

動画補完のための動的背景の除去法

谷田川 達也^{†1} 山口 泰^{†1}

画像補完の応用として研究が進められてきた動画補完は、数年前まで補完領域に適合する類似領域の探索に多くの時間が必要だった。近年では、動画内の動的領域と静的領域をあらかじめ分離しておくことにより、探索を効率化する研究がいくつか紹介され、計算時間は飛躍的に減少した。一方で、いずれの手法も動画内の動的領域が前景、静的領域が背景と明確に区別できる動画のみを対象としており、波打ち際や樹木のそよぎに代表されるような動的背景を含む動画の高速な補完は困難であった。本研究では、これらの動的背景を除去し、補完に重要な前景のみを抽出することで、高速な動画補完法の多様な動画への適応を実現する。

Dynamic background removal for video inpainting

TATSUYA YATAGAWA^{†1} and YASUSHI YAMAGUCHI^{†1}

Previously, many video inpainting techniques were proposed and many of the techniques were based on image inpainting approaches. These methods have computational difficulty. To search most similar fragments which can be acceptable for missing region is time consuming process because of large data size. In these years, some approaches divide videos into dynamic and static region in advance, and they realized efficient similar fragment search. Since these methods handle only videos that can be divided dynamic and static regions easily, it is quite difficult to handle videos with dynamic background region. In this paper, we describe a method to remove unnecessary dynamic background region and to extract only dynamic foreground region, and our method achieves fast video inpainting for various kinds of video scenes.

1. はじめに

近年、YouTubeなどの動画投稿サイトが一般的に普及したことで、これまで、撮影して保存しておくだけにとどまっていた動画が、一般的な人々によって編集される機会が増加しつつある。従来より、静止画像編集処理の一手法として画像中から不要な物体を除去するという技術は多く研究されてきた³⁾。これらの手法は多くのソフトウェアで機能が実装されて

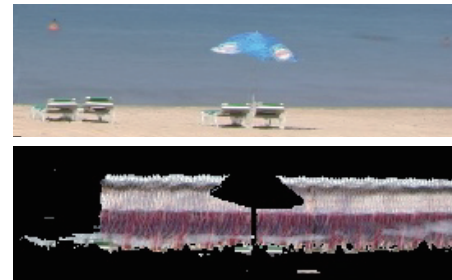


図 1 (上) 背景モザイク。(下) 前景モザイク。
Fig. 1 (top) Background mosaic. (bottom) Foreground mosaic.

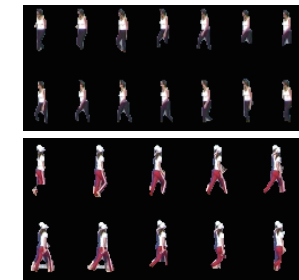


図 2 移動物体テンプレートのデータセット。
Fig. 2 Sample of object templates.

おり、動画像においても静止画像同様に数多くの不要物体除去技術が提案されている^{4),7)-9)}。

動画補完の基本的な考え方は被補完領域に適合する類似領域を動画中から探索することで、被補完領域を埋めていくというものである。その一手法として、動画の自然さを保つために、動画像の時間的な連続性を考慮する手法が提案された⁹⁾。この手法では動画像全体から被補完領域に適切な類似領域を探索するため、頑健な補完結果が得られる一方、多くの計算時間を要するという問題がある。

これを対して、入力動画像において前景と背景をあらかじめ分離し、前景モザイク、背景モザイク(図1)と呼ばれる要約画像を作ることで補完処理の効率化を行う手法が提案された⁷⁾。この手法はモザイク画像上で動画像中のどの位置に類似領域が存在していそうかをあらかじめ探索することで、劇的に探索時間を短縮することに成功した。ただし、モザイク画像は言い換えれば、動画の入力フレームの平均画像であり、入力フレームが多く、補完を行いたい移動物体が、別の移動物体と重なりあうなどした場合には補完の精度が低下するという問題がある。

一方、分離した前景を移動物体テンプレート(object template, 図2)と呼ばれる短冊状の小画像にして利用することで、動的計画法による高速で頑健な動画補完を実現する手法も提案されている⁸⁾。移動物体テンプレートは人の歩行などの周期運動を補完するという目的において非常に優れており、たとえ全てのフレームで正しく前景が抽出されていなくても、ある程度良い前景補完結果を得られる。これは、周期運動を再現するためには最低限1周期分が正確に前景抽出されていれば良いためである。

この他にも動画像上で前景として抽出される領域がフローを持つ動的な領域であることに注目し、その領域を物体追跡することで探索の効率化をはかる手法なども紹介されている⁴⁾。

ただし、いずれの手法も前景と背景の分離の際、オプティカルフローなどの動きを持つ領

^{†1} 東京大学大学院 総合文化研究科 広域科学専攻

The Graduate School of Arts and Cultures, The University of Tokyo

域が前景であるという仮定をおいており、波打ち際や樹木のそよぎなど、背景でありながら、動的な領域を直接扱うことは考慮されていない。また、これらの手法において前景と背景を適切に分離することは補完結果の頑健性だけでなく、計算速度にも影響するため、補完処理にとって、より重要性の高い前景だけを抽出する手法が求められている。

2. 関連研究

前景の分離は主に動画セグメンテーションや背景差分法などの処理によって行われる。動画セグメンテーションは画像セグメンテーションと同様に、ユーザの入力により指定された前景および背景の情報を用いて行われる。まず、ピクセルの色や位置の情報をガウス混合モデルによってモデル化する。その情報を元に得られるギブズエネルギーをグラフィカットなどの手法により最小化することで、セグメンテーションの結果を得る⁶⁾。

