

カメラワークパラメータを利用した背景除去に基づく演出対象の検出

Salient Object Detection based on Background Elimination using Camera Work Parameter

尾崎 昂*¹
Takashi OSAKI

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

In this paper, we describe a method for detecting salient object's region during a pan section in movie. A proposed method is based on the feature that a background region moves at the constant speed, and the feature that movements of background region differ from these of salient object's region.

1. はじめに

近年の映像配信の増加に伴い、映像を蓄積する機会が増え、多くの映像の中から見たい映像に効率よくアクセスできる方法の要求が高まっている。この要求を実現するために、映像の意味内容を解析し、映像に映っている内容の情報を抽出することが必要である。このとき抽出する情報については様々であり、例えば映画やドラマといった映像においては演出に関する情報が重要である。松井ら[1]はカメラワークによる演出効果を抽出しているが、その対象については考慮していない。「映画の文法」[2]によると、カメラワークは重要な演出技法であると位置づけられており、演出の対象によって制作者が伝えたい内容が異なる。

そこで本研究では、カメラワークが存在する区間において演出の対象となっている被写体を検出する手法を提案する。検出した演出対象は区間の特徴がよく表れている情報であるといえ、効率のよいアクセスを実現する、映像の内容検索やインデキシングに利用できる。なお、区間については[1]の手法により決定する。本研究では以下のカメラワークを対象にする。

- パン

カメラを水平方向に回転させる操作

- チルト

カメラを垂直方向に回転させる操作

パン、チルトは(1)対象を見わたす、(2)動く対象を追跡する、という主に 2 つの場面に用いられる[2]。本研究では映像中の演出対象を特定することを目的とするため、(2)における、カメラによって追跡している対象を検出する。

2. 関連研究

パン、チルトといったカメラ操作のある映像からオブジェクトを抽出する研究として、Chang ら[3]はフレームごとに領域分割を行い、領域同士を対応付けることにより、フレーム間で同じ動きをしたオブジェクトを抽出している。奥村ら[4]は、動きベクトル量の少ない部分の検出と肌色領域検出の処理により、カメラ操作によって追跡している人物領域を抽出している。しかし、本研究で対象とする映画やドラマといった映像では、輝度変化が激しいため領域の対応付けが困難であり、また検出する対象の種類や大きさが限定されていない。よって単純な画像処理での検出は難しく、映像自体の性質に着目した手法が必要である。

本研究では映画で用いられるカメラワークに着目する。動く対

象を追跡する区間では、対象の動きとそれ以外の背景の動きは異なっている。またカメラワークによって、背景が一様な動きとなる。そこでカメラ操作量から背景部分の移動量を求め、一様な動きの背景部分を除去することで動く対象を検出する。以下本稿では、カメラワーク区間において演出の対象となっている被写体を、主体と定義する。

3. カメラワークパラメータの抽出

3.1 時空間投影画像

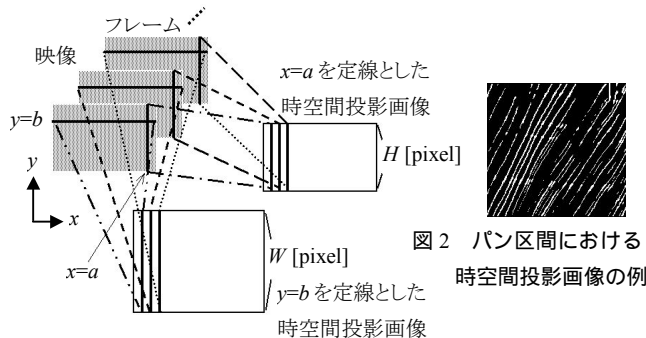


図1 時空間投影画像

時空間投影画像とは、フレーム内の一定位置の直線を各フレームから抽出し、それらを時間軸に沿って並べた画像である(図1)。以降、フレーム内における一定位置の直線のことを定線と表す。

時空間投影画像内のエッジには、カメラワークやその操作量に伴った特徴が現れる。カメラ操作量によって背景の動きの移動量が決まるため、本研究では時空間投影画像を用いることにより、背景部分の一様な動きの移動量を求める。

フレームサイズ $W \times H$ [pixel] の映像において、パン区間では水平方向の直線 $y=b$ ($0 \leq b \leq H-1$)、チルト区間では垂直方向の直線 $x=a$ ($0 \leq a \leq W-1$) を定線とした時空間投影画像にエッジ群が斜めに現れるという特徴がある(図2)。このとき、被写体の動きの影響を受けにくい、フレームの端付近の定線で以降の処理を行う。

3.2 背景移動量の決定

区間内の時空間投影画像の直線エッジを検出し、 e 本の全エッジの角度 θ ($-80^\circ \leq \theta \leq -5^\circ$, $5^\circ \leq \theta \leq 80^\circ$) を求める。背景の動きを表さない垂直、水平方向のエッジは除いている。差分をとるフレーム間隔を dif 、 θ ($i=1, \dots, e$) の最傾角を θ_{mod} とし、背景移動量 mov [pixel] を次の(1)式により決定する。

連絡先: 尾崎 昂, 広島大学大学院工学研究科,
広島県東広島市鏡山 1-4-1, (tel)082-424-7669,
(E-mail)subaru@isl.hiroshima-u.ac.jp

$$mov = dif \times \tan \theta_{mod} \quad (1)$$

フレーム間隔 dif を一定にすると mov の範囲が大きくなり、後述の背景除去処理で除去されない部分の現れ方に違いが生じる。これよりカメラ操作量に依存しない主体検出が困難になるため、 dif は $\tan \theta_{mod}$ によって動的に決定する。

