

于川[†], 佐藤 真一[‡], 浜田 喬^{†‡}Chuan Yu[†], Shin'ichi Satoh[‡] and Takashi Hamada^{†‡}

1 はじめに

シーンチェンジは映像認識と映像理解にとって不可欠な部分であり、シーンチェンジ検出により生映像をシーンを単位に分割し、それぞれのインデクシング情報を付けることができ、従来の映像ストーリーごとに映像を再構成し、より効率的な映像探索空間を作ることができる。本稿では、エッジ検出手法を用い、従来の手法と違い、よりロバストなシーンチェンジ検出手法を研究し、その結果により新たな手法を提案した。この手法はあらゆるシーンチェンジタイプに柔軟に対応できる可能性があると考えている。

2 研究の目的と背景

シーンチェンジ検出は普通三つの種類があり：カット検出、フェイド検出とディゾルブ検出。その他にまたスライドインとスライドアウトなどがある。従来の手法は、フレーム間のヒストグラムの差分を取り、閾値と比較され、シーンチェンジを検出する。従来の手法は、ノイズなどの影響に弱く、ロバストなシーンチェンジ検出を十分に実現しているとはいえなかった。

ここでエッジ検出手法を用い、我々は新たなシーンチェンジ検出手法を実験した。図1のようにXとY方向の各フレームが時間軸tに沿い、フレーム1からiまで映像を形成する。時間軸tの一定の時間帯Tに、フ

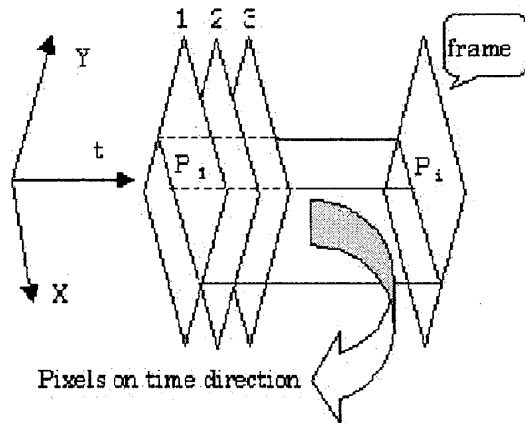


図1: エッジ検出によりシーンチェンジ検出

レームにあるある画素 $P(x, y, i)$ の値をエッジ検出手法を用い、時間軸方向の一次元のエッジ検出を適用する。Yの値を固定させ、X方向のすべての画素のエッジ検出を行い、その結果により、フレームはシーンチェンジフレームであるかどうかを判断する。

3 提案した手法

エッジ検出手法はたくさん存在することで、この中から我々はCANNY[1]手法に近いラプラシアン-ガウシアンオペレータを使用し、一次元のステップエッジの検出を行う。

3.1 ラプラシアン-ガウシアン手法

ここで輝度の計算は次の式を使っている:

$$E(i) = 0.299 * Red + 0.587 * Green + 0.114 * Blue \quad (1)$$

ラプラシアン手法はエッジの方向によらない等方向性の差分オペレータであり、高周波数成分のノイズを強調してしまうため、原画像をまずガウス関数の畳み込みにより高周波数成分を抑制してからラプラシアンオペレータを適用する方法がよく使われる。ここで求めるエッジ関数 $F(x, y, t)$ は次の式で求められる:

$$F(x, y, t) = \int \partial G(\tau) / \partial t E(x, y, (t - \tau)) d\tau \quad (2)$$

$$G(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-t^2/2\sigma^2} \quad (3)$$

ここで $G(t)$ はガウス関数であり、 $E(x, y, t)$ は画素の輝度関数である。 $G(t)$ と $E(x, y, t)$ の畳み込みの計算によりエッジ関数 $F(x, y, t)$ を計算する。実際は離散値で計算する:

$$F(x, y, t) = \sum_{\tau} \partial G(\tau) / \partial t E(x, y, (t - \tau)) \quad (4)$$

ここでガウス関数 $G(t)$ のパラメータ σ の値を1とする、1とする場合はガウス関数 $G(t)$ は高周波数成分を抑制し、より明確なエッジ検出にも適用する。

3.2 シーンチェンジ検出

上記で得たエッジ関数 $F(x, y, t)$ により表される時空間画像は、輝度の時間変化が大きい画素に対して大きい絶対値を持つ。従って、カットに代表されるシーンチェンジ部分では、対応する時間区間において、多くの部分で大きい絶対値となると期待できる。そこで、 $|F(x, y, t)|$ の時間方向に射影したシーンチェンジ関数 $C(t)$ を定義する:

$$C(t) = \sum_{x, y} |F(x, y, t)| \quad (5)$$

シーンチェンジにおいては、 $C(t)$ が大きな値となると考えられるので、適当な閾値などでシーンチェンジ検出が可能になると考えられる。特にカットにおいては、カットの時点のすべての画素で $|F(x, y, t)|$ が大きな値となるので、 $C(t)$ はピークを構成することになる。