背景差分法は、推定背景からの差分情報を用いて前景を抽出する手法が一般的である。この場合の推定背景とは入力動画の各フレームで空間的に同位置なピクセルの平均をとることで作られる平均画像である。入力動画が、移動カメラによるもの場合には画素の平均をとる前にオプティカルフローなどの情報を用い、位置合わせを行う。また、Zivkovicの手法のように動画のフレーム上で同位置に存在するピクセル群の情報をガウス混合モデルにあてはめることで、リアルタイムに背景差分を行う方法も紹介されている¹⁰⁾。

LiuとGleicherは、動画セグメンテーションにキーフレームという考え方を紹介した。この手法は、対象物体の一部のみしか動かない動画のセグメンテーションを目的としている。キーフレームとは、動いている対象物体の一部が持つ情報を正確に取得しうるフレームを指す。対象物体の一部だけにしか動きがない場合、1枚のフレームから対象物体全体の情報を得ることはできない。そこで、入力動画全体から対象物体の各パーツの情報を持つキーフレームを抽出し、その情報を統合することで対象物体全体のセグメンテーションを行う⁵⁾。

3. 提案手法

提案手法では動画を、前景、動的背景、静的背景(静的領域と同義)の3種類の領域から構成されていると考える。従来の背景差分法では前景と背景のみの区別であったが、我々は従来手法における前景をさらに前景と動的背景に分けて考えている。図3に提案手法を用いた動画補完の流れを示す。

入力された動画は、まず背景差分法により動的領域と静的領域に分けられる。この時、動的領域には前景領域と、動的背景領域とが含まれている。前景領域とは、動画における注視物体を指し、入力動画の多くのフレームに写りこんでいるような移動物体に対応する。一方で、動的背景とは、波打ち際や樹木のそよぎなどのように、あるフレームを見たときには動いているとみなすことができるが、別のフレームでは動いていない、あるいは存在していな

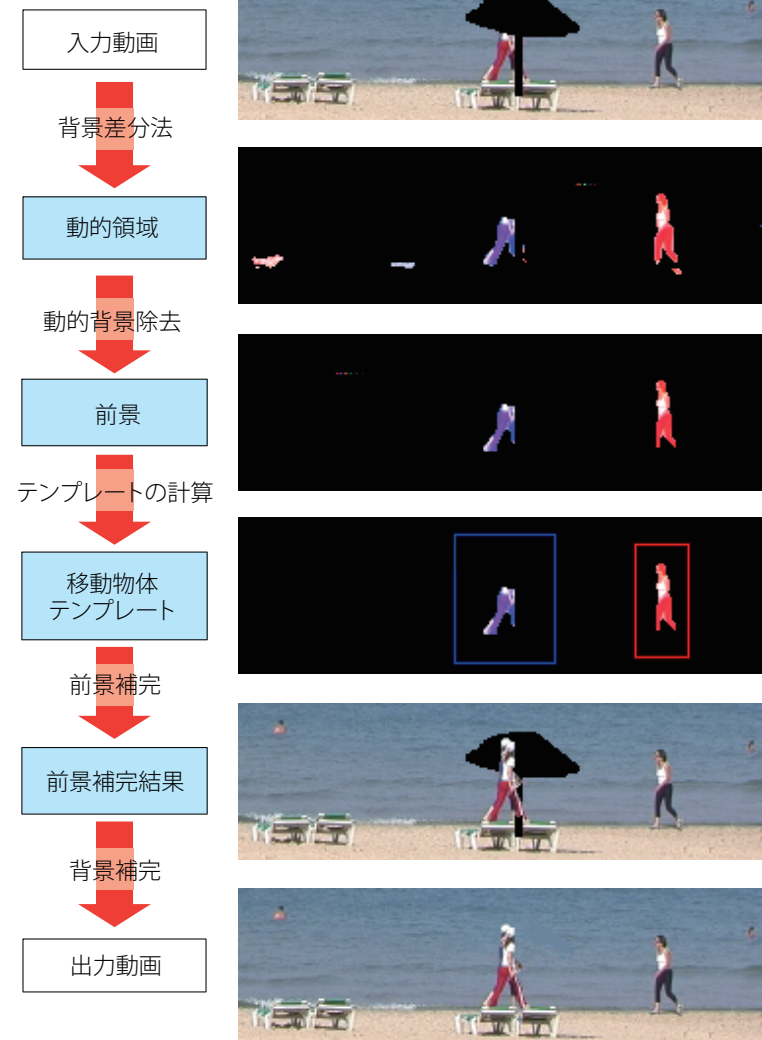


図3 提案手法を用いた動画補完の処理手順。それぞれの画像は34フレーム目に対する処理結果を示したものである。
Fig.3 Flow chart of video inpainting algorithm using our method. Each image shows the results of the 34th frame.



図 4 (a) 入力フレーム . (b) 推定背景画像 .
Fig. 4 (a) Input frame. (b) Estimated background image.

いような領域である .

この点に注目してフレームごとに動的領域を見ていくと、動的背景領域のほとんどないフレームが見つかる . このようなフレームはほぼ前景の情報だけを含むと考えることができる . 我々はこのようなフレームをキーフレームと呼ぶ . キーフレームを用いることの利点は、背景差分法の結果に多く見られる動的背景領域をユーザによる入力なしに比較的安定して除去できるところにある .

