4P-03

道路シーン中の歩行者の置かれている状況の認識
～歩行者の体向き推定の基礎検討～

熊本　浩二†　　山田　啓一‡
名城大学†　　名城大学‡

1. はじめに

車の自動運転や運転支援システムの更なる高度化のためには，道路シーン中の個々の対象の認識・検出に加えて，その対象の置かれている状況の認識が必要だと考えられる．歩行者の状況を認識することは，歩行者の次の行動を予測する鍵になると考えられる．そこで本研究は，車載カメラ画像から道路シーン中の歩行者の置かれている状況を認識することを目的としている．

本論文はこのための基礎検討として，車載カメラ画像からの歩行者の体向き推定に焦点を当てる．本稿では，単眼カメラ画像から，Convolutional Neural Network (CNN)を用いて歩行者の体向きを推定する手法について述べる。なお，歩行者は別手法で検出されていることを前提とする。

2. 関連研究

単眼画像からの歩行者検出は数多くの研究がある。さらに，画像から歩行者の体の向きを推定する研究も取り組まれている．Tao ら[1]は，単眼画像から DCT-HOG 特徴と Part-based random decision forest を用いて歩行者の体向きを推定している。Flohr ら[2]は，ステレオビデオ画像から歩行者の体と顔の位置と向きを同時に推定する確率的手法を提案している。歩行者の体向き推定は，歩行者の体型，服装，所持品，向き，照明環境などの違いによる見えの変化が大きいことが課題となる。また，車載カメラで撮影した実環境下の歩行者は必ずしも解像度が高くないことも課題となる。

本論文では，歩行者の体全体と顔領域のそれぞれから CNN を用いて推定した体向きを確率モデルで融合することにより体向きを推定する方法を提案し，実環境下の歩行者を対象とした評価結果を示す。

3. 提案手法

提案手法の概要を図 1 に示す．車載カメラ画像から歩行者を検出し，この歩行者の矩形領域の画像 I を入力としてその歩行者の体向きを推定する．提案手法は，画像 I 全体を入力として体向き(d1)を推定する CNN1，画像 I 中の顔領域を入力として体向き(d2)を推定する CNN2，および d1 と d2 から最終的に体向き(d)を判定するための確率モデル P(d|d1,d2)から構成される．歩行者の体向きは図 3 に示す 8 方向に分類することにし，それに対応して d1，d2,および d はそれぞれ 8 クラスとする。CNN1 と CNN2 には,VGG16[2]の出力層の素子数を 1000 から 8 に変更したものを用いる．CNN1，CNN2，および確率モデル P(d|d1,d2)は，歩行者の体向きのサンプルから学習する．

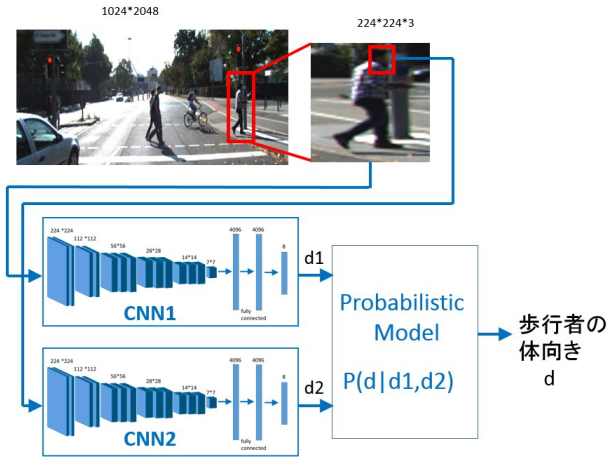

図 1:提案手法の全体構成

4. 実験

4.1 実験方法

学習には PDC dataset [3]を，評価には Cityscapes dataset [4]の中の高さ 100 pix 以上の歩行者の一部を用いた．評価用データの体向きは，(S)こちら向き，(N)むこう向き，(E)右向き，(W)左向きの 4 分類とした．評価用データの画像 1 枚 1 枚に対して，体向きが上記 4 方向のいずれであるかを手作業で分類し，それを ground-truth とした．なお，学習用画像についてはデータセットのアノテーションを ground-truth とした．歩行者はデータセットのアノテーションを利用して抽出し，224×224 pix にリサイズして CNN1 への入力とした．また，この歩行者の顔部分にあたる固定位置の領域を CNN2 への入力とした．

CNN1 と CNN2 は，ImageNet で訓練された VGG16 のパラメータを初期値として訓練した．PDC dataset の 90%を訓練に，残りの 10%を validation に用いた．確率モデルの作成には学習済みの CNN1 と CNN2 の上記 v

