

自撮り動画像の背景変動を利用した人物領域抽出

中川裕太[†] 河野恭之[†]

本研究は手持ちカメラを撮影者自身に向けて撮影した“自撮り”動画像から人物領域を推定して抽出する手法を提案する。撮影者自身と背景を写せる自撮りは手軽に自身の体験や状況を他者と共有できる。しかし一般的な自撮りではレンズと撮影者の距離が腕の長さに制限されてしまうため、どれも背景が広く写らない自撮り特有の構図になってしまう。そこで本研究では自撮り動画像から人物領域だけを自動で抽出し、加工や修正、または合成素材に利用しやすい状態にして編集することで自撮り特有の構図に縛られない自由な自撮り写真の作成に活用するための人物領域抽出手法を提案する。自撮り動画像はカメラを持った腕を前方に伸ばして撮影者自身が写り込むようにし、カメラは撮影者自身に向けたまま身体が中心軸になるように体を回転させて撮影する。撮影者自身はフレームのほぼ中央に写り込み続けてあまり時間軸に対して変動しない人物領域となるが、その周囲の背景領域ではフレーム内に写り込む背景が時間軸に対して大きく変動する。また撮影者の背面に隠れて写り込まなかった背景空間がカメラの回転移動により背景領域に写り込むため、前景である人物領域とその背面に隠れていた背景の差分を検出できる。これら背景領域と人物領域の特徴に着目して各領域を推定し、推定結果を基に GrabCut アルゴリズムを利用して人物領域を抽出する。

1. はじめに

本研究は手持ちカメラを撮影者自身に向けて回転移動させながら撮影した“自撮り”動画像から人物領域を推定して、推定結果を基に GrabCut アルゴリズムを用いて抽出する手法を提案する。

1.1 研究背景

一枚の写真に撮影者自身と景色や建造物などの背景を同時に収められる自撮りは自身の体験や状況を他者と共有するのに手軽な表現手法である。自撮りを行う機会はスマートフォンや SNS の普及で急速に増加している。しかし一般的なカメラ機能を用いた自撮りではレンズの背面側にあるモニタが確認できないため意図した構図で撮影しづらい。またレンズと撮影者の距離が最長でも腕の長さに制限されるため背景を広く映し込むことも困難である。これらの問題はモニタと同じ側のレンズや広角レンズの使用で解決できるが、一脚などの器具を使用せずに自身の腕だけを伸ばしてレンズを自身に向ける撮影手法ではどれも自撮り特有の構図になってしまう。撮影者自身の上半身がフレーム内の領域をほぼ占めてしまい、背景がその周りに少しか写り込んでいる自撮り特有の構図は面白みに欠ける。

撮影した写真を画像編集アプリケーションで加工または修正することは近年当たり前のように行われている。編集する領域を限定すると細かな合成や加工、修正が可能になるため、画像中から任意の領域だけを抽出する機能が求められている。そこで本研究では自撮り写真から前景である人物領域だけを自動で抽出し、自撮り写真を加工や修正、または合成用の素材として利用しやすい形態にする。抽出した人物領域だけを画像編集アプリケーションで加工したり別途撮影した写真と合成したりして、通常の自撮りよりもユーザの意図した情報を付加できる画像を生成する。例えばプリクラのように人物領域を修正して背景領域を自由に差し替えられる写真や、複数の写真を一つにまとめたコ

ラージュ作品などへの利用が考えられる。

1.2 関連研究

主に前景抽出を目的とした領域分割手法に Boykov らの考案した Interactive Graph Cuts [1]が挙げられる。前景と背景のそれぞれ一部分をユーザが指定して、画素間の色差を基に各領域での画素間結合度の重みモデルを生成し、画像をグラフに見立ててグラフカットを行う。グラフカットとは重み付き有向グラフのコストが最小になるように有向グラフを切断することであり、この切断箇所は画像において前景領域と背景領域の輪郭になりうる。またユーザが最初に指定した前景または背景の一部を seed と呼ぶ。Rother らの考案した GrabCut [2]アルゴリズムでは前景と背景の色分布を重み付きで再学習しながら領域分割を繰り返して精度を向上させている。色分布をモデル化するため、複雑な色差を持つ背景であっても高精度に分離できるとされる。GrabCut では前景を含む矩形範囲を指定するだけで矩形範囲外を全て背景 seed とみなして背景の色分布をモデル化するため、とくに細かな seed を与えることなく容易に前景抽出ができる。しかしユーザが前景と背景の一部に正しく seed を与える方が前景抽出の精度が高く、間違った抽出結果を調整するために手で seed を与えている。

