H-072

顔画像クラスタリングとカメラモーション検出を利用した撮影者の注視対象を含むキーフレーム選択

Intentionally Captured Object based Keyframe Selection using Face Clustering and Camera Motion Detection

石川 真澄† Masumi Ishikawa
野村 俊之† Toshiyuki Nomura

1．まえがき

蓄積機器の大容量化と低価格化により、放送局や家庭に大量の映像が蓄積されるようになった。これらの映像から編集点や視聴点を素早く見つけるために、映像の内容を瞬時に把握したいという要望がある。

映像を視聴せずに素早く内容を把握するための技術として、キーフレーム選択技術が研究されている。映像の視覚的特徴の変化を利用する手法[1]では、キーフレームに主要な対象が映っているとは限らない点が問題となる。また、カメラモーションに基づく手法[2]は、パンやズームで何度も撮影された対象についてフレームが冗長に選択される点が問題となる。また、我々はこれまでに、登場人物[3]や、追尾して撮影された移動対象であるフォロー対象[4]を含む少数のフレームを選択する手法を提案した。しかし、[3][4]では人物以外の静止した対象についてキーフレームを選択できない点が問題であった。

そこで本稿では、顔画像クラスタリングとカメラモーション検出を利用することで、撮影者がフィックスやフォローによって注撮影した対象を把握可能とするキーフレーム選択手法を提案する。テスト映像 20 本に本方式と従来法を適用し、キーフレームの抽出精度を比較する。

2．従来法

2.1 視覚的特徴に基づく手法(従来法 1)

ショットの先頭フレームを第一のキーフレームとし、ショット内のフレームを時系列順に走査する。前に選択したキーフレームとの視覚的特徴の変化量が閾値を超えたフレームを、次のキーフレームとして選択する[1]。

この手法は注視対象を特定しないため、選ばれたフレームに注視対象が含まれるとは限らない点が問題である。

2.2 カメラモーションに基づく手法(従来法 2)

ショットをパン・ズーム・フィックスの区間に分割する。閾値以上の長さの区間について、隣接区間との接続パターンと区間内の並進移動量またはズーム量をもとに選択枚数を決定する[2]。

この手法では、フォロー撮影時のように同じ対象をパンやズームで何度も撮影した場合に、同じ対象のフレームが重複して選択される点が問題となる。

2.3 登場人物に基づく手法(従来法 3)

各フレームから顔を検出し、人物ごとにグループ化する。グループ内の顔を含むフレームの時間位置からグループに対応する人物の登場時刻を推定し、登場時間が閾値以上の人物を網羅するフレームセットを選ぶ[3]。

この手法では、顔が検出されない人物や人物以外の対象を検出できない点が問題となる。

† NEC 情報・メディアプロセッシング研究所

2.4 フォロー対象に基づく手法(従来法 4)

移動軌跡と色情報をもとにフォロー領域を検出する[4]。フォロー領域を含むフレームが規定時間以上継続した区間をフォロー区間とし、各区間から 1 枚選択する。

この手法では、静止した対象を検出できない点や、複数回フォローされた人物について重複してフレームが選択される点が問題となる。

3．提案法

3.1 概要

撮影者は、一般に注視対象を画面内に捉えるようにカメラ操作を行う。対象が移動する場合はフォロー、静止している場合はフィックスで撮影する。複数の対象や大きい対象を撮影する際はパンやズームを利用するが、動きのあるカメラ操作の前にフィックスを挿入する撮影ルールが存在し[5]、プロの撮影者はこれに従う傾向がある。

そこで提案法は、移動および静止した注視対象を網羅するように、キーフレーム選択の単位として、フォロー撮影された区間、フィックス撮影された区間、その他の区間に映像を分割する。同じ対象を含むキーフレームを重複して選択しないように、区間の接続パターンと各人物の登場時刻をもとに、他の区間と異なる人物および対象を含むと推定された区間からキーフレームを選択する。

