IIEEJ 論文 -- IoT 社会の進展を支える画像電子関連技術論文特集 --

Structure from Motionを用いた背景の3次元形状復元による移動カメラ画像列からの歩行者消去

八 木 賢太郎† 長谷川 邦 洋† 斎 藤 英 雄†(正会員)

†慶應義塾大学

Diminished Reality for Privacy Protection by Hiding Pedestrians in Motion Image Sequences Using Structure from Motion

Kentaro YAGI†, Kunihiro HASEGAWA†, Hideo SAITO†(*Member*)

†Keio University

〈あらまし〉 近年，画像中から任意の物体を消去し，まるでその物体が存在していないかのような画像を生成する技術が盛んに研究されている．こうした技術は画像中に写り込んだ人物のプライバシーの保護など様々な目的で用いられている．しかしながら既存研究には「複数カメラが必要である.」,「消去対象物体が静止している必要がある.」,「背景が平面である」といった制約があった．そこで本研究ではハンドヘルドカメラ一台のみを用いて撮影され，かつカメラ撮影者と消去対象人物の両方が移動を行うような，動的シーンに対して使用可能な人物消去技術を提案する．また，背景の3次元形状復元を行い，それを元に人物消去を行うため，任意の背景を持つシーンに対して使用可能である．実験では物体消去を行う複数の既存手法との比較を行い提案手法の優位性を示す．

キーワード：diminished reality, structure from motion, 3次元復元, multi-view stereo

<**Summary**> In recent years, techniques for diminishing objects from images are actively researched. These technologies are used for multiple applications such as privacy protection for people who are unconsciously in images. However, conventional methods have constraints such as “multiple cameras are required”, “the object to be diminished is required to be static”, “the background must be a plane.” In this research, we propose a method for diminishing people in an image sequence using a single handheld camera. We suppose that the scene is dynamic of which both photographer and people to be diminished are moving while the video is being recorded. In addition, the background can be arbitrary shape because we reconstruct 3D shape of it. In the experiment, we compare the proposed method with existing methods for different scenes which have various backgrounds.

Keywords: diminished reality, structure from motion, 3D reconstruction, multi-view stereo

1. はじめに

近年，SNSやGoogle Street Viewなどを通して誰もが容易に写真や動画の投稿，閲覧を行うことが可能である．この際，関係の無い人物が画像中に写りこみプライバシーの侵害を受ける場合がある．そこで彼らのプライバシーを保護するために画像に写り込んだ人物に対して，顔にモザイクをかけたり，別の色でマスクするといった対策が行われている．しかしながら，そうした処理により生成される画像は明らかに元画像と異なるため，動画自体の見栄えを損なう恐れがある．そこで自動的に画像中に写り込んだ人物を検出し，あたかもその人物がもとから存在していなかったかのような画像を生成する技術が求められている．

提案手法ではハンドヘルドカメラ1台で撮影された映像から，背景3次元モデルを復元することで画像中の人物消去を行う．本研究の特徴は，従来手法では実現の難しかった「背景が平面と近似できないような奥行きを持つシーン」,「カメラ撮影者も消去対象人物も移動を行うシーン」,「ハンドヘルドカメラ一台で撮影されたシーン」において自然な画像を生成できる点である．

まずStructure from Motionを用いて背景の3次元モデルを取得する．この際，人検出領域外の特徴点を用いること

で人を含まない背景のみの 3 次元点群を取得する．その後，Multi-view Stereo，点群のメッシュ化， Texturing を行うことで密な 3 次元モデルを生成する．最後に 3 次元モデルのレンダリング画像と入力画像を合成することで画像中の人物を消去する．

2. 関 連 研 究

本研究の目的は移動カメラによって撮影された映像中の歩行者を消去することである．本手法では事前に消去対象となる歩行者を前景領域として抽出しておき， Structure from Motion によって復元した背景領域の 3 次元モデルを各カメラ視点へレンダリングすることで人物消去を行う．本章では 2.1 節で前景背景の分離について，2.2 節で物体消去について既存手法との比較を行い，提案手法の位置付けを明確にする．

2.1 前景背景分離

多くの前景抽出手法は，事前に背景となるモデルを定義し，そのモデルに当てはまるか否かによって前景背景領域を決定するという考え方に基づいている．その内最も単純な方法は，時間軸方向に平均化された背景画像に対して，ピクセルごとの差分を求める手法である．例えば Bayart らが提案する AR システムでは，ユーザー視点カメラに CG を投影する際に，背景差分法によってカメラに写りこんだ物体を特定し消去している[1]．またピクセルごとの差分ではなく，背景領域をある統計モデルとして表し，モデルに当てはまらない領域を前景として検出する手法もある．例えば Zivkovic ら， Friedman らは混合ガウスモデルを用いて映像に写り混んだ物体群を複数領域に分類することで前景の抽出を行っている[?]． 玉木らは上記の方法を組み合わせ，フレーム間差分により動的な物体を抽出した後に，事前に定義された人体モデルに当てはまる領域のみを前景領域として抽出することで動画中の人物領域を抽出する手法を提案し有効性を示した[4]．しかしこれらの手法は固定カメラで撮影された映像を入力としており、我々が対象としているような移動カメラによって撮影された映像には対応していない．

