

対象物の動きの一貫性に注目したシーン編集におけるカットの決定法

Cut Decision Method based on Video's Motion Consistency in Scene Editing

竹本 楽^{*1}
Ryo TAKEMOTO

吉高 淳夫^{*1}
Atsuo YOSHITAKA

平嶋 宗^{*1}
Tsukasa HIRASHIMA

^{*1} 広島大学大学院工学研究科
Graduate School of Engineering, Hiroshima University

The video editing is an operation to produce video contents. This operation adds edit effect to video segments and emphasizes impression. In editing videos, it is important to appropriately decide cuts that one shot is replaced with another. In this paper, we propose a method to decide cut between shots based on video's motion consistency in the video editing.

1. はじめに

映像編集とは、素材映像をつなぎ合わせ、場面の印象を強調する演出意図を込める作業である。また、映像編集について一般的に知られていることについて述べられた「映画の文法」[1]などの知識を用いて、編集することで、映像に演出意図を与えることができる。しかし、編集作業は一般の編集者にとって、困難な作業である。なぜなら、編集作業では多くの時間や労力が必要とする。また、熟練した技術を持たずに映像を編集すれば、映像に対して、正確に演出意図を与えることができず、視聴者に違和感を与える映像となってしまう。そこで、編集者が時間や労力、熟練した技術が必要とせず、自動でフレームをつなぎ合わせ、演出意図が付加された映像を生成し、編集作業の負担を軽減することを目指す。

本研究では、編集技法の中で、ショット長遷移による演出に注目する。[1]では、編集技法に従い、フレームをつなぎ合わせることで映像に特定の印象を付与することができると述べられている。そのとき、付加したい印象を保ち、視聴者に違和感を与えない映像とするため、映像を切り替えるタイミングであるカットの決定が重要である。しかし、一般の編集者は編集技法の知識を持たないため、映像に付加する印象を考慮し、カットを決定することは困難な作業である。そこで、時間的に同期した複数の映像を対象として、印象を付加する編集技法に基づく、映像中の対象物の動きに注目したカットの決定法について述べる。

2. 映画の文法に基づく編集技法

編集の際、制作者によって特定の意味や意図を強調、または違和感のない映像を生成する目的で使用される技法がある。本稿では、[1]で述べられている編集技法に基づき、処理を行う。

2.1 ショット長遷移による演出効果

ショット長遷移による演出は4種類の印象を強調する[1]。

(1) アクション

短いショットが連続し、ショット内の画像の変化が激しい。

(2) 落ち着いた

長いショットが連続し、ショット内の画像の変化が少ない。

(3) 緊迫

ショットの長さが徐々に短くなる。

(4) 解放

ショットの長さが徐々に長くなる。

それぞれのショット長遷移の編集技法に従い、フレームをつなぎ合わせることで、特定の印象を強調した映像を生成することが可能である。

2.2 映像の動きに基づくカットの決定

2.1 節において、ショット長遷移により、特定の印象を強調することができる述べたが、ショット長を変化させるとき、映像を切り替えるタイミング、すなわちカットの決定が重要になる。なぜなら、連続するショットで映像の動きの激しさや向き、速さが一致していなければ、視聴者に違和感を与え、混乱させるからである[1]。また、カットの前後で動きを一致させることで、ショット長遷移による演出を保存することが可能である。

そこで、本稿では映像の動きに基づく、カットの決定法について述べる。映像の動きとは、映像の動きの激しさ、向き、速さで定義する。映像に特定の印象を与えるとき、そのショット長遷移規則に従い、かつカット決定の規則を満たすフレームをカットと決定する。

カットの位置は、映像の動きが終了したフレームとする。または、映像の動きがないフレームをカットとする。また、映像の動きがあるときに、映像を切り替える場合、2.3 節で述べる演出に従ったカットの決定を行う。

2.3 演出に従った動きの一致

映像の動きがあるとき、カットを決定する場合、映像に付加したい印象ごとに異なる方法でカットの決定を行う。映像に「アクション」の印象を与えるとき、ショット長遷移規則に従ったショット長であり、かつカットの前後で映像の動きの激しさが最も一致する位置をカットと決定する。また、映像に「アクション」以外の印象を与えるとき、すなわち「落ち着いた」、「緊迫」、「解放」の印象を与えるとき、与える印象のショット長遷移規則に従ったショット長であり、かつカット前後の映像の動きの向き、速さが最も一致する位置をカットと決定する。

3. 映像の動きの解析

編集技法に従い、カットを決定するため、カット決定の特徴量として、映像の動きの激しさ、向き、速さの解析を行い、検出する。動きの向き、速さに関して、スチル区間、カメラワーク区間に分けて、検出を行う。なぜなら、スチル区間とカメラワーク区間では、着目する映像の動きの特徴が異なるため、異なる方法で検出する。

3.1 動きの激しさの検出

文献[2]の手法より、時空間投影画像上に現れるエッジの数から動きの激しさを表す値の算出を行う。映像の動きが激しいほど時空間投影画像上に現れるエッジの数が多くなることから、エッジの数に着目して、その数によって映像の動きの激しさを定義し、動きの激しさの値を算出する。

