

# 卒業論文概要書

## Summary of Bachelor's Thesis

Date of submission: 02/06/2019

学科名 Department	情報通信	氏名 Name	山口 隆太郎	指導 教員 Advisor	渡辺 裕 印 Seal
研究指導 Research guidance	オーディオビジュアル情報処理研究	学籍番号 Student ID number	CD 1W143144-8		
研究題目 Title	単眼 RGB カメラによる姿勢推定情報と歩容特徴量を用いた個人識別手法 Personal Identification Method by Pose Estimation Information and Gait Characteristics from Monocular RGB Camera				

### 1. まえがき

近年、様々な生体認証システムが提案されており、その中に歩き方で個人を識別する歩容認証がある。歩容認証の利点は二つある。一つ目は、指紋認証や虹彩認証と異なり対象者の協力なくデータを収集可能な点である。二つ目は、認証時にも個人からの協力なく認証可能な点である。歩容情報による認証を多くの場面で利用するためには、必要なカメラが安価であり、設置が容易であることが求められる。したがって、より安価で一般的な単眼 RGB カメラの利用が望まれる。一方で、映像中の人物の姿勢情報を特徴点の座標値として出力できる技術が多く提案されている。本研究では、単眼 RGB カメラで姿勢推定可能な OpenPose[1]と 3d-pose-baseline[2]を使用する、OpenPose と 3d-pose-baseline を利用して得られた姿勢情報から、個人を識別する手法を提案する。

### 2. 姿勢推定情報と歩容特徴量を用いた個人識別

#### 2.1 手法概要

提案手法の概要を三つの段階に分けて述べる。

一つ目は、姿勢情報の取得である。撮影した映像に OpenPose を適用し、2次元関節座標の推定結果を JSON ファイルで出力する。OpenPose が出力したデータを 3d-pose-baseline の入力とし、3次元関節座標の推定結果を txt ファイルで出力する。

二つ目は、取得した姿勢情報から特徴量への変換である。JSON ファイルと txt ファイルから数値データを読み取り計算する。JSON ファイルの 2次元データにはスケール変換と位置補正を行う。二つの処理を

加えた 2次元データを特徴量とする。更に、txt ファイルの 3次元データから関節の角度と手首、足首の距離を算出し特徴量とする。

三つ目は算出した特徴量を用いた分類である。CNN を用いて個人を分類し、精度を評価する。

#### 2.2 OpenPose

OpenPose は映像内の複数の人物を同時に認識し、1人につき 18 個の関節の 2次元座標を出力できる。論文発表後、顔、手の平の特徴点を求める機能が追加された。顔については、顔の各パーツ、輪郭の合計 70 個の特徴点を求められる。手の平については、左右の手首から指先までの合計 21 個の特徴点を求められる。OpenPose の適用例を以下の図 2.1, 2.2 に示す。



図 2.1 姿勢推定前画像

図 2.2 姿勢推定成功例

#### 2.3 3d-pose-baseline

3d-pose-baseline は、入力された 2次元の姿勢情報から 3次元の姿勢情報を推定する。3d-pose-baseline では、Human3.6M[3]の 3次元姿勢データを教師データとしてニューラルネットで学習、推定する。OpenPose と同様に画像ではなく関節の座標点を学習

するため、軽量なモデルで実現できる。  
3d-pose-baselineの適用例を以下の図2.3, 2.4に示す。

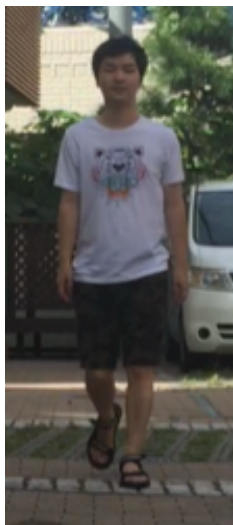


図 2.3 姿勢推定前画像

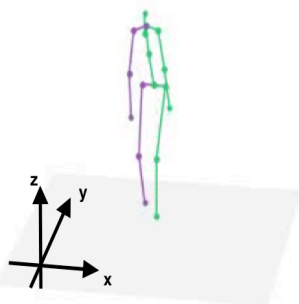


図 2.4 姿勢推定成功例

## 2.4 スケール変換

OpenPose で得られた 2 次元の姿勢情報データは、1 画像中のどこに各関節が存在しているかを 2 次元座標で表したものである。使用した映像では、被験者がカメラに近づいたり、カメラから遠ざかったりしている。したがって、そのままでは座標のスケールが不揃いである。そこで、ある関節間の長さを 1 とし、2 次元データのスケールを一致させる。

提案手法では、様々な方向からの撮影で、映像中の 2 次元の長さの変化が最も少なく、検出失敗が少ない大腿部を選ぶ。右大腿部の長さが 1 になるように姿勢情報をスケール変換する。

## 2.5 位置補正

スケール変換の意義を述べた通り、使用する映像では常に被験者の映る場所が変化する。OpenPose の出力は 2 次元座標であるため、スケール変換に加えて座標の位置補正を行う必要がある。そこで、被験者の首の位置が原点となるように位置補正を行う。

## 2.6 3 次元角度の算出

3d-pose-baseline で得られた 3 次元の姿勢情報から、被験者の歩容特徴が多く表現されると考えられる関節間の 3 次元角度を算出する。この角度を歩容情報とする。対象となる角度は、尾骶骨、尻、膝の三点、尾骶骨、尻、足首の三点、肩、肘、手首の三点がなす角度で合計 6 個である。

