

ビジュアルオドメトリの前処理としての 歩行者除外手法の検討

山崎 達郎^{1,a)} 藤田 茂²

概要: 自律移動ロボットの自己位置推定の手法として、カメラ画像の特徴点から、回転量・並進量を推定するビジュアルオドメトリがある。ビジュアルオドメトリの問題点として、静的な環境を仮定しているため、歩行者などの動体が画像内に存在すると推定に誤差が生じるという点が挙げられる。誤差を抑える手法としてRANSACがある。しかし、RANSACは動体による特徴点を除いて推定を行う手法ではないので、歩行者が多く動体による特徴点が多い状況では、誤差が大きくなる問題がある。そこで本研究では、回転量・並進量を推定する前段階で歩行者による特徴点を除去することにより、推定誤差を抑える手法を検討する。

1. はじめに

ロボット技術の発展にともない、様々な場面でロボットの活躍が期待されており、近年では家庭用の掃除ロボットなども普及しはじめている。少子高齢化と言われている現代では、警備や案内などの生活支援サービス代行が期待されている。また、作業者の安全性を考慮し、人体に有害な影響のある環境や二次災害の危険性のある災害現場での救助活動などにおいてもロボットの必要性が高まってきている。

人間の生活の周りで活用される自律移動ロボットは、工業用ロボットなど限られた特別な環境下でのみ動くものとは異なり、移動体などにより刻々と変化する環境下を走行することができる柔軟性と、人間や障害物に衝突して被害を与えないための安全性、正確性が求められる。

ロボットが安全かつ確実に自律走行を行うのに必要な要素の一つとして、自己位置推定がある。自己位置推定は、ロボット自身がどの場所に存在しているかを推定する技術である。自己位置がうまく推定できないと、いま地図上のどの場所にいるのかを認識できなくなるため自律的に移動する事ができなくなる。

自己位置推定の手法として、左右タイヤの回転量をそれぞれ計測し、その差から移動距離を推定するオドメトリが

ある。しかし、オドメトリはタイヤのスリップによる誤差が生じ、移動距離の推定に誤差が発生するという問題がある。その他に、GPSを用いて自己位置を推定する手法もある。しかし、GPSは屋内では測位を行うことが出来ないため、屋内ではGPSを用いた自己位置推定を行えないという問題がある。

タイヤのスリップに影響されない自己位置推定の手法として、カメラ画像を用いてオドメトリを行う、ビジュアルオドメトリがある [1]。ビジュアルオドメトリの問題点として、静的環境を仮定した手法であるため、人が実際に生活する環境では動体により推定に誤差が生じるという点が挙げられる。

動体の存在する環境でビジュアルオドメトリを用いて自己位置推定を行いながら自律移動ロボットが走行を行うためには、取得したカメラ画像から動体を識別し、除外する必要がある。人が実際に生活する街中には、歩行者や車など多くの動体が存在する。これらのすべてを識別して除外することは難しい。そこで本研究では、人が実際に生活する環境における動体として歩行者に焦点をあて、歩行者による特徴点を除外することにより、自己位置の推定誤差を抑える手法を検討する。

2. 関連研究

2.1 ビジュアルオドメトリ

ビジュアルオドメトリとは、複数枚の連続した画像からカメラの回転行列、並進行列を算出し、これらの行列からカメラの移動量を推定する手法 [1] である。現在のフレーム画像 I_t にみられる特徴点 $\mathbf{p} = (u, v, 1)^T$ と、1つ前のフレーム画像 I_{t-1} 内で \mathbf{p} に対応する特徴点 $\mathbf{p}' = (u', v', 1)^T$

¹ 千葉工業大学大学院 情報科学研究科
Graduate School of Information and Computer Science,
Chiba Institute of Technology

² 千葉工業大学 情報科学部情報工学科
Computer Science, Faculty of Information and Computer
Science, Chiba Institute of Technology