4. 主体の検出

4.1 背景除去処理

背景の動きをキャンセルした差分画像を生成する。本稿では左方向のパンの場合について説明する。 5×5 [pixel]のブロック $B_{kl}(k=0, \dots, W/5, l=0, \dots, H/5)$ において (2) 式の条件を満たすとき、つまり主体の検出を行うフレーム f_n の座標 (i, j) における輝度 $V_n(i, j)$ と、フレーム f_{n+dif} における座標 $(i+mov, j)$ における輝度 $V_{n+dif}(i+mov, j)$ との、ブロック内での差の合計が閾値 Th 以下のとき、フレーム f_n のブロック B_{kl} を背景部分として除去する。

$$\sum_{i=5k}^{5k-45l-4} \sum_{j=5l}^{5l-4} |V_n(i, j) - V_{n+dif}(i+mov, j)| \leq Th \quad (2)$$

背景移動量の誤差を考慮し、実験によって求めた誤差の範囲内で背景移動量を変化させ、最も多くの背景部分が除去されたときを背景除去処理の結果とする。

4.2 領域分割処理との統合による主体検出

(1) 領域分割処理

背景除去処理では、主体領域の内部も一緒に除去されるので、主体領域の全体を抽出できていない。この問題を解決するため、フレームを色特徴量が類似している領域ごとに分割する。

(2) 背景除去処理と領域分割処理の統合

2つの処理の結果を統合し、フレーム内における主体を検出する。背景除去処理によって、実際の主体の輪郭付近に未除去の画素が多く残る。これに基づき、領域分割処理によって得られた領域のうち、領域全体における未除去画素の割合と、領域のエッジ部分における未除去画素の割合について、実験により決定した閾値以上である領域を主体の領域とする。

5. 評価実験

本手法を適用して区間ごとに主体を検出することにより、本手法の有効性を考察する。用いた映像は、フレームサイズ 160×120 [pixel]、フレームレート 30 [fps]、 24 ビットカラーである。

5.1 実験方法

2本の映画の動く主体を追跡しているパン、チルト6区間の映像について実験を行う。区間の10フレーム(約0.3秒)ごとに、検出された領域における実際の主体領域の割合を非過剰率、実際の主体領域における検出された領域の割合を包含率として、6区間の合計54フレームにおいて本手法を適用した。

5.2 実験結果

非過剰率を図3(a)、包含率を図3(b)に示す。全体的に非過剰率に比べて包含率の割合が高いフレームが多い原因として、今後フレーム間で検出した主体を追跡する際に、実際の主体領域より少なめに抽出された場合は主体を追跡していくことが困難であるが、実際の主体領域より多めに抽出された場合は、フレーム間で主体の位置があまり変化しないことを利用して主

体でない領域は除去できるので、実際の主体領域より多めに抽出されるように背景除去処理と領域分割処理を統合する閾値を決定していることが挙げられる。

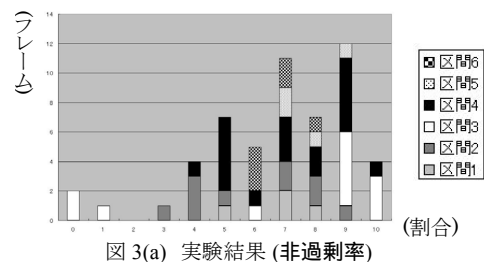


図3(a) 実験結果 (非過剰率)

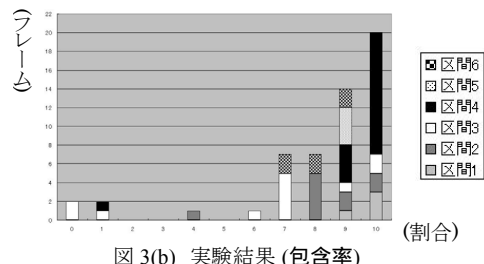


図3(b) 実験結果 (包含率)

5.3 考察

本研究で対象とした映画などの映像では、検出する主体の種類や大きさ、撮影する環境が限定されておらず、一般的に被写体の正確な抽出が困難であることから、検出精度は低くなってしまふ。しかし区間5や区間6では、実験を行った全フレームにおいて非過剰率、包含率ともに高い値であり、区間内のフレームで実際の主体領域の多くが抽出されていることから、これらの区間では主体を検出し、追跡していくことが可能であるといえる。また映像の内容検索やインデキシングは視聴者への閲覧支援が目的であり、最終判断は人間によるため必ずしも正確な抽出は望まれていない。よって本手法は、カメラワーク区間での演出の対象を検出する方法として有効であると考えられる。

6. おわりに

本研究では、映像内における演出対象となる被写体を検出する手法を提案した。その際に、背景の動きと主体の動きが異なり、またカメラワークによって背景の動きが一様になるという性質を利用した。

今後の課題として、現在はパン、チルト区間内における任意の1フレームで主体検出を行っているが、これを区間全体に対して処理を行うことで、パン、チルト区間における同じ主体を追跡していくことが挙げられる。

参考文献

- [1] 松井亮治, 吉高淳夫, 平嶋宗: “カメラワークを利用した演出効果の抽出”, 信学技報, PRMU2004-163-178, Vol.104, No.573, pp.25-30, Jan. 2005.
- [2] 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.
- [3] Chang Yuan, Yu-Fei Ma, Hong-Jiang Zhang: “Extracting Video Object’s Motion Trajectory by Velocity Voting”, in proc. of IEEE Pacific-Rim Conference on Multimedia, pp.452-454, Dec. 2003.
- [4] 奥村真澄, 高木真一, 小館亮之, 富永英義: “動き補償と色情報を組み合わせた MPEG 映像からの人物領域抽出”, 情報処理学会研究報告, 2002-AVM-36, No.036-006, pp.31-36, Mar. 2002.