

見やすい映像提示のための映像特徴抽出法

盛林 達哉, 村上 和人
愛知県立大学大学院情報科学研究科

あらまし 複数のカメラからの映像をひとつのコンテンツとして編集する場合、一般には編集者の主観に基づいてこれを行っている。しかし、映像の自動編集を実現するためには、人間の主観に代わる何らかの客観的な映像編集基準を設定する必要がある。本稿では、映像中の移動物体の特徴に基づく映像編集方法を提案する。提案手法では、映像中の移動物体の特徴量を定め、複数のカメラの中で最も特徴が大となるカメラの映像を繋ぐように映像を編集する。提案手法を RoboCup 小型リーグの映像に適用した結果、映像中の移動物体の特徴を映像編集基準として自動編集を行った場合でも、主観的に編集を行った映像と同様な見やすい映像が編集できることを確認した。

A Method to Switch Cameras Based on the Features of Moving Objects for an Interesting Video Clip

Tatsuya MORIBAYASHI, Kazuhito MURAKAMI

Graduate School of Information Science and Technology, Aichi Prefectural University

Abstract It is necessary to set an objective standards that take the place of director's subjectivity, in order to realize automatic video edit. This paper proposes an method to switch cameras based on the features of the moving objects. We applied our method to the RoboCup's video clips. Experimental results show the effectiveness of the proposed method.

1.はじめに

1台のカメラによる撮影は、3次元空間情報を2次元画像平面に射影するため、隠蔽の発生や、動きの情報を十分に反映できないという本質的な課題を内包する。この問題を解決するため、最近では複数のカメラによる撮影が一般的に行われる。

図1は小売店に取り付けられたカメラの例である。複数のカメラを利用することで店内における隠蔽空間を減らし、効果的な万引きの監視を行っている。また、テレビ放送に代表されるメディアコンテンツの収録にも多くのカメラが使われている。複数のカメラを使用することによって利用可能な映像を増やし、映像の見せ方やコンテンツの幅を広げることができる。



(a)複数のカメラからの映像 (b)設置されたカメラの一例

図1: 撮影に複数のカメラが使われる例

しかし、撮影に複数のカメラを使用し、取得するデータを増やすことは、不要なデータまでも増加させてしまうという問題を引き起こす。この問題は、映像の表示領域や記録領域といったリソースを圧迫するだけでなく、撮影された映像を利用する上で非効率化を招く。そこで、取得した複数の映像をどのように編集(取捨選択)するかが課題となる[1, 2]。人(例えば、テレビ局の人)はこのような作業を経験的に基づいて行っている。しかし、この経験的な作業を自動化できれば、映像編集の効率化の向上につながるだけでなく、人はどのような局面やシーンを興味・関心をもって見ているかを解明することにつながる。

これまで、マルチカメラなどを利用した映像編集における研究では、多くは講義やスポーツ映像など、撮影対象を予め設定した研究であった[3-5]。しかし、より一般的な映像に対して手法を適用する場合、撮影対象を予め決めないことが望ましい。以上を考慮し、本稿では、映像中における、視聴者が見やすいと感じられる条件を明らかにし、画像処理により自動的に条件にあった最適な映像を提示する一手法を提案する。

以下、2章で提案手法について述べ、3章で有効性を確認するための実験例について述べる。そして4章で実験結果を示し、5章で得られた結果について考察する。

2.提案手法

2.1.見やすい映像とは

見やすい映像とは、“視聴目的に適した映像”と考えられる。例えば小売店の万引き監視のためには、監視対象が視野から外れることなく写されている映像が必要であり、見やすさの観点からは対象物ができるだけ大きく写されている映像が好ましいということになる。しかし、見やすい映像というのは主観的な判断であるため、映像編集にあたっては何らかの客観的な特徴量に置き換えた上で編集などの作業を行わなければならない。