^{a)} yamasaki@sf.cs.it-chiba.ac.jp

には、次のような拘束がある。

$$\mathbf{p}^T \mathbf{F} \mathbf{p}' = 0 \quad (1)$$

ここで、 u, u' は画像内の特徴点の横方向の画素位置、 v, v' は縦方向の画素位置をそれぞれ表す。この拘束をエピポラ拘束という。(1)式において、 \mathbf{F} は Fundamental Matrix (F 行列, 基礎行列) と呼ばれる。F 行列はカメラの回転行列、並進行列とカメラの内部パラメータ行列に依る行列である。撮影に用いられるカメラの内部パラメータ行列が分かっているならば、F 行列から特異値分解を用いて回転行列、並進行列を算出することができる。ビジュアルオドメトリでは、F 行列を精度良く求めることが重要である。

特徴点の検出には、Harris, FAST, SIFT, SURF などを用いることが出来る [2]。現在のフレーム画像と 1 つ前のフレーム画像でそれぞれ特徴点の検出し、その後、フレーム画像間で特徴点のマッチングを行う。特徴点マッチングの例を図 1 に示す。図 1 の右下の画像が現在のフレーム画像、左上の画像が 1 つ前のフレーム画像である。マッチングによって組となった特徴点をそれぞれ \mathbf{p}', \mathbf{p} とする。

(1) 式を展開すると、

$$[u'u \quad uv' \quad u \quad u'v \quad v'v \quad v \quad u' \quad v' \quad 1] \mathbf{F}' = 0 \quad (2)$$

となる。ここで、

$$\mathbf{F}' = [f_1 \quad f_2 \quad f_3 \quad f_4 \quad f_5 \quad f_6 \quad f_7 \quad f_8 \quad f_9] \quad (3)$$

であり、 \mathbf{F} は、

$$\mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \quad (4)$$

である。 \mathbf{F}' を算出することで F 行列を求めることができる。(2) 式から \mathbf{F}' を求めるには、特徴点が 8 点あれば

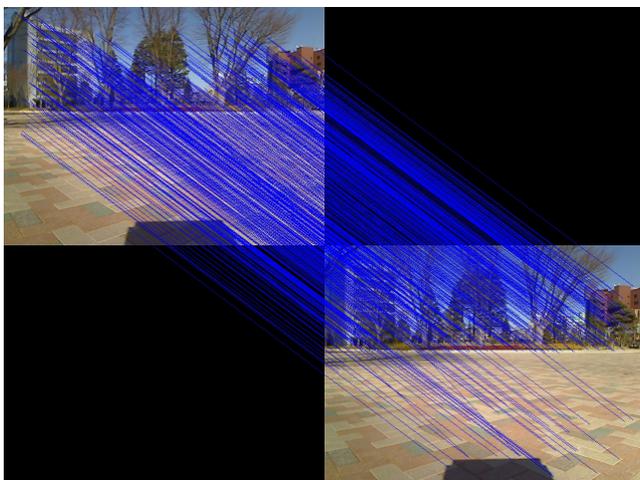


図 1 特徴点マッチングの例
 Fig. 1 An example of feature matichnig

よい。また、5 点の特徴点によって F 行列を求める手法 [3] もある。

動体の存在しない画像間の特徴点マッチングでも、本来とは異なる特徴点とマッチングしてしまうことがある。このような特徴点を F 行列の算出に用いてしまうと、算出結果が本来の F 行列の値とは異なってしまふ。このような問題を解決する手法として、random sample consensus (RANSAC) [4] がある。RANSAC では、用いる特徴点をランダムに選び F 行列の算出し、すべての特徴点について (1) 式を計算し誤差を求める。この計算を一定回数繰り返す、もっとも誤差が少ない F 行列を正解とする。

