

混雑環境歩行時における他者との衝突回避のための 周辺歩行者の接近判定手法の検討

飛田 真武¹ 中村 嘉隆² 稲村 浩³ 石田 繁巳^{3,4}

概要：日常生活において、歩行者は常に周辺歩行者の位置・方向等を把握し、進路を予測することで衝突を回避している。歩行者が様々な方向に入り乱れるような混雑環境においては、多数の周辺歩行者の複雑な行動を把握する必要があり衝突回避の難易度が高いため、これを支援することで歩行者の負担を軽減できる。観測者の周辺に複数の歩行者が存在する環境を想定し、周辺歩行者の行動予測をリアルタイムで行う。衝突可能性のある周辺歩行者が存在する場合に、該当する周辺歩行者の数秒後の移動経路を予測・通知することで、回避行動を促し衝突するリスクを軽減することを目指す。リアルタイム処理の実現が課題であり、処理速度向上のアプローチとして、予測対象を限定することで姿勢推定を用いた行動予測処理の適用回数を削減する。限定する予測対象を選択するために、映像を入力とする人物検出の結果情報のみに基づいて行動予測より軽量の計算で予測対象の位置と進行方向の抽出を行い、観測者に接近する周辺歩行者を判定する。本稿ではこの人物接近判定を、検出した人物の画像上でのサイズ変化から判定する手法を提案する。歩行者1人を撮影した映像に対して接近の判定を行い、障害物の存在する映像においても誤検出を補正することで精度を向上させることができた。

1. はじめに

日常生活において、歩行者は別の歩行者に気づかずに、不用意に接触・衝突してしまうことがある。近年では、歩きスマホやイヤホンの装着等の理由から周辺の状況把握が疎かになり、別の歩行者に衝突するケースが多く見られる。東京消防庁の調査によると、年間7000人以上が対向者や障害物との衝突によって救急搬送されており、2019年には7882人が報告されている(図1)[1]。救急搬送されない場合も含めれば、実際にはより多数の歩行者が衝突を起していると思われる。

歩行者同士の衝突は、両者の様々な位置関係の下で発生する。そのため歩行者は、常に周辺を観察することで周辺歩行者の位置・方向等を把握し、進む経路を予測することで衝突を回避している。しかし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においては、多数の周辺歩行者の複雑な行動を把握する必要がある上に、見通しの悪い事が多いため衝突回避の難易度が高い。このような環境において常に周辺観察を行いながら歩行することは、歩行者に認知負荷がかかる上に前方の不注意に

繋がりがかねない。

歩行者は歩行動作中に周辺観察・経路予測・衝突回避の3つのプロセスを行いながら進路を決定しているが、人間1人の視野で周辺歩行者全員を観察することは難しいため、これらを支援することで歩行者の負担の軽減が可能である。特に正面方向以外など視覚外に存在する周辺歩行者に着目し、該当する周辺歩行者の行動予測を行うことで混雑環境での衝突回避を支援する手法を検討する。行動予測とは対象となる人物が数秒後までにとる行動を予測する技術であり、予測対象の姿勢情報や位置座標を入力として数秒後までの移動経路を得ることができる[2]。衝突回避の際には、周辺歩行者が数秒後に移動する経路に重ならないように自身の経路を決定する必要があるため、予測対象の移動経路が出力される行動予測は適切な手法である。

本研究では、観測者の周辺に複数の歩行者が存在する環境を想定し、周辺歩行者の行動予測をリアルタイムで行うことを目的とする。観測者の頭部にウェアラブルカメラを装着して周辺歩行者を撮影し、映像から周辺歩行者の姿勢を推定し、そこで得られた姿勢情報を用いて予測を行うことを想定する。行動予測によって求めた周辺歩行者の数秒後までの移動経路と位置から、観測者周辺の一定範囲に予測対象が侵入する可能性があるときに衝突可能性を算出する。衝突可能性のある周辺歩行者が存在する場合に、該当

