

# 自撮りパノラマ写真生成のための人物領域抽出

中川裕太† 河野恭之†

**概要**：通常の自撮り写真ではカメラのレンズと撮影者自身と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写しこむことが難しい。自撮り写真の背景を広く映し込むことは写真の情報量を多くするだけでなく、1枚の写真で撮影者自身の様子と周囲の状況を見栄え良く伝えられるため、SNSへの投稿など体験の記録や共有に効果的である。本研究では自撮り動画像から人物領域を抽出し、背景のみを繋ぎ合わせた背景パノラマ画像と人物のみの画像を生成して重ね合わせることで従来では撮影しづらかった背景が広く写った自撮り写真を生成する手法を提案する。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腰を回転させて撮影するものとする。この撮影方法で得られた動画像中の背景領域は横軸方向へ変動するのに対し人物領域はほぼ変動しない。この変動の差異を動画像中の特徴点移動量と分割ブロックごとの輝度の時間軸変動から検出して人物領域を推定し、残りの背景領域を背景の移動量に合わせて重ね合わせた背景パノラマ画像に貼り合わせることで自撮りパノラマ写真を生成する。

## 1. はじめに

本研究は自撮り動画像から人物領域を抽出し、背景のみを繋ぎ合わせた背景パノラマ画像と人物のみの画像を生成して重ね合わせることで従来では撮影しづらかった背景が広く写った自撮り写真を生成する手法を提案する。撮影者が撮影者自身を撮影する“自撮り”は手軽な撮影で撮影者自身と背景を同時に映し込むことができる。背景が写り込んでいる自撮り写真は1枚の写真で撮影者自身の様子と周囲の状況を見栄え良く表現できるため、自身の体験や周囲の状況を記録、共有するのに効果的である。自身を撮影するにはカメラを三脚などの固定器具を用いる方法か周囲の他人に撮影を依頼する方法が一般的であった。しかし固定器具を用いる方法では固定器具の持ち運びやセッティングが必要であり、また周囲の他人に撮影を依頼できない状況も少なくない。手軽に自分自身を記録する自撮りはとくにSNSやスマートフォンの普及でますます増加している。

一般的なカメラを用いて上手く自撮りをするのは難しい。芝ら[1]は顔検出とエッジ抽出を利用して顔が中心に収まる鮮明な自撮り支援システムを提案し、レンズの背面にあるライブビューを確認できないため撮影者が上手く中心に映らない問題を解決している。しかしカメラを手で持って撮影する通常の自撮りではレンズと被写体の距離が最長でも撮影者の腕の長さに制限されるため、映し込める背景の範囲はレンズの画角に依存し、ほとんどの場合は人物領域がフレーム内を占有して背景を広く写し込めない。背景を広く映し込むことは写真の持つ情報量を大きくし、また見栄えの良い写真になるため自分の体験や状況を伝えるのに効果的である。そこで本研究は動画で自撮りをして、動画像中のフレームを背景領域と人物領域に分け、背景領域のみの画像を繋ぎ合わせた背景パノラマ画像と任意のフレームから抽出した人物のみの画像を別に用意して後か

ら重ね合わせることで背景を広く自由に映し込める自撮り手法を提案する。

## 2. 提案手法の概要

本研究は自撮り動画像から人物領域を抽出して、背景領域を繋ぎ合わせた背景パノラマ画像と任意フレームの人物領域を重ね合わせて図1のように自撮りパノラマ写真を生成する。自撮り動画像から人物領域を抽出する場合、背景領域には撮影者の背後にある様々な被写体が写り込む。背景画像が既知であれば背景画像との差分を用いて領域抽出する手法もあるが、背景画像の変動を避けるためにカメラを固定しなければならず、カメラを固定しても照明変動や背景中の被写体の形状変化などの影響を受ける。また松崎ら[2]が提案しているオブジェクト追跡手法であるCondensationと輪郭モデルを用いた人物領域抽出手法では、抽出対象の輪郭モデルが既知でなければならない。本研究では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中の背景領域と人物領域の変動の差異に注目することで、時間軸変動の小さい領域が人物領域だと推定して抽出する。

