



Title	路線バスにおけるカメラ画像を用いたバス乗客OD推定システムに関する研究
Author(s)	西浦, 翼
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第16985号
Issue Date	2026-03-25
DOI	https://doi.org/10.14943/doctoral.k16985
Doc URL	https://hdl.handle.net/2115/99829
Type	doctoral thesis
File Information	Tsubasa_Nishiura.pdf, 全文



路線バスにおけるカメラ画像を用いた バス乗客OD推定システムに関する研究

西浦 翼

北海道大学 大学院情報科学院

2026 年 1 月

目次

1 序論	1
1.1 研究背景	1
1.2 研究目的	2
1.3 本論文の構成	2
2 関連研究	4
2.1 路線バス	4
2.1.1 路線バスを取り巻く環境	4
2.1.2 バス OD データ	6
2.2 Computer Vision に関する研究分野	13
2.2.1 Computer Vision の基礎と代表タスク	14
2.2.2 人物計測	17
2.2.3 人物追跡	18
3 バス OD 推定システムの全体設計	20
3.1 はじめに	20
3.2 路線バス利用者の動画データ	20
3.2.1 動画取得	20
3.2.2 取得動画の分析	22
3.3 提案システム：BAIRD-OD	24
3.3.1 要素技術の検討	25
3.3.2 提案アルゴリズム	26
3.4 提案システムの有効性の検証	33
3.4.1 データセット	33
3.4.2 実験 1：バス停別乗降者数推定	34
3.4.3 実験 2：乗車区間別乗降者数推定	35
3.5 まとめ	38

4 個人識別の精度向上	41
4.1 はじめに	41
4.2 個人識別における精度向上手法の検討	41
4.2.1 個人識別における課題	42
4.2.2 推定精度向上手法	43
4.3 提案手法	44
4.3.1 人物検出へのアプローチ	44
4.3.2 人物追跡へのアプローチ	45
4.4 提案手法の有効性の検証	47
4.4.1 予備実験1：人物検出器の選定	47
4.4.2 予備実験2：人物追跡器の学習	47
4.4.3 実験1：バス停別乗降者数推定	49
4.4.4 実験2：アブレーションスタディ	51
4.4.5 実験3：乗車区間別乗降者数推定	51
4.5 まとめ	52
5 乗降判定の精度向上	54
5.1 はじめに	54
5.2 乗降判定における精度向上手法の検討	54
5.2.1 乗降判定における課題	55
5.2.2 提案手法	56
5.3 提案手法の精度検証	57
5.3.1 データセット	57
5.3.2 実験1：物体検出器の精度確認	57
5.3.3 実験2：DNN モデルの学習	58
5.3.4 実験3：バス停別乗降者数推定	59
5.3.5 実験4：乗車区間別乗降者数推定	60
5.4 まとめ	61
6 結論	62
謝辞	65
参考文献	66

目次

1.1	本論文の構成の概略図	3
2.1	コンパクトシティの例	4
2.2	トリップの例	6
2.3	トリップエンドの例	7
2.4	代表的交通手段の優先順位	7
2.5	10 カードの片利用（国土交通省資料 [1] より引用）	9
2.6	バス利用者における交通系 IC カードの利用率（国土交通省資料 [2] より引用）	10
2.7	年代別のバス利用者における交通系 IC カードの利用率（国土交通省資料 [2] より引用）	11
2.8	交通系 IC カードの導入費用の例（出典：国土交通省「交通系 IC カードの普及・利便性拡大に向けた検討会 とりまとめ」図表 12[3]）	12
2.9	長崎県バス各社における導入費用（出典：国土交通省「月例経済報告関係資料」図 7「長崎県バス各社における導入費用」[4]）	13
2.10	OD データの図	13
2.11	バスの経路	14
2.12	バス OD データの図	14
3.1	動画データを収集した路線図	21
3.2	収集した動画の例	21
3.3	自動車車体規格のガイドラインによるドライブレコーダ設置位置 [5]	22
3.4	乗車口と降車口に設置したカメラ位置	22
3.5	白飛び、黒つぶれが発生した人物	23
3.6	オクルージョンの発生	23
3.7	暗い色・類似した服装の利用者	24
3.8	提案システム：BAIRD-OD	25
3.9	動画取得の概略図	26
3.10	人物検出の概略図	28

3.11 人物追跡の概略図	28
3.12 人物軌跡	29
3.13 扉範囲定義の概略	29
3.14 乗降判定の概略	30
3.15 乗車人物の軌跡と降車人物の軌跡	31
3.16 軌跡間特徴量距離行列	32
3.17 乗車人物の軌跡と降車人物の軌跡のマッチング	33
3.18 扉範囲の設定（左：乗車口，右：降車口）	34
3.19 特徴量距離の分布	37
3.20 乗車口（左）と降車口（右）での明るさの違い	39
3.21 乗車口（左）と降車口（右）での角度の違い	40
3.22 マッチングに失敗した画像例（左：乗車口での ID15,18, 右：降車 口での ID15,18）	40
4.1 本論文における第 4 章の位置づけ	41
4.2 ID スイッチが発生した例	42
4.3 軌跡の断片化が発生した例	42
4.4 大きい矩形	46
4.5 小さい矩形	47
4.6 右側に映る矩形	48
4.7 軌跡の結合	48
5.1 本論文における第 5 章の位置づけ	54
5.2 誤って乗車人物と判定される例	55
5.3 バス内外のアノテーション（緑：inside, 赤：outside）	58

表目次

2.1	交通系 IC カード利用率調査のサンプル内訳	9
3.1	第 3 章で用いるデータセットの概要	34
3.2	乗車口におけるバス停別乗降者数推定結果	35
3.3	降車口におけるバス停別乗降者数推定結果	35
3.4	アルゴリズムごとの乗車区間別乗降者数推定結果	36
3.5	誤ってマッチングしたペア	36
3.6	Rank-N の割合	37
3.7	誤ってマッチングしたペアの特徴量距離	38
4.1	人物追跡モデルの比較	43
4.2	人物検出精度へのアプローチ手法の比較	44
4.3	人物追跡精度へのアプローチ手法の比較	45
4.4	人物検出器の推定精度比較	49
4.5	学習データとテストデータの割当	49
4.6	人物追跡器 MPNTrack の推定精度	49
4.7	データセットの詳細	50
4.8	バス停別乗降者数の誤差率	51
4.9	乗車口における各条件の効果比較	51
4.10	降車口における各条件の効果比較	52
4.11	第 4 章手法を実装した場合の乗車区間別乗降者数推定結果	52
5.1	第 5 章で用いるデータセットの概要	57
5.2	YOLOv10 の検出精度	58
5.3	乗車口における DNN モデルのパラメータのグリッドサーチ結果	59
5.4	降車口における DNN モデルのパラメータのグリッドサーチ結果	59
5.5	乗車口におけるバス停別乗降者数推定の結果	60
5.6	降車口におけるバス停別乗降者数推定の結果	60
5.7	第 5 章手法を実装した場合の乗車区間別乗降者数推定結果	60

第1章 序論

1.1 研究背景

日本国内において路線バスは、地域住民の基礎的な移動を担う重要な公共交通である一方、地方・郊外を中心に赤字路線の増加や運転士不足が深刻化している。限られた経営資源の下で持続的な運行を維持するためには、利用実態に即したダイヤ編成や経路再編、車両・乗務員の効率的な配置が不可欠である。

その基礎情報となるのが、利用者の乗降地点の組を記録したOD(Origin-Destination)データである。バス停単位での需要分布やバス停間の移動分布が把握できれば、時間帯別の増減便、バス停配置の見直し、区間別運賃体系の設計、乗継利便性の評価など、運行計画やサービス設計に直接的に資する。

従来のバス OD データ収集には、乗客アンケート、交通系 IC カード、RFID 等のタグ方式が用いられてきた。しかし、アンケートは労力の割に回収率が低く継続性に乏しい。交通系 IC カードは発行・管理主体とバス事業者が必ずしも一致せずデータアクセスに制約が生じやすい。タグ方式は利用者の能動的協力を前提とし、導入・維持に係る機器費や運用負担が小さくない。このように、費用面・運用面・アクセス面の観点から、広域かつ長期に渡りスケラブルに適用できる収集手段は限定的であった。

近年、防犯・安全目的で普及した既設の車内カメラを活用すれば、上記の費用・アクセス制約を緩和しつつ OD を推定できる可能性がある。ただし、本研究が位置づく Computer Vision (CV) による人物計測、とりわけ人物追跡の観点では、環境特性が性能に大きく影響する。一般に屋内環境では固定視点や安定照明を前提とした背景・幾何情報の活用が有効であるのに対し、屋外環境では照度・天候の変動、スケール変化、群衆による遮蔽などが主要課題となる。路線バスは定義上は屋内であるが、走行に伴う外交の急変、扉開閉に起因する突発的な流入出、高密度・強遮蔽の狭い通路、車体振動によるブラー等、屋外的・動的要因を併せ持つ中間環境であり、汎用ベンチマーク前提の手法とはドメインギャップが生じやすい。このため、既設の車載カメラ映像を前提に、中間環境に適応した人物計測により乗降情報を抽出して OD を推定する枠組みは、ダイヤ編成や経路再編といった運行上の意思決定をデータで支援する情報システムの一形態として社会的需要が高い。さら

に、特定の意思決定状況（路線バス運行）をインスタンスとして、実装制約下で有効な要素技術の組み合わせと評価の方法論を示すことは、情報科学における屋内/屋外の環境にある中間環境を対象とした人物計測研究の発展、および都市交通分野におけるデータ駆動型意思決定支援に必要な入力データの生成可能性と要件の整理に寄与することが期待される。

1.2 研究目的

本論文では、路線バスにおけるカメラ画像を用いた OD データの収集を目的とする。本論文における OD データの詳細は第 2 章に定める。

本研究が想定する運用像は次のとおりである。全国の路線バスを対象に、運行中に取得した車内カメラ映像を用いて、運行後のバッチ処理として OD データを推定する。推定結果は翌々月の運行計画の見直しでの活用を想定し、地域差・季節差といった環境変動に対して、最小限のパラメタ調整で追従可能な運用を前提とする。リアルタイム処理は必須要件としない。

対象とする車両は、国内で一般的な 12m 級・2 ドア（中乗り前降り）の路線バスとする。この前提は視野・動線・乗降位置の分布を規定し、人物計測および乗降判定の要件に直接影響する。当該前提の妥当性や対象路線の特性は第 2 章に整理する。

本研究が目指す OD 推定システムの要件は、以下のとおりである。①統計利用に足る整合的な精度（個人認証を目的としない）、②低コスト・低工数での導入・運用、③既設カメラ活用と運行後処理を前提とした実装容易性、④プライバシー保護、⑤照度変動・オクルージョン・車体揺動等へのロバスト性である。

一方で、カメラ画像を用いた OD データ収集には、実運用を想定しない評価設計、分布交通量への未対応、および屋内と屋外の環境に位置づけられる車内の中間環境に起因する人物追跡・同一判定の困難といった課題がある。本研究は、これらの制約を踏まえ、実運用上の要件に整合した設計判断とモジュール別の精度改善によって、コストと精度の両立を図る。

1.3 本論文の構成

本論文では路線バスを対象として、車載カメラの映像から利用者の OD データを推定するシステムに関する研究を実施した。本論文の概略図を図 1.1 に示す。第 2 章では、本研究の関連研究について示す。第 3 章では、バス OD 推定システムの全体設計についての研究成果を示す。第 4 章では、バス OD 推定システムにおける個人識別の処理の精度向上についての研究成果を示す。第 5 章では、バス OD 推定システムにおける乗降判定の処理の精度向上についての研究成果を示す。第 6 章では、本論文のまとめと今後の展望について示す。

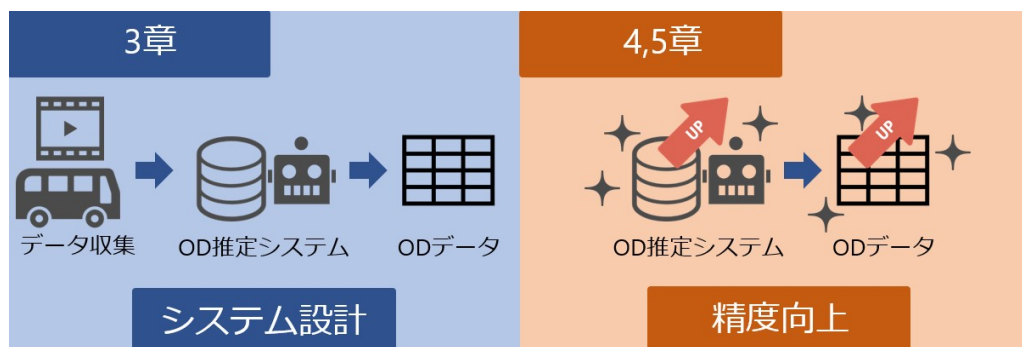


Figure 1.1: 本論文の構成の概略図

第2章 関連研究

本章では、研究対象である路線バスと、研究分野である Computer Vision (CV) に関する周辺知見を概観し、本研究の位置づけを整理する。

2.1 路線バス

本節では、路線バスの運行環境・制度・経営の観点と観測条件、OD データの定義・特性、取得手段とデータ可用性、応用先と要求精度を整理する。

2.1.1 路線バスを取り巻く環境

日本国内において路線バスは、地域住民の基礎的な移動を支える中核的な公共交通である。とりわけ近年推進されているコンパクトシティ化の文脈では、都市間を結ぶ鉄道と、中心市街地と周辺居住地を結ぶ路線バスが役割分担し、地域の移動需要を面的に担保することが想定される。現実にも、都市中心部では路線の重複・競合、

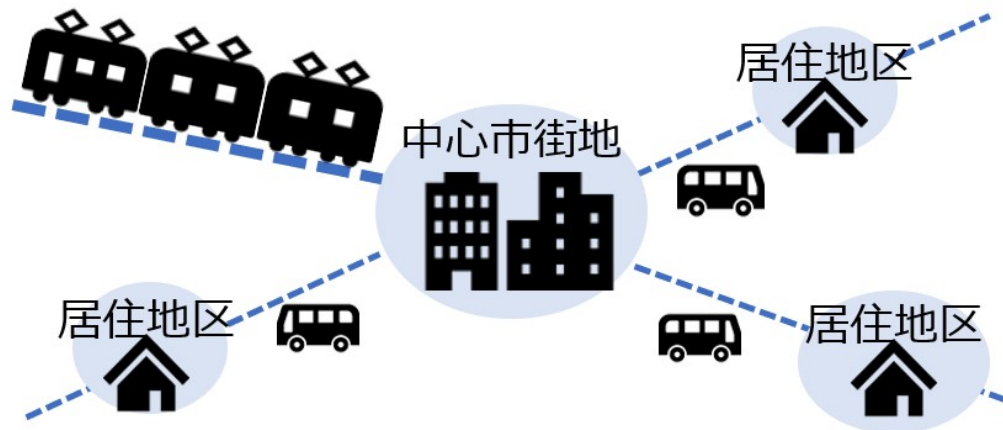


Figure 2.1: コンパクトシティの例

郊外では赤字路線の増加など課題を抱えつつも、移動需要に応じた経路変更・増減便の裁量が比較的大きな交通モードであり、適切な運行調整のために需要の時空間分布把握が重要となる。本研究で対象とするバス乗客 OD (Origin-Destination) データは、運行の効率化に直結する基礎情報であり、活用事例も多い [6, 7, 8]。

福島交通株式会社の取り組み事例として、NORUCA により得られる OD データを居住人口や就業人口、あるいは地理的データと重ね合わせることで、線的、面的な分析を実現し、便数の増減や経路変更等の可能性の検討に活用している [6].

西日本鉄道株式会社の取り組み事例として、nimoca の OD データを活用するとともに、乗継割引やポイント付与制度と組み合わせて利用者の乗り換え抵抗の抑制を図ることにより、福岡市内におけるバス路線再編を実現した [6].

佐賀県の事例として、佐賀都市圏の公共交通は都市圏を東西に横断する JR 長崎本線と佐賀駅バスセンターを起終点に放射状の運行形態となっている、という背景から、公共交通網計画の策定に取り組んでいる [7].

