

修士論文概要書

Summary of Master's Thesis

Date of submission: 01 / 30 /2018 (MM/DD/YYYY)

専攻名（専門分野） Department	情報理工・情報通信専攻	氏 名 Name	加藤 君丸	指 導 教 員 Advisor	渡辺 裕 印 Seal
研究指導名 Research guidance	オーディオビジュアル情報処理	学籍番号 Student ID number	CD 5116F024-3		
研究題目 Title	姿勢情報を用いたカメラ映像からの歩きスマホ検出				

1. 研究目的

近年、スマートフォンの普及とともに、歩きながらスマートフォンを操作する「歩きスマホ」が増えている。特に、駅のホームにおける歩きスマホは、他の歩行者との衝突や線路への転落の原因となり、危険である。我々は、鉄道運行の安全性向上のため、監視カメラからの歩きスマホ検出の実現を目指している。本論文では、姿勢推定を用いた動画からの歩きスマホ検出手法について述べる。

2. 従来の研究

新村ら[1]は、車載カメラの映像による歩きスマホ認識手法を提案している。本手法では、歩行者の全身画像から抽出した HOG 特徴量を学習済みの SVM に入力し、歩きスマホと通常の歩行に分類する。評価実験において、人物が正面向きの場合では 88.8%、右向きでは 73.2%と高い正解率が得られたが、右向きでは 40.9%と正解率が低かった。

本論文では、HOG+SVM による手法を従来手法として、提案する歩きスマホ認識手法と比較を行う。

3. 提案手法

3.1 はじめに

姿勢推定を用いた歩きスマホ検出手法を提案する。姿勢推定のアルゴリズムとして、Realtime Multi-Person 2D Pose Estimation [2]を用いる。本研究では、歩きスマホ検出を、スマホ使用姿勢検出、歩行検出、把持物体認識の3ステップによって行う。図1に提案手法の概要図を示す。また、3.2 - 3.4 で各ステップについて述べる。

3.2 スマホ使用姿勢検出

スマホ使用姿勢検出では、姿勢情報から特徴ベクトルを作成し、SVM によって“スマホ使用”と“スマホ不使用”に分類する。特徴ベクトルの作成手法として、部位座標を位置とスケールについて正規化する手法(部位座標による手法)、肘・肩の関節角度を用いる手法(関節角度による手法 a)、肘・肩・首の関節角度を用いる手法(関節角度による手法 b)の三つを検討する。

部位座標による手法では、姿勢推定によって得られた人物の各部位の画像におけるピクセル座標(部位座標)を位置・スケール正規化を行い、特徴ベクトルと

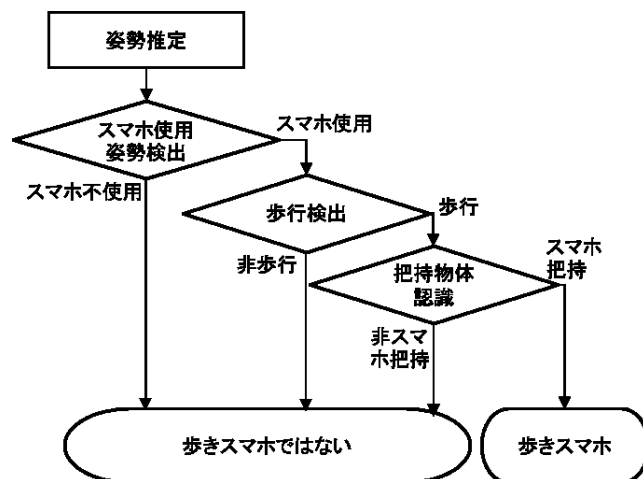


図 1: 提案手法の概要図

する。位置正規化は、人物の首の部位座標が基準となるように正規化を行う。スケール正規化は、人物の両肩の中間座標と首の座標の間の長さが 1 となるように正規化を行う。

関節角度による手法 a では、姿勢推定によって得られた部位座標から関節の角度を計算し、特徴として用いる。用いる関節角度は肘・肩のものである。関節角度は、注目する関節と、隣り合う二つ関節の部位座標から計算される。例えば、右肘の関節角度は画像における右手首-右肘-右肩のなす角度である。

関節角度による手法 b では、関節角度による手法 a と同様に関節角度を特徴として用いる。用いる関節角度は、肘・肩・首のものである。

3.3 歩行検出

歩行検出のステップでは、スマホ使用姿勢検出において“スマホ使用”とされた人物を、複数フレームの姿勢情報から、“歩行”と“非歩行”に分類する。

まず、複数フレームの姿勢情報を用いることから、フレーム間での人物の追跡を行う。追跡は、姿勢情報から得られる人物領域について、色相ヒストグラムの類似度をもとにフレーム間マッチングを行うこととなる。

その後、人物の複数フレームの姿勢情報をもとに、歩行検出を行う。歩行検出手法は、2 フレーム間の足首座標の変化に閾値をかける手法(歩行検出手法 a)、複数フレームの関節角度を SVM により分類する手法(歩行検出手法 b)、関節角度の変化と関節間の長さの

変化率を Random Forest で分類する手法(歩行検出手法 c)の三つを検討する..

歩行検出手法 a では, 左右の足首について, 2 フレーム間での画像におけるピクセル座標の変化に閾値をかけ, 歩行を検出する. 前フレームからの足首の x 座標の変化を C_x , y 座標の変化を C_y , 閾値を T としたとき, 左右片方または両方の足首が

$$\frac{\sqrt{C_x^2 + (W \times C_y)^2}}{S} > T \quad (3.1)$$

を満たす場合, 歩行として検出する. W はデータセットにおける足首の x 座標と y 座標の変化量の比を表す定数である. S は両尻の中間の座標と首との距離であり, 人物の姿勢のスケールを表す定数である.

歩行検出手法 b では, 人物の複数フレームにおける膝・尻の関節角度を特徴とし, SVM により分類する. フレームの数は, 9 を上限として人物を追跡できる限りのフレームを用いる.

歩行検出手法 c では, 複数フレームの時系列上で隣り合う各 2 フレームについて, 人物の関節角度の変化と関節間の長さの変化率を算出し, これを特徴として Random Forest で分類する. 関節角度の変化とは, 2 フレーム間での膝・尻における角度の差である. 関節間の長さの変化率は, 膝・足首間及び尻・膝間について計算する. フレーム n における関節間の長さを l_n としたとき, 変化率 p_n は,

$$p_n = \frac{l_n - l_{n-1}}{l_n} \quad (3.2)$$

によって得られる. 用いるフレームの数は歩行検出手法 b と同様である.

3.4 把持物体認識

最後に, 歩行検出にて“歩行”とされた人物について, 把持物体認識を行う. 把持物体認識では, 人物の手周辺領域の画像認識により, スマホ把持・非把持を判別する.

まず, 両肩の x 座標の大小から人物が正面向きであるかを判定し, 正面向きの場合, 把持物体認識を適用する. これは, 手周辺領域が身体の奥に隠れている場合, 画像認識が正しく行えないためである.

次に, 姿勢情報をもとに人物の手周辺領域を切り出し, CNN による画像認識を行う. これにより, “スマホ把持”と“非スマホ把持”に分類する. また, 把持物体認識が適用できない場合には, “スマホ把持”と同様の扱いとする.

把持物体認識の結果, “スマホ把持”とされた人物を歩きスマホとして検出する.

3.5 おわりに

本章では, 3 ステップからなる歩きスマホ検出手法を提案し, 各ステップについて述べた.

4. 実験

スマホ使用姿勢検出手法と HOG+SVM による手法の比較実験を行った. データセットとして, 16 人の歩行画像 768 枚を用いた. 交差検定法に基づき, 1 人分の画像を評価に, 残りを学習に用いる処理を 16 人

それぞれに行い, 結果を平均する. 結果を表 1 に示す.

表 1: スマホ使用姿勢検出手法と従来手法の比較

手法	Precision	Recall	F1-score
HOG+SVM	0.861	0.823	0.842
部位座標	0.966	0.969	0.967
関節角度 a	0.971	0.956	0.963
関節角度 b	0.955	0.984	0.969

従来手法と比較すると, 提案手法はいずれも優れた性能を示し, 関節角度による手法 b においては F1-score で+12.7%の性能向上がみられた.

また, 歩行検出の各手法と, 把持物体認識の適用の有無についても比較した. 評価用のデータセットとして, 人物の歩行動画から 5783 フレームを用いた. 結果を表 2 に示す.

表 2: 歩きスマホ検出の性能比較

把持物体認識	歩行検出手法	Precision	Recall	F1-score
あり	a	0.835	0.829	0.847
	b	0.778	0.810	0.794
	c	0.812	0.896	0.852
なし	a	0.777	0.861	0.817
	b	0.726	0.842	0.779
	c	0.755	0.931	0.834

表 2 から, 歩行検出手法 c を用い, 把持物体認識を適用した場合に 0.852 と最も高い F1-score を示すことが確認できた. 歩行検出手法を比較すると, Precision においては歩行検出手法 a が, Recall と F1-score においては歩行検出手法 c が優れている. また, 把持物体認識を適用した場合, 適用しない場合と比較して Precision と F1-score は向上し, Recall は低下していることが分かる.

5. 結論

本研究では, 姿勢推定を用いた歩きスマホ検出手法を提案した. 実験にて, 歩きスマホ認識において提案手法が従来手法より優れていることを確認した. また, 動画からの歩きスマホ検出実験においては, F1-score で 0.852 と, 高い精度での検出が可能であることを確認した.

参考文献

- [1] 新村文郷, 川西康友, 出口大輔, 井手一郎, 村瀬洋, 藤吉弘亘, "車載カメラ画像からの「スマホ歩き」認識に基づく歩行者の不注意度推定," 信学技報, PRMU2015-46, vol. 115, no. 98, pp. 83-88, 2015.
- [2] Z. Cao, T. Simon, S.-E. Wei and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *CVPR*, 2017.

2017 年度 修士論文

姿勢情報を用いたカメラ映像からの歩きスマホ検出

Texting While Walking Detection from Camera Image
Using Pose Estimation

提出日 2018 年 1 月 30 日

指導教員 渡辺 裕 教授

早稲田大学大学院 基幹理工学研究科

情報理工・情報通信専攻

5116F024-3

加藤君丸

目次

第 1 章 序論	1
1.1 研究の背景	1
1.2 本研究の目的	2
1.3 本論文の構成	2
第 2 章 関連研究	4
2.1 まえがき	4
2.2 CONVOLUTIONAL POSE MACHINES	4
2.3 REALTIME MULTI-PERSON POSE ESTIMATION	5
2.4 人物行動認識	6
2.5 歩きスマホ認識・検出	6
2.6 むすび	7
第 3 章 姿勢情報によるスマホ使用姿勢検出	8
3.1 まえがき	8
3.2 姿勢推定	8
3.3 部位座標を特徴とした分類	9
3.4 関節角度を特徴とした分類	11
3.5 実験	12
3.5.1 従来手法との比較実験	12
3.5.2 下半身の隠れに対する性能検証	17
3.5.3 歩行していない人物への適用	19
3.5.4 スマートフォンを持っていない人物への適用	19
3.6 むすび	21
第 4 章 歩行検出	22
4.1 まえがき	22

4.2 人物の追跡	22
4.3 足首座標による歩行検出.....	23
4.4 関節角度による歩行検出.....	24
4.5 関節角度の変化と関節間の長さの変化率による歩行検出	25
4.6 実験	27
4.7 むすび.....	29
第5章 把持物体認識	30
5.1 まえがき	30
5.2 把持物体認識.....	30
5.3 実験	32
5.4 むすび.....	36
第6章 姿勢情報を用いた歩きスマホ検出	37
6.1 まえがき	37
6.2 概要	37
6.3 実験	38
6.4 むすび.....	39
第7章 結論	40
謝辞	42
参考文献	43
図一覧.....	46
表一覧.....	47
研究業績	49
受賞	50

第1章

序論

1.1 研究の背景

鉄道は、日本において重要な公共交通機関の一つである。平成 28 年度における鉄道旅客数量は 246.0 億人 [1]と、平成 19 年度からほぼ横這いの数字であり、今後も安定して多量の利用が予想される。このように、鉄道は広く利用されており、その運行において、安全対策は重要である。

鉄道における人身事故発生の原因として、「歩きスマホ」がある。歩きスマホとは、歩きながらスマートフォン（スマホ）の画面を注視し、操作をする行為である。近年、スマートフォンの普及にともない、歩きスマホが増加している。2014 年 12 月に電気通信事業者協会によって行われた調査 [2]では、歩きスマホが以前と比べて増えていると認識している人は全体の 85.5%である。

また、電気通信事業者協会の調査 [2]では、歩きスマホの危険性が表れた複数のデータが報告されている。調査の回答者のうち、自分が歩きスマホをしていて他人にぶつかった経験がある人は 4.5%，ぶつかりそうになった経験がある人は 33.7%である。更に、ぶつかりそうになった、あるいは実際にぶつかった場所として、20.5%の人が駅のホームと回答している。同様に、歩きスマホをしている人にぶつかられた、ぶつかられそうになった経験があると回答した人は全体の 68.3%であり、その場所として駅のホームと回答した人は 24.6%である。

その他、歩きスマホをしている歩行者の視野調査 [3]では、周囲の環境認識が全くできておらず、人によっては蛇行してしまうケースがあることを確認している。また、国土交通省の報告では、2014 年度には 32 人がスマホ・携帯電話使用中にホームから転落している [4]。

以上のように、歩きスマホが他の歩行者との衝突やホームからの転落につながる事例がある。これらは、鉄道車両との接触のような命にかかわる重大事故に発展する可能性があり、駅ホームでの歩きスマホは危険であるといえる。