また、手足の振り幅として、右手首と左足首、左手首と右足首の間の 3 次元距離を算出し歩容情報とする。

## 3. 実験結果

様々な方向から撮影した人物の歩行映像を用いて 4 種類の実験を行なった。前方から撮影した映像での識別精度が最も高かった。斜めから撮影した映像を使用した場合、身体パーツの誤認識による識別精度の低下が見られた。斜め方向からの映像では、身体の陰に入るパーツが多く、3 次元の姿勢情報の推定が困難である。身体が隠れる問題が生じる状況では、2 次元の姿勢情報のみを用いたほうが高い精度で識別できた。

## 4. むすび

本研究では、人物の歩行映像に対して OpenPose と 3d-pose-baseline を用いて姿勢推定を行い、得られた姿勢情報から人物を識別する手法を提案した。CNN を用いた識別実験では、前方からの映像において Accuracy で 98.6%の精度が得られることを確認し、高い精度で個人識別可能なことがわかった。一方で、前方以外から撮影された映像を用いた場合の精度には課題が残される。姿勢推定失敗時に身体パーツの補完を行う仕組みを取り入れることで、使用する映像の自由度が高い個人識別が可能になると考えられる。

## 参考文献

- [1] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Conference on Computer Vision and Pattern Recognition (CVPR2017), pp.1302-1310, July 2017
- [2] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3d human pose estimation," IEEE International Conference on Computer Vision (ICCV 2017), pp.2659-2668, Oct. 2017
- [3] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.36, No.7, pp.1325-1339, July 2014.

2018 年度 卒業論文

単眼 RGB カメラによる姿勢推定情報と  
歩容特徴量を用いた個人識別手法

Personal Identification Method by Pose Estimation Information and  
Gait Characteristics from Monocular RGB Camera

指導教員 渡辺 裕 教授

早稲田大学 基幹理工学部

情報通信学科

1W143144-8

山口 隆太郎

# 目次

第 1 章 序論 .....	1
1.1 研究の背景 .....	1
1.2 本研究の目的 .....	1
1.3 関連研究 .....	2
1.4 本論文の構成 .....	2
第 2 章 OpenPose と 3d-pose-baseline .....	3
2.1 まえがき .....	3
2.2 OpenPose .....	3
2.2.1 概要 .....	3
2.2.2 適用例 .....	4
2.3 3d-pose-baseline .....	6
2.3.1 概要 .....	6
2.3.2 適用例 .....	6
2.4 むすび .....	8
第 3 章 提案手法 .....	9
3.1 まえがき .....	9
3.2 提案手法の概要 .....	9
3.3 姿勢情報の取得 .....	9
3.4 スケール変換 .....	9
3.5 位置補正 .....	10
3.6 3次元角度の算出 .....	10
3.7 CNN を用いた分類 .....	11
3.8 むすび .....	11
第 4 章 実験および結果・考察 .....	12
4.1 まえがき .....	12
4.2 映像データ撮影 .....	12
4.3 実験 1 前方から撮影した映像を用いた実験 .....	12
4.4 実験 2 後方から撮影した映像を用いた実験 .....	14
4.5 実験 3 斜め 4 方向から撮影した映像を用いた実験 .....	16
4.6 実験 4 斜め 4 方向と前後方から撮影した映像を用いた実験 .....	17
4.7 むすび .....	19
第 5 章 結論と今後の課題 .....	20

# 第1章 序論

## 1.1 研究の背景

近年、様々な生体認証システムが提案されており、その中に歩き方で個人を識別する歩容認証がある。歩容認証の利点は二つある。一つ目は、指紋認証や虹彩認証と異なり対象者の特別な行動なくデータを収集可能な点である。二つ目は、認証時にも個人からの認証のための協力なく認証可能な点である。そのため歩容認証は、個人認証だけでなく防犯や犯罪捜査に応用されている[1].

一方で、映像中の人物の姿勢情報を特徴点の座標値として出力できる技術が多く提案されている。Microsoft 社の Kinect や OpenPose[2]がその例である、これらの技術は姿勢情報や歩容情報を用いた個人識別を目的として、多くの研究で利用されている[3][4][5]. しかし、従来の歩容情報に関する研究では、RGB-D カメラや複数のカメラが用いられている。歩容情報による認証を多くの場面で利用するためには、必要なカメラが安価で設置が容易であることが求められる。したがって、より安価で一般的な単眼 RGB カメラの利用が望まれる。

## 1.2 本研究の目的

OpenPose を用いれば、安価で一般的な単眼 RGB カメラで 2 次元の姿勢情報を推定できる。更に、3 次元の歩容情報を利用するために 3d-pose-baseline[6]を用いる。歩容情報で個人を識別することにより、認証のための特別な行動を必要としない。例えば、セキュリティゾーンを歩いたまま通過する認証システムを構築できる。本研究では、室内の扉の前から人物が歩いてくるシーンを想定し、その人物が扉に達する前に個人を識別する場合を考える。また、扉に設置されたカメラによる正面だけからの映像だけでなく、その他の方向からの映像でも個人を識別可能であるかについても検討する。本研究により扉に設置されたカメラ以外からの映像による認証の補助や、街中のカメラを用いた防犯のための個人識別などに応用できると考えられる。

本論文では、OpenPose を用いて人物の 2 次元の姿勢情報を入手する。また、3d-pose-baseline を用いて 3 次元の姿勢情報を入手し、それらのデータから歩容情報を算出する。得られた姿勢情報と歩容情報データを特徴量として Convolutional Neural Network (以下 CNN) で分類し、個人を識別する。