自由移動カメラによる前景背景分離手法として穣田らの研究があげられる[5]．穣田らは特徴量マッチングによって生成されたパノラマ画像との背景差分を行うことによって前景領域の抽出を行う．穣田らの手法は背景が平面と近似できるようなシーンを想定しており，カメラと撮影シーンが近かったり，カメラ向きが大きく変動するシーンには対応していない．Sheikh らは背景形状に依存しない前景背景分離を行うために，特徴点の軌跡を用いた前景背景分類手法を提案した[6]．Sheikh らの手法は，複雑な形状を持つシーンでも使用可能である．しかしながら背景の 3 次元形状を復元しているわけではないため，我々の手法が目的としている前景領域消去のために，各フレームでのカメラ視点から見た背景領域画像を生成する必要がある場合には向いていない．

そこで本論文で提案する手法では移動カメラによって撮影された映像に対しても使用可能であり，かつ背景の 3 次元形状を復元可能な前景背景分離手法として，前景領域の抽出には，物体検出モデルの state-of-the-art である YOLO[7]を用い， 背景の 3 次元復元には Structure from Motion を用いる．背景の 3 次元復元を行うことで各フレームでのカメラ視点から見たレンダリング画像を生成することができ，背景形状によらない前景消去を行うことができる．

2.2 物体消去

一般的に，映像中の物体を消去する場合には異なる位置姿勢から撮影された別の画像を用いる． Enomoto らによる研究ではカメラ前方の障害物を消去ことで対象平面の透視を行う[8]． Enomoto らは AR マーカーによって対象平面を推定し，平面射影変換を施した別視点の画像を対象平面に投影することで障害物を消去する． Flores はプライバシーの保護を目的とし， Google Street View に映り込む人物の消去を行う[9]．対象シーンに移動物体が無いこと，対象人物の背景が別フレームから撮影されていることを前提としており，平面射影変換を行うことで人物を消去する． Hasegawa らは，本研究と同様に消去人物が移動しているような画像列から人物消去を行う[10]．しかし，背景が平面に近似できること，撮影者が 1 か所に留まっていることを前提としており，提案手法のように撮影者も移動するようなシーンには対応していない．Li らはインターネット上に公開されている画像に付加された Geo-Tag 情報を利用して人物消去を行う[11]．まず消去したい人物が含まれた画像の Geo-Tag 情報から大まかな撮影位置を推定する．続いて，その周辺で撮影された画像群を用いて Structure from Motion を行い，各画像のより正確な撮影位置を取得する．その後，Structure from Motio に用いた画像に対して平面射影変換を行うことで画像中の人物消去を行う．これらの研究は，どれも消去対象人物の背景が平面であると仮定し，別視点画像に対して平面射影変換を行うことで対象物体の消去を行っている．そのため背景が平面と近似できるようなシーンにしか対応できず，本研究が対象とするような背景が複雑な 3 次元形状を持つようなシーンでは正しく対象物体を消去できない．

任意の 3 次元形状を持つシーンに対しても使用可能な物体消去手法として， Honda らによる手法があげられる[12]．この手法では RGB-D カメラを用いて生成された背景 3 次元モデルを利用して対象物体の消去を行う．本手法にはユーザー視点映像を撮影する RGB カメラと，背景の 3 次元復元を行うための RGB-D カメラの 2 台のカメラが用いられる．ユーザ視点カメラで撮影された画像と隠背景撮影カメラで撮影された画像のマッチングを行うことで， 2 台のカメラの相対位置を求め，背景 3 次元モデルをレンダリングすることで対象物体の消去を行う．各フレームごとに 3 次元モデルの更新を行うため，消去対象物体が移動するような動的シーンに