こういったホームからの転落事故を防ぐための対策として、ホームドアの設置が進められている。ホームドアとは、駅のホームと線路の境界に設置される柵であり、車両のドアに

合わせて開閉される扉を持つ。ホームドアが設置された駅の数には 2006 年度末の時点で 318 であったが、2016 年度末の時点では 686 と、10 年で倍増している [5]。しかし、日本には 2017 年 7 月の時点で 9256 の駅があり [6]、ホームドアが設置されている駅の数は全体の 1 割に満たない。利用客の多い主要な駅を中心に設置が進められているものの、一駅あたり数億円といった多大な設置費用や大規模な工事の必要性から、2016 年 3 月の時点で、利用客 10 万人以上の駅への普及率は 3 割にとどまっている [7]。このように、ホームドアによる安全対策はコストが大きく、普及が十分に進んでいないという問題点がある。また、歩きスマホによる歩行者どうしの衝突を防ぐことはできない。

以上のような背景から、本研究では動画からの歩きスマホ検出をめざす。

1.2 本研究の目的

駅ホームにおける安全対策として、監視カメラの映像から歩きスマホをしている人物を自動検出するアルゴリズムは有効である。阪神電気鉄道株式会社と KDDI 株式会社による「阪神×au 歩きスマホ防止キャンペーン」[8]では、実際に駅にて歩きスマホをしている人物に対してアナウンスで注意を行い、歩きスマホを止めさせるといった内容の啓発動画が公開された。これは実際の事例をもとに再現して撮影された動画であり、アナウンスによる注意は有効であると思われる。歩きスマホ検出は、検出された人物に対し指向性を持ったスピーカーを用いたアナウンスにて注意を与えることで、歩きスマホを止めさせるといった使い方が考えられる。

また、監視カメラを用いたシステムはホームドアの設置と比較してコストが軽く、早く普及させることが可能である。ホームドアの設置がなされていない駅における線路への転落事故防止や、歩行者どうしの衝突防止のために、監視カメラからの歩きスマホ検出が有効と予想される。

以上のように、駅ホームにおける安全確保のための有効性が考えられることから、本研究では、動画から歩きスマホをしている人物を検出する手法の実現を目的としている。本論文では、姿勢情報によるスマホ使用姿勢検出、歩行検出、把持物体認識の 3 ステップによる歩きスマホ検出手法について述べる。

1.3 本論文の構成

以下に本章以降の構成を示す。

第1章 本章であり、研究の背景及び目的について述べている。

第2章 関連研究について述べる。まず、姿勢推定のアルゴリズムである Convolutional Pose Machines [9]及び Realtime Multi-Person Pose Estimation [10]の概要を述べる。次に、画像や動画から人物の行動認識を行う研究について述べる。最後に、

歩きスマホ認識に関する研究について述べる.

第3章 姿勢情報から, 静止画における人物のスマートフォン使用姿勢を検出する手法について述べる.

第4章 歩きスマホ検出のため, 姿勢情報を用いて人物の歩行を検出する手法について述べる.

第5章 歩きスマホ検出の **Precision** 向上のため, 把持物体認識を行う手法について述べる.

第6章 第3章から第5章の内容をふまえ, 動画から歩きスマホをしている人物を検出する手法について述べる.

第7章 本研究のまとめを行う.

第2章

関連研究

2.1 まえがき

本章では，本研究の関連研究について述べる．

本章の構成としては，まず，本研究で使用した姿勢推定のアルゴリズムとして，Convolutional Pose Machines 及び Realtime Multi-Person Pose Estimation について述べる．次に，人物行動認識の研究について述べる．最後に，歩きスマホに関する研究について述べる．

2.2 Convolutional Pose Machines

Convolutional Pose Machines (CPMs) [9]は，静止画から人物の姿勢を推定するアルゴリズムである．CPMs は，人物の肩，肘，足首などといった身体的な部位について，画像におけるピクセル位置を得る．人体の P 個の部位を想定した場合，その p 番目の部位のピクセル位置は， $\mathbf{Y}_p \in \mathbf{Z} \in \mathbb{R}^2$ と表せる．ここで， $\mathbf{Z}$ は画像における全ての座標 (u, v) の集合である．CPMs の目的は， P 個全ての部位に対してその画像上の座標 $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_P)$ を推定することである．CPMs はこれを Convolutional Neural Network (CNN) を階層的に積み上げることで実現している．

CPMs のアルゴリズムは，複数のステージによって構成される．各ステージ $t \in 1, \dots, T$ において，CNN による推定器 g_t は，各部位 p と背景に対し，各 $\mathbf{z} \in \mathbf{Z}$ について， $\mathbf{Y}_p = \mathbf{z}$ である確からしさを推定し， $\mathbf{b}_t \in \mathbb{R}^{w \times h \times (P+1)}$ と表せる belief map (confidence map と呼ばれる) を作成する．ここで， w は画像の幅， h は画像の高さである．このとき，ステージ 1 においては $\mathbf{z}$ 周辺の画像局所特徴のみをもとに推定を行うが，ステージ 2 以降においては，前のステージにおいて作成された belief map を追加の情報として用い，推定を行う．

CPMs の入力と出力の例を図 2-1 に示す．左の画像を入力すると，人物の各部位の画像

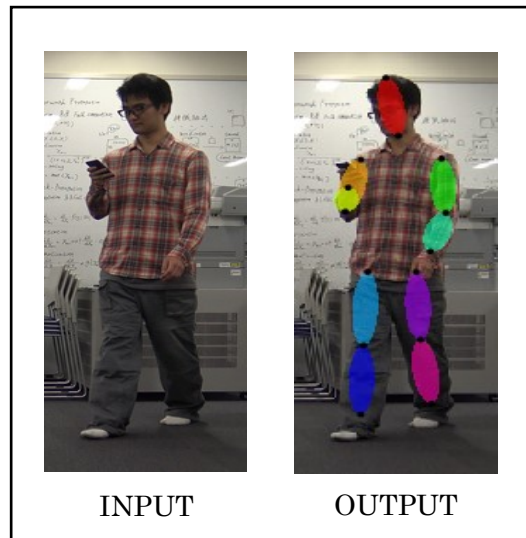


図 2-1: CPMs の入力と出力の例

におけるピクセル座標が得られる。右図の各黒点は、得られた部位の座標を描画したものである。

CPMs は、一つの入力画像に対して、1 人の人物の姿勢推定を行う。一度の処理で複数の人物の姿勢を推定することはできない。

2.3 Realtime Multi-Person Pose Estimation

Realtime Multi-Person Pose Estimation [10]は、CPMs を発展させた手法である。CPMs が一度の処理で 1 人の人物のみの姿勢推定を行うのに対し、Realtime Multi-Person Pose Estimation は複数の人物の姿勢推定が可能である。

Realtime Multi-Person Pose Estimation では、一つのステージで二つの CNN を用いる。一方の CNN は、CPMs と同様に belief map の推定を行う。もう一方の CNN は、Part Affinity Fields (PAFs)の推定を行う。PAFs は、画像上での部位どうしの関連（位置と方向性）を、二次元ベクトル場の集合として表現したものである。各ステージでは、前ステージにて得られた belief map と PAFs の両方を各 CNN に追加情報として入力し、推定を行う。

複数の人物の姿勢を推定する場合、部位の位置を推定するだけでは、それがどの人物の部位なのか見分けがつかない。部位どうしの関連を同時に推定し、部位の接続を行うことで、複数人物の姿勢推定が可能となる。

2.4 人物行動認識

映像から人物の行動を認識するアルゴリズムとして、Two-Stream Convolutional Networks [11]がある。これは、2つのCNNを並行して動かし、最後にその出力を統合して動画における人物の行動を認識する。2つのCNNは、それぞれRGB画像とオプティカルフローを入力とする。RGB画像を入力とするCNNは、動画の現フレームにおける画像特徴を抽出する。それに対して、オプティカルフローを入力とするCNNは、複数フレームでの動きの特徴を抽出する。これらを統合することで、画像特徴と動きの特徴から人物の行動を認識するのである。

また、Two-Stream Convolutional Networks に対して、より局所的な特徴を利用しようとした手法が Deep lOcal Video Features (DOVF) [12]である。DOVFは、CNNの中間層から特徴抽出し、Fisher Vectors (FV)などのグローバル特徴にエンコードすることで、局所的な特徴の利用を検討している。DOVFは、人物行動の動画データセットである HMDB51 [13]及び UCF101 [14]にて実験を行い、Two-Stream Convolutional Networks より良い性能を示している。

このように、動画に対してディープラーニングを用いて行動認識を行う研究はいくつかある。しかし、これらの研究は、動画へのラベル付けを目的としたものである。動画における各人物に対しての行動認識ではないため、歩きスマホ検出にそのまま適用できるものではない。

また、姿勢推定アルゴリズムを用いて人物の行動を解析する手法もいくつか存在する。工藤ら [15]は、人物の姿勢を推定するCNNから抽出した姿勢特徴をLSTM [16]に入力することで人物の行動認識をする手法を提案している。一原ら [17]は、Realtime Multi-Person Pose Estimation によって得られた姿勢情報を、SVM またはニューラルネットワークを用いて分類することで、人物の行動認識を行う手法を検討している。黒瀬ら [18]は、テニスのストロークを対象とし、Convolutional Pose Machines による姿勢特徴をもとにクラスタリングを行い、評価する手法を提案している。

2.5 歩きスマホ認識・検出

ここでは、人物の歩きスマホ検出・認識に関する研究を紹介する。

人物の歩きスマホを検出する手法として、持っているスマートフォンのセンサを活用したものがある。根岸 [15]は、スマートフォンの加速度センサの値を取得し、閾値によって歩行検知を行うアプリケーションを用いた実験を行っている。名坂ら [20]は、スマートフォンの加速度センサの情報と、カメラ映像から算出したオプティカルフローを用い、使用者の歩きスマホを検出する手法を提案している。また、KDDI から歩きスマホ注意アプリ [17]が公開されるなど、実用化されたアプリケーションも存在する。

しかし、使用者の歩きスマホを検出して警告を表示するアプリケーションは、使用者に不満を与え、使用を中止させてしまう可能性が高い。それに対して、スマートフォンのセンサの情報から周囲の危険度を予測し、警告の頻度を減らす研究がなされている。野田口ら [22] は、スマートフォンのカメラから前方の障害物を検出し、スマートフォン使用者に提示することで、衝突を回避する手法を提案している。同じく、平川ら [23] は、スマートフォンのカメラと加速度センサから、衝突が起こりそうな状況が生じた場合のみ使用者に警告する手法を提案している。

このように、スマートフォンのセンサを用いた歩きスマホ検出手法はいくつか提案されているが、スマートフォンへの導入は任意である。使用者の不満を減らす手法も提案されているが、依然として普及率は高くない。駅ホームにおける歩きスマホ検出を考えた場合、各々のスマートフォンのセンサ情報を用いることは難しいといえる。よって、監視カメラなどの駅に設置されたセンサを用いた歩きスマホ検出が必要となる。

カメラの映像を用いた歩きスマホ認識・検出手法はいくつか存在する。皆本ら [20] は、盲導犬ロボットへの搭載を目的とし、カメラ映像からの歩きスマホ検出手法を提案している。実験から、HOG [21] とパーティクルフィルタを用いた手法が最も良いことを確認している。本手法は、歩行者の手の位置を追跡するパーティクルフィルタの停滞時間とパーティクルフィルタの重心による判定に加え、歩行者の全身画像の HOG 特徴量を SVM [22] によって分類した結果を統合し判断する。実験にて、適合率、再現率ともに 90% という優れた結果を確認した。しかし、盲導犬ロボットへの搭載を目的としていることから、人物の正面向きの映像のみを対象とした手法であるため、あらゆる方向の人物が想定される監視カメラの映像に適用することは難しい。

新村ら [23] は、車載カメラの映像からの歩きスマホ検出をめざし、HOG 特徴量による歩きスマホ認識手法を提案している。本手法は、人物の全身画像の HOG 特徴量を SVM によって分類し、歩きスマホ認識を行っている。実験では、人物の正面向き、右向き、左向きの画像を用意し、それぞれ学習と認識を行っている。その結果、正面向きでは 88.8%、右向きでは 73.2% と高い正解率が得られたが、右向きでは 40.9% と正解率が低かった。この結果について、データセットの人物は右手にスマートフォンを持っていたため、右手が身体に隠れてしまいカメラから見えにくくなっていたことを原因としてあげている。

2.6 むすび

本章では、本研究に関連する各研究について述べた。本研究の目的に最も合ったものとして、新村らの HOG と SVM による手法 [23] が挙げられる。次章では、姿勢推定によるスマホ使用姿勢検出手法について述べ、HOG と SVM による手法との比較を行う。

第3章

姿勢情報によるスマホ使用姿勢検出

3.1 まえがき

本章では，歩きスマホ認識手法として，姿勢情報をもとに人物がスマートフォンを使用している姿勢をとっているかを判別する手法について述べる．

本章の構成としては，まず，姿勢推定について述べる．その後，得られた姿勢情報の処理方法として，部位座標による分類手法と関節角度による分類手法について述べる．その後，実験と結果をもとにした考察を述べる．最後に，本章のまとめを行う．

3.2 姿勢推定

スマホ使用姿勢検出の第一段階として，姿勢推定を行う．姿勢推定のアルゴリズムとして，Realtime Multi-Person Pose Estimation [10]を用いる．Realtime Multi-Person Pose Estimation はその実装を公開 [24]している．同時に，MPII Human Pose dataset (以下 MPI) [25]による学習済みモデル(以下 MPI モデル)と COCO 2016 keypoints challenge dataset (以下 COCO) [26]による学習済みモデル(以下 COCO モデル)が公開されている．本章では MPI モデルを用いる．

MPI モデルによって推定されるのは，15 の部位の画像におけるピクセル座標である．同様に，COCO モデルでは 18 の部位の座標が得られる．表 3.1 に MPI モデルと COCO モデルで得られる部位の一覧を示す．

