

前方車載カメラ映像中の移動物体除去のための 基準画像選択に関する検討

小塚 亨[†] 出口 大輔^{††} 井手 一郎[†] 村瀬 洋[†]

[†]名古屋大学 大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町

^{††}名古屋大学 情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町

E-mail: [†]kotsukat@murase.m.is.nagoya-u.ac.jp, ^{††}ddeguchi@nagoya-u.jp, ^{†††}{ide,murase}@is.nagoya-u.ac.jp

あらまし 近年、市街地を撮影した画像とその位置情報を対応付けた市街地画像データベースへの需要が高まっている。この市街地画像データベースを各種サービスで利用する場合、画像中に人や自動車等の移動物体が多数映り込む可能性が高く、プライバシーの観点から大きな問題となっている。このような問題の解決を目指し、本報告では同一経路から得られた複数の前方車載カメラ画像系列を統合することにより、画像中の移動物体を除去する手法を提案する。複数の画像系列を統合して利用するためには、ある画像を基準として他の画像を位置合わせする必要がある。このとき、一方の画像のみに存在する移動物体領域は、他方の画像と正しく位置合わせすることができない。そのため、移動物体領域の位置合わせは、その周囲の位置合わせ結果を用いて推定する必要がある。しかし、移動物体領域が非常に大きい場合は、この推定を適切に行えないため、大きな移動物体を含む画像を基準とした場合は、移動物体除去後の画質が大きく劣化する。この問題に対し、提案手法では位置合わせの際の対応付けの正確さを利用した基準画像選択を導入し、更に背景らしい領域のみを候補として統合を行う。実験の結果、従来手法と同等の移動物体除去の性能を実現しつつ、画質劣化の軽減も可能であることを確認した。

キーワード 車載カメラ, 移動物体除去, 基準画像選択

A study on the reference image selection for the removal of moving objects from frontal in-vehicle camera videos

Toru KOTSUKA[†], Daisuke DEGUCHI^{††}, Ichiro IDE[†], and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

^{††} Information and Communications Headquarters, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

E-mail: [†]kotsukat@murase.m.is.nagoya-u.ac.jp, ^{††}ddeguchi@nagoya-u.jp, ^{†††}{ide,murase}@is.nagoya-u.ac.jp

Abstract In recent years, the demand for a street-view image database combining street-view images and their location information is increasing. When using such a street-view image database for various applications, privacy violation is one of the biggest problems because street-view images may contain moving objects, such as pedestrian and vehicle. To solve this problem, this paper proposes a method for moving objects removal from a street-view image by integration of frontal in-vehicle camera videos captured along the same route. Although image registration plays an important role to integrate such videos, it is difficult to align images correctly if reference image contains moving objects. Therefore, alignment of moving object regions needs to be done by combining it with the image registration of it surrounding regions. However, if the moving object region is large, appropriate alignment is difficult. To overcome this problem, we propose a method of reference image selection using the accuracy of the registration and the extraction of scenery candidate regions. Experimental results showed the proposed method could reduce deterioration of visual image quality in comparison with conventional methods.

Key words In-vehicle camera, removal of moving objects, reference image selection

1. まえがき

車載カメラ画像系列に位置情報を付与した市街地画像データベースが広く利用されている。例えば、近年人気を集めている Google ストリートビュー^(注1) のようなサービスでは、目的地周辺の風景を視覚的に確認する目的などで活用されている。また、走行中に撮影した画像をデータベースと比較することにより、自車位置の推定 [1] や街並み変化の検出 [2]、そして障害物の検出 [3] などへの応用も可能である。しかし、実環境で撮影した市街地画像には人や自動車等の移動物体が多数映り込む可能性が高い。これは、プライバシー保護の観点から大きな問題となっている。また、これらは風景を遮蔽するため、自車位置推定手法において必要な風景情報が得られないという問題がある。このような問題に対し、画像中に存在する移動物体を除去し、背景のみの画像を合成する手法が提案されている。