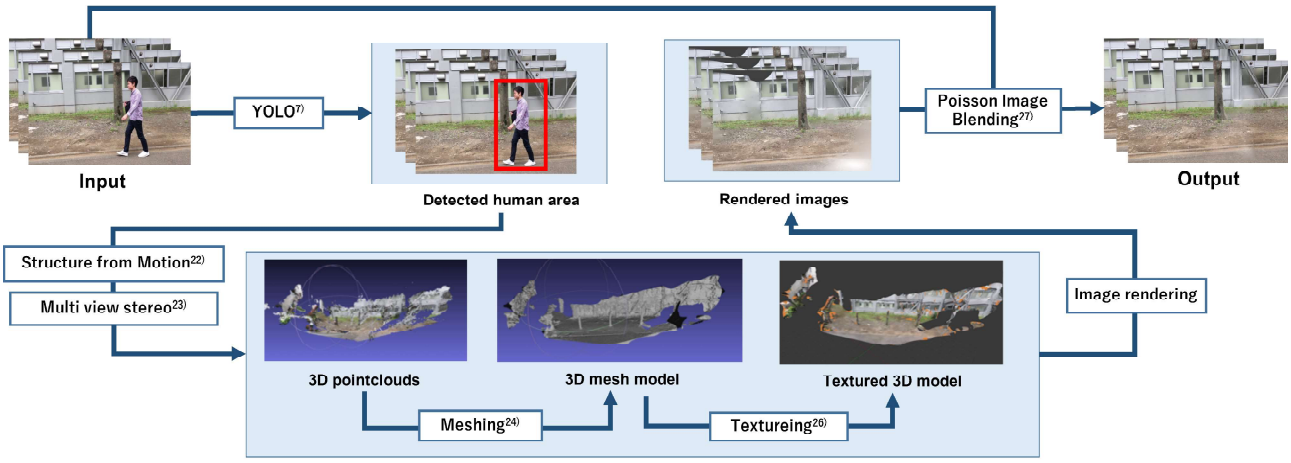

図 1　提案手法の流れ
Fig. 1　Flow of proposed method

対しても使用可能である． Kawai らによる研究では，全方位カメラによって撮影された画像列から人物消去を行う[13]．この研究では，まず Stracture from Motion と Multi-View Stereo を用いて各フレームにおける Depth マップを生成する． Depth マップを用いて背景領域の変換を行なっているため，背景の 3 次元形状を考慮したワーピングが可能である．その後エネルギー最小化を用いて人物消去を行うが，この時対象物体が持つ意味は考慮しておらず，消去対象人物以外の移動物体も全て消去されてしまう．また，消去対象人物が着ている服の色が背景と似ている場合など，背景部分と消去対象部分のエネルギー差が小さい場合は上手く消去が行えない恐れがある． Granados らはハンドヘルドカメラ 1 台で撮影された，背景が奥行きを持つような動画に対して，背景の 3 次元形状復元を行わずに人物消去を行う[14]． Granados らの手法では消去対象人物の背景を複数平面に分割し，それぞれの平面に対して他フレームからの平面射影変換行列を求める．これにより急に奥行きが変わるシーンでも物体間の幾何学的整合性が保たれた，自然な画像を生成することができる．

また他の物体消去手法として，背景可視部分のテクスチャを元に不可視部分の画素値を予測して物体消去を行う方法がある[?]．これらの手法は物体背景のテクスチャが持つ連続性を用いて物体消去を行うため，背景が連続的なテクスチャを持つ場合には有効であるが，背景が不連続である場合には不自然な背景が重畳されてしまう．また，実際の観測に基づいて不可視部分の画素を推定する訳ではないため，重畳された背景は実際の形状とは異なる．

こうした物体消去手法に対して， 提案手法は以下の 3 つの特徴を持つ．「3 次元復元を行うため背景が平面と近似できないような奥行きを持つシーンに対しても使用可能である点」，「カメラ撮影者も消去対象人物も移動を行うような動的シーンに対して使用可能である点」「ハンドヘルドカメラ一台で撮影された映像に対して使用可能である点」．提案手法ではこれまで 1 台のカメラでは実現が難しかった，奥行きのあるシーンでの人物消去を行うため，移動画像列から背景の 3 次元モデルを取得する．動的シーンにおける 3 次元形状復元手法として Tanaja らや， Mustafa らの手法が挙げられる[?]．これらの手法では，異なる視点から複数カメラで運動中の人物を撮影し 3 次元復元を行う．これに対して本手法では移動カメラ中の背景領域の 3 次元モデルを復元する必要があるため Structure from Motion と Multi-View Stereo を用いて 3 次元復元を行う． このとき事前に前景領域を抽出しておくことで前景領域を含まない背景のみの 3 次元モデルを得ることができる．前景消去の際には復元された 3 次元モデルを各カメラ視点へレンダリングすることで背景画像を生成し，背景画像と入力画像を合成することで対象人物の消去を行う．

3. 手　　法

提案手法では人領域を含まない背景 3 次元モデルを生成し，そのレンダリング画像を用いて人物消去を行う．まず初めに画像中の人領域検出を行う．その後，人検出領域外の画素を用いて Structure from Motion， Multi-view Stereo， メッシュ化， Texuturing を行うことで人を含まない背景 3 次元モデルを生成する．その後， 生成された 3 次元モデルのレンダリング画像と入力画像を合成することで，画像中の歩行者消去を行う．図 1 に提案手法の流れを示す．

3.1　背景の 3 次元モデル生成

まず，画像中の人領域検出を行う．本手法では物体検出の CNN モデルである YOLO[7]を用い，人とラベル付けされた矩形領域を人検出領域とする．また YOLO の学習には ImageNet 1000-class competition dataset[20]を用いた．全フレームにおいて人領域検出が終わったら，続いて Structure from Motion を行い疎な背景 3 次元点群を生成する．画像間のマッチングを行う際には人検出領域外の特徴点を用いる．これにより人物が 3 次元復元されるのを防ぐことができ，背景のみの 3 次元点群を得ることができる．各画像での特徴量


(a)Rendered image of pointcloud (b)Rendered image of 3D model

図 2 3 次元点群とメッシュモデルのレンダリング画像
Fig. 2 Rendered images of pointcloud and mesh model

抽出には SIFT 特徴量[21]を用い， Structure from Motion には 3 次元復元に関するオープンソースである OpenMVG を用いる[22]．続いて Multi-View Stereo を行い，より密な 3 次元点群を取得する．本手法では Barnes らが提案した手法を用いる[23]．次に，得られた 3 次元点群同士を繋ぎ，点群のメッシュ化を行う．この時， Structure from Motion, Multi-View Stereo によって得られた点群がどれほど密であるかは入力画像列の総フレーム数，各フレームで検出される特徴点数などに依存する．その為点群のメッシュ化には，入力点群の密度に依存せず，かつガラスや壁などテクスチャレスな環境においても頑健にメッシュ化が可能な Jancosek らの手法を用いる[24]．その後，得られたメッシュに対して Vu らの手法を適用することで，生成された 3 次元モデルの円滑化を行う[25]．最後に，得られた 3 次元モデルに対し Weachter らの手法を用いてテクスチャを施し，背景 3 次元モデルを得る[26]．テクスチャ生成には入力画像群を用いるが，この際消去対象人物がテクスチャに映り込まぬよう，人検出領域に含まれるピクセルはテクスチャ生成には用いないこととする．

