

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

内山 寛之[†] 出口 大輔[†] 高橋 友和^{††} 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

^{††} 岐阜聖徳学園大学 経済情報学部 〒500-8288 岐阜県岐阜市中鶉 1-38

E-mail: †{uchiyama,ddeguchi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp

あらまし 近年、市街地映像を利用して自車位置を推定するなどのドライバ支援システムが着目されている。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、遮蔽やプライバシー等の点で問題となっている。そこで本報告では、複数映像統合により、車載全方位カメラから得られる市街地映像中から移動物体を除去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。次に、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。最後に、移動物体が映っていない部分画像を自動的に選択して、張り合わせることにより、移動物体の映ってい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、良好に移動物体を除去することができることを確認した。

キーワード 移動物体除去, 車載カメラ, ベクトルメディアンフィルタ

Removal of Moving Objects from In-vehicle Camera Image Sequence by Fusing Multiple Image Sequences

Hiroyuki UCHIYAMA[†], Daisuke DEGUCHI[†], Tomokazu TAKAHASHI^{††},

Ichiro IDE[†], and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8601 Japan

^{††} Faculty of Economics and Information, Gifu Shotoku Gakuen University,
Nakauzura 1-38, Gifu-shi, Gifu 500-8288 Japan

E-mail: †{uchiyama,ddeguchi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††ttakahashi@gifu.shotoku.ac.jp

Abstract Applications of street images for driver assistance systems have been studied and services which exhibit street-view images are available on the Internet. However, vehicles and pedestrians in the street images are sometimes considered as a problem. In this report, we propose a method for the removal of moving objects from an in-vehicle camera image sequence by fusing multiple image sequences. First, multiple image sequences are collected by running vehicles equipped with a video camera along the same route many times. Next, spatial and temporal registration are applied between the image sequences. Finally, the images without moving objects are obtained by selecting and mosaicing partial images that contain no moving objects. From results of an experiment using actual image sequences, we confirmed that the proposed method could remove moving objects well.

Key words Removal of moving objects, in-vehicle camera, vector median filter

1. はじめに

カメラを搭載した車両が市街地を巡回することで収集した画像をドライバ支援に利用する研究が行われている。例えば、我々[1],[2] や Yazawa ら [3] は、市街地映像と位置情報で構成

されたデータベースと車載カメラ映像を照合することにより、自車位置推定を行う手法を提案している。また、久徳ら [4] は、市街地映像と、車載カメラ映像を比較することにより、路上の障害物を検出する手法を提案している。これらの手法では、車両等の移動物体が映っておらず、遮蔽のない市街地映像が必要

となる。

一方、Google ストリートビュー [5] は、インターネット上で市街地の全方位カメラ映像を閲覧できるサービスを提供している。このサービスでは、車両等の移動物体が映り込むことで、景観を損ねる、もしくは見たい部分が遮蔽されるといったことが問題となる。また、歩行者の顔や車両のナンバープレートの映り込みは、プライバシーの観点から大きな社会問題となっている。対策として、自動検出してぼかす処理が行われている [6] が、現状では検出漏れが多く、十分な性能とは言えない。

画像中から不要な物体を除去する方法として、(1) 静止画を利用するもの [7], [8] や、(2) 動画画を利用するもの [9] ~ [11] が提案されている。静止画を利用するものとして、1 枚の画像から写真の傷等の不要な領域を除去し、その欠損領域を修復して自然な画像を得る手法 [7] や、多数の学習画像をもとに、統計的に欠損領域の見えを推定する手法 [8] がある。また、動画画を利用するものとして、時空間画像を利用したフィルタリング処理により、除去対象領域を補間するもの [9], [10] がある。しかし、これらの方法には、除去対象領域をあらかじめ人手や自動検出などにより指定することを前提としたもの [7], [8], [10] や、除去対象領域検出処理が手法中に含まれているもの [9], [11] が多い。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、また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。

本報告では、ドライバ支援のための市街地映像データベースを構築することを目的とし、全方位車載カメラ映像から移動物体を除去する手法を提案する。本手法の特徴は、1 つの映像ではなく、複数の映像を利用する点であり、これにより検出処理を行わずに移動物体除去を実現する。