以上の手法を用いて、2 次元データのみで分類する場合と 3 次元の歩容情報を加えた場合の精度を評価する。

### 1.3 関連研究

関連研究としては、戸田らが[3]単眼カメラによる姿勢情報を用いた個人識別手法を提案している。また、三堀ら[4]、森ら[5]が Kinect を用いた姿勢情報による歩容認識を提案している。これらの研究においては、歩容情報と合わせて静的特徴量である姿勢情報を用いて個人を識別している。楨原ら[7]は、複数の多視点同期歩容撮影システムを構築し、広い年齢層の被験者の歩容を撮影することで大規模歩容データベースを構築している。そこから子供、成人男性、成人女性、高齢者という四つのクラスを求めている。

また、歩容認証とは異なるが、顔認証の分野では今岡[8]によると、NEC は顔認証の評価プログラムにおいて、犯罪記録から抽出した顔画像から 92% の 1 位検索率、ビザ申請時に使われた顔画像から 95% の高い検索を達成している。しかし、顔認証にはマスクなどで顔が隠れていると検出が困難となるという欠点が存在する。一方で服装などの制約が存在しない歩容認証は、より簡単な認証方法だと言える。

本研究は、できるだけ安価で簡単に個人を識別することを目的としているため、多くの費用が必要になってしまう深度カメラやカメラの複数使用を前提としない。

### 1.4 本論文の構成

本論文の構成を以下に示す。

第 1 章は本章であり、本研究の背景や目的、関連研究について述べる。

第 2 章では本研究で用いる OpenPose と 3dpose-baseline について述べる。

第 3 章では提案手法について述べる。

第 4 章では手法の比較実験および結果、考察について述べる。

第 5 章では本論文の結論と今後の課題について述べる。

## 第2章 OpenPose と 3d-pose-baseline

### 2.1 まえがき

本章では, 本研究での姿勢推定で用いる OpenPose と 3d-pose-baseline について述べる.

### 2.2 OpenPose

#### 2.2.1 概要

OpenPose は Zhe Cao らによって CVPR2017 で発表された論文「Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields」 [2] で実装されたライブラリである. OpenPose は映像内の複数の人物を同時に認識し, 1 人につき 18 個の関節の 2 次元座標を出力できる. 論文発表後, 顔, 手の平の特徴点を求める機能が追加された. 顔については, 顔の各パーツ, 輪郭の合計 70 個の特徴点を求められる. 手の平については, 左右の手首から指先までの合計 21 個の特徴点を求められる.

OpenPose では, 始めに, 画像内の人物の関節位置を示す confidence map を Deep Learning を用いて推定する. 次に, 関節同士が繋がらうる可能性を推定する Part Affinity Fields を用いて同じ人物の関節同士を繋ぐことで, 複数人の同時姿勢推定を可能にしている. このように, OpenPose では Confidence Map と Part Affinity Fields を用いた二つの逐次プロセスにより, 画像内の人物の関節位置 18 部位と関節間の結合性をボトムアップ的アプローチにより推定している. 以下の表 2.1 で OpenPose の検出部位を示す.

表 2.1 OpenPose 検出部位一覧

検出部位
Nose
Neck
Right Shoulder
Right Elbow
Right Wrist
Left Shoulder
Left Elbow
Left Wrist
Right Hip
Right Knee
Right Ankle
Left Hip
Left Knee
Left Ankle
Right Eye
Left Eye
Right Ear
Left Ear

### 2.2.2 適用例

OpenPose の適用例を以下の図 2.1, 2.2 に示す。また、OpenPose の姿勢推定では誤検出が発生する。誤検出は身体の一部が写っていない場合や、画像内の人物が直立していない場合に発生する。誤検出の適用例を以下の図 2.3, 2.4 に示す。図 2.4 では人物の右前腕が胴体部分の影に隠れているため、検出に失敗している。





图 2.1 姿势推定前画像(1)



图 2.2 2次元姿势推定成功例



图 2.3 姿势推定前画像(2)

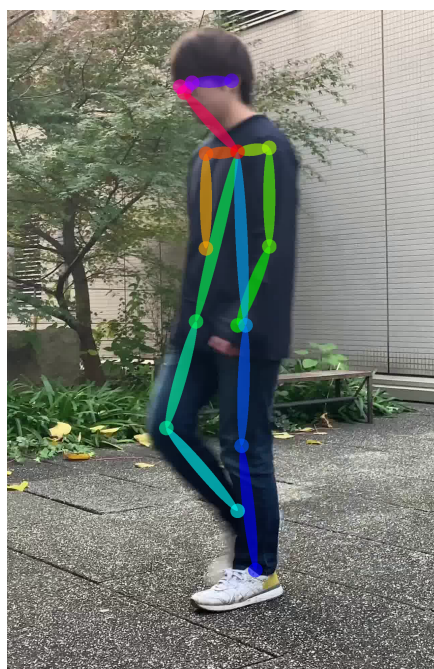


图 2.4 2次元姿势推定失败例(右前腕)

## 2.3 3d-pose-baseline

### 2.3.1 概要

3d-pose-baselineはJulieta MartinezらによってICCV2017で発表された論文「A simple yet effective baseline for 3d human pose estimation」[6]で実装されたライブラリであり、入力された2次元の姿勢情報から3次元の姿勢情報を推定する。3d-pose-baselineでは、データセット「Human3.6M」[9]の3次元姿勢データを教師データとしてニューラルネットで学習、推定する。OpenPoseと同様に画像ではなく関節の座標点を学習するため、軽量なモデルで実現できる。

3d-pose-baselineの姿勢推定方法では、関節点の数を $N$ 、入力となる2次元関節の座標点を $\mathbf{x}_i \in \mathbf{R}^{2n}$ 、出力となる3次元関節の座標点を $\mathbf{y}_i \in \mathbf{R}^{3n}$ として、次の式(2.1)に示す誤差関数を最小化するように関数 $f^* : \mathbf{R}^{2n} \rightarrow \mathbf{R}^{3n}$ を求めている[7]。

$$f^* = \min_f \frac{1}{N} \sum_{i=1}^N L(f(\mathbf{x}_i) - \mathbf{y}_i) \quad (2.1)$$