MPI モデルにより対象の人物の姿勢推定を行い，姿勢情報 $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_{15})$ を得る．ただし， $\mathbf{Y}_p = (u_p, v_p)$ は表 3-1 中の番号 p にあたる部位の画像におけるピクセル座標であり， u_p はその x 座標， v_p はその y 座標を表す．得られた $\mathbf{Y}$ を特徴ベクトルへと変換処理し，SVM に入力することによって人物がスマートフォンを使用しているかを分類する．次節以降，変換処理の手法について述べる．

表 3-1: Realtime Multi-Person Pose Estimation によって推定される部位

番号	MPI モデルで推定される部位	COCO モデルで推定される部位
1	Head (頭)	Nose (鼻)
2	Neck (首)	Neck (首)
3	RShoulder (右肩)	RShoulder (右肩)
4	RElbow (右肘)	RElbow (右肘)
5	RWrist (右手首)	RWrist (右手首)
6	LShoulder (左肩)	LShoulder (左肩)
7	LElbow (左肘)	LElbow (左肘)
8	LWrist (左手首)	LWrist (左手首)
9	RHip (左尻)	RHip (左尻)
10	RKnee (左膝)	RKnee (左膝)
11	RAnkle (左足首)	RAnkle (左足首)
12	LHip (右尻)	LHip (右尻)
13	LKnee (右膝)	LKnee (右膝)
14	LAnkle (右足首)	LAnkle (右足首)
15	Chest (腰)	REye (右目)
16		LEye (左目)
17		REar (右耳)
18		LEar (左耳)

3.3 部位座標を特徴とした分類

本節では、姿勢推定によって得られた部位座標を正規化して SVM で分類する手法について述べる。用いる姿勢情報は、MPI モデルで得られた Chest を除く 14 の部位の座標 $\mathbf{Y}_1 \sim \mathbf{Y}_{14}$ である。

姿勢推定によって得られた部位座標は、画像上のピクセル座標であるから、サンプルごとに基準位置及びスケールが異なる。そこで、部位座標に対し、位置正規化とスケール正規化を行う。

基準位置については、首の位置が基準となるように処理を行う。各部位座標 $\mathbf{Y}_p = (u_p, v_p)$ に対し、

$$\mathbf{Y}_p' = (u_p - u_2, v_p - v_2) \quad (3.1)$$

とすることで、位置正規化された座標 $\mathbf{Y}_p' = (u_p', v_p')$ を得る。

スケール正規化は、右尻と左尻の中間座標と首の間の長さを姿勢情報のスケール s とし、これが 1 となるように正規化を行う。 s は、

$$s = \sqrt{\left(u_2' - \frac{u_9' + u_{12}'}{2}\right)^2 + \left(v_2' - \frac{v_9' + v_{12}'}{2}\right)^2} \quad (3.2)$$

によって得られる．これを用い，

$$Y_p'' = \left(\frac{u_p'}{s}, \frac{v_p'}{s}\right) \quad (3.3)$$

により，スケール正規化を行う．

図 3-1 の入力に対して得られた姿勢情報を正規化した例を表 3-2 に示す．スケール正規化された部位座標 Y'' を SVM にて学習・分類することで，人物がスマートフォンを使用しているかを推定する．

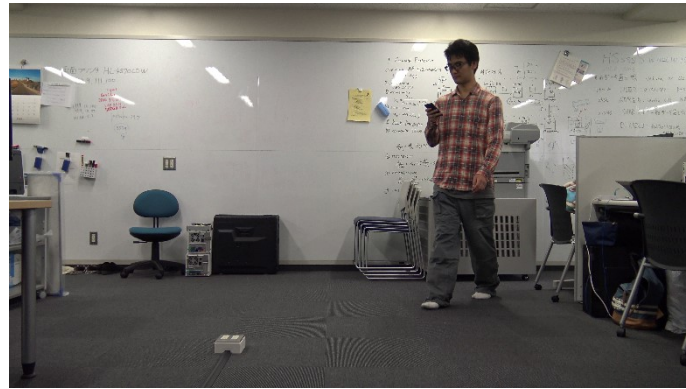


図 3-1: 入力画像

表 3-2: 正規化された部位座標の例

部位	元データ Y		位置正規化 Y'		スケール正規化 Y''	
	u	v	u'	v'	u''	v''
Head (頭)	2511.36	205.586	-46.81	-240.791	-0.08398	-0.43198
Neck (首)	2558.17	446.377	0	0	0	0
RShoulder (右肩)	2446.74	545.972	-111.43	99.595	-0.19991	0.178675
RElbow (右肘)	2370.89	774.547	-187.28	328.17	-0.33598	0.588741
RWrist (右手首)	2365.11	680.922	-193.06	234.545	-0.34635	0.420776
LShoulder (左肩)	2727.85	551.856	169.68	105.479	0.304408	0.189231
LElbow (左肘)	2733.57	792.324	175.4	345.947	0.31467	0.620633
LWrist (左手首)	2669.18	956.502	111.01	510.125	0.199153	0.91517
RHip (左尻)	2458.46	1003.54	-99.71	557.163	-0.17888	0.999557
RKnee (左膝)	2446.75	1303.18	-111.42	856.803	-0.19989	1.537114

RAnkle (左足首)	2417.72	1613.89	-140.45	1167.513	-0.25197	2.094532
LHip (左尻)	2628.24	1003.64	70.07	557.263	0.125706	0.999736
LKnee (左膝)	2645.94	1291.36	87.77	844.983	0.15746	1.515909
LAnkle (左足首)	2686.66	1567.32	128.49	1120.943	0.230513	2.010985

3.4 関節角度を特徴とした分類

本節では、関節角度を特徴とし、SVM にて分類する手法について述べる。関節角度による分類手法として、手法 a と手法 b の二つを提案する。

まず、姿勢推定によって得られた部位座標から、関節角度を求める。求める関節角度と、表 3-2 の $\mathbf{Y}$ から算出された関節角度の例を表 3-3 に示す。関節角度は、対象となる関節位置の部位座標 $\mathbf{Y}_l$ と、隣接する 2 つの部位の座標 $\mathbf{Y}_m, \mathbf{Y}_n$ から計算される。まず、二次元ベクトル $\mathbf{A} = \overrightarrow{\mathbf{Y}_l \mathbf{Y}_m}$, $\mathbf{B} = \overrightarrow{\mathbf{Y}_l \mathbf{Y}_n}$ を定義する。このとき、 $\mathbf{Y}_m \mathbf{Y}_l \mathbf{Y}_n$ のなす角 r_l は、

$$r_l = \arctan\left(\frac{\mathbf{A} \times \mathbf{B}}{\mathbf{A} \cdot \mathbf{B}}\right) \quad (3.4)$$

によって求められる。求められた関節角度を SVM に入力することで、人物がスマートフォンを使用しているかを推定する。

表 3-3: 手法と用いる関節角度

手法 a			手法 b		
関節位置	求める角	関節角度 r の例	関節位置	求める角	関節角度 r の例
右肘	右手首・右肘・右肩	0.382	右肘	右手首・右肘・右肩	0.382
右肩	右肘・右肩・左肩	-1.870	右肩	右肘・右肩・首	-2.621
左肩	右肩・左肩・左肘	-1.616	首 (右側)	右肩・首・頭	2.108
左肘	左肩・左肘・左手首	-2.744	首 (左側)	頭・首・左肩	2.319
			左肩	首・左肩・左肘	-2.151
			左肘	左肩・左肘・左手首	-2.744

関節角度を特徴とした分類は、上半身の部位のみを用いている。このことから、全身の部位座標を特徴とした分類と比較して、隠れなどの要因により下半身の部位座標を推定できない場合に強いと考えられる。

3.5 実験

3.5.1 従来手法との比較実験

部位座標による手法，関節角度による手法 a, b と，従来手法による比較実験を行う．従来手法として，HOG+SVM による手法と CNN による手法を用いる．

データセットとして，16 人の人物による歩行動画から，フレームを静止画像として抽出して用いる．歩行動画は，人物が 45° 毎に 8 つの方向に向かって歩く様子を撮影したものである．人物 1 人につき，歩きスマホをしている動画と通常の歩行の動画の 2 系列を撮影した．図 3-2 に撮影された動画のフレームの例を示す．また，図 3-3 にフレームに対する姿勢推定の結果の例を示す．各動画から，人物の歩く方向 1 つにつき 3 枚のフレームを抽出する．歩きスマホをしている動画から抽出された画像は“スマホ使用”クラス，通常の歩行の動画から抽出された画像は“スマホ不使用”クラスとして扱う．よって，3 フレーム $\times$ 8 方向 $\times$ 16 人 $\times$ 2 クラス = 768 枚の画像が抽出される．このデータセットを用い，各手法の 2 クラス分類の性能を比較する．

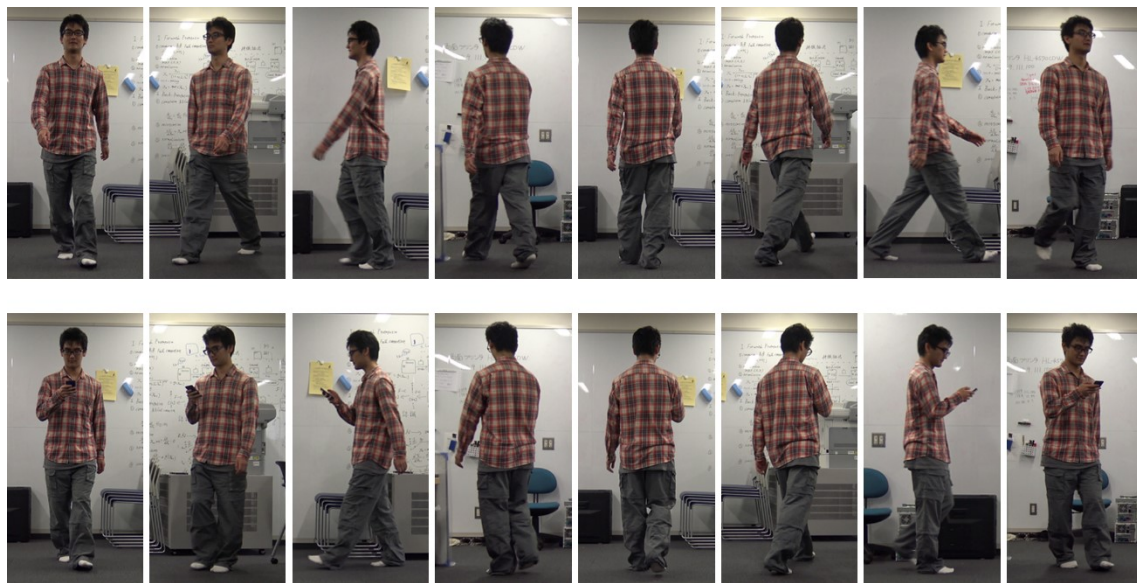


図 3-2: データセットに用いた動画のフレームの例(人物領域のみ)

上: 通常の歩行 下: 歩きスマホ

人物の向きは左から 0° , 45° , 90° , 135° , 180° , 225° , 270° , 315°

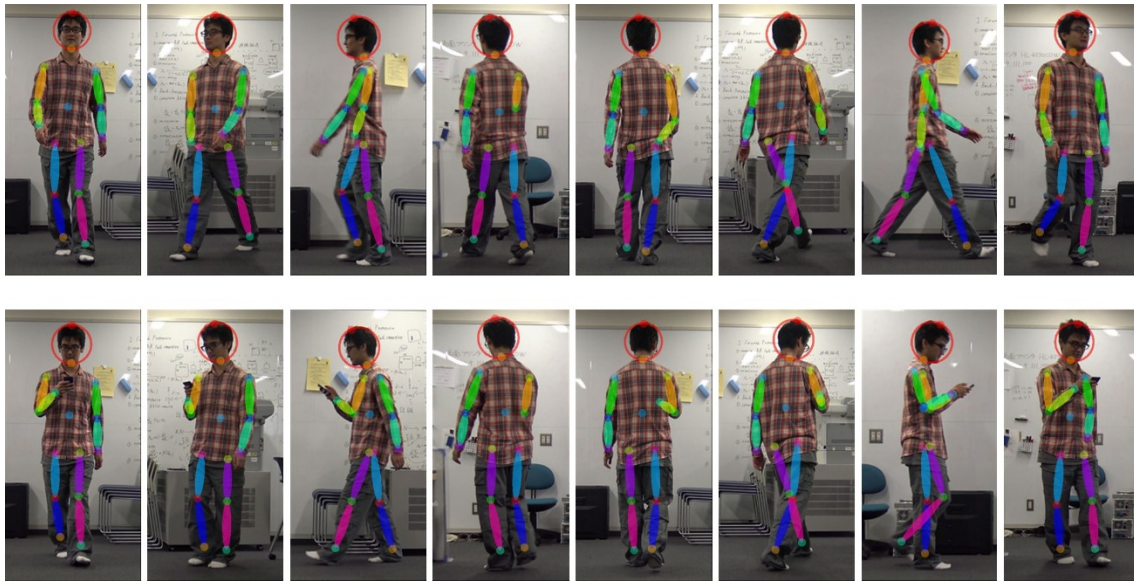


図 3-3: 姿勢推定の結果例（人物領域のみ）

上: 通常の歩行 下: 歩きスマホ

人物の向きは左から 0° , 45° , 90° , 135° , 180° , 225° , 270° , 315°

実験では、交差検定法を用いる．1 人分のデータをテストに，残りの 15 人分のデータを学習に用いる．これを各人物のデータが一回テストに用いられるよう，16 回行う．

HOG と SVM による手法では，画像の人物領域から HOG 特徴量を抽出し，SVM によって分類する．まず，画像の人物領域を幅 64 ピクセル，高さ 128 ピクセルにリサイズし，HOG 特徴量の抽出を行う．HOG 特徴抽出器は，OpenCV [27]による実装を用いる．次に，抽出された HOG 特徴量を SVM によって学習・分類する．HOG 特徴抽出器，SVM のパラメータをそれぞれ表 3-4，表 3-5 に示す．