提案手法では、前景の補完に Venkatesh らの手法⁸⁾を用いており、この手法により前景の補完を行うためには、移動物体テンプレートの計算が必要であるが、この計算においてもキーフレームの考え方は有効である . 動的背景領域を含まないキーフレームにおいては前景物体の位置を容易に特定できるため、キーフレームをもとに移動物体追跡を行うと移動物体テンプレートを求めることができる . 動的背景は、分離された後、静的背景と合わせられ背景として扱われる . 背景の補完は Patwardhan らの手法⁷⁾を用いて補完が行われる . 前景および背景の補完手法の詳細については 3.5 節および 3.6 節で述べる .

3.1 背景差分法による動的領域と静的領域の分離

動的領域を分離するために、まず入力動画から推定背景を求める . 入力動画に対しオプティカルフローを計算し、各フレームで Mean-Shift 法により、フローの分布中心を計算する . この分布中心が各フレームのカメラモーションに対応する . カメラモーションにより位置合わせを行った全フレームにおいて、空間的に同位置となるピクセルを時間方向に平均すると推定背景画像が得られる . 入力フレームと推定背景画像の例を図 4 に示す .

3.2 キーフレームの定義と抽出

キーフレーム抽出の計算は背景差分法により分離された動的領域に対して行われる . 入力動画を I としたとき、動的領域を I^{dyn} 、静的領域を I^{sta} とすると、 $I = I^{dyn} \cup I^{sta}$ (ただし $I^{dyn} \cap I^{sta} = \phi$) である . Liu と Gleicher の定義に従い、入力動画 I における i 番目のフレーム I_i がキーフレームであるかの判定は、次の式によって行われる .

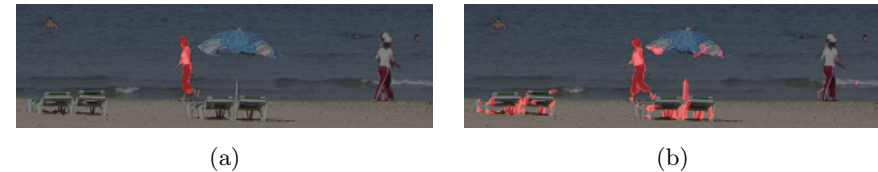


図 5 (a) キーフレーム . (b) 非キーフレーム . それぞれのフレームで抽出された動的領域に色付けしてある .
Fig. 5 (a) Keyframe. (b) Non-keyframe. Each frames are colorized at dynamic regions.

$$Size(I_i^{dyn}) > minArea \quad (1)$$

$$Var((x, y) | (x, y) \in I_i^{dyn}) < maxSpan \quad (2)$$

式 1, 2 において、 $Size(\cdot)$ は集合に含まれるピクセルの総数を返し、 $Var(\cdot)$ は入力されたピクセル位置から求められる共分散行列の行列式を返す . $minArea$ および $maxSpan$ はそれぞれ、キーフレームにおける動的領域の最小面積と最大のばらつき度を表す . 抽出されるキーフレームの数は $minArea$ および $maxSpan$ の値によって変わってくるが、我々の実験では、入力フレーム数に対して、10–15% のキーフレームが抽出されるように $maxSpan$ の値をなるべく小さく調整するとより良い結果が得られた . この定義にしたがって抽出されたキーフレームと、非キーフレームとを図 5 に示す .

3.3 キーフレームにおける前景のモデル化

キーフレームとして抽出されたフレームにおいて、前景領域に含まれるピクセルが持つ情報をガウス混合モデルにあてはめる . キーフレーム上の動的領域において、 $L * a * b$ 色情報、フローのベクトル、ピクセルの位置をモデル化する . ガウス混合モデルは $k = 1, 2, \dots, K$ という K 個のガウス分布 $g_k(\mathbf{x} | \mathbf{m}_k, S_k)$, $\mathbf{x} \in \mathbb{R}^d$ によって表現される .

$$g_k(\mathbf{x} | \mathbf{m}_k, S_k) = \frac{1}{(\sqrt{2\pi})^d \sqrt{|S_k|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_k)^T S_k^{-1}(\mathbf{x} - \mathbf{m}_k)\right)$$

それぞれの情報に対するガウス混合モデルの計算は EM アルゴリズム (Expectation Minimization Algorithm) によって行われる . また、ガウス混合モデルを構成するガウス分布はそれぞれに重み ω_i を持っており、空間内の座標 \mathbf{x} に対するモデル上の確率 $\Pr(\mathbf{x})$ は次のような重みづけ平均として計算される (一般的には $\sum_k \omega_k = 1$ に正規化する) .

$$\Pr(\mathbf{x}) = \sum_{k=1}^K \omega_k g_k(\mathbf{x} | \mathbf{m}_k, S_k)$$

この確率を用いると、キーフレームに隣接したフレームで、動的背景の除去を行うことが可能である . 隣接したフレームにおいて以下の仮定を置く .