¹ 公立はこだて未来大学大学院 システム情報科学研究科

² 京都橘大学 工学部

³ 公立はこだて未来大学 システム情報科学部

⁴ 九州大学システム LSI 研究センター

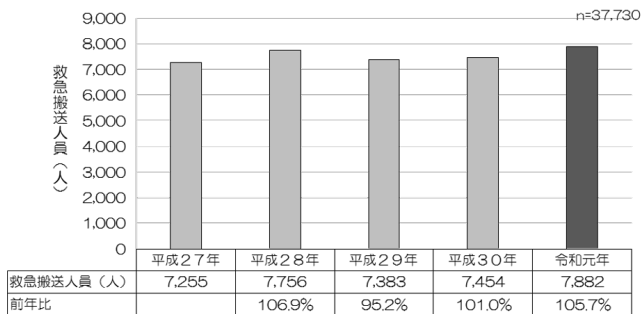


図 1 衝突による年別緊急搬送人数

する周辺歩行者の数秒後の移動経路を予測・通知することで、回避行動を促して衝突するリスクを軽減することを目指す。

これに向けて本稿では、行動予測までの一連の処理のうち姿勢推定についてリアルタイムに推定可能な手法を検討する。この実現に向けてはリアルタイム処理が課題となることから、観測者周辺の一定範囲に侵入する可能性のない周辺歩行者を予め除外して予測対象を限定することで、姿勢推定を用いた行動予測処理の適用回数を削減する。限定する予測対象を選択するために、映像を入力とする人物検出の結果情報のみに基づいて行動予測より軽量の計算で予測対象の位置と進行方向の抽出を行い、観測者に周辺歩行者が接近するかどうかを判定する手法を提案する。

2. 関連研究

2.1 行動予測

行動予測には位置情報や距離等の様々なセンサやシステムを用いた手法が存在するが、映像からは環境情報や予測対象に関する情報を入手することができるため、映像を入力として用いる手法は一般的である。撮影方法によって手法が分かれており、監視カメラや鳥瞰映像といった定点を観測する手法 [3] と、一人称カメラや車載カメラなどの主観映像を用いる手法 [4] がある。定点を観測する手法では決められた場所でしか行動予測を行えないが、障害物等の影響を受けず全体を撮影することができるため、建築物等の環境情報を入手しやすいというメリットがある。一方、主観映像を用いる手法では、予測対象を近い距離で撮影できるため、姿勢や向きといった予測対象に関する情報を比較的簡単に入手できる。歩行者同士の衝突が起こる混雑環境は街中の至る所に存在し、それら全てにカメラを設置することは現実的でない。そのため本研究では、一人称視点映像を用いる手法を採用し周辺歩行者の行動予測を行う。

2.2 姿勢推定

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 [5]。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

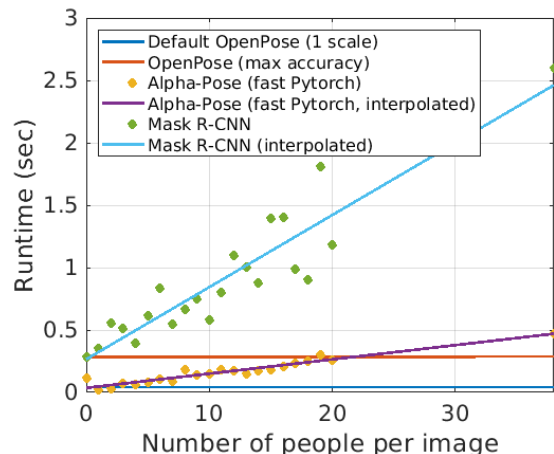


図 2 推定人数と処理時間の推移 ([6] より引用)

定し、関節点を線で結ぶことによって映像中の各人物の全身の姿勢を構成する手法である。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、関節点の対応付けの精度に課題があり、実際の姿勢とは異なる組み合わせの関節点同士が結びついてしまい誤検出が発生する可能性がある。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もある。関節点の推定後に姿勢を構成する段階において、同一人物の関節点同士を適切に結ぶ工程で時間がかかるという問題もある。これらの問題を解決する手法としてZheらはOpenPose[6]を提案している。この手法では関節点の推定とは別に、関節点同士の関係性を2Dベクトルで算出し、姿勢の構成にそれらの情報を用いることで精度を向上させている。さらに処理の構造をシンプルにし、畳み込みのフィルタサイズを小さくすることで処理速度を向上させている。