関数 $f$ は以下に示す図2.5のようなニューラルネットで近似する。ニューラルネットには16個の2D関節座標点が入力され、16個の3D関節座標点が出力される。

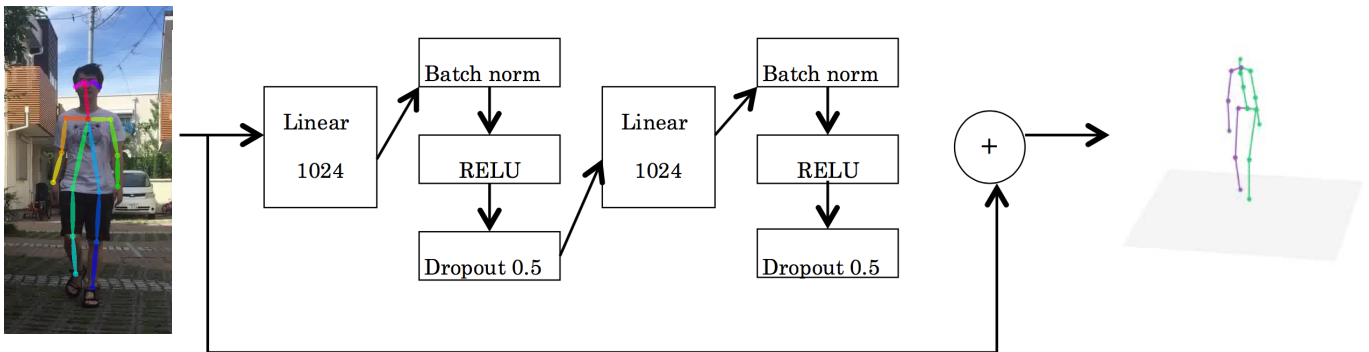


図 2.5 3d-pose-baseline で使用されるニューラルネットワーク構造

### 2.3.2 適用例

3d-pose-baselineの適用例を以下の図2.6, 2.7に示す。また、3d-pose-baselineにおいてもOpenPoseと同様に誤検出が発生する。誤検出は入力の2次元姿勢情報が欠損している場合に発生する。誤検出の適用例を以下の図2.8, 2.9に示す。図2.9では人物の右前腕が胴体部分の影に隠れているため、検出に失敗している。

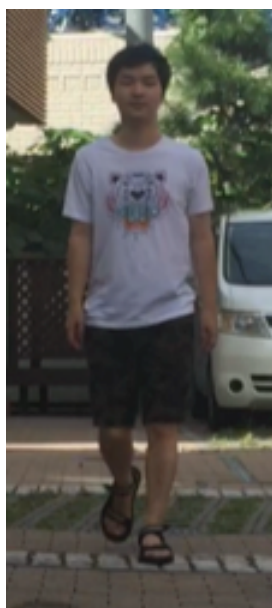


图 2.6 姿势推定前画像(3)

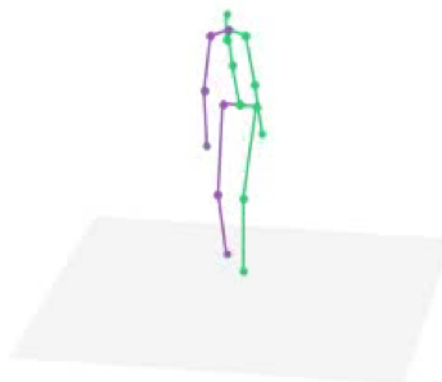


图 2.7 3次元姿势推定成功例



图 2.8 姿势推定前画像(4)

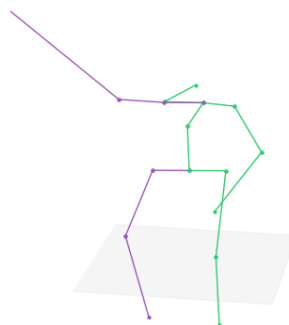


图 2.9 3次元姿势推定失败例(右前腕)

## 2.4 むすび

本章では，本研究で使用するライブラリについて述べた．**OpenPose** は映像中の人物の 2次元姿勢情報を推定できる．**3d-pose-baseline** は 2次元の姿勢情報を入力として，3次元の姿勢情報を推定できる．

## 第3章 提案手法

### 3.1 まえがき

本章では, OpenPose と 3d-pose-baseline を用いて, 歩行中の人物を CNN で分類する手法を提案する. 3.2 節で手法の概要を述べ, 3.2-3.6 で, 処理の詳細について述べる.

### 3.2 提案手法の概要

提案手法の概要を三つの段階に分けて述べる.

一つ目は, 姿勢情報の取得である. 撮影した映像に OpenPose を適用し, 2次元関節座標の推定結果を JSON ファイルで出力する. OpenPose が出力したデータを 3d-pose-baseline の入力とし, 3次元関節座標の推定結果を txt ファイルで出力する. 二つ目は, 取得した姿勢情報から特徴量への変換である. JSON ファイルと txt ファイルから数値データを読み取り計算する. JSON ファイルの 2次元データにはスケール変換と位置補正を行う. 二つの処理を加えた 2次元データの特徴量とする. 更に, txt ファイルの 3次元データから関節の角度と手首, 足首の距離を算出し特徴量とする. 三つ目は算出した特徴量を用いた分類である. CNN を用いて個人を分類し, 精度を評価する.