新井らは GrabCut を用いて前景と背景を分離し、アフィン変換で傾けた背景に前景を合成して擬似立体サムネイルを生成した[3]。大量の画像を管理する上で元の画像を縮小したサムネイルの利用は避けられないが、注目被写体である前景が背景より強調された画像は視認性と被発見性が向上する。前景を含む矩形範囲の選択には検出器を用いているが、必ずしも意図した前景範囲を選択できるとは限らない。しかし構図が限られた自撮りでは前景である人物領域のおおよその位置と範囲を推定できる。カメラを撮影者の身体を中心にして回転させながら撮影すると、フレーム内の背景領域と人物領域で時間軸変動差異が生じ、また前景である人物領域の背面に隠れていた背景と人物の差分から背景領域と人物領域が推定できる。本研究では自撮り動画

[†] 関西学院大学大学院 理工学研究科 人間システム工学専攻

像の特徴を活かして背景領域と前景である人物領域を推定し、推定結果を GrabCut の seed に用いた人物領域抽出手法を提案する。

2. 提案手法

GrabCut を用いて自撮り動画像中の任意のフレームから人物領域を抽出する際に前景である人物領域と背景領域を推定し、推定結果を seed に用いて抽出結果の精度を向上させる。自撮りを行う際にカメラを撮影者の身体を中心にして回転移動させることで発生する動画像フレーム内の背景領域と人物領域の時間軸での変動の差異に着目し、変動の大きい領域が背景領域、変動の小さい領域が前景に相当する人物領域だと推定する。また撮影者の背面に隠れて写り込まなかった背景空間がカメラの回転移動によりフレーム端の背景領域に写り込んでいくため、それらを繋ぎ合わせて生成した擬似背景画像と同じ空間方向での人物が写り込んだフレーム画像との差分が前景である人物領域だと推定する。得られた推定結果をそれぞれ背景領域と人物領域の seed に設定し、GrabCut を用いて人物領域を抽出する。図 1 に提案手法の概要を示す。

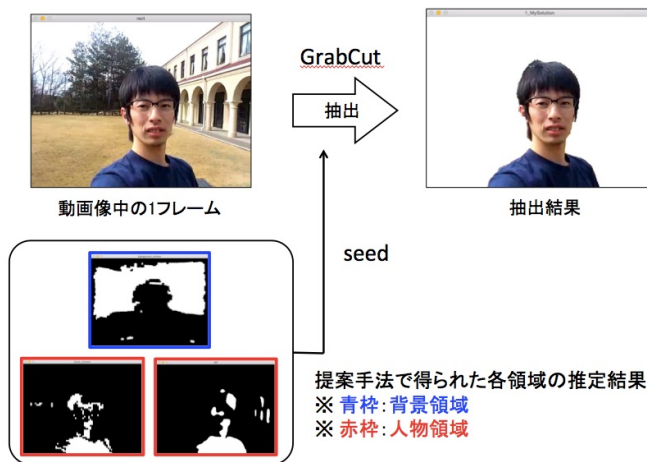


図 1 提案手法の概要

2.1 自撮り動画像

図 2 のようにカメラを持った腕を前方に伸ばして撮影者が写り込むようにし、カメラは撮影者に向けたまま身体が中心軸になるように腰または肩を回転させて撮影した動画像を自撮り動画像とする。本研究では無理なく自然体で撮影するため、上から見て反時計回りに約 45 度から約 90 度の回転を約 2 秒かけて撮影した動画像を対象としている。フレーム内での人物が占める人物領域の割合はレンズの水平面角に依存するが、平均的な肩幅を約 40cm、腕の長さを約 60cm とした場合、レンズの水平面角が約 37 度以上あれば人物の両肩がフレームに収まる。本研究では一般的なスマートフォンの背面カメラを用いて撮影し、640*480 画素、30fps の動画ファイルを PC 上で領域抽出処理プログラムに

読み込んでいる。

2.2 背景領域と人物領域の特徴

自撮り動画像中の各フレーム内は背景領域と撮影者自身が写っている人物領域に分けられる。背景領域はカメラの回転移動に伴って撮影者を介した向こう側にある背景物体がフレーム内で横軸方向へ移動する。上から見てカメラは反時計回りに回転しているため、背景領域中の背景物体はフレームの右から左方向へ移動する。一方、人物領域はフレーム内のほぼ中央でおおむね固定されて移動しない。人物領域の形状は風や手ぶれ変動の影響で変形するが、数秒の撮影時間内での時間軸変動は僅かで大きな変動はない。

屋外の開けた場所で撮影した自撮り動画像の場合、背景領域は図 3 のように背景に含まれる被写体の特徴と存在場所から空領域、背景領域、地面領域の 3 つの領域に分類できる。空領域は高輝度のまま変動しない領域であるが、撮影場所によっては建物や樹木で遮られている箇所が存在するため時間軸変動が検出される。背面領域は撮影者の背面に存在する建物や植物などが占めており、形状や輝度の変化が大きく最も時間軸変動が検出されやすい領域である。地面領域は背景領域の下端部分を占有する地面であり、空間変化が乏しく時間軸変動が検出しづらい領域で、人物領域との変動の差異が現れにくい。背景中の各領域が占める割合は撮影場所やカメラの撮影角度に影響され、またカメラの回転移動で背面に存在する建造物とカメラとの距離が変化するため背面領域の占める割合は時間軸で変化する。