一方で、路線バス事業の経営環境は厳しい。国土交通省の白書等によれば、全国のバス事業者の約 7 割が赤字であり、地方圏では約 9 割が赤字とされる [9]. 加えて人口減少に伴う輸送人員の減少や担い手不足が長期的課題となっている [10, 11, 12]. こうした制約下で持続的運行を維持するには、運行実態に即したデータ利活用と低コストな収集手段の確立が不可欠である。また、バス事業者の赤字状況が IC カード導入・データ取得のボトルネックになっている。

交通系 IC カードは、全国相互利用 (Suica, PASMO, ICOCA 等) の仕組みの下で利便性が高く、近年の利用者調査では鉄道・バスともに決済手段として IC 利用が 7 割程度との結果も報告されている [13]. しかし、発行・管理の主体は多くが鉄道事業者であり、バス単独でデータにアクセス・活用できるとは限らない。さらに、IC の全国的普及を進める政策がある一方で、バス側の機器導入コストは数億円規模とされ、特に地方事業者にとっては負担が大きい [14]. このため、バス事業者自らが追加ハードを最小化しつつ需要データを収集できる代替手段の検討が求められている。

車両の物理的な条件も観測・データ収集に影響する。国内の路線バスは 12m 級が 90 % 以上を占めており [15], 扉配置は中乗り・前降りの 2 ドア構成が 50 % 以上を占める [16]. 車内は通路が狭く、立席密度の上昇や乗降時の滞留により遮蔽が多発する。さらに走行に伴う車体揺動・ブラー、停車中のドア開閉に伴う急激な照度変動や逆光など、車内は「屋内でありながら屋外的要因を強く受ける中間環境」である。この特性は、既設カメラの映像から人物を検出・追跡し乗降を判定する計測の難易度を押し上げるため、後述の CV 要素技術の選択・設計に直接関わる。

バス停の設置・移設・廃止やダイヤ変更には、複数の制度的手続きが関わる。まず、道路上に標柱・上屋・ベンチを設置する場合は、道路法第 32 条に基づく道路占用許可が必要であり、上屋・ベンチは占用物件としての取扱いが整理されている。次に、ダイヤ (運行時刻・本数等) の変更は運行計画の変更届出に該当し、原則として実施予定日の 30 日前までの届出が求められる。また、停留所の名称・位置は事業計画の記載事項であり、変更後は「遅滞なく」事業計画変更の届出 (停留所様式) が必要となる。加えて、運輸規則に基づき、運行に関する主要事項の変更

は少なくとも7日前の公示等が求められる。これらの制度要件は、停留所配置の見直しやダイヤ改編といったODデータの主用途に直結し、データの粒度・更新頻度の要件設計にも影響を与える。

2.1.2 バス OD データ

交通計画学における人の移動

交通計画学や交通工学の文脈では人物の移動の計測が重要であり、トリップという単位で扱われる。トリップとは、人が何か目的を持って、ある地点から別の地点まで移る移動のことと定義される [17, 18, 19, 20]。対象が人ではなく車両のこともある。トリップの例を図 2.2 に示す。トリップは目的トリップと手段トリップに大きく二分される。目的トリップはリンクトリップや純流動トリップとも呼ばれ、1つの目的を達成するための移動である。一般的にトリップとはこちらを指す。手段トリップはアンリンクトリップや総流動トリップとも呼ばれ、リンクトリップ中に使用した交通手段で分割したものを指す。

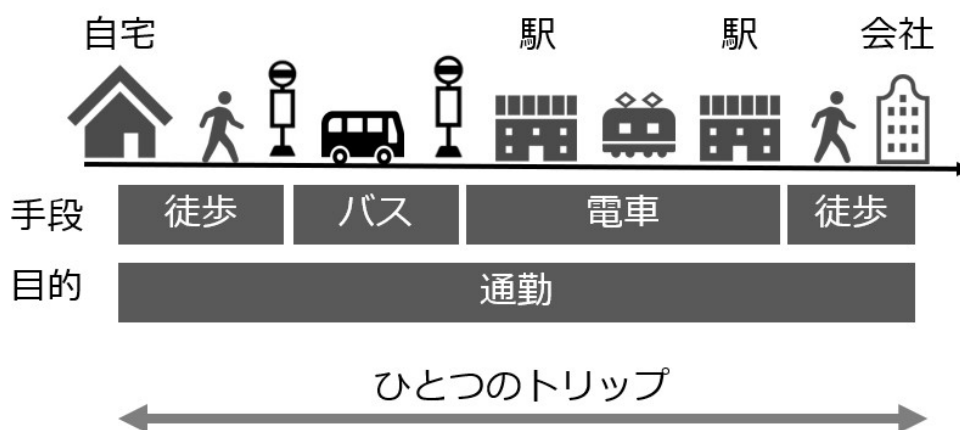


Figure 2.2: トリップの例

トリップの目的は以下の5つに分類される。

- 出勤，通学一日のうち最初に勤務先へ行ったトリップ
- 登校，通学一日のうち最初に通学先へ行ったトリップ
- 自由買い物，食事，レジャーなど私用のトリップ
- 業務配達，会議，農作業など仕事のためのトリップ
- 帰宅自宅へのトリップ

図 2.3 のように，トリップにおける出発地と到着地のことをトリップエンドと呼ぶ [19, 20]．特定のゾーンを出発地として，同一のゾーン内の移動を内々トリップ，ゾーン外への移動を内外トリップ，ゾーン外からの移動を外内トリップ，ゾーン外同士での移動を外外トリップと呼ぶ．

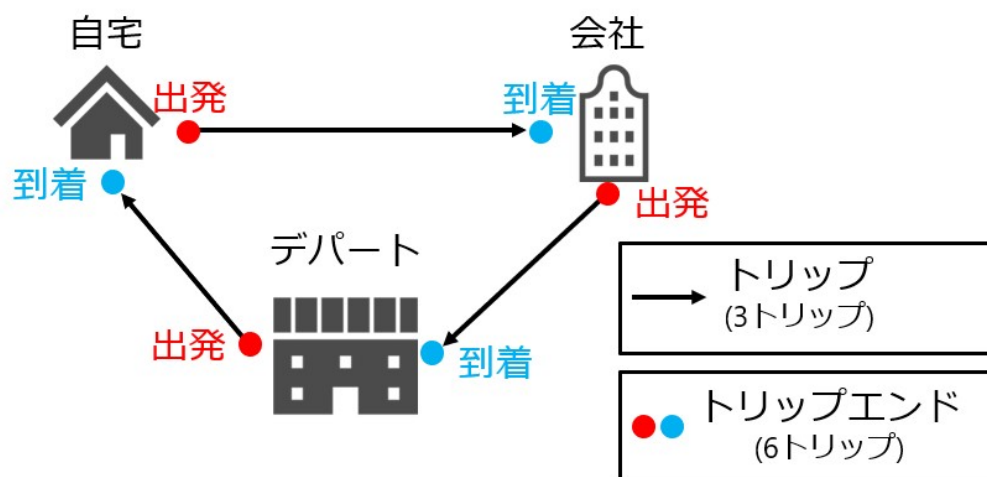


Figure 2.3: トリップエンドの例

トリップには代表的交通手段という概念があり，これは複数の交通手段を利用した場合の主な移動手段を指し，図 2.4 のように優先順位がある．航空機，船舶，鉄道，バス，タクシーは公共交通に分類される [21, 22]．公共交通は，不特定多数の人々が所定の運賃を支払い自由に利用できる交通機関である．それに対して乗用車，自動二輪，原付，自転車は個別交通に分類される．個別交通は個人が所有する交通手段である．また，代表的交通手段に至るまでの交通をアクセス交通，代表的交通手段から到着地までの移動をイグレス交通と呼ぶ．



Figure 2.4: 代表的交通手段の優先順位

トリップ情報を収集する代表的な手段にパーソントリップ調査がある [23]．パーソントリップ調査は1日の移動をアンケート形式で収集する調査で，全国都市交通特性調査，三大都市圏パーソントリップ調査，都市圏パーソントリップ調査がある．全国都市交通特性調査は全国 PT と呼ばれ，調査主体は国である．実施周期は5年に1度で，統計法上の位置づけは一般統計調査である．三大都市圏パーソント

リップ調査は東京、近畿、中京の三大都市圏を対象とした調査で、国と地方公共団体が共同で行っている。実施周期は10年に1度程度で、一般統計調査である。都市圏パーソントリップ調査は三大都市圏以外を対象とした調査で、地方公共団体が主体で行われている。実施周期は10年に1度程度で、統計法上は届出調査である。パーソントリップ調査のアウトプットとして、土木分野への応用やインフルエンザの被害予測などに使われている。

交通系 IC カード

近年ではトリップ情報の収集として、IC カードの利用も進んでいる。IC カードは交通系と流通系に分かれ、交通系 IC カードの中でも特に下記を合わせて10カードと呼ぶ。

- Kitaca：JR 北海道，2008 年 10 月導入
- Suica：JR 東日本，2001 年 11 月導入
- PASMO：PASMO 協議会，2007 年 3 月導入
- taICA：JR 東日本，2006 年 11 月導入
- manaca：トランパス IC 協議会，2011 年 2 月導入
- ICOCA：JR 西日本，2003 年 11 月導入
- PiTaPa：スルッと KANSAI 協議会，2004 年 8 月導入
- nimoca：西鉄，2008 年 5 月導入
- はやかけん：福岡市共通局，2009 年 3 月導入
- SUGOCA：JR 九州，2009 年 3 月導入

10 カードやその他の地域独自の交通系 IC カードは、各地域で相互に交通系 IC カードを利用できるようにするため、2013 年 3 月 23 日に IC カード相互利用センターが導入された。図 2.5 のように IC カード相互利用センターを通して各カードが連携することで、地域ごとに導入されていた交通系 IC カードが全国で利用できるようになっていく。

国土交通省では2022年3月7日から3月11日に、1,457人を対象として交通系 IC カード利用率を調査している [2]。サンプルの内訳を表 2.1 に示す。バスにおいて交通系 IC カードを中心に利用している人は、図 2.6 に示すように 70.7%であった。図 2.7 に年代別のデータも示す。

路線バスにおける交通系 IC カード導入の課題として、機能とコストのバランスが挙げられる。交通系 IC カードは鉄道と路線バスを中心に利用される。鉄道には

<「10カードの片利用」の現状と接続方式>

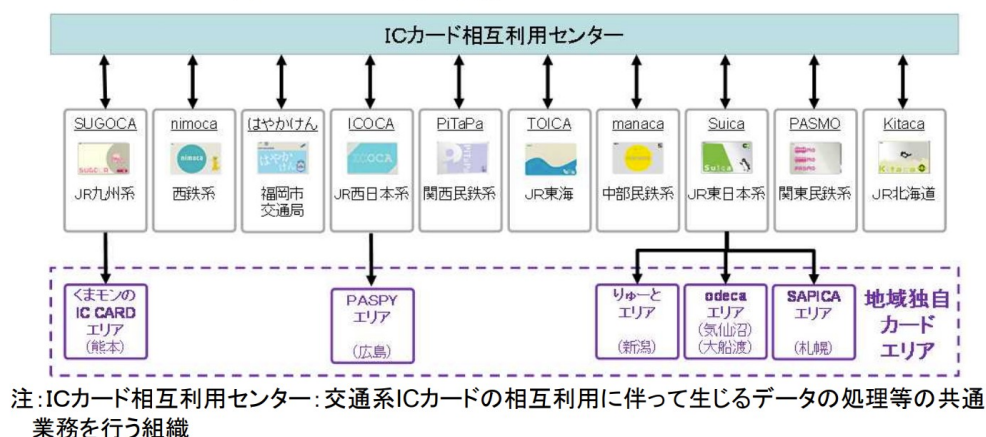


Figure 2.5: 10 カードの片利用（国土交通省資料 [1] より引用）

Table 2.1: 交通系 IC カード利用率調査のサンプル内訳

区分	性年代	n	居住地区分		
			東京都区部・政 令指定都市	人口 10 万人以 上の市	人口 10 万人未 満の市町村
全体		1457	500	500	457
性年代	男性 10 代～20 代	150	50	50	50
	男性 30 代	140	50	50	40
	男性 40 代	145	50	50	45
	男性 50 代	147	50	50	47
	男性 60 代以上	142	50	50	42
	女性 10 代～20 代	150	50	50	50
	女性 30 代	150	50	50	50
	女性 40 代	138	50	50	38
	女性 50 代	145	50	50	45
	女性 60 代以上	150	50	50	50

複雑な相互乗り入れが存在するため交通系 IC カードのメリットが大きく、さらに駅に固定の改札を置くだけでいいという特徴がある。それに対して路線バスでは、割引制度等の柔軟性が求められていることや交通系 IC カードの読み取り機器を車両ごとに設置しなければならないという点で相性が悪い。また、ビジネスモデルという観点で、交通系 IC カードの多くは鉄道事業者によって発行されているという特徴がある。鉄道事業者のビジネスモデルは駅周辺の不動産やリテールからの収益、そして日常の決済手段になれば得られる加盟店手数料が利益の源泉である。それに対してバス事業者は収益の源泉が住民の移動にあり、加盟店手数料はむしろコストになる。人口減少が著しい地方では移動サービスの収入増のための設備投資が難しいのが現状である。さらに旅客事業の公共性からくる完全キャッシュレスへの移行の難しさもある。旅客事業者は地域の移動手段という公共性の高い役割

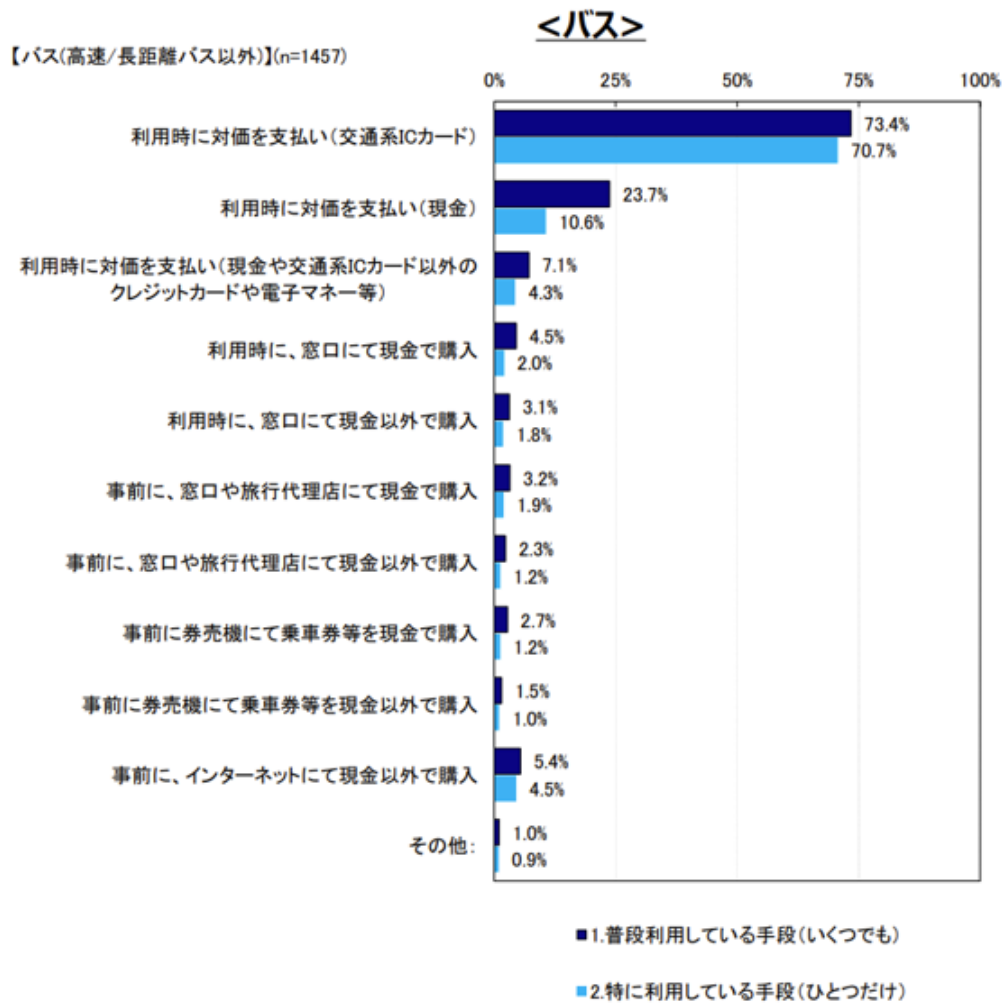


Figure 2.6: バス利用者における交通系 IC カードの利用率（国土交通省資料 [2] より引用）

を担うため、効率改善をうたうにしても、アクセスできなくなる利用者を発生させるわけにはいかない。具体的には、キャッシュレス決済を利用できない人が存在する限り、何らかの配慮が必要となる。かといって、現金での運賃収受も続けられ、現金管理負担が引き続き残存することとなり、効率化の障壁となる。