仮定 1 前景物体には、光源の色変化などによる色調変化が起こらない。

仮定 2 前景物体はほぼ等速で移動しており、途中で大きな速度の変化がない。

仮定 1 を置くことにより、キーフレーム上で色情報のモデルが隣接した非キーフレームにおいても類似した分布を持つと期待できる。同様に、仮定 2 を置くことで、隣接した非キーフレームでもキーフレームと類似したフローのモデルを持つと期待できる。また、仮定 2 を用いることで、非キーフレームにおいて、前景物体がどの位置にあるのかを推測することが可能である。以上の仮定を用いて、非キーフレームの前景モデルをキーフレームのモデルから逐次推定できる。この非キーフレーム上における前景モデル推定を前景モデルの伝播と呼び、次節 3.4 で詳しく述べる。

3.4 前景モデルの伝播

確率モデルの伝播は非キーフレーム上の前景モデルの推定、当該モデルを用いたラベルの更新の 2 段階に分けられる。ここで言うラベルとはフレーム上のピクセルが前景、動的背景のどちらに属しているかを表すものである。今、キーフレームを含む伝播処理済みのフレーム \hat{I}_i から未処理のフレーム I_{i+1} への伝播を考える (処理済みのフレームにはハットをつけて表現してある)。確率モデルの推定は 3.3 節の二つの仮定に基づき行われる。仮定 1 に基づき、色およびフローの情報は隣接した \hat{I}_i の情報をそのまま用いることが可能である。一方、前景物体の位置の情報は、前景物体が移動しているために、 \hat{I}_i の情報をそのまま使うのでは精度が悪い。そこで、 \hat{I}_i において動的領域 (すなわち前景) になっているピクセルが持つフローの分布中心を Mean-Shift 法により計算する。計算されたフローの分布中心は \hat{I}_i, I_{i+1} の両フレーム間での前景物体の移動を表していると考えられる。このフローの分だけ \hat{I}_i 上で計算された確率モデルを平行移動し、 I_{i+1} に適応する。これらの処理によって、 I_{i+1} 上のモデルの推定が行われる。

次に未処理のフレーム I_{i+1} におけるラベルの更新、すなわち動的背景の除去を行う。 I_{i+1} 上の動的領域をグラフ $\mathcal{G}_{i+1} = (\mathcal{V}_{i+1}, \mathcal{E}_{i+1})$ で表すものとする。ここで、 \mathcal{V} は動的領域に含まれるピクセルの集合、 \mathcal{E} は \mathcal{V} 上の隣接したピクセルの組からなる集合である。各ピクセル p のラベル $L_p \in \{“FG”, “dynBG”\}$ を求める問題は、次のギブズエネルギーを最小化する問題と置き換えられる。

$$E(I_{i+1}) = \sum_{p \in \mathcal{V}_{i+1}} f_{data}(\mathbf{m}c_p, \mathbf{c}_p, \mathbf{s}_p) + \sum_{(p,q) \in \mathcal{E}_{i+1}} f_{smooth}(\mathbf{c}_p, \mathbf{c}_q) \quad (3)$$

式 (3) において $\mathbf{c}_p, \mathbf{m}c_p, \mathbf{s}_p$ はそれぞれのピクセル p における色 (color cue)、フロー (motion cue)、位置 (locality cue) の情報を表している。式 (3) で f_{data} はデータ項と呼ばれ、各々のピクセルが、モデルに即した情報を持っているかどうかを対数確率の和として表している。一方、 f_{smooth} は平滑化項と呼ばれ、隣接した二つのピクセルが類似した色情報を持っているかを色空間における距離に応じて数値化している。それぞれの項はさらに次の



図 6 計算された移動物体テンプレート。

Fig. 6 Calculated object templates.

ように書ける。

$$f_{data}(\mathbf{m}c_p, \mathbf{c}_p, \mathbf{s}_p) = -(\ln \text{Pr}_{mc}(\mathbf{m}c_p) + \ln \text{Pr}_c(\mathbf{c}_p) + \ln \text{Pr}_s(\mathbf{s}_p))$$

$$f_{smooth}(\mathbf{c}_p, \mathbf{c}_q) = \exp\left(-\frac{\|\mathbf{c}_p - \mathbf{c}_q\|^2}{2\sigma^2}\right)$$

式 (3) で与えられたエネルギーはグラフカット²⁾ によって最小化することが可能である。グラフカットはこのような 2 種類のラベルづけ (今回の例では前景と動的背景) を行うような問題に対し、高速に最適解を与える手法であり、今回の例でも最適なラベルの更新を行うことができる。

ラベルを更新された I_{i+1} から前景のラベルがついたピクセルだけを残し、この I_{i+1} を新たに処理済みフレーム \hat{I}_{i+1} とする。すなわち $\hat{I}_{i+1} \equiv \{p \in I_{i+1} | L_p = “FG”\}$ とする。この操作を I_{i+2}, I_{i+3}, \dots と続けていくと、最終的に全フレームから動的背景を除去できる。

3.5 前景の補完

提案手法では前景の補完に Venkatesh らの手法を用いる。この手法では前処理として移動物体テンプレートの計算を行い、そのテンプレートを用いることで補完が進行する。

3.5.1 移動物体テンプレートの計算

移動物体テンプレートは前景の物体を含む保守的な矩形領域を指す (図 6)。各テンプレートはキーフレームを用いて計算される。テンプレートの大きさは、キーフレーム上の前景領域にあるピクセルの分布に従って決定される。さらにキーフレームは前景物体以外のノイズをほとんど含まないと考えられるため、前景物体の物体中心を自動的に決定することができる。すなわち、ピクセルの分布中心を物体の中心と考えることにする。中でも確率モデルの伝播を行う前の状態で検出されたキーフレームは最も信頼性の高い情報を持っているため、この情報を元に、移動物体テンプレートのサイズと位置を決定する。その他のフレーム上での物体位置は確率モデルの伝播と同じく、位置情報を Mean-Shift 法により求められたフローの分布中心の値だけ平行移動することで求められる。

