

# 後方二輪車挙動把握のための 人体スケルトン検出を用いた二輪車姿勢推定手法

出水 利治<sup>1,a)</sup> 中村 嘉隆<sup>1</sup> 稲村 浩<sup>1</sup> 木谷 友哉<sup>2</sup>

**概要:** 近年注目されている安全支援運転技術・自動運転技術において、自動車周辺環境の認識は重要な課題である。特に、自動車後方に存在する二輪車については、その形状・移動性から四輪車と比べて複雑な挙動をとるため、その把握が重要と考えられる。これまでに物体としての二輪車を検出する手法は提案されているが、二輪車の挙動の詳細を把握可能な手法は存在しない。そこで、本研究では二輪車の挙動を把握する手法に着目する。本稿では前段階として自動車の後方を走行する二輪車の運転手の姿勢に注目し、我々が準備したテンプレートデータと二輪車ライダーの姿勢を比較することで、二輪車ライダーを検出する手法を提案した。実際の二輪車ライダーを用いた実験結果、二輪車が撮影車両方向を向いている状態に近い場合には、二輪車ライダーの検出が可能であることが分かった。

**キーワード:** 自動運転, 安全運転支援, 二輪車, 人体骨格, 6DoF

TOSHIHARU IZUMI<sup>1,a)</sup> YOSHITAKA NAKAMURA<sup>1</sup> HIROSHI INAMURA<sup>1</sup> TOMOYA KITANI<sup>2</sup>

## 1. はじめに

自動車が自律的に周辺状況をセンシングし、運転行動の決定を行う自動運転は、交通渋滞の解消や、交通事故の根絶など、社会やドライバーに多大な利益をもたらすことが予想され、その実現への期待が高まっている。自動運転のための要素技術において、周辺環境の認識は重要な課題である。中でも道路上の他車両の挙動についての情報は安全な交通のために欠かせない。現在、自動車運転補助技術としての周辺環境のセンシングには視覚センサ及び画像処理技術が主に使用されており、視覚センサから得られた画像を分析することで、周辺障害物の種類や位置の検出を行っている。二輪車はその形状・移動性・運転操作の違いから四輪車と比べて複雑な挙動をとるため、その把握及び予測が難しい。神村らは二輪車の旋回時の傾きの情報から、二輪車の挙動予測を目指している<sup>1)</sup>。このように、二輪車に対しては、四輪車のような車両位置の検出および追跡だけではなく、車両の傾きおよび向きまで検出する必要がある。一方、二輪車は車両形状と乗車方法から、二輪車ライダー自身の身体が外部から観測可能であるという特徴があ

る。そこで本研究では、二輪車運転時の二輪車ライダーの人体骨格（スケルトン）情報に着目し、二輪車ライダーの姿勢から二輪車車両の傾きや向きを検出することを検討する。スケルトン情報の取得には Microsoft 社の Kinect<sup>2)</sup> を使用することが一般的である。しかし、近年の深層学習等による画像処理技術の精度向上にとともに、2次元画像からスケルトンを推定する技術が開発されており、最も一般的なものとして OpenPose<sup>3)</sup> が知られている。本稿で提案するシステムは、まず車載の外部カメラによる後方画像から OpenPose を用いた人体のスケルトン推定を行い、後方に存在する人間のスケルトン情報を得る。取得したスケルトン情報に対し、人体モデルのパターンマッチングを用いて二輪車ライダーを検出する。次に、二輪車ライダーのスケルトン情報に対して、物体の自由度に関する指標の1つである 6DoF(6 degrees of freedom) を用い、二輪車の傾きや向きを推定し挙動の予測を行う。本稿では、この手法に対し、二輪車ライダーを検出するための前段階として、人体モデルのパターンマッチングによる類似度を算出し、実証実験を行うことで提案手法の有効性を確認する。

