

人物の骨格と画像中のエッジを用いた 所持品領域検出に関する一検討

三浦 誠^{1,a)} 阿部 亨^{1,2,b)} 菅沼 拓夫^{1,2,c)}

概要：荷物の置き引きや不審物の放置など人物の所持品が関係した犯罪を監視するため、画像中から所持品の領域を検出することは重要である。近年、(1) 人物領域の雛型を生成し、(2) 画像中の移動物体領域（前景領域）で雛型と異なる箇所（所持品候補箇所）の決定を行い、(3) 所持品候補箇所の特徴から所持品領域を検出するという3つのステップに従い、画像中から所持品領域を検出するアプローチが提案されている。しかし、このアプローチに基づく従来手法では、各ステップの精度が十分でないため、所持品領域を高精度に検出することが困難であった。この課題へ対処するため、本稿では、人物の骨格に基づく雛型の生成、画像中の前景領域のエッジと雛型の輪郭の比較による所持品候補箇所の決定、人物による遮蔽を考慮した所持品領域の検出を行う手法を提案し、所持品領域検出精度の向上を図る。

キーワード：所持品領域検出、人物の骨格、人物領域の雛型、エッジ

A Study on Region Detection of Personal Belongings Using Human Skeleton and Edges in an Image

MAKOTO MIURA^{1,a)} TORU ABE^{1,2,b)} TAKUO SUGANUMA^{1,2,c)}

Abstract: In order to observe crime scenes associated with personal belongings, such as luggage theft or leaving behind suspicious parcels, it is essential to detect personal belonging regions in images. Recently, existing approaches propose a three-step region detection. Initially, (1) they generate a template of the human region, then (2) determine candidate parts of personal belonging regions which are parts of the foreground in an image different to the generated template, and finally, (3) detect personal belonging regions in the determined candidate parts. However, they lack accuracy on detecting the region due to poor accuracy in each step. Therefore, in this article, we propose a highly accurate method for personal belonging region detection, which consists of generating a template of the human region from a human skeleton, determining candidate parts of personal belonging regions by comparing edges in an image with contours of the human region template, and detecting personal belonging regions in consideration of occlusions by the human body.

Keywords: Personal belonging region detection, human skeleton, human region template, edge

1. はじめに

駅や商業施設を始めとするさまざまな場所で、荷物の置

き引きや不審物の放置など人物の所持品が関係した犯罪が数多く報告されている [1]. その対策として、人物の所持品が関係した犯罪を監視カメラの映像等から自動で検出することが期待されており、近年ではその実現に向けての第一歩として、画像中から人物の所持品を検出することへ注目が集まっている [2].

