

I-31 移動体除去による背景画像の生成に関する研究

Study on Automatic Generation of Background Image Removed Moving Object

谷口寿俊¹・田中成典²・古田均²・西田義人³・杉町敏之⁴・加藤雄大⁵Hisatoshi TANIGUCHI, Shigenori TANAKA, Hitoshi FURUTA,
Yoshito NISHITA, Toshiyuki SUGIMACHI and Katsuhiko KATO

抄録：建設用 CAD や GIS の景観テクスチャ、ITS の移動体認識に利用される背景画像は、不要な移動体の写っていないものが望ましい。しかし、人通りや交通量が多い、またはカメラを固定し続けることができない環境で背景のみの画像を取得するには、特別な機器や準備が必要であり容易ではない。そこで、本研究では、画像処理を用いて、フリーハンドで撮影したデジタル動画画像から、撮影条件に影響されることなく移動体を除去した背景画像を自動生成する手法を提案する。本提案手法では、動画画像の各フレームに特徴点を生成し、フレーム間の対応点を取得して位置あわせを行う。そして、位置合わせしたフレーム列から移動体を除去した背景画像を生成する。

Abstract: Regarding a background image to be used for landscape texture of CAD for construction and GIS, or recognition of mobile bodies in ITS, it is desirable to have no unnecessary mobile bodies in the image. However, in order to obtain an image with only the background in such an environment that streets are crowded or with heavy traffic, or where it is impossible to keep the camera fixed, special equipment and preparation are required, which are not easy to provide. For this reason, using the image processing, this study generates a background image from a digital moving image shot by free hand by removing mobile bodies in it without being affected by swing, rotation, or change in scale. First, we generate characteristic points of each frame from a moving image, tracking them to obtain corresponding points among frames. Then, using the affine coordinate transformation with the corresponding points, we perform alignment of each frame. Lastly, we generate a background image by removing the mobile bodies from the sequence of frames kept in proper alignment.

キーワード： デジタル動画画像、画像処理、背景画像、移動体除去

Keywords : Digital video sequence, Image processing, Background image, Moving object removal

1. はじめに

近年、背景画像は建設分野において多くのシステムに利用されている。たとえば、建築や都市開発の計画段階における合意形成や意思決定支援を目的として、3次元 CAD や GIS による景観シミュレーションが利用されており、その核となる構造物は3次元のデータによって表現されるのが一般的であるが、構造物を除く街頭風景等の景観にはデータサイズの小さい2次元の背景画像をテクスチャとして利用することが多い。また、建設工学的な側面では、ITS の画像処理による移動体認識に背景画像を利用するなど、その用途は多岐に渡る。前述した用途に利用される背景画像には、WebGIS の利用普及などで画像が多数の目にさらされる可能性があることや、移動体が画像処理の精度低下

の要因となり得ることから、歩行者や車両などの不要な移動体が画像シーン中に写り込んでいないことが望ましい。そのため、背景画像を撮影し、利用する際は、移動体の侵入に対する配慮が必要となる。しかし、撮影場所が歩行者や交通量の多い場所であった場合、歩行者や車両によるシーン中への侵入を防ぐことは困難である。不要な移動体が写り込んでいない背景画像を取得するには、長時間もしくは早朝や夜間といった人通りの少ない時間帯に撮影を行わなければならない、取得できるシーンが制限される。さらに、撮影場所の環境次第ではカメラを固定して撮影を続けることが困難であるなど、不要な移動体の写り込んでいない背景画像の取得は特別な機器や準備が必要になる。背景画像の撮影における現状として、以下の2つの問題点が挙げられる。

-
- 1 : 学生会員 情修 関西大学大学院 総合情報学研究科
(〒569-1095 大阪府高槻市霊仙寺町二丁目一番一号, Tel :072-690-2154, E-mail : hisarock@fc5.so-net.ne.jp)
 - 2 : 正会員 工博 関西大学 教授 総合情報学部
 - 3 : 非会員 情博 関西大学大学院 総合情報学研究科(現, 三菱電機株式会社先端技術総合研究所)
 - 4 : 正会員 情博 関西大学大学院 総合情報学研究科(現, 神戸大学大学院機械工学科)
 - 5 : 非会員 関西大学大学院 総合情報学研究科

- 時間的制限
 歩行人や車両が途切れるまで待たなければならず、撮影時間が長引く。長期化を回避するために撮影時間帯を早朝や夜間に限定しなくてはならない。
- 環境的制限
 時間的制限における問題を回避するためには、カメラを固定して撮影できる環境が求められる。

移動体に限らず、画像中から不要な領域を除去した画像の取得については古くから多くの研究^{1)~5)}が行われている。代表的な手法には、移動体が写り込む状況下で撮影した動画画像からロバスト推定^{6), 7)}により背景画像を取得する手法^{1), 2)}、移動するカメラに対してEPI (Epipolar Plane Image)⁸⁾を使用して、時空間画像処理により背景の傾きを算出して障害物領域を補完する手法^{3), 4)}や、車載カメラを用いて同一箇所について複数回時間を空けて撮影し、ベクトルメディアンフィルタ (VMF: Vector Median Filter)⁹⁾を用いて移動体を除去する手法⁵⁾が提案されている。ロバスト推定^{6), 7)}を用いた背景画像の生成手法^{1), 2)}では、入力画像列から背景モデルの自動生成と逐次更新を行う手法を提案しており、得られた背景モデルを用いて移動体の検出と追跡を行うシステムを実装している。しかし、これらの研究は、どちらも固定したカメラによる撮影を想定しているため、環境的制限を解決できていない。また、EPI⁸⁾を用いた時空間画像解析による手法^{3), 4)}では、カメラが移動する環境において、背景の傾きを算出して障害物領域を除去した画像を生成しているが、EPIを作成するには等速直線運動でカメラを移動させながら撮影する必要があり、フリーハンドでの撮影は困難である。さらに、VMF⁹⁾を用いて移動体を除去する手法⁵⁾では、時間を空けて複数回撮影する必要があるため、時間的制限を解決できていない。