<sup>1)</sup> 公立はこだて未来大学 大学院システム情報科学研究科

<sup>2)</sup> 静岡大学 大学院総合科学技術研究科

<sup>a)</sup> g2118004@fun.ac.jp

## 2. 関連研究

### 2.1 人体姿勢推定

OpenPose<sup>2)</sup>は、人物画像と Kinect によって深度情報をもとに取得された人物の三次元スケルトン推定結果が合致するように、畳み込みニューラルネットワークによって学習させた結果を用い、画像情報のみからスケルトン情報を出力するライブラリである。このライブラリによって、深度情報を用いずに人の頭部・肩・肘・手・腰・膝・足を検出し、人がどのような姿勢であるかの推定が可能である。深度情報を使用しないため、比較的遠距離でもスケルトンを取得することが可能である。一方、Kinect によって取得された情報との画像マッチングによる学習時に用いる訓練データセットには人体が直立状態である画像データが多く用いられているため、直立以外の姿勢を推定する場合は、それらの姿勢を含むデータセットを用意する必要がある。また、二輪車乗車中のライダーなど、人体の一部が二輪車車体に隠れるような場合は、誤推定の原因となる。例として図??では、右膝より下の骨格が写真左上の座標と誤推定されている。また、二次元座標として取得される関節位置情報から距離情報を含んだ三次元座標への展開も今後の研究課題となっている<sup>3)</sup>。



図 1 誤推定の例 (右膝・右足が車体で隠れている)

### 2.2 画像の三次元人体姿勢推定

Rhodin<sup>4)</sup>らはマルチビューベースでの二次元スケルトン情報から三次元スケルトン情報への拡張を行っている<sup>5)</sup>。複数の映像から個別に人体姿勢推定を行い、これらの視差情報をもとに重回帰式で三次元人体姿勢推定を行う手法である。しかし、ステレオカメラや複数のカメラが必要となるため、拡張に対するコストが高い点が問題となる。工藤らは二次元の関節位置が与えられたときに、三次元人体姿勢のモデルを作成する手法<sup>6)</sup>を提案している。この手法は、 $z$ 成分が正しく推定されれば、出力される三次元姿勢は $y$ 軸周りの回転に対してロバストであるとの仮定に基づき、各

$z$ 成分が正しく推定できるまで学習させて、三次元人体姿勢データの生成に用いるという方法をとっている。また、この手法で推定された三次元人体姿勢を回転させて二次元平面に投影すると、人体の姿勢として破綻のない形状となることが確認されている。本研究では二次元人体画像から三次元スケルトン情報を推定する際に、この手法を用いることとする。

### 2.3 姿勢推定における類似度推定

柴田は人体姿勢の時系列データからコサイン類似度を用いて、人の静動判別を行い動作を切り出している<sup>7)</sup>。Kinect から取得した三次元姿勢データを用い、関節情報から隣り合った体節ベクトルを作成し、毎時刻、直前のフレームとの二つの姿勢データの対応する体節ベクトルについてそれぞれ類似度を計算し、体節ベクトル類似度の和から姿勢全体の類似度を算出している。しかし、Kinect のような三次元スケルトンを取得できるデバイスが必須であることが問題点として挙げられる。西澤は人物画の学習支援のため、描いた人物画から関節座標データを取得し、人体モデルと人物画の類似度を用いて評価を行っている<sup>8)</sup>。この手法では、文献<sup>9)</sup>と同様に、姿勢ベクトルを作成し、各ベクトルのコサイン類似度を用いることで、実際の人物と人物画の類似度を算出している。さらに、姿勢ベクトルの大きさの差に対して重み付けを行うことで、各関節間の類似度を加味した上での全体の姿勢ベクトルの類似度を算出することを可能にしている。しかし、二次元の映像を使用しているため、関節座標を人物の正面から取得する制約があることが挙げられる。

## 3. 提案システム

### 3.1 概要

提案するシステムは四輪車後方の二輪車の挙動予測のために、人体姿勢検出を用いた二輪車姿勢推定を行う。道路運送車両法・第二条列表第一<sup>1)</sup>より二輪車の高さは2.0m以下であるとされている。概ね水平な位置から二輪車ライダーを撮影することで、スケルトン取得時の誤推定を減少させられるため、四輪車後方高さ1.5mの位置への車載外部カメラの設置を想定する。車載外部カメラによる後方画像から OpenPose を用いた人体スケルトン推定を行う。このスケルトン情報に対し、人体モデルのテンプレートマッチングを用いて二輪車ライダーを検出する。次に、検出した二輪車ライダーのスケルトン情報に対して $z$ 成分を推定することで三次元に拡張する。この三次元スケルトン情報の、6DoF(6 degrees of freedom)を適用して、二輪車ライダーの姿勢を検出し、二輪車車体の傾きや向きを推定することで二輪車の挙動の予測を行う。本稿では、人体モデルのテンプレートマッチングを用いて二輪車ライダーを検出するための類似度算出手法について述べる。