3.2 2 次元画像への投影

続いて， 3 次元モデルから 2 次元画像へのレンダリングを行い，背景画像群を生成する．レンダリングには Structure from Motion により得られた，各フレームにおけるカメラの位置姿勢推定値を利用する．

ここで得られた点群をそのままレンダリングした場合と，メッシュ化，テクスチャ生成後の 3 次元モデルをレンダリングした画像を比較する． **図 2** からわかるように Multi-View Stereo を行ったとしても窓や壁など特徴点の少ない箇所は十分な点群が生成できず黒く穴が空いたようになってしまう．それに対して本手法のようにメッシュ化，テクスチャ生成を行うことで欠損領域の少ない背景画像を生成することができる．

3.3 入力画像と背景画像の合成

最後に，得られた背景画像群と入力画像群を合成することで画像中の人物を消去する．ここで，背景画像を最終的な出力とせず，入力画像との合成を行うのは，**図 3** に示すように，背景画像は画像中の一部に欠損領域を含む場合があるためである．ある領域が動画中で異なる視点から観測されなかったり， 3 次元復元に十分な特徴点が得られない場合，こうした欠損領域が生じる．また実際の背景と近しい画像を生成する


(a) Poisson image blending[27] (b) Rendered image of 3D model

図 3 poisson image blending[27]の出力結果とレンダリング画像
Fig. 3 Result of poisson image blending[27]and rendered image


(a) Poisson image blending[27] (b) Superimposing rendered image

図 4 背景画像の重畳結果と poisson image blending[27]の出力結果
Fig. 4 Results of superimposing and poisson image blending[27]

ことを目的としているため入力画像をベースに背景画像との合成を行い最終的な出力とする．

続いて画像合成の手順を説明する．まず同視点にレンダリングされた背景画像から人検出領域に対応する領域を切り抜く．この画像を入力画像に重畳することで人物消去を行う．しかし単純な重畳では入力画像と背景画像の境界線が目立ち，画像に不自然さが残る．そこで poisson image blending[27]を用いて画像を合成することで，より自然な出力を得る．**図 4** に背景画像を人検出領域へ重畳した結果と，poisson image blending[27]による出力結果を示す．

poisson image blending[27]を行う際，入力画像と背景画像を入力とすると，消去対象人物も出力結果に含まれてしまう．そのため poisson image blending[27]の入力は，背景画像重畳後の入力画像，背景画像，人検出領域のマスク画像とする．このときマスク画像は人検出矩形の縦幅，横幅を 10 % 広くした領域とする．こうすることで，合成を行う際に境界部分が滑らかになり，より自然な画像が生成される．

4. 評　　価

本章では 2 つの実験を行った．実験 1 では異なる背景を持つ複数シーンに対して提案手法，既存手法を適用し，定性的に出力結果の比較を行なった．実験 2 では提案手法，既存手法に対して，出力結果と背景画像の近似度を計測することで，どの手法が最も実際の背景と近しい画像を生成できるか評価した．

比較を行う既存手法として，Granados らの手法[14]のように，消去対象人物の背景部分のテクスチャを，背景が撮影されている他フレームに対して平面射影変換を行うことで補完

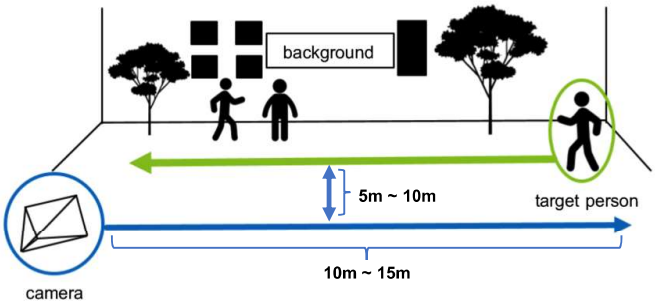

図 5　実験環境
Fig. 5　Experimental environment

する手法 (平面射影変換法)，そして Deep Learing を用いた画像補完手法の state-of-the-art である Iizuka らの手法[17]を選択した．

平面射影変換法については，人検出領域背景領域を複数平面に分割し，それぞれの平面に対して他フレームからのホモグラフィー行列を求めることにより実装した．この手法の具体的なアルゴリズムを次に示す．まず，消去対象人物が含まれるフレーム t に対して，人検出領域の背景を手動で複数平面 $P = p_1, p_2, \ldots, p_n$ に分割する．それぞれの平面 p_k に対して，フレーム t とできるだけ近く，かつ p_k が観測されているようなフレーム t' を手動で選択する．t 中の p_k が含まれる平面と，t' 中の p_k が含まれる平面間でのホモグラフィー行列 H を求める．H によって射影変換した背景画像とフレーム t を poisson image blending[27]を用いて合成する．画像間のマッチングを行う際には AKAZE 特徴量を用い，また H の推定には RANSAC を用いた．poisson image blending[27]を行う際には，提案手法と同様のパラメータを用い，背景画像重畳後の入力画像，背景画像，対象平面 p のマスク画像を入力とした．