一方、トップダウン方式は始めに人物を検出し人物ごとに関節点の推定を行うため、異なる人物の関節点同士を結ぶ可能性が低く人物の誤検出が少ない。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。姿勢推定において複数人の同時推定は課題の一つであり、人物同士の重なりや障害物によって身体の部分遮蔽が起こる場合に推定に失敗することがある。Fangらの提案するAlphaPose[7]は、複数人の姿勢推定を想定した手法であり人物追跡が可能である。この手法では、推定した姿勢の類似性を比較し、誤検出の可能性のある推定結果を除去することで精度を向上させている。歩行者同士の重なり等が原因で姿勢推定精度が低下するという背景から、混雑環境での姿勢推定において効果的なアプローチになると考えられる。しかしトップダウン方式の手法では、図2のように計算時間が推定人数に比例して長くなるデメリットがあり、混雑環境における処理速度の低下が想定される。

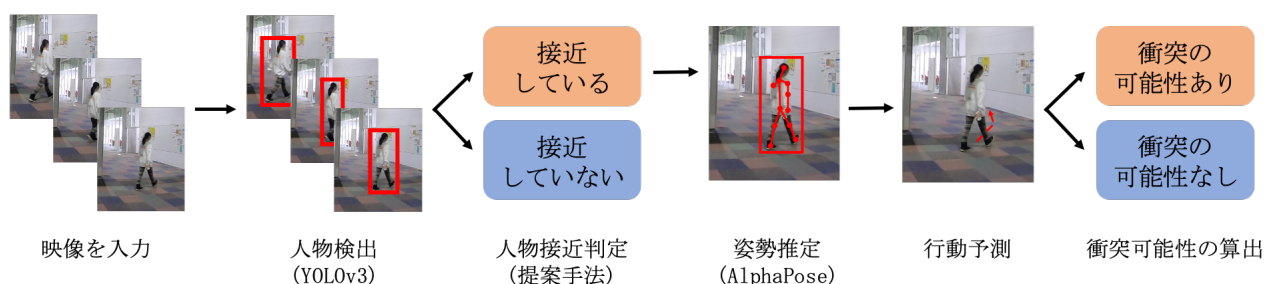


図 3 提案システムの全体図

3. 提案システム

3.1 システム概要

観測者の頭部にカメラを装着して周辺歩行者を撮影し行動予測を行うことを想定する。八木らの手法 [8] を参考に、周辺歩行者の姿勢情報を入力として畳み込みニューラルネットワーク (CNN) を用いて予測を行う。システム全体としては図 3 のような処理が必要となり、人物検出に YOLOv3 [12] を採用し、姿勢推定には複数人の同時推定に優れているトップダウン方式の手法を検討しており、現時点では AlphaPose の採用を想定している。事前学習用のデータセットとして人物が重なり合った混雑シーンが含まれる CrowdPose データセット [9] を用いることで、混雑環境での推定精度向上を図る。処理速度を向上させる要素として計算量が大きく影響することに着目し、CNN の手法の 1 つとして MobileNet [11] を用いることで処理の高速化を検討する。MobileNet では畳み込みの計算を分割して行うことで、全体の計算量を削減し処理の高速化を図っている。予測結果として周辺歩行者が数秒後までに移動する経路を出力し、観測者周辺の一定範囲に予測対象となる歩行者が侵入する可能性があるときに衝突可能性を算出し観測者に通知する。

3.2 研究課題

トップダウン方式の姿勢推定手法では計算時間が推定人数に比例するため、推定人数が多い場合に処理に時間がかかってしまう問題がある。したがって、混雑環境における姿勢推定処理速度の向上を研究課題として設定する。