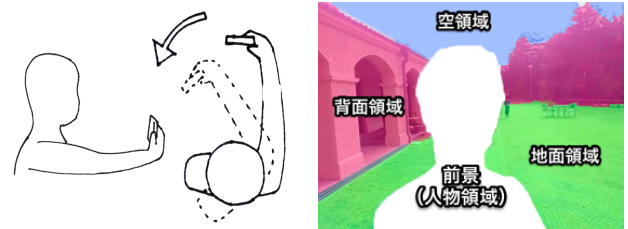


図 2 (左) 自撮り動画像の撮影手法

図 3 (右) 自撮り動画像フレーム内の各領域

2.3 時間軸変動の差異に着目した領域推定

背景領域と前景である人物領域では特徴点移動量と輝度の時間軸変動に差異があるため、これらに着目した 2 つの手法で人物領域を推定する。特徴点移動量の差異に着目する手法では特徴点トラッキングで検出した各特徴点の移動軌跡を背景領域と人物領域にクラス分けをして人物領域の推定を行う。また輝度の時間軸変動に着目する手法ではフレームを 8*8 画素のブロックにメッシュ分割して各ブロック内平均輝度の時間軸での分散を算出し、分散が小さい、つまり輝度変動の小さい領域は人物領域だと推定する。自撮り動画像からフレームを順に取得し、人物領域を推定するための特徴点トラッキング処理と分割ブロックごとの時間軸変動検出処理を行う。フレーム取得では 30fps から

10fps になるようにフレームを間引き、フレーム間での背景領域の変動が現れやすくしている。取得したフレームから背景と人物の時間軸での変動の差異を検出し、背景領域と人物領域をそれぞれ推定する処理の流れを図4に示す。

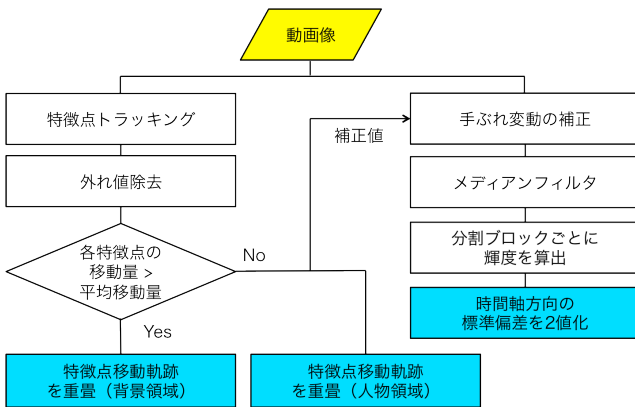


図4 時間軸変動の差異に着目した推定処理の流れ

2.3.1 特徴点トラッキングを用いた領域推定

背景領域の特徴点移動量と人物領域の特徴点移動量の差異から背景領域を推定する。被写体のフレーム内での位置を追跡するとフレーム内での移動量がわかる。自撮り動画では背景に属する空や建造物、地面などの被写体はカメラの回転移動に伴ってフレーム内を移動するが、人物はフレームほぼ中央から動かない。被写体のフレーム内での移動量がわかれば、その被写体が写り込んでいるフレーム内領域が背景領域なのか人物領域なのか判断できる。被写体の動きを追跡するためには連続2フレームに写っている同一の被写体を対応付ける必要があるため、特徴点を用いた対応づけを行って被写体の動きを追跡する。特徴点の抽出は SURF 特徴量を用いて行う。SURF(Speeded-Up Robust Features) [4]は画像中の特徴となる点の抽出をし、その特徴点の周辺領域の輝度勾配の方向とスケールを特徴記述するアルゴリズムである。

フレーム間で特徴点を対応させるため SURF 特徴量で抽出された各特徴点の特徴ベクトルを各次元で比較し、ユークリッド距離が最小のものをフレーム間での特徴点ペアにする。SURF 特徴量は回転やスケール変動、照明変動に不変とされるため、自撮り動画のような手持ち撮影された動画の連続フレームで特徴点を対応づけるのに有効である。この対応づけを連続2フレームで毎回行い特徴点トラッキングをする。対応づけされた特徴点ペアの中には誤対応が含まれているため、特徴点の対応づけ範囲を限定して誤対応を除外する。自撮り動画中の背景領域は横軸方向へ移動しているだけで縦軸方向へ被写体が大きく移動することは基本的にないため、縦軸位置のユークリッド距離が大きい特徴点ペアをしきい値処理で除外する。縦軸位置の大きな変動がない対応セットの中にも誤対応が含まれている。

こうした誤対応を除外するために、前述のしきい値処理で除外されずに残った対応セットの特徴点の横軸位置のユークリッド距離の平均値と分散を求め、その距離が平均値から分散より大きく外れている特徴点ペアは除外した。

