

# 複数画像系列の部分画像選択に基づく移動物体を含まない車載カメラ映像の生成

内山 寛之<sup>†a)</sup>      出口 大輔<sup>†</sup>      高橋 友和<sup>††</sup>      井手 一郎<sup>†</sup>  
村瀬 洋<sup>†</sup>

Generation of an In-Vehicle Camera Image Sequence without Temporary Objects  
by Selection of Partial Images from Multiple Image Sequences

Hiroyuki UCHIYAMA<sup>†a)</sup>, Daisuke DEGUCHI<sup>†</sup>, Tomokazu TAKAHASHI<sup>††</sup>, Ichiro IDE<sup>†</sup>,  
and Hiroshi MURASE<sup>†</sup>

あらまし 近年、市街地映像を利用した運転支援システムが注目されている。また、“Google ストリートビュー”のような、市街地映像をインターネット上で閲覧できるサービスも提供されている。しかし、画像中に車両や歩行者などの移動物体が映り込むことが、遮へいやプライバシー保護の点で問題となっている。そこで本論文では、複数の車載全方位カメラ映像の部分画像を選択して貼り合わせることにより、移動物体を含まない映像を生成する手法を提案する。まず、カメラを搭載した車両が同じ経路を複数回走行することで、異なる時刻の複数の映像を収集する。そして、これらの映像間の時間方向と空間方向のレジストレーションを行う。次に、画像の同位置において移動物体が映っている頻度は少ないという仮定のもと、定義した評価関数を最小化することで、背景らしい部分画像を選択する。最後に、選択された部分画像を張り合わせることで、移動物体を含まない市街地映像を作成する。市街地を撮影した実際の全方位カメラ映像を用いた実験により、移動物体の領域の96.5%を除去した映像を生成することに成功した。

キーワード 移動物体除去、車載カメラ、ITS、市街地映像

## 1. ま え が き

近年、市街地の画像を蓄積したデータベースをドライバ支援に用いる研究が行われている。例えば、位置情報付き画像からなるデータベースと車載カメラ画像を照合することにより、自車位置推定を行う手法が提案されている[1], [2]。しかし、データベース中の移動物体による遮へいが照合の際に問題となる。また、事前に収集した移動物体を含まない画像と入力画像を比較することにより、路上障害物を検出する手法も提案されている[3]。しかし、そのためには移動物体が存在しない画像データベースを用意する必要がある。一方、

“Google ストリートビュー”<sup>(注1)</sup>などの、インターネット上で市街地画像が閲覧できるサービスが提供されている。しかし、人物の顔や車のナンバープレートも画像に含まれており、これらがプライバシー保護の観点から大きな問題となっている。これらを自動検出してぼかす処理[4]が行われているが、現状では十分な精度が得られていない上に、画像中の移動物体で見たい部分が遮へいされるという問題もある。

そこで本論文では、ドライバ支援とプライバシー保護のための市街地映像データベースの構築を目的とし、走行する車両より得られた全方位カメラ映像から移動物体を除去する手法を提案する(図1)。同一経路を複数回走行して得られた映像を統合することにより、移動物体を含まない映像を生成する。

画像中から不要な物体を除去する手法として、同一地点で異なる時刻に撮影した複数の静止画を利用する

<sup>†</sup> 名古屋大学大学院情報科学研究科, 名古屋市  
Graduate School of Information Science, Nagoya University,  
Nagoya-shi, 464-8601 Japan

<sup>††</sup> 岐阜聖徳学園大学経済情報学部, 岐阜市  
Faculty of Economics and Information, Gifu Shotoku  
Gakuen University, Gifu-shi, 500-8288 Japan

a) E-mail: uchiyama@murase.m.is.nagoya-u.ac.jp

(注1) : Google maps, <http://maps.google.com/>

もの[10], [11]が提案されている。しかし、これらは移動カメラを対象とした手法ではない。そのため、大規模な市街地映像データベース構築への直接的な適用は困難である。また、homography 行列を推定することで視点が異なる画像間のレジストレーションを行った後、グラフカットを利用した画像の貼り合わせを行うことで、移動物体が映っていない画像を生成する手法も提案されている[12]。しかし、この手法も車載カメラのような状況を想定していないため、本研究とは状況設定が異なる。また、homography 行列推定を、全方位カメラ映像に直接適用することは難しい。

そのほかに、単一の静止画を利用する手法[5], [6]や、画像系列（動画）を利用する手法[7]~[9]が提案されている。これらの手法は、除去対象領域をあらかじめ人手や自動検出等により指定する必要があったり、除去対象領域の検出処理が手法中に含まれている。しかしながら、膨大な市街地映像中から人手で除去対象領域を指定することは困難であり、一つの静止画像や画像系列から多様に変化する車両や人物の領域を検出すること自体が一つの難しい問題である。

そこで本論文では、複数車載カメラ映像中の部分画像を選択して貼り合わせることにより、移動物体を含まない画像を生成する手法を提案する。単一の映像ではなく、同一経路の複数の映像を用いることにより、検出処理が不要な手法を実現する。複数の映像の収集は、一般車両や公共の交通機関、運輸業者等の車両にカメラを装着することで実現できる。提案手法では、複数の異なる位置・異なる時間で撮影された映像中の多数の部分画像を貼り合わせるという問題を、(1) 非線形かつ精緻な映像間の時間的・空間的レジストレーション手法と、(2) ベクトルメジアンフィルタの出力と隣接する部分画像間の見た目の連続性を考慮した目的関数の最小化による部分画像の選択手法を導入することで解決する。

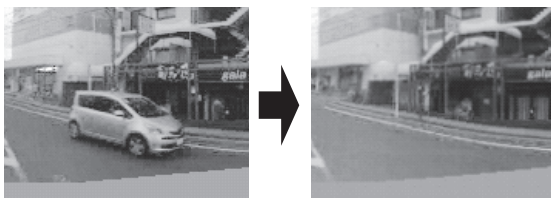


図1 映像中からの移動物体の除去（提案手法の適用例）

Fig. 1 Removal of temporary objects from an image (example of the application of the proposed method).

