Vol.2017-ITS-70 No.5
2017/8/4

頭部トラッキングに基づく車載カメラ映像からの人流推定法の提案

原 佑輔[1]　内山 彰[1]　梅津 高朗[2]　東野 輝夫[1]

概要：人流の把握は都市計画やマーケティング，安全な歩行者誘導などにおいて注目を集めている．本研究では，時空間的解像度の高い人流把握を低コストに実現するため，近年普及が進んでいる車載カメラを用いた人流推定法を提案する．車載カメラ映像では歩行者同士の重なりや障害物による遮蔽が頻発し，常に各歩行者の全身を捉えることは困難である．しかし，車載カメラは移動するため，あるフレームで映像中に現れていない歩行者であっても，前後のフレームでは映像中に現れる可能性が高い．提案手法ではこの特性に着目し，2 段階で人流を推定する．まず，映像の各フレームに対して深層学習により前方および後方の 2 種類に分けて頭部検出を行う．その後，時間的に連続するフレーム間の検出結果に対して，位置および色類似度に基づき人物同定を行う．提案手法の有効性を確認するため，実際に収集した車載カメラ映像に対し評価実験を行った．隣接する交差点間を 1 区間として，5 つの異なる区間に対して評価を行った結果，平均絶対誤差率は前方，後方それぞれ 19.2%，10.6%となり，提案手法の有効性が確認できた．

1. はじめに

都市計画，安全支援，マーケティングなど，様々な目的において都市部における歩行者の移動状況（人流）を把握することは重要である．例えば，把握した人流から人気のあるスポットを検出したり，混雑状況の監視・予測に基づく人流誘導を行うほか，災害時の帰宅困難者の救援計画立案にも活用できると考えられる．

このような人流や人々の分布状況を把握するため，これまでに様々な手法が提案されている．例えばモバイル空間統計 [1] では携帯電話の通信統計情報を用いて区画毎の人口推定を行っている．また，混雑度マップ [2] では GPS 対応の携帯電話利用者から許諾を得て送信される位置情報の分布からの人口推定を行っている．しかし，いずれも 250m メッシュなど比較的広い範囲ごとの人密度を推定するものであり，“ある道路の西側を駅方向に歩く人数” といったスポット的人流を把握する試みは見当たらない．一方，防犯カメラを用いて混雑状況を推定する手法 [3], [4] も存在するが，都市部全体の人流を把握するためには，膨大な数のカメラを設置する必要があり，設置場所やコストの制約上，現実的ではない．

[1] 大阪大学 大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University
[2] 滋賀大学 データサイエンス学部
Faculty of Data Science, Shiga University

そこで本研究では，近年普及が進んでいる車載カメラの映像を用いた歩道レベルでの人流推定法を提案する．様々な道路を走行している複数の車両で撮影された映像に対して，深層学習に基づき歩行者を検出することで人流を推定し，位置情報と共にサーバーで集約・統合する．移動する車載カメラを利用することにより，広範囲に対して低コストで歩道レベルという空間的に解像度の高い人流把握を実現する．

歩行者の検出には安全運転支援を目的とした手法 [5], [6], [7] などの適用が考えられるが，車載カメラ映像では歩行者同士の重なりや障害物による遮蔽が頻発し，常に各歩行者の全身を捉えることは困難である．このため，我々は遮蔽に強い深層学習に基づく Stewart らの手法 [8] を用いて歩行者を検出する．しかしながら，画像を利用するため，遮蔽による検出漏れを完全に防ぐことは本質的にできない．また，歩行者に類似する画像特徴を示す背景などが存在する場合も，誤検出が避けられない．これに対し提案手法では，車載カメラは移動しながら撮影するという特性に着目し，複数フレームにおける歩行者検出結果を統合することにより，検出漏れならびに誤検出を低減する．車載カメラでは移動しながら撮影を行うため，遮蔽により映像中に現れない歩行者であっても，前後のフレームでは捉えられている可能性が高い．このため，一時的な検出漏れに対する堅牢性を高められる．また，単一フレームでは誤検出となる場合であっても，複数フレームの検出

結果に基づき動きの無い背景などを除外できる．

提案手法は以下の 2 段階で人流を推定する．まず，映像の各フレームに対して深層学習により前方および後方の 2 種類に分けて頭部検出を行う．その後，時間的に連続するフレーム間の検出結果に対して，位置および色類似度に基づき人物同定を行う．類似度は検出された頭部領域の位置，および服などを含む周辺領域の色分布により定義される．複数フレーム間の検出結果を統合することにより，単一フレームでは避けられない誤検出や検出漏れを除外したうえで，前方および後方の方向別に移動している歩行者数を推定する．

提案手法の性能を評価するため，実際に大阪市内で収集した車載カメラ映像を用いて実験を行った．隣接する交差点間を 1 区間として，5 つの異なる区間に対して評価を行った結果，平均絶対誤差率は前方，後方それぞれ 19.2%，10.6%となり，提案手法の有効性が確認できた．