導入コストの例を図 2.8 に示す [3]。事業者の規模やバスの台数にも左右されるが、導入費用は数千万円から数億円に上る。実際の例として、図 2.9 に示す長崎県バス各社における導入費用の例では、出入口の端末設置費用、販売用窓口における端末設置費用、ソフト開発費用、IC カード作成費用など、合計で 8 億円以上の費用が掛かっている。

OD データ

交通計画学や交通工学で用いられる OD（Origin – Destination）データは、人やモノの移動について、出発地（Origin）と目的地（Destination）の対応を集計した

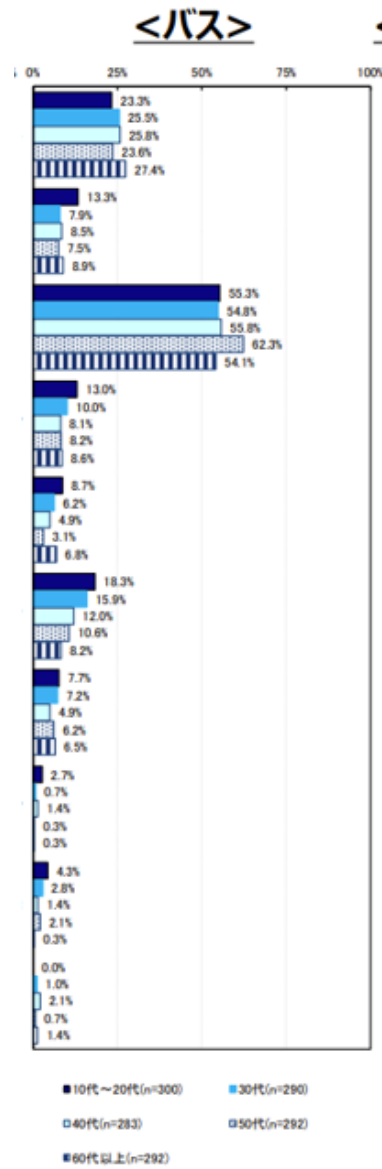


Figure 2.7: 年代別のバス利用者における交通系 IC カードの利用率（国土交通省資料 [2] より引用）

ものであり，都市交通・物流・小売・防災など広い分野で活用されてきた．OD は，時間帯やゾーンに応じて行列表（OD 表）として整理され，行方向に起点，列方向に終点を取り，セル値に分布交通量を格納する．

td_{od} : 分布交通量 (Traffic Distribution)

tg_o : 発生交通量 (Traffic Generation)

ta_d : 集中交通量 (Traffic Attraction)

T : 生成交通量 (Trip Generation)

$$= \sum_{o=1}^{B-1} tg_o = \sum_{d=2}^B ta_d$$

図表 12 交通系 IC カードの導入費用の例

	導入態様	車両数	導入費用	備考
a社	【方策A】	バス 43 台	85 百万円	運賃機等を代替。
b社	【方策A】	バス 76 台	98 百万円	運賃機等を代替。
c社	【方策A】	バス 354 台	313 百万円	運賃機等を一部改修。
d社	地域独自	バス 169 台	280 百万円	運賃機等を代替。地域独自カードのシステムを含む。
e社	地域独自	バス 129 台	261 百万円	運賃機等を代替。地域独自カードのシステムを含む。
f社	【方策B】	バス 137 台	138 百万円	運賃機等を一部改修。10 カードへの接続システムは既存のため費用発生せず。
熊本県バス協会等	【方策B】	バス 約 1,000 台 私鉄 (18 駅、8 編成)	約 800 百万円	運賃機等を一部改修：約 600 百万円 10 カードへの接続システム：約 200 百万円

注：a 社～f 社は、交通系 IC カードの新規導入又は大規模更新を行った案件のうち主なものを掲げている。平成 23(2011)年～平成 26(2014)年の地域公共交通確保維持改善事業による導入支援実績より。なお、熊本県バス協会等に対しても、地域公共交通確保維持改善事業により、一部導入支援を行っている。

資料：国土交通省総合政策局作成

Figure 2.8: 交通系 IC カードの導入費用の例（出典：国土交通省「交通系 IC カードの普及・利便性拡大に向けた検討会 とりまとめ」図表 12[3]）

バス OD データには、①OD 間の経路が位置に定まる、②上流（始点）から下流（終点）への一方向流、③同一バス停への移動（内々トリップ）が生じないという構造上の特性がある。このため、バスの OD 表は路線のバス停順序に沿って上三角構造をとる。集計対象としては、発生・集中交通量に相当しバス停ごとの出入りを数える「バス停別乗降者数」と、分布交通量に相当し乗車バス停と降車バス停の組を数える「バス停別乗降者数」の 2 種類がある。前者は利用状況の把握に、降車は区間需要の把握に対応する。

OD データの収集手段としては、調査票を用いたパーソントリップ調査や調査員派遣による人力カウント、交通系 IC カード履歴、BLE/Wi-Fi や赤外線・圧力センサ、およびカメラ画像が挙げられる。PT 調査や調査員方式は設計自由度が高い一方、費用・工数面の負担が大きい。IC カードは精度と自動収集の面で有利だが、発行主体（多くは鉄道事業者）とのデータアクセスの調整や、機器導入コスト・地域普及率がボトルネックとなるケースが報告されている。こうした状況に対し、既設の車内カメラを活用する手法は、低コストでスケールしやすい代替案として位置づけられている。

バス OD データの応用先としては、運行のダイヤ調整や配車計画、路線の再編・最適化、運賃体系の見直し、混雑把握や乗継評価などがある。運用の時間軸は、当月までのバス乗客 OD データから翌々月の運行改善を行うものとし、原則として季節ごとの需要変動（通学期・降雪期・観光期等）に合わせて見直す。これにより、バス停別乗降者数は停留所配置や便数配分の再評価に、乗車区間別乗降者数は区間需要に基づく増減便や区間別運賃設計に直結する。一方で、工事・イベント・災害等の臨時要因に対しては、臨時改正や軽微変更の枠を用いた短期対応を行う想定で

図7 長崎県バス各社における導入費用

単位：千円)		
	単価	合計
車載機器 一式	493	723,352
降車口カードリーダー（積増機能付）	163	233,579
操作盤 含：メモリーカード・運賃表示器	100	143,300
乗車口カードリーダー	113	190,518
その他	117	155,955
販売窓口関連機器 一式	870	41,760
窓口用リーダーライター	370	17,760
窓口用端末	500	24,000
営業所用関連機器一式	600	18,700
データコンバーター（開錠機）	100	3,200
営業所用端末	500	15,500
本社関連危機 一式	0	10,500
カードエンコーダー	800	4,000
本社用端末（共通精算機能付）	1,300	6,500
ソフトウェア関連 一式		75,248
ICカードシステムソフト		23,844
共通精算ソフト		6,000
定期券発行システムソフト		45,404
合計		
当初合計		824,630

Figure 2.9: 長崎県バス各社における導入費用（出典：国土交通省「月例経済報告 関係資料」図7「長崎県バス各社における導入費用」[4]）

ある。

2.2 Computer Vision に関する研究分野

本節では、本研究を支える Computer Vision（CV）技術の位置づけと、その中でも特に関連の深い人物計測および人物追跡に関する研究動向を整理する。

乗車o \ 降車d	ゾーン1	ゾーン2	...	ゾーンZ	合計
ゾーン1					tg _o
ゾーン2					
⋮					
ゾーンZ					
合計		ta _d			T

Figure 2.10: OD データの図

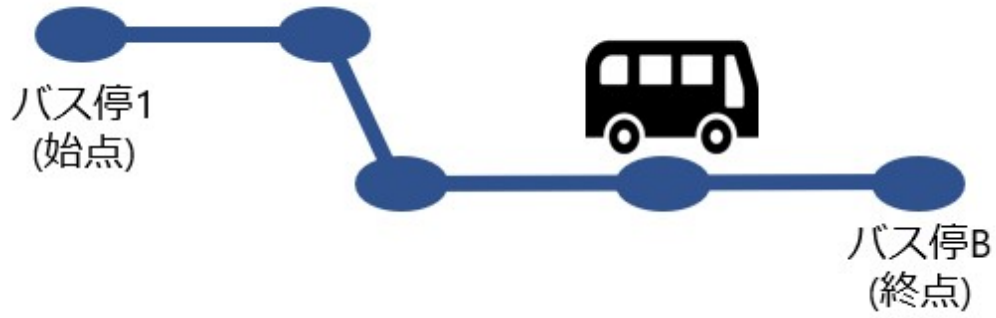


Figure 2.11: バスの経路

乗車 _o \ 降車 _d	バス停 1	バス停 2	...	バス停 B-1	バス停 B	合計
バス停 1						
バス停 2				td_{od}		tg_o
...						
バス停 B-1						
バス停 B						
合計			ta_d			T

Figure 2.12: バス OD データの図

2.2.1 Computer Vision の基礎と代表タスク

近年の Computer Vision (CV) は、深層学習，とりわけ畳み込みニューラルネットワーク (CNN) [24] や Transformer[25] 系モデルの発展により，画像・映像から多様な情報を自動抽出できるようになった．従来は SIFT[26] や HOG[27] といった手工的特徴量と SVM[28] などの識別器を組み合わせる手法が主流であったが，現在は VGG[29] や ResNet[30] などの CNN，Vision Transformer (ViT) [31] に代表される自己注意機構ベースのモデルが高精度な特徴抽出の基盤となっている．こうした汎用特徴抽出器の上に，タスク固有のヘッドを載せることで，多様な CV タスクが定式化されている．本研究で扱う路線バスにおける人物計測も，これらの要素技術を組み合わせた応用の一つである．

画像分類 (Image Classification) 画像分類は，入力画像全体に対して一つ以上のクラスラベルを付与するタスクであり，CV 分野の基盤タスクとして位置付けられる．ImageNet[32] の大規模画像データセット上で VGG[29] や ResNet[30] が高精度を達成したことにより，深層学習が CV の主流となった．応用例としては，医用画像診断（病変の有無判定），交通標識認識，自動検査ラインにおける不良品検出，監視映像からの行動カテゴリー判定などがある．本研究では画像分類を直接用いないが，人物検出や再識別に用いる CNN バックボーンは，画像分類タスクで事前学習されたパラメータを初期値として利用している．

物体検出 (Object Detection) 物体検出は、画像中の物体の位置（バウンディングボックス）とクラスを同時に推定するタスクである。代表的なモデルとして、領域提案に基づく Faster R-CNN[33]、単段検出の YOLO シリーズ [34] や SSD[35] が挙げられる。自動運転における車両・歩行者検出、工場・倉庫の自動ピッキング、監視カメラによる不審物検出、小売店舗における棚上商品の認識など幅広い応用がある。路線バスにおける人物計測では、乗車口・降車口付近の人物領域を安定して抽出することが最初のステップであり、人物検出器の精度がその後の追跡・乗降判定・OD 推定の土台となる。

セマンティック／インスタンス・セグメンテーション セマンティックセグメンテーションは、各画素にクラスラベルを割り当てるタスクであり、FCN[36] や U-Net[37] などが代表的である。インスタンスセグメンテーションは、物体毎に画素レベルの領域を分離するタスクであり、Mask R-CNN[38] が広く用いられている。これらは自動運転のシーン理解（歩道・車線・障害物の識別）、農業における作物と雑草の識別、医療画像における臓器・腫瘍領域の抽出などで実用化されている。路線バスの文脈では、車内の床面・手すり・座席と人物領域を分離することで、乗降動作や立ち位置の解析に応用できるが、本研究では計算コストの観点から採用せず、矩形ベースの人物検出と追跡を主要要素技術として用いる。

骨格推定 (Pose Estimation) 骨格推定は、人体の関節位置（キーポイント）を 2 次元あるいは 3 次元で推定するタスクである。OpenPose[39] や HRNet[40] などが代表的であり、スポーツ動作解析、リハビリ支援、人間工学的評価（作業姿勢のリスク評価）、ジェスチャ認識インタフェースなどで活用されている。バス車内の応用としては、転倒や急ブレーキ時の危険姿勢検知、高齢者・子どもの乗降支援行動の分析などが考えられるが、本研究ではカメラ設置位置とオクルージョンの制約から、骨格レベルの解析までは行わず、人物矩形と軌跡レベルの情報に焦点を当てる。

行動認識 (Action Recognition) 行動認識は、動画中の人物や群集の時空間パターンを解析し、「歩行」「走行」「乗車」「乗り降り」などの行動ラベルを付与するタスクである。3D CNN[41] や Two-Stream ネットワーク [42]、近年では Video Transformer[43] を用いた手法が提案されており、監視映像における異常行動検知、スポーツ映像の自動解析、製造ラインでの作業手順遵守チェックなどに利用されている。路線バス分野においては、乗降動作や危険行動（車内での走行、ドア付近での押し合い）を自動検知することで、安全対策やサービス品質向上に役立つ可能性がある。本研究では行動ラベル自体は扱わないが、将来的な拡張として、乗降イベント検出との統合が考えられる。

人数推定・密度推定 (Crowd Counting / Density Estimation) 人数推定は、群集シーンにおける人物数や密度分布を推定するタスクであり、頭部検出や密度マップ回帰に基づく手法が多い。大規模イベント会場や駅構内での混雑度モニタリング、商業施設の来場者数統計、避難計画の評価などで導入が進んでいる。路線バスにおいても、車内の乗車人数や混雑度をリアルタイムに推定することで、運行ダイヤや車両配備計画の高度化が期待できる。本研究が扱う乗降者数推定は、各バス停の入退出イベントを積み上げることで、区間毎の負荷や混雑度を把握するという点で、人数推定タスクと密接に関連している。

人物検出・人物再識別・人物追跡 (Person Detection / Re-Identification / Multiple Object Tracking; MOT) 人物検出は、画像中から歩行者など人物クラスに属する物体領域を抽出するタスクであり、物体検出の中でも交通監視や自動運転における歩行者検出などで重要な役割を果たしている。人物再識別 (Person Re-ID) は、異なるカメラ間や離れた時刻に撮影された人物画像のペアについて、同一人物かどうかを照合するタスクであり、Market-1501 や DukeMTMC-ReID といったベンチマーク上で、BoT ベースライン [44] のようにバックボーン CNN に識別学習とメトリックラーニングを組み合わせた手法が高い性能を示している。これらを組み合わせた人物追跡 (MOT) は、各フレームで得られた人物検出結果に対し、外観特徴 (Re-ID 特徴量) や運動の整合性に基づいてフレーム間の対応付けを行い、個々の人物に一貫した ID を付与して軌跡を復元するタスクである。Tracking-by-Detection 型の手法としては、検出結果に Kalman フィルタと外観特徴を組み合わせた DeepSORT、検出と Re-ID を単一ネットワークで同時学習する FairMOT、低スコア検出も活用して対応付けの頑健性を高めた ByteTrack などが代表的である [45, 46, 47]。応用例としては、監視カメラ映像における迷子・不審者の追跡や入退館履歴の紐付け、大型商業施設や駅構内での回遊行動解析、交差点における歩行者安全評価、工場・倉庫内での作業員動線の分析、公共交通機関における乗降計数・人流解析などが挙げられる。本研究の路線バスにおける乗車区間別乗降者数推定では、乗車口と降車口の動画に対して人物検出と人物追跡を適用し、さらに乗車動画と降車動画間で Re-ID 特徴量とメトリックラーニングを用いて同一人物を照合することで、MOT 系手法を OD 推定に応用している。

以上のように、CV の各タスクは入力・出力の粒度や時間的スケールが異なり、用途に応じて適切な組み合わせを設計することが重要である。本論文では、特に物体検出・人物追跡・人物再識別および人数推定に関連する技術を組み合わせることで、路線バスにおける乗降映像からバス乗客 OD データを推定する枠組みを構築する。

2.2.2 人物計測

本論文における人物計測とは、画像・映像および必要に応じて幾何・センサ情報を用い、空間内の人物に関する状態量（人数、位置、軌跡、姿勢、属性、行動、入退場イベント等）を推定して、運用・計画の意思決定に資する知見を得るための技術の総称である。用途例として、施設における入退出管理や勤怠管理、小売店舗における顧客行動分析（滞在・回遊・棚前滞留・レジ行列）、動線・人流解析（流量・混雑度・滞留時間・OD）、安全監視（立入・侵入・転倒・危険行動の検知）、イベントや避難時の群集挙動把握、オフィスや商業施設の占有率推定、公共交通の乗降計数や混雑度把握などが挙げられる [48, 49, 50, 51]。近年はこれらの計測結果をデジタルツインやシミュレーションに接続し、レイアウト改善や運行計画の最適化に用いる取り組みも進んでいる。