このように、これまで様々な背景画像を生成する手

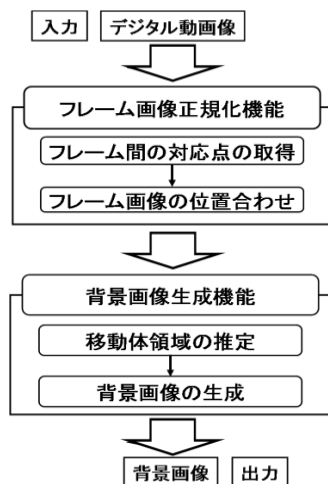


図-1 処理の流れ

法が提案されているが、本研究で述べている背景画像の取得時における時間的制限と環境的制限の2つの問題を解決できていないのが現状である。

そこで、本研究では、カメラを固定して長時間撮影できない環境を想定し、フリーハンドで撮影した時に発生する動画画像の揺れ(平行移動、回転やスケール変化)に対応することで、不安定な撮影条件下においても背景画像を取得できるロバスト性を持ったシステムを開発し、時間的制限と環境的制限の2つの問題を解決する。本研究で定義する背景画像とは、シーン中に歩行人や車両などの移動体が写り込んでいない風景画像とする。

2. 研究の概要

本研究では、背景画像の取得における時間的制限と環境的制限の問題を解決するため、フリーハンドで撮影した移動体を含む動画画像から背景画像を取得する手法を提案する。本研究のシステムの入力データ、出力データと処理の流れに関する概要を次に示す。

(1) 入力データ

本システムでは、市販のデジタルビデオカメラで撮影した動画画像を入力データとする。撮影方法は、カメラを持って立ったまま、一方向を撮影する。特に姿勢を固定せず撮影するため、平行移動に加えて回転やスケールによる画像の変化が起こる。シーン中には、移動体が途切れず常に存在しているものとする。

(2) 出力データ

本システムでは、入力した動画画像から移動体を除去した背景画像を生成する。

(3) 処理の流れ

本システムは、図-1に示すように、動画画像を入力し、フレーム画像正規化機能と背景画像生成機能の2つの機能によって、移動体を除去した背景画像を生成する。フレーム画像正規化機能では、まず、動画画像を構成する各フレーム画像の特徴量を算出して特徴点を自動生成し、フレーム間の対応点を取得する。次に、取得した対応点を用いて、フレーム間の揺れ(平行移動、回転やスケール変化)を推定し、フレーム画像を補正する。最後に、補正したフレーム画像に対して、非剛体レジストレーション¹⁰⁾を適用して画素単位的位置合わせを行う。

背景画像生成機能では、フレーム画像正規化機能で補正したフレーム画像列を用いて、動画中に含まれる移動体を除去した背景画像を生成する。まず、ピクセル状態分析¹¹⁾を用いて、移動体領域の推定を行う。次に、移動体領域の推定結果とフレーム画像列から作成したEPI⁹⁾を用いて背景画像を推定する。

3. フレーム画像正規化機能

本機能では、フレーム間の動きを推定するため、まず、フレーム毎に特徴点を生成してフレーム間の特徴点を追跡する。次に、追跡結果を基にフレーム間のアフィン変換行列を算出し、算出したアフィン変換行列でフレーム画像のズレを補正する。最後に、補正したフレーム画像の画素単位の補正を行うため、非剛体レジストレーション¹⁰⁾を適用する。

(1) フレーム間の対応点の取得

フレーム間の対応点の取得は、まず、フレーム画像から特徴点を生成する。次に、生成した特徴点を基にフレーム間の対応点を取得する。

(2) 特徴点の生成

特徴点の生成手法は、近年、スケールスペースを用いることで、画像の回転、スケールや照明の変化に不変な特徴量を取得する技術(SIFT : Scale-Invariant Feature Transform)^{12), 13)}が提案されている。SIFTは、回転、スケールや照明の変化に不変な特徴量であるため、パノラマ画像生成などに用いられている。前述したように、本研究では、カメラを固定することなく撮影するため、平行移動だけではなく回転やスケールによる画像の変化が考えられる。そこで、本研究でも、SIFT¹²⁾を用いて各フレームの特徴点を生成する。

(3) 対応点の取得

対応点の取得手法において、SIFTを用いた特徴点の追跡^{12), 13)}では、ある特徴点に対して異なるフレーム画像に含まれる全ての特徴点との特徴量間の距離を算出し、その中で距離が最小となる点同士を対応点とする。しかし、この方法は、追跡対象フレームの全特徴点を比較対象としているため、対応点の誤認識が多く発生する。その結果、対応点を取得した後に追跡に失敗した点を削除する処理が必要となる。また、従来の特徴点の追跡方法は、第Nフレームと第N+1フレームの連続するフレーム間で特徴点を追跡している。しかし、連続するフレーム間で特徴点を追跡した場合、フレームが進むにつれて誤追跡による対応点の誤差が蓄積され、後半のフレームになるほど、対応点の誤差が膨らむ。また、取得される対応点の数が減少し、特徴点を再生成する必要がある。

(4) 本研究における対応点の取得方法

本研究では、背景画像生成の精度を向上させるために、動画像中のある1つのフレーム画像を基準画像として、各フレーム画像を基準画像の位置に合わせる。その目的は、フリーハンドで撮影した動画像のフレーム画像間に介在するズレを取り除くことであり、連続したフレーム画像間で特徴点を追跡する必要は無い。本研究では、第1フレーム画像を基準画像として、第1フレーム画像の特徴点と各フレーム画像の特徴点を