画像中の前景領域を除去し、背景を復元する手法はインペインティングと呼ばれており、1 枚の画像のみを入力とする手法と複数枚の画像を入力とする手法に分けることができる。1 枚の画像のみを入力とする手法として、Bertalmio ら [4] は人手で選択した対象領域を、その縁から順に周辺の対象領域外の画素を用いて補完し、これを再帰的に行うことで背景を復元する手法を提案している。しかし、人手で対象領域を指定する必要があるため、膨大なデータベースの画像に対して適用することは困難である。一方、Flores ら [5] は除去対象を歩行者に限定し、人検出手法を用いることで除去対象領域を自動で選択する手法を提案している。しかし、人検出の精度が不十分な場合は、正しく背景画像を復元することはできない。また、車や自転車等の不特定の移動物体を除去することはできない。

複数枚の画像を入力とする手法として、Kawai ら [6] は 1 本の車載カメラ画像系列から 3 次元復元を行い、各画像を画素単位で対応付けし、エネルギー最小化によって移動物体を除去する手法を提案している。この手法は、前述の手法と異なり、人を含む不特定の移動物体全般を自動的に除去し、背景を復元することが可能である。しかし、駐車している自動車など、撮影時に一時的に静止している移動物体は背景と判断されるため、これらを除去することはできない。

上述のような問題を解決するため、内山ら [7] は同一経路から得られた複数の全方位車載カメラ画像系列を統合することにより、画像中の移動物体を除去する手法を提案している。しかし、一般車両に全方位カメラを搭載することは難しい。そこで我々はさらに、一般車両への搭載が容易な前方車載カメラ画像系列に適用可能な移動物体除去手法 [8] を提案した。これらの手法は、複数の画像系列を統合するための前処理として、画像毎の系列対応付けを行う。具体的には、まず全ての系列間での画像同士の対応付けを行う。次に、対応付けられた画像群に対して画素単位の位置合わせを行う。その後、対応付けられた各画像群を統合することにより、移動物体を除去する。このとき、図 1 に示すように、一方の画像のみに存在する移動物体は、他



(a) 基準画像



(b) 入力画像



(c) 画像 (a) を基準として画像 (b) を位置合わせした結果

図 1 基準画像の違いによる位置合わせ結果の差異。基準画像に移動物体が存在するとき、位置合わせができない画素が生じる。一方、基準画像に移動物体が存在しないときは、正しく位置合わせをすることができる

方の画像と位置合わせすることができない。そのため、移動物体領域の位置合わせは、その周囲の位置合わせ結果を用いて推定する必要がある。しかし、移動物体領域が非常に大きい場合は、この推定が適切に行えないため、大きな移動物体を含む画像を基準とした場合は、移動物体除去後の画質が大きく劣化する。逆に、大きな移動物体を含まない画像を位置合わせの基準とすれば、この問題を回避することができる。そこで我々は、移動物体除去画像への寄与度に基づく基準画像選択手法 [9] を提案した。これは、画像群の各画像とその画像を基準として位置合わせした他の画像を組として用意し、組毎に独立に統合を行う。その際に最終的な統合画像に使用された画素が最も多い画像が最も移動物体が少ないと考え、この画像を位置合わせの基準とする。これにより、移動物体除去画像に残存する移動物体の画素数が減少し、かつ画質劣化の抑制効果が得られることを確認した。しかし、各画像にそれぞれ多数の移動物体が存在する場合、適切な基準画像を選択することができないという問題があった。そこで本報告では、位置合わせの際の対応付けの正確さを利用した基準画像選択を導入する。また、統合の際に背景らしい領域のみを候補とすることで、より移動物体を含まない画像の合成を目指す。これにより、従来手法より画質劣化の少ない移動物体除去手法を目指す。

以降、2. で移動物体除去手法の詳細を述べ、3. で提案手法の有効性を評価する実験について述べる。4. に実験の考察を述べ、最後に 5. でまとめと今後の課題について述べる。

2. 移動物体除去手法

本節では、同一経路から得られた複数の前方車載カメラ画像系列の統合による移動物体除去手法について述べる。まず、移動物体の定義を述べ、次に基準画像の選択方法の概略を述べる。