以降、2. で映像からの移動物体を除去する手法を述べ、3. で評価実験とその考察について述べる。最後に、4. でまとめを述べる。

## 2. 映像からの移動物体除去手法

### 2.1 概要

まず、提案手法の基本戦略を述べる。本論文では、路上に存在する自動車（一時的に駐停車しているものを含む）、自転車、歩行者など、恒常的にその場に存在しないものを移動物体と呼び、それらを除去の対象とする。同一地点で撮影した映像中の同一部分を長時間観察した場合、一般に移動物体が映っている時間はわずかである。そこで、異なる時刻に同一地点で撮影した画像を部分画像に分割し、貼り合わせることで、全体として移動物体を含まない画像を得る（図2）。

しかし、実際には、車載カメラにより映像を撮影した場合、走行ごとに車両の速度や軌跡が異なるため、映像ごとに各フレームの撮影位置は異なる。そのため、映像間のレジストレーション処理により、複数の車載カメラ映像から、同一地点で撮影したものと見た目が類似する画像群を生成する。

提案手法は合計  $N$  本の映像を用いる。ここで、移動物体除去の対象となる映像を基準映像（画像系列）、残りの  $N - 1$  本の映像をソース映像（画像系列）と呼ぶことにする。提案手法の流れを図3に示す。カメラの特性に起因し、各映像の明るさや色合いが異なる

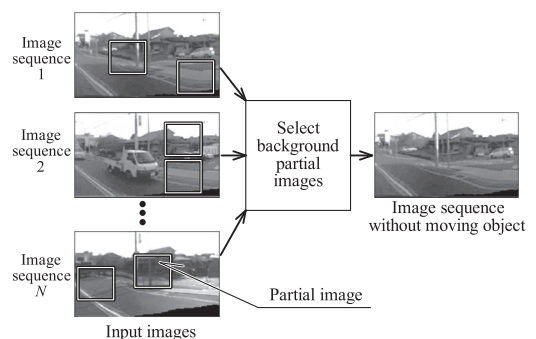


図2 異なる時刻に撮影した同一地点の画像を収集し、移動物体が映っていない部分画像を自動選択して張り合わせることで、全体として移動物体が映っていない画像を生成する。

Fig. 2 First, images captured at the same place on different timings are collected. Next, the proposed method generates an image without temporary objects by selecting and mosaicing the partial images without temporary objects.

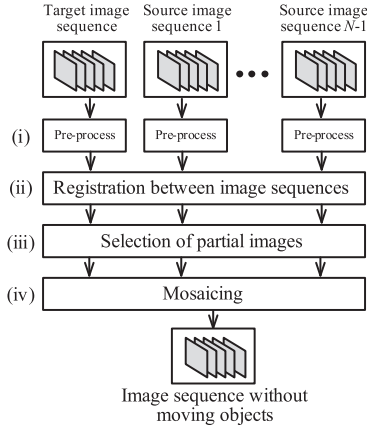


図 3 提案手法の処理の流れ  
Fig.3 Process flow of the proposed method.

ことがある。そこで、(i)の事前処理では、映像の各フレームに色ヒストグラム正規化処理を適用し、画像全体の明るさや色合いを正規化する。次に、(ii)の基準映像とソース映像間の時間的・空間的な位置合せを行うレジストレーション処理により、複数の車載カメラ映像から、同一地点で撮影したものと見た目が類似する画像群を生成する。(iii)の移動物体を含まない部分画像の選択処理では、異なる時刻・同一地点の部分画像の中から、最も背景らしい部分画像を選択する。最後に、(iv)のモザイク処理により、選択された部分画像の貼り合わせを行う。

## 2.2 映像間のレジストレーション

### 2.2.1 時間方向のレジストレーション

まず、全方位カメラ映像間の時間方向の位置合せにより、車両の進行方向の大まかな位置合せを行う(図4)。車両の速度は走行ごとに異なり、同じ走行中でも一定ではない。そこで、DP マッチングに基づく手法[13]を適用することでフレーム間の対応付けを行い、このような非線形な時間軸の伸縮に対処する。本論文では、映像を撮影する車両は同じ車線内を走行し、車線方向の大きな位置ずれはないと仮定する。そのため、DP マッチングのパターン要素間の距離は、単純に画像の画素値の二乗誤差とする。ここで、パターン要素間の距離を計算する際、建物が映る全方位画像の側面の領域だけを利用し、正面・背面の領域は使用しない。また、基準画像の1フレームに対して、ソース映像の複数フレームが対応する場合がある。このような場合には、対応づいた複数フレームの中央のフレームを選択する。また、DP マッチングは、基準映像と

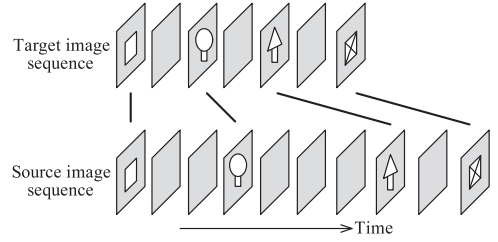


図 4 DP マッチングによる画像系列の時間方向の位置合せ  
Fig.4 Temporal alignment between image sequences performed by DP matching.

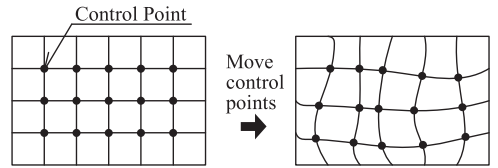


図 5 非剛体レジストレーション  
Fig.5 Non-rigid registration.

ソース映像の  $N - 1$  個の対について、それぞれ適用する。

### 2.2.2 空間方向のレジストレーション

映像を収集する車両は、同じ車線を走行したとしても、車線内での走行位置は走行ごとに異なる。また、カメラのフレームレートの制限により、DP マッチングにより進行方向の撮影位置を合わせた結果にも誤差が含まれる。そのため、映像の時間方向の位置合せだけでなく、空間方向の位置合せも行う必要がある。

市街地環境では建築物が入り組んでいるため、撮影位置が異なると景色の見え方に非線形な伸縮が生じる。また、後の部分画像選択処理で、部分画像をきれいに繋ぎ合わせるために、ピクセル単位で画像の位置合せを行う必要がある。そこで、撮影位置の違いによる画像の見え方の変化を B-spline 曲面で近似し、free-form deformation による非剛体レジストレーション[14]を適用することで、映像の空間方向の対応付けを行う。

非剛体レジストレーションでは、まず図5のように、変形対象となるソース画像上に制御点を格子状に設置する。そして、制御点を移動させることでソース画像を変形する。基準画像との照合コストが最小となるように、ソース画像を変形することで、ソース画像を基準画像に一致させる。ここで、ソース画像と基準画像間の照合コストは、画像間の正規化相互情報量

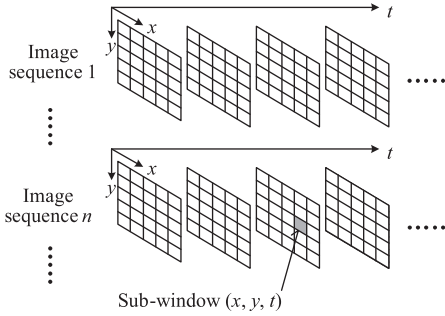


図 6 画像系列とサブウィンドウ  
Fig. 6 Image sequences and sub-windows.