3.5.2 移動物体テンプレートを用いた前景の補完

前景の補完は移動物体テンプレート同士の相違度を比較することで、被補完テンプレートにふさわしい補完用テンプレートを探索する処理である。二つの移動物体テンプレートを

O_i, O_j とし、それぞれのテンプレートは前景に含まれたピクセルを 1, 非前景のピクセルを 0 とするようなマスク M_i, M_j を持つと考える．二つのテンプレートを重ね合わせる位置調整ベクトル (alignment vector) を \mathbf{a} とすると、相違度 $d(O_i, O_j)$ は次の式で表される．

$$d(O_i, O_j) = \min_{\mathbf{a}} (E_O(O_i, O_j; \mathbf{a}) + E_N(O_i, O_j; \mathbf{a})) \quad (4)$$

式 (4) において、右辺はさらに 2 種類の相違度によって表されている． E_O は二つのテンプレートを重ねた際に、前景のピクセル同士が重なったところで輝度の差分を取り、それを合計したものである． E_N は一方のテンプレートで前景となっているピクセルにもう一方のテンプレートでは非前景のピクセルが重なっていた時、そのピクセルの数だけ相違度を上げるというものである．ピクセル $p \in O_i$ の位置を表す位置ベクトルを \mathbf{p} とすると、 E_O および E_N はそれぞれ次のように表される．

$$E_O(O_i, O_j; \mathbf{a}) = \sum_{p \in O_i} [O_i(\mathbf{p}) - O_j(\mathbf{p} + \mathbf{a})]^2 M_i(\mathbf{p}) M_j(\mathbf{p} + \mathbf{a})$$

$$E_N(O_i, O_j; \mathbf{a}) = 255^2 \sum_{p \in O_i} M_i(\mathbf{p}) [1 - M_j(\mathbf{p} + \mathbf{a})]$$

被補完テンプレートは前景の情報が残っているか否かで補完方法を変更する必要がある．前景情報が含まれているものを部分損失 (partially occluded)、全く含まれていないものを完全損失 (completely occluded) と呼ぶ．また、以後は被補完テンプレートにチルダをつけて \tilde{O}_i と表わし、これと最も類似すると考えられるテンプレートを O_{t_i} と表す．

部分損失テンプレートには前景情報が少なからず残っているため、式 (4) を連続した w フレームに適用することで、もっとも相違度の小さいテンプレートを見つけることが可能である．ここで w は探索テンプレート数であり、テンプレート内の移動物体に対し、時間方向の連続性を何フレーム分保証するかを決定する数である．ここで w 枚の被補完テンプレート列を $\tilde{G}_i = \tilde{O}_i, \tilde{O}_{i+1}, \dots, \tilde{O}_{i+w-1}$ 、これと類似するテンプレート列を $G_{t_i} = O_{t_i}, O_{t_i+1}, \dots, O_{t_i+w-1}$ とすると、 \tilde{G}_i に対する G_{t_i} の相違度は次の式で表される．

$$d_w(\tilde{G}_i, G_{t_i}) = \sum_j^{w-1} d(\tilde{O}_{i+j}, O_{t_i+j}) \quad (5)$$

一方で、完全損失テンプレートでは、前景情報が被補完領域により完全に失われているため、前後、数フレームの情報だけでは、適切な補完用テンプレートを見つけることはできない．そこで連続した h 枚の完全損失テンプレートにおける累積相違度 C_{t_i} を次のように再帰的に定義する．

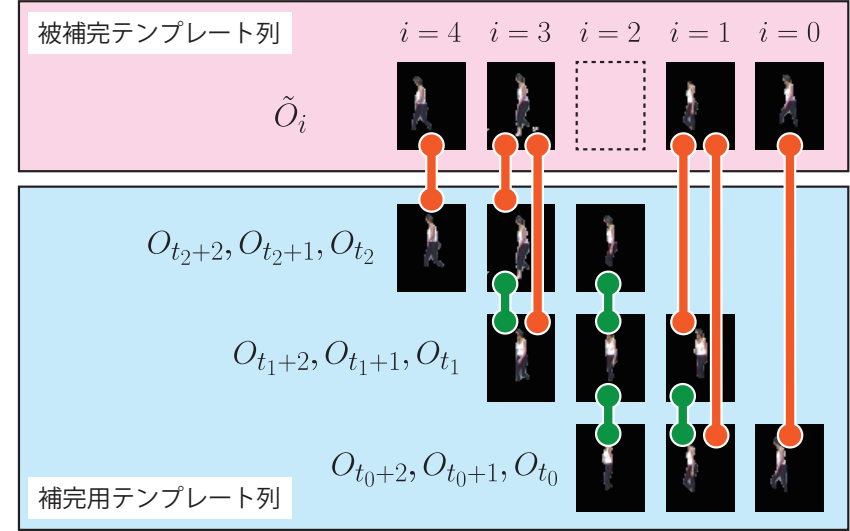


図 7 完全損失テンプレート補完の様式図．赤色のリンクは式 (6) の第 1 項にある相違度に対応し、緑色のリンクは式 (6) の第 2 項にある相違度に対応する．

Fig.7 Illustration of foreground inpainting for completely occluded object. each of links colored red or green indicates the desimilarity in the first term or the second term in Eq. (6).

$$C_{t_i} = \sum_{j=0}^{w-1} d(\tilde{O}_{i+j}, O_{t_i+j}) \cdot 1_{\tilde{O}_{i+j}} + \min_{t_{i-1}} \left(C_{t_{i-1}} + 1_{\{i>0\}} \cdot \sum_{j=0}^{w-2} d(O_{t_i+j}, O_{t_{i-1}+j+1}) \right) \quad (6)$$