(注1) : <http://www.google.co.jp/help/maps/streetview/>



図 2 移動物体の例

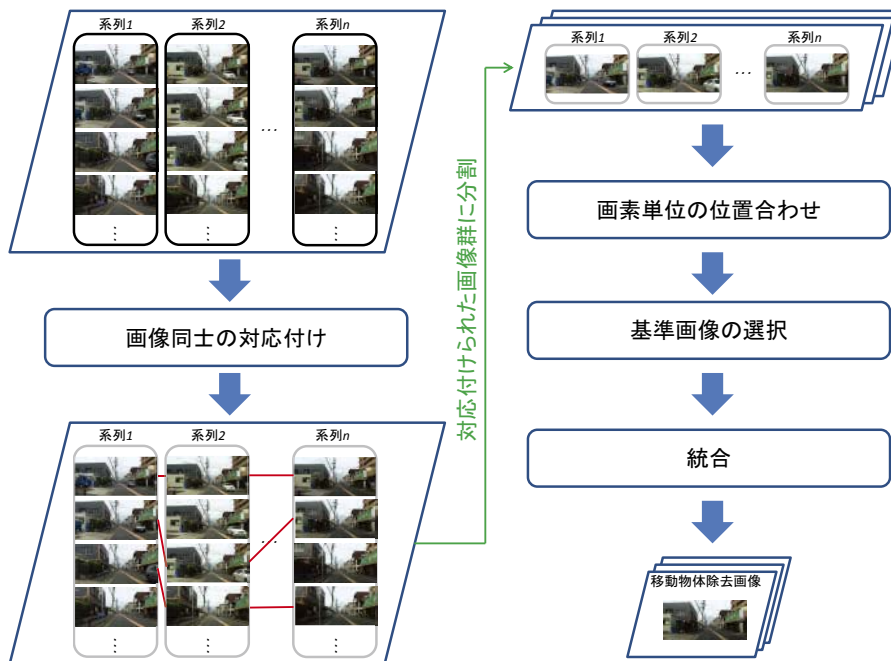


図 3 提案手法の処理の流れ

その後、提案手法の概略と、各手順について詳細に説明する。

2.1 移動物体の定義

本研究で扱う移動物体は、同一地点を複数回撮影した際に恒常的に存在しない物体である。例えば、図 2 のように特定の撮影時のみに存在する人 (画像 (a), 画像 (b)) や車両 (画像 (c), 画像 (d)) を移動物体とする。また、駐車車両 (画像 (e), 画像 (f)) も同様である。一方、複数回の撮影において、常に同じ位置に駐車されている車両は移動物体ではない。

2.2 基準画像選択方針

上述の通り、対応付けられた画像群を位置合わせする際、最も移動物体が少ない画像を基準とすべきである。そのような画像を自動選択する基本方針について述べる。基準画像を選択する問題を、ある画像集合 $I = \{1, 2, \dots, n\}$ に対して以下の目的関数 $G(i)$ が最大となる画像 i を求める処理と考える。

$$G(i) = \sum_{j=1}^n \sum_{k=1, k \neq i}^n \sum_{\mathbf{x}} g(\mathbf{x}, i, j) \quad (1)$$

$$g(\mathbf{x}, i, j) = \begin{cases} 1 & \text{if } \|v_{i \rightarrow j}(\mathbf{x}) + v_{j \rightarrow i}(\mathbf{x} + v_{i \rightarrow j}(\mathbf{x}))\| \leq l \\ 0 & \text{otherwise} \end{cases}$$

ただし、 \mathbf{x} は画像中の各画素、 $v_{i \rightarrow j}(\mathbf{x})$ は画像 i 上のある画素 \mathbf{x} に対応づいた画像 j の画素への変位量とする。つまり、2 枚の画像を位置合わせした際の画素の位置ずれが l 以下であるものを正しく位置合わせできたと判断し、その画素数を計算する。そして、その数が最も多い画像を、その画像集合において最も

移動物体を含まない画像と判断し、それを位置合わせの基準画像として選択する。なお、以下の実験では $l = 4$ とした。

また、同一地点を複数回観察すると、その多くは同一の背景が得られる可能性が高い。そのため、複数の画像にわたって正しく位置合わせができた画素は背景であると考えられる。よって、そのような画素のみを用いて背景の復元を行い、移動物体を除去する。

