

車載カメラを用いた遠方歩行者からのアイコンタクトの検出

畑 隆聖^{1,a)} 出口 大輔¹ 平山 高嗣² 川西 康友^{3,1} 村瀬 洋¹

概要

車両の運転において、アイコンタクトは歩行者の車両への気付きを判断する重要な要素である。従来のアイコンタクト検出手法の多くは眼球計測に基づく直接的な視線推定に依存しており、道路環境のように車両と歩行者の距離が離れるような場合は視線推定が困難である。一方で我々人間は、視線が不明確な歩行者に対しても、顔向きや姿勢等、歩行者の全身の情報から、こちらを見ているか判断できる。本研究ではこの知見に倣い、顔と体の情報を複合的に考慮することで、視線推定が困難な遠方歩行者からでもアイコンタクトの有無を検出可能な手法を提案する。車載カメラ画像を用いた実験により、その有効性を検討した。

1. はじめに

車両の運転において、歩行者がこちらの存在に気づいているかは危険予測などの面で重要である。そして、その気づきの判断をするために、歩行者がこちらを見ているか、すなわちアイコンタクトの有無が重要な判断要素となる。

アイコンタクト検出や、関連タスクである視線推定に関する従来の研究の多くは、眼球計測や目の周りの外観に基づく直接的な視線推定手法に依存しているため、高画質な顔画像が必要である [1]。高画質な顔写真を撮影するためには、カメラと対象人物が至近距離である必要があり、道路環境のように車両と歩行者との距離が離れている場合、アイコンタクト検出は困難である。故に、そのような状況でのアイコンタクト検出をするためには、直接的な視線推定以外の手法が必要である。

一方で我々が普段車両を運転する様子を振り返ると、視線推定が困難な歩行者に対しても顔向きや姿勢等の情報を複合的に用いてアイコンタクトの有無を判断している。この知見に倣うことで、視線推定が困難な遠方歩行者に対しても、アイコンタクト検出が可能になるのではないかと考える。

遠方で視線推定が困難な歩行者からのアイコンタクトを



(i) アイコンタクト無し (ii) アイコンタクト有り

図 1 同様の顔向きの歩行者画像例

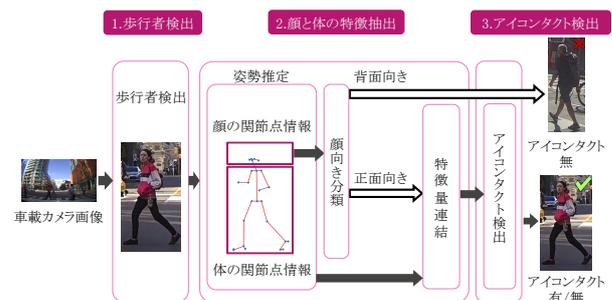


図 2 提案手法の処理手順

判断する際、顔向きなどの「顔の情報」は重要な判断材料である。しかし、図 1 のように顔向きが似たような場合でもアイコンタクトの有無は異なる可能性があるため、顔の情報のみを使ってその違いを判断することは困難である。ここで、体の姿勢を考慮すると、図 1 左の人物は顔と体が同方向であるのに対して、右の人物は足から顔にかけて徐々にカメラ方向に姿勢が向く、すなわち「ねじれ」の状態であることがわかる。このことから、この歩行者はこちらを見ていると予想できる。このように、顔の情報に、姿勢などの「体の情報」を加えることによってアイコンタクトの有無の判断が可能となると考える。よって本研究では、顔の情報と体の情報を複合的に用いることで、遠方歩行者からのアイコンタクト検出の精度向上を図る。

2. 提案手法

2.1 処理の概要

図 2 に提案手法の処理手順を示す。

- (1) 車載カメラで撮影された画像（以下、車載カメラ画像）に対して歩行者検出を行ない、歩行者領域画像（以下、歩行者画像）を取得する。

¹ 名古屋大学

² 人間環境大学

³ 理化学研究所 情報統合本部 GRP

^{a)} hatar@vislab.is.i.nagoya-u.ac.jp

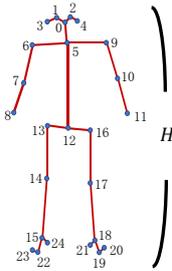


図 3 本研究で使用する関節点

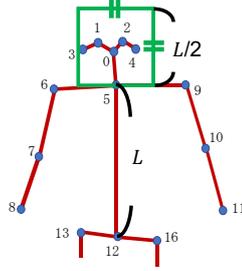


図 4 顔の関節点の正規化

(2) 歩行者画像に対して姿勢推定を行ない、図 3 のような関節点情報 (各関節点の二次元座標とその信頼度) を取得する。関節点情報について、首より上を「顔の関節点情報」、首から下を「体の関節点情報」とし、それぞれ独立に正規化を施す。正規化後の各特徴ベクトルを「顔の情報」、「体の情報」とする。そして、顔の情報について、関節点の位置関係が顔向きを表しているという考えの下、顔の情報を顔向き分類器に入力して、正面向きと背面向きに 2 クラス分類する。ここで、背面向きと判断された画像については、この時点でアイコンタクト無しとして早期棄却を行う。それ以外の画像は、顔の情報と体の情報を結合し、アイコンタクト検出器に入力するための特徴ベクトルを取得する。