### 3.2 二輪車ライダーのテンプレートデータ作成

画像中から二輪車ライダーを検出するためにテンプレートデータを利用する。テンプレートデータとしては二輪車運転時の基本姿勢であるリーンウィズ図(??)と呼ばれる姿勢を用いる。リーンウィズでは図(??)のように、二輪車車体が直立状態であり、二輪車ライダーの上半身も二輪車車体と同様直立状態になっている。また、四輪車後方から撮影した場合、後方二輪車は概ねリーンウィズの状態取得されると考えられる。



図 2 リーンウィズ

### 3.3 二次元人体姿勢データの取得

人体姿勢データの取得には、OpenPoseを使用する。使用カメラのフレームレートが30fpsであるため、1秒間に30枚の画像データを切り出して使用する。OpenPoseによって出力された頭・首・胴体・肩・肘・腰・足の計16点の関節部分の二次元位置座標から図(??)のような二次元人体姿勢データを取得する。

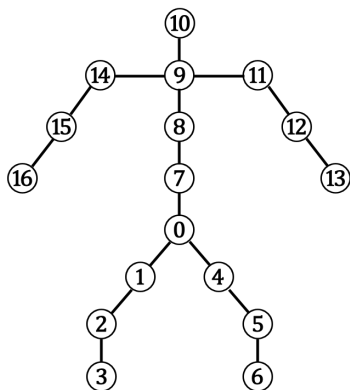


図 3 二次元姿勢データ

### 3.4 三次元姿勢データへの拡張

OpenPoseで取得された二次元人体姿勢データに対し、文献?の手法を用いることで三次元人体姿勢データに拡張する。三次元人体姿勢データへの拡張のための姿勢データの学習時には文献?でも用いているMPII Human Pose Dataset?を用いる。MPII Human Pose Datasetはトレーニング用の250万枚の画像とテスト用の3000枚の画像からなり、40,000人以上の人体を含む25,000の画像が含まれており、410個の人間の活動が含まれ、それぞれに活動ラベルが付いている。さらに画像における人体姿勢データには16箇所の関節の位置がアノテーションされている。各関節の二次元座標からz座標を推定して三次元人体姿勢データを生成した上で再度二次元に投影し、投影した結果が二次元人体姿勢データとして存在するもののみが正しいデータとして学習される(図??)。

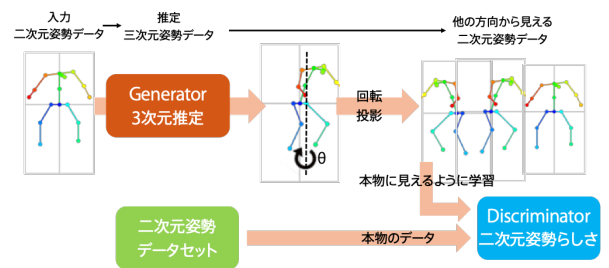


図 4 三次元姿勢データのための学習モデル

この学習結果を用いることにより、任意の二次元人体姿勢データを三次元人体姿勢データに拡張できる(図??)。拡張された三次元人体姿勢データに対し、Human3.6M Dataset?と呼ばれる三次元人体姿勢のデータセットを用いて定性的に評価した結果、二次元人体姿勢データから拡張された三次元人体姿勢データは人体姿勢として破綻がないものであることを確認している?。この拡張手法を用いて車載外部カメラから取得した二輪車ライダーの二次元人体姿勢データをフレームごとに三次元人体姿勢データに拡張する。テンプレートデータを三次元拡張した結果を図??に示す。