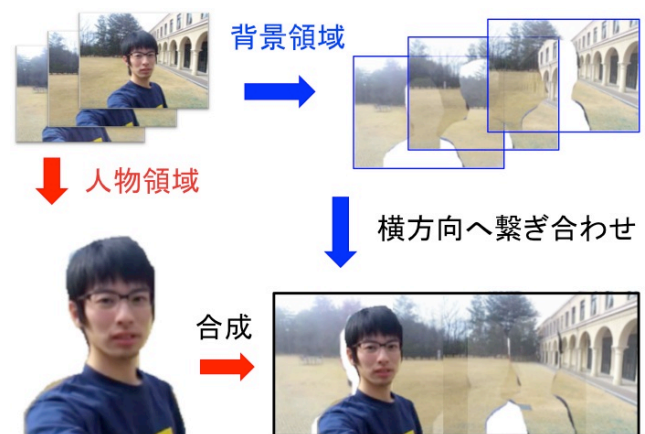


図1 自撮りパノラマ写真の生成手法

† 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

## 2.1 実装

本研究では動画像の処理に OpenCV2.4 を用いる。OpenCV(Open Source Computer Vision Library)は Intel 社が開発したオープンソースの画像処理ライブラリで、画像処理に必要な関数が多数用意されておりマルチプラットフォームで動作する。自撮り動画像の撮影にはソニー・エリクソン・モバイルコミュニケーションズ社のスマートフォン SO-03C のカメラを使用した。カメラは記録画素数約 800 万画素でレンズ画角は水平約 60 度である。動画像は 640\*480 画素の可変フレームレートで mp4 ファイルに保存する。OpenCV2.4 では可変フレームレートの mp4 ファイルが読み込めないため、Adobe 社の Adobe Media Encoder を用いてフレームレート 30fps の QuickTime ファイルに変換したものをシステムの入力とする。

## 2.2 自撮り動画像の撮影方法と特徴

自撮り動画像の制約条件は、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もしくは肩を回転させて撮影したものとする。手ぶれ変動の影響を少なくするため、腕はできるだけ伸ばした状態にする。無理なく自然体で撮影できるようにするため、撮影者を上から見て反時計回りに約 45 度の回転を約 2 秒間で撮影した。

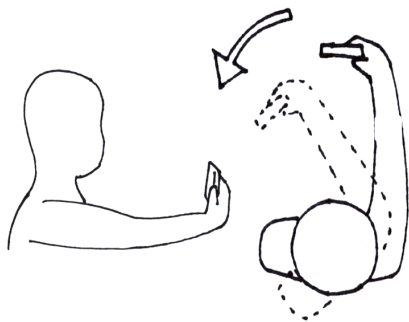


図 2 自撮り動画像の撮影方法

図 3 はこの自撮り動画像中の 1 フレームであり、背景領域と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横軸方向へ移動する。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いる。フレーム内での人物領域の割合はカメラのレンズの水平画角に依存しているが、人間の平均的な肩幅を約 40cm、腕の長さを約 60cm とした場合、レンズの水平画角が約 37 度以上のカメラを使用すると撮影者の両肩までの範囲がフレームに収まる。人物領域の形状は風や手ぶれ変動の影響

で変形することも考えられるが、数秒の撮影時間内での時間軸変動はわずかである。



図 3 背景領域と人物領域

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 4 のように背景中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 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。本研究は背景領域と人物領域の時間軸変動の差異を利用して人物領域を推定するため、空領域は樹木などが遮っている状況、背面領域は空間変化が大きい建物や植物が存在する状況、地面領域は地面上にベンチなどの物体が存在する状況での撮影が望ましい。また空領域と地面領域の占める割合を減らせれば、フレーム全体における背景領域と人物領域の時間軸変動の差異が大きくなる。

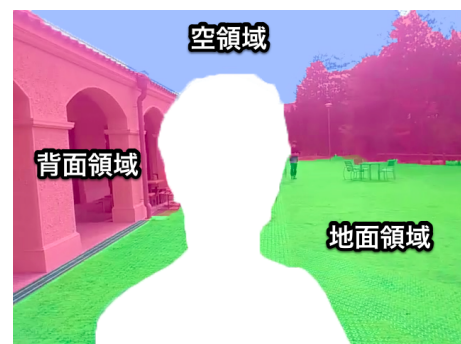


図 4 空領域と背面領域、地面領域

## 3. 人物領域の推定

背景領域と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 つの人物領域推

定手法を実装した。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画素のブロックに分割して各ブロックでの平均輝度から時間軸での標準偏差を算出し、標準偏差が小さい、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人物領域を推定する処理の流れを図5に示す。