エピポラ拘束を用いた F 行列の算出は、フレーム画像内に動体が写っていない静的な環境を仮定している。したがって、動的な環境、例えば歩行者がいる街中などでは、歩行者に起因する特徴点を用いて F 行列の算出を行ってしまふ、本来とは異なる F 行列の値が得られるという問題が起こる。RANSAC は、F 行列の算出に用いる特徴点をランダムに選び、計算回数を多くすることで、マッチングに失敗した特徴点を除いた正確な F 行列の算出を行う手法である。よって、カメラの移動のみによって画素位置が変化したものではない特徴点、つまり動体による特徴点がフレーム画像内に多くなると、特徴点の選択の際に動体による特徴点を選ぶ確率が上がり、RANSAC では正確な F 行列を求めることができない。

2.2 機械学習による歩行者の検出

Advanced Driver Assistance Systems (ADASs) や Intelligent Vehicles の分野において、歩行者と車の衝突事故を未然に防ぐ等の理由で、カメラ画像から歩行者をロバストに検出しトラッキングする研究 [5-8] が行われている。

Prioletti ら [5] は、Haar-Like 特徴を用いた分類器で全身の領域を検出した後、検出されたそれぞれの領域で頭部・胴部・下半身について Histogram of Oriented Gradient (HOG) 特徴量を用いた検出を行っている。全身の検出と体の各部分の検出を組み合わせることで、疑陽性に対する頑健性と遮蔽による頑健性を得ている。Enzweiler ら [6] も同様に、上半身・胴部・下半身・左半身・右半身のそれぞれについて HOG 特徴量を用いた検出を行うことで、遮蔽にロバストな歩行者の検出を行っている。

3. 課題

歩行者が存在する環境で取得したカメラ画像では、歩行者がフレームイン・フレームアウトする瞬間に半身しか写らない。また、カメラと歩行者の距離の関係で、頭部や足元が写らないことがある。図 2 に身体の一部が写らない例を示す。

一方で、機械学習による分類を用いた人検出では、学習データに人の全身が写った画像を多く用いる。図 3 に学習



図 2 身体の一部が写らない例
Fig. 2 Examples of occlusion



図 3 学習データの例
Fig. 3 An example of training datas

データの例を示す。このような学習データによって学習を行った分類器では、フレームイン・フレームアウトの瞬間やカメラとの距離の関係で身体の一部が写っていない歩行者を検出することが出来ないという問題がある。

4. 提案手法

本研究では、車両型自律移動ロボットにおける、歩行者の存在する環境でのビジュアルオドメトリを扱う。ロボットに搭載された単眼カメラから取得した画像を用いて F 行列を算出し、F 行列からカメラの回転行列と並進行列を求め、移動量を推定する。

4.1 動的環境における F 行列の算出手法

F 行列を算出する際に、動体由来の特徴点を用いて算出を行ってしまうと、実際の F 行列とは異なった値を算出してしまう。歩行者の存在する環境では、歩行者は動体となる。そこで、提案手法では、まず、機械学習による人検出を用いて歩行者の検出を行う。図 4 に機械学習による人検出の例を示す。次に、検出された歩行者領域に含まれる特徴点を除いて、F 行列の算出を行う。図 5 に歩行者領域の除外のイメージを示す。

機械学習による分類器は、HOG による上半身・下半身・胴部・左半身・右半身の分類器をそれぞれ作成し、人検出に用いる。身体の一部についてそれぞれ検出を行うのは、フレームイン・フレームアウトやカメラとの距離の関係で歩行者の身体の一部が写らない問題に対応するためである。