人物計測に用いられるセンサ方式は多様であり、大きく非画像系と画像系に分けられる。非画像系には赤外線遮断型（IR ビーム）や受動赤外線（PIR）、圧力マット・荷重センサ、超音波センサ、ToF／ステレオ深度センサ、LiDAR、ミリ波レーダ、CO₂ 濃度センサ、RFID／NFC、BLE／Wi-Fi プローブ等があり、画像系には可視域カメラや熱赤外カメラがある [52, 53]。IR ビームや PIR センサは出入口に設置して通過人数のみを計測する用途で広く利用され、圧力マットはゲートや改札・店舗の出入口などで歩行者を検知するのに適している。Wi-Fi や BLE プローブはスマートフォン等の端末から発せられる信号を観測することで、建物全体の在室者数や人流の大まかな傾向を推定する手法に用いられている。CO₂ センサは室内環境の制御と合わせて在室者数の推定にも応用され、オフィスや教室などでの長時間の占有率推定に用いられる [52]。

どのセンサ方式を採用するかは、対象が登録済みの既知個体か不特定多数かという前提、出力の粒度（単純カウント／位置・軌跡／同一性）、設置・保守コストと電源・通信要件、混雑時の遮蔽に対する強さ、照度や温度、電波環境といった周辺環境への感度、さらにプライバシーや法令順守の観点などを総合して決まる。非画像系センサは構成が簡素で単価も低く、顔画像を扱わないことからプライバシー上のリスクが小さいという利点を持つ一方で、個体の連続性や同一性に関する情報がほとんど得られない。このため、区間 OD の再構成や属性別（性別・年齢層別など）の詳細分析には限界がある。RFID／NFC はタグ携行者については高精度な追跡が可能であるが、利用者側の能動的協力（カード携行やタッチ操作）を前提としており、不特定多数が乗降する公共交通機関やオープンな施設では適用範囲が限定される。BLE／Wi-Fi も端末の保有率や OS 仕様に依存した標本の偏りが避けられず、母集団全体の人数や属性を厳密に推定することは難しい [53]。

これに対して、可視カメラに基づく人物計測は、既設の監視カメラ資産を活用しやすく、人物検出・追跡・再識別・入退出判定を組み合わせることで、詳細な時

空間データを取得できる点に特徴がある。深層学習に基づく物体検出・人物追跡・人物再識別の発展により、外観の変化や複雑な背景環境下においても高精度な人物計測が可能となり、駅ホームや商業施設、空港、スタジアムなどの混雑環境における入退出計測や群集カウント、公共交通の乗降計数といった応用が報告されている [48, 50, 51, 49]。一方で、カメラ方式ではオクルージョンや照度変化への頑健性、顔や体の画像を扱うことに伴うプライバシー保護といった課題が残る。これに対処するため、顔領域へのモザイク処理や低解像度化、エッジデバイス内での匿名化処理、プライバシーを考慮した特徴量のみをサーバに送信する構成などが提案されている。

CV を用いた入退出 (Entry-Exit) 計測は、監視・小売・交通など多様な領域で研究が行われている [54, 55, 49]。典型的な枠組みは、人物検出と多目標追跡 (MOT) によりフレーム間の軌跡を推定し、その軌跡に対して空間的ルール (ドア周辺に設定したゾーンや仮想トリップワイヤ) と時間的ルール (ドア開閉状態、先頭・末尾フレーム時刻、滞在時間しきい値など) を適用して入場・退出イベントを同定するものである。この枠組みは単眼カメラ・複数カメラのどちらにも適用され、ホモグラフィにより画像座標を床面へ射影してゾーン交差を安定化する方法、ドア状態の認識やイベント検出を状態遷移として統合し誤検出を抑える方法、強い遮蔽や群集時の断片化に対処するため再識別 (Re-ID) や外観類似度を用いて軌跡の連続性を補完する方法などが提案されている [48, 50]。本研究が対象とする公共交通の乗降計数・OD 推定は、この系譜に属する応用であるが、路線バス車内は通路が狭く遮蔽が多いことに加え、扉の開閉に伴う急激な照度変化が生じるなど、中間的な環境特性を持つ。このため、前段の人物追跡の品質に強く依存し、ID スイッチやトラック断片化が入退出判定の主要な誤り要因となり得る。次節では、これらの課題に直接関わる人物追跡技術について概説する。

2.2.3 人物追跡

人物追跡 (Multiple Object Tracking; MOT) とは、フレーム毎に推定された人物矩形を時間方向に対応付け、個々の人物に一貫した ID を付与して軌跡を復元する処理である。実装上は、まず人物検出で各フレームの矩形を抽出し、ついで外観 (Re-ID 特徴) や運動の整合性に基づくデータアソシエーションでフレーム間を連結する二段構成 (Tracking-by-Detection) が広く用いられる。

研究の潮流としては、MOTChallenge に代表されるベンチマークの整備を背景に (MOTA, IDF1 などの共通指標が普及)、まず Tracking-by-Detection (TBD) 系が主流となり、その後、検出と識別を単一ネットワークで同時学習する試み (例: FairMOT) が登場した。さらに、前フレーム情報を入力して同時検出・追跡を行う方式 (例: CenterTrack)、低スコア検出も活用してアソシエーションを強化する方式 (例: ByteTrack) が精度と頑健性を押し上げた。近年は Transformer を用いたエンドツーエンド (E2E) 型 (例: TrackFormer / TransTrack / MOTR) が台

頭し、後処理に依存しない一体型学習を志向している。これらの系譜は、オンライン／オフラインや速度・ID 一貫性のトレードオフを異にし、環境や要件に応じた選択が求められる。

環境特性との関係では、屋内固定視点は背景・幾何の安定性を活かせるのに対し、屋外は照度・天候変動やスケール変化、群集による強遮蔽が主要課題となる。路線バスは定義上は屋内であるが、走行に伴う揺動・ブラー、扉開閉に起因する照度急変／人流の突発的流入出、狭い通路における高密度遮蔽など屋外的・動的要因を併せ持つ中間環境であり、追跡品質（特に ID スイッチ と断片化）への影響が大きい。

第3章 バスOD推定システムの全体設計

3.1 はじめに

本章では、路線バスに設置したカメラで取得した利用者の乗降映像からバス OD データを推定するシステムの全体設計を行う。対象とする出力はバス停別乗降者数と乗車区間別乗降者数であり、当月までのバス乗客 OD データから翌々月の運行見直しに資する粒度での推定を目的とする。

本章の焦点は、個々の手法の改良ではなく、データ取得から前処理、モジュール間の入出力設計、最終的な OD 集計と評価までを一貫したパイプラインとして定義することで、再現性と実装可能性を担保する設計原理を明確化することにある。これにより、第2章で定めた前提の下で、OD 推定の対象・粒度・評価観点をシステムとして整合的に扱えるようにする。

本章では、3.2 節で実際に運行している路線バスで収集した動画データの概要と分析を示す。3.3 節では、バス OD データ推定システムとして BAIRD-OD を提案する。3.4 節では提案システムの有効性に関する研究成果を示す。3.5 節では、本章のまとめを示す。

3.2 路線バス利用者の動画データ

本節では、システムの検証に必要な路線バス利用者の動画データの取得と、その動画を分析した結果について示す。

3.2.1 動画取得

札幌市内を実際に運行している路線バスにて動画データを収集した。車載器やカメラ等の機器は株式会社シーズ・ラボ、路線バスは株式会社じょうてつにご協力いただいた。図 3.1 のように、南 4（黒：真駒内駅→市立病院前）、南 95（赤：真駒内駅→硬石山）、南 96（青：真駒内駅→中の沢）の 3 路線で収集した。収集時期は 2021 年 12 月 5 日～6 日の 7 時台～22 時台である。冬の札幌であるため利用者は図

3.2 のようにコートや帽子・マフラー等を身に着けており、コロナ禍につきマスクを着用している。カメラは図 3.3 のように自動車車体規格のガイドラインによるドライブレコーダ設置位置 [5] を参考にし、図 3.4 のように設置した。乗車口の動画に関して、始点のバス停あるいは人が待っているバス停で乗車口のドアが開き、動画を撮影する。降車口の動画に関して、終点のバス停あるいはバス車内の人物が降車ボタンを押したバス停で降車口のドアが開き、動画を撮影する。撮影はドアが開く 10 秒前から開始され、ドアが閉まってから 10 秒後に終了する。個人情報の取り扱いの観点から、バス車内に動画を撮影していることを示す張り紙をし、撮影と同時に顔にモザイク処理を施してから動画を保存した。各動画の解像度は乗車口で縦 882 px・横 886 px, 降車口で縦 1,024 px・横 768 px, フレームレートは 4 fps である。

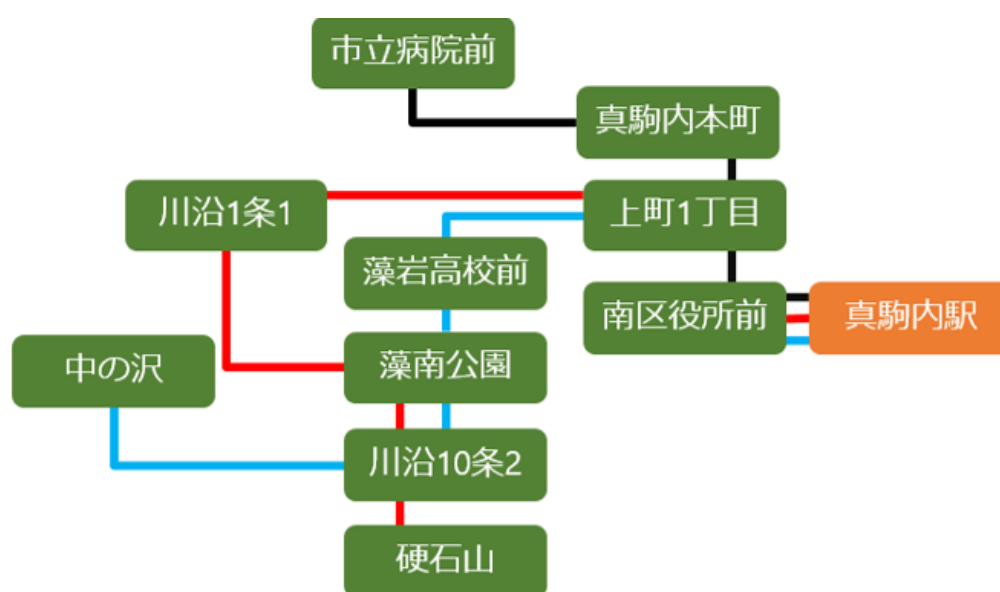


Figure 3.1: 動画データを収集した路線図



Figure 3.2: 収集した動画の例

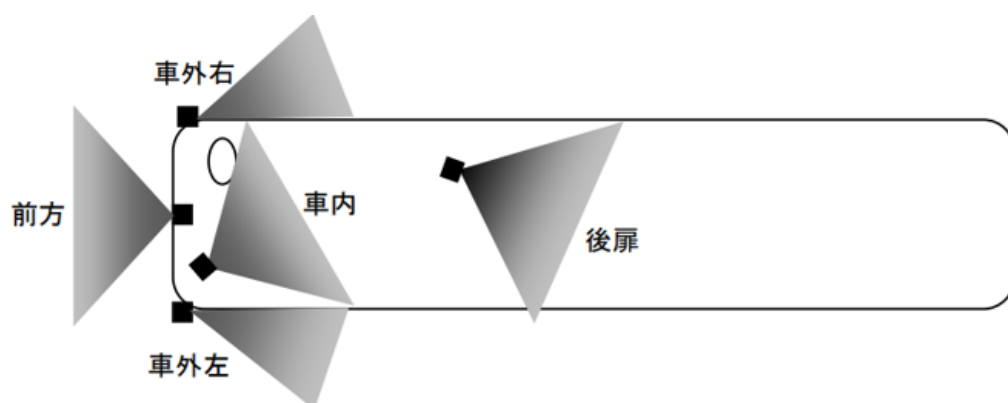


Figure 3.3: 自動車車体規格のガイドラインによるドライブレコーダ設置位置 [5]

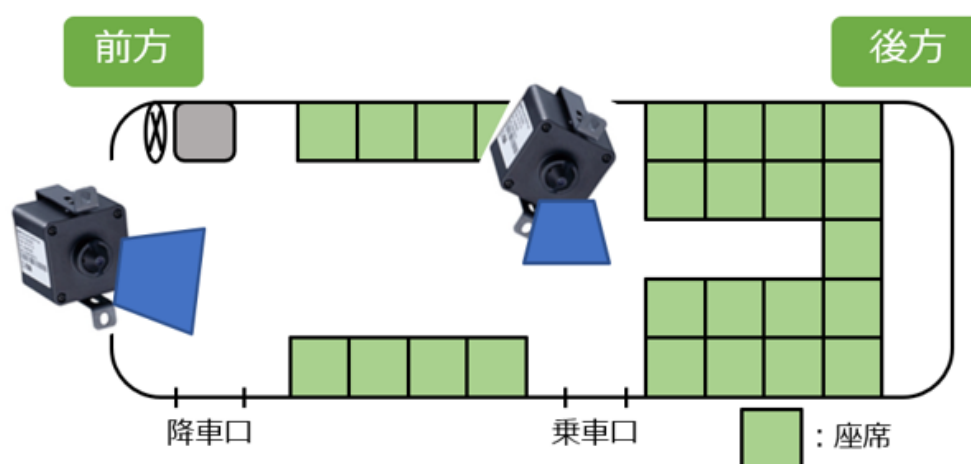


Figure 3.4: 乗車口と降車口に設置したカメラ位置

3.2.2 取得動画の分析

取得した動画を確認し分析したところ、CV タスクとして難しい点が 3 点確認された。

照明条件 バスの外側で強い光が当たっている場所では、図 3.5 のようにバスに乗車する直前で白飛びが発生してしまう。その後バスに乗車すると明るい車外と暗い車内のギャップで、逆光による黒つぶれが発生してしまう。このように厳しい照明条件下では、人物の形や色など外観情報が欠落してしまい、CV タスクとしての難易度が上がる。

オクルージョン 路線バス車内ではオクルージョンが多発し、人物の姿が完全に映らない状況が多い。特に乗車・降車に際して人物が流入出するときに発生し、図 3.6 のように人物同士や手すり等の車内オブジェクトとの間にオクルージョンが発生する。乗車口では続けて乗車してくる人物同士や、降車するためにバス後方から

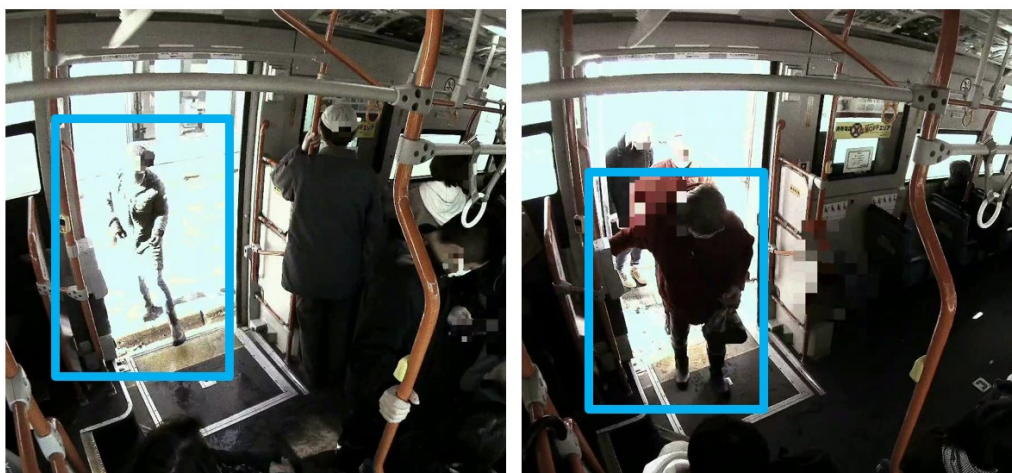


Figure 3.5: 白飛び，黒つぶれが発生した人物

前方に移動する人物の間で発生する．降車口では通路が狭いため，料金箱付近で並んでいる人物の間で発生する．



Figure 3.6: オクルージョンの発生

服装 コロナ禍の冬の札幌でデータ収集を行ったため、利用者の多くはコート、帽子、マフラー、マスクを着用している。図 3.7 のように、特に黒やグレー、ブラウンなど暗い色の服装をしている利用者が多く、暗い車内では人物の形を認識するのが難しい。また、似たような服装の人物が多く乗車口の動画と降車口の動画で同一人物を判定するタスクも、他の Re-ID タスクの環境と比較して難易度が高い。