2. 移動物体除去手法

2.1 提案手法の概要

まず、提案手法の基本戦略について述べる。本報告では、駐車しているものを含む路上に存在する自動車、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画像中の同一部分を観察した場合、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、同じ経路を複数回走行することで、異なる時刻に撮影した同一地点の画像を収集する。次に、それらの画像の同一部分画像を比較し、移動物体の映っていない時刻の部分画像を自動選択し、張り合わせることによって、全体として移動物体の映っていない全方位カメラ画像を生成する (図 1)。

次に、処理手順の概略について述べる (図 2)。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、残りの $N - 1$ 本の映像をソース映像と呼ぶことにする。提案手法は、基準映像とソース映像間の時間的・空間的な位置合わせを行うレジストレーション処理と、移動物体が映っていない部分画像の選択処理の 2 つの処理に分けられる。全方位カメラを搭載した車両は路上を移動しているため、各映像の各時刻においてカメラの視点は異なる。そこで、全方位カメラ映像間の時間的・空間的なレジストレーションを行うこと

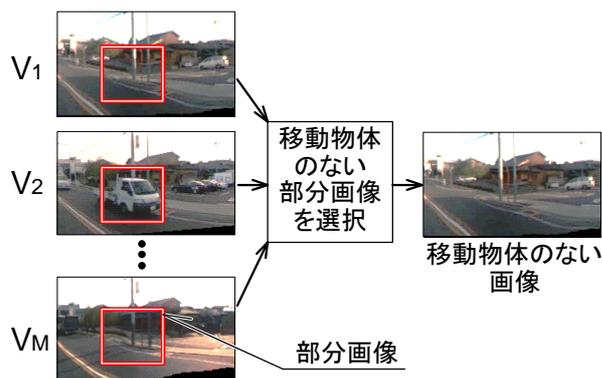


図 1 異なる時刻に撮影した同一地点の画像を収集し、移動物体の映っていない部分画像を自動選択して張り合わせるることにより、全体として移動物体の映っていない画像を生成する。

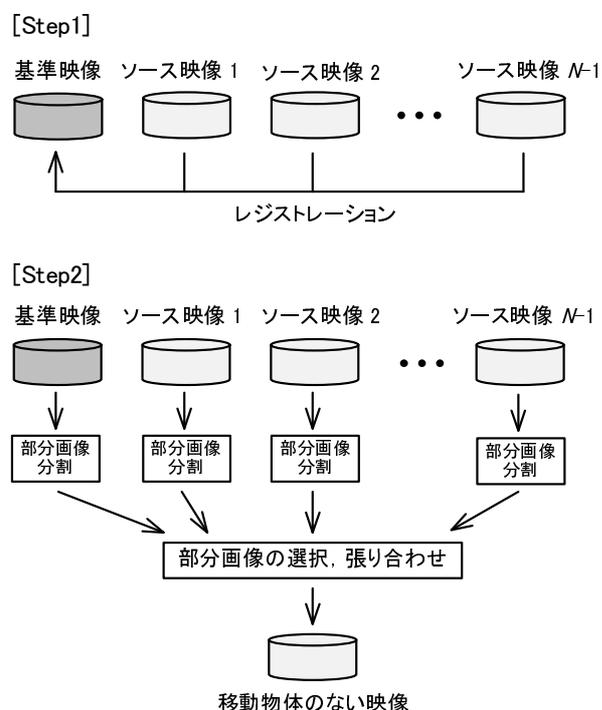


図 2 提案手法の処理の概略。

により、基準映像と同一の地点で観測した画像群を生成する。そして、基準映像とレジストレーションを行った各ソース映像を部分画像に分割する。最後に、移動物体が映っている頻度は少ないという仮定の下、最も背景らしい部分画像を自動選択し、張り合わせ処理を行うことで、移動物体が映っていない全方位カメラ映像を得る。以降、これらの処理の詳細を説明する。

2.2 全方位カメラ映像間のレジストレーション

2.2.1 DP マッチングによる時間方向の位置合わせ

まず、全方位カメラ映像間の時間方向の位置合わせにより、車両の進行方向のたまかな位置合わせを行う (図 3)。

車両の速度は走行毎に異なり、また場所により変化する。そこで、DP マッチング [12] 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。ここで、DP マッチングのパターン要素間の距離は、画像の画素