$C_{\text{NMI}}$  と B-spline 曲面の滑らかさに関するペナルティ  $C_{\text{smooth}}$  を用いて次式のように計算する.

$$C_{\text{cost}} = -C_{\text{NMI}} + C_{\text{smooth}} \quad (1)$$

ここで,  $C_{\text{smooth}}$  は B-spline 曲面上の各点における二次微分の二乗和に重みを乗じたものである.  $C_{\text{smooth}}$  の項は, 画像の極端な変形により, 見た目が不自然になることを防ぐ効果がある.

### 2.3 移動物体を含まない部分画像の選択

#### 2.3.1 部分画像選択のための目的関数

レジストレーションを行った画像系列を部分画像単位に分割する. そして, 複数の画像系列の同座標の部分画像群の中から, 移動物体が映っていない部分画像を選択する. 以降, 特定の位置において, 部分画像をいずれかの画像系列から選択する枠のことを, サブウィンドウと呼ぶ. まず, 各画像上に  $W \times W$  画素のサブウィンドウを, 隣接するもの同士が半分程度重なり合うように設置し, サブウィンドウ内の画素値をベクトルとして表す. このベクトルを以降, 部分画像ベクトルと呼ぶ. 部分画像ベクトルは, 各画素の RGB 値を並べたものであり,  $3W^2$  次元である. サブウィンドウは, 画像系列の空間軸及び時間軸に沿って並べられており, その位置を  $(x, y, t)$  と表記する (図 6). ただし,  $x = 1, 2, \dots, X$ ,  $y = 1, 2, \dots, Y$ ,  $t = 1, 2, \dots, T$  である. そして, 移動物体が映っている頻度は少ないという仮定のもと, 以下の処理により背景らしい部分画像を含む画像系列  $n$  を決定する.

まず, 背景らしい部分画像を含む画像系列を選択する基準として, 次の二つを用いる.

- (i) 移動物体らしさ
  - (ii) サブウィンドウ間の不連続度
- 具体的には, あるサブウィンドウ  $(x, y, t)$  に注目し,

画像系列  $n$  を選択した場合における移動物体らしさ  $f_{xyt}(n)$  と, サブウィンドウ間の不連続度  $g_{xyt}(n)$  を導入する. そして,  $f_{xyt}(n)$  と  $g_{xyt}(n)$  の重み付き和を考え, 各サブウィンドウにおいて選択される最適な画像系列  $\mathbf{n} = \{n_{xyt}\}$  ( $x = 1, 2, \dots, X$ ,  $y = 1, 2, \dots, Y$ ,  $t = 1, 2, \dots, T$ ) を決定するために, 次の目的関数を  $\mathbf{n}$  に関して最小化する.

$$F(\mathbf{n}) = \sum_{x,y,t} [(1-\lambda)f_{xyt}(n_{xyt}) + \lambda g_{xyt}(n_{xyt})] \quad (2)$$

ここで,  $\lambda$  は重みである.

移動物体らしさ  $f_{xyt}(n)$  が小さいほど, その部分画像中に移動物体が存在しにくい. また, サブウィンドウ間の不連続度に関する項  $g_{xyt}(n)$  を導入することにより, 出力画像系列の隣接するサブウィンドウ間で, 入力画像系列の照明条件の関係を保存する. 以降,  $f_{xyt}(n)$  と  $g_{xyt}(n)$  について説明する.

#### 2.3.2 移動物体らしさ $f_{xyt}(n)$

移動物体は同じ地点にとどまらないため, 異なる時刻かつ同一地点で撮影した画像中の同じ部分を観察した場合, 一般的に背景の出現頻度が最も高い. そこで, 移動物体らしさ  $f_{xyt}(n)$  をベクトルメジアンフィルタ [15] を用いて計算する. ベクトルメジアンフィルタはメジアンフィルタの多次元への拡張であり, 他のベクトルとの距離の和が最小となるベクトルを選択するフィルタである.  $N$  個のベクトル  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$  を入力とするベクトルメジアンフィルタの出力  $\mathbf{m}$  は, 次式で表される.

$$\mathbf{m} = \arg \min_{\mathbf{v} \in \{\mathbf{v}_1, \dots, \mathbf{v}_N\}} \sum_{i=1}^N |\mathbf{v} - \mathbf{v}_i| \quad (3)$$

ここで,  $|\cdot|$  はベクトルの L2 ノルムを表す.

ベクトルメジアンフィルタは外れ値を除外する性質がある. そのため, 図 7 のように, 中央値に近い部分画像ベクトルほど移動物体が存在しにくいと考えられる. そのため, 中央値との距離を, その部分画像ベクトルにおける移動物体らしさ  $f_{xyt}(n)$  とする.

$$f_{xyt}(n) = |\mathbf{v}_{xyt}(n) - \mathbf{m}_{xyt}| \quad (4)$$

ここでは, サブウィンドウ  $(x, y, t)$  での  $n$  番目の画像系列の部分画像ベクトルを  $\mathbf{v}_{xyt}(n)$ , 部分画像ベクトルのメジアンを  $\mathbf{m}_{xyt}$  と表記する.

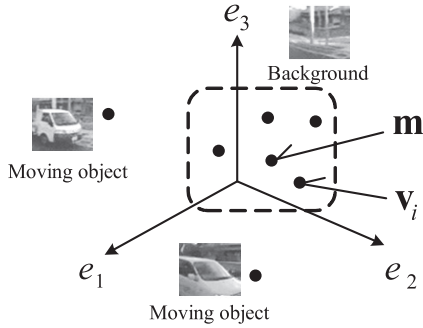


図7 移動物体らしさの計算. メジアンに近い部分画像ベクトルほど, 移動物体が存在しにくい.

Fig.7 Calculation of likelihood of temporary objects. A vector closer to the median tends to be the background without a temporary object.

ベクトルのメジアンフィルタを用いることにより, 背景画像の選択を画素の RGB 成分単位ではなく, 領域単位で行うことが可能となる. そのため, 背景画像選択処理において, 単一の画素では背景と移動物体の判別が困難な場合でも, 複数の画素を用いることにより, 正しい判別が可能となる.