画像中から所持品領域を検出する方法として、図1のようなアプローチが提案されている。このアプローチでは、

¹ 東北大学大学院情報科学研究科
Graduate School of Information Sciences, Tohoku University

² 東北大学サイバーサイエンスセンター
Cyberscience Center, Tohoku University

a) miura@ci.cc.tohoku.ac.jp

b) beto@tohoku.ac.jp

c) suganuma@tohoku.ac.jp

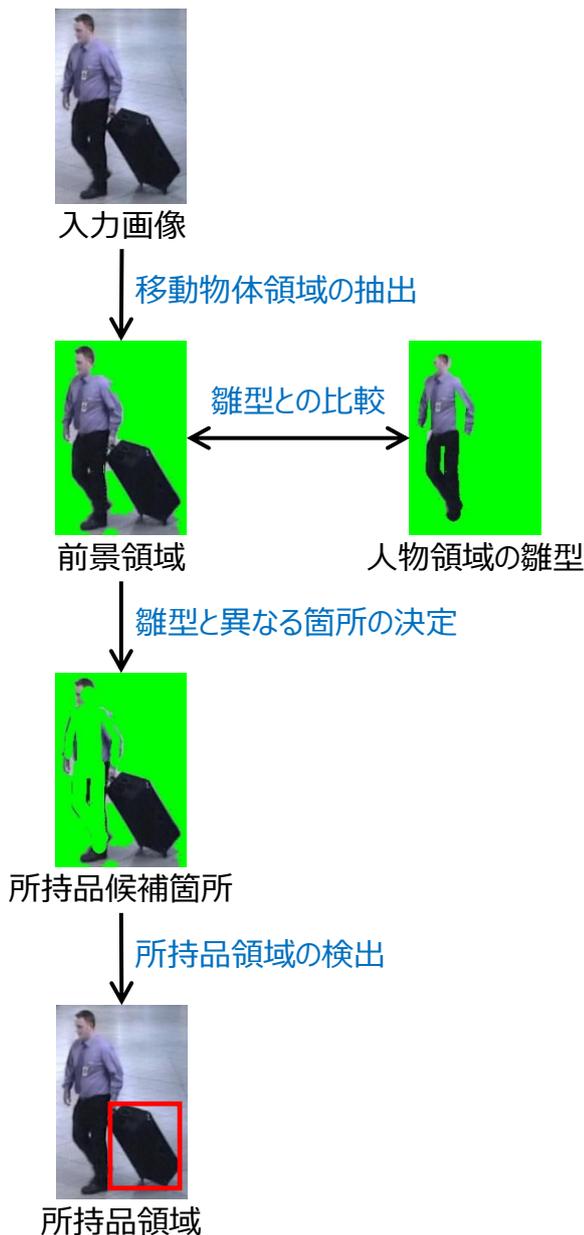


図 1: 所持品領域検出のアプローチ

まず、画像中での人物領域の雛型を生成する。次に、入力画像から抽出した移動物体領域(前景領域)と人物領域の雛型との比較を行い、人物領域の雛型と異なる前景領域の箇所を所持品候補箇所として決定する。最後に、決定された所持品候補箇所の特徴を分析することで、所持品領域を検出する。このアプローチに基づく手法は、様々な所持品の検出に適用可能であるものの、既存の手法 [2-4] では、主に以下の 3 つの課題により、所持品領域を高精度に検出することが困難である。

- 人物領域の雛型を生成する精度が不十分
- 前景領域から所持品候補箇所を決定する精度が不十分
- 候補箇所から所持品領域を検出する精度が不十分

そこで本研究では、これらの課題を解決するため、画像

中の人物の骨格とエッジを利用することで所持品領域の高精度な検出を図る手法を提案する。本稿では、提案手法の詳細と、提案手法の動作を検証するために行った実験について述べる。

2. 関連研究

先に述べたように、画像中から所持品領域を検出するため、人物領域の雛型を生成、前景領域から所持品候補箇所を決定、所持品候補箇所から所持品領域を検出という 3 つのステップに従うアプローチが提案されている。本章では、既存手法の概要と課題を各ステップに分けて述べる。

2.1 人物領域の雛型の生成

人物領域の雛型を生成するアプローチは大きく 2 つに分けられる。

1 つ目は、事前に用意した人物の画像を人物領域の雛型として利用するアプローチである [2, 3]。このアプローチでは、事前に用意した画像と入力画像で人物の姿勢が大きく異なる場合、人物領域の雛型と入力画像中の前景領域を正しく比較できず、所持品候補箇所の正確な決定が困難になるという問題がある。

2 つ目は、入力画像を解析して人物領域の雛型を生成するアプローチである [4, 5]。このアプローチは、あらかじめ用意した人物の画像を利用するのではなく、入力画像をその都度解析し人物領域の雛型を生成するため、人物のさまざまな姿勢に対応できるという利点がある。しかし、Tavanai の手法 [4] では、画像中の色情報をもとに人物領域を推定しており、人体の形状に関する情報を利用していないため、人物領域の雛型を正確に生成することは困難である。また、Xia らの手法 [5] では、Convolutional Neural Network(CNN) を用い、人体の形状と画像中の色に関する情報の両方を考慮した人物領域の推定を行っているが、人物の骨格を推定する CNN と人体の各部位の領域を推定する CNN を学習させるため、大量の学習データが必要という課題がある。

2.2 前景領域での所持品候補箇所の決定

Damen らの手法 [2] や Tavanai の手法 [4] では、前景領域のうち、人物領域の雛型と重なる部分を人物の一部とみなして除去し、残った部分を所持品候補箇所として決定する手法を用いている。これらの手法では、前景領域と人物領域の雛型の位置関係のみに基づいて判定を行っているため、人物領域の雛型を十分正確に生成できない場合は、所持品候補箇所の決定精度が大きく低下することになる。