3.2 処理の流れ

提案法の処理は 3 ステップで構成され、視覚的特徴量の軌跡[6]に基づいて区切られたショットごとに実行される。

(1) カメラモーション検出

(1-1) フォロー区間検出

フォロー領域を含むフレームの連続区間を検出する[4]。

(1-2) 静止区間検出

[4]の過程で算出される隣接フレーム Fi-1 と Fi の間のズーム量および水平および垂直方向の各並進移動を表す各カメラワークパラメータ{*zi*,*dxi*,*dyi*}が以下を満たすとき、フレーム Fi を静止フレームとする。

$$|zi-1| \le THz \text{ かつ } |dxi| \le THp \text{ かつ } |dyi| \le THp \quad (1)$$

フォロー区間以外で、静止フレームが規定値以上継続する区間を、静止区間として検出する。

(1-3) 動区間検出

フォロー区間と静止区間以外の区間を動区間とする。

(2) 顔画像クラスタリング

[3]を利用して各人物の登場時刻を推定する。

(3) キーフレーム選択

区間の種別に応じて以下のように処理を分ける。

(3-1) フォロー区間の場合

区間内に人物が登場しない場合、フォロー領域の重心位置が画面中心に最も近いフレームを選択する。

区間内に人物が登場する場合、フォローされた時間が最

も長い人物を特定する。この人物が選択済みのキーフレームに含まれていない場合、この人物の顔を判別しやすい同一ショット内のフレームを選択する。顔領域から検出される特徴点が少なく顔領域全体がフォロー領域になりにくいことから、顔領域がフォロー領域と重なりを持つ場合に、その人物がフォローされたとする。顔の判別しやすさは、正面性や輝度、大きさ、顔の欠損率に基づいて評価する。

(3-2)　静止区間の場合

フォロー区間に隣接する静止区間は、移動速度が低下した隣接区間中のフォロー対象を撮影した区間であることが多いため、選択処理を行わない。

区間内に人物が登場しない場合、区間の中間フレームを選択する。区間内に人物が登場する場合、両目の中間座標の区間平均がフレームの中心に最も近い人物を特定し、(3-1)と同様にこの人物を含むフレームを選択する。

(3-3)　動区間の場合

フォロー/静止区間に隣接する動区間は、次の対象に向かってカメラを旋回した区間や対象の一部を撮影した区間であることが多いため、選択処理を行わない。ショット全体が動区間の場合は、区間前後のフィックスの撮影漏れと判断し、区間の開始点と終了点からそれぞれ区間の長さの $1/THm$ 離れたフレームを選択する。

4．評価実験

4.1　実験条件

素材映像 10 本とホームビデオ 10 本(各 1 分, 320x240[pix], 29.97[fps])を用いて、提案法と従来法 4 種のキーフレーム抽出精度を比較する。なお、従来法 1 では、[6]の視覚的特徴量を用いた。従来法 2 の併進移動量とズーム率には[4]で算出するカメラワークパラメータを利用した。実験では THz=0.01,THp=1.0,THm=5 を用いた。

4.2　評価尺度

キーフレーム抽出精度を再現率・適合率で評価する。

$$\text{再現率} = \frac{\text{正検出されたキーフレーム数}}{\text{正解のキーフレーム数}} \quad (2)$$

$$\text{適合率} = \frac{\text{正検出されたキーフレーム数}}{\text{抽出されたキーフレーム数}} \quad (3)$$

なお、手動付与した注視対象の登場区間をもとに、注視対象を網羅する最小枚数のキーフレームセットを選び、正解とする。注視対象の登場区間から検出されたキーフレームを正検出とし、同じ注視対象の登場区間からキーフレームが複数選択された場合は一枚のみ正検出とする。