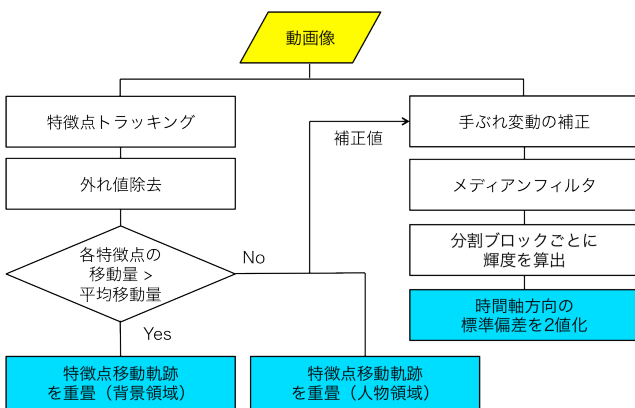


図5 人物領域推定の処理の流れ

### 3.1 特徴点トラッキングを用いた推定

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人物領域を推定する。背景領域と人物領域に写っている被写体を追跡することで被写体の動きがわかるため、自撮り動画の条件からその被写体が背景領域と人物領域どちらに属しているか判断することができるが、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。本研究では特徴点を用いた対応づけを行うことで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) [3]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、特徴点の対応付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なの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。しかし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標準偏差を求め、その距離が平均値から標準偏差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あり、移動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6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背景領域でも検出されている。移動量が小さな特徴点が背景領域で検出されるのは、特徴点の誤対応を全て除去しきれていないためである。

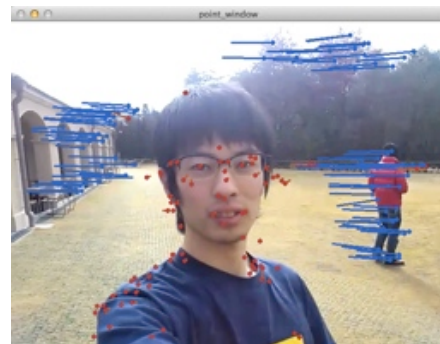


図6 特徴点移動のクラス分け結果

前記の処理を動画の全ての連続する2フレームに対して行い、各クラスの特徴点移動軌跡を全て重畳した結果が図7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が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。一方、人物領域で検出された特徴点はほぼ同じ位置で検出され続けるため、特徴点移動軌跡が輝度勾配の大きい特定の箇所集中している。また、背景領域にあたる領域で誤対応されてしまっている特徴点の移動軌跡も多く現れてしまう。しかし人物領域はフレーム内のおおむね中央にあることを前提としているため、フレームの左右両端に近い領域で検出される特徴点は誤対応の可能性が高い。移動軌跡の重畳ではフレーム中央と特徴点位置の横軸方向距離に応じて軌跡を細く重畳することで誤対応の影響を少なくしている。



図 7 背景領域(左)と人物領域(右)の特徴点移動軌跡

本来ならば背景領域にあたる領域にも人物領域クラスの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ることで人物領域のみを抽出できる。誤対応の軌跡や特徴点が検出できずに軌跡が現れなかった領域が特徴点移動軌跡にノイズとして現れているので、膨張収縮処理でノイズを除去する。ノイズ除去した背景領域の特徴点移動軌跡の否定値と人物領域の特徴点移動軌跡の論理積をとった結果が図 8 である。背景領域の移動軌跡では地面領域に特徴点移動軌跡が現れていないが、人物領域の移動軌跡にも地面領域にあたる領域に特徴点が検出されにくいため誤対応の影響を与えない。



図 8 特徴点移動軌跡を用いた推定結果

### 3.2 分割ブロックごとの時間軸変動を用いた推定

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

度から時間軸での標準偏差を算出し、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。カメラを横軸方向へ移動させている最中に縦軸方向へも僅かにカメラが上下移動しているため、各フレーム間で被写体の縦軸方向の位置ずれが発生している。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る。

被写体の移動がない領域でも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標準偏差を算出する。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標準偏差を算出する。輝度の時間軸変動を確認しやすくするために標準偏差の値を 256 階調に正規化した結果を図 9 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標準偏差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標準偏差が小さくなっている。



図 9 輝度の標準偏差

輝度の時間軸変動の標準偏差を 2 値化することで人物領域を推定する。2 値化で用いるしきい値は標準偏差の大きい背景領域と標準偏差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標準偏差も小さくなってしまいうため、大津の手法 [4] 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標準偏差が現れている。この領域の輝度変動の標準偏差値の標準偏差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標準偏差を2値化できる。2値化した結果を図10に示す。