一方、Ghadiri らは、エッジ抽出を行った前景領域を小領域(ブロック)に分割し、各ブロックに含まれているエッジの特徴を分析することで、ブロック毎に所持品候補箇所であるか否かを決定する手法を提案している [3]。この手

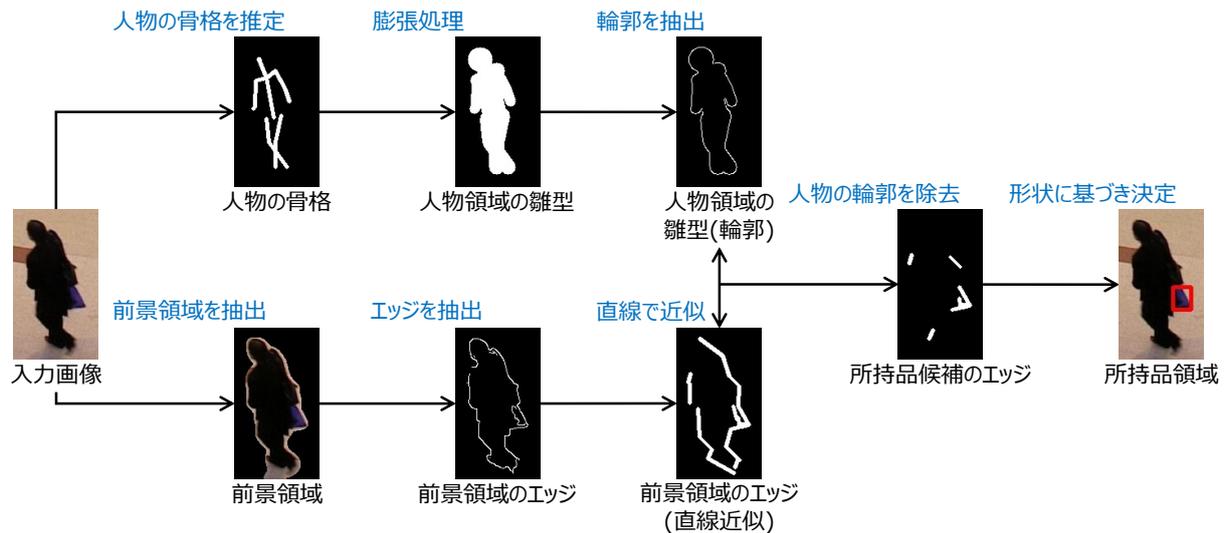


図 2: 提案手法の概要

法では、まず前景領域のエッジと人物領域の雛型の輪郭を格子状にブロックへ分割し、エッジ(輪郭)の傾きの分布である Shape Context [6] を各ブロックで求める。次に、前景領域と人物領域の雛型で対応するブロック毎に Shape Context の比較を行い、Shape Context が異なる前景領域のブロックを人物以外とみなし、そのブロック内のエッジを所持品候補箇所として決定する。以上のように、この手法では、所持品候補であるか否かの決定をブロック単位で行っているため、前景領域のブロック内に所持品のエッジと人物のエッジが共に含まれる場合、所持品候補のみを正確に決定することは困難である。

2.3 所持品候補箇所からの所持品の検出

Damen らや Ghadiri らは、所持品候補箇所の人物領域に対する位置に基づいて所持品候補箇所から所持品領域を検出する手法を提案している [2,3]。しかし、所持品の種類や持ち方等によって所持品の位置はさまざまに変化するため、これらの手法では所持品領域を高精度に検出することは困難であると考えられる。

一方、Tavanai は、所持品候補箇所のエッジが構成する多角形の形状に基づいて所持品候補箇所から所持品領域を検出する手法を提案している [4]。この手法では、人物が持ち運ぶことを想定し作成された所持品の形状は多くが凸状となることに着目し、エッジが構成する多角形の途切れが少なく形状が凸状に近いものほど所持品領域に対応する可能性が高いと判定している。しかし、所持品の一部が人体により遮蔽されると、エッジが構成する多角形の途切れが増え、凹状となる箇所も増えるため、所持品領域を正確に検出することが困難となる。

3. 提案

本稿では、画像中の人物の骨格とエッジに基づいて所持

品領域を検出する手法を提案する。提案手法の概要を図 2 に示す。提案手法では、以下に示す 3つのステップにより、人物領域の雛型の生成、前景領域からの所持品候補箇所の決定、所持品候補箇所からの所持品領域の検出をそれぞれ高精度に実現することで、既存手法の課題への対処を図る。