Figure 3.7: 暗い色・類似した服装の利用者

3.3 提案システム：BAIRD-OD

提案システムの BAIRD-OD (Boarding-Alighting Identification, Re-identification and Data Association for OD Estimation) について、概要を図 3.8 に示す。BAIRD-OD は、①動画取得、②個人識別、③乗降判定、④同一判定の 4 つのプロセスから構成される。

①動画取得 本システムは路線バスの既設カメラから動画を取得し、バス乗客 OD データを推定することで運行の効率化を図る状況を想定している。そのため、リアルタイムでの OD 推定が不要であり、運行後のサーバでのバッチ処理が望ましい。各バス停で利用者の映像を撮影し、プライバシー保護の観点から顔にモザイク処理を施した後、車載器に動画を保存する。

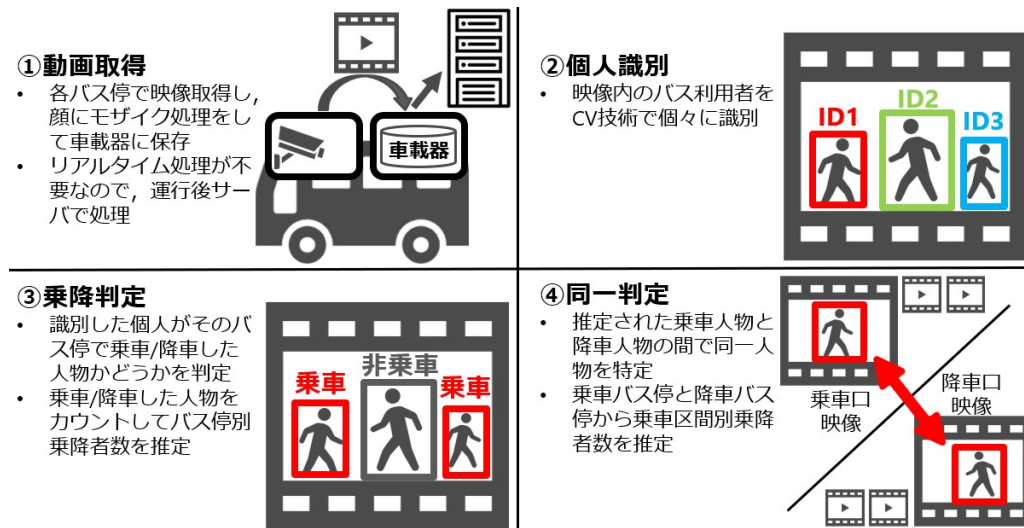


Figure 3.8: 提案システム：BAIRD-OD

②個人識別 取得した動画内に映っている人物を、CV 技術を用いて個々に識別していく。

③乗降判定 動画には様々な人物が映っており、個人識別の段階では区別ができていない。本研究では撮影した動画内で実際に乗車または降車した人物を対象人物、それ以外のバスに乗車しているものの当該バス停では乗降しない人物やバス車外の通行人物などを不適人物と呼称する。ここでは識別された個人からそのバス停で乗車または降車した対象人物を判定する。対象人物の人数をカウントし、そのバス停におけるバス停別乗降者数を推定する。

④同一判定 推定された乗車人物と降車人物の間で同一人物を特定することで、各利用者の乗車バス停と降車バス停を特定する。バス停間の移動をカウントし、乗車区間別乗降者数を推定する。

3.3.1 要素技術の検討

本項では提案システム BAIRD-OD の各処理において用いる要素技術について検討する。

②個人識別 まず、画像内の物体の位置とクラスを特定する第一歩の技術として、物体検出が存在する。物体検出はその派生技術として、各タスクに特化した人物検出や頭部検出なども存在する。頭部検出は検出の対象となる部分が小さく、オクルージョンが多発するバス車内では不適である。そのため、本研究ではオクルージョンで体の一部が映らなくても検出できる人物検出を採用する。その後、個人を識別する処理としては顔認識、骨格推定、人物追跡などが存在する。本システムで

は顔にモザイク処理を施すため顔認識は使用できず、オクルージョンが多発するバス車内では全身が映りにくいため骨格推定も難しい。そのため、本研究では人物追跡を採用し、人物矩形を入力としてフレームをまたいだ同一人物に同一 ID を割り当てた、軌跡として出力する。

③乗降判定 個人識別により出力された軌跡に対して乗降判定を行う手法として、ルールベースと機械学習が考えられる。工数や計算資源の面ではルールベースが有効であり、推定精度の面では機械学習が有効である。地域や季節ごとにチューニングが求められる本研究の想定状況では、ルールベースでの乗降判定が最適と判断した。

④同一判定 乗車人物の軌跡と降車人物の軌跡から同一人物を判定する、という本タスクは乗車人物集合と降車人物集合の二部グラフのマッチングとして捉えることができる。重みとしては軌跡同士の外観特徴の特徴量距離を使用することができるため、メトリックラーニングを適用することができる。メトリックラーニングは、同一人物の画像を特徴量距離が小さくなるようにエンベディングしてベクトル化する技術である。

3.3.2 提案アルゴリズム

①動画取得 図 3.9 に動画取得の概要を示す。バス停 bs_β にて利用者 pa_ρ が乗降する動画 $MO_{\beta,g}$ を取得する。 g は扉を表し、 $1 \leq \beta \leq B$, $1 \leq \rho \leq P$, $g = bo, al$ である。動画 $MO_{\beta,g}$ は $K_{\beta,g}$ 枚の画像 $im_{\beta,g,\kappa}$ から構成されており、 κ はフレーム番号を表し、 $1 \leq \kappa \leq K_{\beta,g}$ である。

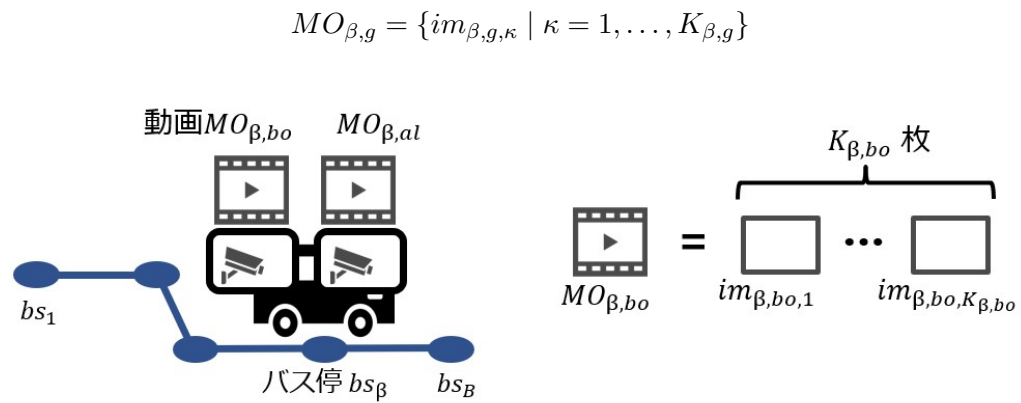


Figure 3.9: 動画取得の概略図

利用者の乗車バス停を取得する関数を $Gb()$ 、降車バス停を取得する関数を $Ga()$ とする。

$$Gb(pa_\rho) = bs_\beta$$

$$Ga(pa_\rho) = bs_\beta$$

発生交通量 tg_δ , 集中交通量 ta_ε , 分布交通量 $td_{\delta,\varepsilon}$ は以下のようなになる.

$$tg_\delta = |\{pa_\rho \mid Gb(pa_\rho) = bs_\delta\}|$$

$$ta_\varepsilon = |\{pa_\rho \mid Ga(pa_\rho) = bs_\varepsilon\}|$$

$$td_{\delta,\varepsilon} = |\{pa_\rho \mid Gb(pa_\rho) = bs_\delta \wedge Ga(pa_\rho) = bs_\varepsilon\}|$$

②個人識別 図 3.10 に人物検出の概要を示す. 人物検出器 $De()$ に画像 $im_{\beta,g,\kappa}$ を入力して矩形群 $BB_{\beta,g,\kappa,d}$ を出力する. 矩形群 $BB_{\beta,g,\kappa,d}$ には $A_{\beta,g,\kappa,d}$ 個の矩形 $bb_{\beta,g,\kappa,d,\alpha}$ を含んでおり, d は人物検出器, α はその画像内の矩形 ID を表し, $1 \leq \alpha \leq A_{\beta,g,\kappa,d}$ である. ここで矩形 $bb_{\beta,g,\kappa,d,\alpha}$ は, 幾何情報 (矩形の中心座標, 幅, 高さ) と外観情報 (人物画像のピクセル値) を持っている.

$$De(im_{\beta,g,\kappa}) = BB_{\beta,g,\kappa,d}$$

$$BB_{\beta,g,\kappa,d} = \{bb_{\beta,g,\kappa,d,\alpha} \mid \alpha = 1, \dots, A_{\beta,g,\kappa,d}\}$$

図 3.11 に人物追跡の概要を示す. 人物追跡器 $Tr()$ に動画内の矩形をすべて入力して ID 付き矩形群 $BI_{\beta,g,d,t}$ を出力する. ID 付き矩形群 $BI_{\beta,g,d,t}$ にはその動画内の全フレーム分 $\sum_{\kappa=1}^{K_{\beta,g}} A_{\beta,g,\kappa,d}$ 個の ID 付き矩形 $bi_{\beta,g,\kappa,d,t,\gamma}$ を含んでおり, t は人物追跡器, γ は軌跡 ID を表し, $1 \leq \gamma \leq \Gamma_{\beta,g,d,t}$ である. ここで, $\mathcal{K}_{\beta,g,d,t,\gamma}$ は軌跡 $TR_{\beta,g,d,t,\gamma}$ が存在するフレーム番号の集合とする.

$$Tr\left(\bigcup_{\kappa=1}^{K_{\beta,g}} \{bb_{\beta,g,\kappa,d,\alpha} \mid \alpha = 1, \dots, A_{\beta,g,\kappa,d}\}\right) = BI_{\beta,g,d,t}$$

$$\mathcal{K}_{\beta,g,d,t,\gamma} \subseteq \{1, \dots, K_{\beta,g}\}$$

$$BI_{\beta,g,d,t} = \{bi_{\beta,g,\kappa,d,t,\gamma} \mid \gamma = 1, \dots, \Gamma_{\beta,g,d,t}, \kappa \in \mathcal{K}_{\beta,g,d,t,\gamma}\}$$

図 3.12 のように ID 付き矩形を同一軌跡 ID の集合にして軌跡 $TR_{\beta,g,d,t,\gamma}$ を作成する.

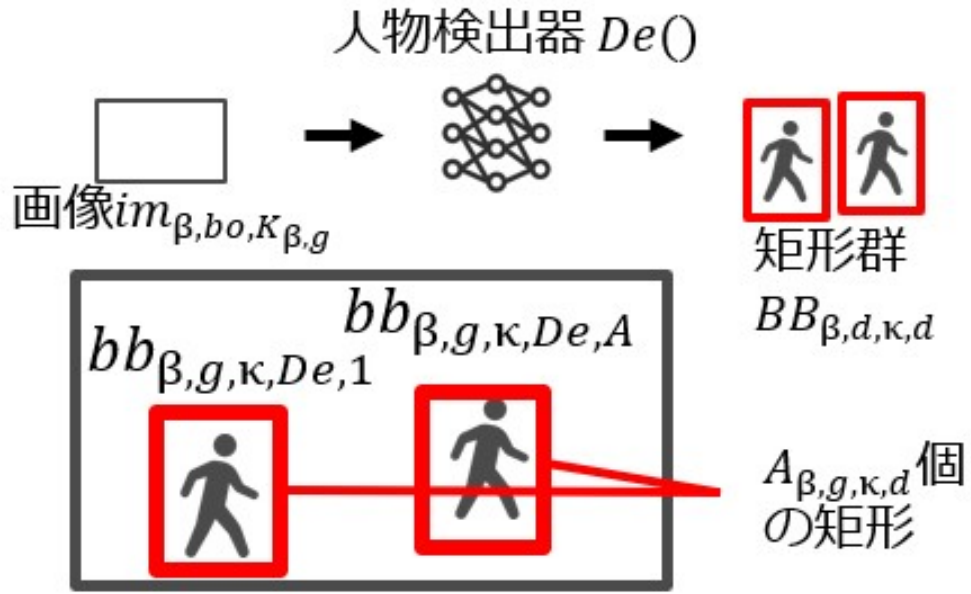


Figure 3.10: 人物検出の概略図

$$TR_{\beta,g,d,t,\gamma} = \{ bi_{\beta,g,\kappa,d,t,\gamma} \mid \kappa \in \mathcal{K}_{\beta,g,d,t,\gamma} \}$$

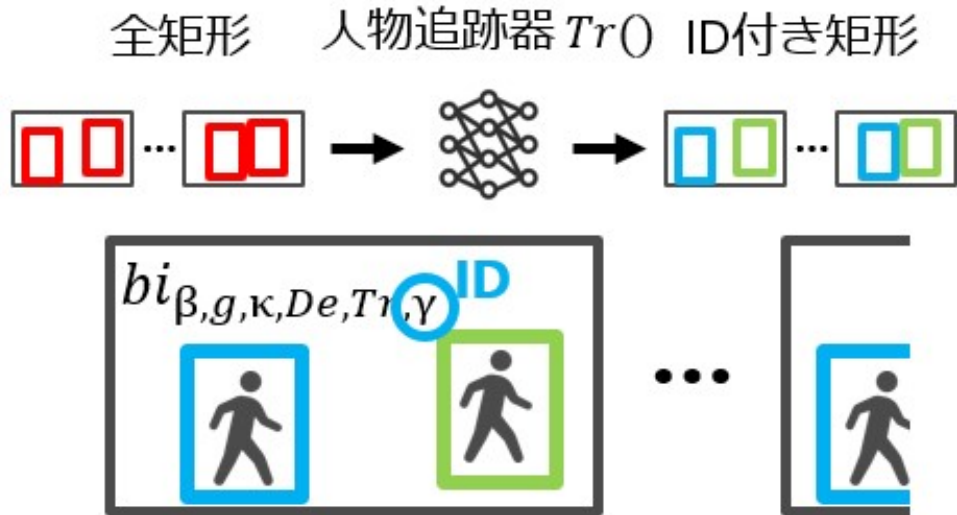


Figure 3.11: 人物追跡の概略図

③乗降判定 推定した軌跡ごとにその動画内で乗車または降車したかどうかを判定する。まずは図 3.13 乗車口と降車口で別々に画像内の扉範囲を矩形で設定し、そ

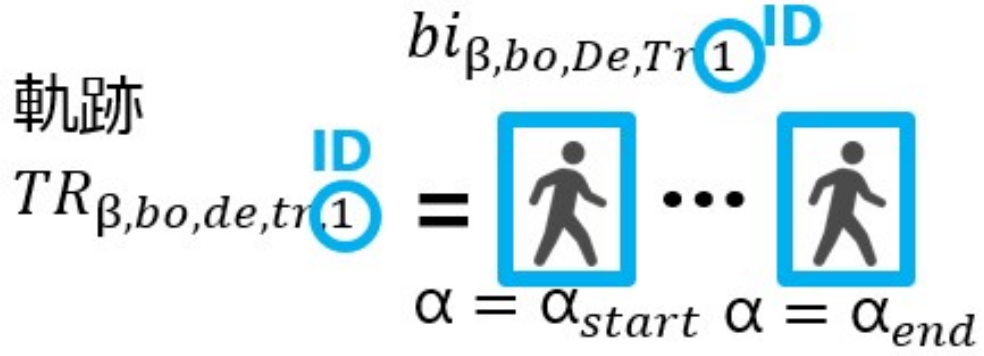


Figure 3.12: 人物軌跡

の辺の座標を乗車口では $bleft$, $bright$, $btotop$, $bbottom$, 降車口では $aleft$, $aright$, $atop$, $abottom$ とする. 人物矩形を入力として中心座標 x, y を取得する関数を $Gx()$, $Gy()$ とし, κ_{start} および κ_{end} はそれぞれ軌跡が観測される最初と最後のフレーム番号とする.