### 3.3 姿勢情報の取得

被験者の歩行映像から OpenPose を用いて 2次元の姿勢情報を推定する, 得られたデータを 3d-pose-baseline の入力として, 3次元の姿勢情報を推定する. 得られた数値データを用いて 3.3-3.5 の処理を行うことで, CNN で利用可能な特徴量ベクトルを作成する.

### 3.4 スケール変換

OpenPose で得られた 2次元の姿勢情報データは, 1 画像中のどこに各関節が存在しているかを 2次元座標で表したものである. 使用した映像では, 被験者がカメラに近づいたり, カメラから遠ざかったりしている. したがって, そのままでは座標のスケールが不揃いである. そこで, ある関節間の長さを 1 とし, 2次元データのスケールを一致させる. 提案手法では, 様々な方向からの撮影で, 映像中の 2次元の長さの変化が最も少なく, 検出失敗が少ない大腿部を用いる. 更に, 右大腿部の長さが 1 になるように姿勢情報をスケール変換する.

変換関数を  $F$ , 右大腿部の長さを 1 とした時に, 右尻の座標を  $X_{righthip}(x_{righthip}, y_{righthip})$ , 右膝の座標を  $X_{rightknee}(x_{rightknee}, y_{rightknee})$ , とすると姿勢情報  $X_i(x_i, y_i)$ , ( $i$  は OpenPose の出力パーツ名) は以下の式 3.1 で表される.

$$f(X) = \frac{X(x,y)}{(|x_{righthip} - x_{rightknee}|, |y_{righthip} - y_{rightknee}|)} \quad (3.1)$$

この式 3.1 を用いて全ての 2 次元データをスケール変換する.

### 3.5 位置補正

スケール変換の意義を述べた通り, 使用する映像では常に被験者の映る場所が変化する. OpenPose の出力は 2 次元座標であるため, スケール変換に加えて座標の位置補正を行う必要がある. そこで, 被験者の首の位置が原点となるように位置補正を行う.

首の位置を原点とした時に, 首の座標を  $X_{neck}(x_{neck}, y_{neck})$ , 姿勢情報を  $X_i(x_i, y_i)$  とすると, 位置補正後の姿勢情報  $X(x, y)$  は以下の式 3.2 で表される.

$$X(x, y) = X_i - X_{neck} = (x_i - x_{neck}, y_i - y_{neck}) \quad (3.2)$$

この式(3.1)を使って全ての 2 次元データを位置補正する. スケール変換と位置補正を両肩, 両肘, 両手首, 両尻, 両膝, 両足首の 12 点において行い, 2 次元姿勢情報の 24 次元ベクトルの特徴量とする.

### 3.6 3 次元角度の算出

3d-pose-baseline で得られた 3 次元の姿勢情報から, 被験者の歩容特徴が多く表現されると考えられる関節間の 3 次元角度を算出する. この角度を歩容情報とする. 対象となる角度は, 尾骶骨, 尻, 膝の 3 点, 尾骶骨, 尻, 足首の 3 点, 肩, 肘, 手首の 3 点がなす角度で合計 6 個である.

また, 手足の振り幅として, 右手首と左足首, 左手首と右足首の間の 3 次元距離を算出し歩容情報とする.

3 次元空間上の 3 点 A, B, C が与えられた時,  $\angle ACB$  を  $\theta^\circ$  とする. この時, 点 C から点 A へのベクトルを  $\vec{CA} = (a_1, a_2, a_3)$ , 点 C から点 B へのベクトルを  $\vec{CB} = (b_1, b_2, b_3)$  とすると,  $\theta$  は以下の式 3.3 で表される.

$$\theta = \cos^{-1} \frac{a_1 b_1 + a_2 b_2 + a_3 b_3}{\sqrt{a_1^2 + a_2^2 + a_3^2} \sqrt{b_1^2 + b_2^2 + b_3^2}} \quad (3.3)$$

また, 3 次元空間上の 2 点 A( $A_x, A_y, A_z$ ), B( $B_x, B_y, B_z$ ) が与えられた時, 2 点の距離を D とすると, D は以下の式 3.4 で表される.

$$D = \sqrt{(A_x - B_x)^2 + (A_y - B_y)^2 + (A_z - B_z)^2} \quad (3.4)$$

式(3.3)と(3.4)を用いて、得られたデータから特徴量を算出する。これらを歩容情報の 8 次元ベクトル特徴量とする。

### 3.7 CNN を用いた分類

提案手法により得られた 32 次元の特徴量を CNN で分類する。CNN は入力層, 中間層, 出力層の 3 層からなり, 中間層の入力は 256 次元, 出力は 64 次元である。モデルの実装は Keras の Sequential[10]モデルを使用し, 最適化アルゴリズムは Adam(Adaptive Moment Estimation)[11]を用いる。

### 3.8 むすび

本章では, 提案手法の概要述べた。提案手法の詳細として, Openpose から得られた姿勢情報の位置補正とスケール変換, 3d-pose-baseline から得られた姿勢情報を用いた関節角度の算出について述べた。更に, 得られた情報を用いて CNN で個人を識別する手法について述べた。

## 第4章 実験および結果・考察

### 4.1 まえがき

本章では，様々な方向から撮影した映像を用いて 4 種類の実験を行った結果を示す．実験の結果を評価し，考察を述べる．実験 1 では，歩行中の人物を前方から撮影した映像を学習データとテストデータに用いて実験を行なった．実験 2 では，歩行中の人物を後方から撮影した映像を学習データとテストデータに用いて実験を行なった．実験 3 では，歩行中の人物を右斜め前方，右斜め後方，左斜め前方，左斜め後方の 4 方向から撮影した映像を学習データとテストデータに用いて実験を行なった．実験 4 では，歩行中の人物を右斜め前方，右斜め後方，左斜め前方，左斜め後方の 4 方向から撮影した映像を学習データに用い，前方と後方から撮影した映像をテストデータに用いて実験を行なった．