表 3-4: HOG+SVM による手法における HOG 特徴抽出器のパラメータ

セルサイズ	8*8 ピクセル
ブロックサイズ	16*16 ピクセル
ブロックストライド	8*8 ピクセル
ヒストグラムビン数	9

表 3-5: HOG+SVM による手法における SVM のパラメータ

カーネル	RBF
C	50
γ	0.001

また，一般物体認識などの分野で近年 HOG+SVM に代わり主流となっている CNN による手法を比較手法として用いる．CNN として，GoogLeNet [28]を用いる．実装には Chainer [29]を用いる．精度向上のため，学習データの水増し(データオーギュメンテーション)を行う．データオーギュメンテーションに用いた手法とパラメータのリストを表 3-6 に示す．このデータオーギュメンテーションにより，学習データは 90 倍の枚数となる．また，学習のパラメータを表 3-7 に示す．

表 3-6: CNN による手法におけるデータオーギュメンテーションの手法とパラメータ

ガンマ値による画像の明るさ変化	$\gamma=[0.5, 0.75, 1.0, 1.5, 2.0]$
左右反転	[反転あり, 反転なし]
回転	回転角= $[-5^\circ, 0^\circ, 5^\circ]$
スケール変化	倍率= $[0.95, 1.00, 1.05]$

表 3-7: CNN による手法における学習パラメータ

バッチサイズ	32
学習レート	0.01
学習反復回数	30000

部位座標を特徴とした手法では，画像から推定された人物の部位座標を正規化し，SVM により分類する．学習の際には，学習データから得られた姿勢情報に加え，それらを左右反転したデータを用いる．左右反転した姿勢における部位座標 Y_p''' は，スケール正規化された各部位座標 $Y_p'' = (u_p'', v_p'')$ の x 座標を表す u_p'' の正負を反転させることで得られる．よって，

$$Y_p''' = (-u_p'', v_p'') \quad (3.5)$$

である．また，SVM のパラメータを表 3-8 に示す

表 3-8: 部位座標による手法における SVM のパラメータ

カーネル	RBF
C	10
γ	5

関節角度を特徴とした手法では，画像から推定された人物の部位座標をもとに関節角度を計算し，SVM により分類する．手法 a では肘と肩，手法 b では肘と肩に加えて首の関節角度を用いる．手法 a で用いる SVM のパラメータを表 3-9 に，手法 b で用いる SVM のパラメータを表 3-10 に示す．

表 3-9: 関節角度による手法 a における SVM のパラメータ

カーネル	RBF
C	1
γ	1

表 3-10: 関節角度による手法 b における SVM のパラメータ

カーネル	RBF
C	5
γ	3

以上の 5 手法により実験を行った。評価指標として、Precision, Recall, F1-score を用いる。各手法の分類結果を表 3-11~表 3-15 に示す。また、“スマホ使用”クラスにおける各手法の比較を表 3-16 に示す。

表 3-11: HOG+SVM の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	0.830	0.867	0.848
スマホ使用	0.861	0.823	0.842

表 3-12: CNN の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	0.937	0.930	0.933
スマホ使用	0.930	0.938	0.934

表 3-13: 部位座標による手法の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	0.969	0.966	0.967
スマホ使用	0.966	0.969	0.967

表 3-14: 関節角度による手法 a の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	0.956	0.971	0.964
スマホ使用	0.971	0.956	0.963

表 3-15: 関節角度による手法 b の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	0.984	0.953	0.968
スマホ使用	0.955	0.984	0.969

表 3-16: “スマホ使用”クラスにおける各手法の性能比較

手法	Precision	Recall	F1-score
HOG+SVM	0.861	0.823	0.842
CNN	0.930	0.938	0.934
部位座標	0.966	0.969	0.967
関節角度 a	0.971	0.956	0.963
関節角度 b	0.955	0.984	0.969

Precision においては関節角度による手法 a が, Recall と F1-score においては関節角度による手法 b が最も良い結果を示した. HOG+SVM による従来手法と比較すると, 関節角度による手法 b は F1-score で+12.7%とより高い性能を示した. 部位座標による手法と関節角度による両手法を F1-score で比較すると, 差は 1%未満であり, ほぼ同等の性能を有すると評価できる.

また, 関節角度による手法 b における人物の向きと誤分類数を表 3-17 に, 誤分類された人物画像の例を図 3-4 に示す. 人物の向きは図 3-2 の例と同様である. 表 3-17 から, 135°の向き(左奥向き)の人物に誤分類が多いことが分かる. これは, 右利きの人物の右手が身体に隠れてしまうことが原因として考えられる. 特に, 図 3-4 下段右の例ではスマートフォンを持つ右手がほぼ完全に隠れてしまい, 正しく姿勢を推定できていない.

表 3-17: 関節角度による手法 b における人物の向きと誤分類数

人物の向き(°)	正解クラス		合計
	スマホ不使用	スマホ使用	
0	0	0	0
45	0	0	0
90	3	1	4
135	7	4	11
180	3	0	3
225	1	1	2
270	3	0	3
315	1	0	1

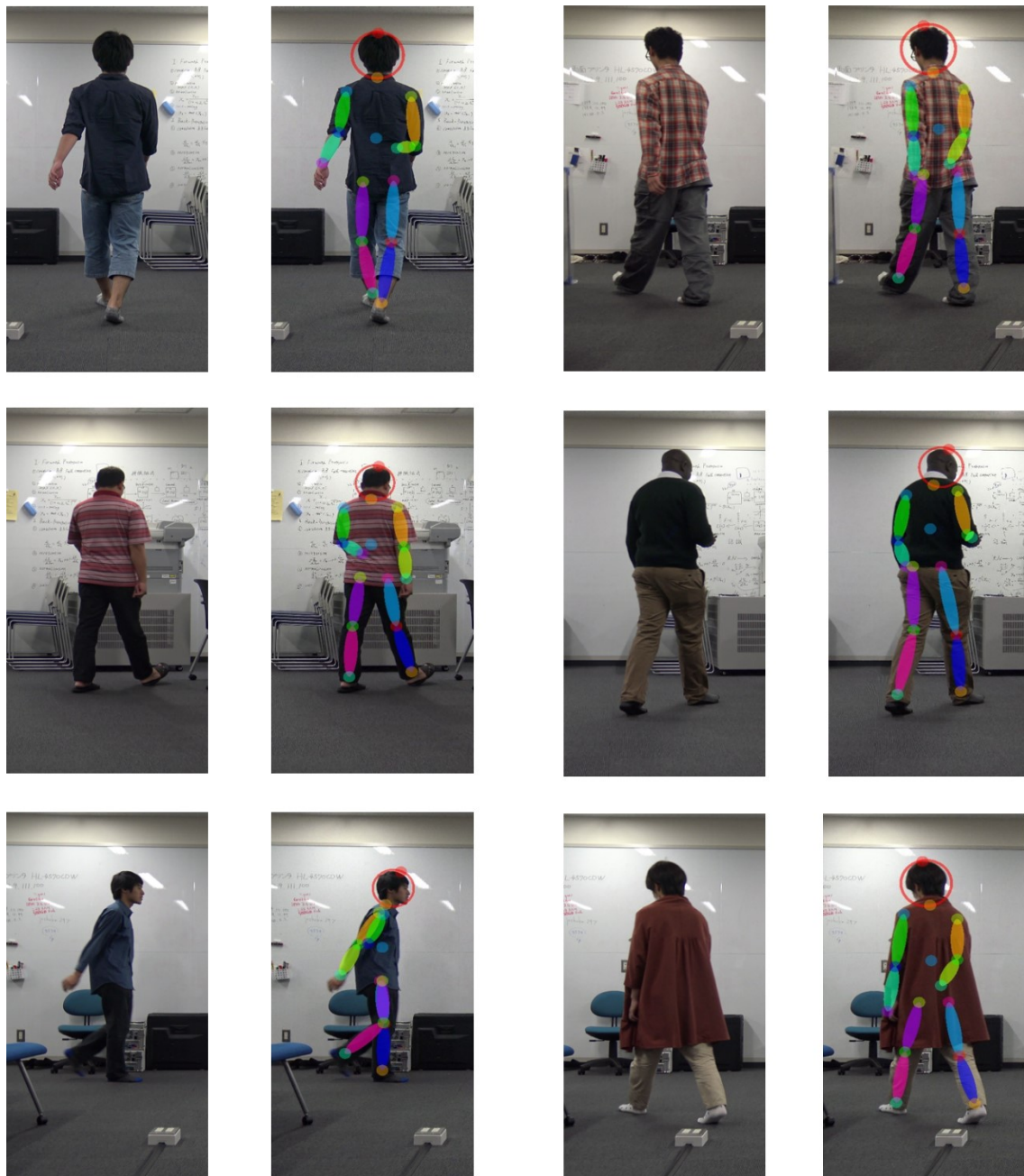


図 3-4: 誤分類された画像と姿勢推定の例

左 2 列: “スマホ不使用” 画像の誤分類の例 右 2 列: “スマホ使用” 画像の誤分類の例

3.5.2 下半身の隠れに対する性能検証

各手法の人物の下半身が隠れている状況における性能を比較し、関節角度による手法の優位性を検証する。比較する手法は部位座標による手法、関節角度による手法 a、関節角度による手法 b の 3 手法である。

評価用データとして、下半身が隠れた人物の歩行動画のフレームを用いた。フレーム数は、“スマホ使用”が216フレーム，“スマホ不使用”が183フレームである。学習データとして、3.5.1において用いたデータセットから8人分の画像を使用した。また、SVMのパラメータも3.5.1と同じものを用いた。

実験の結果を表3-18～表3-20に示す。なお、 $\text{True Positive} + \text{False Negative} = 0$ の場合のPrecisionは1とした。



図 3-5: 下半身が隠れた人物の画像例

左: スマホ不使用 右: スマホ使用



図 3-6: 下半身が隠れた人物の姿勢推定の例

左: スマホ不使用 右: スマホ使用

表 3-18: 部位座標による手法の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	1.000	0.000	0.000
スマホ使用	0.541	1.000	0.702

表 3-19: 関節角度による手法 a の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	0.971	0.721	0.828
スマホ使用	0.806	0.981	0.885

表 3-20: 関節角度による手法 b の分類結果

クラス	Precision	Recall	F1-score
スマホ不使用	1.000	0.694	0.819
スマホ使用	0.794	1.000	0.885

表 3-21: “スマホ使用”クラスにおける各手法の性能比較

手法	Precision	Recall	F1-score
部位座標	0.541	1.000	0.702
関節角度 a	0.806	0.981	0.885
関節角度 b	0.794	1.000	0.885

結果より、下半身が隠れている状況においては、部位座標による手法と比較して関節角度による手法が優位であることが示された。部位座標による手法は、全フレームの人物を“スマホ使用”と分類しており、下半身の隠れの影響が大きいことが分かる。また、関節角度による手法においても、下半身が隠れる状況においては Precision の低下が確認された。これは、下半身と同様に手首が隠れ、姿勢推定が難しかったことが原因であると考えられる。

3.5.3 歩行していない人物への適用

本手法は、人物がスマートフォンを使用しているかを検出している。人物が歩行しているかを考慮していないことから、本手法をそのまま歩きスマホ検出手法として適用すると、歩いていない人物を歩きスマホとして検出してしまうと予想される。確認のため、実験を行った。

評価用データとして、立ち止まってスマートフォンを操作している人物の動画から抽出したフレームを用いる。図 3-7 に画像と姿勢推定の例を示す。

動画から 1947 フレームに関節角度による手法 a を適用した。学習データとパラメータは 3.5.2 と同様である。その結果、全体の 91.0%にあたる 1771 フレームで人物が“スマホ使用”と分類された。スマートフォンを使用しているかどうかを検出する場合は問題ないが、これをそのまま歩きスマホ検出としてしまうと、歩いていない人物を歩きスマホとして検出してしまうことになる。よって、歩きスマホ検出のためには、人物の歩行検出が必要である。歩行検出については第 4 章で述べる。

3.5.4 スマートフォンを持っていない人物への適用

本手法は、人物のスマホ使用姿勢を検出する手法である。人物がスマートフォンを使用していない場合でも、類似する姿勢であれば、“スマホ使用”と検出されると予想される。確



図 3-7: 立ち止まってスマートフォンを使用する人物の画像と姿勢推定の例



図 3-8: 手提げ袋を持った人物の画像と姿勢推定の例

認のため、実験を行う。

歩きスマホに類似した姿勢の画像として、手提げ袋を提げて歩く人物の動画から抽出したフレームを用いる。図 3-8 に用いた動画のフレームと姿勢推定の例を示す。図 3-8 の人物は右手が胸の高さに来ており、スマートフォンを使用する姿勢に類似しているといえる。

動画から 300 フレームに関節角度による手法 a を適用した。学習データとパラメータは 3.5.2 と同様である。その結果、正解率 0.227 と非常に低い値となった。このことから、スマホ使用姿勢検出のみでは、スマートフォンを持っていない、類似した姿勢の人物を誤検出してしまう可能性が高いといえる。これを解決するためには、人物がスマートフォンを持っ

ているかを認識する把持物体認識が必要である。把持物体認識については第 5 章で述べる。

3.6 むすび

本章では、歩きスマホ検出を目的とし、姿勢推定を用いたスマホ使用姿勢の検出手法を提案した。また、実験により従来手法との比較を行った。実験結果より、姿勢推定を用いた手法はいずれも従来手法より優れていることが示された。更に、F1-score の比較から、姿勢推定を用いた手法には大きな性能の差がないことを確認した。

また、関節角度による手法が部位座標による手法と比較して下半身の隠れに強いことを確かめた。部位座標による手法は下半身の隠れにより分類がほぼ不可能となったのに対し、関節角度による手法は、性能の低下は見られたものの、高い精度を保った。