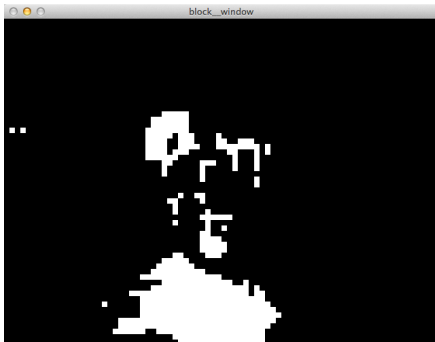


図10 2値化した推定結果

### 3.3 推定結果の統合

2つの推定手法から得られた人物領域を統合してマスク画像を作成する。特徴点トラッキングを用いた手法で得られた人物領域は輝度勾配が大きく特徴点が検出されやすい領域であるが、分割ブロックごとの時間軸変動を用いた手法で得られた人物領域は輝度勾配が小さく時間軸変動の小さい領域である。2つの推定手法ではそれぞれ人物領域内の輝度勾配の大きい領域と小さい領域をそれぞれ推定しているため、互いに欠けている人物領域を補うことができ、2つの推定結果を統合することで人物領域全体が推定できる。図8と図10の論理和を求めて推定結果を統合するが、このときフレームの両端と上端に特徴点軌跡の周辺ノイズ残っている。人物領域はフレームの下端から中央にかけて存在していることは撮影方法からして既知なため、フレームの下辺の中央を中心とした楕円状のマスク画像を用いて周辺ノイズを除去する。人物領域の推定で得られたマスク画像が図11である。

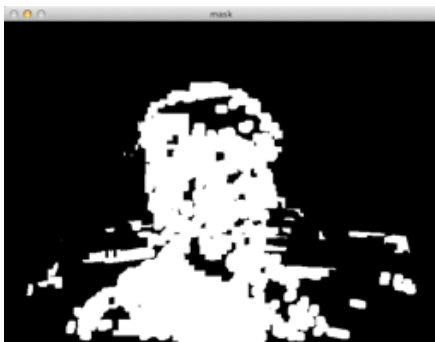


図11 マスク画像

### 4. 自撮りパノラマ写真の生成

図11のマスク画像を用いて自撮り動画の各フレーム

から背景領域のみを抽出し、全フレームの背景領域を背景の移動量に合わせて平行移動させて重ね合わせることで背景パノラマ画像を生成する。背景の移動量は特徴点トラッキングから得られたフレーム間での誤対応を覗いた特徴点の平均移動量を用いる。背景パノラマ画像に任意フレームの人物領域を重ね合わせた自撮りパノラマ写真が図12である。マスク画像の不完全さから図12の背景領域に誤って人物領域が含まれてしまっているが、輪郭に沿って正確に人物領域が抽出できていればおおむね綺麗な自撮りパノラマ写真が生成できることが示された。図12ではフレームの中央に任意フレームから抽出した人物領域を重ね合わせているが、重ね合わせる位置やスケールを変更することで違った構図の自撮りパノラマ写真を生成することも可能である。

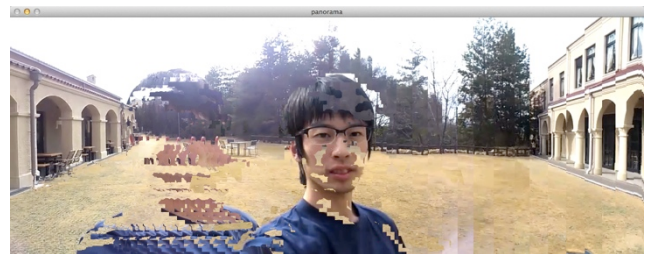


図12 自撮りパノラマ写真

### 5. 課題と展望

本研究は背景領域と人物領域の時間軸変動の差異に着目し、特徴点トラッキングと分割ブロックごとの輝度変動から人物領域を推定し、自撮り動画の各フレームを背景領域と人物領域に分離して自撮りパノラマ写真を生成した。撮影状況の異なるいくつかの自撮り動画をj用いて実験した結果、背景領域と人物領域の時間軸変動の差異を検出にくい状況で誤った推定結果が得られた。原因は主に特徴点の誤対応と、空や地面などの輝度変動が小さい領域と人物領域を分割できないことである。人物領域の推定処理中の様々なパラメータを調整することで推定精度は向上するが、理想的な撮影状況ではおおむね正しく人物領域を推定できた。