4. 提案手法

4.1 提案手法概要

混雑環境における姿勢推定処理速度の向上のためには、図 2 に見られるように推定対象の人数を削減することが有効である。そこで、撮影した映像の中には観測者周辺の一定範囲に侵入する可能性のない周辺歩行者も存在するという想定のもと、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で、予測対象を限定し処理速度を向上させる。予測対象を限定する条件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。人物接近判定は人物検出の後の段階で行うため、特徴量として用いることができるのは人物検出の段階で得られる情報のみである。本研究で採用する YOLOv3 は物体検出の手法であり人物以外の物体検出も行うことができる。物体の検出と識別を同時に行うことで処理速度を向上させており、特徴として背景の影響を受けにくいことが挙げられる。映像を入力すると、検出した物体の種類、バウンディングボックス、信頼度スコアが出力として得られる。ここで得られるバウンディングボックスのサイズ変化から接近を判定する手法を提案する (図 4)。各フレームで検出した人物のバウンディングボックスから特徴量を抽出し、時系列における特徴量の変化量から相対距離の変化を求め接近を判定する。判定はカットオフ値を設定し二値分類で行い、特徴量の変化量がカットオフ値より大きい場合には接近、小さい場合には接近していないと判定する。1 秒を 30 フレームとして判定はフレーム毎に行い、1 フレームでは変化が小さく判定が難しいため 10 フレーム前からの変化量を用いる。検出に失敗した場合には 1 つ前のフレームの変化量を代入することで補正する。特徴量として (A) バウンディングボックスの縦軸の長さであるスケール値 (以下、特徴量 A)、バウンディングボックス下端の y 座標である位置座標 (以下、特徴量 B) の 2 種類を検討し、5.1 にて評価を行い採用する特徴量を決定する。

4.2 部分遮蔽の補正

障害物や画角の影響で起こる身体の部分遮蔽によって、検出した人物のバウンディングボックスのサイズが減少し誤検出が発生する。この誤検出は人物の実際の移動を反映していないため、人物の移動速度から変化量の上限値を設定しヒューリスティクスとして用いることで検知をする。移動速度に対して大きな値をその閾値としてバウンディングボックスの 1 フレームでの急峻な値の変化を部分遮蔽によるものとみなす。部分遮蔽が発生したタイミングでの 1 フレーム前からの変化をオフセットとし、部分遮蔽の起きている間はこれを検出したバウンディングボックスのサイズに加えることで補正する。

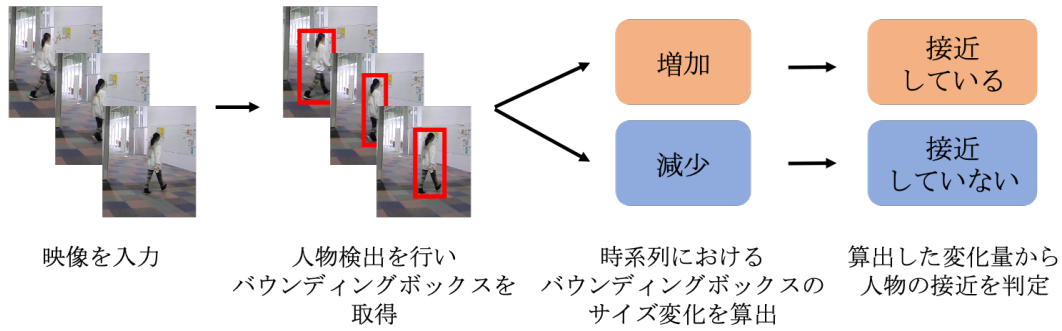


図 4 提案手法における映像の入力から人物接近判定までの流れ

表 1 カメラの仕様

使用カメラ	GoPro HERO7
解像度	1080p
フレームレート	30fps
焦点距離	6mm
深度視野	(水平) 87.6° (垂直) 56.7°

表 2 判定結果

		実際の移動	
		接近	接近していない
判定結果	接近	True Positive (TP)	False Positive (FP)
	接近していない	False Negative (FN)	True Negative (TN)

4.3 外れ値の補正

4.2 節の手法では 1 フレームでの変化が閾値を超えない段階的なものに対しては検知できないため、10 フレームでのバウンディングボックスの変化量に上限値を設定することで外れ値を検知する。移動速度に対して大きな値をその閾値として 10 フレームでの急峻な値の変化を誤検出とし、1 つ前のフレームのバウンディングボックスの変化量を代入することで補正する。

5. 特徴量ごとの人物接近判定の精度比較実験

5.1 実験概要

人物接近判定において (A) バウンディングボックスの縦軸の長さであるスケール値、(B) バウンディングボックス下端の y 座標である位置座標の 2 種類の特徴量のうち、実際に採用する特徴量を決定するために判定精度の比較実験を行った。1 人の歩行者が様々な方向に移動する映像に対して判定を行った。カメラは GoPro HERO7 を使用し、カメラの仕様は表 1 に示す。