更に、スマホ使用姿勢検出のみによる歩きスマホ検出では、歩行していない人物やスマートフォンを持っていない人物に適用した場合に誤検出する場合があることを確認した。このことから、歩きスマホ検出の実現には歩行検出及び把持物体認識が必要であるといえる。これをふまえ、第 4 章では歩行検出について述べる。また、第 5 章では把持物体認識について述べる。

第4章

歩行検出

4.1 まえがき

第3章にて、姿勢推定を用いたスマホ使用姿勢検出手法を提案したが、人物の歩行の有無を考慮したものではなかった。そのため、歩きスマホ検出のためには、歩行検出をあわせて行う必要がある。本章では、歩きスマホ検出のための歩行検出手法について述べる。

本章の構成としては、まず、複数フレームにわたる人物の追跡について述べる。次に、2フレーム間の足首座標の変化に注目した歩行検出手法について述べる。その後、複数フレームの下半身の関節角度を特徴とし、SVMによって歩行検出する手法について述べる。そして、関節角度の変化と関節間の長さの変化率を特徴とし、Random Forestによって歩行検出する手法について述べる。最後に、実験とまとめを行う。

4.2 人物の追跡

歩行検出においては、複数フレームにわたる人物の姿勢情報を用いる。そのため、フレーム間での人物の追跡を行う必要がある。各フレームにおける人物の位置は姿勢推定によって得られる。よって、2フレーム間での人物の同定を行うことで、追跡が可能となる。2フレーム間の人物の同定のため、人物領域の色の類似度を用いる手法について述べる。

まず、姿勢情報から人物領域を求める。ここでの人物領域は、人物の姿勢に外接する矩形と定義する。すなわち、姿勢情報 $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_{14})$ において x 座標の最小値・最大値が $u_{min} \cdot u_{max}$ 、 y 座標の最小値・最大値が $v_{min} \cdot v_{max}$ であるとき、 (u_{min}, v_{min}) , (u_{max}, v_{min}) , (u_{max}, v_{max}) , (u_{min}, v_{max}) の四点を頂点とする長方形領域を人物領域 $\mathbf{R}$ とする。

次に、人物領域 $\mathbf{R}$ の色情報を抽出する。ここでは、色相ヒストグラムを用いる。 $\mathbf{R}$ の領域をHSV画像に変換し、Hueのチャンネルをもとに32ビンのヒストグラムを生成する。

その後、フレーム間で人物領域同士のヒストグラムの比較を行う。ヒストグラムの比較は、

Correlation によるスコアを用いる。Correlation は二つのヒストグラム H_1, H_2 において,

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}} \quad (4.1)$$

によって算出される [30]。ただし,

$$\bar{H}_k = \frac{1}{N} \sum_J H_k(J) \quad (4.2)$$

であり, N はビンの総数である。Correlation が大きいほど, 二つのヒストグラムは類似しているといえる。

フレーム n において l 個の人物領域があるとき, それを $R_{n,1}, \dots, R_{n,l}$ とする。同様に, フレーム $n-1$ における m 個の人物の領域を $R_{n-1,1}, \dots, R_{n-1,m}$ とする。また, $R_{a,b}$ の色相ヒストグラムを $H_{a,b}$ とする。この時, フレーム n における各人物領域について, フレーム $n-1$ の各人物領域とのスコアを計算する。つまり, $D = (d(H_{n,1}, H_{n-1,1}), d(H_{n,1}, H_{n-1,2}), \dots, d(H_{n,1}, H_{n-1,m}), d(H_{n,2}, H_{n-1,1}), \dots, d(H_{n,l}, H_{n-1,m}))$ を計算する。

D における最大値が $d(H_{n,c}, H_{n-1,d})$ であるとき, $R_{n,c}$ と $R_{n-1,d}$ は同一の人物を表す領域であるとする。その後, D から $H_{n,c}, H_{n-1,d}$ をもとに計算された $d(H_{n,c}, H_x), d(H_x, H_{n-1,d}) \forall x$ を取り除いた D' を用い, 同様にしてスコアが最大となる人物領域の組み合わせを同一人物によるものと認める。これを D が空になるまで l 回, あるいは m 回繰り返す, 人物の同定を行う。フレーム n においてフレーム $n-1$ との同定が行われなかった人物領域は新たな人物の領域, フレーム $n-1$ において同定が行われなかった人物領域は画面から消えた人物の領域であると考え。

時系列上での一つ前のフレームとの同定を続けることで, 各フレームにおける特定の人物の領域を追跡することができる。次節以降では, 追跡によって得られた複数フレームの姿勢情報から, 人物の歩行を検出する手法について述べる。

4.3 足首座標による歩行検出

ここでは, 2 フレーム間の足首の座標の変化による歩行検出手法について述べる。これを歩行検出手法 a とする。左右の足首についてそれぞれ 2 フレーム間での座標の変化を計算し, 片方, あるいは両方の足首の座標変化が閾値を超えた場合, 歩行として検出する。

ここで, 人物のフレーム n における姿勢情報を $Y_n = (Y_{n,1}, \dots, Y_{n,14})$ とする。ただし $Y_{n,p} = (u_{n,p}, v_{n,p})$ であり, p は表 3-2 の MPI モデルによる推定での番号に対応する。このとき, 右足首の x 座標の変化 $C_{r,x}$ 及び y 座標の変化 $C_{r,y}$ は,

$$\begin{cases} C_{r,x} = u_{n,11} - u_{n-1,11} \\ C_{r,y} = v_{n,11} - v_{n-1,11} \end{cases} \quad (4.3)$$

である。同様に左足首は

$$\begin{cases} c_{l,x} = u_{n,14} - u_{n-1,14} \\ c_{l,y} = v_{n,14} - v_{n-1,14} \end{cases} \quad (4.4)$$

によって座標変化を計算する．歩行の検出は，閾値 T を用いて，

$$\begin{cases} \text{Walking} & \left(\text{if } \frac{\sqrt{c_{l,x}^2 + (W \times c_{l,y})^2}}{s} > T \text{ or } \frac{\sqrt{c_{r,x}^2 + (W \times c_{r,y})^2}}{s} > T \right) \\ \text{Not walking} & (\text{else}) \end{cases} \quad (4.5)$$

によって行われる．ただし， W は歩行時の足首の x 座標と y 座標の変化の比率であり，学習データから計算される．また， s は式(3.2)によって計算される姿勢のスケールである．

4.4 関節角度による歩行検出

本節では，複数フレームにわたる関節角度情報を SVM により歩行と非歩行に分類する手法を述べる．これを歩行検出手法 **b** とする．

まず，注目するフレームとそこから過去の $N-1$ フレーム，合計 N フレームについて，人物の下半身の関節角度を計算する．求める関節角度は表 4-1 に示す．角度の計算方法は 3.4 と同様である．

表 4-1: 用いる関節角度のリスト

関節位置	求める角
右膝	右足首-右膝-右尻
右尻	右膝-右尻-左尻
左尻	右尻-左尻-左膝
左膝	左尻-左膝-左足首

その後， N 個のフレームについて計算した関節角度を一つのベクトルに統合し，SVM にて“歩行”・“非歩行”の 2 クラス分類を行う．関節角度による歩行検出の概要図を図 4-1 に示す．

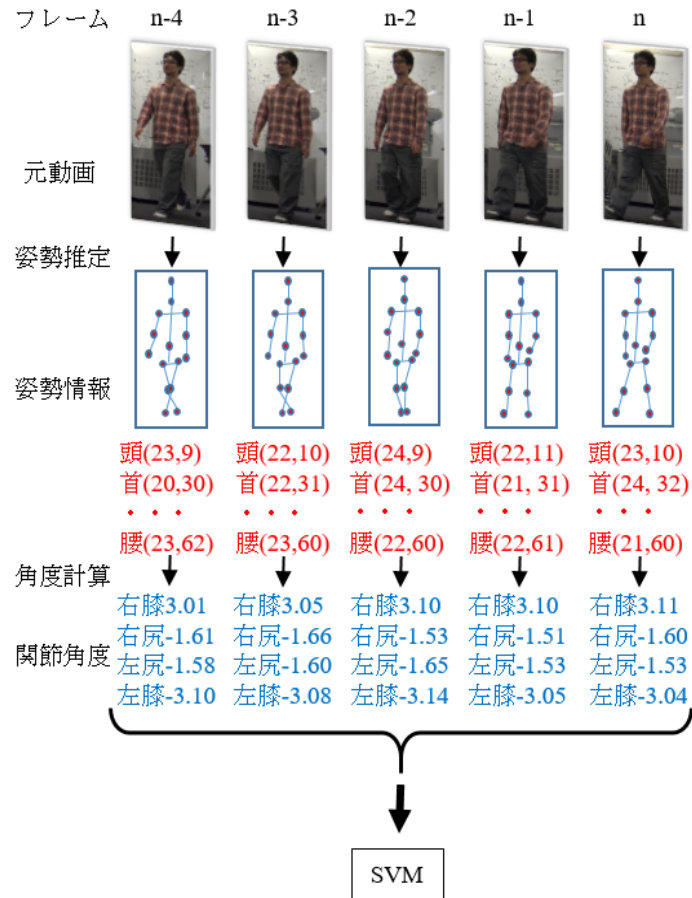


図 4-1: 関節角度による歩行検出の概要図 [31]

4.5 関節角度の変化と関節間の長さの変化率による歩行検出

本節では、関節角度の変化と関節間の長さの変化率を特徴とし、Random Forest [31]により歩行と非歩行に分類する手法を述べる。これを歩行検出手法 c とする。

まず、各フレームについて関節角度の変化を計算する。フレーム n における関節角度は、4.4 と同様にして表 4-1 に示したものが計算される。フレーム n における関節角度の変化は、フレーム n における各関節位置での関節角度とフレーム $n-1$ における同関節位置での関節角度の差である。図 4-2 に関節角度の差の例を示す。

同時に、各フレームにおける関節間の長さの変化率を計算する。まず、関節間の長さは、画像における関節の部位座標間の長さとして計算される。例として、 p 番目の部位と q 番目の部位の間の長さ l は、それぞれの座標 $Y_p = (u_p, v_p)$ と $Y_q = (u_q, v_q)$ から、

$$l = \sqrt{(u_p - u_q)^2 + (v_p - v_q)^2} \quad (4.6)$$

によって得られる。フレーム n における関節間の長さを l_n とすると、フレーム n における関節間の長さの変化率 p_n は、

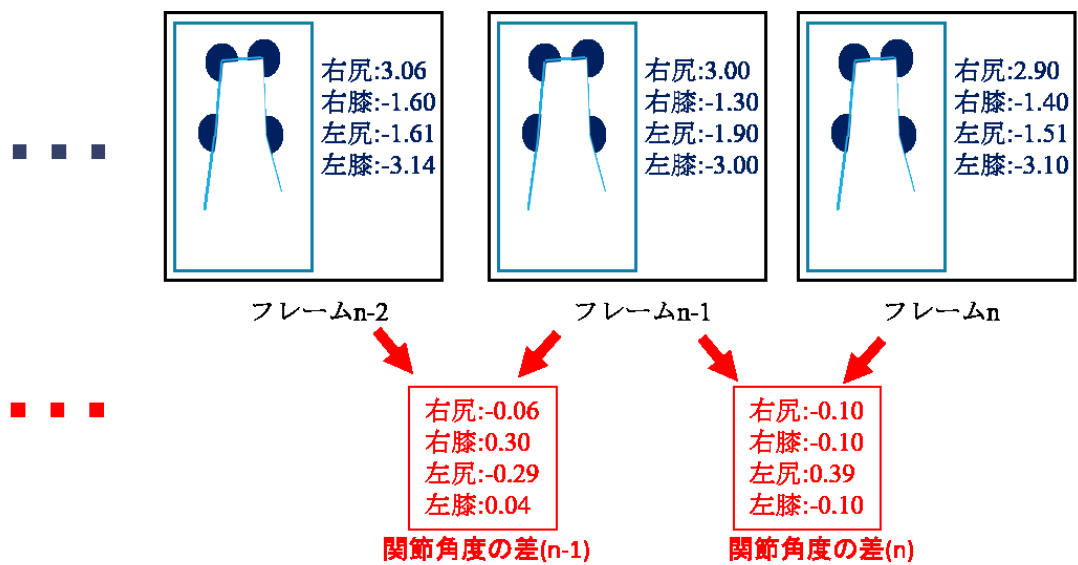


図 4-2: 関節角度の差の例

$$p_n = \frac{l_n - l_{n-1}}{l_n} \quad (4.7)$$

によって得られる．表 4-2 に用いる関節間の長さのリストを示す．また，図 4-3 に関節間の長さの変化率の例を示す．

表 4-2: 用いる関節間の長さのリスト

関節位置	求める長さ
右上腿	右尻-右膝
右下腿	右膝-右足首
左上腿	左尻-左膝
左下腿	左膝-左足首

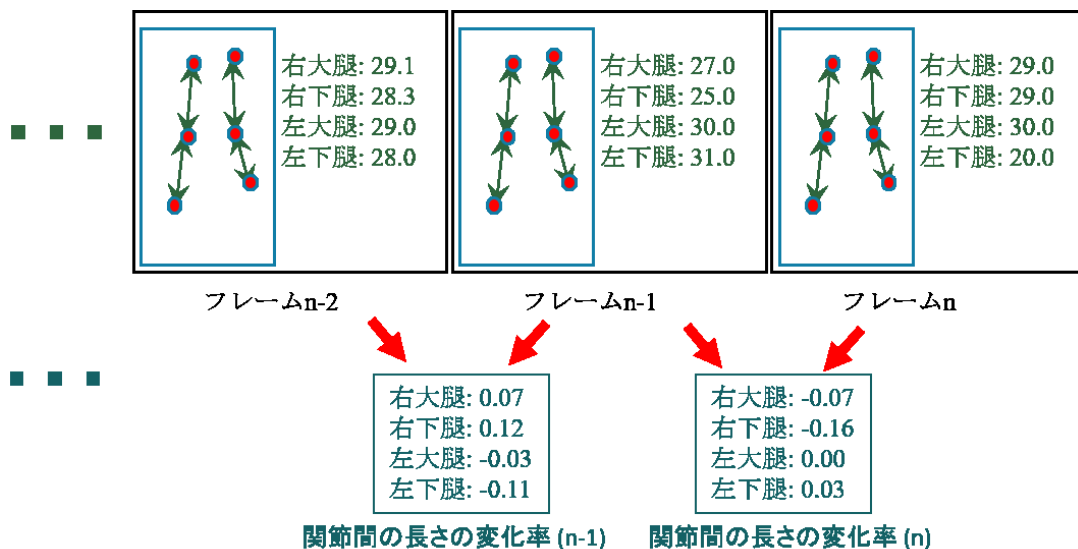


図 4-3: 関節間の長さの変化率の例

最後に、得られた関節角度の差と関節間の長さの変化率を N フレーム分統合し、これを特徴ベクトルとして“歩行”・“非歩行”の 2 クラス分類を行う。ここでは、2 種類の特徴が混在していることから、SVM ではなく Random Forest を用いる。Random Forest は、特徴ごとに重要性を学習し、重みづけて判別を行うことができるという特性を持つ。

