. 提案手法

3.1 概要

ワイプ・DVE を構成するフレームには切り替え前後の映像が空間的に共存し、それぞれの表示領域の間に境界線(主に直線)を形成する¹。さらに境界線は以下 2 つのいずれかの性質を有する。

タイプ A: 1本の境界線がフレーム間を連続的に移動する

タイプ B: 複数の境界線が四角形の画像枠を形成する²

ワイプや一部の DVE(スライドインなど)はタイプ A の性質を有する(図 3 参照)。一般的な DVE はタイプ B の性質を有する(図 4 参照)。カメラモーションなどの他の映像変化はこれらの性質を有しない。提案手法では、フレームから境界線を抽出し、境界線が上記の性質を有するフレーム区間をワイプ・DVE として検出する。

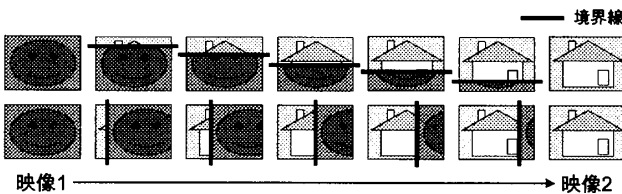


図 3: タイプ A の例

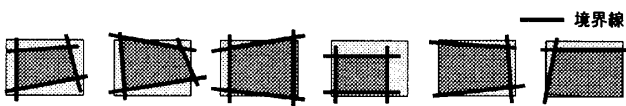


図 4: タイプ B の例

3.2 処理手順

映像からワイプ・DVE のフレーム区間を検出する処理手順を図 5 に示す。まず、映像をワイプ・DVE が含まれる候補となる区間に絞り込む。次に、絞り込まれた区間の各フレームから境界線を抽出する。続いて、境界線がタイプ A またはタイプ B の性質を有するフレーム区間を検出する。最後に、検出されたフレーム区間に含まれる境界線の境界線性の検証により過剰検出を除去する。

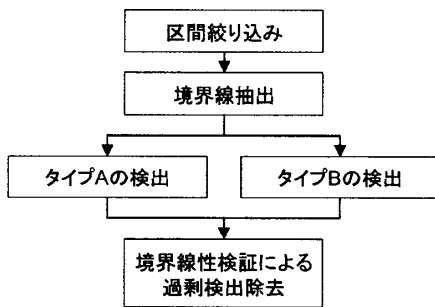


図 5: 処理手順

¹ 白線などの明示的な境界線もあるが、映像の重畳によって自然に形成される境界線が多い。

² それぞれの境界線はタイプ A の性質を有するが、より複雑な動きをするため、タイプ A の性質を用いた検出が困難である。

3.2.1 区間絞り込み

フレーム間の特徴量の差分値に基づいて、映像が緩やかに変化する区間を検出することでワイプ・DVE が含まれる候補となる区間に絞り込む。フレームの特徴量としては HSV 表色系による色ヒストグラム(HSV ヒストグラム)を用いる。Zhang らの Twin-Comparison 法[1]と同様に大小 2 つの閾値 th_1 と th_2 ($th_1 > th_2$) を設け、近接するフレーム間の特徴量の差分値が連続的に th_2 以上であり、かつ区間前後のフレーム間の差分値が th_1 以上となる区間を検出する。ただしカットを含む区間は除外する。また絞り込みが目的のため、ワイプ・DVE を含む区間の漏れがないように閾値は甘めに設定する。

3.2.2 境界線抽出

各フレームからの境界線の抽出は、(1)境界線の候補となる画素の検出、(2)ハフ変換による直線抽出、(3)境界線の選定、の 3 段階の処理により行う。各段階の処理結果の例を図 6 に示す。

移動する境界線を抽出するために、境界線の候補画素の検出ではまずフレーム間差分を行う。フレーム間差分値の大きい画素に対して Canny のエッジ検出法[7]により³エッジを検出し、検出されたエッジを候補画素とする。

次に候補画素を入力としてハフ変換による直線抽出を行う。ハフ変換では直線式を $\rho = x \cos \theta + y \sin \theta$ と表し、 (ρ, θ) の 2 次元パラメータで記述する。ここで ρ は原点から直線へおろした垂線の長さ、 θ は垂線と水平軸とのなす角度である。各候補画素 (x_i, y_i) に対して (x_i, y_i) を通る全ての直線 $\rho = x_i \cos \theta + y_i \sin \theta$ を $\rho - \theta$ の 2 次元パラメータ空間に投票する。投票されたパラメータ空間 $\rho - \theta$ に対して 2 次元のウィンドウをかけていき、ウィンドウ内の投票のピーク値がある閾値以上となる場合に、対応する直線を抽出する。