2.3 移動物体除去画像合成手順

ある移動物体に注目して考えると、同じ移動物体が異なる時刻において同一地点に存在することは稀である。つまり、一般に、異なる時刻に同一地点を複数回撮影した場合、ある小領域について見ると背景のみが映る確率が高い。この仮定に基づき、異なる時刻に撮影された複数本の画像系列から対応付けられた画像群毎に基準画像を選択し、それによって位置合わせされた画像を統合することで移動物体の除去を行う。図 3 に提案手法の処理の流れを示す。提案手法は大きく分けて 3 つの処理に分かれる。まず、全ての系列間での画像同士の対応付けを行う。次に、対応付けられた画像群での画素単位の位置合わせを行う。このときに各画像群における基準画像の選択を行う。最後に、対応付けられた各画像群を統合することにより、移動物体を除去する。以下、各処理について詳述する。

2.3.1 系列間での画像同士の対応付け

エピポーラ幾何に基づく画像間距離を用いた DP マッチング [10] により、系列間での画像同士の対応付けを行う。具体的

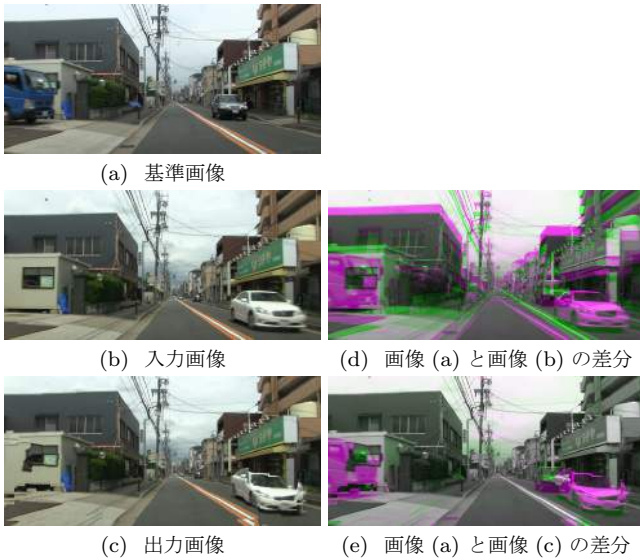


図4 SIFT flow による位置合わせ結果の例. 画像 (a) を基準に画像 (b) を位置合わせしたものが画像 (c). また, 画像 (a) と (b), 画像 (a) と (c) それぞれの差分を明示的に表したものがそれぞれ画像 (d), 画像 (e) である

には, 各画像系列中の全画像において基準系列と入力系列の画像対からエピポールをそれぞれ算出する. このとき, 撮影位置が近いほど, 画像中心・エピポール間距離が大きくなる, という性質がある. これを利用して, 画像中心・エピポール間距離を尺度とし, DP マッチングによって系列間での画像同士の対応付けを行う.

2.3.2 画像群での画素単位の位置合わせ

2.3.1 により対応付けられた画像群は, 撮影位置の微妙な違いにより見えが異なる. これを修正するために, SIFT flow [11] を用いた画素単位の位置合わせを行う. SIFT flow は, 各画素に対して SIFT 特徴量を算出し, 確率伝播法を用いて全画素の対応付けを行う. 輝度値そのものではなく SIFT 特徴量を用いるため, SIFT と同様に回転, スケール変化に対して不変であり, また照明変化に頑健であるため, 従来の Optical flow よりも正確に位置合わせを行うことが可能であることが知られている. 図4に位置合わせ結果の例を示す. 画像 (d) と画像 (e) を比較することにより, SIFT flow による高精度な位置合わせが可能であることがわかる.

2.3.3 画像群における基準画像選択

画像群に含まれる全ての画像を基準として 2.3.2 の操作を行う. その際, 2.2 で説明した処理を行うことにより, 各画像群における基準画像を選択する.

2.3.4 画像群の統合

2.3.2, 2.3.3 によって画素単位で位置合わせされた画像群を用いて, 移動物体の除去を行う. 2.3 の仮定により, 同一地点・同一領域に対応するベクトル群に対して, 多数決の原理により移動物体の除去が可能である. 具体的には, まず各画像を W 個のパッチに分割し, それぞれパッチ内の画素値を並べたベクトルに変換する. 次に, 以下の目的関数 $F(\mathbf{n})$ を最小化するパッチを選択する.