### 4.2 映像データ撮影

早稲田大学学生会館 10 階の廊下，及び 2 階の中庭で映像データを撮影する．映像は 60fps, 1920×1080[pe]である．

### 4.3 実験 1 前方から撮影した映像を用いた実験

5 人の被験者を前方向から撮影した映像を用いて，識別実験を行った．学習データとテストデータはそれぞれ別の場所で撮影した．学習データは 1 人につき 1000 件，テストデータは 1 人につき 100 件で合計 5500 件である．姿勢情報のみ，歩容情報のみ，両方の情報を用いる 3 パターンで実験した．結果は accuracy と confusion matrix により評価する．実験の結果を表 4.1-4.4 と図 4.1 に示す．

表 4.1 実験 1 精度結果

	骨格情報のみ	歩容情報のみ	姿勢情報と歩容情報
Accuracy	92.0%	84.8%	98.6%

表 4.2 姿勢情報のみを用いた識別結果の confusion matrix(実験 1)

	被験者 A	被験者 B	被験者 C	被験者 D	被験者 E
被験者 A	98	1	1	0	0
被験者 B	5	95	0	0	0



被験者 C	3	0	88	4	5
被験者 D	1	1	0	82	16
被験者 E	1	0	0	4	95

表 4.3 歩容情報を用いた識別結果の confusion matrix(実験 1)

	被験者 A	被験者 B	被験者 C	被験者 D	被験者 E
被験者 A	98	1	0	0	1
被験者 B	12	61	0	19	8
被験者 C	7	1	78	13	1
被験者 D	6	4	3	100	0
被験者 E	4	1	2	1	92

表 4.4 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 1)

	被験者 A	被験者 B	被験者 C	被験者 D	被験者 E
被験者 A	100	0	0	0	0
被験者 B	3	97	0	0	0
被験者 C	1	0	95	1	3
被験者 D	0	0	0	100	0
被験者 E	0	0	0	0	100

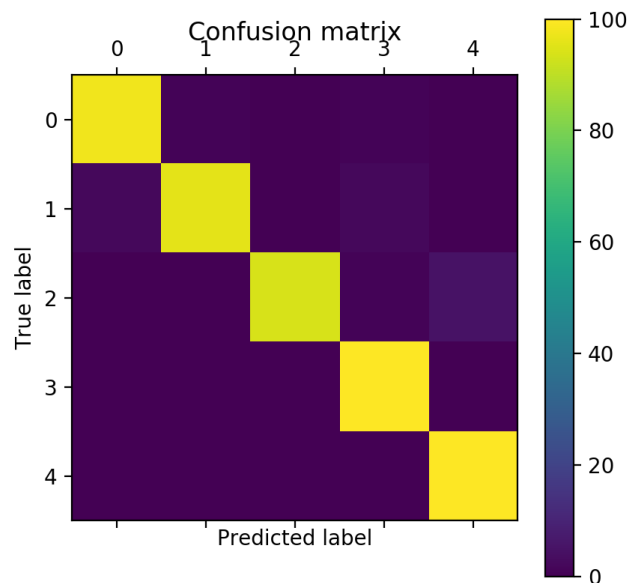


図 4.1 姿勢情報と歩容情報を用いた分類結果の confusion matrix(実験 1)

表 4.1 に示した通り, 姿勢情報と歩容情報両方を用いた識別の精度は 98.6% と非常に高い結果となった. 前方からの映像では, 誤検出が発生しにくいため, 高い精度で識別できると考えられる. 歩容情報のみを用いた識別の精度は, 姿勢情報のみを用いた時に比べて劣る. しかし, 両方を用いた場合に最も高い精度を示しているため, 姿勢情報に歩容情報を加えて識別することは有用であると言える.

#### 4.4 実験 2 後方から撮影した映像を用いた実験

5 人の被験者を後ろ方向から撮影した映像を用いて, 分類実験を行った. 学習データとテストデータはそれぞれ別の場所で撮影した. 学習データは 1 人につき 800 件, テストデータは 1 人につき 100 件で合計 4600 件である. 姿勢情報のみ, 歩容情報のみ, 両方の情報を用いる 3 パターンで実験した. 結果は accuracy と confusion matrix により評価する. 実験の結果を表 4.5-4.8 と図 4.2 に示す.

表 4.5 実験 2 精度結果

	骨格情報のみ	歩容情報のみ	姿勢情報と歩容情報
Accuracy	80.4%	54.0%	82.2%

表 4.6 姿勢情報のみを用いた識別結果の confusion matrix(実験 2)

	被験者 A	被験者 B	被験者 C	被験者 D	被験者 E
被験者 A	92	2	4	0	2
被験者 B	9	82	0	0	9
被験者 C	1	25	63	0	11
被験者 D	0	7	0	85	8
被験者 E	6	15	0	0	79

表 4.7 歩容情報のみを用いた識別結果の confusion matrix(実験 2)

	被験者 A	被験者 B	被験者 C	被験者 D	被験者 E
被験者 A	31	44	15	1	9
被験者 B	0	47	0	17	36
被験者 C	8	12	74	1	5
被験者 D	0	7	0	90	3
被験者 E	3	23	1	35	38

表 4.8 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 2)

	被験者 A	被験者 B	被験者 C	被験者 D	被験者 E
被験者 A	82	0	18	0	0
被験者 B	0	90	0	0	10
被験者 C	10	14	68	0	8
被験者 D	3	4	0	90	3
被験者 E	0	17	0	0	83