ハフ変換で抽出された直線から、直線を構成する画素(ハフ変換でその直線に投票した画素)の勾配方向の検証に基づき、境界線として適切なものを選定する。理想的な条件下において、境界線を構成する画素の勾配方向は境界線に対して垂直になる。そこで、抽出された直線に投票した画素の勾配方向が直線の垂直方向と近くなる直線を境界線として選定する。具体的には抽出された直線 (ρ, θ) に投票した画素の勾配方向 (g_1, g_2, \dots, g_N) (ただし N は直線に投票した画素数)を求める。直線の垂直方向 θ と画素の勾配方向 g_i との差の 2 乗平均 MSD を以下の式で算出する。

$$MSD = \frac{1}{N} \sum_{n=1}^N (\theta - g_n)^2 \quad (1)$$

ただし $\theta = \theta + \pi$ とし、 $0 \leq \theta - g_n \leq \pi/2$ とする。MSD がある閾値以下となる場合に、その直線を境界線として選定する。

³ ただし計算コストの関係上、2 つの閾値を用いて強いエッジに隣接する弱いエッジを検出する処理は省く。

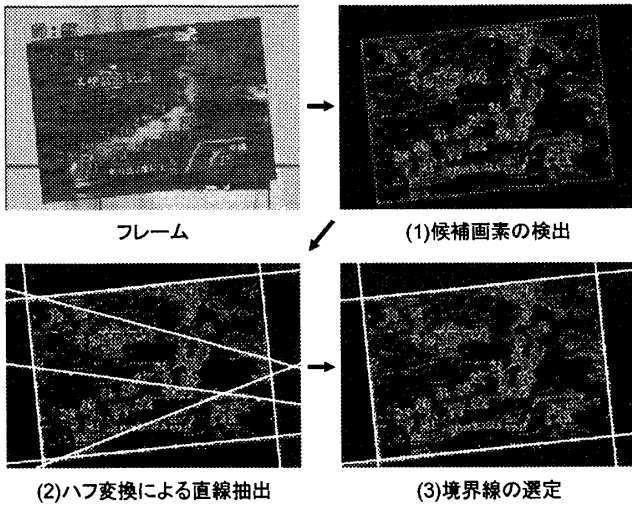


図 6: 境界線抽出の例

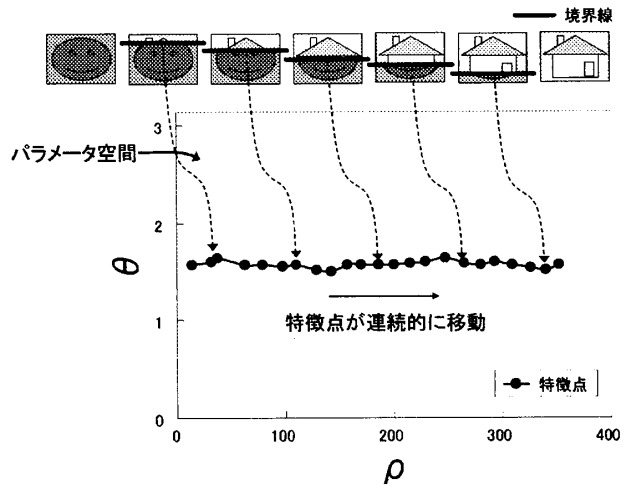


図 7: パラメータ空間での特徴点の連続移動

3.2.3 タイプ A の検出

1本の境界線が連続的に移動するフレーム区間をタイプ A の性質を有する区間として検出する。このために各フレームから抽出された境界線 (ρ, θ) を 2 次元パラメータ空間 $\rho-\theta$ の特徴点として表し、パラメータ空間内で特徴点が時間と共に連続的に移動していくフレーム区間を検出する(図 7 参照)。特徴点が連続的に移動するフレーム区間を検出するために、パラメータ空間内で逐次的に特徴点を追跡し、隣接フレーム間の特徴点の連続性を判定していく。逐次的な特徴点の追跡では、過去フレームの特徴点から現フレームの特徴点の予測点を算出する。現フレーム i の予測点 $(\hat{\rho}_i, \hat{\theta}_i)$ は、過去 2 フレームの特徴点 $(\rho_{i-1}, \theta_{i-1})$ 、 $(\rho_{i-2}, \theta_{i-2})$ を用いて、