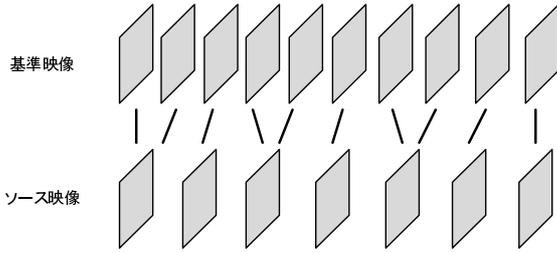


図 3 DP マッチングによる映像の時間軸方向のレジストレーション．これは、車両の進行方向のままかな位置合わせを行うことに相当する．

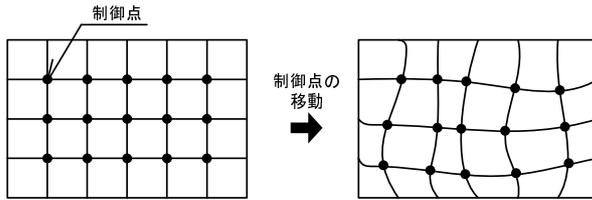


図 4 非剛体レジストレーション．

値の二乗誤差とする．DP マッチングは、基準映像とソース映像の $N - 1$ 個の対について、それぞれ適用する．

2.2.2 非剛体レジストレーションによる空間方向の位置合わせ

映像を収集する車両は、同じ車線を走行しても、車線内での走行位置は走行毎に異なる．また、カメラのフレームレートが制限されるため、DP マッチングによる進行方向の撮影位置合わせの結果にも誤差が生じる．そのため、映像の時間方向の位置合わせだけでなく、空間方向の位置合わ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おり、撮影位置が異なると画像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、部分画像をきれいつなぎ合わせるために、ピクセル単位で画像の位置合わ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 [13] 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、まず変形対象となるソース画像上に格子状に制御点を設置する (図 4)．そして、制御点を移動することでソース画像を変形させる．基準画像との距離が最小となるように、ソース画像を変形させることで、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距離は二乗誤差とする．非剛体レジストレーションは、DP マッチングと同様に、各ソース映像を基準映像に合わせる．

非剛体レジストレーションを全方位カメラ映像のあるフレームに適用した例を図 5 に示す．図中では、ソース画像と基準画像をチェッカーボード状に並べて表示してある．(a) では、例えば左上の建築物の境界や右下の白線の位置がずれている．これを、非剛体レジストレーションにより、(b) のように補正する．

2.3 ベクトルメディアンフィルタによる部分画像選択処理レジストレーションを行ったある地点の画像群を部分画像に分割し、その中から移動物体が映っていない部分画像を選択し

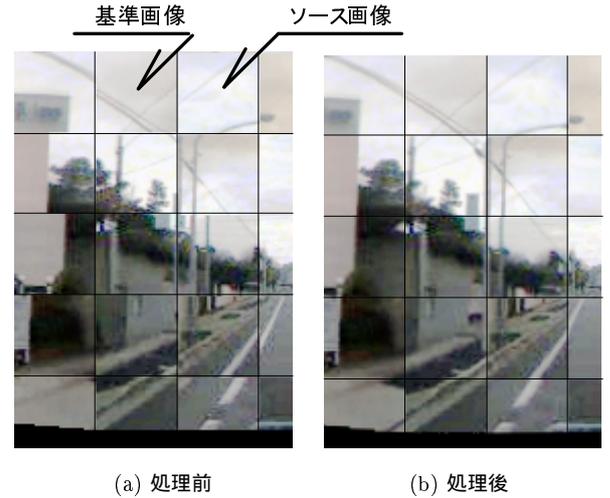


図 5 非剛体レジストレーションによる映像の空間方向のレジストレーション．ソース画像と基準画像をチェッカーボード状に並べて表示してある．(a) では、例えば左上の建築物の境界や右下の白線の位置がずれている．これを、(b) のように補正する．

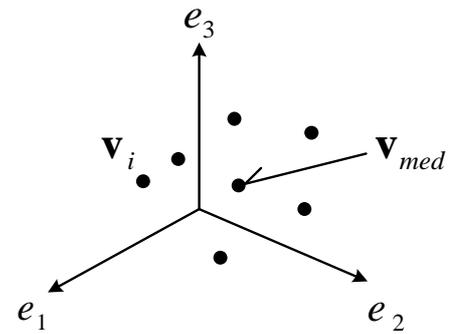


図 6 ベクトルメディアンフィルタ．特異なベクトルを除外し、最も代表的なベクトルを選択する性質がある．

ていく．まず、各画像上に $W \times W$ 画素の窓を、縁の一部分が窓の大きさの半分程度重なり合うように設置し、窓内の部分画像をベクトルとして表す．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．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下、同じ位置の窓のベクトルに対し、ベクトルメディアンフィルタ [14] を適用することで、最も背景らしい部分画像を選択する．