### 2.3.3 サブウィンドウ間の不連続度 $g_{xyt}(n)$

式 (2) の目的関数にサブウィンドウ間の不連続度  $g_{xyt}(n)$  を導入することにより, 出力画像系列の隣接するサブウィンドウ間において, 入力画像系列の照明条件の関係を保存する. 前述のように, 図 3 (i) の事前処理により画像全体の明るさと色の正規化を行うが, 屋外環境では時間帯や天候により, 局所的な陰影の有無が異なる. それぞれのサブウィンドウにおいて独立に部分画像選択処理を行った場合, 隣接するサブウィンドウ間で異なる照明条件の画像系列から部分画像が選択されやすくなる. そのため, 局所的に周囲と色が異なる部分が発生し, 生成される画像系列の画質が劣化する. そこで, サブウィンドウ間の不連続度に関する項  $g_{xyt}(n)$  を加えることにより, 隣接するサブウィンドウ間での照明条件の関係を保存しつつ, 移動物体を含まない部分画像を選択する.

図 8 のように, 周辺サブウィンドウ (Sub-window of neighbor) で画像系列  $n_a$  が選択されている場合, 注目サブウィンドウ (Sub-window of interest)  $(x, y, t)$  で選択される画像系列  $n$  を決定する場合を考える. 具体的には, 出力画像系列の周辺サブウィンドウと注目サブウィンドウの照明条件の関係を, 画像系列  $n_a$  のものと同じにする. そのためには, 注目サブウィンド

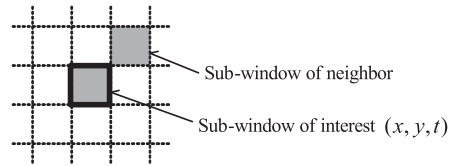


図8 不連続度の計算

Fig.8 Calculation of the discontinuity.

ウ  $(x, y, t)$  で選択される画像系列  $n$  の部分画像として, 画像系列  $n_a$  中の注目サブウィンドウ  $(x, y, t)$  の位置にある部分画像と照明条件が類似する部分画像を選択すればよい. したがって, 注目サブウィンドウ  $(x, y, t)$  とその周辺のサブウィンドウ間の不連続度  $g_{xyt}(n)$  は次式により計算する.

$$g_{xyt}(n) = \frac{1}{|A_{xyt}|} \sum_{n_a \in A_{xyt}} |\mathbf{v}_{xyt}(n) - \mathbf{v}_{xyt}(n_a)| \quad (5)$$

ここで,  $A_{xyt}$  はサブウィンドウ  $(x, y, t)$  の時間・空間方向を含めた 26 近傍で選択されている画像系列の集合を表し,  $|A_{xyt}|$  はその要素数を表す. また,  $\mathbf{v}_{xyt}(n)$  は注目サブウィンドウ  $(x, y, t)$  で選択される画像系列  $n$  の部分画像を表し,  $\mathbf{v}_{xyt}(n_a)$  はある周辺サブウィンドウで選択されている画像系列  $n_a$  の注目サブウィンドウ  $(x, y, t)$  と同じ位置の部分画像を表す.

### 2.4 モザイクキング

部分画像同士が重なり合う領域を  $\alpha$  ブレンディングすることで, 境界部分が目立ちにくい合成画像を生成する. 大量の映像を利用すれば, 類似する照明条件の画像が多く存在すると考えられる. そのため, 単純な  $\alpha$  ブレンディングでもモザイクキングが十分可能である. ここで, 部分画像の縁の近くは透明度が高く, 縁から遠いほど透明度が低くなるように,  $\alpha$  ブレンディングの係数を設定する.  $\alpha$  ブレンディングの係数に応じた透明度で, 部分画像同士を貼り合わせることで, 一つの画像系列を生成する.

## 3. 実験及び考察

### 3.1 実験条件

提案手法の有効性を確かめるために, 実際に市街地で撮影した映像を用いて次の二つの実験を行った.

**実験 1:** 移動物体除去性能の評価

**実験 2:** レジストレーションの精度評価

実験 1 では, 従来の画素成分単位のメジアンを用いた



図9 実験に用いるデータ。様々な照明条件の映像を含む。  
Fig. 9 Dataset for the experiments. It contains a variety of light conditions.

背景選択手法と提案手法を比較する。また、使用する映像数と移動物体の移動物体除去性能の関係を調査し、その結果について考察する。実験2では、DPマッチングと非剛体レジストレーションの精度評価を行う。

全方位カメラとして Point Grey Research 社製 Ladybug2 を使用し、フレームレート 15 fps で撮影した。実験のために合計 15 本の映像を収集した。その中の 1 本を基準映像として用い、残りの 14 本をソース映像として用いた。映像の長さは約 2 分で、走行距離は約 1.0 km である。映像は全て異なる日時に撮影されており、様々な照明条件の映像が含まれる。図 9 に実験に用いた画像の例を示す。実験には、球面投影によりパノラマ展開した映像を用いた。解像度は視野角 360 deg 当り 1,024 画素であった。

式 (2) の最小化問題について、逐次的なアルゴリズムにより近似解を求める。具体的には、ランダムに一つのサブウィンドウを選択し、目的関数の値が減少するように総当りで  $n_{xyt}$  を変化させる。この操作は、次式で表される。

$$n_{xyt}^{new} \leftarrow \arg \min_{n_{xyt}} [(1 - \lambda)f_{xyt}(n_{xyt}) + \lambda g_{xyt}(n_{xyt})] \quad (6)$$

いずれのサブウィンドウを選択しても  $n_{xyt}$  が変化しなくなった時点で、逐次処理を終了する。

### 3.2 実験1：移動物体除去性能の評価

#### 3.2.1 実験方法

従来の画素成分ごとにメジアンを用いた背景選択手法と提案手法中の背景画像選択手法を比較した。また、ソース映像として使用する映像の数を変化させ、使用する映像数と移動物体の除去性能の関係を調査した。

実験手順として、まず  $N$  本の入力画像系列に対し、提案手法によるレジストレーション処理を行った。そして、提案手法と比較手法により、それぞれ背景画像選択処理を行った。比較手法として用いるスカラ値に

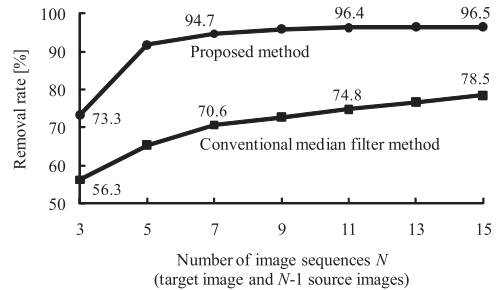


図10 使用した映像数  $N$  と移動物体除去成功率の関係  
Fig. 10 Relationship between the number of image sequences  $N$  and the removal rate.