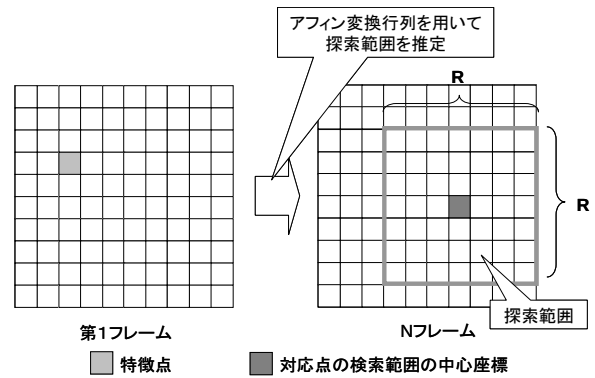


図-2 対応点の探索範囲

対応付ける。具体的には、まず、第1フレームと第2フレームの対応点を取得する。対応点の探索には第1フレームの特徴点の位置を中心に探索範囲($R \times R$)を指定する。そして、初期対応点群を基にしてアフィン変換行列を算出する。次に、第1フレームと第3フレームにおける対応点の位置推定は、第2フレームを補正する際に算出したアフィン変換行列を使用し、図-2に示すように、対応点の探索範囲($R \times R$)を限定する。そして、第Nフレームにおける特徴点の座標が推定した探索範囲内に含まれるかどうかを判定する。最後に、探索範囲内に含まれる特徴点の特徴量と追跡対象の特徴点の特徴量との距離 D を次式(1)により算出する。

$$D = \sqrt{\sum_{i=1}^n (v1_i - vN_i)^2} \quad (1)$$

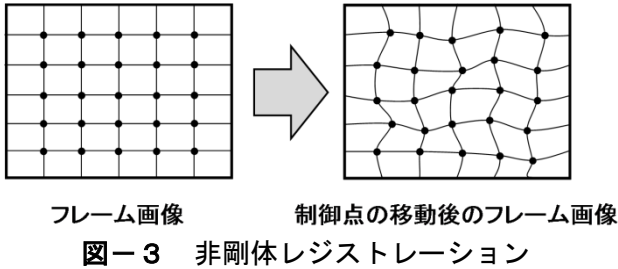
ここで、 $v1_i$ は第1フレームの特徴点、 vN_i は、第Nフレームの特徴点、 n は特徴量の次元数を表し、SIFTの特徴量は128次元であるため、 $n=128$ となる。最後に、算出した距離の中で最小となる特徴点を対応点として取得する。以上の処理を探索範囲内のすべての特徴点に繰り返し、対応点を取得する。

(5) フレーム画像の位置あわせ

フレーム画像の位置合わせでは、取得した対応点を用いて、フレーム間のアフィン変換行列を推定し、フレーム画像を補正する。さらに、補正したフレーム画像の画素単位での位置合わせを行うため、非剛体レジストレーション¹⁰⁾を用いて補正する。

a) アフィン変換行列によるフレーム画像の補正

本研究では、フレーム間の平行移動、回転、スケール変化の動きを補正するため、第1フレーム目の特徴点と第Nフレーム目の対応点を用いてアフィン変換行列を算出する。そして、各フレームで算出した変換行列を施すことで、基準とのフレーム間のズレを補正する。アフィン変換行列は、原画像と変換画像の対応座標値が3個以上判明していれば算出できる。



アフィン変換行列を算出する方法として、 n 個の特徴点の座標 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ とそれに対応する特徴点の座標 $(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)$ があるとき、RANSAC¹⁴⁾により3個以上の対応座標値を取得し、最小二乗法¹⁵⁾を用いてアフィン変換行列を算出する。そして、算出したアフィン変換行列を用いて第 N フレーム目の画像を変換させ、基準画像と変換画像の輝度値の差が最少となるときのアフィン変換行列を算出する。

b) 非剛体レジストレーションによるフレーム画像の補正

アフィン変換行列を用いて補正したフレーム画像に対して、画素単位での位置合わせを行う。アフィン変換は、フレーム間の位置補正として広く利用されている。しかし、撮影対象が平面近似できない場合や非剛体の場合、これらの変換方法では、高精度に補正することができない。そこで、本研究では、非剛体レジストレーション¹⁰⁾を用いてアフィン変換で補正した画像をさらに位置補正する。非剛体レジストレーションは、医療分野において、MRIやCTから得られた人体の画像同士の補正に使用され、画素単位での補正が必要な分野に高い効果を発揮している。非剛体レジストレーションでは、図-3に示すように、まず、変形対象となるフレーム画像に格子状の制御点を設定する。そして、制御点を移動することでフレーム画像を変形させる。フレーム画像の変形には、B-Spline関数を用いた制御点の移動により画像を変形させ、基準画像との輝度値の差が最少となる制御点の座標を算出して画像を補正する。

4. 背景画像生成機能

本機能では、補正したフレーム画像群から移動体領域を除去した背景画像を生成する。本研究では、様々な背景画像の推定手法の中から安定して背景画像を推定することができるEPI⁸⁾を使用する。しかし、使用する動画画像中に占める移動体の数が多い場合や動画画像の撮影時間が短い場合、生成した背景画像の中に移動体が残ってしまい、違和感のある画像が生成される。そこで、この問題を解決するため、補正したフレーム画像群から移動体領域を抽出する。そして、抽出した移

動体領域とEPIを使用して背景画像を生成する手法を提案する。

(1) 移動体領域の推定

移動体領域の抽出手法として、古くから背景差分^{16), 17)}が有名であり、最も抽出精度が良いとされている。しかし、背景差分には、移動体が写っていない背景画像が必要であるが、本研究では、予め背景画像を撮影することが困難な状況を想定しているため、背景差分は使用できない。その他の手法として、近年、AdaBoostなどの統計的学習を用いて移動体を抽出する研究^{18), 19)}が注目されている。しかし、統計的学習による手法は、予め膨大な量の基礎データが必要であり、また、移動体の種類や適応する環境毎に基礎データを作成する必要がある。

一方、Fujiyoshiらは、時系列に並んだ画像群を使用して各ピクセルの状態を分析することで、移動体領域を抽出する手法¹¹⁾を提案している。本研究では、デジタルビデオカメラを用いて時系列に並んだ画像群を使用して背景画像を生成することを目的としている。また、カメラの揺れに対応するため、各フレーム画像を基準画像に位置合わせしている。そこで、本研究では、補正されたフレーム画像群を使用してピクセル状態分析による移動体領域の抽出手法を提案する。