このような機械学習による歩行者の検出では、歩行者の他に、移動をしていない人も検出してしまう。また、人が存在しない領域を誤検出することもある。図に誤検出の例

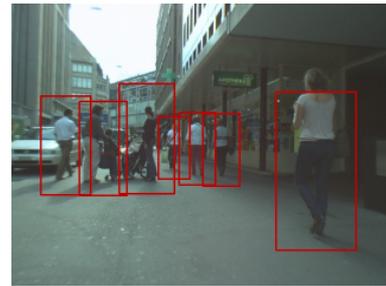


図 4 機械学習による人検出の例
Fig. 4 An example of people detection



図 5 歩行者領域の除外のイメージ
Fig. 5 An example of exception pedestrians

を示す。移動をしていない人や人が存在しない領域を誤って検出してしまうことは、歩行者の検出としては失敗である。しかし、これらの誤検出を含めて歩行者領域とし、この領域に含まれる特徴点を除いて F 行列を算出しても、算出に用いられる動体由来でない特徴点の候補数が減少するだけで、F 行列の誤算出にはつながらない。本研究では、ビジュアルオドメトリに向けた、動的な環境での F 行列の正確な算出を目的としている。このことから、歩行者の検出における、移動をしていない人や人が存在しない領域の誤検出については問題としないこととする。

4.2 想定するロボットについて

本研究では、車両型ロボットの自律移動に向けたビジュアルオドメトリを想定している。本研究で想定するロボットの外観を図 6 に示す。

ビジュアルオドメトリに用いるカメラ画像は、ロボット前方に設置した Logicool Qcam Pro 9000 を用いて撮影する。

本研究室では、ロボットの自律移動へのチャレンジとして、つくばチャレンジに参加してきた。つくばチャレンジは、実際に人が生活する街の中を、安全にかつ確実に動くことを目標としている。つくばチャレンジのように実際の街中を走行する際には、タイヤのスリップに影響を受けて、タイヤの回転量によるオドメトリでは自己位置の推定に誤差が生じてしまう。このような問題に対して、タイヤのスリップに影響を受けないビジュアルオドメトリは有効である。しかし、実際の街中では歩行者などの動体が存在する



図 6 車両型ロボットの外観
 Fig. 6 Appearance of vehicle robot

ため、動体の影響によって誤った F 行列の値を算出するという問題がある。この動体に関する問題について、本研究では提案手法により改善を行う。

5. 評価実験

データセットを用いて、任意のシーンにおける、

- (1) 手動で動体領域を指定し算出した F 行列による (1) 式の二乗誤差
- (2) 動体領域を指定せずに算出した F 行列による (1) 式の二乗誤差
- (3) 提案手法によって算出した F 行列による (1) 式の二乗誤差

のそれぞれの比較を行う。

カメラ画像のデータセットとして、歩行者が存在する環境で撮影されたデータセットであること、撮影カメラの高さが本研究で対象としている車両型ロボットに搭載されたカメラと近い等の理由で、ETH Linthescher [9] を用いた。ETH はステレオカメラで撮影された。左カメラと右カメラの画像データが公開されている。本研究では単眼カメラを対象としているため、左カメラの画像データのみを用いた。

また、歩行者検出に用いる分類器の学習データのセットとして、Daimler Detection Benchmark [8] の学習用歩行者画像データを利用した。Daimler Detection Benchmark の学習用歩行者画像データは、 20×40 サイズと 48×96 サイズの二種類が用意されている。今回は、 48×96 サイズのものを用いた。カメラ画像のデータセットとは異なるデータセットの学習データを用いた理由は、ETH に学習用のデータが用意されていなかったことと、実験の再現性確保

表 1 画像の組み合わせ

Table 1 Sets of images

	現在のフレーム画像	1 つ前のフレーム画像
試行 1	image_00000931_0.png	image_00000930_0.png
試行 2	image_00000951_0.png	image_00000950_0.png
試行 3	image_00000985_0.png	image_00000984_0.png
試行 4	image_00001401_0.png	image_00001400_0.png