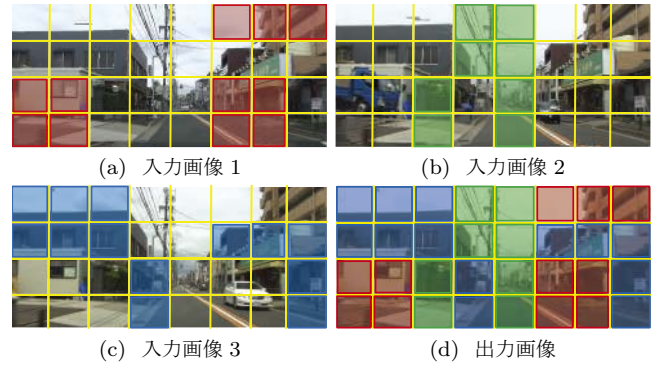


図5 背景領域の選択と合成. 画像 (a), 画像 (b), 画像 (c) を入力として, 画像 (d) を出力する

$$F(\mathbf{n}) = \sum_{w=1}^W [(1-\lambda)f_w(n_w) + \lambda g_w(n_w)] \quad (2)$$

ここで, n_w は w 番目のパッチである. また, f_w は移動物体らしさに関するペナルティ項であり, その計算にはベクトルメディアンフィルタ [12] を用いる. ベクトルメディアンフィルタは, 入力とする複数ベクトルのうち中央となるベクトルを出力するフィルタである. ここで, 中央ベクトルは他のベクトルとの距離の和が最小になるベクトルとして定義される. そこで, f_w は中央ベクトルと他のベクトルとの距離の和が出力される. 仮定より, 異なる時刻に同一地点を複数回撮影した場合, ある小領域について見ると背景のみが映る確率が高いため, ベクトルメディアンフィルタにより出力される中央ベクトルは, パッチの多数決結果に対応する. また, ベクトルメディアンフィルタの入力には 2.2 の処理で得られた正しく位置合わせできている画素を中心としたパッチのみとすることで, 移動物体除去画像の画質劣化を抑制する. しかし, ベクトルメディアンフィルタによる各パッチの選択は独立に行われるため, 近傍パッチの不連続度に関するペナルティ項 g_w を導入することで, 照明条件を保存する. 詳しくは文献 [7] を参照されたい. また, λ は f_w と g_w に対する重みであり, 以下の実験では $\lambda = 0.6$ とした.

図5に各入力画像から $F(n)$ の式により出力に使われるパッチを選択して統合する様子を示す. なお, 実際は各パッチは互いが重複するように配置し, α ブレンディングを用いて統合することで, 移動物体を除去した画像を合成する.

3. 実験

提案手法の有効性を確認するために, 実験を行った. まず 3.1 で基準画像の選択精度の評価実験について述べ, 次に 3.2 で移動物体除去の精度評価実験について述べる.

入力として同一経路を異なる時刻に走行して撮影した前方車載カメラ画像系列 5 本を用いた. 解像度は 720×340 pixels とした. フレームレートは 23.98fps であり, 約 40 秒間撮影した. また, 撮影経路は直線状で, 移動物体が疎らに映る市街地であった. 各画像系列に含まれる画像枚数はそれぞれ 1,000 枚であり, そのうち 25 枚毎に選んだ 40 枚を精度評価の対象とした.

比較手法 1 は基準画像の選択を行わない手法である. つまり, 画像系列の画像同士の対応付けをした際の基準画像系列の各画

表 1 基準画像に含まれる移動物体の画素数の平均 (画素)

最適基準画像	提案手法	比較手法 1	比較手法 2
693	2,106	4,906	3,868

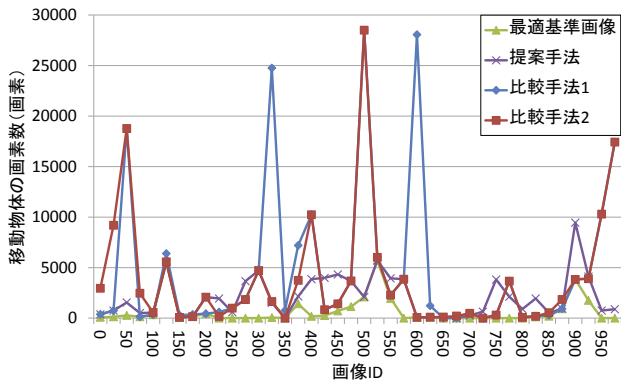


図 6 各画像群における基準画像に含まれる移動物体の画素数の遷移

像を、画像群での画素単位の位置合わせの基準画像とした。比較手法 2 は文献 [9] の手法である。

3.1 基準画像選択の評価実験

提案手法によって選択された各画像群における基準画像のもっともらしさを実験により評価した。実験方法として、各画像群における基準画像に含まれる移動物体の画素数を調べた。正解データとして、各画像群における最適基準画像を用意した。ここで最適基準画像とは、人手で移動物体を入力した画像を用い、各画像群において最も移動物体の画素数が少ない画像である。