4.3　結果と考察

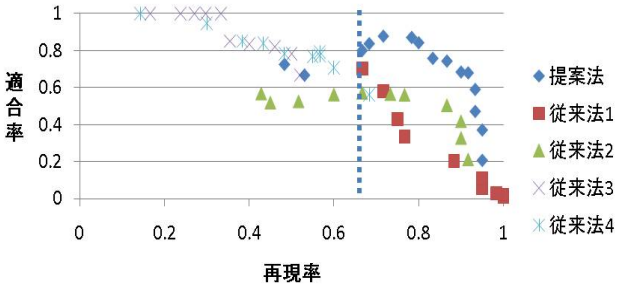
図 1 に、各手法における閾値を変化させた際のキーフレーム抽出精度を、ホームビデオと素材映像に分けて示す。

提案法は、再現率 0.8 のときの適合率は、素材映像で 0.84、ホームビデオで 0.63 であった。素材映像の方がホームビデオよりも高い精度を得た理由は、素材映像ではフィックスで手ぶれが少なく、フォローやパンやズームの際に一定速度でカメラを動かしているため、フォロー/静止/動区間の判定が精度良く行えるためである。

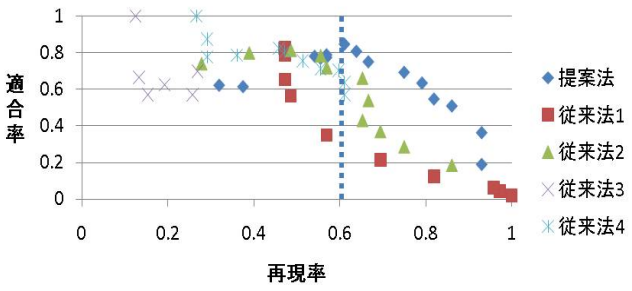
提案法と従来法 1,2 とを比較すると、再現率が素材映像で 0.67 以上、ホームビデオで 0.60 以上のとき、提案法の方が高い適合率であった。これは、提案法ではフォロー区間内では視覚的特徴やカメラモーションの変化の有無に拠らずフレームを選択しないことによる効果である。また、提案法と従来法 3,4 とを比較すると、提案法の方が高い再現率を得た。これは、提案法が顔やフォローに加えて、静止対象についてフレームを選択するためである。

5．まとめ

本稿では、カメラモーションと顔画像クラスタリングを利用した撮影者の注視対象を含むキーフレーム選択法について報告した。本方式は、カメラモーションの種別をもとに区切られた区間をキーフレームの選択単位とすることで、フォローや静止で撮影された対象を含むフレームの選択を可能とする。冗長フレームの選択を抑制するために、顔画像クラスタリングで推定される各人物の出現時刻と各区間の接続パターンに基づき、他の区間と異なる対象を含む区間からフレームを選択する。素材映像 10 本とホームビデオ 10 本に適用した結果、再現率 0.80 のときの適合率は素材映像で 0.84、ホームビデオで 0.63 であり、従来法より高い値を得た。



(a)　素材映像



(b)　ホームビデオ

図 1. キーフレーム抽出精度

参考文献

[1]Minerva M. et al., “Efficient Matching and Clustering of Video Shots”, ICIP’95, vol.1, pp.338-341, October 23-26, 1995.
[2]M. Guironnet et al., “Video Summarization Based on Camera Motion and a Subjective Evaluation Method”, EURASIP Journal on Image and Video Processing, vol.2007, April 23 2007.
[3]石川他, “ショット毎の登場人物の把握に適した代表画像抽出・閲覧システムの構築”, FIT’08, H-008.
[4]石川他, “特徴点の移動軌跡と色情報を利用したフォロー対象把握のためのキーフレーム抽出方式”, IEICE’10, D-12-32.
[5]横田栄治,”ビデオ制作技法”, 映像新聞社, 1987.
[6] 岩元浩太,山田昭雄, “多次元特徴空間解析に基づく映像のカット検出手法”, FIT’05, I-027.