また、映像は動きの情報が多く、この情報を多く含

むほど映像としての価値が高く、より見やすい映像に近づくのではないかと推測される。つまり、映像中における特徴として、移動物体の特徴を重要視する必要がある。

そこで、映像中の移動物体による特徴のうち定量的に求めることができるものを特徴量として設定し、複数のカメラからの映像を切り換えて編集を行う。以下、提案手法の基本的な処理について述べる。

2.2.提案手法

本手法では、まず映像特徴として使用する“移動物体の特徴量”をフレーム毎に求める。 N 台の各カメラの映像から求めた特徴量 x_c ($c = 1, 2, 3, \dots, N$) を比較し、各時点において特徴量が最大となるカメラの映像を選択するように編集する。 $N = 3$ の場合の様子を模式的に図 2 に示す。時刻 $0 \sim t_1$ の間はカメラ 1 の映像を、時刻 $t_1 \sim t_2$ の間はカメラ 2 の映像を、時刻 $t_2 \sim t_3$ の間はカメラ 3 の映像を、それ以降はカメラ 2 の映像を繋いでいる例である。以下同様に、特徴量が最大のカメラの映像を繋ぐように編集する。

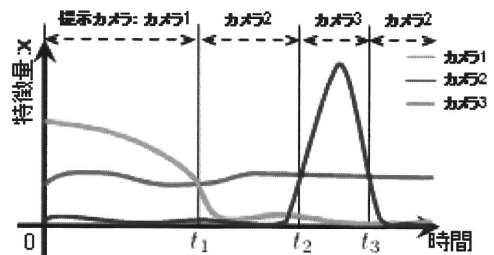


図 2:カメラの切り替え例(特徴量最大のものを選択)

また、2つのカメラの特徴量が拮抗すると、特徴量が最大になるカメラの映像を選択という単純な処理では、カメラの切り換わりが頻繁に起こってちらつくことになり、映像として好ましくない。そのため、カメラの特徴量が最大になった後、0.5秒以上続けて最大となった場合のみカメラを切り替えることとする。

3.映像特徴の抽出による映像提示実験

3.1.実験方法

移動物体を対象として、複数のカメラにより同時に撮影を行う。そして、提案手法で示した方法で一つの映像に編集する。提案手法の有効性を確認するために、この映像を被験者に提示し、図 1(a)のように画面を分割し複数のカメラからの映像を同時に見せた場合と、どちらが見やすいのかを比較する形で行う。また、特徴量によって編集結果が変わるため、どの特徴量を用いるのが有効かの評価も併せて行った。

3.2.実験に用いた映像

今回の実験では、RoboCup 小型リーグの試合の様子を撮影したものを実験素材として用いた。図 3 に試合風景を示す。



図 3:RoboCup 小型リーグの試合風景

RoboCup 小型リーグでは、5.0m×3.5m のフィールド上でオレンジ色のゴルフボールを用い、1 チーム 5 台以内のロボットがサッカーのルールに従って対戦する。

3.3.映像特徴量

映像中の移動物体の動きを総合的に表す特徴量としてフレーム間差分を、また、映像中の移動物体のみに着目した特徴量として、移動物体の面積を特徴量とする。さらに、後者の例としてロボットの面積と、ボールの面積を考える。以下に、それぞれの特徴量の求め方をまとめる。

① フレーム間差分

現在のフレームと前のフレームとの画像全体での濃度値変化を合計し、これを特徴量とする。

② ロボットの面積

映像中のロボット領域を、ロボットが黒色であることを利用して抽出し、画像全体でロボットの面積を合計し、これを特徴量とする。

③ ボールの面積

映像中のボール領域をボールがオレンジ色であることを利用して抽出し、ボールの領域のうち、最大のものの面積を特徴量とする。

3.4.実験環境

実験には同一仕様のカメラ 3 台を使用した。カメラは SANWA SUPPLY の TC-112 (図 4)を使用した。

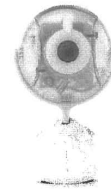


図 4:撮影に使用したカメラ(TC-112)

撮影にあたり、カメラは固定し、カメラの自動調整機能はすべて OFF にした。また、画像サイズは 320 × 240 とした。設置したカメラの各々の視野を図 5 に示す。