- 人物の骨格に基づく人物領域の雛型の生成
入力画像から推定した人物の骨格に対して膨張処理を施すことで人物領域の雛型を生成する。これにより、人物領域の多数のサンプルを用いた事前の学習を行うことなく、入力画像中の様々な状態の人物に対し、人物領域の雛型の高精度な生成を図る。
- 画像中のエッジと雛型の輪郭の特徴に基づく所持品候補箇所の決定
前景領域のエッジを直線で近似し、人物領域の雛型の輪郭と比較を行う。人物領域の雛型の輪郭と位置や傾きの特徴が類似した前景領域のエッジ(直線)を除去し、残ったエッジを所持品候補箇所として決定する。前景領域のエッジと雛型の輪郭の比較を直線単位で行い、比較の際に特徴として位置だけでなく傾きも考慮することで、所持品領域によるエッジのみを所持品候補箇所として決定する精度の向上を図る。
- 人体による遮蔽とエッジが構成する多角形の形状に基づく所持品領域の検出
所持品候補のエッジが構成する多角形の形状(凸の度合い・途切れの度合い)に基づいて所持品候補箇所から所持品領域を検出する。人物領域の雛型を参照し、人体により所持品が遮蔽されている状態を推定することで、所持品領域を検出する精度の向上を図る。

以下では、各ステップの詳細を述べる。

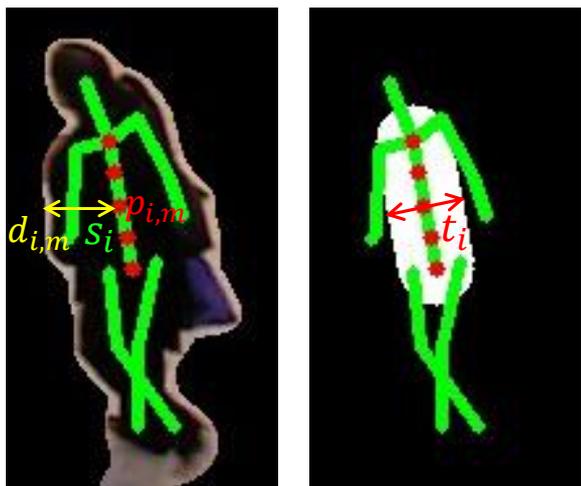


図 3: 骨格 s_i に対する膨張処理

3.1 人物の骨格に基づく人物領域の雛型の生成

人物領域の雛型を生成するため、まず人物の骨格を推定する。人物の骨格の推定には、OpenPose [7] を利用する。OpenPose により取得した人体の特徴点 15 個をもとに、各部位を直線で結ぶことで人物の骨格 12 個を推定する。

次に、人物の骨格に対して膨張処理を施すことで、人物領域の雛型を生成する。そのイメージを図 3 に示す。骨格を膨張させる太さ t_i は式 (1) に従って決定する。

$$t_i = \frac{\sum_{m=1}^M d_{i,m}}{M} \quad (1)$$

ここで、 $d_{i,m}$ は骨格 s_i を等分する M 個の点 $p_{i,m}$ から背景領域までの最小距離を表している。ただし、腕を構成する 4 つの骨格 (左右の上腕, 前腕) と、足を構成する 4 つの骨格 (左右の上腿, 下腿) については、それぞれ 4 つの骨格の太さの平均値を最終的な太さとして設定する。また、頭については、顔の中心点のみをもとに太さを決定する。

3.2 画像中のエッジと雛型の輪郭の特徴に基づく所持品候補箇所の決定

前景領域のエッジから人物に相当する部分を除去し、残ったエッジを所持品候補箇所と決定する。

まず、直線近似した前景領域のエッジ e_j 毎に、人物の雛型の輪郭と特徴 (位置・傾き) が類似しているか比較することで、 e_j が人物の輪郭であるか否かを判定する。そのイメージを図 4 に示す。具体的には、 e_j を等分する N 個の点 $p_{j,n}(x_{j,n}, y_{j,n})$ のうち、 T_q 個以上の点で、式 (2), (3) を満たす人物の雛型の輪郭上の点 $p(x, y)$ が存在する場合、 e_j は人物の輪郭であるとみなし除去する。

$$\begin{cases} |x - x_{j,n}| \leq T_d & (y = y_{j,n}), \\ |y - y_{j,n}| \leq T_d & (x = x_{j,n}) \end{cases} \quad (2)$$

$$\left| \frac{\mathbf{v}_{j,n} \cdot \mathbf{v}}{|\mathbf{v}_{j,n}| |\mathbf{v}|} \right| \leq T_a \quad (3)$$