完全損失テンプレートの補完は完全損失が起こる前後 $w-1$ フレームの情報を用いて行われるため、式 (6) において入力テンプレート列は $\tilde{O}_i, (i = 0, \dots, h+2w-3)$ となる．式 6 において、第 1 項は無損失あるいは部分損失のテンプレートと類似テンプレートとの相違度を表す項であり、 $1_{\tilde{O}_{i+j}}$ とはそれを意味する指示関数 (indicator function) である．すなわち \tilde{O}_{i+j} が無損失あるいは部分損失であるときに 1 を完全損失である時に 0 をとる．第 2 項は類似テンプレート同士の相違度を表す項であり、選ばれた補完用テンプレート同士の連続性を保証している．式中の $1_{\{\cdot\}}$ はある条件 $\{\cdot\}$ を満たす時に 1 を、満たさない時に 0 をとる指示関数である．この累積相違度を $i = h+2w-3$ について計算した際に得られた類似テンプレート列 $O_{t_0}, O_{t_1}, \dots, O_{t_{h+2w-3}}$ を連続した完全損失テンプレートに対する補完用テンプレートとする．式 (6) を図によって表すと図 7 のようになり、図中の赤色、緑色のリンクはそれぞれ式 6 における第 1 項、第 2 項に対応している．部分損失、完全損失ともに被補完

テンプレート \tilde{O}_i に対する補完は、補完用テンプレート $O_{t_i - \lfloor w/2 \rfloor + \lfloor w/2 \rfloor}$ を用いて行われる。これに加え、被補完テンプレートに対して、どの位置に補完用テンプレートを配置するかを何らかの方法で決定しなくてはならない。部分損失テンプレートの中心位置 \hat{z}_i は補完用テンプレートの位置 z_{t_i} に対し、式 (5) で求められた位置調整ベクトル \mathbf{a}_{t_i} を用いて次のように表わされる。

$$\hat{z}_i = z_{t_i} + \mathbf{a}_{t_i}$$

一方、完全損失テンプレートには前景情報がないため、式 (6) でも被補完テンプレートに対する相違度や位置調整ベクトルは計算されていない。そのため、補完用テンプレート間での位置調整ベクトルを \mathbf{a}_{t_i} を次のように積算することで、テンプレートの位置を決定する。

$$\hat{z}_i = z_{t_i} + \sum_{j=w-1}^i \mathbf{a}_{t_j}$$

ここでは、位置調整ベクトル \mathbf{a}_{t_i} を、以下のように求める。

$$\mathbf{a}_{t_i} = \arg \min_{\mathbf{a}} (E_O(O_{t_i+1}, O_{t_i+1}; \mathbf{a}) + E_N(O_{t_i+1}, O_{t_i+1}; \mathbf{a}))$$

ただし、この積算を行うと完全損失テンプレート列内で誤差が蓄積してくため、次のようにして、テンプレートごとの平均誤差 e を求めて、位置の補正を行う。

$$e = \frac{1}{h} (z_{h+w-1} - \hat{z}_{h+w-1})$$

$$\hat{\hat{z}}_i = \hat{z}_i + e \quad (7)$$

完全損失テンプレートの最終的な位置は式 (7) の $\hat{\hat{z}}_i$ を用いる。

3.6 背景の補完

背景の補完には Patwardhan らの動的領域補完法を用いる。Patwardhan らの手法における背景補完は近接したフレームから、画像補完の要領で類似断片を探索することにより行っている。だが、提案手法では背景にも動的な領域があることを仮定しているため、この背景補完法では時間方向の動きの連続性が不自然になることが考えられる。そこで提案手法では背景においても、Patwardhan らの動的領域補完法を用い補完を行う。

Patwardhan らの動的領域補完法は 2 種類のモザイク画像を用いて行われ、それぞれ色情報、フロー情報の平均をとったものである。色情報の平均によって作られるモザイク画像は背景差分法において計算された推定背景と同じものである。フロー情報を平均して得られるフローモザイクは Wexler らの手法によって示された簡易オプティカルフローを用いて計算される⁹⁾。簡易オプティカルフローとは x 方向、 y 方向のそれぞれで $I_t/I_x, I_t/I_y$ と表される。なおここで、 I_t, I_x, I_y は各方向に対する輝度値の偏微分を表している。

提案手法における背景の補完は非補完領域の境界部分から順に行われる。 i 番目のフレームにおいて、被補完領域との境界部分にあるピクセル p を中心とした $b \times b \times b$ の領域を被補完断片 $\Psi_{p,i}$ (target fragment) と呼ぶ。ここで b は探索断片のサイズを表す定数である。背景補完の処理はこの $\Psi_{p,i}$ に対し、次の順序で進行する。

- (1) 背景差分法を行う段階で計算したカメラモーションを用いて、動画上でのピクセル位置 p をモザイク上でのピクセル位置 p_{mos} に変換する。
- (2) モザイク上のピクセル p_{mos} を中心とした 2 次元の断片 $\Psi_{p_{mos}}^{(2)}$ と最も類似した 2 次元候補断片 $\Psi_{p_{mos}}^{cand(2)}$ を探索する。
- (3) モザイク上での 2 次元候補断片の中心位置 p_{mos}^{cand} とカメラモーションから、 j 番目のフレームにおいてのピクセル位置 p_j^{cand} を求める。
- (4) 全てのフレーム $j = 1, 2, \dots, n$ において、被補完断片 Ψ_p と候補断片 $\Psi_{p_j^{cand}, j}$ との相違度を計算し、もっとも相違度の低い候補断片を求める。