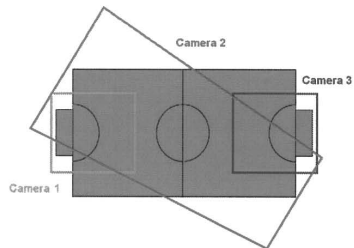


図 5:実験に使用したカメラの視野

撮影は 3 回行い、撮影毎に図 5 の視野を少しずつ変化させた。RoboCup 小型リーグのためのマルチカメラ配置を扱った研究[3]もあるが、今回のカメラ配置は一般的に注目度の高いゴール前と、全体を写すカメラという一般的な構成をとった。図 6 はキッ

クオフ前の各カメラからの入力映像である。

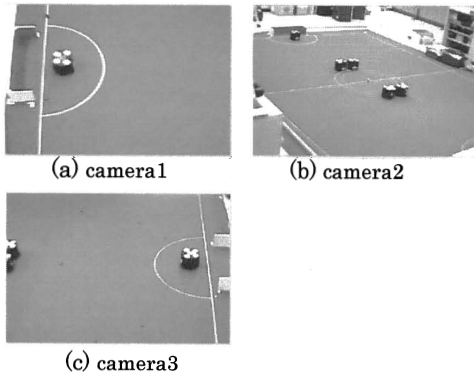


図 6: 各カメラからの映像の一例 (キックオフ時)

撮影された映像から特徴量を求める際の中間処理結果の一例を図 7 に示す。図 7(a)は撮影された映像であり、これから、同図(b)のようなボール領域の候補を求めている。

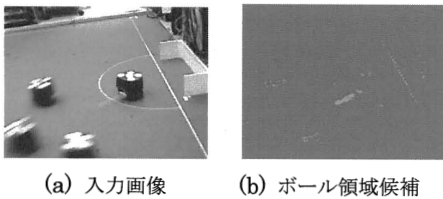


図 7: 特徴量抽出結果の一例

3.5. 評価方法

撮影は 3 回行い、そのそれぞれに対して特徴量別の動画を 3 本作成した。また、3 つのカメラからの映像を同時に収録した映像も同時に作成した。

特徴量別に編集した映像が 3 つと、3 つのカメラからの映像をそのまま合成した映像 (画面を分割して同時表示した) 映像の計 4 つの動画を作成した。評価はこの 4 つの映像を 1 セットとして被験者に提示し、アンケートにより主観的な評価を行う。

撮影は 3 回行い、被験者による評価は 3 セットに対して行ってもらった。3 セットの評価を 20 人の被験者に対してアンケート調査を行った (有効回答数 $n=60$)。

被験者には“3 つすべての映像を同時に提示した場合の映像”を基準評価値 5 とし、0~10 の範囲で、

7 つの項目について回答をもらった。質問項目を表 1 に示す。この中で、“別カメラの映像が見たいところがあった”と、“カメラの切り換えが遅いところがある”については基準値を設けず 0~10 の範囲で回答してもらった。なお、すべての項目において数値が高いほどよい評価とする。

表 1: アンケート項目と質問番号対応表

評価項目	ラベル
試合状況が把握しやすい	A
プレーの内容がわかりやすい	B
映像に迫力があつた	C
プレーのスピーディーさが伝わつた	D
ストレスなくみられた	E
別カメラの映像が見たいところがあつた	F
カメラの切り換えが遅いところがある	G

4. 実験結果

4.1. 評価結果

図 8 に 60 回答の平均点を、図 9 に一番高い得点が付いた回数をまとめた。図 8、図 9 の評価項目 A~G は表 1 のラベルに対応している。また、図 9 では重複回答を認めている。