4.6 実験

歩行検出手法 a, 歩行検出手法 b, 歩行検出手法 c による実験を行う。

データセットは、人物が歩行している動画 8 系列と直立している動画 8 系列を用い、それぞれの 4 系列を学習データ、残りの 4 系列を評価用データとして用いる。一つの動画には、人物の向きとして図 3-2 と同様に 8 方向が含まれている。学習・評価には各動画から人物の一つの方向につき 10 フレームをランダムに抽出し用いる。よって、8 系列×8 方向×10 フレームの 640 フレームが学習と評価それぞれに用いられる。

また、実験環境における Realtime Multi-Person Pose Estimation の動作速度がおよそ 5 フレーム/秒であることから、歩行検出において過去のフレームの情報を用いる場合、5 フレーム/秒での動作を想定して過去のフレームを取得した。よって、歩行検出器がフレーム n を処理しているとき、フレーム $n-1, n-2$ として取得されるフレームは、30 フレーム/秒の動画におけるフレーム $n-6, n-12$ となる。

実験における歩行検出手法 a のパラメータを表 4-3 に示す。また、歩行検出手法 b の各 N における SVM のパラメータを表 4-4 に示す。このときのカーネルは全て RBF である。歩行検出手法 c においては、Random Forest のパラメータは N に関わらず一定とした。表 4-5 に歩行検出 c における Random Forest のパラメータを示す。実験結果を表 4-6 に示す。

表 4-3: 歩行検出手法 a のパラメータ

T	2.20
W	0.04

表 4-4: 歩行検出手法 b における SVM のパラメータ

N	C	γ
1	50	1
2	50	1
3	50	3
4	30	0.01
5	30	0.01
6	50	0.01
7	50	0.01
8	50	0.01
9	30	0.001

表 4-5: 歩行検出手法 c における Random Forest のパラメータ

木最大深さ	ノード最小サンプル数	木最大個数	収束閾値
10	2	100	0.001

表 4-6: 歩行検出手法の正解率

N	歩行検出手法 a	歩行検出手法 b	歩行検出手法 c
1	NA	0.678	NA
2	0.988	0.847	0.964
3	NA	0.811	0.977
4	NA	0.892	0.983
5	NA	0.900	0.994
6	NA	0.911	0.998
7	NA	0.900	0.992
8	NA	0.891	0.989
9	NA	0.905	1.000

歩行検出手法 a は N=2 において正解率 0.988 と、最も良い性能を示した。N が 5 以上の場合は、歩行検出手法 c が N=2 における歩行検出手法 a の性能を上回った。歩行検出手法

b の正解率は最高でも 0.911 であり，歩行検出手法 a, c いずれよりも性能が低いことが確認された．

4.7 むすび

本章では，歩行検出手法について述べた．足首座標による手法，関節角度による手法，関節角度の変化と関節間の長さの変化率による手法の三つを提案し，実験によって比較を行った．実験結果から，用いるフレーム数が少ない場合には足首座標による手法が，多い場合には関節角度の変化と関節間の長さの変化率による手法が良い性能を示すことが確認された．歩行検出の各手法については，第 6 章においても，歩きスマホ検出のアルゴリズムに組み込んで比較を行う．

第5章

把持物体認識

5.1 まえがき

第3章にて、姿勢情報からのスマホ使用姿勢検出では、スマートフォンを持っていない、歩きスマホに類似した姿勢の人物を誤検出してしまう可能性があることが課題として指摘された。これを解決するため、人物がスマートフォンを持っているかを判別する把持物体認識が必要である。本章では、把持物体認識について述べる。

本章では、まず把持物体認識手法の概要を述べる。その後、把持物体認識による実験について述べる。最後にまとめる。

5.2 把持物体認識

把持物体認識は、1. 手領域の可視の判別、2. 人物の手領域の座標計算、3. 画像認識の3ステップにて行われる。把持物体認識の概要図を図5-1に示す。1. 手領域の可視の判別では、人物の手の領域がカメラから見えているかを判別する。手領域が見えていない場合、画像認識による把持物体認識は難しいことから、把持物体認識を行わない。

判別は、人物の向きをもとに行う。人物がカメラからみて正面向きである場合は可視と判別し、そうでない場合は不可視と判別する。これは、右肩の座標 $Y_3 = (u_3, v_3)$ と左肩の座標 $Y_6 = (u_6, v_6)$ を用いて、

$$u_3 < u_6 \quad (5.1)$$

である場合に可視と判別し、それ以外の場合は不可視と判別する。

2. 人物の手領域の座標計算では、姿勢推定によって得られた部位座標から、画像認識に用いる人物の手の領域の座標を計算し、切り出しを行う。

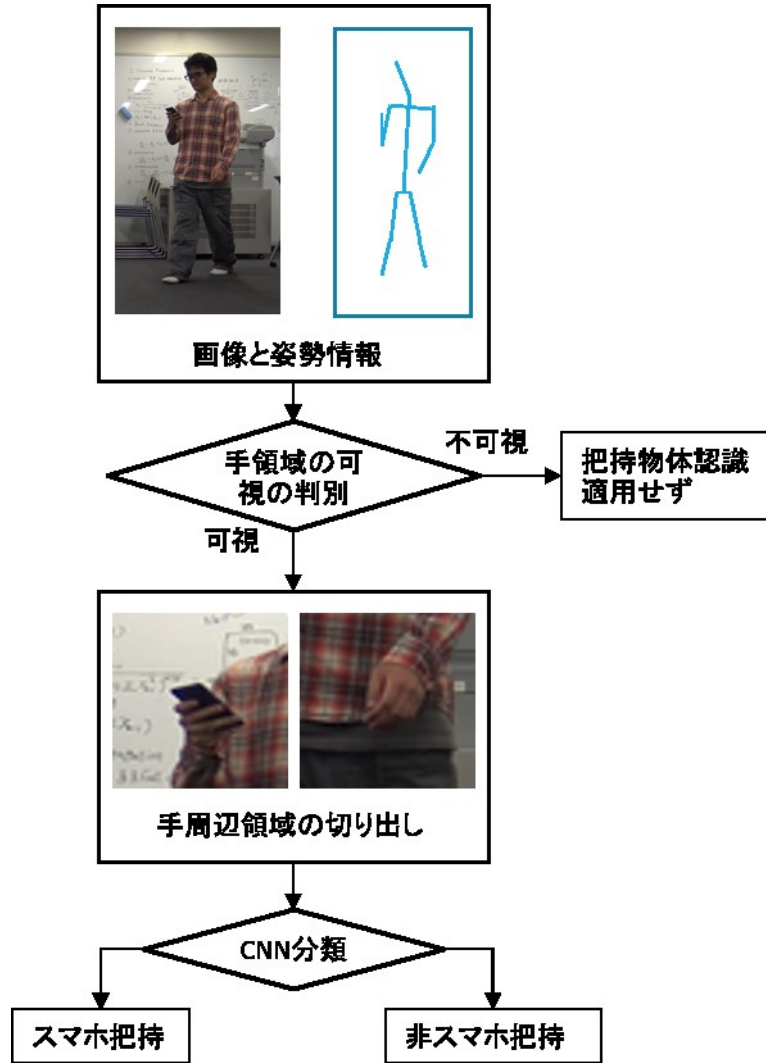


図 5-1: 把持物体認識の概要図

まず，切り出す領域の中心座標 $\mathbf{C}$ を肘の座標と手首の座標から計算する．右手を例にした場合，右肘の座標 $\mathbf{Y}_4 = (u_4, v_4)$ と右手首の座標 $\mathbf{Y}_5 = (u_5, v_5)$ から，

$$\mathbf{C}_r = (u_5 + H(u_4 - u_5), v_5 + H(v_4 - v_5)) \quad (5.2)$$

によって中心座標 $\mathbf{C}_r$ を求める． H は定数である．左手の場合も同様にして，左肘の座標 $\mathbf{Y}_7$ と右手首の座標 $\mathbf{Y}_8$ から中心座標 $\mathbf{C}_l$ を求める．

次に，切り出す範囲 l を計算する．これは，式(3.2)によって得られた姿勢情報のスケール s を用い，

$$l = As \quad (5.3)$$

によって範囲 l を求める． A は定数である．

これにより， $\mathbf{C}$ を中心とした $l \times l$ ピクセルの正方形の範囲を人物の手領域として切り出す．

3. 画像認識のステップでは、切り出された手領域の画像を CNN によって“スマホ把持”と“非スマホ把持”の 2 クラスに分類する。このとき、画像は CNN の入力サイズにリサイズされてから入力される。

5.3 実験

CNN により人物の手領域を“スマホ把持”と“非スマホ把持”に分類する実験を行う。

CNN のアーキテクチャとして、CaffeNet [30]を用いる。また、式(5.2)の定数 H には 1/3 を、式(5.3)の A には 0.3 を用いた。

学習データとして、3.5.1 にて用いたデータセットのうち 8 人分の画像から 5.2 の手法によって手領域を切り出した。これによって得られた画像は、“スマホ把持”が 143 枚、“非スマホ把持”が 535 枚であった。更に、インターネット上から収集した画像から手動により手領域を切り出し、250 枚の“スマホ把持”画像を追加した。

得られた合計 928 枚の画像にデータオーギュメンテーションを行い、144 倍の 133632 枚とした。これを CNN の学習データとして用いる。データオーギュメンテーションとして用いた手法とパラメータを表 5-1: 把持物体認識におけるデータオーギュメンテーションの手法とパラメータに示す。また、CNN の学習時のパラメータを表 5-2 に示す。更に、切り出された手領域の例を図 5-2 に示す。

表 5-1: 把持物体認識におけるデータオーギュメンテーションの手法とパラメータ

手法	パラメータ
左右反転	[反転あり, 反転なし]
回転	回転角=[0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°]
ガウスぼかし	σ =[適用なし, 11, 25]
スケール変化	倍率=[1.0, 1.05, 1.10]

表 5-2: 把持物体認識における CNN 学習パラメータ

バッチサイズ	64
学習開始時の学習レート	0.001
学習レート変更率	0.1 倍
学習レート変更ステップサイズ	90000
学習反復回数	450000
Cropping	あり

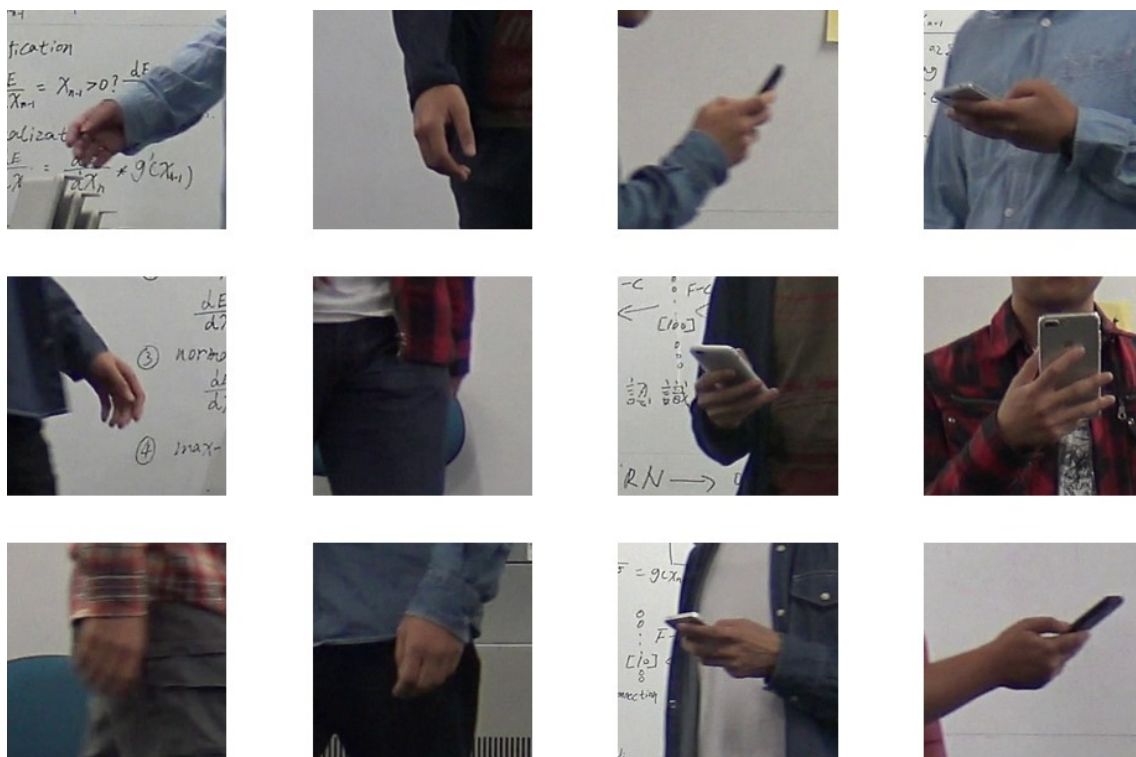


図 5-2: 切り出された手領域の例
左: 非スマホ把持 右: スマホ把持

評価用データとして、3.5.1 にて用いたデータセットのうち、学習データに用いなかった 8 人分の画像から 5.2 の手法によって手領域を切り出した。学習データが CNN の学習時に Cropping の処理を受けていることから、評価用画像の切り出し時には $A=0.2$ を用いた。これによって得られた画像は、“スマホ把持”が 161 枚，“非スマホ把持”が 517 枚であった。表 5-3 及び表 5-4 に詳細な実験結果を示す。