対するメジアンは、値を昇順にソートした中央値とし、画素のそれぞれの RGB 成分に対して独立に処理を行った。

提案手法における式 (2) の重み  $\lambda$  は予備実験により 0.6 とした。評価方法として、基準映像中で移動物体が存在する領域の画素のうち、除去映像中に残った画素の数を比較した。具体的には、まず人手で移動物体の画素をマーキングした基準映像・ソース映像を用意した。そして、それぞれの画素の位置において、提案手法で選択される映像と同じマーキング映像から画素を取得することにより、除去映像中に残った画素を計数した。本実験では、ランダムに生成した 50 の組合せについて実験を行い、その平均を評価に用いた。ただし、ソース映像を 14 本用いた場合では一つの組合せのみを試行した。評価には、映像中から無作為に選んだ車両が映っている 11 地点を用いた。この実験における部分画像選択処理の際のサブウィンドウのサイズは  $30 \times 30$  画素とした。

#### 3.2.2 実験結果と考察

提案手法と比較手法について、用いた映像数（基準映像数＋ソース映像数）に対する除去成功率を図 10 に示す。提案手法では、映像数が 15 のとき 96.5% の移動物体領域の除去に成功した。一方で、比較手法では 78.5% であった。提案手法では、残った移動物体領域の多くが遠くの小さなものや、移動物体の断片であり、得られた移動物体除去映像は目的とするドライバ支援のためのデータベースやプライバシー保護に十分利用可能であると考えられる。

図 10 から、映像数が多いほど移動物体除去性能が向上することが確認できる。提案手法では、最低 3 本の映像があればベクトルメジアンフィルタをもとにした部分画像選択は可能であるが、9 割以上の移動物体



(a) 基準映像 (入力)



(b) 15 本の映像の統合結果 (提案手法)

図 11 移動物体除去の結果の例

Fig. 11 Example of the removal of temporary objects.

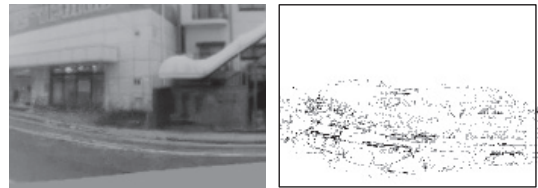
除去の性能を得るためには、5 本以上の映像が必要であった。移動物体除去の例を図 11 に示す。(a) が入力となる基準映像で、(b) が提案手法による 15 本の映像の統合結果である。

提案手法と比較手法による生成画像の一部分を拡大した画像と、除去できなかった移動物体の画素を黒色で表した画像の例を図 12 に示す。提案手法は良好に移動物体が除去できることが分かる。一方、比較手法では、複数の入力画像系列に存在する移動物体の画素が粒子状に多く残留した。これにより、画素成分単位ではなく領域単位で処理することにより、除去性能が向上することが確認できる。

式 (2) の重み  $\lambda$  が 0.0 と 0.6 の場合において生成された画像の一部分を拡大したものを、それぞれ図 13 (a), (b) に示す。使用した映像数は 9 本である。図 13 の建築物の白い壁に注目すると、本来は一樣な色の壁であるにもかかわらず、(a) では局所的に周囲と色が異なる部分が見られる。これは、式 (2) のサブウィンドウ間の不連続度  $g_{xyt}(n)$  の比重が小さい場合には、照明条件の異なる画像系列から部分画像が選択されやすくなるためである。一方、適切な重み  $\lambda$  を設定した場合には、(b) のように、改善されている。このように、サブウィンドウ間の連続性を考慮することにより、生成画像系列の画質が向上することが確認できる。また、重み  $\lambda$  が 0.0 のときの移動物体除去率は 96.7 %, 0.6



(a) 提案手法 (領域単位)

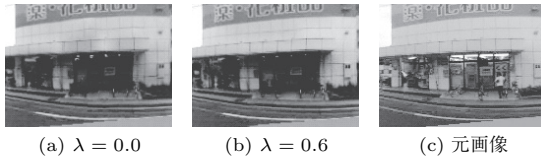


(b) 比較手法 (画素成分単位)

図 12 生成された画像と残留した移動物体の画素。画素成分単位ではなく、領域単位で処理することにより、除去性能が向上する。

Fig. 12 Result images and the residual pixels of the temporary objects. Removal ability is improved by processing an area instead of a pixel element.

のときは 95.9 %であった。これは、不連続度  $g_{xyt}(n)$  を導入した場合には、サブウィンドウ中に移動物体が存在する場合でも、元の入力映像の連続性を保存しようとするため、移動物体が存在する部分画像も選択されやすくなるためである。しかしながら、サブウィンドウ間の不連続度  $g_{xyt}(n)$  を導入した場合でも、移動



(a)  $\lambda = 0.0$  (b)  $\lambda = 0.6$  (c) 元画像

図 13 部分画像選択の重み  $\lambda$  による生成画像の違い  
Fig. 13 Difference in the result image by the weight  $\lambda$  for the partial image selection.

物体除去率の低下が十分に小さいことを確認した。また、各映像で建物の形状は変化しないが、上空は雲の有無や配置により大きく見た目が変化する。そのため、サブウィンドウ間の不連続度  $g_{xyt}(n)$  を導入しない場合では、上空では部分画像の境界が目立つ画像が生成される。一方、不連続度  $g_{xyt}(n)$  を導入した場合は、元の空のテクスチャの連続性を保持しようとするため、部分画像の境界の見た目は滑らかになる。

実験に使用した映像は、それぞれ撮影日時や撮影環境が異なる。利用するソース映像の違いの影響を調べるために、ソース映像を様々に変化させて移動物体除去率の標準偏差を調べた。その結果、映像数が 7 本で 2.0% (図 10 より平均は 94.7%)、映像数が 9 本で 1.1% (平均は 95.9%) であった。利用する映像数が多いほど、ソース映像の違いによる移動物体除去率のばらつきが小さくなり、移動物体の除去性能が安定することが確認できる。

また、選択する基準映像の違いによる影響を調べるために、基準映像として昼 (午前 11 時) に撮影した映像を選択した場合と、夕方 (午後 4 時) に撮影した映像を選択した場合で実験を行った。使用した映像数が 9 本のとき、基準映像が夕方の場合には移動物体除去率は 95.9% であり、昼の場合には 93.5% であった。このように、選択する基準映像により移動物体除去率が異なった。ただし、それぞれの実験で使用した映像群は同じである。そのため、このときの移動物体除去性能の違いは、映像間のレジストレーションの精度によるものと考えられる。よって、基準映像の選び方の指針として、レジストレーションの精度が向上するように選択することが考えられる。例えば、映像群中で色合いや明るさが中間的な映像を選択する方法が挙げられる。

提案手法は、逐次的に部分画像を選択することにより、移動物体除去のための目的関数の最小化を行っている。他に目的関数最小化の方法として、冒頭で紹介したグラフカットを利用する手法 [12] がある。この手

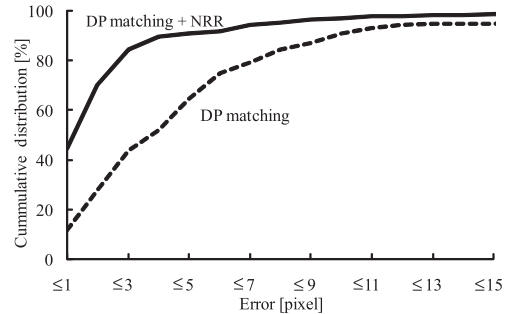


図 14 レジストレーション誤差の累積分布. DP マッチングに加え、非剛体レジストレーション (NRR) も用いることで、レジストレーション精度の大きな向上が見られた。

Fig. 14 Cumulative distribution of the registration errors. The registration accuracy was improved by applying non-rigid registration (NRR) in addition to DP matching.