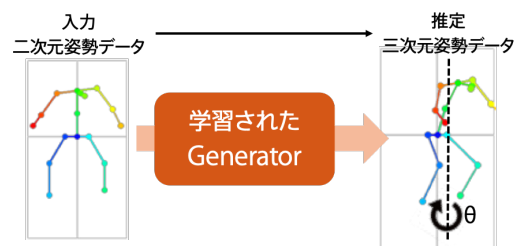


図 5 三次元姿勢データ推定

同様に、カメラより取得した二輪車ライダーの姿勢も毎フレーム三次元拡張処理を行う。

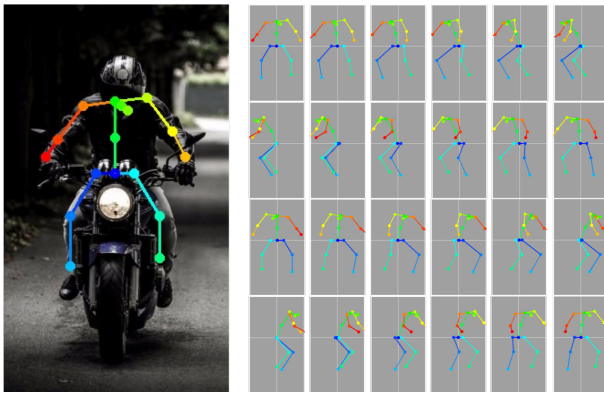


図 6 三次元拡張されたテンプレートデータ

### 3.5 三次元人体姿勢データの類似度

車載外部カメラ画像から取得された二輪車ライダーの姿勢データは様々な向きのものであることが想定される。そのため提案システムでは、どのような向き・傾きをもつ姿勢データからであっても二輪車ライダーが検出可能である必要がある。そのため、二輪車ライダーのテンプレートデータと取得した姿勢データの類似度を算出し、ある程度高い類似度が得られた姿勢データを二輪車ライダーとみなすことで、二輪車ライダーの検出を行う。人体姿勢に関する類似度推定手法にコサイン類似度を用いた類似度推定??があるが、提案システムでは人体姿勢データを三次元に拡張しており、さらに取得されるスケルトン情報は向き・傾き・距離による大きさの変化があるため、コサイン類似度を用いた類似度推定は適用することが難しい。そのため提案システムではテンプレートデータと拡張された三次元人体姿勢データのそれぞれに対して正規化を行う。

正規化は以下の(1)~(4)4つの処理に分類される。以下拡張されたテンプレートデータを  $A$ 、拡張された三次元人体姿勢データを  $B$  とする。図??に三次元人体姿勢データの正規化手法を示す。

- (1) 人体の重心計算
- (2) 遠近感によるスケールの正規化
- (3) 回転方向の規格化
- (4) 各骨格の距離算出

#### (1) 人体の重心計算

三次元に拡張されたテンプレートデータと取得した三次元人体姿勢データの位置合わせを行う。三次元に拡張されたテンプレートデータと拡張された三次元人体姿勢データそれぞれについて、各関節の座標からそれぞれのデータにおける重心の座標を求め、三次元データを重心から原点に向かって並行移動することで、重心が原点 ( $A_0, B_0$ ) になるようにしてそれぞれのデータの座標空間を同一にする。位置合わせのための式は以下の通りである。ここで  $\bar{A}$ 、 $\bar{B}$  は原点 (0,0,0) に対しての距離を表している。

$$A_0 = A - \bar{A}, B_0 = B - \bar{B} \quad (1)$$

#### (2) 遠近感によるスケールの正規化

三次元テンプレートデータと取得した三次元人体姿勢データそれぞれに正規化処理を行うことで遠近差によるデータの差を排除する。 $F$  はフロベニウスノルムを示す。フロベニウスノルムとは、行列の大きさを表す量である。 $\|X_0\|_F$  は関節座標の行列のノルムを示す、つまり三次元人体姿勢データのスケールを示す。以下に式を示す。

$$A_{norm} = \frac{A_0}{\|A_0\|_F}, B_{norm} = \frac{B_0}{\|B_0\|_F} \quad (2)$$

#### (3) 回転方向の規格化

正規化した三次元テンプレートデータと取得した三次元人体姿勢データを回転して人体の向きをあわせる。三次元人体姿勢データを三次元テンプレートデータの人体の向きに合わせる際の回転行列  $R$  は以下のように定義できる。

$$R = \operatorname{argmin}_{\Omega} \|\Omega A - B\|_F \text{ s.t. } \Omega^T \Omega = I \quad (3)$$