$$Gx(bi_{\beta,bo,\kappa_{start},d,t,\gamma}) = x_{bi}$$

$$Gy(bi_{\beta,bo,\kappa_{start},d,t,\gamma}) = y_{bi}$$

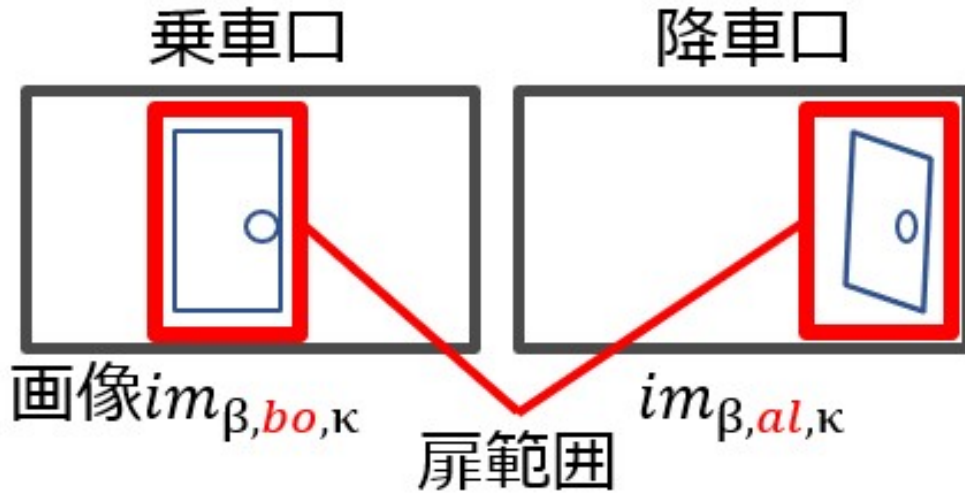


Figure 3.13: 扉範囲定義の概略

乗車判定器を $Jb()$ とし, 図 3.14 のように, 軌跡の最初の矩形の中心座標が扉範囲内にあれば, その人物を乗車人物と判定する.

$$Jb(TR_{\beta,bo,d,t,\gamma}) = \begin{cases} \text{True,} & bleft \leq x_{bi} \leq bright, \quad btop \leq y_{bi} \leq bbottom \\ \text{False,} & \text{otherwise.} \end{cases}$$

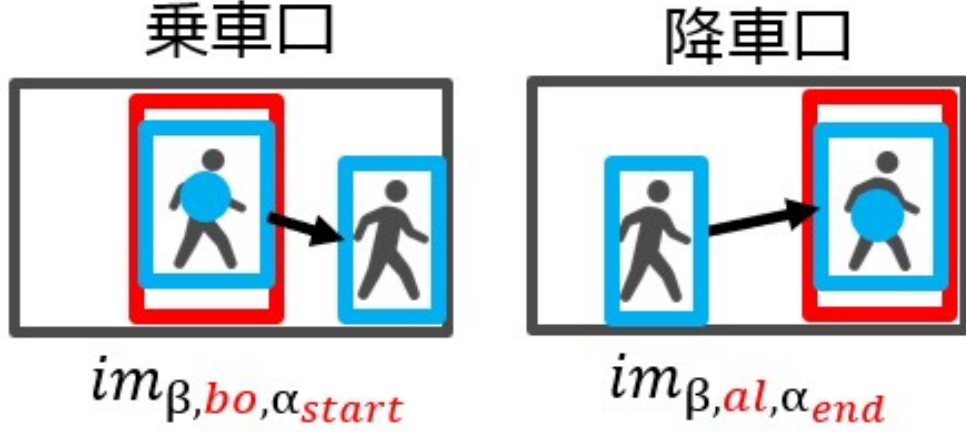


Figure 3.14: 乗降判定の概略

降車判定器を $Ja()$ とし、軌跡の最後の矩形の中心座標が扉範囲内にあれば、その人物を降車人物と判定する。

$$Gx(bi_{\beta,al,\kappa_{end},d,t,\gamma}) = x_{bi}$$

$$Gy(bi_{\beta,al,\kappa_{end},d,t,\gamma}) = y_{bi}$$

$$Ja(TR_{\beta,al,d,t,\gamma}) = \begin{cases} \text{True,} & aleft \leq x_{bi} \leq aright, \quad atop \leq y_{bi} \leq abottom \\ \text{False,} & \text{otherwise.} \end{cases}$$

乗車人物と判定された軌跡を BT とし、その集合を $BT_{s,d,t}$ 、降車人物と判定された軌跡を AT とし、その集合を $AT_{s,d,t}$ で表す。

$$BT_{s,d,t} = \{ TR_{\beta,bo,d,t,\gamma} \mid Jb(TR_{\beta,bo,d,t,\gamma}) \}$$

$$AT_{s,d,t} = \{ TR_{\beta,al,d,t,\gamma} \mid Ja(TR_{\beta,al,d,t,\gamma}) \}$$

推定発生交通量を $\hat{t}g_{\delta}$ 、推定集中交通量を $\hat{t}a_{\varepsilon}$ として、以下のように計算する。

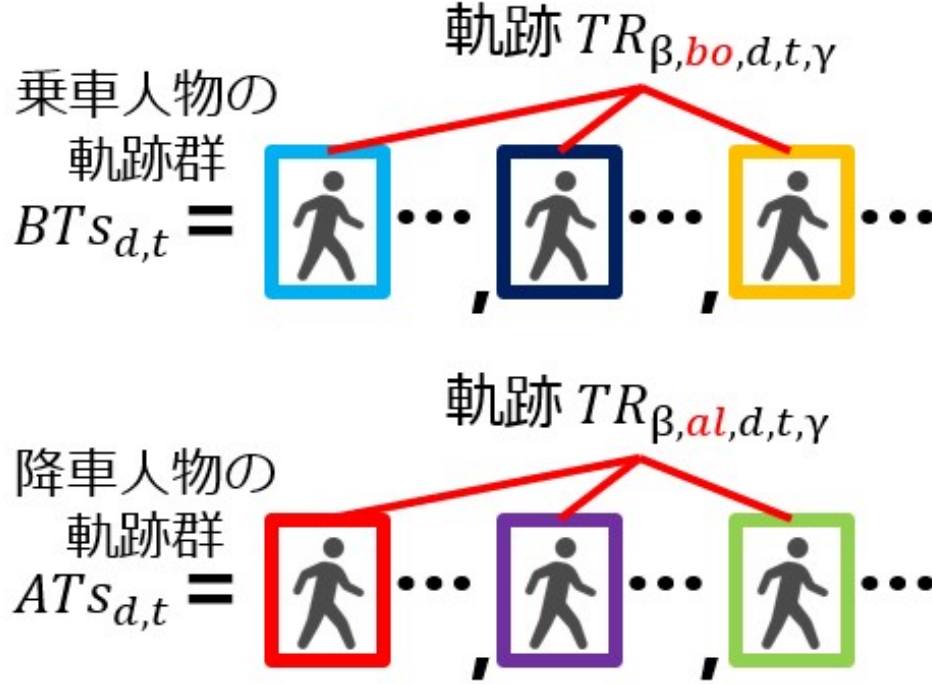


Figure 3.15: 乗車人物の軌跡と降車人物の軌跡

$$\hat{t}g_{\delta} = |\{TR_{\beta,bo,d,t,\gamma} \in BTs_{d,t} \mid \beta = \delta\}|$$

$$\hat{t}a_{\varepsilon} = |\{TR_{\beta,al,d,t,\gamma} \in ATs_{d,t} \mid \beta = \varepsilon\}|$$

④同一判定 推定した乗車人物の軌跡と降車人物の軌跡をペアリングするため，図 3.16 のように軌跡間特徴量距離行列を作成する．乗車人物の軌跡は $I(= |BTs_{d,t}|)$ 個，降車人物の軌跡は $J(= |ATs_{d,t}|)$ 個推定されるため， $I \times J$ の行列で表される．

$$BTs_{d,t} = \{BT_i \mid i = 1 \dots I\}$$

$$ATs_{d,t} = \{AT_j \mid j = 1 \dots J\}$$

乗車人物の軌跡 BT_i と降車人物の軌跡 AT_j を入力として，軌跡間特徴量距離 $ts_{i,j}$ を計算する関数 $Ts()$ を使用し，軌跡間特徴量距離 $ts_{i,j}$ を総当たりで計算していく．

$$Ts(BT_i, AT_j \mid BT_i \in BTs_{d,t}, AT_j \in ATs_{d,t}) = ts_{i,j}$$

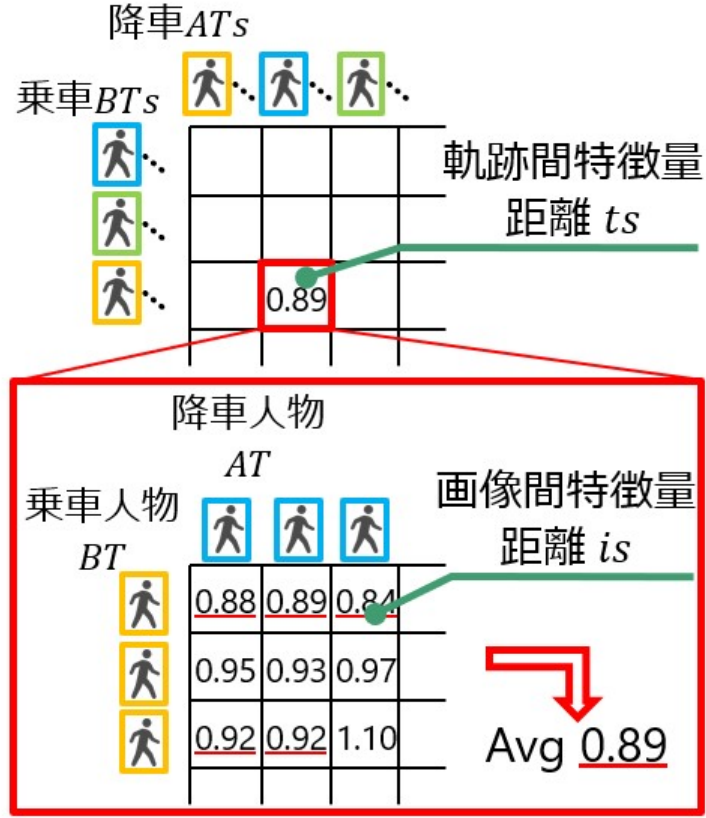


Figure 3.16: 軌跡間特徴量距離行列

軌跡間特徴量距離を計算する関数 $Ts()$ について詳細を説明する．乗車人物の軌跡 BT_i は $M_i(=|BT_i|)$ 枚の ID 付き矩形（人物画像），降車人物の軌跡 AT_j は $N_j(=|AT_j|)$ 枚の ID 付き矩形（人物画像）から構成されている．各軌跡とそれに含まれる人物画像 $bi_{\beta,g,\kappa,d,t,\gamma}$ は以下のように表せる．

$$BT_i = \{bi_m^i \mid m = 1, \dots, M_i\}$$

$$AT_j = \{bi_n^j \mid n = 1, \dots, N_j\}$$

メトリックラーニングを用いて画像間特徴量距離 $is_{m,n}$ を計算することで， $M_i \times N_j$ の画像間特徴量距離行列を作成できる．画像間特徴量距離 $is_{m,n}$ を計算する関数を $Is()$ として，総当たりで計算していく．

$$Is(bi_m^i, bi_n^j \mid bi_m^i \in BT_i, bi_n^j \in AT_j) = is_{m,n}$$

軌跡の中の画像枚数は軌跡によって大きく異なること，また，ひとつの軌跡の中でも光の当たり方や体の向きによっても特徴に違いがでるため，画像間特徴量距離 $is_{m,n}$ のうち上位 5 つを平均することで軌跡間特徴量距離 $ts_{i,j}$ とする．

図 3.17 にマッチングの概要を示す。バス OD データの特徴として、乗車人物のバス停よりも前のバス停で降車した人物はペアリングすることがないため、軌跡間特徴量距離 $ts_{i,j}$ を ∞ で置換する。最後に作成された軌跡間特徴量距離行列に対して割当アルゴリズムを適用し、マッチングした乗降ペアを作成する。



Figure 3.17: 乗車人物の軌跡と降車人物の軌跡のマッチング

3.4 提案システムの有効性の検証

この節では提案システム BAIRD-OD が、実際の路線バスで収集した動画データに対して有効に機能するかを検証する。

3.4.1 データセット

収集した動画データにアノテーション作業を行い、実験用のデータセットを作成する。すべてのフレームに対して人物矩形と人物 ID のアノテーションを行う。人物 ID は乗車動画と降車動画で同一人物に対して同一のものが振られる。その高いアノテーションコストから、以下の条件を満たす 2 路線の動画を選定し、それぞれ運行データ 1、運行データ 2 とする。データセットの詳細を表 3.1 に、使用する動画の例と扉範囲の設定を図 3.18 に示す。

- 極端に密集しすぎない中程度の人数 (30 人前後)

- 暗すぎない日中の時間帯
- 同じ制服を着た学生が大量に利用する通学路でない

Table 3.1: 第3章で用いるデータセットの概要

データ名	発車時刻	利用者数	総画像数	バス停数（乗車/降車）
運行データ 1	15:43 発	25 人	3,174 枚	7/12
運行データ 2	12:40 発	32 人	6,937 枚	12/15

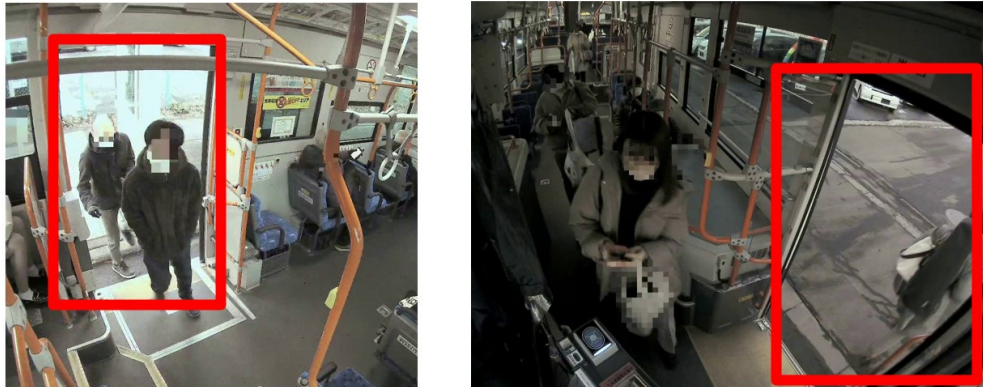


Figure 3.18: 扉範囲の設定（左：乗車口，右：降車口）

3.4.2 実験1：バス停別乗降者数推定

実験目的

提案システム BAIRD-OD の有効性を検証するため、バス停別乗降者数を推定する。これは個人識別と乗降判定という2つの処理の精度検証にあたる。

実験設定

人物検出器として YOLOR[56] を使用する。YOLOR は物体検出モデル YOLO[34] シリーズの CNN 系モデルで、矩形のクラスと同時にその確信度を出力する。モデルの学習データには MSCOCO2017[57] の物体検出を使用する。MSCOCO は Microsoft が作成した物体検出・セグメンテーション・画像キャプションなどのベンチマークデータセットで、その中の物体検出は 80 クラス、11.8 万枚の画像から構成されている。人物追跡器として MPNTrack[58] を使用する。MPNTrack は物体追跡のベンチマーク MOT Challenge[59] の MOT20 で当時最も ID スイッチが少なかった物体追跡モデルで、オクルージョンが多発するバス環境に最適なモデルである。モデルの学習データには運行データ 2 の動画を使用する。

Table 3.2: 乗車口におけるバス停別乗降者数推定結果

バス停	1	2	3	4	6	8	合計
正解人数	12	2	1	1	8	1	25
推定人数	12	3	1	1	8	1	26

Table 3.3: 降車口におけるバス停別乗降者数推定結果

バス停	4	5	6	7	8	9	10	11	12	13	14	15	合計
正解人数	1	1	4	2	3	2	1	1	4	2	3	3	25
推定人数	1	1	4	2	3	2	1	0	4	2	3	3	24

実験結果テストデータに運行データ 1 の動画を使用し、バス停別乗降者数推定を行った結果を表 3.2, 3.3 に示す。乗車口では 6 か所のバス停（6 本の動画）で推定を行い、25 人中 24 人を正確に推定することができた。バス停 2 では 1 件の ID スイッチが発生し、推定人数は 26 人となった。降車口では 12 か所のバス停で推定を行い、24 人を正確に推定することができた。バス停 11 では 1 人検出ミスが発生し、推定人数は 24 人となった。