(2) ピクセル状態分析による移動体領域の抽出

ピクセル状態分析(PSA: Pixel State Analysis)¹¹⁾は、ピクセルの時間変化をモデル化することにより、各画素を背景、静状態、動状態の三状態に判別する手法である。そして、これらの状態を捉えるには、輝度値の変化と輝度値の安定度に着目する。ピクセル状態分析のアルゴリズムを図-4に示す。

図-4の $I(x, y)$ は、時刻 t における各画素の状態を表し、背景のとき bg 、動状態のとき tr 、静状態のとき st となる。 th_t は、輝度値の変化を判定する閾値、 th_s は、安定性を判定する閾値、 $background$ は背景画像を表す。既研究¹²⁾では、輝度値の変化量 T を求め

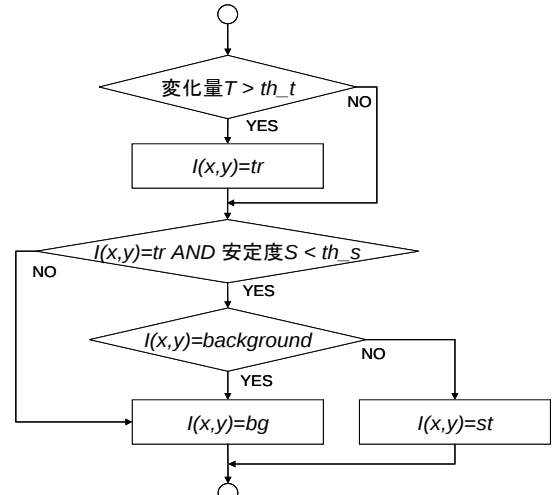


図-4 ピクセル状態分析のアルゴリズム

るには、時刻 t より前の画像の輝度値の差の最大値を使用しているが、本研究では、時刻 t の基準画像より後の画像を補正するため、時刻 t より後の画像の輝度値の差の最大値を使用する。変化量 T の算出式(2)を以下に示す。

$$T = \max(|f_t - f_{t+i}|, \forall_j [1, k]) \quad (2)$$

ここで、 f_t は時刻 t 、 $f_{(t+i)}$ は時刻 t から i 秒後の各画素の輝度値を表し、 k は変化量 T の算出に使用するフレーム数を表す。また、輝度値の安定度 S を算出するには、既研究¹¹⁾と同様に時刻 t より後のフレーム画像に着目する。安定度 S の算出式(3)を以下に示す。

$$S = \frac{k \sum_{i=1}^k f_{(t+i)}^2 - \left(\sum_{i=1}^k f_{(t+i)} \right)^2}{k(k-1)} \quad (3)$$

安定度 S は、 f_t から $f_{(t+k)}$ までの輝度値の分散であり、 S が小さい程、ピクセルは安定した状態であることを示す。

(3) 本研究における移動体領域の抽出

本研究では、前述(2)のピクセル状態分析¹¹⁾を改良した移動体領域の抽出手法を提案する。ピクセル状態分析を改良する理由として、本研究では、カメラを固定することができない環境下で撮影するため、常にカメラが微動する。そのため、フレーム画像正規化機能によって各フレーム画像を基準画像に補正している。しかし、補正した画像群は必ずしも基準画像と完全に一致せず、多少のずれが生じることがある。そのため、単純にピクセル状態分析を使用すると、移動体が存在する画素以外が動状態として判定される。

そこで、本研究では、画素単位ではなく、**図-5**に示すように、画像を縦 $h \times$ 横 w ピクセルで構成された矩形領域で分割し、領域単位でピクセル状態分析を行う。具体的には、まず、画像を縦 $h \times$ 横 w ピクセルの矩形領域で分割する。次に、各領域内において、輝度値のヒストグラムを作成する。そして、作成したヒストグラムの中央値を取得する。最後に、取得した中央値を用いて状態を分析する。ただし、本研究では、予め背景画像を用意することができないため、**図-4**にある背景画像 *background* との比較は行わず、**図-6**のように、安定度 S から直接ピクセルの状態を判定するアルゴリズムに変更する。

(4) 背景画像の生成

本研究で改良したピクセル状態分析によって抽出した移動体領域と EPI⁸⁾を使用して背景画像を生成する。EPI を用いた背景画像の推定手法では、まず、フレーム画像群から各座標の画素値を時系列(**図-7**)に並

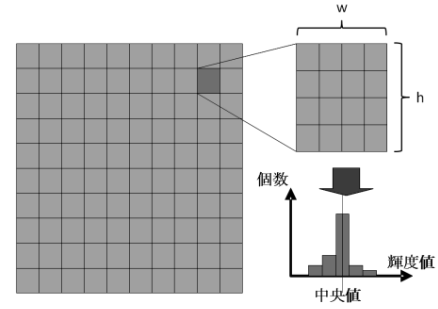


図-5 $h \times w$ ピクセルの領域毎による画素値の取得

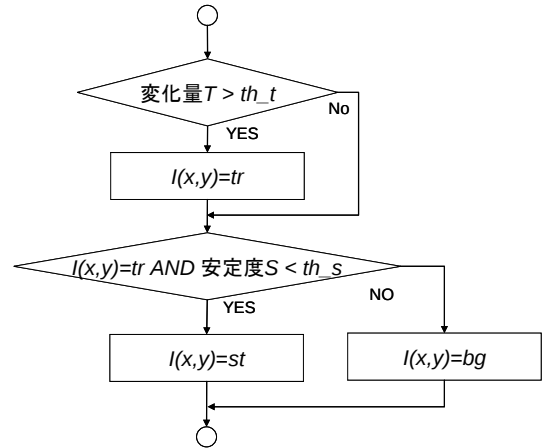


図-6 本研究のピクセル状態分析のアルゴリズム

べる。次に、各座標のヒストグラムを作成する。最後に、作成したヒストグラムから各座標の最頻値を算出し、最頻値を背景画素の輝度値であるとして、背景画像を生成する。ただし、本研究では、**図-8**に示すように、抽出した移動体領域をマスク化することで、移動体領域が除かれたヒストグラムから各座標の輝度値の最頻値を算出し、背景画像を生成する。

