

# 周辺歩行者の行動予測を用いた他者との衝突回避のための 姿勢推定処理速度向上に関する検討

飛田真武<sup>†</sup> 中村嘉隆<sup>†</sup> 稲村浩<sup>†</sup>

公立はこだて未来大学 システム情報科学部<sup>†</sup>

## 1. はじめに

日常生活において、都市の歩道や駅構内など様々な方向に歩行者が入り乱れるような混雑環境では歩行者同士が接触・衝突する危険がある。歩行者は常に周辺を観察することで衝突を回避する必要があるが、周辺観察を行いながらの歩行は、歩行者に認知負荷がかかる。これに対し、周辺歩行者の行動予測を用いて衝突回避の支援を行う手法が提案されている。ここでの行動予測とはある人物が数秒後までにとる行動を予測する技術であり、歩行者が進む経路を予測する経路予測もこの一種である。行動予測の主な流れとしては、カメラ撮影映像中の人物の姿勢を推定し、姿勢情報を用いて予測を行う。しかし、既存の姿勢推定手法では混雑環境において、歩行者同士の重なり等の問題から十分な精度を達成できていない。また、リアルタイムな推定に必要な処理速度も得られていない。

本稿では、行動予測のプロセスの一つである姿勢推定において、推定精度を維持させたまま処理速度を向上させることを目的とする。これにより、行動予測の処理速度の向上も可能となる。

## 2. 関連研究

現在確立されている人物姿勢推定の手法は、ボトムアップ方式とトップダウン方式の二種類に大別できる。ボトムアップ方式は始めに映像中にある人物の関節点を全て推定し、映像中の各人物の全身の姿勢を構成する手法であり、OpenPose[1]が代表的である。OpenPoseでは複数人を同時に推定する場合でも計算時間が推定人数に比例しないというメリットがある。しかし、時系列における同一人物のタグ付けを行っていないため、混雑環境では同一人物の追跡に失敗する場合がある。

一方、トップダウン方式は、始めに人物を検知し人物ごとに関節点の推定を行うため、人物の誤認識が少なくボトムアップ方式に比べて人物追跡に優れていると言える。代表的なものにAlphaPose[2]があり、人物ごとにタスクを分割でき、ボトムアップ方式に比べて少ない計算量での推定が可能である。しかし、計算時間が推定人数に比例して長くなる。更にこれらの姿勢推定手法では映像中の人物全ての推定を行うため、混雑環境では推定速度が低下する。

## 3. 提案手法

複数人が映る映像から姿勢推定が必要な人物を抽出することを研究課題とする。観測者の頭部にカメラを装着し周辺歩行者の姿勢推定を行う。姿勢推定手法としてAlphaPoseを用いることを想定し、カメラによる撮影から姿勢推定までは図1のような処理が必要となる。行動予測の必要があるのは観測者に衝突する可能性のある周辺歩行者のみであるため、該当する周辺歩行者の抽出を行うことで推定人数を削減する。本稿では、観測者への衝突の可能性を表す指標として観測者と周辺歩行者の相対距離に着目し、観測者に接近する周辺歩行者を抽出する(図1(a))。接近の判定には映像中の人物の大きさであるスケールの変化量を用い、時間経過とともにある歩行者のスケールが大きくなっていけばその歩行者は観測者に接近していると判定する。

人物検知にはAlphaPoseに組み込まれているYOLOv3[3]を用いる。YOLOv3は人物以外の物体の検知も行う手法であるが、人物を検知した場合のみスケール変化量の計算を行う。検知した人物の映像中における高さや幅が出力されるため、それらの変化量から相対距離の変化を求める。映像は1秒を30フレームに分けて処理する。1フレームでの変化量が小さいため10フレームでの変化量から判定する。リアルタイム処理を想定し、接近の判定はフレーム毎に行う。

A Study on Fast Pose Estimation for Collision Avoidance with Others Using Behavioral Prediction of Pedestrians

<sup>†</sup>Masamu Tobita, Yoshitaka Nakamura, Hiroshi Inamura, School of Systems Information Science, Future University Hakodate