3.4.3 実験 2：乗車区間別乗降者数推定

実験目的

提案システム BAIRD-OD の有効性を検証するため、乗車区間別乗降者数を推定する。これは同一判定の精度検証にあたる。

実験設定

入力として実験 1 で推定した軌跡を使用する。実際の利用者が 25 人に対して、推定乗車人数 26 人、推定降車人数 24 人である。メトリックラーニングモデルとして BoT[44] を使用する。BoT は Person Re-ID においてベースラインとなっている ResNet[30] ベースのモデルである。学習データとして DukeMTMC データセット [60] を使用する。DukeMTMC は Duke 大学で撮影された Re-ID の大規模データセットで、702ID、約 1.7 万枚の画像から構成されている。割当アルゴリズムとしてハンガリアン法 [61, 62] とランキング法 [63] を比較する。ハンガリアン法はコスト行列を入力とした割当問題を必ず最適解で解くアルゴリズムで、ランキング法はコスト行列を各行ごとにランクに変換してからハンガリアン法を適用する手法である。評価指標としては以下で定義される Recall, Precision, F-measure を使用する。

$$\text{Recall (再現率)} = \frac{TP (\text{正解した乗車・降車ペア数})}{TP + FN (\text{実際の全乗車・降車ペア数})}, \quad (3.1)$$

$$\text{Precision (適合率)} = \frac{TP (\text{正解した乗車・降車ペア数})}{TP + FP (\text{推定した全乗車・降車ペア数})}, \quad (3.2)$$

$$F\text{-measure (F 値)} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}. \quad (3.3)$$

実験結果

乗車区間別乗降者数推定を行った結果を表 3.4 に示す。いずれの指標でもハンガリアン法の方が推定精度が高かった。ハンガリアン法では 18 ペア正解することができ、いずれの評価指標も 7 割を超える精度となった。

Table 3.4: アルゴリズムごとの乗車区間別乗降者数推定結果

アルゴリズム	正解ペア数	Recall	Precision	F measure
ハンガリアン法	18	0.75	0.72	0.74
ランキング法	14	0.58	0.56	0.57

ハンガリアン法によるマッチングにおいて、乗車人物と降車人物の割り当てに失敗したペアの ID を表 3.5 に示す。ID1 の降車時の軌跡が ID20 の乗車時の軌跡とマッチングしている。ID20 は降車時に追跡できなかったため、ID1 の乗車に対応する軌跡が存在しなくなっている。ID15 と ID18 が入れ替わってマッチングしている。ID17, ID19, ID22 は 3 人で入れ替わってマッチングしている。

Table 3.5: 誤ってマッチングしたペア

on id	1	15	17	18	19	20	22
off id	-	18	22	15	17	1	19

推定した乗車人物 26 人と降車人物 24 人との特徴量距離に関するヒストグラムを図 3.19 に示す。正しいペアの特徴量距離のうち、正しくマッチングできた値を青、誤ってマッチングしたものをオレンジで表示している。異なる ID 同士の特徴量距離を緑で表示している。正しいペアの特徴量距離は概ね値が小さい方に分布しているが、正しいペアでも距離が遠くなっている場合や異なる ID のペアでも距離が近くなっている場合もある。

また、ある乗車人物に対して全降車人物との特徴量距離を昇順に並べて、正解となる降車人物の順位を Rank とし、全乗車人物の中で各降車人物が N 位以内に入っている割合を Rank-N とする。その結果を表 3.6 に示す。Rank-1 に示すように、全体の 52% で正しいペアの特徴量距離が最も小さくなった。誤ってマッチングしたペアでは Rank が 5,6,10 となっているペアもある一方で、Rank が 1 にも関わ

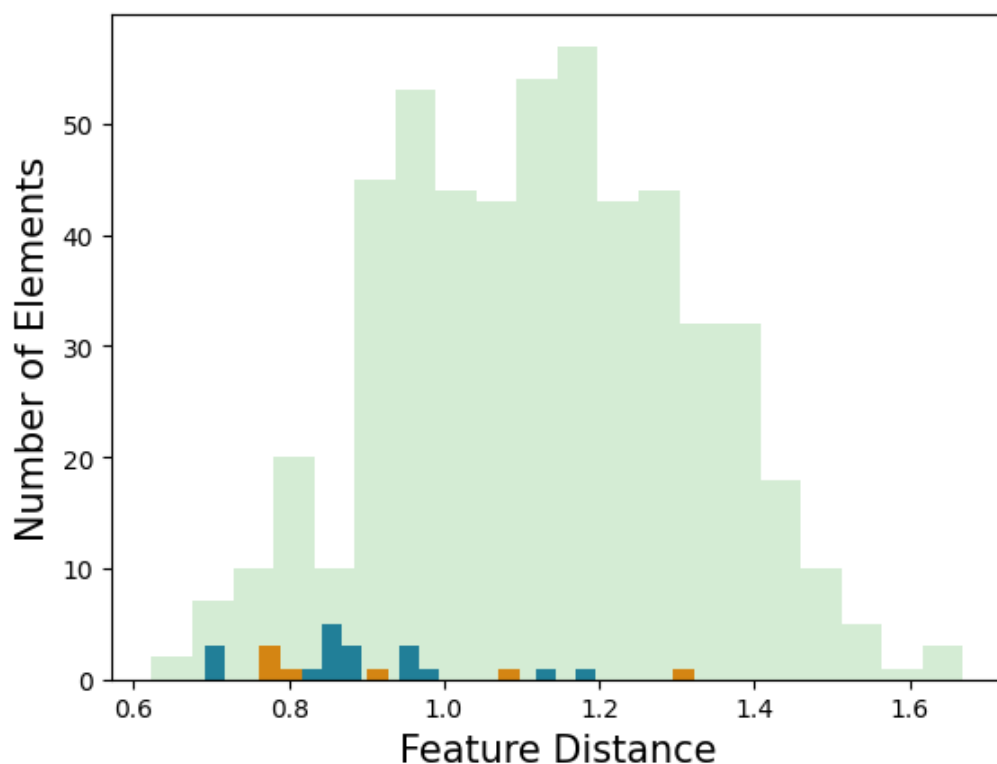


Figure 3.19: 特徴量距離の分布

らず正しくマッチングできていないペアもあった。それを分析するため、誤ったペアでの特徴量距離を表 3.7 に示す。ID1, 20 に関しては、降車する ID1 に対して、乗車する ID20 の方が距離が近くなっている。ID15, 18 に関して、乗車する ID15 と降車する ID18 の距離が近すぎたため、入れ替わってマッチングされている。ID17, 19, 22 に関して、乗車する ID17 と降車する ID22 の距離が近すぎるため、乗車する ID22, 乗車する ID19 に影響が出ている。

Table 3.6: Rank-N の割合

Rank	1	2	3	4	5	6	10
Correct	12	1	2	1	1	1	-
Miss	1	0	1	1	2	1	1
Rank-N	52%	56%	68%	76%	88%	96%	100%

考察 精度を下げている原因として、乗車人物の候補が ID スイッチにより増加していること、降車人物の候補が人物追跡の段階で 1 人追跡できなかったことが第一に挙げられ、バス停別乗降者数の推定精度を向上させる必要がある。次に、乗車人物と降車人物の映り方の差異が挙げられる。図 3.20 に示すように乗車口では自然光が多く入り比較的明るかったり白飛びした画像になったりする傾向にあるが、降車口ではバス中央の通路から料金箱付近では暗くなる傾向にある。こういった明

Table 3.7: 誤ってマッチングしたペアの特徴量距離

miss	20, 1	15, 18	18, 15	17, 22	19, 17	22, 19
	0.963	0.770	0.567	0.593	0.903	0.942
true	1, 1	15, 15	18, 18	17, 17	19, 19	22, 22
	1.321	0.898	0.761	0.904	0.761	0.764
Rank	5	10	3	5	1	5

るさの違いが乗車人物と降車人物との特徴量距離に影響を与えていることが考えられる。さらに図 3.21 に示すように、乗車口では乗車人物を正面で撮影した画像が多いのに対して、降車口では通路部分では正面、料金箱付近では頭上、ドア付近では後方から撮影され、体の角度も斜めになることがある。冬の札幌というデータ収集状況もあり、見た目がよく似た利用者が多かった。マッチングに失敗したペアの例を図 3.22 に示す。2 人の利用者はともに茶色の帽子、白いマスク、黒いコート、グレーのリュックという酷似した外観特徴を持っており、人の目で見ても同一判定の難易度が高い。加えて乗車口と降車口で照明条件も異なるため色味も変わり、その難易度はさらに上がる。

3.5 まとめ

本章ではバス OD 推定システムの全体設計を行って BAIRD-OD を提案するとともに、その有効性の検証を行った。本システムは①動画取得、②個人識別、③乗降判定、④同一判定という 4 つの処理から構成されている。各処理の内容と先行研究を分析し、必要な要素技術を検討した。動画取得に関して、冬の札幌という環境で実際に運行している路線バスにカメラを取り付け、利用者が乗降する映像を収集し、アノテーションを行ってデータセットを作成した。個人識別に関して、人物検出器 YOLOR と人物追跡器 MPNTrack を用いて人物軌跡を推定した。乗降判定に関して、先行研究の内容からルールベースで条件を設定した。バス停別乗降者数推定の実験を行い、一定の精度を確認した。同一判定に関して、メトリックラーニングモデル BoT と割当アルゴリズムを用いて人物軌跡同士のマッチングを行った。乗車区間別乗降者数推定の実験を行い、一定の精度を確認した。2 つの実験を通して提案システムに一定の有効性があることを確認するとともに、推定に失敗したデータを分析することで人の目で見ても同一人物の判別が難しい画像が存在することが判明した。



Figure 3.20: 乗車口 (左) と降車口 (右) での明るさの違い



Figure 3.21: 乗車口 (左) と降車口 (右) での角度の違い



Figure 3.22: マッチングに失敗した画像例 (左：乗車口での ID15,18, 右：降車口での ID15,18)

第4章 個人識別の精度向上

4.1 はじめに

本章の目的は、第3章で提案したバス乗客 OD 推定システム BAIRD-OD における②個人識別部分の精度を向上させることである。本章の位置づけは図 4.1 のとおりである。本システムは処理が直列につながっており、各処理の推定精度が次の処理の推定精度に直結している。そのため、前段の処理の精度向上が最終的な出力の精度向上につながる。一方で、後段の処理の改善から着手すると、精度が十分でない入力に過度に適合してしまうおそれがある。以上より、本章では前段に位置する個人識別から改善を行うこととした。

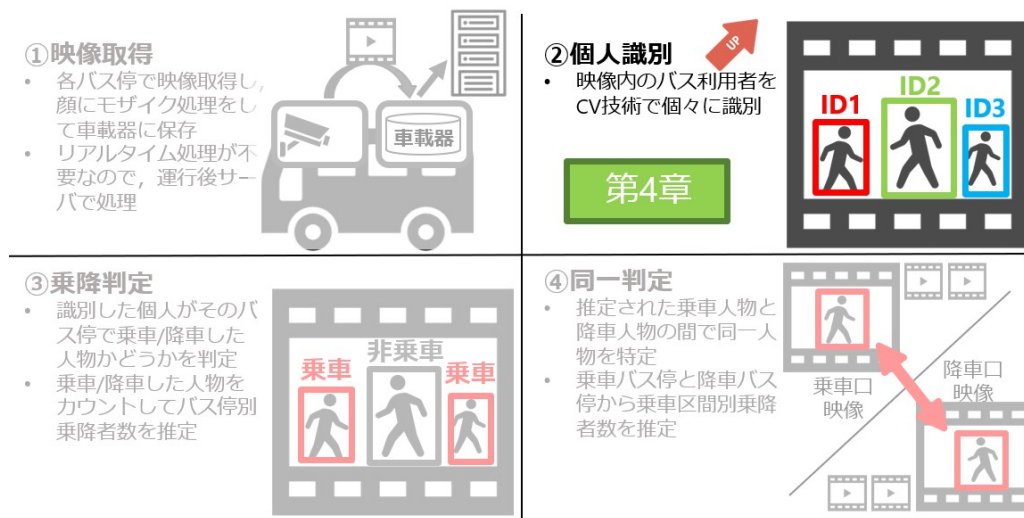


Figure 4.1: 本論文における第4章の位置づけ

本章では、4.2 節で個人識別における精度向上手法の検討を示す。4.3 節では提案手法を述べる。4.4 節では、提案手法の有効性を実験により検証し、4.5 節では、本章のまとめを示す。

4.2 個人識別における精度向上手法の検討

本節では、個人識別において発生している課題の分析と、課題を解決し推定精度を向上させる手法についての検討を行う。

4.2.1 個人識別における課題

提案システム BAIRD-OD を他の路線の動画に適用したところ、十分な精度が得られず、汎化性能が十分でないことが判明した。発生している問題は ID スイッチと軌跡の断片化である。ID スイッチは図 4.2 のように、ある人物に付いていた ID が他の人物に付いてしまい、2 人の人物が 1 本の軌跡にまとめられてしまうものである。軌跡の断片化は図 4.3 のように、ある人物の軌跡がオクルージョンや検出ミスによって 2 つの軌跡に分かれるものである。軌跡のフレームが飛んでしまうため、前半の軌跡の最後のフレームと後半の軌跡の最初のフレームを比較した際に、幾何情報や外観情報が変化しやすく、同一人物に 2 つの ID が付いてしまう ID スイッチが特に発生しやすくなる。

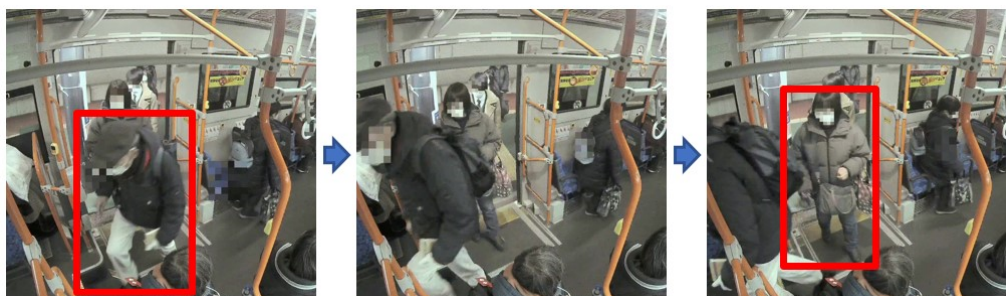


Figure 4.2: ID スイッチが発生した例

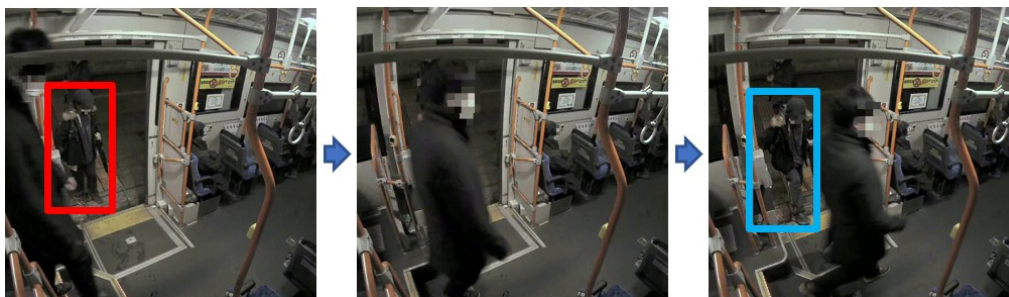


Figure 4.3: 軌跡の断片化が発生した例

これらが発生している要因として、不適人物（当該バス停で乗降しない人物）の軌跡が多いこと、本来の軌跡の一部しか追跡できていないことが考えられる。個人識別における最終的な出力である人物軌跡の前段階として、入力である人物矩形の段階で体の一部のみが検出されていること、無関係な車外人物が検出されていること、さらに白飛びや逆光による検出失敗が発生していることが確認された。

4.2.2 推定精度向上手法

人物追跡モデルの比較

人物追跡タスクは人物検出と Re-ID という 2 つの処理を組み合わせたタスクであるが、人物追跡モデルの学習方法のアプローチとしては大きく End-to-End (E2E) と Tracking-by-Detection (TbD) に分かれる。E2E と TbD の比較を表 4.1 に示す。

Table 4.1: 人物追跡モデルの比較

	End-to-End	Tracking by Detection + ヒューリスティクス
概要	検出と ReID の統合モデル	検出+ReID の 2 段階構成とヒューリスティクスの補正
学習データ	大規模なデータが必要，ドメインシフトに弱い	小規模データでも稼働，ドメイン知識によるチューニングが容易
計算資源	重いモデルが多い	軽量のモデルもある
遮蔽	強いが学習データに依存	比較的弱い但し補正可能
説明可能性	低く，失敗要因の分析が難しい	高く，失敗パターンを規則化できる