法の生成される画像が連続となるように移動物体が映っていない画素を選択するという方針は、本論文の手法と共通する。そのため、今後両者の利点を活かした手法の検討を行いたい。

CPU Core i7 870 2.93 GHz, メモリ 8 GB の計算環境で、1 対の画像系列の DP マッチングに要する計算時間は平均 6.2 分、1 対の画像の非剛体レジストレーションに要する計算時間は平均 4.7 分、使用する映像が 9 本の場合の部分画像選択処理に要する計算時間は約 6 分であった。

### 3.3 実験 2: レジストレーションの精度評価

#### 3.3.1 実験方法

映像間のレジストレーションとして、DP マッチングのみを用いた場合と、DP マッチングに加え、非剛体レジストレーションも併用した場合について、画像の位置合せ誤差を比較した。手順として、まず基準画像とソース画像中の建物の角や窓枠等の特徴的な点を人手で対応づけた。そして、基準画像とソース画像の対応する特徴点間の画像上の距離を計算し、それをレジストレーションの誤差とした。1 枚の画像につき約 10 箇所の特徴点を選択し、15 地点の画像対について計測を行い、その誤差の平均を評価した。

#### 3.3.2 実験結果と考察

誤差の累積分布を図 14 に示す。誤差が 4 画素以下の割合が DP マッチングのみの場合は 52% であったのに対し、非剛体レジストレーションも適用した場合は 89% であった。DP マッチングに加え、非剛体レジストレーションも用いることで、レジストレーション

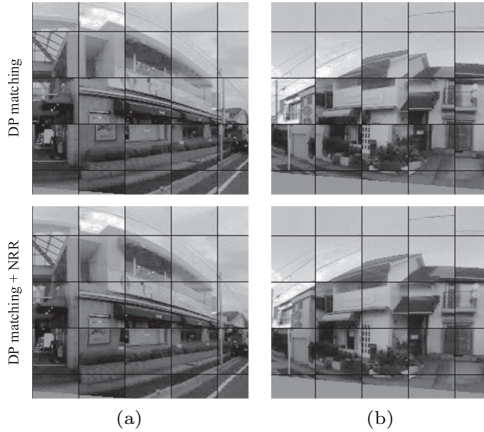


図 15 映像間のレジストレーションの適用例。上段に DP マッチングの適用結果を示す。上段の画像に非剛体レジストレーション (NRR) を適用した結果を下段に示す。ソース画像と基準画像をチェッカボード状に並べて表示してある。

Fig. 15 Example of the registrations between image sequences. The results of DP matching and non-rigid registration (NRR) are shown in top and bottom rows respectively. The source and the target images are shown as an checkerboard.

精度の向上が見られた。

映像間のレジストレーションの適用例を図 15 (a), (b) に示す。上段が DP マッチングのみを適用した場合の結果であり、下段が非剛体レジストレーションも用いた場合の結果である。基準画像とソース画像をチェッカボード状に並べて表示している。例えば、(b) の上段の画像では、建築物の屋根と空の境界がずれているが、非剛体レジストレーションにより、下段の画像のように補正することが可能となる。

映像中に移動物体が大きく映るフレームが存在する場合でも、DP マッチングへの影響は確認できなかった。これは、全方位カメラを用いているため、移動物体が視野の大部分を遮へいすることは少ないからと考えられる。また、実験に使用した映像の長さはそれぞれ異なるが、映像の長さのマッチングへの影響は見られなかった。また、提案手法では、DP マッチングのパターン要素間の距離として、単純に画像の画素値の二乗誤差を用いたため、車線内の走行位置が異なることにより、画像系列間の対応付け精度が低下することが懸念される。しかし、実験に用いた映像ではそのような問題は見られなかった。異なる車線を走行する状況も想定する場合には、車線方向の位置の違いによる

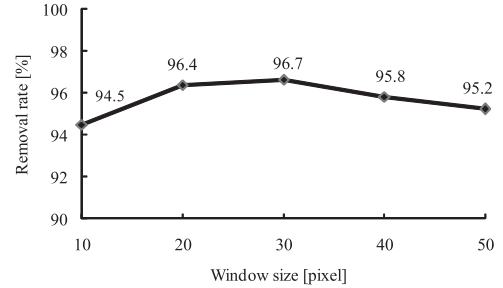


図 16 サブウィンドウの大きさと移動物体除去成功率の関係

Fig. 16 Relationship between the window size and the removal rate.

対応付け精度の低下が懸念されることから、DP マッチングのパターン要素間の距離について、更なる検討が必要である。

### 3.4 実用化へ向けた検討

#### 3.4.1 サブウィンドウの大きさについての検討

サブウィンドウの大きさの違いによる移動物体除去率について調査した。映像は実験 1 と同じものを用い、重みは  $\lambda = 0$  とした。サブウィンドウの大きさと移動物体除去率の関係を図 16 に示す。図 16 のように、サブウィンドウの大きさによって移動物体除去率が異なるという結果が得られた。また、サブウィンドウの大きさより移動物体が小さく映る場合、除去されない例が見られた。そのため、最適なサブウィンドウの大きさの決め方に関する指針として、除去対象とする移動物体の大きさに合わせる方法が考えられる。

#### 3.4.2 本手法の対象とする範囲

ドライバ支援のデータベースといった用途では、時間的に変化しない物体のみから構成された市街地画像が必要となる。また、プライバシー保護の目的では、人物の顔やナンバープレートが映っていない映像が必要である。3.2.2 の実験結果では 96.5% の移動物体を除去できており、また図 11 のように、目的とする移動物体除去がほぼ達成できた。しかしながら、除去に失敗した例が一部で見られたため、その考察を以下で行う。

提案手法で移動物体除去に失敗した例を図 17 に示す。図 17 (a) は常に駐車場の同じ位置に同じ車両が存在したため、除去に失敗した例である。ドライバ支援のデータベースといった用途では、時間的に変化しない物体のみから構成された市街地画像が必要となるため、このような駐車場の車両は問題とならない。しか



(a) 駐車場

(b) 交通量

図 17 移動物体除去の失敗例

Fig. 17 Example of failures.