各画像群における基準画像に含まれる移動物体の画素数の平均を表 1 に示す。提案手法が比較手法に比べて最適基準画像に近い値を得られたことがわかる。次に、各画像群における基準画像に含まれる移動物体の画素数の時間的な遷移を図 6 に示す。提案手法では、比較手法に比べて最適基準画像に近い画像を選択できたことがわかる。

3.2 移動物体除去精度の評価実験

次に、提案手法による移動物体除去精度を評価した。具体的には、移動物体除去後に残存した移動物体の画素数を調べた。

移動物体除去画像に残存した移動物体の画素数の平均を表 2 に示す。次に、移動物体除去後に残存した移動物体の画素数の時間的な遷移を図 7 に示す。提案手法では、比較手法に比べて残存した移動物体の画素数が減少したことがわかる。

4. 考 察

4.1 基準画像選択の評価実験

図 6 より、比較手法 1 では各画像群における基準画像の選択を適応的に行わないため、移動物体を多く含む画像が基準画像として選択される場合がある。一方、比較手法 2 ではそのような選択をある程度修正できたことがわかる。しかし、画像 ID : 50 及び画像 ID : 500 では、基準画像の選択に失敗したことがわかる。画像 ID : 500 における各系列の画像を図 8 に示す。これは、それぞれの画像に多数の移動物体が存在したため、適

表 2 移動物体除去画像に残存した移動物体の画素数の平均 (画素)

提案手法	比較手法 1	比較手法 2
16	36	25

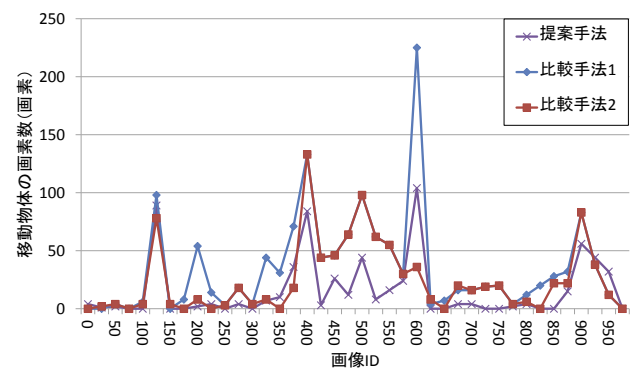


図 7 移動物体除去後に残存した移動物体の画素数の遷移



図 8 画像 ID : 500 における各系列の画像



(a) 提案手法 (b) 比較手法 2

図 9 移動物体除去後の画像例

切な基準画像の選択に失敗したからであると考えられる。具体的には、比較手法は系列 1 の画像を選択した。つまり、背景領域の選択で使用したベクトルメディアンフィルタを基準画像選択に利用することは、適切な基準画像を選択するという観点では不十分であると考えられる。それに対し、提案手法では系列 3 の画像を選択しており、このような場合でも、適切な基準画像の選択を行えていることがわかる。以上より、基準画像選択における提案手法の有効性を確認した。

4.2 移動物体除去精度の評価実験

比較手法及び提案手法のいずれについても、移動物体除去前に比べて移動物体除去後に残存した移動物体の画素数は大きく減少した。これは、背景領域の選択で使用したベクトルメディアンフィルタによる背景領域の選択が正しく行われたことを示



図 10 移動物体除去後の画像例

している。提案手法及び比較手法 2 により合成された移動物体除去後の画像例を図 9 に示す。比較手法 2 では除去に失敗して残存してしまった人が、提案手法では除去できたことがわかる。以上により、提案手法の有効性を確認した。