#### (4) 各骨格の距離算出

正規化及び回転処理を終えた三次元テンプレートデータおよび三次元人体姿勢データ間の類似度を算出する。通常、二輪車ライダーの骨格には二輪車車体によって隠される部分が存在するため、OpenPose による検出時に誤検出の可能性がある。そこで、頭部・左右の足首の三次元骨格情報は用いず、その他の14点の関節座標のみを使用し類似度を算出する。正規化および回転処理完了後の両データの対応する関節座標間の距離の最小二乗和を計算し、その値を両データ間の類似度とする。このときの計算式は式(4)のようになる。ここで  $i$  は  $(x, y, z)$  座標であり、 $j$  は14点の関節座標を表している。

$$\sum_{i=1}^3 \sum_{j=1}^{14} (A_{ij} - B_{ij})^2 \quad (4)$$

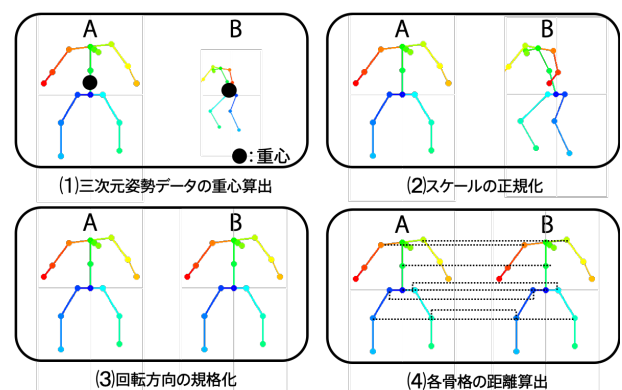


図 7 三次元人体姿勢データの正規化

三次元テンプレートデータと三次元人体姿勢データの類似度を算出するにあたって、プロクラステス分析?を用いる。プロクラステス分析とは点同士の対応がとれた2つ

の点群に対して、並進・回転・一様なスケーリングの変換のもとで、点群間の二乗誤差が最小になるように重ね合わせる処理である。この処理を正規化された三次元人体姿勢データと三次元に拡張されたテンプレートデータの各関節の座標を元に解析を行う。

## 4. 実験

提案システムによって算出された類似度により、二輪車ライダーとそれ以外の姿勢の人体を判別できるかどうかについて、実際にモデルを作成して評価する。二輪車ライダーのテンプレートデータを用い、二輪車ライダーの姿勢データとの類似度が高い、すなわち姿勢データ間の距離が0に近く、歩行者を想定した姿勢データとの類似度が低い、すなわち姿勢データ間の距離が大きくなっていけば、有効に働いていると評価できる。また、姿勢データを回転させた場合の類似度の変動を評価することで、実環境における様々な方向の二輪車および歩行者を判別できるかどうかについても評価する。

### 4.1 実験環境

本実験は、2019年4月公立はこだて未来大学の研究室で行なった。撮影に用いたカメラは地上から1.2mの位置であり、実験車両・被験者までの距離は2.0mの位置に設置を行った(図??)。カメラのスペックを表??に、実験車両の寸法を表??に示す。

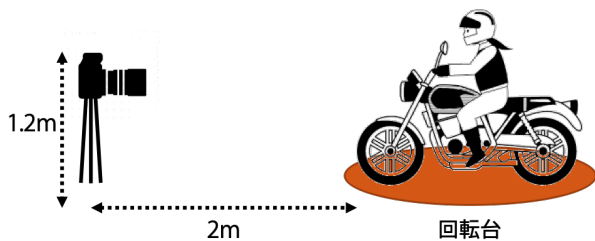


図 8 実験車両・被験者の撮影位置

表 1 使用カメラの仕様

使用カメラ	iPhoneX
解像度	1080p
フレームレート	30fps
焦点距離	4mm
深度視野	(水平)63° (垂直)50°

表 2 実験車両の寸法

車種名	車両種	全長 (mm)	全幅 (mm)	全高 (mm)
セロー 225	二輪車	2070	805	1140

### 4.2 実験シナリオ

外部カメラを用いて図??のように二輪車ライダーおよび歩行者を想定した3種類の姿勢の歩行者の撮影を行う。様々な角度からの撮影を考慮するため、これら4つの姿勢データはそれぞれ水平方向に30°ずつ回転させ、4×12パターンの計52パターンの画像を撮影する。