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。カメラと歩行者の間に障害物のない映像と、障害物によって歩行者の身体が部分遮蔽された映像をそれぞれ 14 パターン、合計 28 パターン撮影し、接近移動と接近していない移動の 2 種類に分類して用いた。障害物の大きさ、位置はランダムであり、障害物自体は移動しないものとする。Precision-Recall 曲線 (PR 曲線) を作成し、Area Under the Curve (AUC), Precision, Recall の 3 つの指標を用いて評価を行った。AUC はグラフの曲線下の面積で 0 から 1 までの値をとり、値が 1 に近いほど判別能が高いことを示す。一般的には値が 0.8 以上であることが望ましい。Precision, Recall は表 2 に示す True Positive (TP), False Negative (FN), False Positive (FP),

True Negative (TN) の 4 つの数値を用い、それぞれを式 1, 2 によって計算する。本実験では、接近移動を接近と判定した数を TP, 接近していないと判定した数を FN, 接近していない移動を接近と判定した数を FP, 接近していないと判定した数を TN に分類している。

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

5.2 結果

特徴量ごとの人物接近判定の結果を PR 曲線に表したものを図 5 に示す。特徴量 A では AUC は 0.71 であるのに対し、特徴量 B では 0.84 と高い精度を達成した。特徴量 A はカットオフ値が増加しても Precision に大きな変化がなくどの値をカットオフ値に設定した場合でも 0.8 以下であり、最も低い時には 0.5 以下になることが確認された。特徴量 B ではカットオフ値の増加とともに Precision も増加しており、どの値をカットオフ値に設定しても特徴量 A より高い値となった。

各特徴量を利用した場合における障害物の有無による結果を図 6, 7 に示す。どちらの特徴量においても障害物の影響による数値の低下が確認できた。特に特徴量 A は障害物の有無による結果の差が顕著であり、AUC に 0.19 の差が生まれた。また、障害物ありの映像ではカットオフ値にかかわらず Precision がほぼ一定であることも確認できた。どちらの特徴量においても障害物ありの映像では、カットオフ値を高く設定した場合に Precision の急激な低下が確認できた。障害物なしの映像ではどちらの特徴量でも右肩下がりの曲線を描いているが、特徴量 A の AUC が 0.84

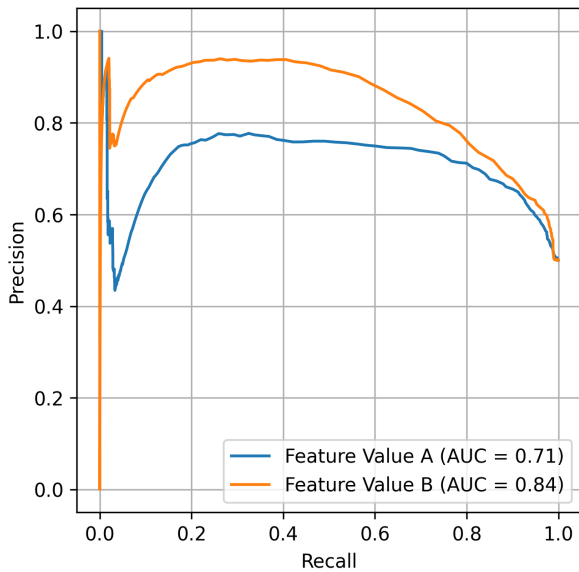


図 5 特徴量ごとの人物接近判定の PR 曲線

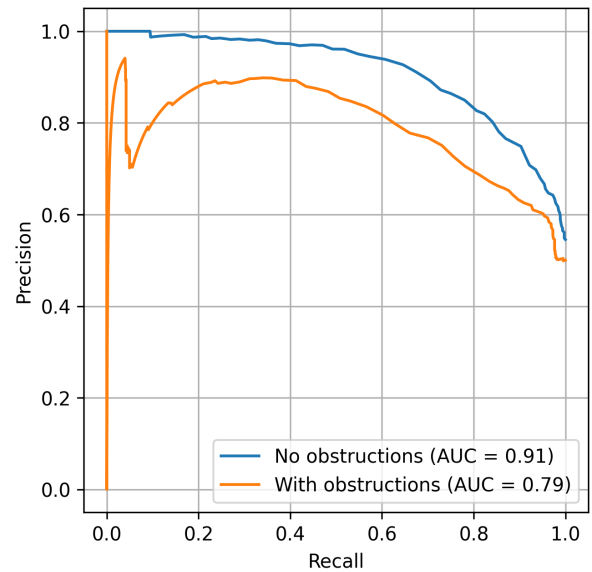


図 7 障害物の有無による PR 曲線 (特徴量 B)