表 2 試行 1 の結果

Table 2 A result of trial 1

手法	二乗誤差
手動	32.64
RANSAC のみ	211.49
提案手法	17.99

表 3 試行 2 の結果

Table 3 A result of trial 2

手法	二乗誤差
手動	14.159
RANSAC のみ	11.629
提案手法	84.081

表 4 試行 3 の結果

Table 4 A result of trial 3

手法	二乗誤差
手動	9.830
RANSAC のみ	23.712
提案手法	44.5478

表 5 試行 4 の結果

Table 5 A result of trial 4

手法	二乗誤差
手動	3.855
RANSAC のみ	9.409
提案手法	1.553

の為である。

歩行者検出に用いる、上半身・下半身・左半身・右半身・胴部のそれぞれの分類器の学習データは、Daimler Detection Benchmark の学習用データから作成した。Daimler Detection Benchmark の学習用データを縦・横にそれぞれ二等分し、上半身・下半身・左半身・右半身の学習データとした。胴部の学習データは、Daimler Benchmark の学習用データを上から 16%, 34%, 50% の領域に分け、真ん中の 34% の領域を学習データとした。作成したそれぞれの学習データの例を図 7 に示す。それぞれの分類器の学習には、正解画像 2000 枚、非正解画像 6000 枚を用いた。

6. 実験結果

ETH Linthescher の画像の中からの 4 枚の画像を選び、それらの画像を現在のフレームとして実験を行った。実験



図 7 作成した学習データの例. 左から元画像, 左半身の学習データ, 右半身の学習データ, 上半身の学習データ, 胴部の学習データ, 下半身の学習データ.

Fig. 7 Examples of training datas. Original data(the left), Left-half(the second one form the left), Right-half(the third one from the left), Top(the third one from the right), Torso(the second one from the right), Bottom(the right)

に使われた画像のセットを表 1 に示す.

表 1 の各試行の結果を, 表 2 から 5 に示す.

また, 試行 1 において手動で除外した領域と提案手法で除外した領域を図 8 から図 11 に, 試行 2 において手動で除外した領域と提案手法で除外した領域を図 12 から図 15 に, 試行 3 において手動で除外した領域と提案手法で除外した領域を図 16 から図 19 に試行 4 において手動で除外した領域と提案手法で除外した領域を図 20 から図 23 に, それぞれ示す.

7. 考察

表 2 と表 5 において, 手動で動体を除外して算出した F 行列を用いて求めた (1) 式の二乗誤差と, 他の 2 つの二乗誤差の値を比較すると, RANSAC のみに比べて, 提案手法を用いて求めた二乗誤差の方が手動で動体を除外して求めた二乗誤差の値に近い. このことから, 試行 1 と試行 4 では提案手法によって実際の F 行列に近い値を求められたことがわかる.

一方, 表 3 と表 4 では, 提案手法を用いて求めた二乗誤差よりも RANSAC のみを用いて求めた二乗誤差の方が, 手動で動体を除去して算出した F 行列を用いて求めた二乗誤差に近い. このことから, 試行 2 と試行 3 では RANSAC のみを用いる方が提案手法よりも実際の F 行列に近い値を求められたことがわかる.

試行 2 と試行 3 で, 提案手法を用いて算出した F 行列が, RANSAC のみを用いて算出した F 行列よりも実際の値から外れた原因として, 歩行者領域の誤検出が考えられる. 図 13 をみると, 画像右側の大きな枠が誤検出をしていることが分かる. この誤検出により, 背景の建物による特徴点の多くが除外される. 一方で, 図 15 では, 画像右側に背景の建物を覆うような誤検出の領域はなく, 背景の建物による特徴点は除外されない. この特徴点の量の差によって特徴点のマッチングミスが発生すると, 求める F 行列の値が実際の値から外れる原因となる. 図 12 と図 13,

図 14 と図 15 をそれぞれ比べてみると, 手動で除外した領域は提案手法でも除外されていることがわかる. 試行 3 においても, 図 16 と図 17, 図 18 と図 19 をそれぞれ比較することで同様のことがわかる. このことから, 歩行者領域の誤検出が F 行列の算出結果の誤差に影響していると考えられる. 歩行者領域の誤検出が, F 行列の算出結果の誤差にどう影響しているかは, 検討の必要がある.