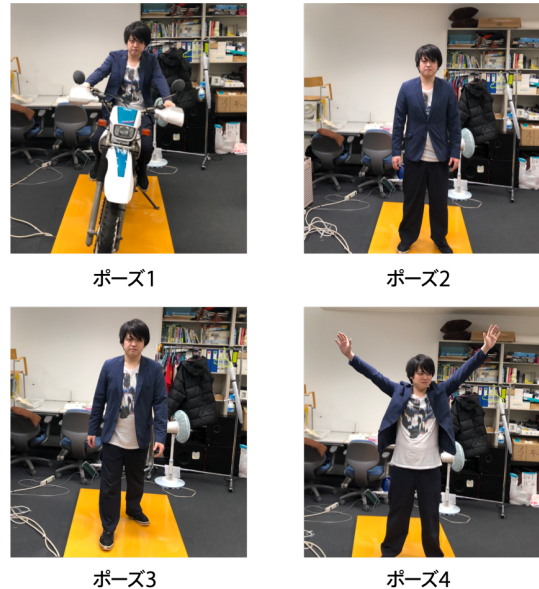


図 9 実験に使用する4つのポーズ

### 4.3 類似度判定結果

まず、二輪車ライダー(ポーズ1)についてテンプレートデータとの類似度評価を行う。水平方向に30°ずつ回転させた12パターンの姿勢データそれぞれについての類似度算出結果を図??に示す。縦軸は類似度、横軸は水平方向に回転した際の角度を示している。テンプレートデータと各姿勢データとの関節座標間の距離の最小二乗和を類似度として算出しており、類似度の値が大きいほど二輪車ライダーと類似しておらず、0に近いほど二輪車ライダーに近いと判定される。水平方向の回転が0°~120°の間では、類似度は0.1~0.24の間の値をとっており、150°~240°の間では類似度は0.32~0.41の間、270°~330°の間では0.8~1.2の間の値をとった。

テンプレートデータと歩行者を想定した3種類の姿勢データ(ポーズ2, 3, 4)、計42パターンそれぞれとの類似度の算出結果を図??に示す。ポーズ2は0°~330°で0.24~0.56の値であり、ポーズ3は0°~330°で0.35~0.62の値であり、ポーズ4は0°~330°で0.23~0.68の値をとった。

## 5. 考察

二輪車ライダーの姿勢データとの類似度は、水平方向の回転の回転が0°~120°および270°~330°の場合は類似

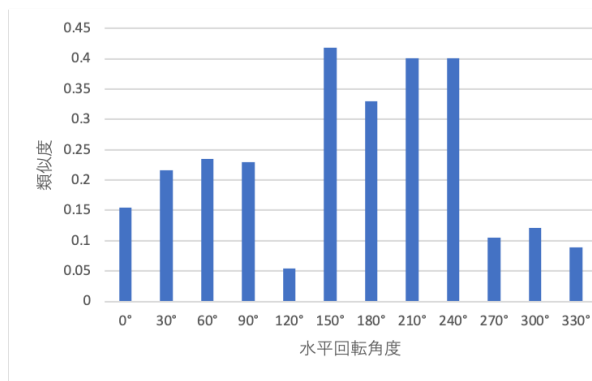


図 10 テンプレートデータと取得された二輪車ライダーの姿勢データとの類似度

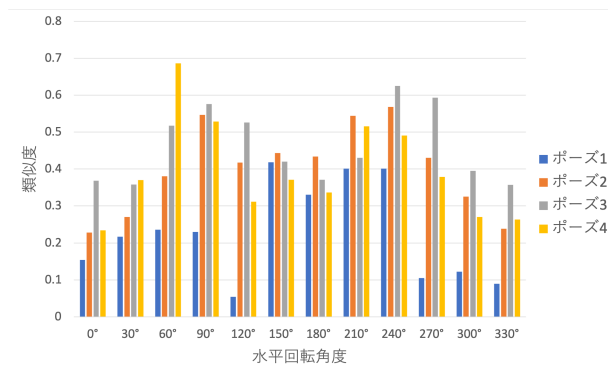


図 11 テンプレートデータと歩行者を想定した姿勢データの類似度

度が0.2前後に収まっており概ねテンプレートデータとの類似度が高いと言え、二輪車ライダーであると推定可能である。しかし、150°～240°の間では類似度の値が大きくなり、類似姿勢であるという推定が難しい。これは、三次元に拡張された人体姿勢データを確認すると、この角度の間では推定することができなかつたため、このような結果が出たと考えられる。また歩行者を想定した3パターンのポーズでは、全ての角度で0.2以上の値を取っており、閾値を0.2で設定することで二輪車ライダーを検出することは可能であると考えられる。今回の実験では歩行者の姿勢データは正しく生成されたものの、二輪車ライダーと同じく姿勢データが誤推定された場合に、類似度の値が変動し、二輪車ライダーと誤判定する可能性があるため、多くの種類の人体姿勢データを用いた類似度評価実験が必要となる。