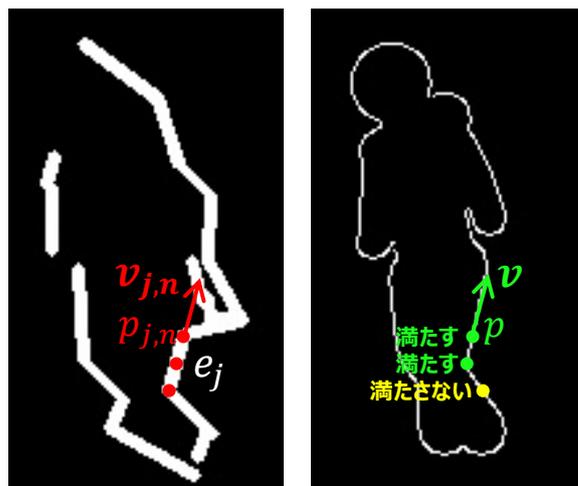


図 4: 前景領域のエッジ e_j が人物の一部であるかの判定

ここで、 $\mathbf{v}_{j,n}$, \mathbf{v} は前景領域のエッジ e_j 上の点 $p_{j,n}$, 人物領域の雛型の輪郭線上の点 p における接ベクトルを、 T_d , T_a は式 (2), (3) における閾値を各々表している。

また、頭部や足先周辺に所持品が存在する可能性は低いという仮定に基づき、頭部や足先周辺のエッジは除去する。

これらの処理を通して、最終的に残ったエッジを所持品候補箇所として決定する。

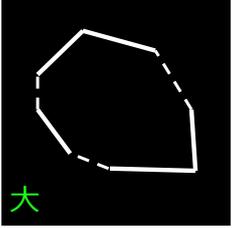
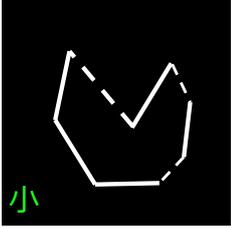
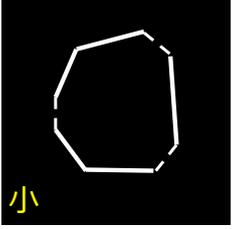
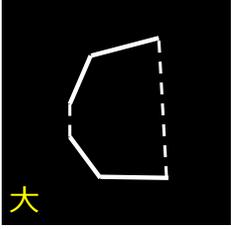
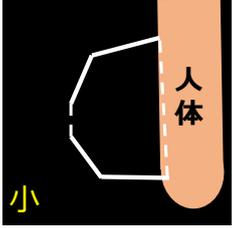
3.3 人体による遮蔽とエッジが構成する多角形の形状を考慮した所持品領域の検出

Tavanai の手法 [4] に基づき、所持品候補のエッジが構成する多角形の形状 (凸状の度合い・途切れの度合い) を考慮し、凸状の度合いが大きく途切れの度合いが小さい多角形を所持品と判定する。その際、提案手法では、人物領域の雛型を参照することで、所持品の一部が人体で遮蔽されてしまう場合へ対処する。そのイメージを表 1 に示す。表 1 において、実線は多角形を構成する所持品候補のエッジ、破線は途切れている辺を表している。

初めに、所持品候補のエッジ e'_i から多角形 P を構成する方法について述べる。そのイメージを図 5 に示す。まず全ての e'_i の両端点の重心 g を求める。次に、 g とある 1 つの端点を結ぶ線を基準線として、 g と残りの端点を結ぶ直線とのなす角度をそれぞれ求める。この時、各 e'_i において、角度が小さい側の端点を e'_i の始点とみなす。最後に、 g と各 e'_i の始点を結ぶ直線と基準線のなす角度が小さい順に各 e'_i を結ぶことで、多角形 P を構成する。