しながら、プライバシー保護の観点では問題となるため、その場合には、他の手法を併用する必要がある。

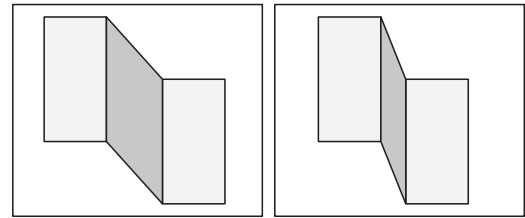
図 17 (b) は、交通量が多い地点であり、大部分の入力映像中に移動物体が映っていたため、移動物体除去に失敗した例である。本実験の映像には見られなかったが、交通量によっては、すべての入力映像中で前後方向に他車両が映る場合も考えられる。このような問題を防ぐには、交通量が少ない時間帯の映像を利用するなどの対策が必要となる。

また、遠景の移動物体では、サブウィンドウの大きさに比べて移動物体が小さく、除去されない例が見られた。しかしながら、移動物体が小さい場合には、ドライバ支援のデータベースやプライバシー保護といった応用では問題にならないと考える。

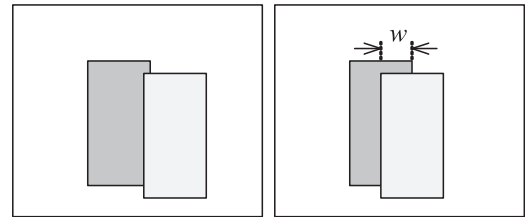
### 3.4.3 非剛体レジストレーションの妥当性

提案手法では、非剛体レジストレーションにより、複数の車載カメラ映像から、同一地点で撮影したものと見た目が類似する画像群を生成した。従来から、異なる視点で撮影した三次元物体の画像間を非剛体変形により対応づけることは広く行われている [16]。以降で、非剛体レジストレーションを市街地画像に適用することの妥当性を考察する。

市街地の建造物の多くは、壁面と道路面で構成されており、特に壁面は道路方向に平行なものと垂直なものが多い。図 18 (a) のように 2 視点間で観測可能な壁面領域が共通している場合には、非剛体変形に対応可能である。一方、図 18 (b) のように手前の物体が奥の物体を遮っている場合は対応付けが難しい。しかしながら、車線方向への視点位置の変化が十分小さい場合や、車両の速度に対して映像のフレームレートが十分高い場合は、遮へい部分の影響は無視できると考えられる。車線方向の位置の違いがない場合では、車速が 40 km/h、フレームレートが 15 fps、カメラと手前の物体との距離が 3 m、奥の物体との距離が 5 m



(a) 遮へいがない場合



(b) 遮へいがある場合

図 18 非剛体レジストレーションを市街地映像に適用することの妥当性。(a) の場合には、非剛体変形に対応可能である。(b) の場合には、視点の位置の違いが小さければ、遮へいは無視できる。

Fig. 18 Appropriateness for applying non-rigid registration to street-view images. (a) Free-form deformation is sufficient for matching. (b) An occlusion is ignorable in the case that the difference of the view point is small.

と仮定すると、進行方向の視点間の距離は最大 0.37 m となり、遮へい領域の幅  $w$  は最大 2.8 度の視野角に相当すると見積もることができる。この場合、遮へい領域の幅は十分に小さく、非剛体変形でも十分対応できる。特に、奥行の違いが小さい場合や、奥行の違いが大きくても、これらの物体が遠方にある場合には遮へい領域の幅が小さくなる。一方、奥行の違いが極端に大きい場合には、遮へい領域の幅が増加するため、非剛体変形では対応が難しくなる。本実験で、電柱と背後の建物でこのような事例が見られることがあったため、今後それに対応した手法の開発が必要である。

カメラの設置位置の違いや車線方向の位置の違いについても、進行方向の視点位置の違いと同様である。そのため、カメラの設置位置がほぼ同じ、かつ同じ車線内を走行する限り、大部分の典型的な市街地の構造において、非剛体変形で十分対応可能であると考えられる。

### 3.4.4 複数映像の収集方法

使用する映像数が多いほど移動物体除去性能が高くなるという結果が得られた。しかしながら、映像数が多いほど、その収集コストは大きくなる。この問題の解決策として、一般車両や公共の交通機関、運送業者

等の車両にカメラを装着することで、低コストに映像を収集する方法が考えられる。収集する映像としては、**3.4.2** や **3.4.3** の議論から、交通量が少ない時間帯の映像や、車線等の走行条件が類似する映像が適すると考えられる。

使用する映像数を削減するという方向性も挙げられる。例えば、移動物体らしさの計算の際に、映像の連続性を生かし、注目サブウィンドウに対する時間方向・空間方向の周辺画像領域を使用する方法が考えられる。

#### 4. む す び

複数車載カメラ映像の部分画像選択に基づく、移動物体を含まない映像の生成手法を提案した。具体的には、まず同一経路を複数回走行して得た映像に対して、時間的・空間的なレジストレーションを行った。これにより、映像間の撮影位置の違いの問題の解決を図った。次に、サブウィンドウ単位で、移動物体が映っていない部分画像を目的関数最小化により選択し、張り合わせることで、全体として移動物体の映っていない全方位カメラ映像を生成した。その際、バクトルメジアンフィルタによる移動物体らしさに基づく基準を提案し、隣接するサブウィンドウ間で、入力画像系列の照明条件の関係が保存されるような拘束を加えた。15本の映像を用いた実験の結果、96.5%の移動物体の除去に成功した。使用する映像の本数が多いほど、除去性能が向上するという結果を得た。また、サブウィンドウ間の連続性を考慮することにより、画質の向上が見られた。

今後、同一経路の大量の映像を収集するために、カメラを搭載した複数の車両が協調するという枠組みが考えられる。その他の課題として、移動物体除去性能と画質の両立が可能な手法の開発や、移動物体が除去された映像を利用した実際の応用例の検討が挙げられる。

**謝辞** 日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に深く感謝する。本研究の一部は、JST 戦略的創造研究推進事業 CREST 及び科学研究費補助金による。また、本研究では画像処理に MIST ライブラリ (<http://mist.murase.m.is.nagoya-u.ac.jp/>) を使用した。

#### 文 献

[1] 内山寛之, 出口大輔, 高橋友和, 井手一郎, 村瀬 洋, “拡張 DP マッチングを用いた視野角の異なるカメラ映像間の時空間対応付けによる自車位置推定,” 信学論 (D),

vol.J93-D, no.12, pp.2659–2665, Dec. 2010.