表 5-3: 把持物体認識用 CNN による混同行列

クラス	スマホ把持	非スマホ把持	合計
非スマホ把持	56	461	517
スマホ把持	144	17	161

表 5-4: 把持物体用 CNN による分類結果

クラス	Precision	Recall	F1-score
非スマホ把持	0.964	0.892	0.927
スマホ把持	0.720	0.894	0.798

実験の結果，“スマホ把持”クラスについて 0.798 の F1-score を得た。また、全体の正解

率は 0.892 であった。

また、スマホ使用姿勢検出と把持物体認識の組み合わせによる実験を行った。スマホ使用姿勢検出にて“スマホ使用”とされた人物で、CNNにより両手が“非スマホ把持”と分類されたものを“スマホ不使用”とする。データセットは 3.5.1 にて用いた画像である。CNN は上記の実験で学習したものを用いる。CNN と同様の 8 人分の画像で SVN の学習を行い、残りの 8 人分を評価に用いた。SVM のパラメータは 3.5.1 にて用いたものと同様である。関節角度による手法 a 及び b での結果を表に示す。

表 5-5: 関節角度による手法 a と把持物体認識による実験結果

把持物体認識	クラス	Precision	Recall	F1-score
あり	“スマホ不使用”	0.944	0.969	0.956
	“スマホ使用”	0.968	0.943	0.955
なし	“スマホ不使用”	0.954	0.964	0.959
	“スマホ使用”	0.963	0.953	0.958

表 5-6: 関節角度による手法 b と把持物体認識による実験結果

把持物体認識	クラス	Precision	Recall	F1-score
あり	“スマホ不使用”	0.968	0.932	0.950
	“スマホ使用”	0.935	0.969	0.951
なし	“スマホ不使用”	0.978	0.906	0.941
	“スマホ使用”	0.913	0.979	0.945

実験結果より、把持物体認識を適用すると“スマホ使用”クラスの Precision が向上し、Recall が低下することが分かる。F1-score は関節角度による手法 b においては向上が見られる。また、把持物体認識によって正しい分類へと変更された例を図 5-3 に、誤った分類へと変更された例を図 5-4 に示す。図 5-3 左の人物は、右手が隠れて正しく姿勢推定できなかったことから“スマホ使用”と分類されたものを把持物体認識により“スマホ不使用”と変更されたと思われる。また、右の人物は左手が高く上がっている姿勢から“スマホ使用”と分類されたものを、スマートフォンを持っていないことから“スマホ不使用”と変更された例である。図 5-4 の人物はいずれもスマートフォンを持っているにもかかわらず把持物体認識での誤認識により“スマホ不使用”とされた例である。



図 5-3: 把持物体認識によって正しい分類に変更された例



図 5-4: 把持物体認識によって誤った分類に変更された例

更に、同様の学習条件で 3.5.4 にて用いた画像による実験を行った。用いた手法は関節角度による手法 a であり、把持物体認識の適用の有無により比較を行う。結果を表に示す。

表 5-7: 手提げ袋を持った人物の画像に対する把持物体認識による正解率の変化

把持物体認識の適用	正解率
なし	0.227
あり	1.000

スマホ使用姿勢検出のみでは、スマホ使用姿勢に類似した姿勢の、手提げ袋を持った人物に対する正解率は低い。それに対し、把持物体認識を適用した場合には正解率は 1.000 であり、類似する姿勢の人物に対する性能が大きく向上していることが分かる。

以上の各実験結果から、把持物体認識の適用は歩きスマホ検出の Precision 向上に有効であるといえる。

5.4 むすび

本章では、把持物体認識の手法について述べ、実験を行った。実験にて、把持物体認識が歩きスマホ検出の **Precision** 向上に有効であることが確認された。その一方、把持物体認識による **Recall** の低下も確認された。

Precision の向上は、アルゴリズムによる誤検出を減らすことにつながり、信頼性が高まる。しかし、**Recall** の低下は、アルゴリズムによる歩きスマホの未検出を増やすことになるため、問題である。把持物体認識のアルゴリズム改良や用いる CNN の性能向上によるさらなる **Precision** の向上と **Recall** 低下の抑制が今後の課題であるといえる。

次章では、第 3 章から第 5 章にて述べた手法を組み合わせた、動画からの歩きスマホ検出手法について述べる。

第6章

姿勢情報を用いた歩きスマホ検出

6.1 まえがき

第3章から第5章にて、歩きスマホ検出のための各手法について詳しく述べた。これらをふまえ、本章では姿勢推定を用いた動画からの歩きスマホ検出手法について述べる。

本章の構成としては、まず、歩きスマホ検出手法の概要を述べる。その後、実験を行い、アルゴリズムの性能を確かめる。同時に、用いる歩行検出手法と把持物体認識の有無による性能の比較を行う。最後に、本章のまとめを行う。

6.2 概要

姿勢推定を用いた歩きスマホ検出は、スマホ使用姿勢検出、歩行検出、把持物体認識の3ステップによって行われる。

まず、**Realtime Multi-Person Pose Estimation** によって姿勢情報が得られた人物について、スマホ使用姿勢検出を行う。スマホ使用姿勢検出の詳細は第3章にて述べた。これにより、人物を“スマホ使用”・“スマホ不使用”の2クラスに分類する。

次に、“スマホ使用”とされた人物について、歩行検出を行う。歩行検出の詳細は第3章にて述べた。これにより、人物を“歩行”・“非歩行”の2クラスに分類する。ただし、人物を追跡可能なフレーム数が十分でないなどの要因により、歩行検出が行えない場合がある。その場合、“歩行”と分類された場合と同様の扱いとする。

その後、スマホ使用姿勢検出にて“スマホ使用”，歩行検出にて“歩行”と分類された人物について、把持物体認識を行う。これにより、人物を“スマホ把持”・“非スマホ把持”の2クラスに分類する。手領域が不可視であると判別され、把持物体認識が適用できない場合には、“スマホ把持”と同様の扱いとする。

以上の3ステップにより、把持物体認識のステップにて“スマホ把持”，あるいはそれと

同様の扱いとなった人物を“歩きスマホ”として検出する．なお，6.3 での実験においては，把持物体認識を適用しない手法についても検討を行っているが，その場合，歩行検出にて“歩行”とされた時点で“歩きスマホ”として検出する．図 6-1 にアルゴリズム全体の概要図を示す．

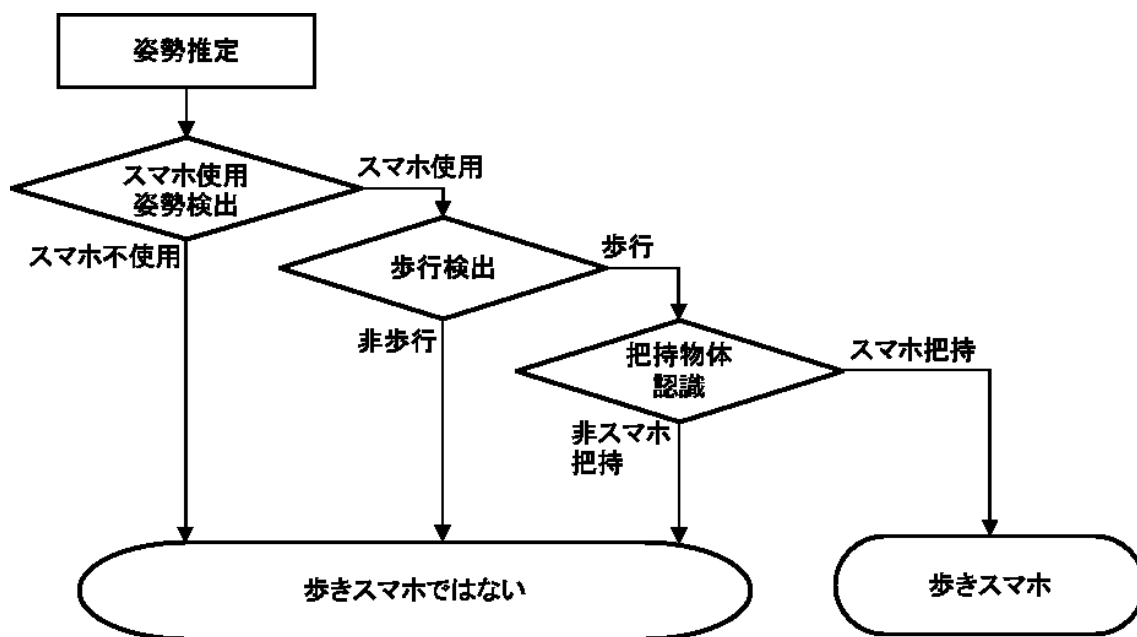


図 6-1: 歩きスマホ検出手法全体の概要図

6.3 実験

動画から歩きスマホをしている人物を検出する実験を行う．実験に用いるデータセットとして，8 人の人物が 45° 感覚で 8 方向を向いて歩行している動画を用いた．各人物について，歩きスマホをしている動画(歩きスマホ動画)と歩きスマホをしていない動画(非歩きスマホ動画)が存在する．うち 1 人の非歩きスマホ動画は，手提げ袋を腕から下げた状態で歩行しているものである．ただし，歩きスマホ動画においても，立ち止まっている状態など，人物が歩きスマホをしていないフレームが存在する，それらのフレームの人物については“歩きスマホ”としては扱わず，これを“歩きスマホ”として検出した場合には誤検出となる．

スマホ使用姿勢検出には，関節角度による手法 a を用いる．学習データとパラメータは 3.5.2 と同様である．

歩行検出には，歩行検出手法 a(足首座標による手法)，歩行検出手法 b(関節角度+SVM による手法)，歩行検出手法 c(関節角度の変化と関節間の長さの変化率+Random Forest)の各手法を用い，比較を行う．学習データとパラメータは 4.6 と同様である．N は上限を 9 と

し、各フレームにおいて過去に向かって人物を追跡できる最大のフレーム数を用いる。

把持物体認識は、その適用の有無を比較する。学習データとパラメータは 5.3 と同様である。

アルゴリズムによって処理されたフレームは、歩きスマホ 2210 フレーム、非歩きスマホ 3157 フレームで、合計 5783 フレームであった。図 6-1 に各手法による実験結果を示す。

表 6-1: 歩きスマホ検出の実験結果

把持物体認識	歩行検出手法	Precision	Recall	F1-score
あり	歩行検出手法 a	0.835	0.829	0.847
	歩行検出手法 b	0.778	0.810	0.794
	歩行検出手法 c	0.812	0.896	0.852
なし	歩行検出手法 a	0.777	0.861	0.817
	歩行検出手法 b	0.726	0.842	0.779
	歩行検出手法 c	0.755	0.931	0.834

実験結果より、歩きスマホ検出手法は、F1-score において最大で 0.852 の性能を示した。歩行検出手法を比較すると、Precision においては歩行検出手法 a が、Recall と F1-score においては歩行検出手法 c が優れていることが分かる。また、把持物体認識を適用した場合、適用しない場合と比較して Precision と F1-score は向上し、Recall は低下していることが分かる。

6.4 むすび

本章では、姿勢情報を用いた歩きスマホ検出手法について述べた。第 3 章から第 5 章の内容をふまえ、3 ステップによる歩きスマホ検出手法を提案し、性能評価のための実験を行った。また、歩行検出手法の比較及び把持物体認識の有無についての検討を行った。その結果、提案手法は、歩行検出手法 c を用い、把持物体認識を適用した場合に 0.852 の最大 F1-score を示すことが確認できた。

第7章

結論

本研究では、駅ホームにおける歩きスマホの危険性から、カメラ映像からの歩きスマホ検出手法を研究している。本論文では、姿勢推定を用いたカメラ映像からの歩きスマホ検出手法について述べた。姿勢推定によって得られた姿勢情報をもとに、スマホ使用姿勢検出、歩行検出、把持物体認識の3ステップによる歩きスマホ検出手法を提案し、各ステップ及びアルゴリズム全体について検証を行った。アルゴリズム全体では、従来の歩きスマホ検出・認識ではなされていない8方向を向いた人物の動画による検証を行い、F1-scoreにして0.852という高い精度での検出が可能であることを確認した。

第1章では、本研究の背景と目的について述べた。また、本論文の構成を述べた。

第2章では、本研究に関連する技術や研究について述べた。まず、姿勢推定のアルゴリズムとして Convolutional Pose Machines とその発展である Realtime Multi-Person Pose Estimation について述べた。その後、人物行動認識に関する研究を紹介した。人物行動認識の分野では、Two-Stream Convolutional Networks を用いた手法が現在主流であるが、姿勢推定を用いた研究についても紹介した。最後に、歩きスマホ認識・検出の研究について述べた。本研究に深く関連するものとして、新村らの HOG と SVM による手法を紹介した。

