一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
IMQ2024-17,IE2024-95,MVE2024-34(2025-03)

[ショートペーパー] 複数歩行者間のインタラクション理解に向けた気づき認識データセットの構築

辰己　弘征†　出口　大輔†　村瀬　洋†　伊藤　誠悟†

† 名古屋大学 大学院情報学研究科

あらまし　歩行者間のインタラクション理解において気づきは重要な要素であり，インタラクション理解は行動予測にとっても重要な特徴である．一般に，撮影後に歩行者が真に見た対象をアノテーションするのは困難である．そのため，データセット構築にあたり，注視対象が明確な場面を設計し，複数の歩行者に指示通りの動きと注視を演じてもらうことにより，注視対象を推定するタスクのためのデータセットを構築した．本報告では，画像中に存在する複数の歩行者がそれぞれ見ている対象をアノテーションしたデータセットの構築について述べる．

キーワード　学習データセット、機械学習

Hiroyuki TATSUMI†, Daisuke DEGUCHI†, Hiroshi MURASE†, and Seigo ITO†

† Graduate School of Informatics,Nagoya University, Furo-cho, Chikusa-ku, Nagoya,　Aichi

1. はじめに

ショッピングモールや駅構内など，複数の歩行者が存在する状況では，各歩行者は接触事故を防ぐために付近の他者に注意を向けながら移動する．このような環境でロボットが自律移動する場合，歩行者の気づきや注意から歩行者間のインタラクションを理解すれば，歩行者の行動予測や同じグループの推定などによって円滑な移動を実現できると考えられる．気づきや注意は，歩行者の注視対象から読み取ることができるが，歩行者の注視対象は他者の行動や視線から大きな影響を受ける．図 1 に示す歩行者 C に着目すると，歩行者 B が歩行者 A に気を取られていることから歩行者 B に注意を払っていることがわかる．このような明示的，暗黙的な歩行者間のインタラクションの理解を目的として，注視対象を推定するためのデータセットを構築した．

2. 関連研究

交通シーンにおける歩行者を含んだデータセットは様々提案されている．Holger ら [1] は交通シーン画像に Bounding Box と物体の種類を付与した nuScenes を公開している．また Sun ら [2] も同様に交通シーン画像に Bounding Box と物体の種類を付与した Waymo Open Dataset（waymo）を公開している．しかしこれらに歩行者の注視対象は含まれない．そこで Murakami ら [3] は，歩行者の行動予測を目的として，これらの一般公開

図 1　データセットに含まれる画像例

されたデータセットに注視対象をアノテーションした「歩行者の注視対象データセット」を構築した．しかし，第三者目線で注視対象を推測したものを歩行者の注視対象としているため，アノテーションが真の注視対象と一致している保証はない．

3. 撮影条件

データセット構築にあたり，複数の歩行者が自由に移動できるような空間を想定してデータ撮影を実施した．具体的には，4 人の被験者に歩行者役を演じてもらい，あらかじめ用意した 20 種類のシーン（注視対象と歩く方向を指定）に沿って撮影を行った．これらのシーンを表 1 に示す．シーン中の注視対象は常に同じであるという条件で撮影を行うことにより，映像中のすべてのフレームに対して同じアノテーションを付与することで，コスト削減を行なった．撮影環境は表 2 の通りである．各シーンにおける歩行者は，他の歩行者を注視する歩行者，スマートフォン（スマホ）を注視する歩行者，注視対象がない歩

表 1　撮影シーン一覧

シーン名	注視関係（例）	シーン数
すれ違い	反対方向に進む歩行者同士がすれ違うまで互いを注視	5
追い越し	追い越される静止歩行者を追い越す歩行者が注視	2
追い越し（スマホ）	スマホを見る静止歩行者を追い越す歩行者が注視	2
追従	追従する歩行者が追従される歩行者を注視	1
追従（スマホ）	追従される歩行者がスマホを注視、追従する歩行者は上に同じ	1
注視なし	全歩行者が注視対象を持たない	1
全員がスマホを注視	全歩行者がスマホを注視	1
交差	後に通る歩行者が、自分の前を先に横切る歩行者を交差完了まで注視	2
交差（スマホ）	先に通る歩行者はスマホを注視し、後に通る歩行者は先に通った歩行者を注視	2
自由移動	静止する歩行者と自由に移動する歩行者がお互いを注視	2
自由移動（スマホ）	スマホを注視して静止する歩行者を他の歩行者が静止しながら注視	1