8. おわりに

本研究では, 車両型自律移動ロボットの動的環境での自己位置推定に向けて, ビジュアルオドメトリの前処理として歩行者を除外する手法について検討した.

実験により, 提案手法が RANSAC のみを用いた F 行列の算出よりも実際の F 行列に近い値を算出した実験結果を得たが, 一方で, RANSAC のみを用いて F 行列を算出した方が提案手法よりも実際の F 行列に近い値を得られた実験結果もあった. 歩行者の誤検出が F 行列の算出の誤差にどのように影響しているかは, 今後, 検討する予定である.



図 8 試行 1 で手動で除外した領域 (前フレーム)

Fig. 8 Excerpted regions in trial 1 by manual operation. (prior frame)

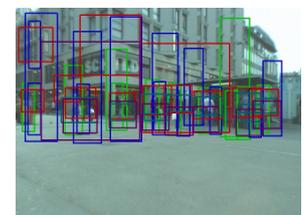


図 9 試行 1 で提案手法で検出し除外した領域 (前フレーム)

Fig. 9 Excerpted regions in trial 1 by proposal method. (prior frame)

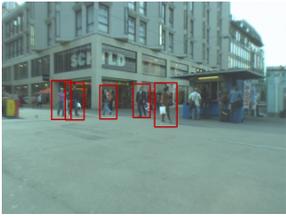


図 10 試行 1 で手動で除外した領域 (現フレーム)

Fig. 10 Excluded regions in trial 1 by manual operation. (current frame)

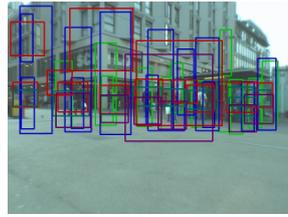


図 11 試行 1 で提案手法で検出し除外した領域 (現フレーム)

Fig. 11 Excluded regions in trial 1 by proposal method. (current frame)



図 16 試行 3 で手動で除外した領域 (前フレーム)

Fig. 16 Excluded regions in trial 3 by manual operation. (prior frame)

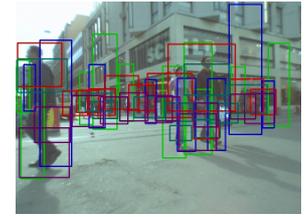


図 17 試行 3 で提案手法で検出し除外した領域 (前フレーム)

Fig. 17 Excluded regions in trial 3 by proposal method. (prior frame)

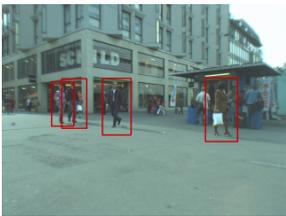


図 12 試行 2 で手動で除外した領域 (前フレーム)

Fig. 12 Excluded regions in trial 2 by manual operation. (prior frame)



図 13 試行 2 で提案手法で検出し除外した領域 (前フレーム)

Fig. 13 Excluded regions in trial 2 by proposal method. (prior frame)

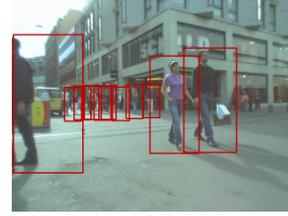


図 18 試行 3 で手動で除外した領域 (前フレーム)

Fig. 18 Excluded regions in trial 3 by manual operation. (prior frame)

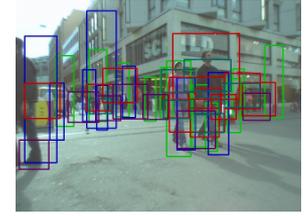


図 19 試行 3 で提案手法で検出し除外した領域 (現フレーム)

Fig. 19 Excluded regions in trial 3 by proposal method. (current frame)



図 14 試行 2 で手動で除外した領域 (現フレーム)

Fig. 14 Excluded regions in trial 2 by manual operation. (current frame)