$$\hat{\rho}_i = 2\rho_{i-1} - \rho_{i-2} \quad (2)$$

$$\hat{\theta}_i = \theta_{i-1} \quad (3)$$

と算出する。予測点 $(\hat{\rho}_i, \hat{\theta}_i)$ と実際に現フレームで抽出された特徴点 (ρ_i, θ_i) との誤差を算出し、誤差がある閾値以内であれば連続であると判定する。こうして逐次的に特徴点を追跡していき、連続的な特徴点系列の ρ の通算の移動量がある閾値以上となる場合に、そのフレーム区間の前後のフレームの特徴量(HSV ヒストグラム)の差分値を算出する。その差分値がある閾値以上となる場合に、当該フレーム区間をタイプ A の性質を有する区間として検出する。

3.2.4 タイプ B の検出

複数の境界線がフレーム内で四角形を形成するフレーム区間をタイプ B の性質を有する区間として検出する。ただし、四角形の画像枠の 4 辺全てがフレーム内に収まるとは限らないため(図 4 参照)、四角形を構成する隣り合う 2 辺の境界線を検出の対象とする。各フレームで抽出された境界線の中からフレームの四隅を中心とした探索領域内(図 8 参照)で交差し、かつ交差の角度がある範囲内である境界線のペアを検出する。検出された境界線のペアに対して、交点を出発点として両境界線に沿ってエッジを追跡していく。連続するエッジの長さがある閾値以上となる場合に、当該フレームがタイプ B の性質を有すると判定する。タイプ B

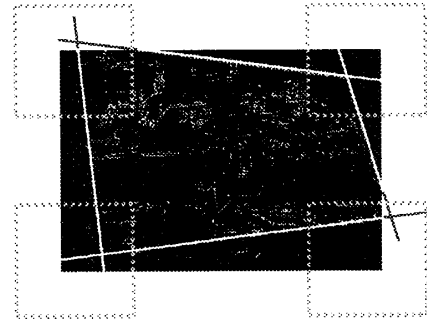


図 8: 境界線ペアの探索領域(点線内部)

の性質を有するフレームが連続する⁴フレーム区間を、タイプ B の性質を有するフレーム区間として検出する。

3.2.5 境界線性検証による過剰検出除去

検出されたタイプ A, B の性質を有するフレーム区間の中には、境界線に類似した直線状のエッジ(例えば柱など)がカメラモーションなどにより移動するフレーム区間を誤って検出したものが含まれる場合がある。そこで検出されたフレーム区間の各フレームの境界線上の動きベクトルの性質を用いた境界線性の検証に基づき、こうした過剰検出を除去する⁵。直線状のエッジがカメラモーションなどによって移動した場合、一般的にその直線上の動きベクトルの方向と大きさは揃う。一方、ワイプ・DVE における境界線は 2 つの動いている映像の境界であるため、境界線上の動きベクトルはばらつきが大きい(図 9 参照)。この性質を利用すると、境界線上の動きベクトルのばらつき度合いが小さい場合は過剰検出と判断できる。具体的には、まず境界線上の等間隔にサンプリングされた点において、ブロックマッチング法により動きベクトル $(\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_N)$ (ただし N は動きベクトルの数を求め、その平均ベクトルを $\bar{\mathbf{m}}$ と

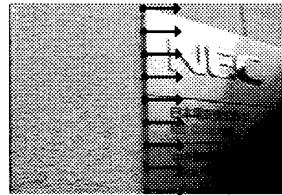
⁴ 画像枠がある大きさ以上となるフレームは DVE の区間の一部(開始直後または終了間際)であるため、検出に必要な連続フレーム数は少なめに設定する(本稿の実験では 1 フレームとしている)。

⁵ この処理は境界線を検出する段階で行うこともできるが、動きベクトル算出の計算コストが大きいため、この段階で行う。

する。境界線上の動きベクトルのばらつき度合い V は、平均ベクトル $\bar{\mathbf{m}}$ からの誤差(ユークリッド二乗距離)の平均値として求める。

$$V = \frac{1}{N} \sum_{n=1}^N |\bar{\mathbf{m}} - \mathbf{m}_n|^2 \quad (4)$$

検出されたタイプ A, B のフレーム区間の各境界線に対して動きベクトルのばらつき度合い V を求め、フレーム区間全体での平均値を算出する。ばらつき度合いの平均値がある閾値以下となる場合に、検出されたフレーム区間を除去する。こうして残ったフレーム区間をワイプ・DVEとする。