Understanding the situation of pedestrian in a road scene
～A basic study on direction estimation of pedestrian～
†Koji Kumamoto・Meijo University
‡Keiichi Yamada・Meijo University

alidation 用データに対する出力を用いた．体全体から推定した体向き d1 と顔部分から推定した体向き d2 の結果の組み合わせは，それぞれ体向きが 8 クラスに分類されるため，64 通り存在する．それぞれの組み合わせに対して，最も予測率の高いクラスをその組み合わせの分類先クラスとすることで確率モデルを作成した．

体向き推定結果の正誤の判定は，評価用データが 4 方向で定義されているのに対し，学習用は 8 方向に定義されているため(図 3)，SW，S，SE のいずれかに分類された場合は S というような対応付けをした．なお，歩行者が複数写っているものや，歩行者の体が完全に隠れてしまっているような画像は評価対象から除外した．比較手法として，体全体から CNN1 のみを用いて体向きを推定するものを用いた．

(S)　(N)　(E)　(W)

図 2:歩行者の例([4]より)

図 3:方向定義

4.2 実験結果

提案手法と比較手法の評価結果(confusion matrix)を表 2 および表 3 にそれぞれ示す．また，確率モデル P(d|d1,d2)による判定規則を表 1 に示す．

提案手法の平均正答率は 89%となった．(E)右向き及び(W)左向きクラスの正答率が最も高い結果となり，誤って判定されたケースが 1 つずつしかなかった．最も正答率が低いクラスは(S)こちら向きであり，75%となった．

比較手法（体全体だけから推定した場合）の平均正答率は 84%となり，最も正答率が低かった(S)こちら向きの結果は 65%，次に正答率が低かった(N)むこう向きの結果は 83%となった．(E)右向きと，(W)左向きの結果は，顔部分の情報を用いた場合と用いない場合での違いは出なかった．

このことから，体全体と顔部分のそれぞれからの推定結果を組み合わせた方が，体全体のみから推定した場合より良い成果が得られることがわかった．特に歩行時の体向きにおいては，顔向きが体の向きに直接関係する可能性が高いことから，顔部分のみの情報を加える本手法の方が高い正答率を得られたと考えられる．図 4 に，体全体からの推定(d1)は誤ったが，体全体からの推定(d1)と顔部分からの推定(d2)を組み合わせる(d)と正しく推定された画像の例を示す．また，両手法とも誤って判定された画像の例を図 5 に示す．

表 1:確率モデルからの分類クラス

argmax d	P(d|d1,d2)	体全体からの分類結果 d1 N	NE	E	SE	S	SW	W	NW
	N	N	NE	NE	SE	S	NW	NW	NW
	NE	N	NE	NE	E	該当なし	該当なし	該当なし	NW
顔部分	E	N	NE	E	SE	S	SW	W	N
からの	SE	N	SE	SE	SE	S	NW	NW	NW
分類結果	S	N	NE	E	SE	S	SW	SW	該当なし
d2	SW	SW	NE	NE	SE	S	SW	W	SW
	W	NE	NE	E	SE	S	SW	W	NW
	NW	NW	NW	NE	該当なし	該当なし	SW	NW	NW

表 2:結果(提案手法)

C.matrix		Predicted Class N	NE	E	SE	S	SW	W	NW	正答率(%)
Actual Class	N	21	6	1	1	1	1	0	8	90
	E	0	12	10	4	1	0	0	0	96
	S	5	4	1	12	12	12	0	2	75
	W	0	0	0	0	1	3	7	4	93
										89

表 3:結果(比較手法)

C.matrix		Predicted Class N	NE	E	SE	S	SW	W	NW	正答率(%)
Actual Class	N	22	6	1	1	1	2	2	6	83
	E	0	10	13	3	1	0	0	0	96
	S	7	4	2	10	12	9	3	1	65
	W	0	0	0	0	1	4	8	2	93
										84

d1=NE
d2=SE
d=SE
GT=S

d1=N
d2=S
d=N
GT=S

図 4:正しく判定された例　図 5:誤り例

5. まとめ

車載カメラによる単眼画像から，道路シーン中の歩行者の体向きを推定する手法を提案し評価結果を示した．今後は向き推定の正答率改善に加え，歩行者の主行動などの推定を行う予定である．

参考文献

[1] J.Tao *et al.*:Part-based RDF for Direction Classification of Pedestrians and a Benchmark, ACCV2014 Workshops.
[2] F.Flohr, *et al.*:A Probabilistic Framework for Joint Pedestrian Head and Body Orientation Estimation, IEEE Trans. ITS, 2015.
[3] K.Simonyan *et al.*:Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv: 1409.1556, 2014.
[4] M.Cordts *et al.*:The Cityscapes Dataset for Semantic Urban Scene Understanding, CVPR2016.