5. 実証実験

実証実験では、本研究の有効性を確認するため、「対応点の取得精度の検証」、「フレーム間における対応点数の検証」、「移動体領域の抽出精度の検証」と「背景画像の生成精度の検証」の4つの実験を行った。

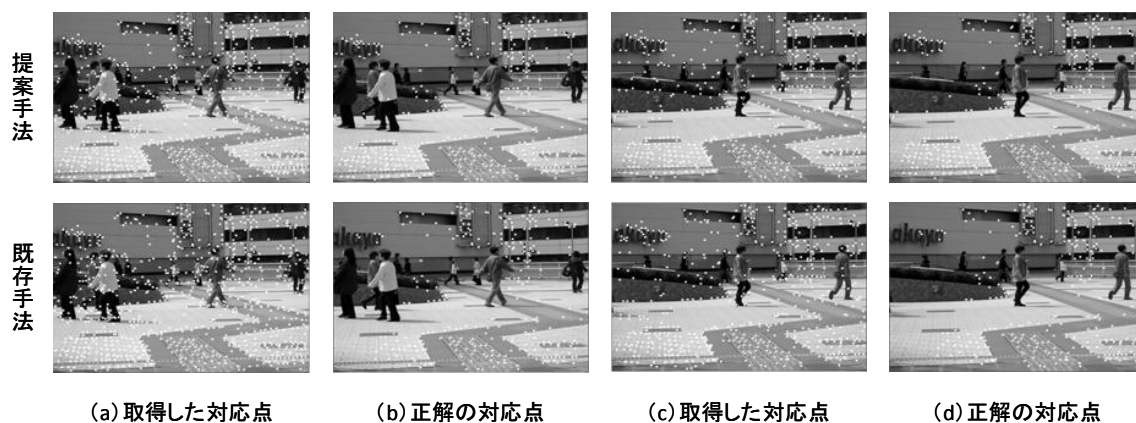
(1) 対応点の取得精度の検証

a) 実験内容

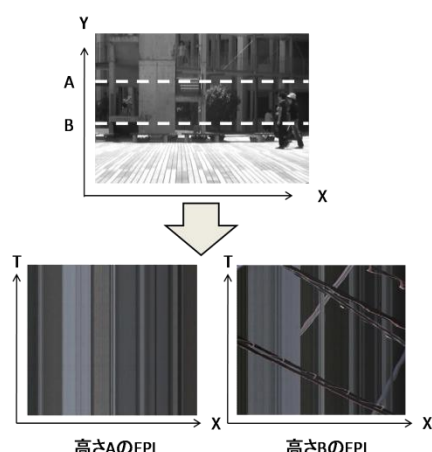
本実験では、提案手法と従来手法を用いて背景画像の生成に使用した画像から対応点を取得する。そして、本提案手法で探索範囲の絞込みだけを行った結果と従来の SIFT を用いた対応点の取得¹³⁾との結果を比較し、その有効性を確認する。結果の評価として、取得した対応点を目視で確認し、成功率を算出する。

b) 結果と考察

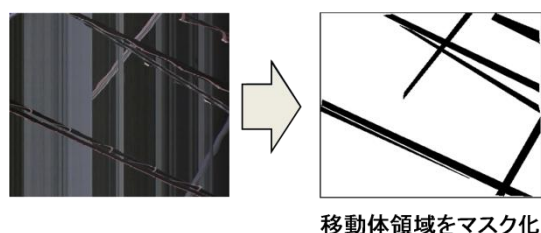
対応点の取得精度を表-1に、取得結果の例を**図-9**に示す。ここで、正解の対応点(**図-9**(b)と(d))



図－9 対応点の取得結果



図－7 EPI の作成

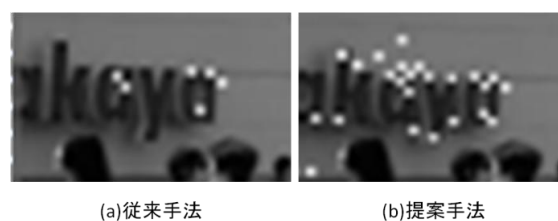


図－8 移動体領域をマスク化

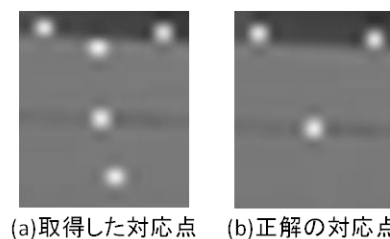
とは、取得した対応点（図－9（a）と（c））に対して、目視で確認し、一致している対応点のみを表示したものである。取得精度としては、従来のSIFTを用いた対応点の取得¹³⁾が約66%，提案手法である探索範囲の絞込みの手法が約81%という結果となった。さらに、図－9（b）と（d）の特徴箇所（エッジ箇所）を追跡した結果に注目したところ、図－10（a）と（b）に示すように、従来手法では、比較的追跡が容易であるはずの特徴箇所をほとんど正確に追跡できていなかった。従来手法では、ある特徴点に対して画像全体の特徴点を比較対象とし、特徴点間の特徴量のユークリッド距離が最小となる点を対応点としているため、実際の対応点ではない点を対応点として判定したことが誤追跡の原因と考えられる。一方、提案手法では、エッ

表－1 従来手法と提案手法の比較結果

手法	成功率
本研究における提案手法	81.44%
従来研究の手法 ¹³⁾	66.58%



図－10 特徴箇所の追跡結果



図－11 誤追跡が発生した領域

ジなどの特徴箇所を正確に追跡できていることが確認できた。この理由として、一定時間の連続性に着目し、1つ前の対応点群から取得したアフィン変換行列を用いて対応点位置の推定を行い、探索範囲を限定することにより、対応点の誤判定を防ぐことができたと考えられる。以上の結果から、本研究で提案した探索範囲の絞込みによる対応点の取得方法は、対応点の取得精度を向上させる手法として有効であると言える。提案手法で約19%見受けられた誤追跡の原因として、図－9（b）と（d）の誤追跡が多い領域（図－11）に注目する。誤追跡が発生した領域（図－11）は、同一色でエッジなどの特徴箇所がほとんど存在せず、その範囲も比較的広い。そのため、周辺画素間の輝度差がほとんど無く対応点候補を絞りきれなかったため、誤追跡が発生したと考えられる。提案手法で約19%見受けられた