動画の撮影には Apple 社の iphone6s を用いた．解像度は 1,920×1,080 でありキャプチャ速度は 30fps である．また，図 5 に示すようにカメラ撮影者，消去対象人物の両方が移動を行うシーンで撮影を行なった．その際，歩行者とカメラ撮影者は最も近づいた時で 5m-10m 程の距離を保ち，カメラ撮影者は歩行者の進行方向と逆方向に 15m ほど歩行しながら撮影を行った．また背景 3 次元モデルから 2 次元画像へのレンダリングには Blender 2.78b を使用した．

4.1　実験 1

実験 1 では異なる背景を持つ 8 つのシーンに対して人物消去を行いその出力結果を定性的に評価した．図 6 に各手法による出力結果を示す．左から順に入力画像列中のある 1 フレーム，同フレームに対する提案手法の適用結果，平面射影変換法の適用結果，Iizuka らの手法[17]の適用結果である．

4.1.1　実験結果

提案手法では背景の 3 次元モデルを生成しているため背景が平面に近いシーン，奥行きを持つシーンの両方で整合性を保ったまま人物の消去を行うことができた．scene 1，scene 4，scene 5，scene 6 に示すように消去対象人物の背景に奥行きを持つ物体が存在していても，背景にズレが生じていないことがわかる．それに対して平面射影変換法では図中の赤枠で示す箇所において不自然な結果となっている．scene 1 では木の位置が背景とズレてしまっている．これは元画像中において人物の背景に存在する木が隠れてしまっているために，木の平面に対するホモグラフィー推定がうまくいかなかったためである．同様に scene 6 でも人の背景に完全に隠れてしまったポールに対してはホモグラフィーを求めることができず歪んだ画像となっている．また scene 5 では背景の滑り台にズレが生じている．これは滑り台が十分なテクスチャを持たないために，ホモグラフィー推定が上手く行かなかったためである．その一方で scene 2，scene 3，scene 7，scene 8 のように消去対象人物の背景において急な奥行きの変化が生じていないシーンでは平面射影変換法でも自然な画像が生成できることがわかる．

Iizuka らの手法[17]では，複数シーンにおいて，消去対象人物の背後に存在する物体が復元されなかった．scene 1，scene 4，scene 6，scene 7 における結果を見てみると，本来存在するはずの木やポールが復元されていないことがわかる．これは Iizuka らの手法[17]では実際の観測に基づいて不可視領域を補完するのではなく，他領域の画素値を元に領域内部の画素値を求めているためである．このように Iizuka らの手法[17]をはじめとする inpainting 手法では消去対象人物の背景に完全に隠れてしまった物体は復元できないという問題がある．また，隠背景領域が大きくなったり，テクスチャが複雑になるほど背景の復元は難しくなるため，人検出領域が比較的大きい scene 1，scene 8 では背景が歪んだような画像が生成された．これに対して，本手法では背景の 3 次元モデルを用いて人物消去を行なっているため，人物の背景に隠れた物体も復元が可能である．また，消去対象領域の大きさや背景のテクスチャに依存せずに自然な人物消去が可能である．

また，提案手法は scene 8 に示すように消去対象人物が複数人存在するようなシーンであっても使用可能である．人物消去ができるかどうかは背景の 3 次元復元ができるかどうかに依存しているため，複数人の消去対象人物がいても人物背景が複数フレームによって撮影できていれば，消去対象人物の数には関係なく本手法を適用できる．

scene 6 では背景とズレることなく人物消去できているが，背景と消去対象人物の背景に存在する支柱が短く投影されてしまっている．これは 3 次元点群に対するメッシュ化，メッシュの円滑化の過程でポールの先端が削られてしまったためである．3.2 節で示したように，メッシュ化を行うことで点群が得られないような箇所に対しても補完ができるが，このシーンに含まれるポールのように急激に形状が変わるような箇所はノイズとして除去されてしまう場合もある．そのため，急に形状が変化するような箇所に対しても使用できるような