次に、人体による遮蔽を考慮して所持品候補のエッジ e'_i が構成する多角形 P の凸状の度合いを判定する方法について述べる。凸状の度合いを判定するため、まず所持品候補のエッジ e'_i に対する凸包 C を求める。この際、凸包 C と人物領域の雛型が重なっている領域を、所持品が人体で遮蔽されている領域 H とみなす。 P の面積 SP , C の面積 SC , H の面積 SH が式 (4) の関係を満たせば、凸状の度

表 1: エッジが構成する多角形の形状による所持品の判定

	所持品である	所持品でない
凸状の度合い		
		
途切れの度合い		
		

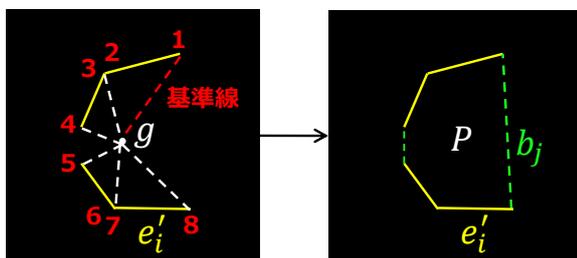


図 5: 所持品候補のエッジ e'_i から多角形 P を構成

合いが大きいと判定する。

$$\frac{SP}{SC - SH} \geq T_{convex} \quad (4)$$

ここで、 T_{convex} は式 (4) における閾値である。式 (4) によって、人体で遮蔽されている領域を除いて凸状の度合いを求めることで、所持品の一部が人体で遮蔽されて凹状と

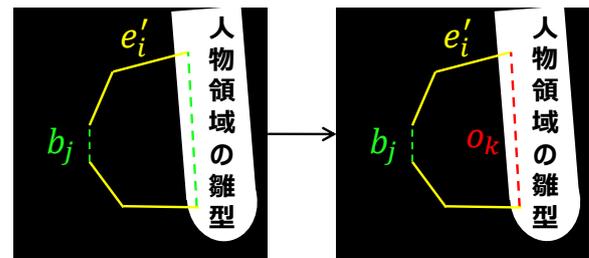


図 6: 人体による遮蔽が原因で途切れている辺の探索

表 2: 実験で設定した各パラメータの値

M	20
N	3
T_q	1
T_d	人物領域の雛型の高さの 15%
T_a	$\cos 15^\circ$
T_{convex}	0.85
$T_{connectivity}$	0.45

なる場合も凸状の度合いが大きいと判定することができる。

次に、人体による遮蔽を考慮して P の途切れの度合いを判定する方法について述べる。まず P を構成する途切れている辺 b_j から、人体による遮蔽が原因で途切れている辺 o_k を探索する。これは、 b_j のうち、人物領域の雛型と重なっている辺を o_k とみなすことで実現する。ただし、 P を構成する全ての辺が人物領域の雛型と重なっている場合は、 o_k は存在しないとする。そのイメージを図 6 に示す。 P を構成する辺が、式 (5) を満たせば、途切れの度合いが小さいと判定する。

$$\frac{Le}{Le + Lb + Lo} \geq T_{connectivity} \quad (5)$$

ここで、 Le は所持品候補のエッジ e'_i の長さの総和を、 Lb は途切れている辺 b_j の長さの総和を、 Lo は人体による遮蔽が原因で途切れている辺 o_k の長さの総和を表している。また、 $T_{connectivity}$ は式 (5) における閾値を表している。式 (5) によって、人体による遮蔽が原因で途切れている辺を除いて途切れの度合いを求めることで、所持品の一部が人体で遮蔽される場合も、途切れの度合いが小さいと判定することができる。