図 1　人流推定結果の例

2. 関連研究

2.1 車載カメラを用いた人検出

自動運転車に関連する技術の発展とともに，安全運転支援を目的として，車載カメラを用いた歩行者検出法が数多く研究されている．これらの手法は，人の動きを検出する方式と人の形状を検出する方式の 2 種類に大別される．文献 [5] では人特有の動きのパターンを特徴量として歩行者を検出する．しかし，この手法は動きのパターンを抽出するために歩行者の足が一定時間見えている必要がある．また，人の動きを用いて検出を行っているため静止している歩行者は検出することができない．

一方，人の形状を特徴量として歩行者を検出する手法は移動している人と静止している人の両方を検出することができる．文献 [6] ではウェーブレット解析 [9] と Support Vector Machine(SVM)[10] を用いて歩行者検出を行っている．これらの手法は運転支援を目的としており，群衆中では人同士の重なり (オクルージョン) が大きく影響し，検出精度が低くなるという問題が生じる．

2.2 CNN を用いた物体検出

CNN を用いた物体検出は，ImageNet[11] で注目を集めて以来，様々な方式が考案されている．中でも CNN を用いた画像中に複数存在する可能性のある複数クラスの対象物検出は Localization and Classification と呼ばれ，難しい問題の一つである．R-CNN[12] は CNN を用いた複数クラスの対象物検出手法の一つであり，Selective Search[13] により物体の候補領域を抽出したうえで，CNN による分類を行う．これによって，単純な sliding window を用いた総当たりでの分類よりも高速に物体検出を行うことができる．しかしながら，歩行者同士の重なりが頻発する場合，Selective Search により抽出された候補領域が正しく複数の歩行者を捉えることができず，検出漏れが多発するなど，依然として課題が残されている [14]．一方，Stewart らの LSTM に基づく人検出法 [8] では，人同士の重なりが生じる場合でも精度良く人検出を行う手法を提案している．Stewart らの手法では，候補領域の抽出処理を必要とせず，入力画像全体に対して人が存在する可能性が最も高い領域を一つずつ順番に検出することで，重なりがある場合でも高い精度を実現している．また，入力画像全体から歩行者の頭部検出を行うため，上半身や足，腕など，頭部以外の身体の一部だけでも画像に写っている場合に，頭部のみを利用した手法よりも高い性能を発揮する．このため，提案手法では Stewart らの手法を車載カメラ映像向けにチューニングし，歩行者頭部の検出を試みる．

3. 提案手法

3.1 概要

3.2 想定環境

本研究では，少数の協力ユーザーや自治体職員などがカメラを搭載した車両で対象領域を走行し，得られた映像を用いることを想定する．車載カメラは，ダッシュボードにマウントされたスマートフォンや一般のドライブレコーダーを利用する．ドライブレコーダーの中にはスマートフォンや車載器などと WiFi により接続できる製品が存在する．したがって，携帯通信網により外部ネットワークに接続されたスマートフォンや車載器をゲートウェイとすることで，ドライブレコーダーの映像をサーバーに送信できる．ただし，通信量をできるだけ抑えることが望ましいため，取得した映像に対して，スマートフォンや車載器で処理を行い，方向別の歩行者数を推定した後，その結果の

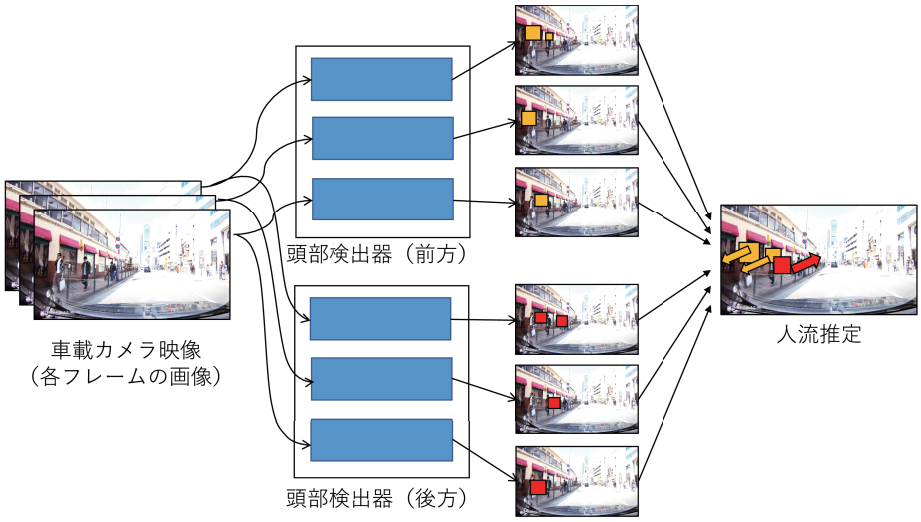

図 2 提案手法の概要

みをサーバーに送信することを想定している．なお以降では，方向別の歩行者数を人流と表記する．人流の推定結果は撮影位置および時刻とともにサーバーに送信され，サーバーでは，複数車両から送られてきた各地の歩道における人流を統合し，地図上にマッピングする（図 1）．