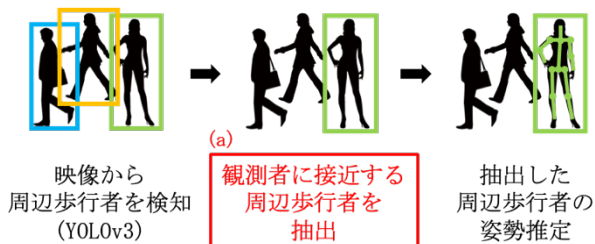


図1 提案手法の全体図

#### 4. 実験

接近人物の抽出精度評価のために、一人の歩行者が様々な方向に移動する映像を撮影した。撮影は2020年12月に公立ほこだて未来大学の室内通路にて行い、カメラはGoPro HERO7を使用し解像度を1080pに設定した。本実験では、カメラを観測者の身長と同じ高さである165cmの定点に設置した。カメラの周辺に8つの地点を設置し、歩行者は決められた2地点間を移動した。撮影はカメラと歩行者の間に障害物のない環境で行った。映像は全15パターン撮影し、それぞれを以下の3種類の移動に分類した。

- ・接近：歩行者がカメラに近づく
- ・停滞：歩行者がカメラと一定の距離を保つ
- ・離脱：歩行者がカメラから遠ざかる

更に撮影した映像を編集し、障害物によって歩行者の体の一部が映っていない映像を15パターン作成した。合計30パターンの映像から歩行者の各移動に対し、カメラへの接近の有無を判定し、判定の精度を評価した。

#### 5. 結果

接近判定結果を障害物の有無によって分けたものを表1に、移動の種類ごとの正解率を表2に記す。接近移動では接近と判定された場合を正解に、その他の移動では接近と判定されなかった場合を正解に分類している。

表1 接近判定結果

| 障害物     | 無    | 有    | 両方   |
|---------|------|------|------|
| 再現率 (%) | 86.9 | 76.6 | 81.7 |
| 適合率 (%) | 62.3 | 59.4 | 60.8 |
| 正解率 (%) | 75.3 | 69.1 | 72.2 |

表2 移動の種類ごとの正解率

| 移動      | 接近   |      | 停滞   |      | 離脱   |      |
|---------|------|------|------|------|------|------|
|         | 無    | 有    | 無    | 有    | 無    | 有    |
| 正解率 (%) | 86.9 | 76.6 | 51.1 | 46.3 | 75.9 | 73.0 |

障害物のない映像における再現率が最も高く9割近い数値であった。障害物なしとありの両方を合わせた正解率が約7割であるのに対し、再現率は8割以上の数値を達成したが、適合率は約6割であった。移動の種類ごとに比較すると、停滞映像の正解率が最も低く、障害物の有無に関係なくおよそ半分の確率で接近と判定された。障害物のある映像では障害物のない映像に比べて全体的に正解率が低くなった。

#### 6. 考察

停滞移動の正解率が最も低かったが、これは歩行者の姿勢変化によってスケールが変化するため誤判定が起きてしまったことが原因だと考えられる。その他の誤判定に関しても歩行者がカメラの正面の向きに移動した映像では正解率が高かったことから、歩行時の腕の振りによって歩行者の幅の大きさが変化したことが原因だと推測できる。障害物の有無によって正解率が変化した。特に接近映像に関しては障害物の影響が大きく、歩行者の全身が映っている状態から徐々に障害物に重なることで、体の一部が隠蔽されスケールが減少したことが要因だと思われる。以上から、歩行者の姿勢の変化によるスケール変化および歩行者の体の一部の隠蔽への対応が今後の課題である。

#### 7. おわりに

本稿では歩行者一人の移動に対してカメラへの接近の判定を行い、障害物の有無によって数値に変化はあるものの、接近移動に対して高い正解率を得られた。今後は姿勢変化と障害物への対応を考え、複数人が映った映像での接近判定評価を行い、手法の有用性を評価する。

#### 参考文献

[1] Zhe Cao, et al.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, arXiv:1812.08008(online) (accessed 2020-10-27).  
 [2] Fang, et al.: RMPE: Regional multi-person pose estimation, arXiv:1612.00137(online) (accessed 2020-06-27).  
 [3] Redmon, et al.: YOLOv3: An incremental improvement, arXiv:1804.02767(online) (accessed 2020-06-27).