式 (4)、(5) を共に満たす多角形が存在する場合、その多角形を内部に含むような最小の矩形を、所持品領域として決定する。

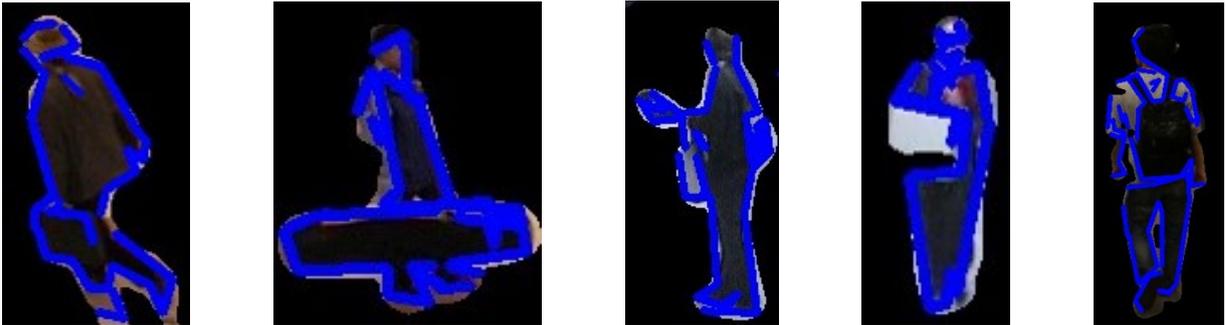
4. 実験

提案手法の動作を検証するため、PETS 2006 benchmark data [8], i-Lids dataset for AVSS 2007 [9], CMUSR: Surveillance Research Dataset [10] 中の画像を入力画像として用いて、所持品領域を検出する実験を行った。実験では、前景領域の抽出には背景差分手法の LOBSTER [11] を、エッジの抽出には Canny の手法 [12] を、エッジの直線

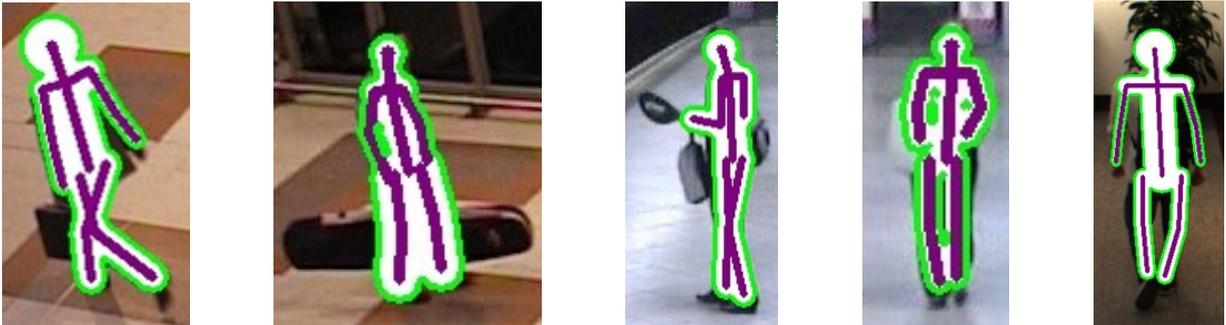
入力画像



前景領域とそのエッジ



人物の骨格と人物領域の雛型



所持品候補のエッジ



検出された所持品領域



(a)

(b)

(c)

(d)

(e)

図 7: 検出に成功した例

入力画像



前景領域とそのエッジ



人物の骨格と人物領域の雛型



所持品候補のエッジ



検出された所持品領域



(a)

(b)

(c)

(d)

(e)

図 8: 検出に失敗した例

近似には Kovesi の手法 [13] を用いた。また、各パラメータは表 2 の通りに設定した。

実験結果の例を図 7, 8 に示す。各図の 1 段目は入力画像を、2 段目は前景領域とそのエッジ (青い線) を、3 段目は人物の骨格 (紫の線) と人物領域の雛型 (緑の線に囲まれた白い領域) を、4 段目は所持品候補のエッジ (黄色い線) を、そして 5 段目は検出された所持品領域 (赤い矩形) を表している。図 7(a)~(e) は、所持品領域の検出に成功した例を示しており、この結果から分るように、様々な姿勢の人物や所持品に対して、人物領域の雛型の生成、前景領域からの所持品候補箇所の決定、所持品候補箇所からの所持品領域の検出の各ステップを概ね正確に実行できていることが確認できた。このことから、提案手法を所持品領域の検出に十分利用できると思われる。

しかし、提案手法により所持品領域を正確に検出できない例もいくつか見られた。その例を図 8 に示す。

図 8 (a), (b) では、人物の衣服の一部を所持品として誤検出されていることが確認された。これは、提案手法では人物の各骨格をそれぞれ一定の太さで膨張させ人物領域の雛型を生成しているため、人体に密着していない衣服は人物領域とみなされず所持品候補箇所として残ったことが原因であると考えられる (図 8 (a) ではコートの裾が、(b) ではスカートの裾が所持品候補箇所として判定されている)。