本研究では簡単のため大まかな歩道位置が事前に分かるものとし，手動で決定した 640x480 ピクセルの領域を検出対象とする．実際には，いくつかの方法により歩道位置の大まかな推定は実現可能であるが，本研究では対象外とする．例えば，画像認識により車線を判別したり，一定の期間，画像全体に対して歩行者検出を行い，検出結果位置の分布から一定以上の頻度で歩行者が現れている領域付近を歩道として推定する，といった方法が考えられる．さらに，GPS により得られた車両位置と車線数などの道路情報や，カメラの設置位置に関する情報を併用することもできる．

また，ほとんどの歩行者は前方または後方のどちらかに移動することを想定している．実際の歩行者は前方，後方以外の方向にも移動したり，立ち止まったりする可能性があるが，遮蔽を含む短い映像から歩行者の様々な状態を推定することは困難である．しかし，交差点間の区間においては，多くの歩行者が前方または後方のいずれかに移動していると考えられるため，人流推定という目的においてはそれ以外の歩行者を無視しても大きな影響は無いと考える．ただし，交差点においては立ち止まる歩行者が多数存在するため，本研究では交差点付近を検出対象外として手動で除外している．

3.3 概要

図 2 に示すように提案手法では，以下の 2 ステップに分けて人流推定を行う．

(1) フレームごとの方向別頭部検出
(2) 複数フレームにおける頭部検出結果の位置および画像類似度に基づく人物同定

まず，車載カメラにより撮影された映像の各フレーム（静止画）において，方向別の歩行者頭部検出を行う．これは画像内の物体の場所とクラス（どこに何があるか）を決定することに等しく，画像処理の分野では Localization and Classification と呼ばれる問題である．これに対して，本研究では文献 [8] の手法を適用し，前方，後方の方向別頭部検出器を構築する．次に，各フレームの頭部検出結果における検出漏れや誤検出の影響を軽減するため，時間的に連続する複数フレームにおける検出結果の位置関係や画像の濃度分布で定義される類似度に基づき，移動方向別の歩行者人数を推定する．

3.4 深層学習による方向別頭部検出

提案手法では，前方，後方の 2 種類の方向別頭部検出を行うため，Stewart らが提案した人検出法 [8] を適用する．図 3 に頭部検出の概要を示す．まず，縦 480 ピクセル，横 640 ピクセルの入力画像を GoogLeNet に与えることで，20×15 のセルそれぞれにおける 1024 次元のベクトルを特徴量として得る．各セルのベクトルは画像中の対応領域における特徴を要約しており，物体の位置情報も含んでいると考えられる．この各セルの特徴量を LSTM に入力することで，検出対象（頭部前方または後方）が存在する位置を bounding box として検出信頼度とともに出力する．LSTM では信頼度が高い bounding box から順に出力がなされる．この時，直前の出力結果を次のユニットに入力することにより，同一対象の重複検出が起こらないようにしている．これを信頼度が閾値 T 以上の bounding box が見つからなくなるまで繰り返す．最終的に，得られた複数の検出結果を統合することで，一つの入力画像に対する検出結果が得られる．学習は，文献 [8] で提案されている損失関数に従って行うものとした．検出後，重なりの生じ

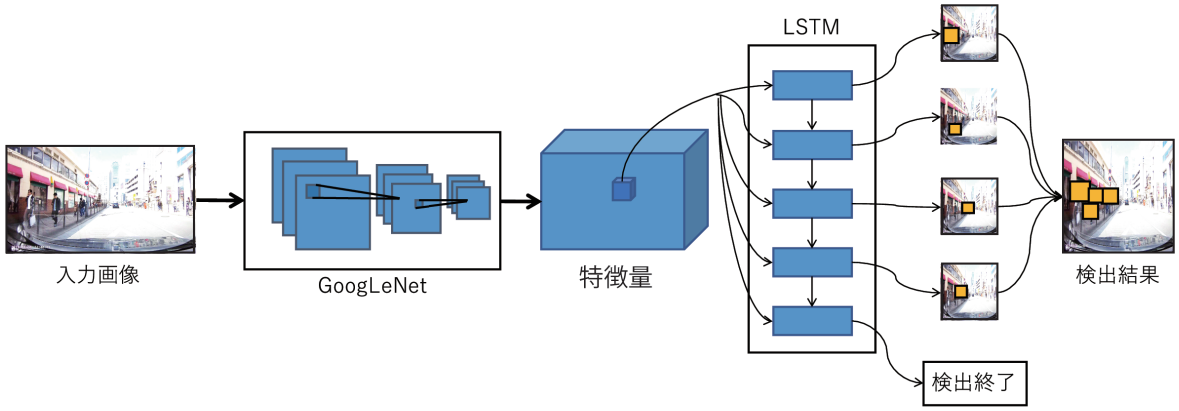

図 3　深層学習による方向別頭部検出の概要

ているウィンドウを除外する．あるフレームにおいてウィンドウの中心がどちらか一方に入っていた場合にスコアが高いものを用いるものとする．

3.5　人流推定

3.5.1　人物同定アルゴリズム

複数フレームにおける方向別の頭部検出結果を統合し，歩行者 1 人 1 人の移動軌跡を推定することによって，人流推定を行う．各フレームにおける頭部検出結果は，その瞬間の画像特徴量のみを用いているため，誤検出や検出漏れが避けられない．そこで，頭部検出結果の時空間的な特徴や画像特徴を考慮することによって，誤検出や検出漏れに対する堅牢性の向上を図る．