第3章では、姿勢推定を用いたスマホ使用姿勢検出について述べた。まず、Realtime Multi-Person Pose Estimation による姿勢推定について述べた。次に、部位座標を位置とスケールについて正規化し、SVM によって分類する手法について述べた。また、上半身の関節角度を計算し、それを特徴として SVM で分類する手法について述べた。その後、実験について述べた。実験では、HOG+SVM による手法、CNN による手法と姿勢情報による手法を比較し、姿勢情報による手法が優れていることを確認した。また、下半身の隠れた人物に対する適用において、関節角度による手法が部位座標による手法よりも優れていることを確認した。更に、歩行していない人物及びスマートフォンを持っていない人物への適用により、歩行検出と把持物体認識の必要性を確認した。

第4章では、歩きスマホ検出のための歩行検出について述べた。まず、歩行検出に人物の

複数フレームにわたる姿勢情報を用いることから、人物の追跡について述べた。次に、足首の座標の変化に閾値処理を行うことで歩行検出する手法について述べた。その後、人物の下半身の複数フレームにわたる関節角度を特徴ベクトルとし、SVMにより分類する手法について述べた。更に、人物の下半身の関節角度の変化と関節間の長さの変化率を特徴ベクトルとし、Random Forestによって分類する手法について述べた。実験を行い、足首座標の変化による手法、関節角度の変化と関節間の長さの変化率による手法が有力であることが確かめられた。

第5章では、把持物体認識について述べた。まず、把持物体認識の手法について述べた。把持物体認識は、人物の手周辺領域が可視であるとき、姿勢情報から手周辺領域の座標を計算し、画像認識を行うことで、スマートフォンを持っているかを判別する手法である。実験により、歩きスマホ検出のPrecision向上に有効であることが確認された。

第6章では、第3章から第5章で詳細に述べた3つのステップによる歩きスマホ検出手法について述べた。まず、アルゴリズムの概要を述べた。スマホ使用姿勢検出にて“スマホ使用”とされた人物について、歩行検出を行う。更に、“歩行”とされた場合には把持物体認識を行い、“スマホ把持”とされた人物を“好きスマホ”として検出する手法である。実験により、高い精度での検出が可能であることを確認した。また、歩行検出手法の比較と把持物体認識の有無についての検討を行い、各指標における優劣を確認した。

第7章は結論であり、本論文の総括を行った。

謝辞

本研究成果は、独立行政法人情報通信研究機構(NICT)の委託研究「ソーシャル・ビッグデータ利活用・基盤技術の研究開発」によって得られたものであります。

本研究の必要機材や実験環境，快適な研究環境を与えてくださり，研究及び学生生活において素晴らしいご指導を頂いた渡辺教授に心より感謝いたします。

日頃から貴重なアドバイスや示唆を頂いている渡辺研究室の皆様に感謝いたします。

最後に，私をここまで育ててくださり，生活を支えてくださっている家族に感謝いたします。

参考文献

- [1] 国土交通省, “鉄道輸送統計年報,” 国土交通省総合政策局情報政策課交通経済統計調査室, 2017.
- [2] 一般社団法人 電気通信事業者協会, “「歩きスマホ」に関する調査,” 一般社団法人 電気通信事業者協会, 2015.
- [3] 小塚一宏, “歩行中・自転車運転中の“ながらスマホ”時の視線計測と危険性の考察,” 電子情報通信学会 基礎・境界ソサイエティ *Fundamentals Review*, vol. 10, no. 2, pp. 129-136, 2016.
- [4] 国土交通省, “「プラットホーム事故0（ゼロ）運動」について,” 国土交通省鉄道局安全監理官室, 2015.
- [5] 国土交通省, “ホームドアの設置状況,” 2017.
- [6] 国土地理協会, “緯度経度付き全国沿線・駅データベース,” 2017.
- [7] 小林広幸, “鉄道ホームドアの整備加速も、課題は「1駅4—5億円」の費用,” *日刊工業新聞・ニュースイッチ*, 47 2017.
- [8] 阪神電気鉄道株式会社, KDDI 株式会社, “阪神電車と au の共同企画「阪神×au 歩きスマホ防止キャンペーン」を展開,” 14 12 2015. [オンライン]. Available: <http://news.kddi.com/kddi/corporate/newsrelease/2015/12/14/1494.html>. [アクセス日: 27 12 2017].
- [9] S.-E. Wei, V. Ramakrishna, T. Kanade, Y. Sheikh, “Convolutional Pose Machines,” *CVPR*, 2016.
- [10] Z. Cao, T. Simon, S.-E. Wei and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” *CVPR*, 2017.
- [11] K. Simonyan and A. Zisserman, “Two-Stream Convolutional Networks for Action Recognition in Videos,” *NIPS*, 2014.

- [12] Z. Lan, Y. Zhu and A. G. Hauptmann, "Deep Local Video Feature for Action Recognition," *CVPR workshop*, 2017.
- [13] H. Kuehne, H. Jhuang, E. Garrote, T. A. Poggio, T. Serre, "HMDB: A Large Video Database for Human Motion Recognition," *ICCV*, 2011.
- [14] K. Soomro, A. R. Zamir, M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," *CRCV*, 2012.
- [15] 工藤雄太, 指田岳彦, 青木義満, "CNN を用いた姿勢特徴抽出と行動遷移を考慮した人物行動認識," *電子情報通信学会和文論文誌 D*, Vols. J100-D, no. 7, pp. 681-691, 2017.
- [16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [17] 一原賢吾, 竹内健, 金井謙治, 甲藤二郎, "姿勢推定技術を活用した人物の動作認識手法の精度評価," *映像情報メディア学会冬季大会*, 2017.
- [18] 黒瀬龍之介, 林昌希, 石井壮郎, 岡村麻人, 青木義満, "姿勢推定を用いたテニス映像の姿勢傾向分析," *映像学技報*, vol. 40, Nn. 43, pp. 21-24, 2016.
- [19] 根岸匠, "「ながらスマートフォン」抑止システムの開発," 筑波大学, 2014.
- [20] 名坂康平, 加藤岳久, 西垣正勝, "スマートフォン使用時の不注意による事故防止システムの提案," *研究報告マルチメディア通信と分散処理 (DPS)*, vol. 150, no. 28, pp. 1-6, 2012.
- [21] "歩きスマホ注意アプリ," KDDI, [オンライン]. Available: <https://www.au.com/mobile/service/smartphone/safety/aruki-sumaho/>.
- [22] 赤池英夫, 角田博保, 野田口宗, "歩行中のスマートフォン使用時における障害物検出および提示手法の提案と評価," *研究報告ヒューマンコンピュータインタラクション (HCI)*, vol. 151, no. 3, pp. 1-7, 2013.
- [23] 平川哲也, 小倉加奈代, ベッドバハドゥールビスタ, 高田豊雄, "「ながらスマートフォン」状況下での衝突事故防止のための障害物検知手法の提案," *第77回全国大会講演論文集*, vol. 2015, no. 1, pp. 243-244, 2015.
- [24] 皆本光, 佐野睦夫, "歩行者の不注意行動認識 -歩きスマホ検出-, 2015. [オンライン]. Available: <http://www.y-adagio.com/public/committees/vhis/confs/vhis6/vhis6-2.pdf>. [アクセス日: 27 12 2017].
- [25] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *CVPR*, vol. 1, pp. 886-893, 2005.
- [26] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, p. 273-297, 1995.

- [27] 新村文郷, 川西康友, 出口大輔, 井手一郎, 村瀬洋, 藤吉弘亘, "車載カメラ画像からの「スマホ歩き」認識に基づく歩行者の不注意度推定," *信学技報, PRMU2015-46*, vol. 115, no. 98, pp. 83-88, 2015.
- [28] Z. Cao, T. Simon, S.-E. Wei and Y. Sheikh, "GitHub:Realtime Multi-Person Pose Estimation," [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/caffe_rtpose. [Accessed 29 12 2017].
- [29] M. Andriluka, L. Pishchulin and P. Gehler, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," *CVPR*, 2014.
- [30] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick and P. Dollár, "Microsoft COCO: Common Objects in Context," *ECCV*, 2015.
- [31] "OpenCV," [オンライン]. Available: <https://opencv.org/>.
- [32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going Deeper with Convolutions," *CVPR*, 2015.
- [33] 西川徹, 岡野原大輔, 長谷川順一, "Chainer," [オンライン]. Available: <https://chainer.org/>.
- [34] "OpenCV3.10 Histograms," [オンライン]. Available: https://docs.opencv.org/3.1.0/d6/dc7/group__imgproc__hist.html. [アクセス日: 8 1 2018].
- [35] 加藤君丸, 渡辺裕, "姿勢推定による歩きスマホ認識のための歩行検出," *FIT2017(第16回科学技術フォーラム)*, 2017.
- [36] L. Breiman, "Random forests," in *Machine learning*, vol. 45, 2001, pp. 5-32.
- [37] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, pp. 1097-1105, 2012.

図一覧

図 2-1: CPMs の入力と出力の例.....	5
図 3-1: 入力画像.....	10
図 3-2: データセットに用いた動画のフレームの例(人物領域のみ).....	12
図 3-3: 姿勢推定の結果例 (人物領域のみ).....	13
図 3-4: 誤分類された画像と姿勢推定の例	17
図 3-5: 下半身が隠れた人物の画像例.....	18
図 3-6: 下半身が隠れた人物の姿勢推定の例.....	18
図 3-7: 立ち止まってスマートフォンを使用する人物の画像と姿勢推定の例.....	20
図 3-8: 手提げ袋を持った人物の画像と姿勢推定の例	20
図 4-1: 関節角度による歩行検出の概要図 [31]	25
図 4-2: 関節角度の差の例	26
図 4-3: 関節間の長さの変化率の例	27
図 5-1: 把持物体認識の概要図.....	31
図 5-2: 切り出された手領域の例	33
図 5-3: 把持物体認識によって正しい分類に変更された例	35
図 5-4: 把持物体認識によって誤った分類に変更された例	35
図 6-1: 歩きスマホ検出手法全体の概要図	38

表一覧

表 3-1: Realtime Multi-Person Pose Estimation によって推定される部位	9
表 3-2: 正規化された部位座標の例	10
表 3-3: 手法と用いる関節角度	11
表 3-4: HOG+SVM による手法における HOG 特徴抽出器のパラメータ	13
表 3-5: HOG+SVM による手法における SVM のパラメータ	13
表 3-6: CNN による手法におけるデータオーギュメンテーションの手法とパラメータ	14
表 3-7: CNN による手法における学習パラメータ	14
表 3-8: 部位座標による手法における SVM のパラメータ	14
表 3-9: 関節角度による手法 a における SVM のパラメータ	15
表 3-10: 関節角度による手法 b における SVM のパラメータ	15
表 3-11: HOG+SVM の分類結果	15
表 3-12: CNN の分類結果	15
表 3-13: 部位座標による手法の分類結果	15
表 3-14: 関節角度による手法 a の分類結果	15
表 3-15: 関節角度による手法 b の分類結果	16
表 3-16: “スマホ使用” クラスにおける各手法の性能比較	16
表 3-17: 関節角度による手法 b における人物の向きと誤分類数	16
表 3-18: 部位座標による手法の分類結果	18
表 3-19: 関節角度による手法 a の分類結果	18
表 3-20: 関節角度による手法 b の分類結果	19
表 3-21: “スマホ使用” クラスにおける各手法の性能比較	19
表 4-1: 用いる関節角度のリスト	24
表 4-2: 用いる関節間の長さのリスト	26
表 4-3: 歩行検出手法 a のパラメータ	28
表 4-4: 歩行検出手法 b における SVM のパラメータ	28

表 4-5: 歩行検出手法 c における Random Forest のパラメータ	28
表 4-6: 歩行検出手法の正解率	28
表 5-1: 把持物体認識におけるデータオーギュメンテーションの手法とパラメータ ..	32
表 5-2: 把持物体認識における CNN 学習パラメータ	32
表 5-3: 把持物体認識用 CNN による混同行列	33
表 5-4: 把持物体用 CNN による分類結果	33
表 5-5: 関節角度による手法 a と把持物体認識による実験結果	34
表 5-6: 関節角度による手法 b と把持物体認識による実験結果	34
表 5-7: 手提げ袋を持った人物の画像に対する把持物体認識による正解率の変化	35
表 6-1: 歩きスマホ検出の実験結果	39

研究業績

- [1] 加藤, 渡辺: “鉄道運行映像に対する背景差分法の適用“, 電子情報通信学会総合大会, D-11-20, Mar. 2016
- [2] 加藤, 渡辺: “線路脇カメラと鉄道車載カメラによる障害物検出の検討“, PCSJ/IMPS 2016, No. P-3-16, Nov. 2016
- [3] 加藤, 渡辺: “Convolutional Pose Machines を用いた歩きスマホ認識“, IEICE 総合大会, D-12-56, Mar. 2017
- [4] 加藤, 渡辺: “姿勢推定による肘と肩の角度情報を用いた歩きスマホ認識“, 映像情報メディア学会スポーツ情報処理時限研究会第 4 回, No.7, June2017
- [5] 加藤, 渡辺: “関節角度情報を用いた歩きスマホ認識“, 映像情報メディア学会年次大会, 13D-4, Sep. 2017
- [6] 加藤, 渡辺: “姿勢推定による歩きスマホ認識のための歩行検出“, FIT2017(第 16 回科学技術フォーラム), H-036, Sep. 2017
- [7] 加藤, 渡辺: ”姿勢推定と把持物体認識を用いた歩きスマホ認識手法の検討“, 映像情報メディア学会冬季大会, 22B-2, Dec. 2017
- [8] 加藤, 渡辺: ” 姿勢推定を用いた歩きスマホ検出に用いる歩行検出手法の改善“, 電子情報通信学会総合大会, Mar. 2018 (発表予定)

受賞

- [1] ITE 年次大会 学生優秀発表賞，映像情報メディア学会，(Dec. 2017)