誤対応を除外した各対応セットの特徴点の横軸位置のユークリッド距離をその特徴点の横軸方向への特徴点移動量として、特徴点を背景領域と人物領域にクラス分けする。横軸方向のみに着目する理由は、縦軸方向への特徴点移動は手ぶれ変動の影響を受けているのみで背景領域と人物領域の特徴点移動量の差異が発生しないためである。得られた特徴点ペアのうち背景領域で検出された特徴点は横軸方向へほぼ等しい距離を移動しているのに対して、人物領域で検出された特徴点ペアはほぼ移動していないため、特徴点移動量の平均値をしきい値に用いて特徴点を背景領域と人物領域にクラス分けした。背景領域にクラス分けされた特徴点移動を青色、人物領域にクラス分けされた特徴点移動を赤色で表示したものが図5である。移動量が大きい特徴点はおおむね背景領域で検出されているが、移動量が小さな特徴点は人物領域だけでなく誤対応を全て除去できずに背景領域でも検出されている。とくに背面領域中の樹木では特徴点の誤対応が発生しやすい。

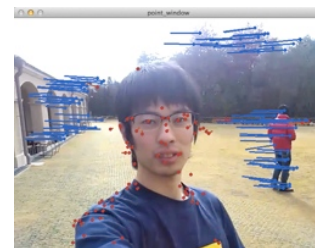


図5 特徴点移動のクラス分け結果

前記の処理をフレーム間引きした動画の全ての連続する2フレームに対して行い、背景領域クラスの特徴点移動軌跡を全て重畳した結果が図6である。背景領域にクラス分けされた特徴点移動軌跡の重畳は空領域と背面領域、地面領域で異なる。空領域は背面領域の被写体が横切ることによって特徴点の軌跡が現れ、背面領域は建物や植物など輝度勾配が現れやすい被写体が多いため全体的に特徴点移動軌跡が検出される。しかし地面領域は輝度勾配が小さいため特徴点検出されにくいという問題があり、特徴点移動軌跡が現れていない。また誤対応で背景領域にクラス分けされた特徴点移動軌跡が人物領域を横切るように現れてしまう。図7は人物領域クラスに分けられた特徴点移動軌跡の重畳結果である。人物領域クラスでは本来ならば背景領域にあたる領域にも特徴点誤対応の影響で短い特徴点移動軌跡が現れてしまっているが、背景領域にクラス分けされた特徴点移動軌跡の否定値と論理積をとれば人物領域のみを残せる。地面領域では特徴点検出されづらいため、背景領域

とも人物領域とも推定できない領域が残ってしまう。



図 6 (左) 背景領域の特徴点移動軌跡の重畳結果
 図 7 (右) 人物領域の特徴点移動軌跡の重畳結果

2.3.2 輝度の時間軸変動を用いた人物領域推定

各フレームを 8*8 画素に分割したブロックごとの平均輝度から時間軸での分散を算出し、時間軸変動の小さい領域を人物領域として推定する。自撮り動画は手持ち撮影で発生した手ぶれ変動の影響を受けている。各領域の横軸方向への変動の差異に着目するにはこの縦軸方向への被写体の変動を取り除く必要があるため、縦軸方向への手ぶれ変動の補正処理を行う。各フレーム間での特徴点トラッキングで得られた特徴点のうち、人物領域にクラス分けされた特徴点の縦軸方向への移動量の平均値を縦軸方向への位置ずれとみなし、フレームを移動量に合わせて上下に平行移動させて補正する。また被写体の移動がない領域でも撮影角度による照明変動や手ぶれ変動の影響で各画素単位での輝度はフレーム間で変動するため、各フレームをメディアンフィルタで平滑化して輝度値の変動を抑える。平滑化で注目する各画素の周辺領域は周辺 9*9 画素とした。

手ぶれ変動の補正とメディアンフィルタで平滑化したフレームを 8*8 画素のブロックに分割し、各ブロック内画素の平均輝度の時間軸での分散を算出する。この処理を全フレームで行ってブロックごとの時間軸での分散を算出する。輝度の時間軸変動を確認しやすくするために分散の値を 256 階調に正規化した結果を図 8 に示す。被写体の移動で輝度の時間軸変動が大きくなる背景領域は分散が大きくなっているのに対して、輝度の時間軸変動が小さい人物領域は分散が小さくなっている。

背景領域と人物領域での輝度の時間軸変動の差異は必ずしもはっきりするとは限らないため、既知の背景領域での値を基に 2 値化する。2 値化で用いるしきい値は分散の大きい背景領域と分散の小さい人物領域を分割できる値にする必要があるが、この手法では地面領域と人物領域を分割するしきい値を設定する。地面領域は輝度の時間軸変動が小さく分散も小さくなってしまいうため、大津の手法[5]のようにフレーム全体の値からしきい値を算出すると地面領域と人物領域が同じクラスに属してしまう。しかし地面領域はフレームを縦方向に 4 分割した一番下の領域をさらに横方向に 3 分割した両端の領域を全体にわたって占めており、