フレーム t において検出された i 番目の bounding box を $b_i^t \in B^t$ とする．B^t はフレーム t で検出された bounding box の集合である．また，b_i^t により検出された人物の ID を $y(b_i^t)$ とする．提案手法では，b_i^t と b_j^u が同一人物であるか否かを判定するため，類似度 $\mathrm{sim}(b_i^t, b_j^u)$ を以下のように定義する．

$$\mathrm{sim}(b_i^t, b_j^u) = w_1 l(b_i^t, b_j^u) + w_2 v(b_i^t, b_j^u) \quad (1)$$

ここで，$l(b_i^t, b_j^u), v(b_i^t, b_j^u)$ はそれぞれ b_i^t, b_j^u の位置関係，画像特徴量に基づき定義される類似度であり，w_1, w_2 は重みである．これらの類似度の定義は続く 3.5.2 節，3.5.3 節で述べる．

同一人物の判定および移動軌跡の推定は以下の手順で行う．まず，類似度の定義に基づき，全ての隣接するフレーム間において，同一人物の判定を行う．ここで，$y(b_i^t)$ を，b_i^t と同一人物のものと見なせる bounding box の集合と定義し，各 b_i^t に対して，$y(b_i^t) = \{b_i^t\}$(同一人物と見なせるのはそれ自体のみ) と初期化する．具体的には，フレーム $t, t+1$ 間の bounding box ペアのうち，$\mathrm{sim}(b_i^t, b_j^{t+1})$ が最大のペア $(b_i^t, b_j^{t+1}) \in B^t \times B^{t+1}$ について，$\mathrm{sim}(b_i^t, b_j^{t+1})$ が閾値 S 以上であれば両 bounding box は同一人物のものと見なし，$y(b_i^t) = y(b_j^{t+1}) = y(b_i^t) \cap y(b_j^{t+1})$ として，b_i^t, b_j^{t+1} をそれぞれ B^t, B^{t+1} から除外する．これを閾値を越えるペアが無くなるまで繰り返す．同様にして，フレーム t と $t+2, t+3, \ldots, t+W$ 間の bounding box ペアに対して，フレーム間隔を 1 ずつ増やしながら順に人物を同定する．これによって，ある程度の検出漏れを許容した人物同定を実現している．

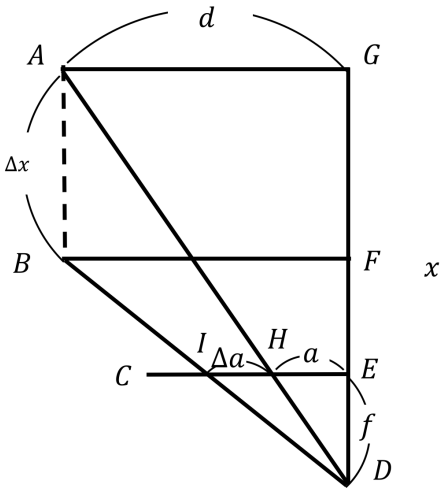

図 4　車載カメラと歩行者の関係図

最後に，同一人物と推定されたものが検出されたフレームが W 以下であったものは除外する．ただし，W は誤検出を除外する為のパラメータである．以上により歩行者一人一人の移動軌跡が推定されるため，歩道における移動方向別の歩行者数が得られる．

3.5.2　位置類似度

位置類似度 $l(b_i^t, b_j^u)$ は時刻 t における b_i^t の中心座標 p_i^t と時刻 $u(t < u)$ における b_j^u の中心座標 p_j^u により定義する．車両と歩行者の相対移動速度に基づき p_i^t にいた歩行者が時刻 u で画像中に現れるべき位置 $p_i^{t\to u}$ を求め，$p_i^{t\to u}$ と p_j^u の距離が近いほど，高い類似度が与えられるようにする．具体的には，位置類似度は以下の関数で定義される．

$$min\left(\frac{1}{|p_i^{t\to u} - p_j^u|}\right) \quad (2)$$

または

$$max\left(0, 1 - \frac{|p_i^{t\to u} - p_j^u|}{(|\Delta a| + M)}\right) \tag{3}$$

ただし，Δa は後述の値，M は許容するマージンの値である．経験的に $M = 50$ とした．これらの 2 つの性質の異なる関数を用いた．

$$l(b_i^t, b_j^u) = -\frac{p_j^u}{p_i^{t\to u}} + 1 \tag{4}$$

$p_i^{t\to u}$ は以下のようにして幾何的に求める．図 4 に車載カメラと歩行者の関係を上から見た図を示す．歩行者 P はある時刻 T で点 A にいるものとする．点 D にカメラの焦点があり，カメラは D から E の方向を向いているものとする．線分 CE はカメラの投影面であり，点 A にいた歩行者は画像中の点 H(画像中心から a[pixel]) に位置する．歩行者 W は ΔT 秒後にカメラを基準として点 B に相対的に移動するものとする．線分 AB の長さ Δx は GPS 情報から求めた車両の速度と一般的な歩行者の速度から求める．この時の歩行者 P は画像中 I(画像中心から $a + \Delta a$ [pixel]) に位置するものとする．d はカメラから歩行者までの距離，f は焦点距離をピクセル単位で表したものである．