(a)境界線の場合 →動きベクトルがばらつく
(b)直線状のエッジの移動の場合 →動きベクトルが揃う

図9: 境界線上の動きベクトルの性質

4. 評価実験

ニュース番組の映像を用いて、提案手法および従来手法についてワイプ・DVEの検出精度を評価した。

4.1 実験条件

実験で使用した映像の諸元を表1に示す。合計15時間20分のニュース番組の映像に含まれる計237箇所のワイプ・DVEを検出する。なお、映像は352×240画素サイズのMPEG2(30fps, 3Mbps)で符号化されている。

提案手法で用いる閾値などの各種パラメータは、約10時間分の学習用の映像を用いて設定した。従来手法としては、汎用的に適用できるZhangらのTwin-Comparison法[1]を採用した。Twin-Comparison法ではフレームの特徴量としてRGBヒストグラムを用い、パラメータ(2つの閾値)は同じ学習用の映像を用いて設定した。また提案手法、従来手法ともに、短い時間間隔で連続的に過剰検出が発生するのを抑えるために、固定フレーム長のウィンドウ区間において出力されるワイプ・DVEが1つになるようにフィルタリングした。

表1: 映像諸元

映像種別	ニュース番組(CM区間は削除)
映像本数	41
映像総時間	15時間20分
ワイプ・DVEの数	237(タイプA: 149, タイプB: 88)

4.2 結果

提案手法と従来手法のそれぞれに対して、ワイプ・DVEが検出された否かを評価し、ワイプ・DVEの検出の再現率と適合率を測定した。それぞれ以下の式で算出される。

$$\text{再現率(\%)} = \frac{\text{正しく検出されたワイプ・DVEの数}}{\text{ワイプ・DVEの総数}} \times 100 \quad (5)$$

$$\text{適合率(\%)} = \frac{\text{正しく検出されたワイプ・DVEの数}}{\text{検出の総数}} \times 100 \quad (6)$$

評価実験の結果を表2に示す。提案手法は再現率91.6%、適合率54.7%を達成した。提案手法は従来手法と比較して再現率を30.0%、適合率を43.0%改善した。ワイプ・DVEのフレーム区間が有する境界線に着目することで、ワイプ・DVEをパターンに依存せず汎用的に検出することができた。また、特にカメラモーションやオブジェクトの動きによる過剰検出を大幅に抑えることができた。

表2: 実験結果

手法		再現率	適合率
従来手法		61.6%	11.7%
提案手法	タイプA	93.3%	60.4%
	タイプB	88.6%	46.7%
	総合	91.6%	54.7%

5. まとめ

ワイプとDVEによる映像切り替えをパターンに依存せず汎用的に検出する手法を提案した。提案手法では、ワイプ・DVEのフレーム区間に共存する切り替え前後の映像の表示領域を分ける境界線に着目し、各フレームから境界線を抽出する。境界線が連続的に移動するフレーム区間、および複数の境界線が四角形の画像枠を形成するフレーム区間を、ワイプ・DVEのフレーム区間として検出する。提案手法は再現率91.6%、適合率54.7%を達成し、従来手法と比較して再現率を30.0%、適合率を43.0%改善した。

参考文献

- [1] H. J. Zhang, A. Kankanhalli, S. W. Smoliar, "Automatic Partitioning of Full-Motion Video", *Multimedia Systems* 1, pp. 119-128, 1993.
- [2] 中島康之, 氏原清乃, 米山暁夫, "部分復号を用いたMPEGデータからのカット点検出", *信学論D-II*, Vol. J81-D-II, No. 7, pp. 1564-1575, July 1998.
- [3] I. Koprinska, S. Carrato, "Detecting and Classifying Video Shot Boundaries in MPEG Compressed Sequences", *Proc. of EUSIPCO98*, pp. 1729-1732, Sept. 1998.
- [4] 河合吉彦, 馬場口登, 北橋忠宏, "放送型スポーツ映像におけるデジタルビデオ効果に着目したリプレイシーン検出の一手法", *信学論D-II*, Vol. J84-D-II, No. 2, pp. 204-212, February 2001.
- [5] 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フェースへの応用", *信学論D-II*, Vol. J79-D-II, No. 4, pp. 538-546, April 1996.
- [6] Ramin Zabih, Justin Miller, Kevin Mai, "A Feature-Based Algorithm for Detecting and Classifying Production Effects", *Multimedia Systems* 7, pp. 119-128, 1999.
- [7] J. Canny, "A Computational Approach to Edge Detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, pp. 679-698, November 1986.