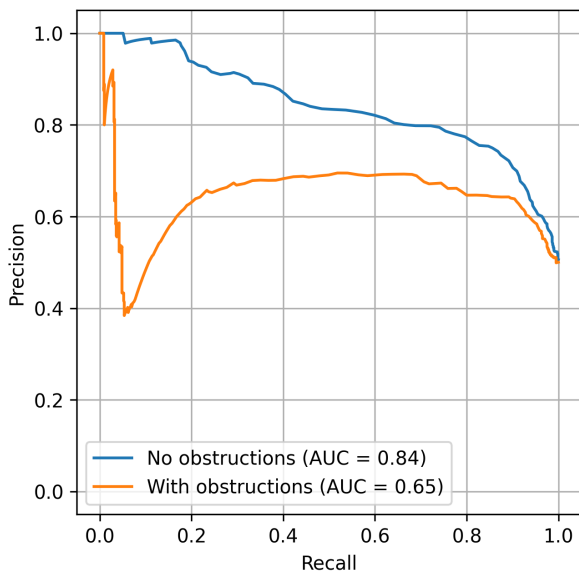


図 6 障害物の有無による PR 曲線 (特徴量 A)

であるのに対し特徴量 B では 0.91 であり、障害物の影響を除いても特徴量 B の方が高い精度を達成した。

5.3 考察

特徴量 A ではカットオフ値にかかわらず *Precision* がほぼ一定であったことから、あらゆる値に誤検出が分散していることが分かる。障害物の有無による結果を比較すると特徴量 A では *AUC* の値に大きな開きがあり、障害物の影響を大きく受けたことが確認できる。特徴量 A ではバウンディングボックスの上下端の値を用いているため、どちらか一方に誤検出が発生すると判定に影響があり、下端のみを用いる特徴量 B と比較して障害物の影響を大きく受けたと考えられる。本実験ではカメラを観測者の頭の位置

に設置したため、下端と比較して人物の移動による上端の数値の変化が少なく、反対に姿勢変化等による数値の変化が大きくなったことも原因と思われる。人物の歩行においては基本的にどちらかの足が地面に接しているため、足の動きによる下端の数値の変化は小さかったが、頭や身体全体の傾きによって上端の数値が変化していた。特徴量 B では姿勢変化等の人物の実際の移動を反映していない変化が少なかったため、特徴量 A と比較して高い精度を達成できたと考えられる。特徴量 B においては下半身が遮蔽された場合に移動の変化の検出難易度が高いことが問題であり、それが原因で障害物ありの映像では精度の低下が起きたと推測する。どちらの特徴量においてもカットオフ値を高く設定した場合に *Precision* の急激な低下が確認でき、障害物の影響により変化量が極端な値になったことが原因と考えられる。特徴量 B においては障害物ありの映像においてカットオフ値を低く設定した場合に *Recall* の低下が確認でき、これも同様の原因と考えられる。この問題に対しては 4.2 節、4.3 節にて述べたアプローチによって補正を行い 6 章にて効果を検証する。図 5 より特徴量 B では、人物検出の段階で得られるバウンディングボックスの情報のみから人物接近判定を行い、歩行者 1 人を撮影した映像において *AUC* が 0.84 を達成することができた。よって人物接近判定には特徴量 B を採用する。

6. 誤検出の補正に関する効果検証実験

6.1 実験概要

4.2 節、4.3 節にて述べた部分遮蔽と外れ値の補正に関して、効果を検証するために実験を行い補正ありとなしの場合を比較した。1 人の歩行者が様々な方向に移動する映像に対して判定を行い接近の判定漏れ・誤判定の少なさと判