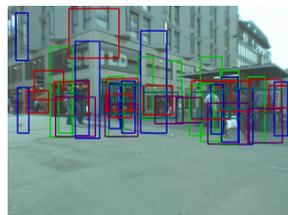


図 15 試行 2 で提案手法で検出し除外した領域 (現フレーム)

Fig. 15 Excluded regions in trial 2 by proposal method. (current frame)

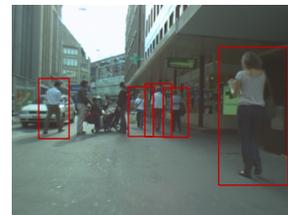


図 20 試行 4 で手動で除外した領域 (前フレーム)

Fig. 20 Excluded regions in trial 4 by manual operation. (current frame)

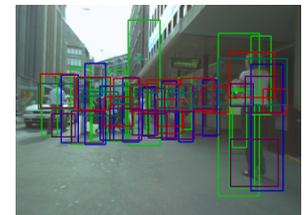


図 21 試行 4 で提案手法で検出し除外した領域 (前フレーム)

Fig. 21 Excluded regions in trial 4 by proposal method. (prior frame)

参考文献

- [1] Scaramuzza, D. and Fraundorfer, F.: Visual Odometry [Tutorial], *Robotics Automation Magazine, IEEE*, Vol. 18, No. 4, pp. 80-92 (online), DOI: 10.1109/MRA.2011.943233 (2011).
- [2] Fraundorfer, F. and Scaramuzza, D.: Visual Odometry : Part II: Matching, Robustness, Optimization, and Applications, *Robotics Automation Magazine, IEEE*, Vol. 19, No. 2, pp. 78-90 (online), DOI: 10.1109/MRA.2012.2182810 (2012).
- [3] Nister, D.: An efficient solution to the five-point relative pose problem, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 26, No. 6, pp. 756-770 (online), DOI: 10.1109/TPAMI.2004.17 (2004).
- [4] Fischler, M. A. and Bolles, R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM*, Vol. 24, No. 6, pp. 381-395 (online), DOI: 10.1145/358669.358692 (1981).
- [5] Prioletti, A., Mogelmose, A., Grisleri, P., Trivedi, M., Broggi, A. and Moeslund, T.: Part-Based Pedestrian Detection and Feature-Based Tracking for Driver Assistance: Real-Time, Robust Algorithms, and Evaluation (2013).
- [6] Wojek, C., Walk, S., Roth, S., Schindler, K. and Schiele, B.: Monocular Visual Scene Understand-

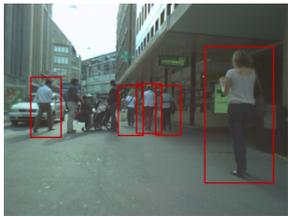


図 22 試行 4 で手動で除外した領域 (現フレーム)

図 23 試行 4 で提案手法で検出し除外した領域 (現フレーム)

Fig. 22 Excepted regions in trial 4 by manual operation. (current frame)

Fig. 23 Excepted regions in trial 4 by proposal method. (current frame)

ing: Understanding Multi-Object Traffic Scenes, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 35, No. 4, pp. 882–897 (online), DOI: 10.1109/TPAMI.2012.174 (2013).

- [7] Dollar, P., Wojek, C., Schiele, B. and Perona, P.: Pedestrian Detection: An Evaluation of the State of the Art, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 34, No. 4, pp. 743–761 (online), DOI: 10.1109/TPAMI.2011.155 (2012).
- [8] Enzweiler, M. and Gavrilu, D.: Monocular Pedestrian Detection: Survey and Experiments, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 31, No. 12, pp. 2179–2195 (online), DOI: 10.1109/TPAMI.2008.260 (2009).
- [9] Ess, A., Leibe, B., Schindler, K. and Van Gool, L.: A mobile vision system for robust multi-person tracking, *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8 (online), DOI: 10.1109/CVPR.2008.4587581 (2008).