ベクトルメディアンフィルタはメディアン多次元への拡張である geometric median によって計算され、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 (図 6)． M 個のベクトル $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M$ を入力とするベクトルメディアンフィルタの出力 \mathbf{v}_{med} は、次式で表される．

$$\mathbf{v}_{med} = \arg \min_{\mathbf{v} \in \{\mathbf{v}_1, \dots, \mathbf{v}_M\}} \sum_{i=1}^M |\mathbf{v} - \mathbf{v}_i| \quad (1)$$

ベクトルが 1 次元の場合、geometric median はスカラのメディアンと等価である．提案手法では、ベクトル間の距離としてユークリッド距離を用いる．

ベクトルメディアンフィルタは、特異なベクトルを除外し、

与えられたベクトル群から最も代表的なベクトルを選択する性質がある．そのため，移動物体が映っている頻度は少ないという仮定により，移動物体が映っている部分画像よりも，背景画像の方が選択されやすくなる．

2.4 画像の張り合わせ

部分画像同士が重なり合う領域を ブレンドリングすることで，境界部分が目立ちにくい合成画像を生成する．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，ブレンドリングの係数を設定する．

3. 実験および考察

次の2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: 部分画像選択の際に最適な窓サイズに関する検討

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．基準映像用に 1 本，ソース映像用に 14 本の合計 15 本の映像を用いた．映像は全て異なる日に撮影した．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 あたり 1024 画素であった．

3.1 実験 1: 移動物体除去性能の評価

ソース映像として使用する映像の数を変化させ，使用する映像の数と移動物体の除去性能の関係を調査した．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．用いた指標は次のとおりである．

$$\text{残留} = \frac{|A \cap B|}{|A|}$$

$$\text{増加} = \frac{|\bar{A} \cap B|}{|A|}$$

$$\text{合計} = \frac{|B|}{|A|} = \text{残留} + \text{増加}$$

ただし， A は基準映像中で移動物体が存在する領域， \bar{A} は基準映像中で移動物体が存在しない領域， B は除去映像中に移動物体が残る領域を示し， $|\cdot|$ は領域中の画素数を表す．14 本のソース映像の全ての組み合わせについて実験を行い，その平均を評価に用いた．例えば，ソース映像を 7 本用いる場合は， ${}_{14}C_7$ 通りの組み合わせについて実験した．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．この実験での部分画像選択処理の際の窓のサイズは 30×30 画素とした．

用いた映像数に対する除去失敗率を図 7 に示す．映像数が 15 のとき，合計の除去失敗率は 3% 以下であり，97% 以上の移動物体領域の除去に成功した．残った移動物体領域のほとんどが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として十分利用可能であると考えられる．

図 7 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．さらに多くの映像を収集することで，より高い除去効果が得られると考えられる．しかし，使用映像数が少ない場合は，残留の率が特に大きかった．最低 3 本の映像があ