モザイク上、動画上のどちらにおいても、相違度の計算は 5 次元ベクトルに対する SSD (Sum of Squared Difference) を用いて行われる。ここで 5 次元ベクトルとは RGB の色情報 3 次元と簡易オプティカルフローの情報 2 次元を指す。こうして求められた最も相違度の低い候補断片を類似断片 (source fragment) と呼ぶ。補完はこのような類似断片の探索と、被補完領域に対する情報のコピーの繰り返しにより行われる。

4. 実験結果

本節では提案手法による動的背景除去の結果および、それを用いた動画補完の結果を示す。なお提案手法の特徴を示すため、結果を示した動画は全て背景に動的な領域を含んでいる。

4.1 動的背景除去の結果

図 8 に提案手法の結果および一般的な背景差分法の結果の比較を示す。図に示された結果は左から順に 20 フレーム目、53 フレーム目、86 フレーム目の結果である。この動画においては背景に上から下への移動する白波が存在しているほか、中央にあるビーチパラソルが風になびいて動いている。閾値処理を用いた背景差分法、および Zivkovic の手法による結果では白波およびビーチパラソルの一部が前景として抽出されていることが分かる。一方、提案手法ではこれらの動的背景が除去され、歩行を行う女性のみがほぼ正確に抽出されている。

ただし、図 8 において黄色の矩形に囲まれた部分を拡大すると、提案手法において、女性の足元が正しく抽出できていないことが分かる。Venkatesh らは、このように前景の抽出に一部失敗していたとしても、多くの場合は正確な補完結果が得られると述べており、我々の行った実験においても、ほとんどの場合で正確な補完結果が得られた。しかしながら、正しく前景抽出できていないテンプレートを用いて補完を行う際、欠けのある箇所が被補完領域と重なれば、不完全な前景がそのまま補完結果に反映されてしまう。図 9 において赤色で示



図 8 前景抽出結果の比較 . (a) 入力フレーム . (b) 推定背景を用いた閾値処理による背景差分法の結果 . (c) Zivkovic の手法による結果 . (d) 提案手法による結果 . 黄色で示した箇所を拡大すると , 提案手法では白波の動的背景が除去できている一方 , 女性の足の部分に欠けが見られる .

Fig. 8 Result comparison of extracted foreground. (a) Input frame. (b) Background subtraction by thresholding using estimated background. (c) Result of Zivkovic. (d) Our result. In our result, white waves are correctly removed in the yellow rectangle. On the other hand, absence at the lady's foot can be seen.

表 1 補完に用いた動画と計算時間 .

Table 1 Example for video inpainting and computational time.

図の番号	フレームサイズ	フレーム数	計算時間
10	271 × 80	98	< 10 min
11	170 × 80	88	< 6 min

した部分を拡大すると , 提案手法で女性の顔および左腕の部分に不自然な欠けが見られる .

4.2 動画補完の結果

図 10 では , 提案手法による前景抽出結果を用いた動画補完の結果と Wexler らの結果との比較を示す . 図に示された結果は左から , 34 フレーム目 , 65 フレーム目 , 72 フレーム目の結果である . 提案手法はいずれのフレームも , Wexler らの手法による結果に劣らない結果になっていることが確認できる . また , 提案手法は C++ で実装し , 2.86GHz の Core i7 で 1 コアのみを用いて計算を行ったが , 271 × 80 のフレーム 98 枚により構成される動画において 10 分以下で計算が終了した . 表 1 に , 実験を行った動画とその計算時間を示す .

5. まとめおよび今後の課題

本報告では , 動画補完のための動的背景分離手法について提案した . 提案手法は , これまで建造物など静止している背景を含む動画像にしか適応が難しかった前景抽出に基づく高速な動画補完法を , 波打ち際などの動的な背景領域を持つ動画にも適応できることを示した .



図 9 結果画像の拡大図 . (上) Wexler らの手法による結果 . (下) 提案手法による結果 . 提案手法による結果では , 女性の顔および左腕の部分に不自然な欠けがあらわれている .

Fig. 9 Zoomed result image. (top) Result of Wexler et al. (bottom) Our result. Our result contains unnatural missing part around the lady's face and left arm.



図 10 動画補完結果の比較 . (a) 入力フレーム . (b) Wexler らの手法による結果 . (c) 提案手法による結果 . 入力動画サイズ , 縦 80 px × 横 271 px × 98 フレーム . 補完領域 , 4220 px × 98 フレーム .

Fig. 10 Result comparison of video inpainting. (a) Input frames. (b) Result of Wexler et al. (c) Our result. Input video size is 80 px × 271 px × 98 frames. Missing region is 4220 px × 98 frames.

また , これまで動的な背景を持つ動画にも適応可能だった Wexler らの手法に対し , 提案手法では数十倍の速度で同等の結果が得られるものと思われる .

現在 , 既知の問題点として , 図 11 に示したような動的背景領域の補完が失敗する事例が確認されている . 背景の補完において , 動的背景と静的背景を区別せずに補完を行う際に , 平均化したモザイク画像を用いると , 動的な領域がどこに存在しているかという情報が消されてしまうことが問題であると考えられる . 今後は , 動的背景と静的背景を区別しながら , 背景補完を行う手法あるいは , 動的静的にかかわらず , 背景領域を包括的に扱うことのできる補完法について提案を行っていきたい .