特徴点移動軌跡と輝度変動の推定結果をそのままマスク画像にj用いて人物領域を抽出しているため輪郭に沿って人物領域を抽出できていないが、おおむね正しく人物領域が推定できていれば Waterched[5]や GrabCut[6]などのアルゴリズムを利用して輪郭に沿った抽出が可能だと考えられる。また自撮り動画中の各フレームを空間微分して得られた輪郭の変動に着目することで、より人物領域の輪郭に沿った推定を行い、人物領域を抽出する必要がある。

本研究はスマートフォンのカメラで自撮りを行うことを前提としているため、提案手法をスマートフォン用にアプ

り開発し、より高精度な自撮りパノラマ写真の生成を目指したい。一般的なスマートフォンには加速度センサや地磁気センサが内蔵されているため、これらのセンサ情報からカメラの回転速度や角度が得られ、より高精度な手ぶれ変動の補正や背景パノラマ画像の生成が可能と考えられる。カメラの回転移動を正確に追うことができるとフレーム内の背景領域中の被写体の移動を追うことができるため、人物領域に遮られているかどうかの検出が可能となり、人物領域を推定することができる。また自撮り動画はスマートフォンの標準カメラアプリで撮影しているが、画面タッチで録画開始と終了の操作を行うため、操作時に手ぶれ変動を起こしてしまう問題がある。録画開始操作を行うと一定時間後に自動で撮影を開始し、カメラの回転移動が終わると自動で録画を終了する機能を組み込むことで、自撮り動画像への手ぶれ変動の影響を少なくできる。

背景領域をつなぎ合わせることで背景の広く写り込んだ自撮りパノラマ写真を生成したが、自撮りとは別に撮影した背景画像や、背景ではない別の写真と重ね合わせることで今までになかった新たな自撮り表現が可能である。以下に例を示す。



図 13 別の背景との合成例



図 14 背景以外の写真との合成例

### 別の背景との合成

人物領域を抽出するために背景領域と人物領域の時間軸変動の差異が現れやすい環境で自撮りをした場合、背景領域を繋ぎ合わせた背景パノラマ画像があまり美しくない可能性がある。そういった場合でも好きな場所やアングルで撮影した背景画像を用いることで、自撮り動画像に写り込んだ背景に依存しない写真が生成できる。

図 13 は別に撮影した背景画像との合成結果である。

### 背景以外の写真との合成

何かを紹介する自撮りを行う場合、従来は紹介したい物を手で持つなどの工夫をして撮影していたが、紹介したいものを別に撮影して人物のみの画像を重ねることで従来よりも紹介したいものを強調した自撮り表現が可能である。1 枚の写真に紹介したいものと撮影者自身の表情を同時に収めることで、感情を含めた紹介写真を生成できる。図 14 は食べ物と合成した例である。

本研究の例では撮影者 1 人だけで自撮りを行っているが、2 人以上での自撮りでも対象人物全員が含まれた人物領域を抽出できるように対応したい。実際に 1 人で自撮りをする人は少なく、多くの人が自撮りする写真は 2 人以上での記念写真である。

## 6. おわりに

本研究は、自撮り動画像から人物領域を抽出して自撮りパノラマ写真を生成する手法を提案した。自撮りする際にカメラを移動させることで発生する背景領域と人物領域の時間軸変動の差異から人物領域を推定した。人物領域を抽出できれば背景領域を繋ぎ合わせた背景パノラマ画像に重ね合わせるだけでなく、別に撮影した背景画像と重ね合わせて、また違った構図の写真も生成することも可能である。従来の写真は記録するという側面が強かったが、SNS の普及でリアルタイムに撮影した写真を共有する文化が広がり、新たに写真をアピールするという側面が加わってきた。提案手法は記録するという側面とアピールするという側面の両方において写真表現の幅を広げ、従来よりも創造性に富んだエンタテインメント性のある自撮り写真を生み出せる。

### 参考文献

- 1) 芝星帆, 入江英嗣, 吉永努, “顔検出とエッジ抽出を利用した携帯端末による自撮り支援システムの提案” WISS2012 論文集, 2012.
- 2) 松崎慧介, 目黒光彦, 古閑敏夫, “変形輪郭モデルに基づく動画像からの任意オブジェクト領域の抽出” 信学技報, SIS2006-78, pp39-44, 2007
- 3) 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.
- 4) 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.
- 5) Fernand Meyer, “Topographic distance and watershed lines” Signal Processing, Vol. 38, pp113-125, 1994.
- 6) Carsten Rother, Andrew Blake, Vladimir Kolmogorov, “GrabCut - interactive foreground extraction using iterated graph cuts” ACM Trans. on Graphics, Vol.23, No.3, pp309-314, 2004.