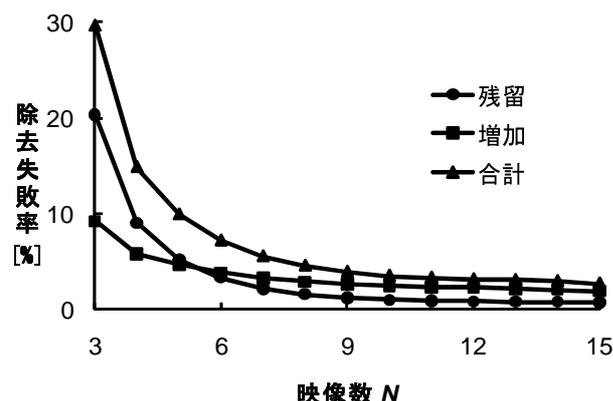


図 7 映像数と除去失敗率の関係．

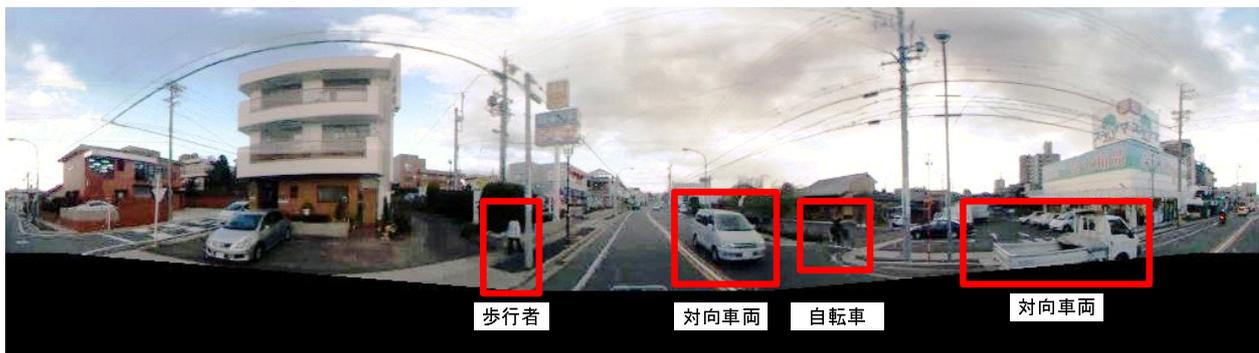
ればベクトルメディアンフィルタによる部分画像選択は可能であるが，9 割以上の移動物体除去の性能を得るためには，5 本以上の映像が必要であった．

移動物体除去の例を図 8 に示す．目視により合成後の画質について確認したところ，映像数の違いによる画質の差を確認することはできなかった．また，元の画像に対して画質が大きく劣化することはなかった．しかし，背景中の電線やボールなど，細く，空間中に孤立した物体をきれいに再現できないことが一部確認された．これは，撮影位置の違いによる画像の見えの違いを，画像の 2 次元的な変形により近似したことおよび，非剛体レジストレーションがこのような形状間の照合を苦手とすることが原因として考えられる．

3.2 実験 2: 部分画像選択の際に最適な窓サイズの検討

ベクトルメディアンフィルタによる部分画像選択処理において，最も除去効果の高い窓のサイズについて検討した．使用する映像数を 7 本に固定し，窓のサイズを変化させ，除去失敗率を調べた．窓のサイズと除去失敗率の関係を図 9 に示す．窓サイズが 30×30 画素のとき，残留と増加の率が共に最小となり，移動物体の除去の観点で最も効果があることを確認した．なお， 30×30 画素の矩形は，立体角では 0.0216 sr に相当する．

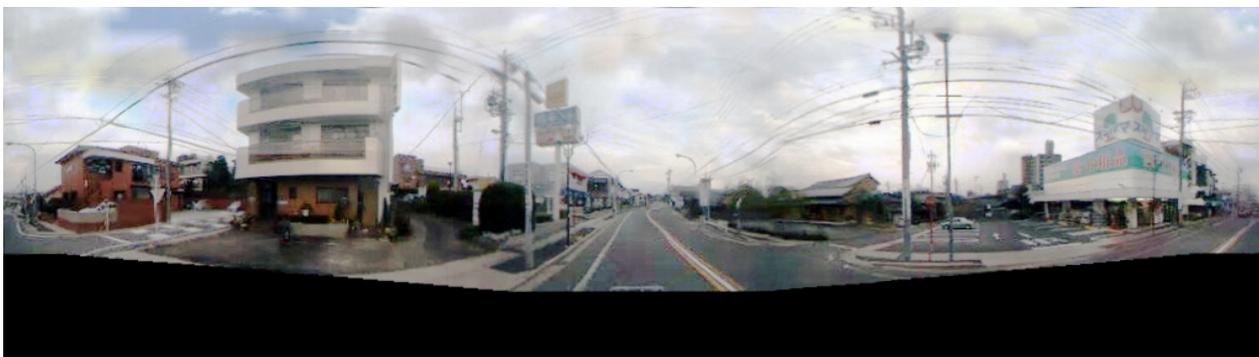
しかし，窓のサイズによる除去効果の違いは，窓サイズと映像中の移動物体の見かけ上のサイズに関係があり，移動物体の見かけ上のサイズによっては 30×30 画素では効果的に除去できない場合があると考えられる．例えば，図 10 のように，移動物体とカメラとの距離が遠い場合などが挙げられる．図のように，移動物体の見かけ上のサイズが小さい場合には， 15×15 画素のときは除去に成功したが，最も除去性能の高い 30×30 画素では失敗することがあった．これは，窓内に占める移動物体の領域の割合が小さく，部分画像選択処理において，移動物体領域と背景物体領域の区別ができなかったためである．残存する移動物体の領域は小さいため，目的とするドライバ支援のためのデータベースには支障はないと考えられる．しかし，より除去効果を高めるならば，移動物体が大きく映りやすい自車の側面については窓サイズを大きく設定し，小さく写りやすい前後面は窓サイズを小さく設定する方法が考えられる．



(a) 基準映像



(b) 5本の映像の統合結果



(c) 9本の映像の統合結果



(d) 15本の映像の統合結果

図8 移動物体除去の結果の例.