[2] M. Jabbour, V. Cherfaoui, and P. Bonnifait, “Management of landmarks in a GIS for an enhanced localisation in urban areas,” Proc. 2006 IEEE Intelligent Vehicles Symposium, pp.50–57, Sept. 2006.

[3] 久徳遙矢, 出口大輔, 高橋友和, 目加田慶人, 井手一郎, 村瀬 洋, “車載カメラ映像の道路面経時差分による前方障害物検出,” 画像の認識・理解シンポジウム (MIRU) 2009 論文集, IS1-55, pp.734–738, July 2009.

[4] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent, “Large-scale privacy protection in Google street view,” Proc. IEEE 12th Int. Conf. on Computer Vision, pp.2373–2380, Sept. 2009.

[5] N. Kawai, T. Sato, and N. Yokoya, “Image inpainting based on energy minimization,” Proc. SPIE Electronic Imaging, vol.6498, pp.649813–649813-9, Jan. 2007.

[6] A. Levin, A. Zomet, and Y. Weiss, “Learning how to inpaint from global image statistics,” Proc. IEEE 9th Int. Conf. on Computer Vision, pp.305–312, Oct. 2003.

[7] 栗林宏輔, 川崎 洋, 小野晋太郎, 池内克史, “移動カメラ映像の障害物除去のための時空間画像フィルタの提案,” 画像の認識・理解シンポジウム (MIRU) 2009 論文集, IS2-32, pp.1065–1072, July 2009.

[8] Y. Wexler, E. Shechtman, and M. Irani, “Space-time video completion,” Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, vol.1, pp.120–127, June 2004.

[9] A. Yamashita, I. Fukuchi, T. Kaneko, and K. Miura, “Removal of adherent noises from image sequences by spatio-temporal image processing,” Proc. 2008 IEEE Int. Conf. on Robotics and Automation, pp.2386–2391, May 2008.

[10] J. Böhm, “Multi-image fusion for occlusion-free façade texturing,” International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol.35, Part B5, pp.867–872, July 2004.

[11] 高橋英之, 堀磨伊也, 神原誠之, 横矢直和, “全天球画像データベース作成のための色調統一と動物体の除去,” 信学技報, MVE2009-40, Oct. 2009.

[12] 飯吉建彰, 三橋 渉, “画像モザイク生成のための部分的な前景の自動除去,” 信学論 (D), vol.J92-D, no.8, pp.1403–1413, Aug. 2009.

[13] 佐藤准嗣, 高橋友和, 井手一郎, 村瀬 洋, “GPS 座標付き全方位映像群からの市街地映像マップの構築と街並変化の検出,” 信学論 (D), vol.J90-D, no.4, pp.1085–1095, April 2007.

[14] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes, “Nonrigid registration using free-form deformations: Application to breast MR images,” IEEE Trans. Med. Imaging, vol.18, no.8, pp.712–721, Aug. 1999.

[15] J. Astola, P. Haavisto, and Y. Neuvo, “Vector me-

dian filters,” Proc. IEEE, vol.78, no.4, pp.678–689, April 1990.

- [16] Y. Yaguchi, K. Iseki, N.T. Viet, and R. Oka, “3D object reconstruction using full pixel matching,” Proc. 13th Int. Conf. on Computer Analysis of Images and Patterns, pp.873–880, Sept. 2009.

(平成 22 年 11 月 9 日受付, 23 年 4 月 28 日再受付)



内山 寛之 (学生員)

平 19 名工大・工・電気情報卒. 平 21 名大大学院情報科学研究科博士前期課程了. 修士 (情報科学). 現在, 同研究科博士後期課程在学中. 画像処理・パターン認識の分野に興味をもつ. 平 22 電気関係学会東海支部大会 IEEE 名古屋支部学生奨励賞受賞.



出口 大輔 (正員)

平 13 名大・工・情報卒. 平 18 同大大学院博士後期課程了. 平 16~18 日本学術振興会特別研究員. 平 18 名大大学院情報科学研究科研究員, 平 18 名大大学院工学研究科研究員, 平 20 より同大助教, 現在に至る. 主に画像処理・パターン認識技術の開発とその ITS 及び医用応用に関する研究に従事. 博士 (情報科学). 日本医用画像工学会, 日本コンピューター外科学会, 日本生体医工学会各会員.



高橋 友和 (正員)

平 9 茨城大・工・情報卒. 平 12 同大大学院理工学研究科博士前期課程了. 平 15 同研究科博士後期課程了. 博士 (工学). 同年より 2 年間名古屋大学大学院情報科学研究科 COE 研究員. 平 17 より 3 年間日本学術振興会特別研究員. 平 20 より岐阜聖徳学園大学経済情報学部准教授, 現在に至る. 画像認識の基礎研究並びにその応用に興味をもつ. 画像電子学会会員.



井手 一郎 (正員: シニア会員)

平 6 東大・工・電子卒. 平 8 同大大学院工学系研究科情報工学専攻修士課程了. 平 12 同研究科電気工学専攻博士課程了. 博士 (工学). 同年国立情報学研究所助手. 平 16 名古屋大学大学院情報科学研究科助教授, 平 19 より准教授. この間, 平 14~16 総合研究大学院大学数物科学研究科助手併任, 平 16~22 情報・システム研究機構国立情報学研究所客員助教授・准教授兼任, 平 17, 18, 19 フランス情報学・統計システム研究所 (IRISA) 招聘教授. 平 22~23 オランダアムステルダム大学情報学研究所上級訪問研究員. パターン認識技術の実応用や映像メディア処理全般に興味をもっている. 情報処理学会, 映像情報メディア学会, 人工知能学会, 言語処理学会, IEEE Computer Society, ACM 各会員.



村瀬 洋 (正員: フェロー)

昭 53 名大・工・電気卒. 昭 55 同大大学院修士課程了. 同年日本電信電話公社 (現 NTT) 入社. 平 4 から 1 年間米国コロンビア大客員研究員. 平 15 から名古屋大学大学院情報科学研究科教授, 現在に至る. 文字・図形認識, コンピュータビジョン, マルチメディア認識の研究に従事. 工博. 昭 60 本会学術奨励賞, 平 6 IEEE-CVPR 最優秀論文賞, 平 7 情報処理学会山下記念研究賞, 平 8 IEEE-ICRA 最優秀ビデオ賞, 平 13 高柳記念奨励賞, 平 13 本会ソサイエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞, 平 16 IEEE Trans. MM 論文賞, ほか受賞. IEEE フェロー, 情報処理学会会員.