ほぼ同じ輝度変動の分散が現れている。この領域の輝度変動の分散値の分散を算出してしきい値に用いることで、地面領域を除外して輝度変動の分散を 2 値化できる。2 値化した値を反転した結果を図 9 に示す。輝度の変動が小さかった領域が人物領域だと推定できる。



図 8 (左) 輝度の時間軸変動の分散
 図 9 (右) 地面領域を基に 2 値化した結果

2.4 擬似背景画像との差分を用いた人物領域推定

人物領域の左右の領域は背景領域であるため、その背景領域だけを用いて擬似背景画像を生成し、同じ空間方向でのフレームとの差分が人物領域だと推定できる。フレームの左右両端は背景領域が上端から下端までを占めるため、この背景領域矩形を 2.3.1 節の特徴点トラッキングで得た背景の横軸方向への移動量に従って繋ぎ合わせて擬似背景画像を生成する。図 10 に擬似背景画像の生成手法の概要を示す。カメラを撮影者自身に向けて身体を中心に回転させながら撮影する自撮り動画では撮影者の周囲を広範囲に渡り写し込めるため、フレームと同サイズの擬似背景画像が生成できる。手ぶれの影響でフレーム貼り合わせ位置が上下左右ともに少量ずれているが、およその背景画像が得られる。

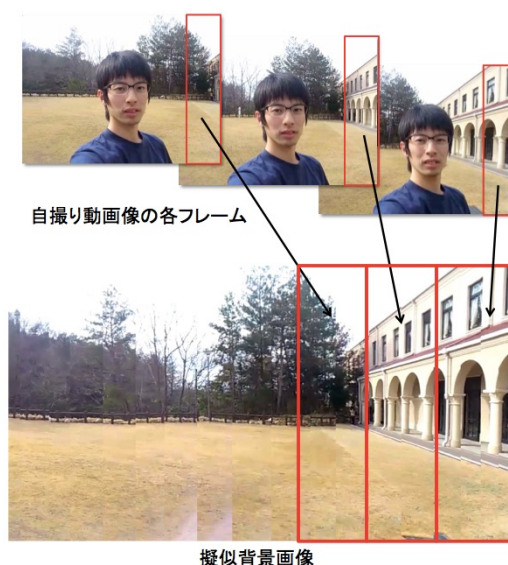


図 10 擬似背景画像の生成手法

一般的な背景差分法とは違って固定カメラではなく手持ちカメラで撮影したフレームを繋ぎ合わせて擬似的に生成した背景画像を用いるため、背景領域でも差分が発生する。背景領域ではとくに被写体間の領域境界で大きな差分が発生する。背景領域での差分を小さくするため、擬似背景画像と人物が写っているフレームそれぞれにガウシアンフィルタで平滑化処理を行った。図 11 に平滑化した擬似背景画像と人物が写り込んだ同じ空間方向のフレームを示す。また差分値を 256 階調に正規化した画像と、中央値をしきい値に用いて 2 値化した結果を図 12 示す。平滑化処理で領域境界が曖昧になったため大きな差分値は現れなくなった。差分を人物領域の seed に用いるため 2 値化する必要があるが、背景領域と人物領域で差分値の差異が明確ではないため差分値の中央値をしきい値に用いた。

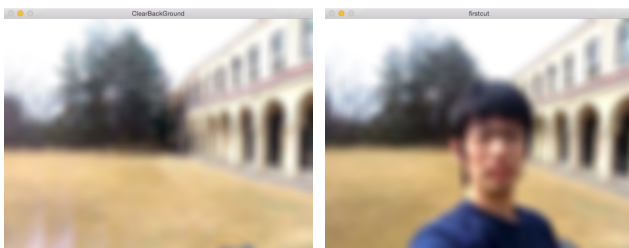


図 11 平滑化した擬似背景画像とフレーム

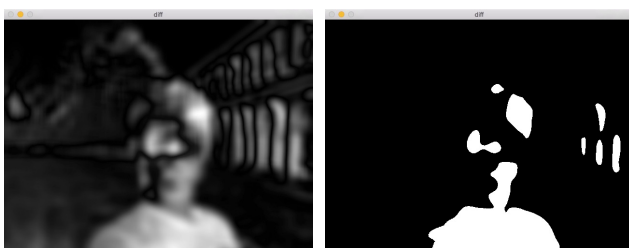


図 12 差分をとった結果(256 階調)と 2 値化した結果

2.5 人物領域の抽出

2.3 節と 2.4 節で得られた推定結果を GrabCut の seed に与え、自撮り動画中のフレームから人物領域の抽出を行う。GrabCut は画素の RGB 値から色分布をモデル化してグラフカットを繰り返すため、複雑な背景でも高精度に領域分割ができるとされる。通常の GrabCut では前景を含んだ矩形範囲を選択して矩形範囲外を全て背景と仮定するが、前景は矩形範囲内に存在するという初期情報しか与えられない。背景領域と前景である人物領域を概ね推定できる場合は矩形範囲に加えて推定結果を seed として与えると、従来よりも正確な領域分割が期待できる。