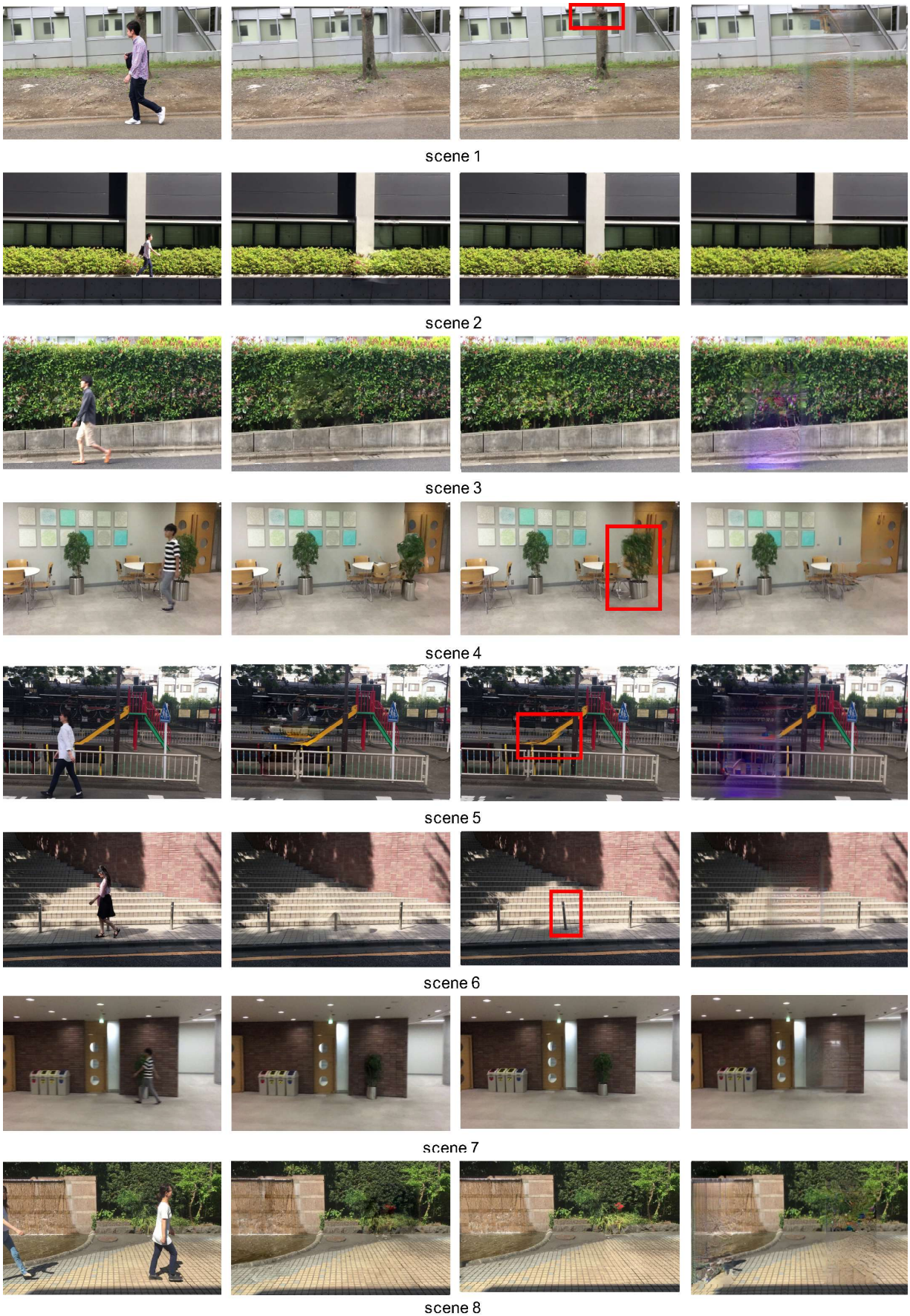

図 6　提案手法の適用結果
Fig. 6　Results of proposed method

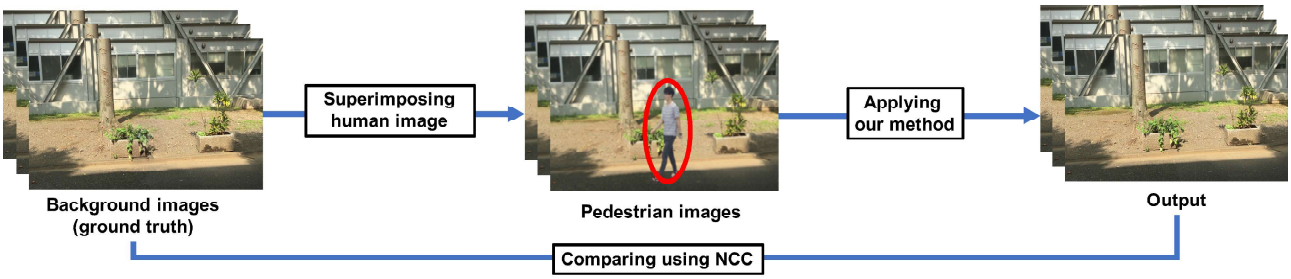

図 7　実験 2 の流れ
Fig. 7　Flow of experiment 2

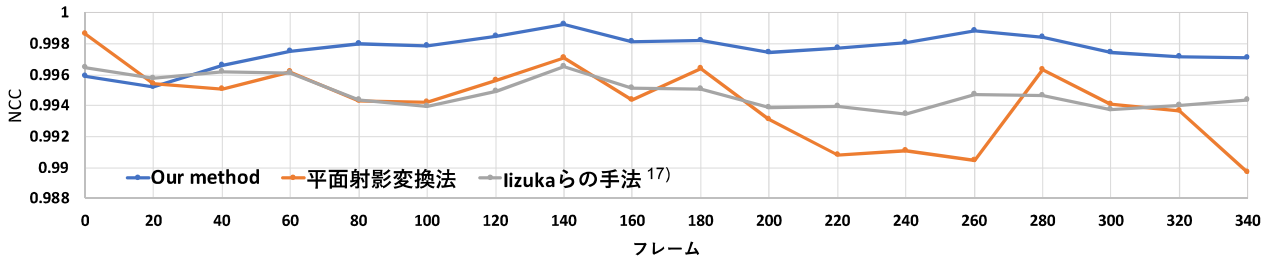

図 8　各手法の NCC 値
Fig. 8　Value of NCC of each method

メッシュ化の方法を検討することが今後の課題である．

4.2　実験 2

実験 2 では各手法の出力結果と，人の写っていない背景画像を比較することで，どれだけ実際の背景と近しい画像を得られるかを定量的に評価した．まず ground truth となる背景画像を取得するために歩行者がいないシーンで動画の撮影を行った．その後，動画中の各フレームに人物画像を重畳することで擬似的に歩行者が写り込んだ動画を生成した．この動画を入力として人物消去を行い，その出力結果と人物画像頂上前の映像を比較した．図 7 に実験 2 の流れを示す．