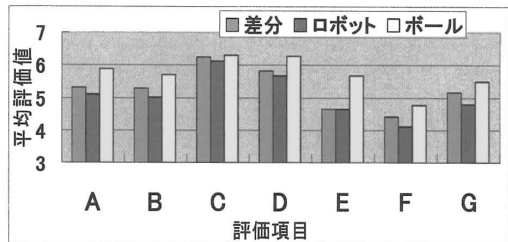


図 8: 評価結果平均値グラフ

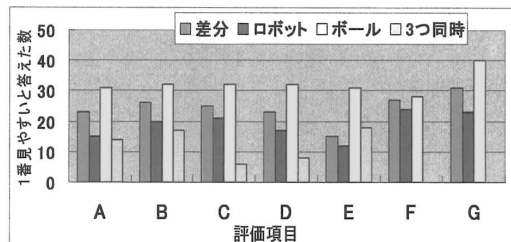
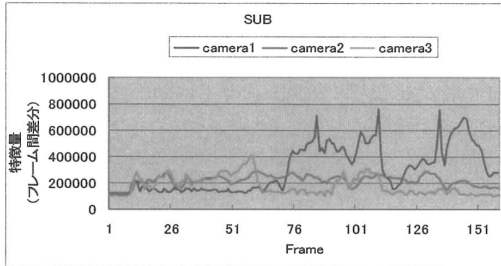


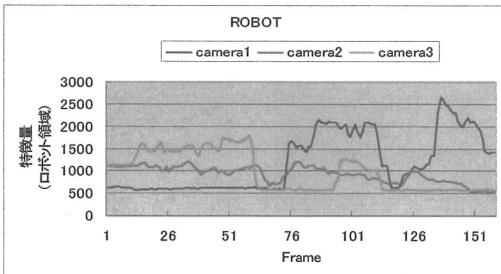
図 9: 一番見やすいと回答した数グラフ

4.2. 評価結果考察

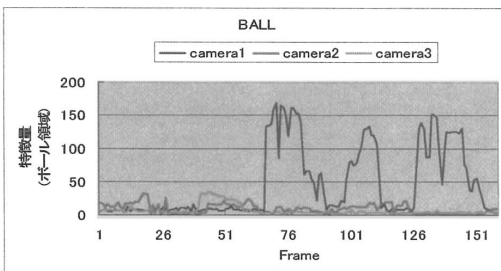
図 8 より、特徴量同士であまり差が開かない結果となった。これは、ボールの集まる場所にロボットが集まりやすいという RoboCup 小型リーグの特性のためと推測される。図 10 に、ある区間での特徴量の時系列変化を示す。



(a) フレーム間差分



(b) ロボット領域の面積の総和



(c) 最大ボール領域の面積

図 10: 特徴量の時系列変化グラフ

図 10 より(a) フレーム間差分と、(b) ロボットの面積では、グラフの変化が類似していることがわかる。これは、映像中における差分発生要因のうち最も大きいものがロボットであるため、ロボットが多く集まる場所で差分が大きくなるためである。

また、(c) ボールの面積では 60frame 辺りまでは各カメラからの特徴量に変化が少ない。ボールという小さな対象物を検出しなければならぬため、一番視野の広い camera2 においては最大でも 30 ピクセルほどしか領域が検出できておらず、小さな対象物の面積を特徴量とするとノイズに左右されやすい場面があることがわかった。

3 つのカメラからの映像をそのまま提示するより、特徴量により編集を行った映像のほうが、主観的評価で良い結果が得られた。また、基準として設定する特徴量によっても評価結果に差が出ることも確認された。今回の実験で使用した特徴量に関しては、ボールの面積を特徴量とした場合が最もよい結果であり、ボールが見えている映像では視聴者にストレスを与えにくいこともわかった。

5. 考察

5.1. 提案手法の有効性についての考察

今回の実験では、移動物体の面積に基づいた映像特徴を特徴量として使用すれば、複数の映像を同時に表示するよりも見やすい映像への自動編集が可能であることがわかった。これは、図 9 の評価項目 A~E で、ほとんどの場合において“3 つのカメラからの映像を同時に提示した場合”よりも移動物体の面積を特徴量として自動編集した映像のほうが評価が高いことからわかる。