(3) 上記で得られた特徴ベクトルをアイコンタクト検出器に入力し、最終的な出力結果を得る。
以下、手法の詳細について述べる。

2.2 歩行者検出

車載カメラ画像に対して、歩行者検出器によって歩行者領域を検出する。この時、得られる歩行者を囲む矩形が $H \times W$ 画素の大きさの場合、歩行者全体を囲えるよう矩形を上下左右に 5% ずつ拡張した $1.1H \times 1.1W$ 画素の大きさの矩形で歩行者画像の切り出しを行なう。

2.3 顔の情報と体の情報の抽出

姿勢推定器を用いて、歩行者画像から図 3 に示すような 25 個の関節点について、「関節点情報」を取得する。ある関節点 i の関節点情報は、歩行者画像における左上を原点とする、サブピクセル単位の x 座標と y 座標、及びその推定の信頼度 c を要素とする 3 次元特徴ベクトル $(\mathbf{x}_i; \mathbf{y}_i; \mathbf{e}_i)$ である。ここで、 \mathbf{e}_i は $[0;1]$ の範囲の実数値をとり、 $\mathbf{e}_i = 0$ の場合は関節点を検出できなかったものとみなし、 \mathbf{x}_i と \mathbf{y}_i の値はどちらも 0 とする (関節点の欠損)。そして、全ての関節点の特徴ベクトルを結合し、 $\mathbf{p} = (\mathbf{x}_0; \mathbf{y}_0; \mathbf{e}_0; \dots; \mathbf{x}_{24}; \mathbf{y}_{24}; \mathbf{e}_{24})$ の 75 次元の特徴ベクトルとする。 \mathbf{p} のうち、鼻、両目、両耳 (図 3 の 0, 1, 2, 3, 4) を顔の関節点 \mathbf{q}_{face} として、それ以外を体の関節点 \mathbf{p}_{body} として扱う。

ここで、各歩行者画像の大きさごとに関節点の位置関係のスケールが異なるため、表 1 のように顔と体で別々の手

表 1 提案手法における顔と体の正規化手法の違い

	顔の関節点情報	体の関節点情報
関節点位置の正規化の基準	座高 L	歩行者画像の縦サイズ H
関節点位置の正規化の中心	$(\tilde{x}_5, \tilde{y}_5)$	$(\tilde{x}_{12}, \tilde{y}_{12})$
信頼度	連続値	2 値

法で、各値が $[-1,1]$ の範囲に収まるように正規化を行なう。

まず、体の関節点 i に関する正規化の式を以下に示す。

$$\mathbf{w}_i = \begin{cases} < \frac{\tilde{x}_i - \tilde{x}_{12}}{H/2} = \frac{2\tilde{x}_i - 2\tilde{x}_{12}}{H} & (\mathbf{e}_i > 0) \\ 0 & (\mathbf{e}_i = 0) \end{cases} \quad (1)$$

$$\mathbf{y}_i = \begin{cases} < \frac{\tilde{y}_i - \tilde{y}_{12}}{H/2} = \frac{2\tilde{y}_i - 2\tilde{y}_{12}}{H} & (\mathbf{e}_i > 0) \\ 0 & (\mathbf{e}_i = 0) \end{cases} \quad (2)$$

$$\mathbf{b}_i = \begin{cases} < 1 & (\mathbf{e}_i > 0) \\ 1 & (\mathbf{e}_i = 0) \end{cases} \quad (3)$$

これにより、正規化された体の情報 \mathbf{p}_{body} を得る。

また、顔の関節点 i に関する正規化の式を以下に示す。

$$\mathbf{w}_i = \begin{cases} < \frac{\tilde{x}_i - \tilde{x}_5}{L/4} = \frac{4\tilde{x}_i - 4\tilde{x}_5}{L} & (\mathbf{e}_i > 0) \\ 0 & (\mathbf{e}_i = 0) \end{cases} \quad (4)$$

$$\mathbf{y}_i = \begin{cases} < \frac{\tilde{y}_i - (\tilde{y}_5 - L/4)}{L/4} = \frac{4\tilde{y}_i - (4\tilde{y}_5 - L)}{L} & (\mathbf{e}_i > 0) \\ 0 & (\mathbf{e}_i = 0) \end{cases} \quad (5)$$

$$\mathbf{b}_i = \frac{\mathbf{e}_i - 0.5}{0.5} = 2\mathbf{e}_i - 1 \quad (6)$$

これにより、図 4 における緑枠の正方形の範囲が $[-1;1]$ になるよう正規化された、顔の情報 \mathbf{q}_{face} を得る。

信頼度 \mathbf{e}_i を別々に正規化することに関して、体の関節点では、各関節点の位置関係、すなわち体の姿勢の形からアイコンタクトの有無を判断するために、 $\mathbf{e}_i > 0$ の関節点は全て信用するものとして、これを 1 とする。一方で、顔の関節点では連続値にしているのは、例えば「左目の信頼度は、顔が正面から左向きになるにつれて検出が難しくなるため低下する」など、顔の各関節点の信頼度と顔向きに相関があると考えられるためである。