## 6. おわりに

本稿では、四輪車後方の二輪車の挙動予測のために、人体姿勢検出を用いた二輪車姿勢推定を行うシステムのうち、二輪車ライダーの三次元姿勢データを用いた二輪車の姿勢推定手法について提案した。提案システムの基礎評価として、二輪車ライダーの三次元姿勢データと歩行者の三次元姿勢データの間で類似度に生じる差について検証実験を行った。カメラによって取得した姿勢データが正面を向いている状態に近い場合は、二輪車ライダー検出が可能で

あることがわかった。一方、二輪車の車体の影等で特徴点となる関節座標が取得できない場合には、生成される姿勢データに影響があり、類似度の値が大きくなるため正しく判定されない。そのため、取得できない特徴を考慮した姿勢データの生成方法および類似度算出方法を検討する必要がある。また本稿では、静止画像から姿勢データを取得しているが、実際は連続した姿勢データから挙動を推定する必要があるため、動画を用いた姿勢データ生成手法の検討及び、類似度評価実験を行う。二輪車の挙動把握のためには三次元人体姿勢データを用いて、二輪車ライダーの向き・傾きを推定することで挙動を推定する必要があるため、今後は二輪車の向き・傾きを推定する手法を検討する。

## 参考文献

- [1] 神村吏, 木谷友哉, 渡辺尚: スマートフォン搭載の加速度センサジャイロセンサを使用した二輪車の挙動収集, 情報処理学会研究報告, Vol.2012-ITS-48, No.2, pp.1-8 (2012).
- [2] マイクロソフト Windows デベロッパセンター: Kinect for Windows, 入手先 <<https://developer.microsoft.com/ja-jp/windows/kinect>> (参照 2019-04-01).
- [3] Cao, Zhe, et al.: Realtime multi-person 2D pose estimation using part affinity fields, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.7291-7299 (2017).
- [4] 渡辺裕: 映像情報メディア年報 2018 シリーズ (最終回) 「スポーツ情報処理の研究開発動向」, 映像情報メディア学会誌, Vol.72, No.6, pp.905-908 (2018).
- [5] Rhodin, Helge, et al.: Learning monocular 3D human pose estimation from multi-view images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.8437-8446 (2018).
- [6] Kudo, Yasunori, et al.: Unsupervised Adversarial Learning of 3D Human Pose from 2D Joint Locations. arXiv preprint arXiv:1803.08244 (2018).
- [7] 柴田佳幸: コサイン類似度を用いた人の姿勢の時系列データに基づく日常動作の自動分類とその認識, 大学院研究年報, 中央大学大学院理工学研究科篇, 第46号 (2016).
- [8] 西澤博大, 浦正広, 宮田一乗: 姿勢推定を援用した実人物モデルの描画学習支援システム, 映像情報メディア学会技術報告, Vol.42, No.12, pp.87-90 (2018).
- [9] 国土交通省: 道路運送車両法施行規則, 道路運送車両法・第二条別表第一
- [10] 佐川健太郎: 今日から使えるライテック実践講座【初級編】「車体を安定させるためのライディングフォームとは?」, BikeBros. マガジンズ (オンライン), 入手先<<http://www.bikebros.co.jp/ridetech/index.php?e=1>> (参照 2019-04-01).
- [11] Andriluka, Mykhaylo, et al.: 2D human pose estimation: New benchmark and state of the art analysis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.3686-3693 (2014).
- [12] Ionescu, Catalin, et al.: Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.36, No.7, pp.1325-1339 (2014).
- [13] Ross, Amy.: Procrustes analysis, Course report, Department of Computer Science and Engineering, University of South Carolina (2004).