[†] 東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technology,
The University of Tokyo

[‡] 国立情報学研究所
National Institute of Informatics

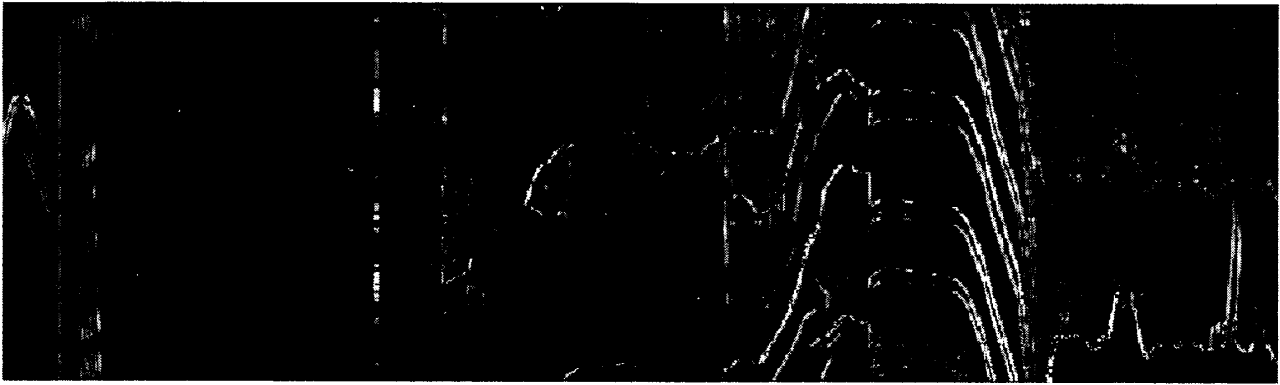


図 2: エッジ画像

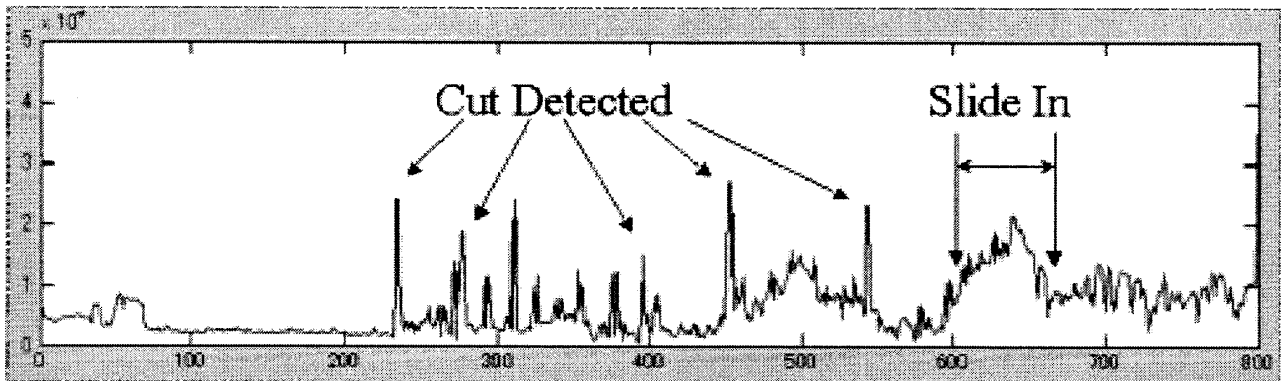


図 3: シーンチェンジ関数

4 実験結果

ここで実験結果を見せる。これは連続のやく55秒の映像をこの手法により検出した結果である。1600フレームを800フレームずつエッジ関数 $F(x, y, t)$ を計算し、図2は、ある映像の特定の800フレームの区間から時空間エッジ画像 $F(x, y, t)$ を求め、その絶対値を $y=100$ 画素の平面で切り出した画像を表している。

また、図3はそれから求めたシーンチェンジ関数 $C(t)$ を表している。図中の線は実際のシーンチェンジを表している。カットにおいて、 $C(t)$ がピークとなっていることが示されている。すべてのカットを検出できるが、カットではないのピックのところもある。図のように610-660フレーム付近では特殊効果によるシーンチェンジ(スライドイン)が存在し、カットのような明確なピークは観測できないが、他の映像部分よりも特殊効果のシーンチェンジ関数 $C(t)$ の変化が見られる、その変化の特徴でスライドインの検出は現在実験中である。他のタイプのシーンチェンジを検出できるため、シーンチェンジ関数 $C(t)$ をまだ改善する必要がある。

5 終わりに

以上の実験結果により、この新たなシーンチェンジ検出手法は従来の手法、よりロバストである手法である。これからの課題は各種のシーンチェンジを柔軟な対応ができることを実現させることを目指す。より大量の映像を使い、この手法を検証し、より早い、正確的な実装したアプリケーションを開発予定が今後の仕事である。

参考文献

- [1] John, Canny: A Computational Approach to Edge Detection, PAMI-8 No.6, November 1986