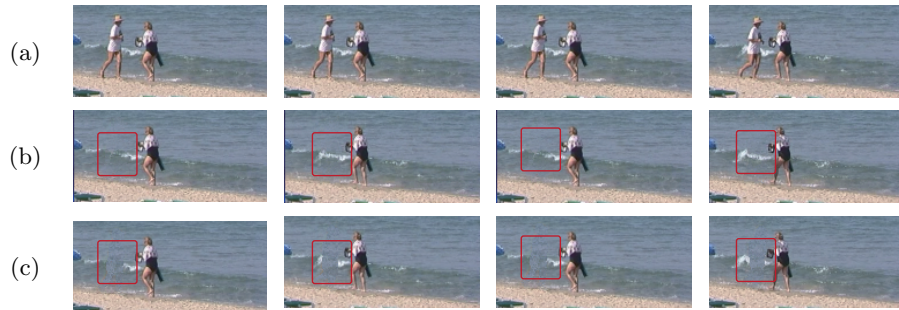


図 11 動的背景部分の補完失敗例 . (a) 入力フレーム . (b)(Wexler 2007) による結果 . (c) 提案手法による結果 . 提案手法による補完では女性の左側にある白波が途切れてしまっている .

Fig. 11 Failed example of dynamic background inpainting. (c) Input frames. (b) Result of Wexler et al. (c) Our result. White wave at the left of a lady has broken up in our result.

付 録

A.1 複数の移動物体を含む際の処理

複数の移動物体を含む場合、提案手法では k-means++法¹⁾を用いたクラスタ分割により移動物体の同定を行っている . このクラスタ分割は動的領域として抽出された動画上のピクセル $p^{(3)} \in I^{dyn}$ が持つオプティカルフローの情報に対し適応される .

k-means++法は k-means 法の抱える、初期のクラスタ中心がクラスタ分割結果に強く影響するという問題点を解決した手法である . k-means 法は一般的に、与えられたサンプル集合から k 個のサンプルを初期中心としてランダムに与えるところから処理を開始するが図 12 (b) のように誤った初期中心を与えると不自然なクラスタ分割結果が得られてしまう .

そこで k-means++法では最初の初期中心のみをランダムに与え、以後の初期中心は入力サンプルとすでに確定した初期中心との距離を計算し、いずれの初期中心からも遠い入力サンプルを新たな初期中心とする . こうすることで、よりクラスタ間の距離が離れたクラスタ分割結果を得ることが可能である . また、初期中心を選び終わった時点で、それらが互いに離れて分布しているため、クラスタ中心が収束する過程で、クラスタに属するサンプルの入れ替わりが起こりづらくなる . 結果としてより早くクラスタ中心が収束する .

ただし、k-means++法を用いたとしても移動物体をフローにより同定する方法には限界がある . 移動物体が極端に多い場合や、移動物体同士の動きが似ている場合には、分離が非常に困難となる . この場合には、あらかじめユーザ入力などにより、移動物体ごとの対応関係を付けておくなどの処理が必要になる .

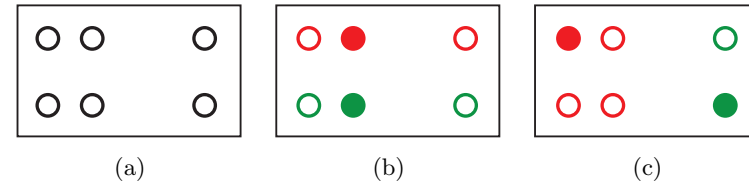


図 12 k-means 法によるクラスタリング結果の例 . (a) 入力サンプル . (b) 初期中心を誤って選択した場合の結果 , (c) 初期中心を正しく選択した場合の結果 . (b), (c) において塗りつぶされている円は選ばれた初期中心である .

Fig. 12 Examples of k-means clustering result. (a) Input samples. (b) Result from wrong initial centroids. (c) Result from correct initial centroids. Filled circles in (b) and (c) are initial centroids.

参 考 文 献

- 1) Arthur, D. and Vassilvitskii, S.: k-means++: the advantages of careful seeding, *Proc. of the 18th annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07*, pp.1027–1035 (2007).
- 2) Boykov, Y. and Jolly, M.-P.: Interactive graph cuts for optimal boundary amp; region segmentation of objects in N-D images, *Proc. of the 8th IEEE International Conf. on Computer Vision*, Vol.1, pp.105–112 (2001).
- 3) Drori, I., Cohen-Or, D. and Yeshurun, H.: Fragment-based image completion, *ACM Trans. on Graphics*, Vol.22, pp.303–312 (2003).
- 4) Jia, Y.-T., Hu, S.-M. and Martin, R.R.: Video completion using tracking and fragment merging, *The Visual Computer*, Vol.21, pp.601–610 (2005).
- 5) Liu, F. and Gleicher, M.: Learning color and locality cues for moving object detection and segmentation, *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp.320–327 (2009).
- 6) Nagahashi, T., Fujiyoshi, H. and Kanade, T.: Video Segmentation Using Iterated Graph Cuts Based on Spatio-temporal Volumes, *ACCV 2009*, Vol.5995, pp.655–666 (2009).
- 7) Patwardhan, K., Sapiro, G. and Bertalmio, M.: Video Inpainting Under Constrained Camera Motion, *IEEE Trans. on Image Processing*, Vol.16, No.2, pp.545–553 (2007).
- 8) Venkatesh, M.V., ching SamsonCheung, S. and Zhao, J.: Efficient object-based video inpainting, *Pattern Recognition Letters*, Vol.30, No.2, pp.168 – 179 (2009).
- 9) Wexler, Y., Shechtman, E. and Irani, M.: Space-Time Completion of Video, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.29, No.3, pp.463–476 (2007).
- 10) Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction, *Proc. of the 17th International Conf. on Pattern Recognition*, Vol.2, pp.28–31 (2004).