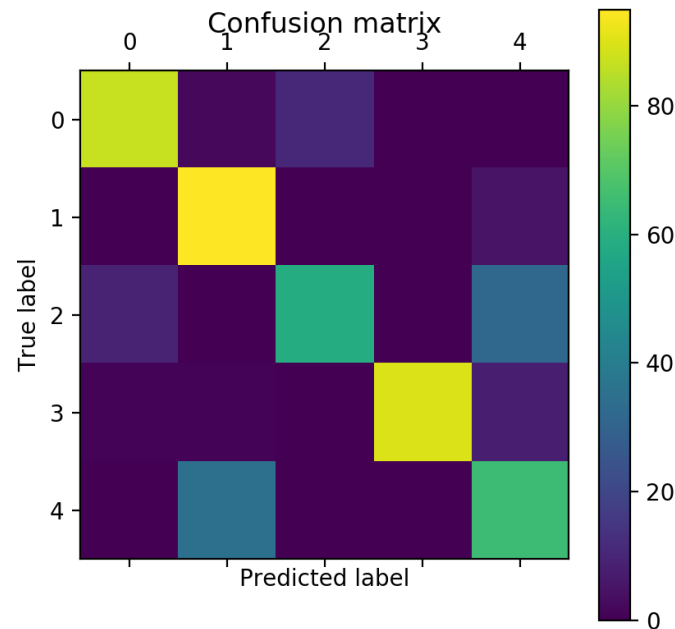


図 4.2 姿勢情報と歩容情報を用いた分類結果の confusion matrix(実験 2)

表 4.1 に示した通り，姿勢情報と歩容情報両方を用いた識別の精度は 82.2% となった．歩容情報のみを用いた識別精度は低下しているが，姿勢情報に歩容情報を加えた場合に精度が向上している．後ろからの映像では腕を前に振る，足を前に出すといった大きな動作が画像の奥に向かって映っている．そのため 3 次元の姿勢が 3d-pose-baseline により推定し難く，歩容情報の精度が低いと考えられる．

#### 4.5 実験 3 斜め 4 方向から撮影した映像を用いた実験

3 人の被験者を右前方，右後方，左前方，左後方の 4 方向から撮影した映像を用いて分類実験を行った．テストデータには，学習データと同じ 4 方向から撮影した映像を用いた．学習データは 1 人につき 200 件，テストデータは 1 人につき 150 件で合計 1050 件である．姿勢情報のみ，歩容情報のみ，両方の情報を用いる 3 パターンで実験した．結果は accuracy と confusion matrix により評価する．実験の結果を表 4.9-4.12 に示す．姿勢推定に失敗したデータが存在したため，分類時に削除した．

表 4.9 実験 3 精度結果

	骨格情報のみ	歩容情報のみ	姿勢情報と歩容情報
Accuracy	74.1%	76.9%	88.2%

表 4.10 姿勢情報のみを用いた識別結果の confusion matrix(実験 3)

	被験者 A	被験者 B	被験者 C
被験者 A	93	12	2
被験者 B	39	71	10
被験者 C	11	8	109

表 4.11 歩容情報のみを用いた識別結果の confusion matrix(実験 3)

	被験者 A	被験者 B	被験者 C
被験者 A	73	12	25
被験者 B	20	84	10
被験者 C	15	10	106

表 4.12 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 3)

	被験者 A	被験者 B	被験者 C
被験者 A	94	7	2
被験者 B	18	95	4
被験者 C	7	3	125

表 4.9 に示した通り、姿勢情報と歩容情報を用いた識別結果は 88.2%であった。このことから学習済みの方向からの映像であれば、ある程度識別可能なことが分かる。しかし、学習データと同じ方向からの映像を用意できない場合も想定しなければならない。その場合の実験を次の実験 4 で行う。

#### 4.6 実験 4 斜め 4 方向と前後方から撮影した映像を用いた実験

3 人の被験者を右前方、右後方、左前方、左後方の 4 方向から撮影した映像を用いて分類実験を行った。テストデータには、学習データと異なる方向から撮影した映像を用いた。学習データは 1 人につき 200 件、テストデータは 1 人につき 150 件で合計 1050 件である。姿勢情報のみ、歩容情報のみ、両方の情報を用いる 3 パターンで実験した。結果は accuracy

と confusion matrix により評価する。実験の結果を表 4.13-4.16 に示す。姿勢推定に失敗したデータが存在したため、分類時に削除した。

表 4.13 実験 4 精度結果

	骨格情報のみ	歩容情報のみ	姿勢情報と歩容情報
Accuracy	88.0%	14.8%	64.4%

表 4.14 姿勢情報のみを用いた識別結果の confusion matrix(実験 4)

	被験者 A	被験者 B	被験者 C
被験者 A	0	0	0
被験者 B	36	14	0
被験者 C	30	1	68

表 4.15 歩容情報のみを用いた識別結果の confusion matrix(実験 4)

	被験者 A	被験者 B	被験者 C
被験者 A	0	0	0
被験者 B	1	22	0
被験者 C	11	88	68

表 4.16 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 4)

	被験者 A	被験者 B	被験者 C
被験者 A	0	0	0
被験者 B	4	0	46
被験者 C	11	1	87

表 4.13 に示した通り、姿勢情報と歩容情報を用いた識別結果は 64.4%であった。このことから、学習していない方向から撮影した映像では正確な識別は困難であることが分かる。原因としては、歩容情報のみを用いた識別結果が非常に悪いことが挙げられる。斜め方向