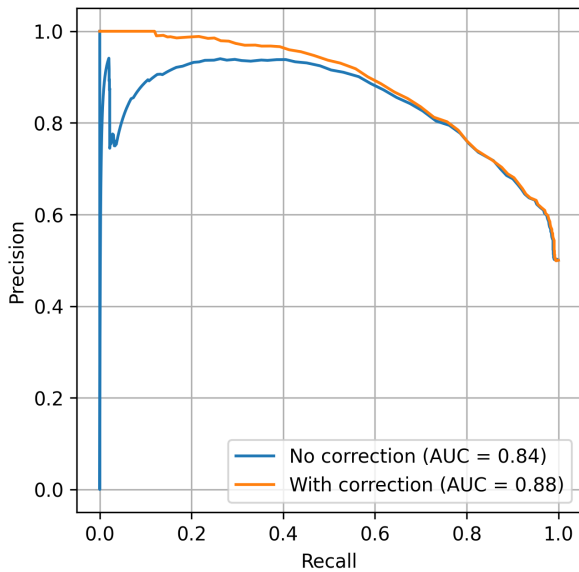


図 8 補正ありとなしの場合の人物接近判定の PR 曲線

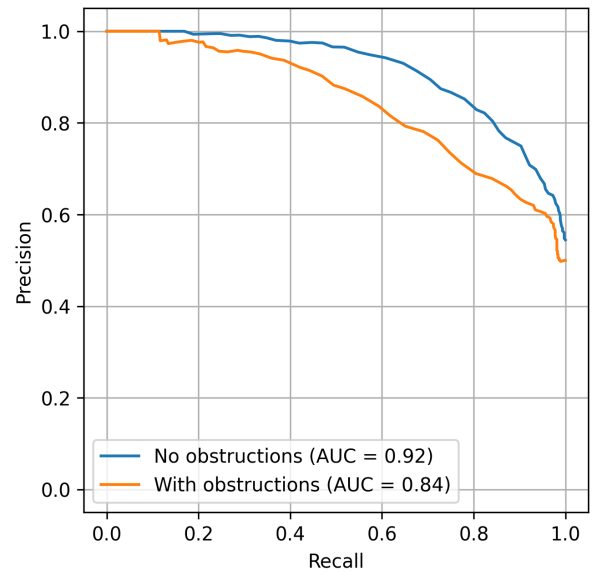


図 9 障害物の有無による PR 曲線 (補正あり)

定精度の評価を行った。映像の撮影環境と評価方法は 5.1 節での実験と同様であり、人物接近判定には特徴量 B を用いた手法を採用した。

6.2 結果

補正の有無による人物接近判定の結果の違いを PR 曲線に表したものを図 8 に示す。補正を行うことによってカットオフ値を高く設定した場合の *Precision* の低下が発生しておらず、誤検出を抑制できたことが確認できる。AUC の数値も 0.04 上昇しており、補正を行うことで全体的に精度を向上させることができた。

補正ありの判定における障害物の有無による結果を図 9 に示す。カットオフ値を高く設定した場合においても障害物ありの映像で *Precision* が低下しておらず、障害物の影響で発生していた誤検出を抑制できたことが確認できる。一方で、障害物ありの映像においてカットオフ値を低く設定した場合に *Recall* の低下が発生しており、全体的には障害物の有無による精度の差が確認できた。障害物ありとなしの映像では AUC に 0.08 の差が生まれた。

6.3 考察

補正ありの判定ではカットオフ値を高く設定した場合に *Precision* が 100% を記録しており、補正の効果が現れたことが確認できる。一方でカットオフ値を低く設定した場合に *Recall* の低下が確認できたことから、マイナスの値に対しての補正が不十分であることが考えられる。4.2 節の手法では人物を検出した段階で既に部分遮蔽が起きてる場合に検知ができず、4.3 節で外れ値を検知することで補正を行おうとしたが、ここで検知漏れが発生した可能性がある。4.2 節の手法ではバウンディングボックスの 1 フレームでの値の変化を比較しており、人物の移動による値の変