また、提案手法により合成された移動物体除去後の画像のその他の例を図 10 に示す。画像 ID : 275 では概ね良好に移動物体を除去できたことがわかる。一方、画像 ID : 400 では道路上の車両の除去に一部失敗した。これは、似た移動物体が異なる時刻において同一地点に存在したため、その領域が正しく位置合わせできず、統合の際にそのようなパッチのみをベクトルメディアンフィルタの入力としたためだと考えられる。

5. ま と め

本報告では、同一経路から得られた複数の前方車載カメラ画像系列を統合することにより、画像中の移動物体を除去する手法を提案した。その際、位置合わせの対応付けの正確さを利用した基準画像選択を導入し、更に統合の際に背景らしい領域のみを候補として統合を行った。実験により、提案手法の有効性を確認した。今後の課題として、交通量が多い市街地画像系列への本手法の適用が挙げられる。実験で移動物体の除去に失敗したものとして、いずれの入力画像においても同じ位置に自動車が存在し、背景に相当する領域が存在しない例が見られた。このような失敗は、交通量の多い市街地で本手法を適用した場合にも起こり得ると考えられる。そのため解決策として、入力画像系列を増やしたり、文献 [6] のような時系列情報を用いたりするなどが考えられる。またそのために、移動物体を多く含む画像に対して他の画像を位置合わせするときの工夫も必要である。

謝 辞

本研究の一部は、JST 戦略的創造研究推進事業 CREST, JST 名古屋大学 COI, 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.nagoya-u.ac.jp/>) を使用した。

文 献

- [1] H. Uchiyama, D. Deguchi, T. Takahashi, I. Ide, and H. Murase, "Ego-localization using Streetscape Image Sequences from In-vehicle Cameras," Proc. 2009 IEEE Intelligent Vehicles Symposium (IV2009), pp. 185–190, Jun. 2009.
- [2] J. Sato, T. Takahashi, I. Ide, and H. Murase, "Change detection in streetscapes from GPS coordinated omni-directional image sequences," Proc. 2006 International Conference on Pattern Recognition (ICPR2006), pp. 935–938, Aug. 2006.
- [3] H. Kyutoku, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase, "Subtraction-Based Forward Obstacle Detection Using Illumination Insensitive Feature for Driving-Support," Proc. ECCV2012 Workshop on Computer Vision in Vehicle Technology : From Earth to Mars, pp. 515–525, Oct. 2012.
- [4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image Inpainting," Proc. 27th Int. Conf. on Computer Graphics and Interactive Techniques, pp. 417–424, Jul. 2000.
- [5] A. Flores and S. Belongie, "Removing Pedestrians from Google Street View Images," Proc. Int. Workshop on Mobile Vision, pp. 53–58, Aug. 2010.
- [6] N. Kawai, N. Inoue, T. Sato, F. Okura, Y. Nakashima and N. Yokoya, "Background Estimation for a Single Omnidirectional Image Sequence Captured with a Moving Camera," IPSJ Trans. on Computer Vision and Applications, Vol. 6, pp. 68–72, Jul. 2014.
- [7] 内山寛之, 出口大輔, 高橋友和, 井手一郎, 村瀬洋, "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," 信学論 (D), Vol. J94-D, No. 12, pp. 2093–2104, Dec. 2011.
- [8] 小塚亨, 出口大輔, 井手一郎, 村瀬洋, "前方車載カメラ映像中の移動物体除去に関する予備的検討," 2014 年信学総大, D-12-56, pp. 131, Mar. 2014.
- [9] 小塚亨, 出口大輔, 井手一郎, 村瀬洋, "適応的な基準画像選択に基づく複数の前方車載カメラ映像を用いた移動物体の除去," 画像の認識・理解シンポジウム (MIRU) 2014, SS1-39, Jul. 2014.
- [10] 久徳遙矢, 出口大輔, 高橋友和, 目加田慶人, 井手一郎, 村瀬洋, "自転車位置推定のための車載カメラ映像と市街地映像データベースの位置ずれや遮へいに頑健なフレーム対応付け," 信学論 (D), Vol. J95-D, No. 11, pp. 1973–1982, Nov. 2012.
- [11] C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense Correspondence Across Scenes and its Applications," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 33, No. 5, pp. 978–994, May 2011.
- [12] J. Astola, P. Haavisto, and Y. Neuvo, "Vector Median Filters," Proc. IEEE, Vol. 28, No. 4, pp. 678–689, Apr. 1990.