背景領域と人物領域それぞれの推定結果が得られたが、一部領域では推定結果が重なっており、また別の一部領域ではどちらの推定結果も得られていない。推定結果が重なっている場合はどちらかが誤った推定結果である。輝度の時間軸変動を用いた人物領域推定や擬似背景画像との差分

を用いた人物領域推定では、カメラの手ぶれなどの影響で自撮り動画のフレーム内の時間軸変動が大きい場合に背景領域を誤って人物領域だと推定しまう可能性が高い。しかし特徴点トラッキングを用いた背景領域推定では、背景領域の特徴点移動軌跡が誤対応の影響で人物領域に現れることもあるが、縮小処理と膨張処理を用いると点在する移動軌跡を除去して密集している移動軌跡の隙間を埋めることができる。特徴点トラッキングを用いた領域推定で人物領域が誤って背景領域だと推定される可能性は他の人物領域推定手法より低いため、推定結果が重なった場合は背景領域を優先する。またフレーム両端は既知な背景領域であるため、フレーム両端から 40 画素の範囲を全て背景領域だと設定した。

背景領域とも人物領域とも推定されなかった領域は輝度の時間軸変動は大きかったが特徴点移動が検出されなかった領域である。この領域は空間上の輝度勾配が小さい被写体が写っている空や地面などの背景だと仮定できるため背景領域の seed を与える。しかし上記のような背景領域の一部にも特徴点トラッキングを用いた領域推定で特徴点の誤対応の影響で人物領域だと推定された領域が多く存在するため、背景領域の一部が誤って人物領域に含まれる。しかし輝度の時間軸変動を用いた人物領域推定と擬似背景画像との差分を用いた人物領域推定では推定できなかった人物領域が推定できている可能性があるため、単純に推定結果を seed から除外できない。本研究では特徴点トラッキングを用いた領域推定で人物領域だと推定された領域を含める “seed_1” と含めない “seed_2” の 2 種類の seed を用意した。特徴点トラッキングから得た人物領域の推定結果も背景領域の推定結果と同様に縮小処理と膨張処理の組み合わせで点在する移動軌跡を除去して密集している移動軌跡の隙間を埋めている。

GrabCut に seed_1 を与えて人物領域を抽出した結果の例を図 13 に示す。灰色の seed が背景領域だと推定された領域、白色の seed が人物領域だと推定された領域、黒色の領域は推定結果が得られなかったため背景だと仮定して seed を与えた領域である。また GrabCut に seed_2 を与えて図 13 と同じ自撮り動画から人物領域を抽出した結果が図 14 である。特徴点トラッキングから得た人物領域の推定結果を含める seed_1 では人物領域の大部分を正しく推定できているが、人物領域の seed が背景領域にも多く現れているため誤って背景領域の一部を人物領域に含んで抽出してしまっている。seed_2 では人物領域の一部を推定できていないが、背景領域に人物領域の seed が現れていないため概ね正しく人物領域を推定できている。背景領域とも人物領域とも推定されなかった人物領域の一部には、左右の地面領域と同じく背景領域の seed が与えられている。図 14 の例では正しく人物領域が抽出されているが、抽出するフレームの人物領域と地面領域の空間上の差異が小さい場合には