化は小さい一方で部分遮蔽による値の変化は大きいため誤検出を検知しやすい。しかし外れ値の検知では 10 フレームでの値の変化を比較しており、部分遮蔽による値の変化であったとしても人物の移動による値の変化と同程度の変化量である場合も多く、誤検出の検知難易度が高いため検知漏れが起これば補正が十分に機能しなかったと考えられる。全体的に障害物の有無による精度の差が確認でき、これは補正後の値が人物の実際の移動を反映できていないことが原因であると考えられる。特徴量 B を用いた手法では位置座標のみを用いているため、部分遮蔽が発生すると人物の移動を検出できる材料がなくなってしまうことが理由である。本稿で行った補正は明らかな誤検出を抑制することが目的あり、それによって補正した値は実際の人物の移動を反映できているわけではない。よって更なる判定精度の向上のためには補正方法の再検討を行う必要があると考えられる。図 8 から歩行者 1 人を撮影した映像において、補正を行うことで AUC の値を 0.04 向上させることができた。

7. おわりに

本稿では、混雑環境における歩行者同士の衝突回避のためのリアルタイムな行動予測を目的とし、リアルタイム処理の実現のために予測対象を限定する方法として観測者に接近する周辺歩行者を判定する手法の提案と評価を行った。人物検出の段階で得られるバウンディングボックスの情報のみを用い、サイズの変化によって人物接近判定を行った。歩行者 1 人を撮影した映像において AUC が 0.84 を達成することができた。また障害物による部分遮蔽の補正を行うことで補正を行わない場合と比較して、静止した障害物の存在する環境において AUC の値を 0.05 向上させることができた。

本稿で提案した人物接近判定の検証実験は、簡易的なものであり複雑な条件設定を行っていない。障害物自体が移動しないことを前提にしているが、実際には他の周辺歩行者等が障害物となることが想定されるため、今後は移動する障害物が原因で発生する部分遮蔽の補正とその検証を行う。また、更なる判定精度の向上のために補正方法の再検討を行う。提案手法を適用することによる姿勢推定の処理速度の変化を調査し、手法の効果を検証する。

参考文献

- [1] 東京消防庁防災部防災安全課. 緊急搬送データからみる日常生活事故の実態. 2019
- [2] 平川翼, 山下隆義, 玉木徹, 藤吉弘巨: 動画像を用いた経路予測手法の分類, 電子情報通信学会論文誌 D, vol.J102-D, no.2, pp.53-67 (2019).
- [3] Kitani, K.M., Ziebart, B.D., bagnell, J.A., and Hebert, M.: Activity Forecasting, European Conference on Computer Vision, pp.201-214 (2012).
- [4] Park, H.S., Hwang, J.-J, Niu, Y., and Shi, J.: Egocentric Future Localization, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.4697-4705 (2016).
- [5] Xiu, Y., Li, J., Wang, H., Fang, Y., and Lu, C.: Pose Flow: Efficient Online Pose Tracking, arXiv:1802.00977(online), available from <http://arxiv.org/abs/1802.00977> (accessed 2021-05-10).
- [6] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, arXiv:1812.08008(online), available from <http://arxiv.org/abs/1812.08008> (accessed 2021-05-10).
- [7] Fang, H.-S., Xie, S., Tai, Y.-W., and Lu, C.: RMPE: Regional multi-person pose estimation, arXiv:1612.00137(online), available from <http://arxiv.org/abs/1612.00137> (accessed 2021-05-10).
- [8] 八木拓真, マンガラムカーティケヤ, 米谷竜, 佐藤洋一: 一人称視点映像における人物位置予測, 情報処理学会研究報告, vol.2018-CVIM-211, no.3, pp.1-7 (2018).
- [9] Li, J., Wang, C., Zhu, H., Mao, Y., Fang, H.-S., and Lu, C.: CrowdPose: Efficient Crowded Scenes Pose Estimation and A New Benchmar, arXiv:1812.00324(online), available from <http://arxiv.org/abs/1812.00324> (accessed 2021-05-10).
- [10] 内田祐介, 山下隆義: 畳み込みニューラルネットワークの研究動向, 電子情報通信学会技術研究報告, vol.117, no.362, pp.25-38 (2017).
- [11] Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv:1704.04861(online), available from <http://arxiv.org/abs/1704.04861> (accessed 2021-05-10).
- [12] Redmon, J. and Farhadi, A.: YOLOv3: An incremental improvement, arXiv:1804.02767(online), available from <http://arxiv.org/abs/1804.02767> (accessed 2021-05-10).