以上により得られた顔の情報 \mathbf{q}_{face} と体の情報を \mathbf{p}_{body} 用いて、アイコンタクト検出を行う。

2.4 顔向き分類

2.3 節で取得した顔の情報 \mathbf{q}_{face} を顔向き分類器に入力し、正面向き/背面向きの 2 クラスに分類する。この時点で背面向きに分類された歩行者画像については、アイコンタクト無しを出力し早期棄却する。それ以外の場合は 2.5 節、及び 2.6 節でアイコンタクト検出を行なう。

2.5 特徴量の連結

2.3 節で取得した顔の情報 \mathbf{q}_{face} と体の情報を \mathbf{p}_{body} を

表 2 実験データの内訳

データの種類	データ数
Train	23,000
Validation	2,500
Test	15,260

連結して、75次元の特徴ベクトル \mathbf{v} を取得する。

2.6 アイコンタクト検出

2.5節で取得した \mathbf{v} を4層の全結合型 Neural Network で構成したアイコンタクト検出器に入力し、アイコンタクト有り/無しの2クラスに分類し、その結果を最終的な結果として出力する。活性化関数には LeakyReLU を用い、出力層では softmax 関数を用いて正規化する。損失関数はクロスエントロピー誤差を用い、最適化関数は AdamW [2] を用いる。

3. 実験

3.1 実装

本実験では、提案手法を以下のように実装した。歩行者検出については、実際の歩行者検出処理の代わりに、後述する PIE データセットの矩形情報を歩行者検出結果とした。姿勢推定器には、CNN を用いた手法である OpenPose [3] の BODY_25 モデルを使用した。顔向き分類器の実装は、アイコンタクト検出器と同様であり、入力層のみ異なる。

3.2 データセット

本実験では、車載カメラ画像として、Pedestrian Intention Estimation dataset(以後、PIE データセット)[4]を用いた。このデータセットは、車内に設置された157の広角レンズを装備した車載カメラによって撮影された、約900,000枚の画像系列によるデータセットである。PIE データセットには複数の歩行者が存在し、そのうち1,842人の歩行者には、歩行者IDや、カメラを見ているか否か、外接矩形、遮蔽具合などの情報がアノテーションされている。

本実験では問題設定の簡単化のため、歩行者画像のうち以下の条件を満たす画像のみを使用した。

- (1) 歩行者の遮蔽率が25%以下
- (2) 切り出した歩行者画像の縦サイズが150画素以上
- (3) 切り出した歩行者画像の矩形が、元の車載カメラ画像の枠(1;920 1;080)を超えない

さらに、アイコンタクト有り無しデータの数が同数となるように抽出した。以上によって得られた歩行者画像を実験で用いた。データの内訳を表2に示す。

また、顔向き推定器については、表2から1,125枚を無作為抽出し、人手で顔向きに関するアノテーション(正面、背面)を付与し、それを用いて学習をした。

表 3 実験結果

手法	使用する情報	正解率
比較手法	視線情報のみ	50.0
提案手法 1	顔の情報のみ	84.1
提案手法 2	顔の情報+体の情報(同一の正規化)	82.2
提案手法 3	顔の情報+体の情報(異なる正規化)	84.8

表 4 顔向きとアイコンタクトの有無の関係

顔向きクラス	データ数	
	アイコンタクト有	アイコンタクト無
正面向き	155	21
斜め向き	226	217
横向き	0	281
背面向き	0	204

3.3 実験方法

以下の4つの手法で比較実験を行なった。

- (1) **提案手法 1** 顔の情報のみを用いてアイコンタクト検出を行う。
- (2) **提案手法 2** 顔の情報と体の情報で同じ正規化手法(体の関節点の正規化手法)を施してアイコンタクト検出を行う。
- (3) **提案手法 3** 顔の情報と体の情報で異なる正規化手法を施してアイコンタクト検出を行う。
- (4) **比較手法** OpenFace 2.0 [1] による視線推定結果を用いて、アイコンタクト検出を行う。

比較手法におけるアイコンタクト検出器は、提案手法と同様に4層の全結合型 Neural Network を用いた。なお、視線推定結果が得られなかった画像については、アイコンタクト無しに分類した。実験では各試行を5回行ない、正解率の平均値を比較した。

3.4 実験結果

比較手法及び提案手法によるアイコンタクト検出の正解率を表3に示す。比較手法について、本実験で使用した歩行者画像は総じて低解像度であり、視線推定が可能な画像が非常に少なく、ほぼ全ての画像がアイコンタクト無しに分類された。また提案手法3は、正解率が最も高く、提案手法1と比較して0.7%、提案手法2と比較して2.6%向上したことを確認した。

4. 考察

顔の情報と体の情報を組み合わせることの有用性、及び、顔と体で別々に正規化をすることの有用性について考察する。

4.1 顔と体の情報を考慮することについての考察

図5左は、顔と体の情報を用いた場合にアイコンタクト検出が成功した例である。顔の情報のみを用いた場合、こ