歩行者の画像上の（見かけ上の）移動距離を求めるには $a, d, f, \Delta x$ が与えられた場合の Δa を求めればよい．$\triangle DEH \sim \triangle DGA$ より

$$a : f = d : x \tag{5}$$

$\triangle DEC \sim \triangle DFB$ より

$$(a + \Delta a) : f = d : (x - \Delta x) \tag{6}$$

式 5，6 より

$$\Delta a = \frac{a^2 \Delta x}{fd - a\Delta x} \tag{7}$$

となる．ただし，f は w を画像の横幅，θ を水平方向の画角とすると

$$f = \frac{w}{2\tan\frac{\theta}{2}} \tag{8}$$

である．

以上より，

$$p_i^{t\to u} = a + \frac{a^2 \Delta x}{fd - a\Delta x} \tag{9}$$

となる．

3.5.3 画像類似度

画像特徴量に基づく類似度は，歩行者の服の濃度に着目して以下のように定める．服の画像領域は，検出された頭部のウィンドウに対し，縦方向のウィンドウサイズを地面の方向に 3 倍した領域を対象とする．今回は画像をグレースケールに変換した後にヒストグラムを計算する．基準となる頭部のヒストグラムと，対象の頭部のヒストグラムの相関を計算することで，類似するヒストグラムであれば大きな値を出力するようになる．具体的には，以下の式で定義する．

図 5 実験で走行した道路 (大阪市茶屋町周辺)

$$v(b_i^t, b_j^u) = \frac{\sum_I (H_i(I) - \bar{H}_i)(H_j(I) - \bar{H}_j)}{\sqrt{\sum_I (H_i(I) - \bar{H}_i)^2 \sum_I (H_j(I) - \bar{H}_j)^2}} \tag{10}$$

ただし H_k は b_k^s のヒストグラム， $\bar{H}_k = \frac{1}{N}\sum_J H_k(J)$ であり，N はヒストグラムのビンの総数を表す．

4. 性能評価

4.1 実験環境

提案手法の性能評価を行うため，大阪市茶屋町周辺の道路をドライブレコーダ (ユピテル社製 DRY-WiFiV5c) を設置した自動車で複数回通行し，映像を撮影した．撮影は，図 5 の地点 1 から地点 2 間の水色で示されている道路において，多くの人が行き交う休日の正午頃に行った．学習用データは前方の画像が 2560 枚， 後方の画像が 2695 枚である．

評価用データは頭部検出の評価には車載カメラ画像 100 枚，人流推定の評価には 5 箇所の歩道の動画を用いた．5 箇所の歩道の歩行者数は前方 93 人，後方 57 人である．評価指標には Precision, Recall, Accuracy を用いた．また，方向別頭部検出の比較対象として，Open CV により実装した Haarlike 特徴量を用いて学習用データから頭部を切り出して学習させて作成した検出器の場合との比較を行った．

学習を行う際のパラメータは文献 [8] に記載されている設定を用い，学習用ライブラリ及び学習ソフトは著者らが公開しているものを用いた[*1]．学習に用いたワークステーションのスペックは CPU が Intel(R) Xeon(R) CPU E5-1680 v3 @ 3.20GHz，メモリ 128GB，GPU は GeForce GTX 1080 である．

[*1] https://github.com/Russell91/ReInspect

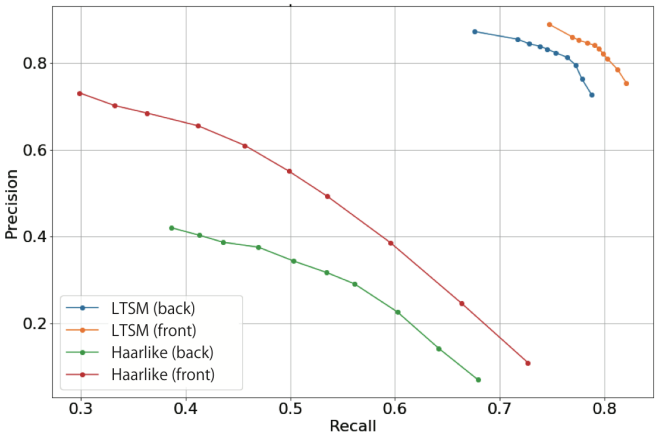

図 6　閾値の影響（Precision と Recall）

4.2　評価結果

4.2.1　方向別頭部検出

図 6 に頭部検出における LSTM の閾値 T を変化させた時の Precision と Recall を示す．Haarlike 特徴量を用いた場合，Precision, Recall はそれぞれ 0.1〜0.7，0.7〜0.3 程度となっており，Precision が最も高い 0.7 程度の場合でも Recall が 0.3 程度まで低下しているため，十分な性能が出ているとは言い難い．一方で，提案手法の Precision, Recall はそれぞれ 0.7〜0.85，0.85〜0.68 程度に収まっている．これは前方，後方どちらの場合も共通であり，Haarlike 特徴量を用いた場合と比べて，提案手法が Precision, Recall ともに大きく上回っていることが分かる．このような結果となった理由は，人同士の重なりが頻発する場合においても，提案手法により誤検出や検出漏れを抑えることができているためと考えられる．