輪郭に沿って人物領域を抽出できず、人物領域の一部が背景領域だと判断されて欠けてしまう。

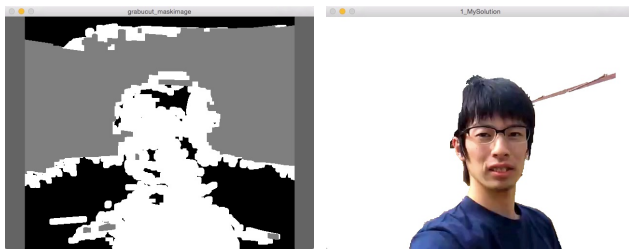


図 13 seed_1 と GrabCut 実行結果

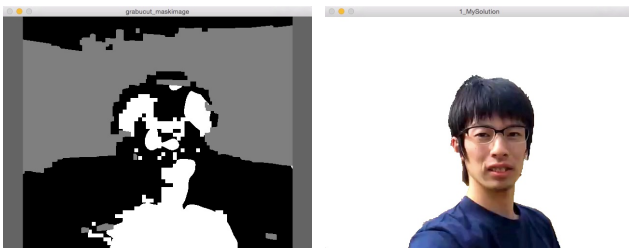


図 14 seed_2 と GrabCut 実行結果

3. 実験と考察

人物領域が含まれた矩形範囲を選択する従来の GrabCut と提案手法で自撮り動画画像中のフレームから人物領域を抽出して精度を比較し、提案手法の評価と考察を行った。以下、矩形範囲の選択だけで人物領域の抽出を行う従来の GrabCut を従来手法と呼ぶ。実験では様々な状況で撮影した 2.1 節の自撮り動画画像 30 本を対象にした。従来手法では人物領域を全て含む矩形はできるだけ小さくした方がよい結果が得られるが、人物領域の大きさは自撮り動画画像毎に違って、理想的な矩形範囲を選択するには手動で範囲選択しなければならない。本実験では自撮り動画画像毎の条件を同じにするため、人物領域を含む範囲はフレーム両端から 40 画素内側の矩形範囲に設定した。また提案手法でも同じくフレーム両端から 40 画素の範囲は背景領域だと仮定する seed を与えている。この条件だと従来手法より提案手法の方が GrabCut に与える seed を多く持っているため抽出精度が高いと考えられるが、誤った推定結果に基づく seed の影響などを従来手法と比較して検証する。2 つの手法で用いる GrabCut の繰り返し回数は事前の知見から 2 回で十分だと判断した。さらに 2.5 節で述べた seed_1 を与えた場合と seed_2 を与えた場合とで人物領域の抽出結果を比較して考察する。

抽出結果を提案手法と従来手法で比較すると 4 つのパターンに分けられるため、それぞれパターン A~D と呼称して表 1 にまとめた。4 つのパターンは提案手法と従来手法の抽出結果の比較のみで分類した。与えた seed の違いで抽出結果が異なるものがあったため、与えた seed ごとの比較結果

を表 2 にまとめた。また seed_1 を用いた抽出結果の例を図 15 に、Seed_2 を用いた抽出結果の例を図 16 に示す。左から順に人物領域を抽出するフレームと与えられた矩形範囲、提案手法で得られた seed、提案手法の領域抽出結果、従来手法の領域抽出結果である。パターン A は抽出結果がほぼ同じだった例で、seed_1 を与えた場合は 30 本中 4 本、seed_2 を与えた場合は 30 本中 3 本が該当する。前景と背景に明確な差があるか、または領域を分断できないほど差がなく、領域の推定結果にかかわらず同等の抽出結果が得られたと考えられる。パターン B は提案手法の抽出精度が従来手法より優れていた例で、それぞれ抽出精度に差はあるが seed_1 を与えた場合は 30 本中 17 本、seed_2 を与えた場合は 30 本中 10 本が該当する。これらは複雑な背景だったため従来手法では上手く人物領域を抽出できなかったが、提案手法では背景の複雑さゆえに時間軸変動が現れやすく、人物領域と背景領域のほぼ全域をそれぞれ推定できたと考えられる。パターン C は両手法とも抽出精度が低い、A と違って両手法の結果が大きく異なる例で、seed_1 を与えた場合は 30 本中 9 本、seed_2 を与えた場合は 30 本中 16 本が該当する。これらは単純な背景と複雑な背景が混合する場合である。従来手法は複雑な背景が苦手な一方、提案手法では単純な空領域や地面領域が存在する場合にそれらの領域を背景領域だと推定できず、時間軸変動の小さい人物領域だと誤推定する欠点がある。また提案手法で人物領域の大部分を推定できていない場合は人物領域を抽出できていない、従来手法では発生しにくい人物領域の欠けが起こる。パターン D は従来手法が提案手法より優れていた場合だが、本実験では該当する抽出結果はない。

表 1 提案手法と従来手法の抽出結果の比較パターン

パターン	抽出結果の比較
A	ほぼ同じ抽出結果
B	提案手法が従来手法より優れている
C	抽出精度が悪いうえ、抽出結果が大きく異なる
D	提案手法が従来手法より劣っている

表 2 提案手法と従来手法の抽出結果比較

従来手法との比較結果のパターン				
与えた seed	A	B	C	D
seed_1	4 本	17 本	9 本	0 本
seed_2	3 本	10 本	16 本	0 本

全体を通して提案手法が従来手法に比べて劣っている場合はなく、およそ半分の抽出結果で提案手法が優れていた。従来手法では人物領域を含む矩形範囲しか情報が与えられ

ず、人物領域と隣接する背景領域が人物領域だと判断されやすいため、提案手法の方が優れた結果を出すのは当然だと考えられる。しかし抽出結果のおよそ半分を占めたパターンCのように提案手法と従来手法で結果が大きく異なる場合も多かった。従来手法では背景領域の情報のみ与えられるため、背景領域の一部が人物領域だと判断されて残ってしまう。提案手法では地面領域を背景領域だと推定することが難しいため推定結果が得られなかった領域を全て背景だと仮定したが、この場合は推定結果が得られなかった人物領域の一部が背景領域だと判断されて欠けてしまう。このパターンCはseed_2を用いた場合に多い。