4.2.1　実験方法

まず歩行者が写っていないシーンで撮影を行う．この時，図 5 に示すシーンを再現するためにカメラ撮影者は 15m 程歩行しながら撮影を行なった．またここで得られた画像に人物画像を重畳する際には，カメラ撮影者の進行方向と逆方向に歩行者が移動するシーンを再現するために，人物重畳領域を動画の全フレーム数 x，画像幅 w ピクセルに対して，w/x ピクセルずつカメラ撮影者の移動方向と逆向きに変化させた．また出力結果と背景画像の比較では 2 画像間の類似度を示す値である NCC(Normalized Cross-Correlation) を指標とした．NCC の値が 1 に近いほど，2 枚の画像は類似しているといえる．

4.2.2　実験結果

各手法の出力結果と背景画像の NCC を図 8 に示す．本実験では全 341 フレームの動画に対して 20 フレームごとに出力結果を取得し背景画像との比較を行った．図 8 が示すように，多くのフレームにおいて提案手法による出力結果が最も NCC 値が大きく，実際の背景と類似度が高い画像を復元できた．提案手法の NCC が既存手法を大きく上回ったシーンでは消去対象人物の背景に植木や木などの物体が存在していた．4.1.1 項に示したように，背景の 3 次元復元を行わない手法では消去対象人物の背景に完全に隠れた物体やテクスチャの少ない物体は復元することが難しい．それに対して，提案手法のように背景の 3 次元モデルを取得する方法では消去対象人物の背景に完全に隠れる物体であっても，その物体の形状に関わらず復元することが可能である．逆に消去対象人物の背景に奥行きのある物体が存在していないシーンでは，平面射影変換法の方が提案手法に比べて NCC の値が大きくなった (0，20，40，60 フレーム目)．

図 9　1 フレーム目における ground truth と提案手法の出力結果
Fig. 9　Ground truth and result of proposed method of the first frame

また，動画開始直後のフレームでは提案手法に比べて既存手法の方が良い結果となった．これは図 9 に示すように，その領域の背景 3 次元モデルが生成できなかったことが原因である．Structure from Motion において，正確な 3 次元形状を復元するためには，その特徴点が多視点から撮影されている必要がある．しかし動画開始直後，動画終了間際でしか写っていないような背景に関しては，複数フレームからの撮

影が行えないために，３次元点が正確に復元できない場合がある．そうした点群はメッシュ化を行う際に外れ値として除去されてしまうため，部分的に３次元モデルが生成できなかったと考えられる．

5. むすび

ハンドヘルドカメラによって撮影された動画を対象として，画像中に写り込んだ人物を消去した映像を生成する手法を提案した．従来手法は「複数カメラが必要である」,「消去対象物体が静止している必要がある」,「背景が平面と仮定できる」といった制約を持ち，本研究が対象とするようなハンドヘルドカメラ１台のみを用いて撮影され，かつカメラ撮影者と消去対象人物の両方が移動を行い，任意の背景を持つ動的シーンに対しての使用は難しかった．それに対して提案手法では，Structure from Motion などを用いて背景の３次元モデルを生成し，それを用いて人物消去を行うことで，今まで対応できなかった条件下での人物消去を可能とした．実験では既存の人物消去手法である，平面射影変換を用いた人物消去手法と Deep Learing を用いた画像補完手法との比較を行い提案手法の優位性を示した．