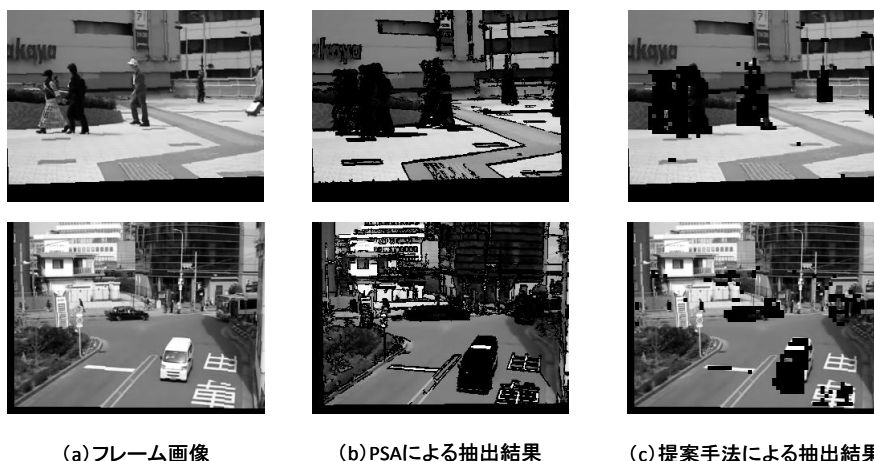


図-1 3 移動体領域の抽出結果

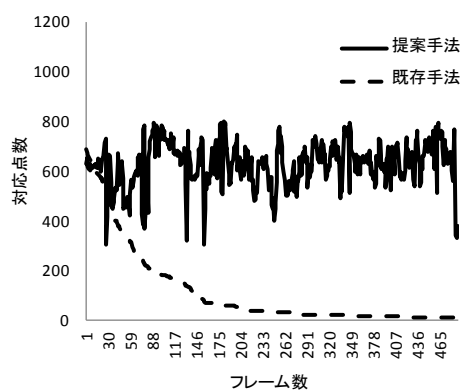


図-1 2 特徴点追跡による対応点数

誤追跡の原因として、図-9 (b) と (d) の誤追跡が多い領域 (図-1 1) に注目する。誤追跡が発生した領域 (図-1 1) は、同一色でエッジなどの特徴箇所がほとんど存在せず、その範囲も比較的広い。そのため、周辺画素間の輝度差がほとんど無く対応点候補を絞りきれなかったため、誤追跡が発生したと考えられる。

(2) フレーム間における対応点数の検証

a) 実験内容

本実験では、カメラを固定することなく撮影した動画画像に対して対応点の取得を行い、フレーム間における対応点数を比較する。4つの異なる動画画像に対して、第Nフレームと第N+1フレームの画像の特徴点を追跡する方法と本研究の追跡手法を用いて対応点を取得した結果を比較することで提案手法の有効性を確認する。

b) 結果と考察

既研究の手法¹³⁾と提案手法による特徴点の追跡結果を図-1 2に示す。これは、異なる動画画像において各フレームで取得した対応点数の平均値の推移を表す。既研究の手法による特徴点の追跡では、フレームが進むにつれて、取得できた対応点数が大きく減少している。これは、追跡できた対応点を基に次フレームの追

跡を行っていることが原因として挙げられる。そのため、300フレーム以降では、アフィン変換係数を算出するために必要な3点の対応点を確保できない結果となった。一方、提案手法では、平均して約600個の対応点を取得できている。対応点数の減少は、全てのフレームを第1フレームの特徴点から追跡することで回避できた。以上の結果から本研究の提案手法は、本研究で想定する条件下において、従来手法と比べてフレームが進むにつれても安定して対応点の取得を実現できたと言える。

(3) 移動体領域の抽出精度の検証

a) 実験内容

本実験では、補正した画像群を用いて移動体領域を抽出し、従来手法と抽出精度を比較する。従来のピクセル状態分析¹¹⁾と本研究で改良したピクセル状態分析によって10枚の補正した画像から移動体領域を抽出し、その結果を比較することで提案手法の有効性を目視で確認する。

b) 結果と考察

移動体領域の抽出結果を図-1 3に示す。図-1 3 (a) は、フレーム画像、(b) は従来のピクセル状態分析による移動体領域の抽出結果、(c) は本研究で改良したピクセル状態分析による移動体領域の抽出結果を表す。図-1 3 (b) に示すように、従来のピクセル状態分析を使用した場合、移動体領域の抽出に成功しているが、それ以外に多くの領域が抽出されている。その理由として、非剛体レジストレーションを用いてフレーム画像を補正しても完全に基準画像と一致するわけではなく、画像中に存在するエッジ付近の少しのずれでPSAの変化量と安定度が大きくなる結果、実際の移動体領域以外の領域が移動体領域として抽出されたと考えられる。一方、図-1 3 (c) に示すように、本研究で改良したピクセル状態分析では、(b) の従来のピクセル状態分析に比べて移動体領域以外の領域の抽出を防いでいる。その理由として、本研究では、画

表－２ 動画像の仕様

ファイル形式	AVI
AVIコーデック	Indeo Video 5.0
フレームレート	30 fps
動画サイズ	640×480 pixel

像を縦 h × 横 w ピクセルの矩形領域に分割し、領域内の中央値を分析に使用している。そのため、画像間にずれが生じた場合でもその影響を受けにくくなり、その結果、従来のピクセル状態分析に比べ正確に移動体領域を抽出できたと考えられる。