また、図 8 (c), (d) のように、服の模様やしわを所持品として誤検出した場合も確認された。提案手法では、前景領域のエッジから人物領域の雛型の輪郭と特徴が類似しているエッジを除去し残ったエッジを所持品候補箇所と判定しているが、この手法で除去できるのは人物の外側の輪郭だけであるため、人物の内側にある服の模様 ((c) のジャケットのベルト状の模様) やしわ ((d) の T シャツのしわ) が所持品領域として誤検出される可能性が高くなると考えられる。

さらに、図 8 (e) のように、実際は所持品 (人物が背負ったリュックサック) であるにも関わらず検出できていない場合も確認できた。これは、前景領域のエッジから人物の輪郭を除去する際に、人物の境界付近に位置している所持品の輪郭も一緒に除去されたことが原因であると考えられる。

5. おわりに

本稿では、画像中から所持品領域を検出する精度を向上させることを目的として、人物の骨格に基づく人物領域の雛型の生成、画像中のエッジと雛型の輪郭の特徴 (位置・傾き) に基づく所持品候補箇所の決定、人物による遮蔽とエッジが構成する多角形の形状を考慮した所持品領域の検出を行う手法を提案した。提案手法の動作を検証するために行った実験では、所持品領域の検出に失敗した場合もいくつか見られたが、様々な姿勢の人物や所持品に対して、

所持品領域を正確に検出できていることが確認できた。このことから、提案手法を所持品領域の検出に十分利用できると思われる。

今後は、提案手法をさらに洗練するため、今回の実験において検出に失敗した場合へ対処する方法の検討・実装を進める予定である。その例として、所持品候補箇所から所持品領域を検出する際に人物の骨格と所持品の位置関係を考慮することや、前後のフレームに対する検出結果を用いて現在のフレームに対する検出結果を補正するなどの処理の追加を考えている。また、評価用のデータセットを用いて、提案手法と既存手法の検出精度の比較を行う予定である。

参考文献

- [1] 警察庁: 犯罪統計 | 警察庁 Web サイト, available from <https://www.npa.go.jp/publications/statistics/sousa/statistics.html>.
- [2] Damen, D. and Hogg, D.: Detecting carried objects from sequences of walking pedestrians, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 34, No. 6, pp. 1056–1067 (2012).
- [3] Ghadiri, F., Bergevin, R. and Bilodeau, G.-A.: Carried object detection based on an ensemble of contour exemplars, *Eur. Conf. Comput. Vision*, Vol. VII, pp. 852–866 (2016).
- [4] Tavanai, A.: Tracking in the context of interaction, PhD Thesis, School of Computing, The University of Leeds (2016).
- [5] Xia, F., Wang, P., Chen, X. and Yuille, A.: Joint multi-person pose estimation and semantic part segmentation, *CoRR*, Vol. abs/1708.03383 (2017).
- [6] Belongie, S., Malik, J. and Puzicha, J.: Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 24, No. 4, pp. 509–522 (2002).
- [7] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Real-time multi-person 2D pose estimation using part affinity fields, *CoRR*, Vol. abs/1611.08050v2 (2017).
- [8] ISCAPS consortium: PETS 2006 benchmark data, available from <http://www.cvg.reading.ac.uk/PETS2006/data.html>.
- [9] AVSS2007: i-Lids dataset for AVSS 2007, available from http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html.
- [10] Hattori, K., Hattori, H., Ono, Y., Nishino, K., Itoh, M., Boddeti, V. and Kanade, T.: CMUSR: Surveillance research dataset, available from <http://www.consortium.ri.cmu.edu/projSRD.php>.
- [11] St-Charles, P.-L. and Bilodeau, G.-A.: Improving background subtraction using local binary similarity patterns, *IEEE Winter Conf. Appl. Comput. Vision*, pp. 509–515 (2014).
- [12] Canny, J.: A computational approach to edge detection, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-8, No. 6, pp. 679–698 (1986).
- [13] Kovese, P. D.: MATLAB and Octave functions for computer vision and image processing, available from <https://www.peterkovese.com/matlabfns/>.