参考文献

1) B. Bayart, J. Y. Didier, A. Kheddar: “Force Feedback Virtual Painting on Real Objects: A Paradigm of Augmented Reality Haptics”, Proc. of International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, Vol.5024, pp.776–785 (2008).
2) Z. Zivkovic: “Improved Adaptive Gaussian Mixture Model for Background Subtraction”, Proc. of IEEE International Conference on Pattern Recognition, Vol.2, pp.28–31 (2014).
3) N. Friedman, S. Russell: “Image Segmentation in Video Sequences: A Probabilistic Approach”, Proc. of Uncertainty in Artificial Intelligence, Vol.13, pp.175–181 (1997).
4) 玉木徹, 山村毅, 大西昇: “画像系列からの人物領域の抽出”, 電気学会論文誌, Vol.119, No.1, pp.137–43 (1999).
5) 譲田 賢治, 坪内 貴之, 菅谷 保之, 金谷 健一: “移動ビデオカメラ画像からの運動物体の抽出”, Computer Vision and Image Media, Vol.2004, No.26, pp.41–48 (2004).
6) Y. Sheikh, O. Javed, T. Kanade: “Background Subtraction for Freely Moving Cameras”, Proc. of IEEE International Conference on Computer Vision, Vol.12, pp.1219–1225 (2009).
7) J. Redmon, S. Divvala, R. Girshick, A. Farhadi: “You Only Look Once: Unified, Real-time Object Detection”, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.779–788 (2016).
8) A. Enomoto, H. Saito: “Diminished Reality Using Multiple Handheld Camera”, Proc. of Asian Conference on Computer Vision Workshop, Vol.7, pp.130–135 (2007).
9) A. Flores, S. Belongie: “Removing Pedestrians from Google Street View Images”, Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp.53–58 (2010).
10) K. Hasegawa, H. Saito: “Synthesis of A Stroboscopic Image from A Hand-held Camera Sequence for A Sports Analysis”, Proc. of International Symposium on Mixed and Augmented Reality Workshop, Vol.2, No.3, pp 277–289 (2016).
11) Z. Li, Y. Wang, J. Guo, L.-F. Cheong, S.-Z. Zhou: “Diminished Reality Using Appearance and 3D Geometry of Internet Photo Collections”, Proc. of IEEE International Symposium on Mixed and Augmented Reality, pp.11–19 (2013).
12) T. Honda, H. Saito: “Real Time Diminished Reality Based on 3d Measurement of Environment Using An Rgbd Camera”, Trans. on Virtual Reality Society of Japan, Vol.19, No.2, pp.105–116 (2014).
13) N. Kawai, N. Inoue, T. Sato, F. Okura, Y. Nakashima, N. Yokoya: “Background Estimation for A Single Omnidirectional Image Sequence Captured with A Moving Camera”, IPSJ Trans. on Computer Vision and Application, Vol.6, pp68–72 (2014).
14) M. Granados, K. I. Kim, J. Tompkin, J. Kautz, C. Theobalt: “Background Inpainting for Videos with Dynamic Objects and A Free-moving Camera”, Proc. of European Conference on Computer Vision, Vol.7572, pp682–695 (2012).
15) J. Herling, W. Broll: “Pixmix: A real-time approach to high-quality diminished reality”, Proc. of IEEE International Symposium on Mixed and Augmented Reality, pp.141–150 (2012).
16) N. Kawai, T. Sato, N. Yokoya: “Towards Iinternetscale Multi-view Stereo”, IEEE Trans. on Visualization and Computer Graphics, Vol.22, No.3, pp.1236–1247 (2016).
17) S.Iizuka, E.Simo-Serra, H. Ishikawa, “Globally and Locally Consistent Image Completion”, ACM Trans. on Graphics, Vol.36, No.4, pp.107:1–107:14 (2017).
18) A. Taneja, L. Ballan, M. Pollefeys. “Modeling Dynamic Scenes Recorded with Freely Moving Cameras”, Proc. of Asian Conference on Computer Vision, Vol.3, pp.613–626 (2010).
19) A. Mustafa, H. Kim, J.Y. Guillemaut, A. Hilton: “General Dynamic Scene Reconstruction from Multiple View Video”, Proc. of IEEE International Conference on Computer Vision, pp.900–908 (2015).
20) O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei: “ImageNet Large Scale Visual Recognition Challenge”, International Journal of Computer Vision, Vol.115, No.3, pp211–252 (2015).
21) D.G. Lowe: “Distinctive Image Features from Scale-invariant Keypoints”, International Journal of Computer Vision, Vol.60, No.2, pp91–110 (2004).
22) P. Moulon, P. Monasse, R. Marlet: “Adaptive Structure from Motion with A Contrario Model Estimation”, Proc. of Asian Conference on Computer Vision, Vol.7727, pp257–270 (2012).
23) C. Barnes, E. Shechtman, A. Finkelstein, D.B. Goldman: “PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing”, ACM Trans. on Graphics. Vol.28, No.3, pp24:1–24:10 (2009).
24) M. Jancosek, T. Pajdla: “Exploiting Visibility Information in Surface Reconstruction to Preserve Weakly Supported Surfaces”, International Scholarly Research Notices, Vol.2014, pp1–20 (2014).
25) H.-H. Vu, P. Labatut, J.P. Pons, R. Keriven: “High Accuracy and Visibility-consistent Dense Multiview Stereo”, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.34, No.5, pp889–901 (2012).
26) M. Waechter, N. Moehrle, M. Goesele: “Let There Be Color! Large-scale Texturing of 3d Reconstructions”, Proc. of European Conference on Computer Vision, pp.836–850 (2014).
27) P. P’erez, M. Gangnet, A. Blake: “Poisson Image Editing”, ACM Trans. on Graphics, Vol.22, pp.313–318 (2003).

(2018 年 5 月 31 日 受付)
(2018 年 12 月 14 日 再受付)

八　木　賢太郎
2017 年 慶應義塾大学理工学部情報工学科卒業．2019 年 現在，慶應義塾大学理工学研究科前期博士課程在学し，コンピュータビジョンやその応用に関する研究に従事．

長谷川　邦　洋
2007 年 慶應義塾大学理工学部情報工学科卒業．2009 年 同大大学院理工学研究科前期博士課程修了．同年キヤノン (株) 入社．2016 年 慶應義塾大学大学院理工学研究科後期博士課程修了．現在，同大大学院理工学研究科特任助教．コンピュータビジョンやその応用に関する研究に従事．博士 (工学)．

斎　藤　英　雄（正会員）
1987 年 慶應義塾大学理工学部電気工学科卒業．1992 年 同大学院理工学研究科電気工学専攻博士課程修了．その後, 同大学助手，専任講師，助教授を経て 2006 年 より教授．博士 (工学)．この間，1997 年 から 99 年 までカーネギーメロン大学ロボティクス研究所訪問研究員．コンピュータビジョンとその応用に関する研究に従事．