一方，Haarlike 特徴量と提案手法のどちらの場合でも，後方の性能は前方よりも低下している．この原因として，前方の場合は目や鼻，口など様々な顔の部位が画像に表れるため，検出に有益な特徴量が得られやすいが，後方の場合は髪の毛で頭部が覆われてしまい，画像から十分な特徴量を得られにくいことが考えられる．それでもなお，提案手法は方向別頭部検出において高い Precision, Recall を達成しており，その有効性が確認された．

提案手法では，複数フレームにおける頭部検出結果を統合して人流推定を行うため，単一フレームにおける誤検出は除外できる．以上の評価結果より，Recall が最も高い $T = 0$ に設定しても Precision は 0.7 を超えていることから，以降の評価では $T = 0$ を用いた．

4.2.2　人流推定

前節の評価で得られたフレームごとの方向別頭部検出結果に対して，人物同定を行い，人流の推定を行った．

交差点から交差点までの歩道 5 箇所についての人数の絶対誤差率の平均で評価を行った．類似度の閾値 S と最小の検出フレーム W を変化させて，方向別に結果をプロットした．S は大きくなるほど同一人物の判定が厳しくなるパラメータであり，パラメータ W は大きくなるほど 1 人とカウントするのに多くのフレームで検出する必要がある．

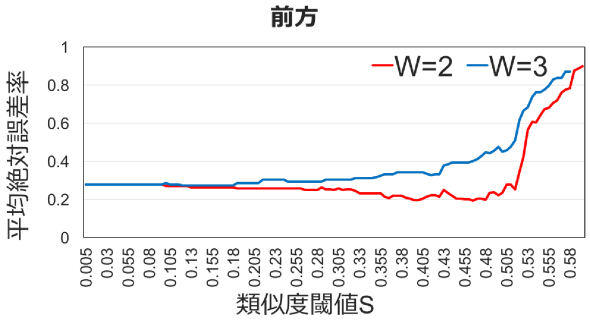

図 7　人流の誤差 (前方)

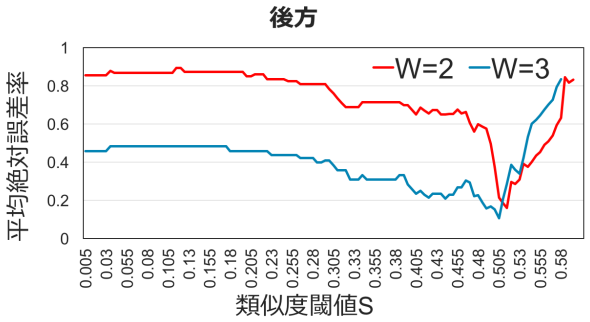

図 8　人流の誤差 (後方)

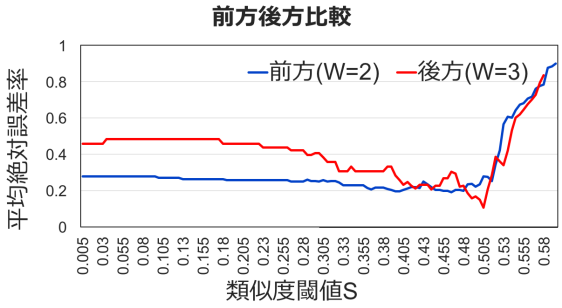

図 9　人流の誤差 (前方後方比較)