(4) 背景画像の生成精度の検証

a) 実験内容

本実験では、カメラを固定せずフリーハンドで撮影した動画像から、本研究のシステムを用いて背景画像を生成する。実験には、シーン中の移動体の移動が確認でき、かつ撮影者への時間的負担が少ない5秒(150フレーム)程度撮影した動画像を使用する。撮影した動画像中には、背景のみのシーンは存在せず、全シーンに渡って途切れることなく移動体が存在しているものとする。実験に使用した動画像の仕様を表－２、本実験に使用した動画像を図－１４に示す。また、背景画像の生成では、まず、補正なしの動画像と提案手法によって補正した動画像から各々背景画像を作成し、その結果を比較することで提案手法の有用性を確認する。ただし、補正した動画像からの背景画像の生成については、従来研究の手法であるロバスト推定による手法^{1),2)}、VMFを用いた手法⁵⁾、EPIのみを用いた手法^{3),4)}、そして、本研究で提案した手法の4つの生成結果を比較することによって、本研究で提案した背景画像の生成手法の有用性を確認する。

b) 結果と考察

フレーム画像の補正を行わず生成した背景画像を図－１５に、補正したフレーム画像に対して各々の手法を用いて生成した背景画像を図－１６に示す。図－１５に示すように、補正を行わずに背景画像を生成した場合、強いぼけを残した背景画像が生成された。この結果から、背景画像の推定において、撮影時の揺れ(移動、回転やスケール変化)から強い影響を受けていることが明らかである。一方、図－１６に示すように、補正した画像を用いて背景画像を生成した場合、図－１５に示すような強いぼけや画像情報の劣化も見受けられない。しかし、手法によっては、移動体を完全に除去することができていない。たとえば、ロバスト推定^{1),2)}とEPI^{3),4)}による背景画像の生成結果(図－１６

(a) と (c)) では、移動体は見受けられないが、細かなノイズが残っており違和感のある画像になった。また、VMF⁵⁾による背景画像の生成結果(図－１６(b))では、ある程度移動体を除去できている画像もあるが、

画像によって移動体の一部が除去しきれずに残っている。一方、本研究で提案した手法(図－１６(d))では、他の3つの手法に比べて移動体を正確に除去できており、違和感のない背景画像が生成できた。その理由として、EPIによる背景画像の生成を行う前に各フレーム画像中に存在する移動体を抽出し、マスクしたことで、抽出した移動体領域の画素を背景画像の生成処理から省くことができたと考えられる。また、ロバスト推定やVMFを用いた手法においても背景画像の生成の際に、抽出した移動体領域の画素を省いて実験を行ったが、本研究で提案した手法に比べて細かなノイズが多く残る結果となった。以上のことから、本研究で提案した手法が本研究で想定している時間的制限と環境的制限を解決できる手法であると言える。

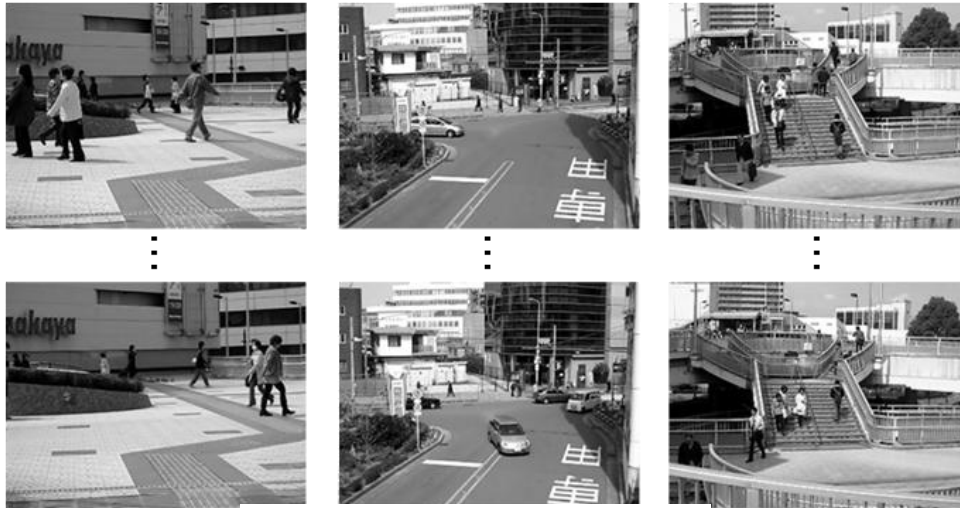
6. 総括

本研究で作成したシステムは、動画像の揺れ(移動、回転やスケール変化)に対応することで、カメラを固定することなく撮影した動画像からも背景画像を生成できた。そのため、本研究によって、カメラを固定して撮影しなければならないという環境的制限を軽減できたと言える。また、5秒程度の短い動画像を用いて移動体の存在しない背景画像を生成できたことから背景画像の取得における時間的制限についても概ね解決できたと考える。これにより、時間的制限と環境的制限の両制限に対してロバストな背景画像生成手法を実現できたと考える。さらに、安価な市販のデジタルビデオカメラで撮影した動画像から背景画像を生成するため、特別な機器や準備を必要しないことから、他の手法やシステムとの連携も容易であるため拡張性も高い。

しかし、本研究の背景画像生成手法は、サンプリング対象である全フレーム画像の画素中に存在する移動体の数が実行結果に影響を及ぼす。たとえば、移動体の数が増えると背景領域が動画像中に一度も映らない箇所がある。その場合、移動体を抽出しても背景領域が存在しないため、その箇所の背景画像を推定することができない。この問題については、撮影時間を10秒、20秒と増やすことや時間を空けて再び撮影した画像群を追加することで移動体を除去した背景画像を生成することができた。しかし、この方法では、短時間とはいえ撮影場所の状況に影響を受けて追加の撮影時間やタイミングが変動するため、制限の解決に向けてさらなる手法の改良が必要であると考えられる。

7. おわりに

本研究では、カメラの揺れに対応することで、不安



図ー 1 4 本実験で使した動画像



図ー 1 5 補正を行わず生成した背景画像

(a) ロバスト推定



(b) VMF



(c) EPI



(d) 提案手法
(PSA+EPI)