さらに、同図評価項目 C (映像に迫力があつた)、D (プレーのスピーディーさが伝わった) より、特徴量別に編集した映像は、より“スピーディーさと迫力の伝わる映像”という面白みのある映像に仕上がったといえる。

以上より、提案手法を RoboCup の映像に適用した範囲では、限定的ではあるが提案手法が有効であるといえる。

5.2. 映像特徴の一般化

本稿では、特徴量を映像中の移動物体の面積など 3 種類に絞って実験を行ったが、注目する物体に

よって視聴者に与える印象に違いが生じたのも事実である。複数の移動物体が映像中に含まれる場合など、特に設定する特徴量を明確にする必要があると考えられる。より効果的な映像編集をするためには、人が注目すべき移動物体やその特徴を、例えば視線追跡技術を応用するなどして分析し、実際の映像収録現場の視覚感性的な状態を反映させることも重要となってくる。今後はこれらのことを検討する必要がある。

5.3.映像提示法への応用

今回の映像編集では、特徴量を比較し、特徴量が最大となるカメラの映像のみを提示する方式をとった。しかし、図 10(a)の 60frame 辺りまでに見られるように、2 つのカメラで特徴量が拮抗する場面がみられた。このような場合には 2 つのカメラを切り替えながら提示するより、どちらか一方のカメラ映像を重視する、あるいは、この 2 つのカメラからの映像を同時に提示するほうが状況理解しやすくなると考えられる。

また、図 10(c)の 60frame 付近で見られるように、特徴量に急激な変化が現れることがあった。今回、映像中の移動物体の面積に注目した特徴量を使用しているため、このような急劇な変化が起こることは、映像中の物体に急激な変化があった事を意味する。そのため、急激な動きに対し、映像をスロー再生するような配慮は、見やすい映像提示を行う上で有効と思われる。このように、本手法を通じて、映像特徴を使用したさらに効果的な映像提示への応用が考えられる。

6.おわりに

本稿では、複数カメラからの映像編集・映像提示を行うにあたり、映像中の移動物体の特徴量を基準として自動で行う方法を提案した。そして、RoboCup 小型リーグの映像に適用し有効性を確認した。アンケートから、提案手法の有効性を見て取ることが出来た。しかし、提案手法はベターな手法であるもの

の、ベストな方法である保障が得られたわけではない。今後は目的にあった映像編集に効果的と思われる映像特徴の選択を自動的に行えるようにすることが必要になる。

参考文献

- [1] 尾形 涼, 尾関 基行, 中村 裕一, 大田 友一, “制約と評価関数に基づいた映像編集モデル”, 電子情報通信学会技術研究報告 IE, 画像工学, Vol.103, No.205, pp.13-18 (2003) .
- [2] 亀田 能成, 新 康孝, 西口 敏司, 美濃 導彦, “撮影対象の運動履歴に基づく固定ショット切り替え式撮影法”, 電子情報通信学会技術研究報告 PRMU, パターン認識・メディア理, Vol.103, No.207, pp.1-6 (2003) .
- [3] 井口 泰典, 土居 元紀, 眞鍋 佳嗣, 千原 國宏, “スポーツ映像放送のための実時間映像解析によるマルチカメラの自動制御と自動スイッチング”, 映像情報メディア学会誌, Vol.56, No.2, pp.271-279 (2002) .
- [4] 錦織 修一郎, 菅沼 明, 谷口 倫一郎, “黒板講義を対象とした講義自動撮影システムの構築”, 電子情報通信学会 PRMU2000-212, Vol.100, No.701, pp.79-86 (2001).
- [5] 先山 卓朗, 大野 直樹, 椋木 雅之, 池田 克夫, “遠隔講義における講義状況に応じた送信映像選択”, 電子情報通信学会論文誌, Vol.J84-D-II, No.2, pp.248-257 (2001) .
- [6] 成田 良太, 村上 和人, “複数カメラの適応配置とロボカップへの応用について”, 映像情報メディア学会技術報告 ME2007-129, Vol.31, No.39, pp.31-34 (2007).