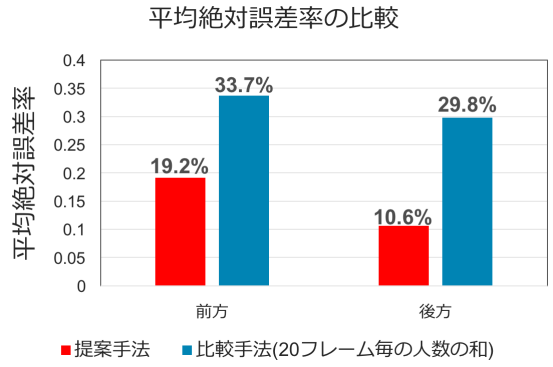

図 10　ナイーブ手法との比較

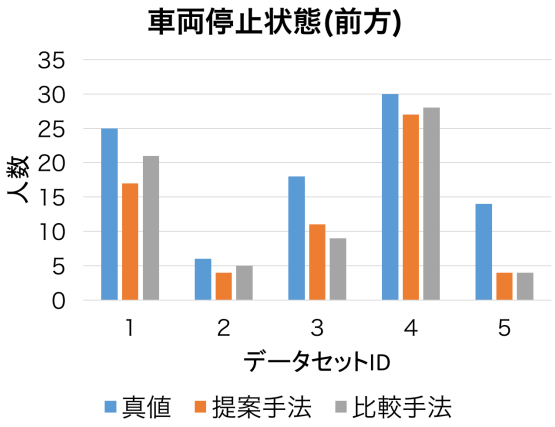

図 11　停止状態での前方歩行者人数推定

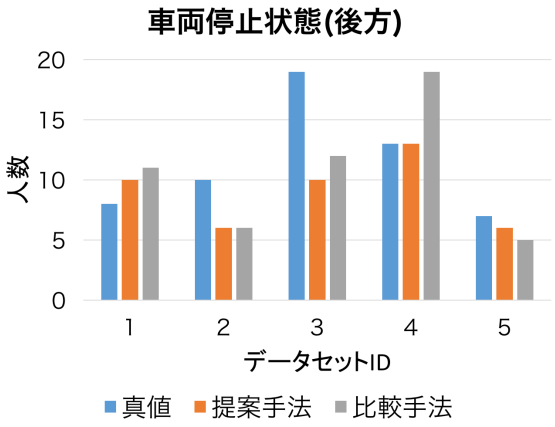

図 12　停止状態での後方歩行者人数推定

また，3.5.2 節で述べた位置類似度関数は式 2 を用いており，$p_i^{t\to u}$ は走行中車両の歩行者の平均的な画像中の移動量を固定値として用いている．車両速度の変化に対しての本手法の評価は 4.2.3 節で行っている．式 (1) の頭部位置予測で用いるパラメータは $w_1 = w_2 = 1$ とした．

図 7 に前方の結果を示す．W が増加するほど平均誤差率が大きくなっていることが分かる．これは，前節の頭部検出の結果から前方の頭部検出は後方と比べ誤検出が少ないため，W を大きくすることで真値である歩行者も除外していると考えられる．

図 8 に後方の結果を示す．こちらは W が増加するほど平均誤差率が小さくなっていることが分かる．これは，後方の頭部検出は前方と比べ誤検出が多いので W を大きくすることで誤検出が除外できていると考えられる．

図 9 に前方と後方の比較結果を示す．類似度閾値 S が小さいときは前方の方が後方と比べ，平均誤差率が小さくなっている．これは，S が小さい場合は誤検出であっても同一人物であると紐付けをする場合が多くなり，結果として検出器の性能差が人流の推定性能に表れていると考えられる．

図 10 にナイーブ手法との比較を示す．ナイーブ手法として，ダブルカウントが起こらないフレーム間隔での頭部検出結果の人数の和を取るといったものを用いた．フレーム間隔は 20 フレームとした．この結果から提案手法の方が平均誤差率が小さくなっており，移動軌跡を推定することで誤検出，検出漏れが補正されていることがわかり，本手法の有効性が確認された．

4.2.3　車両速度変化への対応

本節では車両速度が変化した場合での性能評価を行う．車両が停止している場合の 5 箇所の動画と車両が走行している場合の交差点間 5 箇所の歩道の動画それぞれにおいて評価を行った．提案手法に対し $p_i^{t\to u}$ を走行中車両の歩行者の平均的な画像中の移動量を固定値として用いた手法で性能比較を行う．図 11，12，13，14 に個別のデータセットにおける車両停止状態の前方歩行者，後方歩行者，走行状態の前方歩行者，後方歩行者数の真値と推定結果を示す．また，それぞれの場合においての平均相対誤差を表 1 に示す．

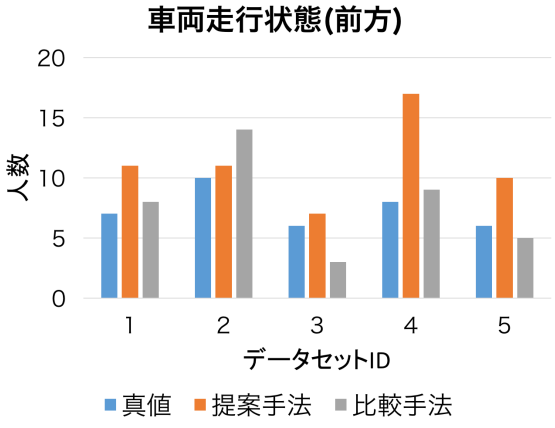

図 13　走行状態での前方歩行者人数推定

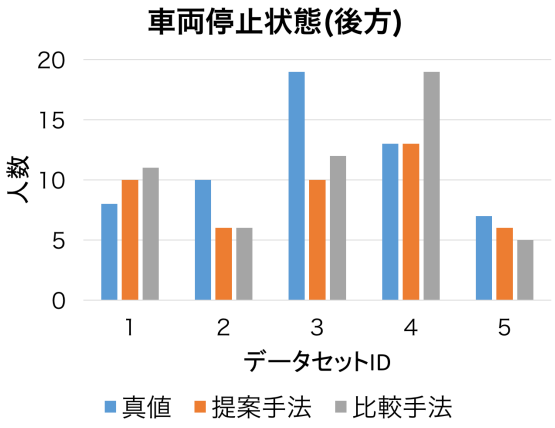

図 14　走行状態での後方歩行者人数推定

停車時は比較手法と比べ提案手法は後方の歩行者の推定誤差が小さくなっている．これは比較手法が走行時の車両から見た歩行者の相対速度の平均を移動量としているため，歩行者は車両と反対向きに進むと予測される．したがって