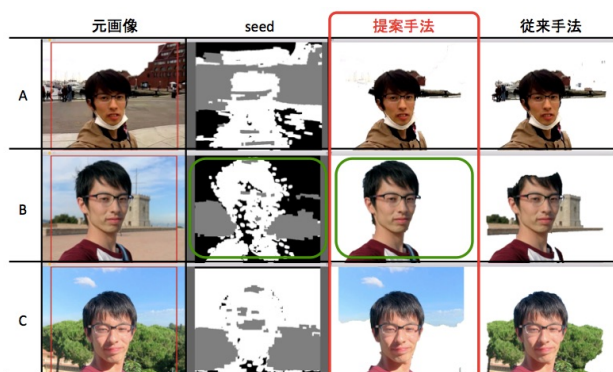


図 15 seed_1 を用いた人物領域の抽出結果

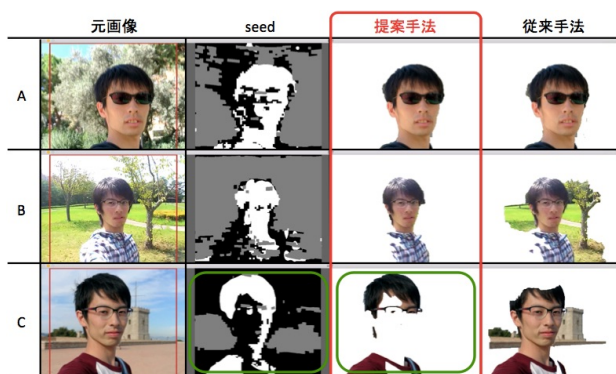


図 16 seed_2 を用いた人物領域の抽出結果

与えた seed の違いで抽出結果のパターンが違うものは 30 本中 14 本であった。seed_1 での抽出結果がパターン B であったが seed_2 ではパターン C になったものが 17 本中 10 本あった。緑色で囲った図 15 の B と図 16 の C がその例である。これらは特徴点トラッキングで人物領域だと推定された領域が他の人物領域推定で人物領域だと推定されずに seed_2 だと人物領域の一部が抽出されなかったが、seed_1 だと抽出されている。また地面領域の一部が特徴点トラッキングで人物領域だと推定されたため、seed_2 ではパターン B だったが seed_1 ではパターン C になったものは 10 本中 3 本あった。本実験では seed_2 を与える方が seed_1 を与えるより従来手法と比較した抽出精度が向上した。特徴点

トラッキングで得られた人物領域の推定結果を seed に含めると、人物領域の一部が人物領域だと推定されずに人物領域の抽出結果の一部が欠ける現象が抑えられる。

人物領域の抽出精度に注目すると、少し背景領域が残ったり人物領域が欠けたりしているが合成などの画像編集に耐えられる許容範囲内で人物領域を抽出できている結果が seed_1 を与えた場合だと 30 本中 10 本、seed_2 を与えた場合だと 30 本中 7 本あった。また従来手法だと 30 本中 1 本のみだった。空領域と地面領域でも空間上の輝度勾配がある自撮り動画像だと背景領域の推定精度が高いため、領域抽出すると背景領域ほぼ全域を除外できる。とくに地面領域と人物領域で明確な差異がある場合、輝度の時間軸変動や擬似背景画像との差分を用いた推定で人物領域のほぼ全域を推定できる。

4. まとめと課題

自撮り動画像中の背景変動を利用して前景である人物領域と背景領域を推定し、GrabCut アルゴリズムに推定結果を seed として与えて自撮り動画像からの人物領域の抽出精度を向上させた。背景の時間軸変動が十分に検出できる場合、提案手法の抽出精度は従来手法に比べて向上した。しかし空領域や地面領域が単純な背景で時間軸変動や人物領域との差異が小さい場合は人物領域が正しく推定されずに抽出精度が低下する。単純な背景であっても前景と背景の時間軸変動の差異を検出できるような領域推定の手法が課題である。また光学的特性を踏まえてより高精度な擬似背景画像を生成できれば同じ空間位置でのフレームとの差分が明確に検出できる。提案手法においては一様ではない地面の上で背後に樹木や建造物のある場所で自撮り動画像を撮影すれば、人物領域と背景領域をそれぞれ推定できるため、そのような場所で自撮りを行えば提案手法は有効である。また自撮りはスマートフォンで行う機会が多いため、加速度センサや地磁気センサなどカメラ以外のセンサ情報を利用するとより高精度な人物領域抽出が期待できる。

参考文献

- 1) Y. Boykov, and M. P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images", ICCV2001, Vol.1, 105, 2001
- 2) C. Rother, A. Blake, and V. Kolmogorov, "GrabCut - interactive foreground extraction using iterated graph cuts", ACM Trans. on Graphics, Vol.23, No. 3, 309-314, 2004
- 3) 新井啓介, 武井宏将, 山名早人, "ThumbPop: 注目物体を強調した疑似立体サムネイル生成", 信学技報. HIP, 111(500), 177-182, 2012
- 4) Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "SURF: Speed Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol.110, No.3, pp.346-359, 2008.
- 5) 大津展之, "判別および最小 2 乗法に基づく自動しきい値選定法", 電子通信学会論文誌, J63-D-4, pp349-356, 1980.