表 2　撮 影 環 境

撮影機器	handycam(sony)，$a7$(sony)
画素数	1280 × 720（handycam），3840 × 2160（$a7$）
フレームレート	30fps
撮影場所	名古屋大学 IB 電子情報館中棟

行者，の三つに分類される．表 1 に示すシーン数は，同じ基本動作に対して避ける方向等を変化させて撮影したものの総数である．すれ違いのシーンには，4 人全員が独立して歩くパターンに加えて，図 1 のように 2 人が連れ立って横並びに歩くパターンを含んでいる．追従や追い越しのシーンでは，対象となる歩行者との距離を 1m 以上保ち，回避が必要な場合は 1.5m〜2.0m の距離で回避動作を開始するよう設定している．また自由移動のシーンでは注視対象のみを指示し，移動は歩行者に任せている．なお，動画全体としては現実に即さない動作も含まれるが，切り出される各フレームは現実のどこかの場面で存在しうるものとして撮影を行った．

4.　データセット構築

今回は，静止画に対して注視対象を推定するタスクに利用するデータセットの構築のため，シーンを全て 30Hz で分割をした．分割した画像のうち，注視対象が画面内に存在する歩行者もしくは，注視対象を持たない歩行者が存在する画像のみを抽出し，それらの画像に対してアノテーションを行った．すべての歩行者に対し，bbox の座標と注視対象の座標のアノテーションを行った．注視対象がスマホの場合は，図 2 のようにスマホ自身が明確に写らない場合が多いため，スマホの座標ではなく，スマホを見ているラベルを歩行者自身に付与した．歩行者の注視対象のアノテーションは歩行者が以下の条件を満たす場合のみ行った．

（1）　頭部が全て写っている歩行者

（2）　注視対象の一部が画面内に写っている，もしくは明示的に注視対象を持たない歩行者

上記を条件として，アノテーションした結果，32,863 枚の画像がアノテーション対象となった．他者を注視する歩行者を表す bbox が 36,568 個，注視物を持たない歩行者を表す bbox が

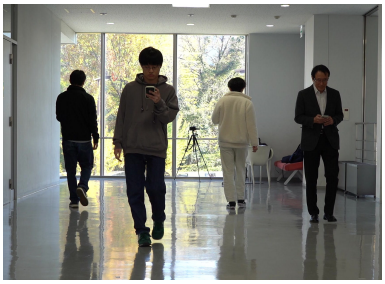

図 2　全員がスマホを注視しているシーン

28,317 個，スマホを注視するラベルをつけた歩行者の bbox が 18,704 個となった．

5.　む　す　び

本報告では，複数歩行者間のインタラクション理解を目的として，注視対象推定に必要となるデータセットの構築について述べた．歩行者の注視対象推定を目的とした既存のデータセットの中で，注視対象の真値がわかるものが存在しないため，注視対象を指定した 20 種類のシーンを作成し，フレームに分割したものにアノテーション作業を行った．今後の課題として，データセットに含まれる物体の種類が少ないことが挙げられる．今回は，他の歩行者，もしくはスマートフォンを注視対象としたが，実際にはより多くの注視対象となる物体が存在するので，それらを含めたデータセット構築を検討する．

謝辞　本研究の一部は JSPS 科研費 23H03474 による

文　　献

[1] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In CVPR, pp. 11618–11628 (2020).

[2] Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al. Scalability in perception for autonomous driving: Waymo open dataset. In CVPR, pp. 2443–2451 (2020).

[3] 村上大斗，出口大輔，平山高嗣，川西康友，村瀬洋，歩行者の注視対象データセットの構築，情報処理学会研究報告，Vol.2023-CVIM-233 No.57，(2023)