図ー 1 6 各手法で生成した背景画像

定な撮影環境下において動画像から移動体を除去した背景画像を生成できる背景画像生成手法を提案した。まず、本研究では、フレーム画像正規化機能において、SIFTによって取得した特徴点に対して探索範囲を限定することで、単純なSIFTによる特徴点追跡¹³⁾に比べ高精度な特徴点追跡を実現した。そして、これらの結果から、各フレームのアフィン変換行列を算出してフレーム画像のズレを補正し、非剛体レジストレーション¹⁰⁾を行うことで、アフィン変換や射影変換では補正することが困難な画素単位での位置合わせを行った。また、背景画像生成機能において、ピクセル状態分析¹¹⁾を用いた移動体領域の抽出結果とEPI⁸⁾を用いて移動体を除去した背景画像の自動生成を実現した。

これらの手法の有効性を実証するため、4つの実証実験を行った。特に、「背景画像の生成精度の検証」では、他の背景画像の生成手法¹¹⁾⁵⁾と比較し、本研究で提案した手法の有効性を実証した。しかし、本研究では、移動体が常に画像中を移動することを前提としたが、移動体が停止した場合、本提案手法では、移動体領域を抽出することができず、生成した背景画像に移動体が残ってしまう可能性がある。そのため、今後は、これらの問題について検討する予定である。

謝辞：本研究の一部は、平成 20～24 年度私立大学戦略的研究基盤形成支援事業（研究課題「セキュアライフ創出のための安全知循環ネットワークに関する研究」）から助成を受け、その成果を公表するものである。

参考文献

- 1) 栗田多喜夫, 梅山伸二, 島井博行: 適応的なロバスト推定を用いた背景のモデリング, 情報ソサイエティ大会講演論文集, 電子情報通信学会, D-12-92, pp.314, 1998 年 7 月.
- 2) 島井博行, 栗田多喜夫, 梅山伸二, 田中勝, 三島健稔: ロバスト統計に基づいた適応的な背景推定法, 電子情報通信学会論文誌, 電子情報通信学会, Vol.86, No.6, pp.796-806, 2003 年 6 月.
- 3) 羽下哲司, 鷺見和彦, 橋本学, 関明伸, 黒田伸一: 時空間画像処理による注目距離対象の強調に基づく遠隔監視システム, パターン認識・メディア理解技術研究報告, 電子情報通信学会, Vol.97, No.325, pp.41-46, 1997 年 10 月.
- 4) 栗林宏輔, 小野晋太郎, 川崎洋, 池内克史: 時空間画像解析を利用した車載カメラ映像からの障害物除去手法, コンピュータビジョンとイメージメディア研究報告, 情報処理学会, Vol.2009, No.167, pp.1-8, 2009 年 6 月.
- 5) 内山寛之, 出口大輔, 高橋友和, 井手一郎, 村瀬洋: 複数映像統合による車載カメラ映像からの移動物体除去, パターン認識・メディア理解技術研究報告, 電子情報通信学会, Vol.109, No.306, pp.57-62, 2009 年 11 月.
- 6) 栗田多喜夫: ロバスト統計に基づく画像解析, 電子情報通信学会誌, 電子情報通信学会, Vol.76, No.12, pp.1293-1297, 1993 年 12 月.
- 7) R. Hampel, M. Ronchetti, J. Rousseeuw, A. Stahel: Robust Statistics The Approach Based on Influence Functions, Wiley-Interscience, 1986.
- 8) R. Bolles, H. Baker, D. Marimont: Epipolar plane image analysis an approach to determining structure from motion, International Journal on Computer Vision, Kluwer Academic Publishers, Vol.1, pp.7-55, 1987.
- 9) J. Astola, P. Haavisto, Y. Neuvo: Vector Median Filters, Proceedings of the IEEE, IEEE, Vol.78, No.4, pp.678-689, 1990.
- 10) D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, D. Hawkes: Nonrigid Registration Using Free-Form Deformations, IEEE Transactions on Medical Imaging, IEEE, Vol.18, No.8, 1999.
- 11) Fujiyoshi, H., Kanade, T.: Layered Detection for Multiple Overlapping Objects, IEICE Transactions on Information and systems, IEICE, Vol.87, No.12, pp.2821-2827, 2004.
- 12) D. Lowe: Distinctive image features from scale-invariant keypoints, Journal of Computer Vision, Springer, Vol.60, No.2, pp.91-110, 2004.
- 13) 藤吉弘亘: Gradient ベースの特徴抽出-SIFT と HOG-, コンピュータビジョンとイメージメディア研究報告, 情報処理学会, Vol.2007, No.160, pp.211-224, 2007 年 8 月.
- 14) M. Fischler, R. Bolles: Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography, Communications of the ACM, ACM, Vol.6, No.24, pp.381-395, 1981.
- 15) 金谷健一: これなら分かる応用数学教室, 共立出版, 2003 年.
- 16) 中井宏章: 事後確率を用いた移動体物体検出手法, コンピュータビジョン研究会研究報告, 情報処理学会, Vol.94, No.81, pp.1-8, 1994 年 9 月.
- 17) 村中徳明, 徳丸正孝, 今西茂, 伊達啓: 安全運転支援システムにおける拡張背景差分を用いた歩行者検出法, 電気学会論文誌, 電気学会, Vol.124, No.8, pp.1575-1584, 2004 年 8 月.
- 18) Wu, B. and Nevatia, R.: Detection of Multiple, Partially Occluded Humans in a Single, International Journal of Computer Vision, Springer, Vol.75, No.2, pp.247-266, 2007.
- 19) 松島千佳, 山内悠嗣, 山下隆義, 藤吉弘亘: 人検出のための Real AdaBoost に基づく HOG 特徴量の効率的な削減法, コンピュータビジョンとイメージメディア研究会研究報告, 情報処理学会, Vol.167, No.32, pp.1-8, 2009 年 6 月.

(2010. 5. 28 受付)