後方に進む歩行者は車両から見たら車両と同じ側に進むために比較手法の予測とは逆方向に進んでいることとなり，類似度が小さい値となり人物同定がされにくくなる．

一方提案手法は車両の速度と歩行者の速度の相対速度によって予測移動量を変化させるため，後方へ進む歩行者でも人物同定が正しく行えているものと考えられる．停車時は歩行者の速度が車両速度と比べ相対的に大きくなるため，性能差が顕著に表れている．

一方で車両停車時の前方歩行者推定性能は比較手法に劣っている．これは．車両速度を GPS から取得しており，GPS 誤差の影響で車両停車時でも車両の速度が 0 でない値が取得されているからだと考えられる．これについてスマホ等の機器から加速度を取得し，車両の停止状態を把握することができれば対策可能だと考えられる．

表 1 平均相対誤差

方向	状態	提案手法	比較手法
前方	停止	37.1%	32.1%
	走行	52.6%	26.7%
後方	停止	25.3%	37.8%
	走行	37.3%	62.9%

5. おわりに

本研究では，街中を走行する車両の車載カメラ映像を利用した歩道レベルでの人流推定法を提案した．車載カメラ映像では歩行者や障害物による遮蔽が頻発するため，常に全ての歩行者を捉えられるとは限らない．そのため，連続する複数フレームでの頭部検出結果に対し，時空間的な位置関係及び画像特徴量による人物の同定を行い，映像中の歩行者の移動軌跡を推定する．歩道上に存在する歩行者の移動方向は車の進行方向に対して前方と後方に大別されるため，2 種類に分けて頭部を検出する．頭部検出では遮蔽が頻発する環境でも堅牢性の高い LSTM (Long Short-Term Memory) に基づく手法を適用する．提案手法の有効性を確認するため，実際に収集した車載カメラ映像に対し評価実験を行った．また，その結果を時系列的に処理を行い，検出位置の特徴と服の色の特徴を用いて同一人物判定を行い，人物の移動軌跡推定を行い歩道の人流推定を行った．5 箇所の歩道に対し人数の平均絶対誤差率で評価を多なったところ，前方は 19.2%，後方は 10.6%となり，本手法の有効性を確認する事ができた．

今後，様々な場所におけるデータに対する評価や，車速が変化した場合でも対応できるように改良を進めていく予定である．

謝辞

本研究は JSPS 科研費 JP26220001，JP26700006，JP16K00123 の助成を受けたものです．

参考文献

[1] 寺田雅之，永田智大，小林基成：モバイル空間統計における人口推計技術 (社会・産業の発展を支える 「モバイル空間統計」：モバイルネットワークの統計情報に基づく人口推計技術とその活用)，NTT DoCoMo テクニカル・ジャーナル， Vol. 20, No. 3, pp. 11–16 (2012).

[2] 株式会社ゼンリンデータコム：混雑度マップ，`http://lab.its-mo.com/densitymap/`.

[3] Silveira Jacques Junior, J., Musse, S. and Jung, C.: Crowd Analysis Using Computer Vision Techniques, *IEEE Signal Processing Magazine*, Vol. 27, No. 5, pp. 66 – 77 (2010).

[4] Wu, Z., Thangali, A., Sclaroff, S. and Betke, M.: Coupling detection and data association for multiple object tracking, *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1948–1955 (2012).

[5] Wöhler, C., Anlauf, J. K., Pörtner, T. and Franke, U.: A Time Delay Neural Network Algorithm for Real-Time Pedestrian Recognition, *Proceedings of International Conference on intelligent vehicle*, pp. 247–251 (1998).

[6] Papageorgiou, C., Evgeniou, T. and Poggio, T.: A Trainable Pedestrian Detection System, *Proceedings of Intelligent Vehicles*, pp. 241–246 (1998).

[7] Lee, K.-H., Hwang, J. N., Okapal, G. and Pitton, J.: Driving recorder based on-road pedestrian tracking using visual SLAM and Constrained Multiple-Kernel, *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 2629–2635 (2014).

[8] Stewart, R., Andriluka, M. and Ng, A. Y.: End-To-End People Detection in Crowded Scenes, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).

[9] 山口昌哉：ウェーブレット解析，科学，Vol. 60, pp. 398–405（オンライン），入手先 ⟨http://ci.nii.ac.jp/naid/10006233574/⟩ (1990).

[10] Bradski, G. and Kaehler, A.: 詳解 OpenCV: コンピュータビジョンライブラリを使った画像処理・認識，O'Reilly Media, Inc. (2009).

[11] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database, *CVPR09* (2009).

[12] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587 (online), DOI: 10.1109/CVPR.2014.81 (2014).

[13] Uijlings, J. R., Sande, K. E., Gevers, T. and Smeulders, A. W.: Selective Search for Object Recognition, *Int. J. Comput. Vision*, Vol. 104, No. 2, pp. 154–171 (online), DOI: 10.1007/s11263-013-0620-5 (2013).

[14] 原 佑輔，小島颯平，Elhamshary, M. M.，内山 彰，梅津高朗，東野輝夫：車載カメラを用いた CNN による方向別歩行者頭部検出法の提案，情報処理学会研究報告，24, Vol. 2016-MBL-81, pp. 1–8 (2016).