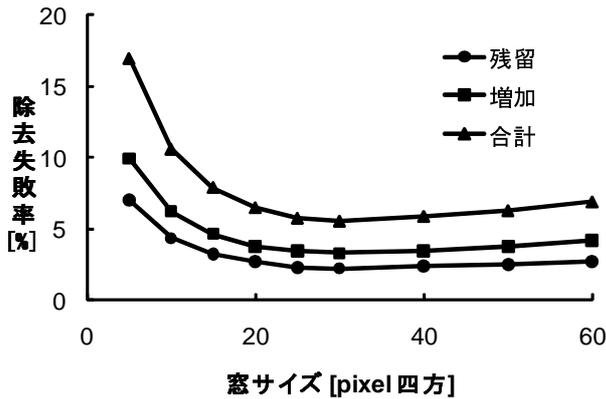


図9 窓サイズと除去失敗率の関係。

4. ま と め

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ついて、時間的・空間的なレジストレーションを行った。次に、部分画像単位で、移動物体の映っていない画像を選択し、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。15本の映像を用いた実験の結果、97%の移動物体の除去に成功した。また、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。今後の課題として、少ない映像数で移動物体除去性能を高める手法の開発や、細かい部分の再現性を高める方法の検討等が挙げられる。

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 およ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, 高橋友和, 井手一郎, 村瀬洋, “自車位置推定のための車載全方位カメラ映像と通常カメラ映像のフレーム間対応付け,” 信学技報, PRMU2007-272, Mar. 2008.
- [2] 内山寛之, 出口大輔, 高橋友和, 井手一郎, 村瀬洋, “街並み画像系列と2台の車載カメラの画像系列との照合による自車の走行位置推定,” 信学技報, PRMU2008-159, Dec. 2008.
- [3] N. Yazawa, H. Uchiyama, and H. Saito, “Image Based View Localization System Retrieving from a Panorama Database by SURF,” Proc. IAPR Conf. on Machine Vision Applications (MVA2009), pp.118-121, May 2009.
- [4] 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, IS1-55, July 2009.
- [5] “Google Maps,” <http://maps.google.co.jp/>
- [6] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, L. Vincent, “Large-scale Privacy Protection in Google Street View,” Proc. IEEE 12th Int. Conf. on Computer Vision (ICCV2009), pp.2373-2380, Sept. 2009.
- [7] 河合紀彦, 佐藤智和, 横矢直和, “パターン類似度に基づくエネルギー最小化による画像修復,” 信学技報, PRMU2006-163, Dec.



(a) 15 × 15 画素



(b) 30 × 30 画素

図10 窓サイズによる除去効果の違い (15本の映像を使用)。窓サイズが大きい場合、遠くの小さく映る移動物体の除去を失敗しやすい。(b)の矩形の辺の長さが30画素に相当する。

- [8] A. Levin, A. Zomet, and Y. Weiss, “Learning How to Inpaint from Global Image Statistics,” Proc. IEEE 9th Int. Conf. on Computer Vision (ICCV2003), pp.305-312, Oct. 2003.
- [9] 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, IS2-32, July 2009.
- [10] Y. Wexler, E. Shechtman, and M. Irani, “Space-Time Video Completion,” 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR2004), Vol.1, pp.120-127, June 2004.
- [11] 福地功, 山下淳, 金子透, “時空間画像処理による動画からの視野妨害ノイズ除去,” 情処学研報, 2006-CVIM-154, May 2006.
- [12] 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, Apr. 2007.
- [13] 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. on Medical Images, Vol.18, No.8, pp.712-721, Aug. 1999.
- [14] J. Astola, P. Haavisto, and Y. Neuvo, “Vector Median Filters,” Proc. of the IEEE, Vol.78, No.4, pp.678-689, Apr. 1990.