からの映像では、身体の陰に入るパーツが多く、3次元の姿勢情報の推定が困難である。この問題を解決しない限りは、2次元の姿勢情報のみを用いたほうが高い精度で識別できる。

#### 4.7 むすび

本章では、4種類の実験の結果とそれに対する考察を述べた。前方から撮影した映像での識別精度が最も高かった。斜めから撮影した映像を使用した場合、身体パーツの誤認識による識別精度の低下が見られた。

## 第5章 結論と今後の課題

本研究では、人物の歩行映像に対して OpenPose と 3d-pose-baseline を用いて姿勢推定を行い、得られた姿勢情報から人物を識別する手法を提案した。

CNN を用いた識別実験では、前方からの映像において Accuracy で 98.6%の精度が出ることを確認し、高い精度で個人識別可能なことがわかった。従来手法との比較においても、単眼 RGB カメラでデプスカメラやセンサと同じように個人を識別できることがわかった。一方で前方以外から撮影された映像を用いた場合の精度には課題が残った。姿勢推定失敗時に身体パーツの補完を行う仕組みを取り入れることで、使用する映像の自由度が高い個人識別が可能になると考えられる。



**謝辞**

本研究の環境を与えてくださり，研究のテーマ設定や方向性，また問題点の提起や整理など丁寧かつ適切なお指導をくださった渡辺裕教授と国際情報通信研究センター石川孝明氏に深く感謝申し上げます。また，日頃から問題点解決の相談に乗っていただき，貴重なご意見をくださる研究室の皆様に深く御礼申し上げます。

## 参考文献

- [1] 岩間, 村松, 榎原, 八木, “犯罪捜査支援のための歩容に基づく人物鑑定システム”, 情報処理学会研究報告, Vol. 2013-CVIM-186(3), pp.1-10, Apr. 15, 2013
- [2] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), pp.1302-1310, July 2017.
- [3] 戸田, Moro, 梅田, “単眼カメラから得られる骨格情報を用いた人物識別 -CNN と SVM の比較-”, The Japan Society for Precision Engineering, Vol. D69, pp.267-268, May 2018
- [4] 三堀, 花泉, “Kinect V2 を用いる歩容認識に基づく個人識別手法”, 情報処理学会第 79 回全国大会, 4P-89, pp.2-439-2-440, May 2017
- [5] 森, 菊池, “深度センサによる歩容特徴量を用いた個人識別・追跡方式の提案”, Computer Security Symposium 2017, pp.972-979, Oct. 2017
- [6] J. Martinez, R. Hossain, J. Romero, and J. J. Little, “A simple yet effective baseline for 3d human pose estimation,” IEEE Conference on Computer Vision (ICCV 2017), pp.2659-2668, Oct. 2017.
- [7] Y. Makihara, H. Mannami, and Y. Yagi, “Gait Analysis of Gender and Age Using a Large-Scale Multi-View Gait Database,” Proc. the 10th Asian. Conf. on Computer Vision, Queenstown, New Zealand, pp. 975-986, Nov. 2010.
- [8] 今岡仁, “NEC の顔認証技術と応用事例”, 情報処理学会研究報告, Vol. 2013-CVIM-187, No.38, pp.1-4 May. 2013
- [9] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments,” IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.36, No.7, pp.1325-1339, July 2014.
- [10] 「Sequential モデル - Keras Documentation」,  
<<https://keras.io/ja/models/sequential/>>, 2019/01/20
- [11] P. Diederik, and J. B. Kingma, “Adam: A Method for Stochastic Optimization,” International Conference on Learning Representations, arXiv:1412.6980 [cs.LG] , Dec. 2014

## 図一覧

図 2.1 姿勢推定前画像(1) .....	5
図 2.2 2次元姿勢推定成功例 .....	5
図 2.3 姿勢推定前画像(2) .....	5
図 2.4 2次元姿勢推定失敗例(右前腕) .....	5
図 2.5 3d-pose-baseline で使用されるニューラルネットワーク構造 .....	6
図 2.6 姿勢推定前画像(3) .....	7
図 2.7 3次元姿勢推定成功例 .....	7
図 2.8 姿勢推定前画像(4) .....	7
図 2.9 3次元姿勢推定失敗例(右前腕) .....	7
図 4.1 姿勢情報と歩容情報を用いた分類結果の confusion matrix(実験 1) .....	14
図 4.2 姿勢情報と歩容情報を用いた分類結果の confusion matrix(実験 2) .....	16

## 表一覧

表 2.1 OpenPose 検出部位一覧.....	4
表 4.1 実験 1 精度結果 .....	12
表 4.2 姿勢情報のみを用いた識別結果の confusion matrix(実験 1) .....	12
表 4.3 歩容情報を用いた識別結果の confusion matrix(実験 1).....	13
表 4.4 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 1).....	13
表 4.5 実験 2 精度結果 .....	14
表 4.6 姿勢情報のみを用いた識別結果の confusion matrix(実験 2) .....	15
表 4.7 歩容情報のみを用いた識別結果の confusion matrix(実験 2) .....	15
表 4.8 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 2).....	15
表 4.9 実験 3 精度結果 .....	16
表 4.10 姿勢情報のみを用いた識別結果の confusion matrix(実験 3).....	17
表 4.11 歩容情報のみを用いた識別結果の confusion matrix(実験 3).....	17
表 4.12 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 3) .....	17
表 4.13 実験 4 精度結果.....	18
表 4.14 姿勢情報のみを用いた識別結果の confusion matrix(実験 4).....	18
表 4.15 歩容情報のみを用いた識別結果の confusion matrix(実験 4).....	18
表 4.16 姿勢情報と歩容情報を用いた識別結果の confusion matrix(実験 4) .....	18