3.2 スチル区間における映像の動きの検出

映像中のスチル区間では、視聴者が注目する動物体の動きの向き、速さをスチル区間における映像の動きとする。動きの向きは右、左、上、下の4つの向きに分類する。

(1) フレーム間差分画像の生成による動物体領域の抽出

画像中の動物体領域をフレーム間差分により、抽出する。本研究ではフレームサイズ $W \times H$ ($W=320, H=240$) [pixel] の画像を、実験により求めた最適なブロックサイズ $M \times N$ ($M=5, N=5$) [pixel] に分割し、ブロック内の画素の輝度値の平均を算出する。比較するフレームの対応する位置にあるブロックと比較し、輝度値の平均の差が閾値 $Th_t (=4)$ 以上であれば、画像が変化しているブロックとして抽出する。さらに動物体領域を正確に抽出するため、注目しているフレーム f_i とフレーム f_{i-d}, f_{i+d} との差分画像を生成し、それらの論理積を計算し、フレーム f_i の差分画像とする(図1)。このとき、時間 t 、フレーム間隔 $d (=5)$ とする。

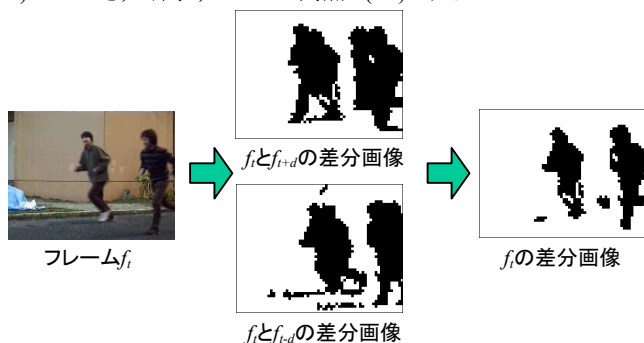


図1 フレーム間差分画像

(2) 動物体領域の動きの向き、速さの検出

3.2(1)で抽出した動物体領域の動きの向き、速さを検出するため、3.2(1)で抽出したブロックごとにブロックマッチングを行い、フレーム f_{i+d} におけるフレーム f_i 中の注目しているブロックに対応するブロックを抽出する。ブロックマッチングとは、画像中のあるブロックをテンプレートとし、あるフレーム間隔後の画像中を探索し、差分評価関数 $D(i, j)$ の値を最小とするブロックを対応ブロックとする。テンプレートブロックの左下端の座標を (x_0, y_0) 、フレーム f_i の画素 (x, y) の輝度値を $V(x, y, f_i)$ 、対応するブロックの探索範囲を i, j とすると、差分評価関数 $D(i, j)$ は次式(1)で表される。

$$D(i, j) = \sum_{m=0}^M \sum_{n=0}^N |V(x_0 + m, y_0 + n, f_i) - V(x_0 + m \pm i, y_0 + n \pm j, f_{i+d})| \quad (1)$$

テンプレートであるブロックと探索された対応するブロックから、動きベクトルを検出する。そして、検出された動きベクトルを用いて、動物体領域の代表ベクトルを検出する。代表ベクトルの向きの検出は、検出された動きベクトルの向きから4階級のヒストグラムを算出し、ヒストグラムの最頻となる階級の向きを動物体領域の動きの向きと決定する。

次に、動きの速さの検出は、動物体領域の動きの向きと同じ向きである動きベクトルを x 軸, y 軸成分に分解し、ベクトルの長

さの平均を動物体領域の画像上の移動距離、すなわち動きの速さとして検出する。

4. 実験

ビデオカメラを用いて撮影した映像に対して実験を行った。なお、使用したビデオデータの画像形式は、フレームサイズ 320×240 [pixel], フレームレート 30 [fps], 24 ビットカラーである。

4.1 映像の動きの向きの検出精度

スチル区間における映像の動きの向きの検出結果と、主観による評価を照らし合わせて検出精度を求めた。実験では5フレーム間隔で動きの向きの検出を行った。表1にその結果を示す。なお、 Nd : 提案手法により検出された数, Nc : 提案手法により正しく検出された数, Na : 主観により求めた数とする。結果として、Recallが82%となり、動物体領域の動きの向きを検出することができた。しかし、Precisionが63%であった。原因として、動物体が存在しない区間において、フレーム間差分画像を生成した際、ノイズ領域が発生したため、誤検出が増加したことが挙げられる。

表1 動きの向きの検出精度

Precision(Nc/Nd)	Recall(Nc/Na)
0.63(123/196)	0.82(123/150)

4.2 決定されたカットの妥当性評価

本システムにより決定されたカットが、印象を保ち、視聴者に違和感を与えない位置に決定されたものであるかを評価するための実験を行った。大学生5名に、本システムでカットを決定した3本の映像を視聴してもらい、そのカットが適切な位置に決定されているか5段階で評価してもらった。適切である場合の評価値を5、少し適切である場合を4、どちらでもない場合を3、あまり適切でない場合を2、不適切である場合を1とした。表2にその結果を示す。表2に示しているのは、5名の平均評価値である。結果として、平均以上の評価を得た。映像2では、動物体の動きの向きに加え、速さもカットの前後で一致していたため、高い評価を得た。逆に映像1では、動物体の動きの向きは一致していたが、動きの速さがカットの前後で異なっていたため、低い評価になったと考えられる。

表2 カットの妥当性の実験結果

	映像1	映像2	映像3
評価値	3.4	4.6	3.8

5. 今後の展開

本稿では、[1]に述べられたショット長遷移による演出により、印象を付加した映像の編集を実現するため、その際重要である、映像の動きの一貫性に注目したカットの決定について提案し、評価した。今後は、今回提案したカット決定法を用いて、ショット長遷移による演出により、印象を付加した映像の編集システムの構築を行っていく予定である。

参考文献

- [1] ダニエル・アリホン著, 岩本憲児, 出口丈人訳, "映画の文法", 紀伊国屋書店, 1980.
- [2] 出口嘉紀, 吉高淳夫, "意味内容に基づく映像の自動編集", インタラクティブシステムとソフトウェア XI: 日本ソフトウェア科学会 WISS2003, pp. 31-40, 2003.