E2E は推定精度や，情報系分野における最先端のアプローチという点で強力なモデルだが，実用上の制約下では弱点が顕在化している。大規模なモデルにすべてを一括して学習させるよりも，タスクごとのドメイン知識を活用することの有効性を主張する研究も多い [64, 65, 66, 67, 68]。本研究の路線バスという環境においては，利用者の移動経路・速度・方向などの行動ルールや，カメラの設置位置と人・ドア・車内オブジェクトの座標といった空間的制約などのドメイン知識が重要な役割を持つ。そのため，TbD とヒューリスティクスを組み合わせたアプローチを採用する。

実用的なシステムの制約

理想的な実験環境とどまらず，実用的なシステムを考えたときの制約として以下のことが考えられる。

①**実環境におけるロバスト性** 複雑なオクルージョン，多様な照明変化，センサ機器の位置や角度の制約など様々な問題が発生する。これらの条件に対してロバストに動作し得る手法を検討する必要がある。

②**コスト効率** 社会実装を前提とした場合，費用や工数の面でコスト効率に関する制約が課される。利用可能なセンサや計算資源といったハードウェア面の制約（費用），実運用段階で必要となる専門的なチューニング作業数（工数）の両面で効率性が求められる。この観点からも，TbD+ヒューリスティクスのアプローチは適している。

③**プライバシーへの配慮** 利用者の動画収集という点で個人情報や肖像権の扱いには十分な配慮が必要である。OD データの収集という目的上リアルタイム処理が不要であり、サーバにおけるバッチ処理のためには動画の保存が必要である。これらの点については、第3章のデータ収集プロトコルにおいて撮影の告知と顔領域へのモザイク処理を行うことで対応している。

推定精度向上手法の比較

人物検出精度へのアプローチとして、検出器の改良、マルチモーダル、フィルタリングについて比較した表を表4.2に示す。課題としては部分遮蔽、不適人物の検出、光変化による検出精度低下があり、制約としては費用面、工数面、動画ブレがある。本研究においては人物矩形をフィルタリングするアプローチが適切であると判断した。

Table 4.2: 人物検出精度へのアプローチ手法の比較

	課題			制約		
	部分遮蔽	不適人物	光変化	費用	工数	動画ブレ
検出器の改良	○	×	○	○	△	△
マルチモーダル	○	○	○	×	×	○
フィルタリング	○	○	△	○	○	○

人物追跡精度へのアプローチとして、データ拡張、高難度データセットによる学習、短い軌跡の除去、トラック保持戦略、軌跡の結合について比較した表を表4.3に示す。課題としては断片化とIDスイッチによる追跡精度低下があり、制約としては費用面、工数面がある。本研究においては短い軌跡の除去と軌跡の結合によるアプローチが適切であると判断した。

4.3 提案手法

本研究では、推定精度低下の要因になっている矩形や軌跡を不適データセグメントと呼称する。本節では、不適データセグメントの説明とそれらに対する処理手法の提案について示す。

4.3.1 人物検出へのアプローチ

人物検出により出力される矩形の中で、不自然なもの・精度を下げていると考えられるものに対してフィルタリングを行う。矩形の中心座標を x, y 、幅を w 、高さを h とする。

Table 4.3: 人物追跡精度へのアプローチ手法の比較

	課題		制約	
	断片化	ID スイッチ	費用	工数
データ拡張	×	○	△	×
高難度データセット	△	○	○	△
短軌跡除去	○	○	○	○
トラック保持戦略	○	×	○	○
軌跡の結合	○	○	○	○

①**大きい矩形の削除** 図 4.4 のように、バス後方から移動する人物や乗車口ドアから入ってカメラ手前に来る人物を対象としてフィルタリングする。体の一部が見切れてしまうことや、カメラに近い分移動速度が速く、ぶれてしまうことが精度低下につながるからである。閾値を bt として、以下の式に該当する矩形を削除する。ここでは閾値 bt を 150,000 px に設定した。

$$w \times h > bt$$

②**小さい矩形の削除** 図 4.5 のように、バスの外で遠くにいる人物や体の一部のみ見切れる人物を対象としてフィルタリングする。不適人物の矩形や、体の一部のみの矩形は人物追跡時に精度低下につながるからである。閾値を st として、以下の式に該当する矩形を削除する。ここでは閾値 st を 35,000 px に設定した。

$$w \times h < st$$

③**右側に映る矩形の削除** 図 4.6 のように、後ろ向きの人物や手すりとおクルージョンする人物を対象としてフィルタリングする。後ろ向きの人物は乗降人物のマッチングにおいて有効でないことと、手すりとおクルージョンすることで人物追跡の精度低下につながるからである。閾値を rt として、以下の式に該当する矩形を削除する。ここでは閾値 rt を 300 px に設定した。

$$x - \frac{w}{2} > rt$$

4.3.2 人物追跡へのアプローチ

④**軌跡の結合** 図 4.7 のように、断片化、ID スイッチした軌跡を対象として軌跡の結合を行う。単純な ID スイッチや軌跡の断片化（フレームジャンプ）による ID スイッチが発生した際、幾何情報（矩形の中心座標、幅、高さ）が大きく変わらな

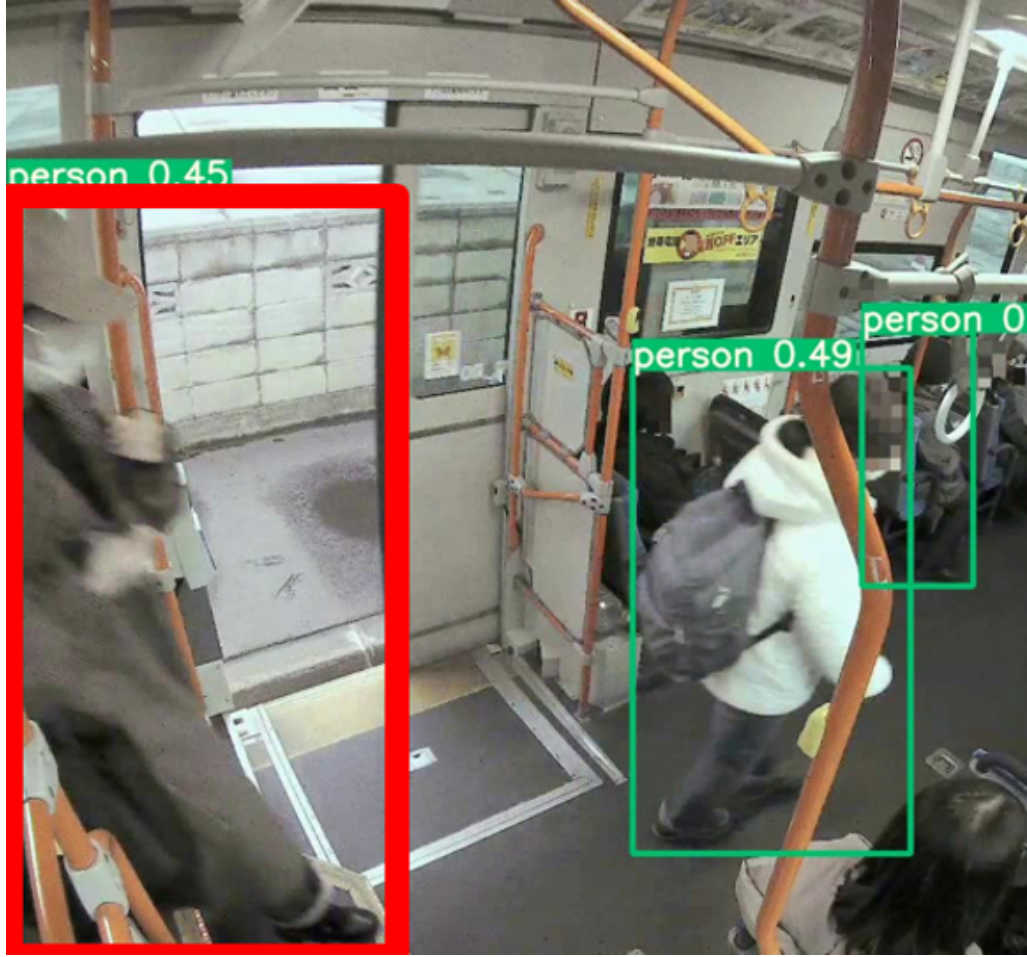


Figure 4.4: 大きい矩形

いことが多い．この特性を利用して動画内の全軌跡の最初の矩形と最後の矩形を総当たりで比較し，フレームの差が tt 未満かつ IoU（Intersection over Union）の値が it より大きい軌跡同士に同一の ID を割り当てて結合する．条件は以下の式で表され， tt を 3 フレーム， it を乗車口では 0.8，降車口では 0.4 に設定した．

$$\kappa_{\text{start}}^p - \kappa_{\text{end}}^q < tt \wedge \text{IoU}_{\kappa_{\text{end}}^q}^{\kappa_{\text{start}}^p} > it$$

⑤短い軌跡の除去 オクルージョンを原因の中心として，断片化により発生した短い軌跡を削除する．フレームの閾値 ft を 3 に設定した．

$$\kappa_{\text{end}}^q - \kappa_{\text{start}}^p < ft$$

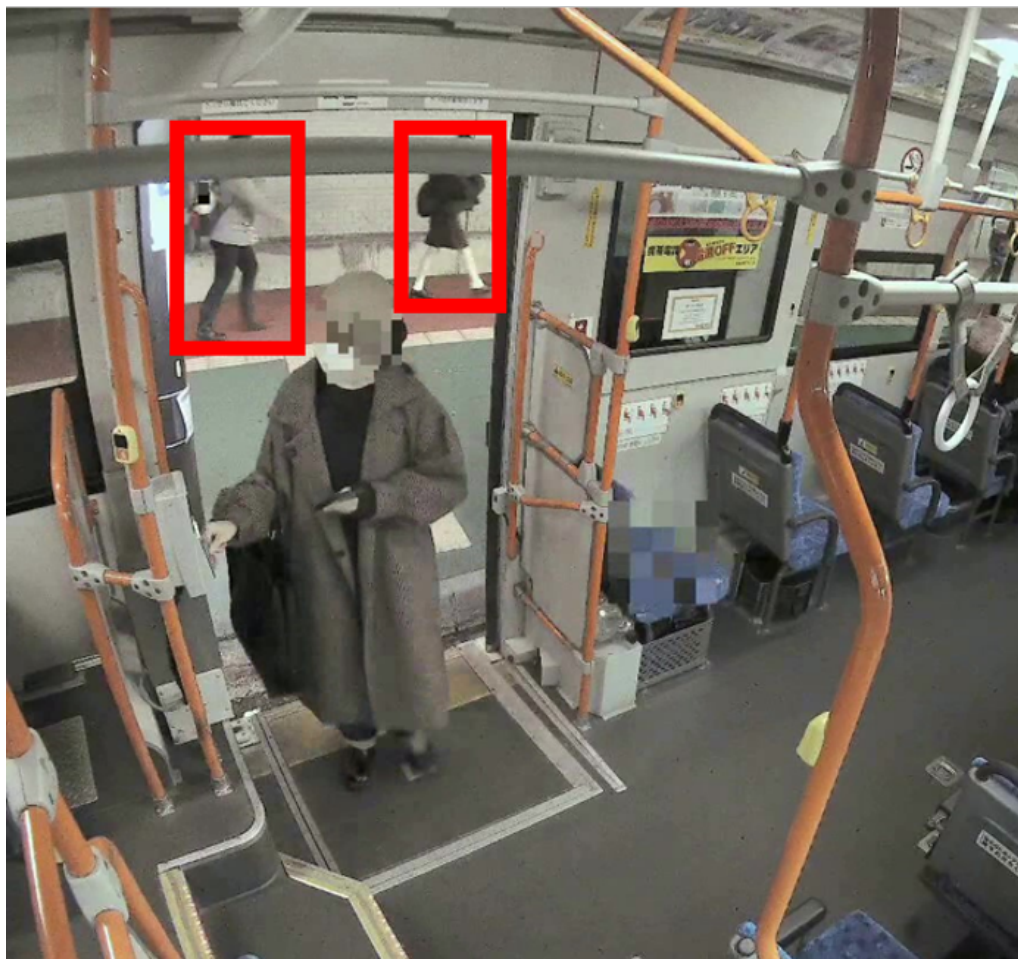


Figure 4.5: 小さい矩形

4.4 提案手法の有効性の検証

4.4.1 予備実験 1：人物検出器の選定

より精度の高い人物検出器を選定するため、YOLOv5[69], YOLOv8[70], YOLORの推定精度比較実験を行う。各モデルはMSCOCOで学習させ、運行データ1と運行データ2のすべてのフレームを用いてテストを行う。評価指標は物体検出において最も一般的な指標であるAP(AveragePrecision)を用いる。YOLOシリーズの物体検出器は推論を行う際、各矩形のクラス確信度を同時に出力するため、本実験では「Person」クラスの確信度の閾値を0.2~0.6の間で0.1ずつ変化させてAPを計測する。その結果を表4.4に示す。今回の実験ではYOLORを確信度0.4に設定した場合がAP0.61と最も高くなったため、本章ではこのモデルと閾値を採用する。

4.4.2 予備実験 2：人物追跡器の学習

人物追跡器MPNTrackの学習及びその評価を行う。本システムは人物検出器で推定した人物矩形を用いて人物追跡を行うが、本実験ではMPNTrackがバス車内と



Figure 4.6: 右側に映る矩形

いう状況での人物同士の Re-ID を正しく行えるかを確認するため、学習も評価もアノテーションを行った Ground Truth の矩形情報を用いる。データセットとして運行データ 1 と運行データ 2 の動画を使用する。学習・評価に用いるデータセットの組み合わせを表 4.5 に示す。条件 1～4 は乗車口と降車口でモデルを分けることを想定し、条件 5, 6 は同一のモデルで処理することを想定している。本実験の評価

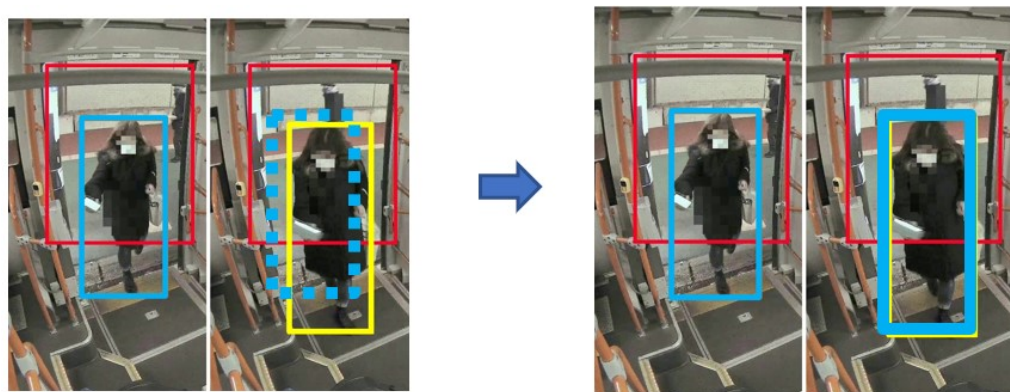


Figure 4.7: 軌跡の結合

Table 4.4: 人物検出器の推定精度比較

モデル	0.2	0.3	0.4	0.5	0.6
YOLOv5	0.55	0.55	0.55	0.54	0.53
YOLOv8	0.60	0.60	0.52	0.58	0.55
YOLOR	0.48	0.55	0.61	0.58	0.54

指標には物体追跡の評価指標として一般的な IDSW, MOTA, IDF1 を用いる。実験の結果を表 4.6 に示す。MOTA, IDF1 の値が高く、十分に学習できていることが確認できた。乗車口と降車口でモデルを分けて処理した条件 1~4 と同一のモデルで処理した条件 5, 6 では大きな差が見られなかった。

Table 4.5: 学習データとテストデータの割当

条件	運行データ 1 乗	運行データ 1 降	運行データ 2 乗	運行データ 2 降
1	Train		Test	
2	Test		Train	
3		Train		Test
4		Test		Train
5	Train	Train	Test	Test
6	Test	Test	Train	Train

Table 4.6: 人物追跡器 MPNTrack の推定精度

条件	利用者数	IDSW	MOTA	IDF1
1	32	4	0.995	0.937
2	25	0	0.996	0.957
3	32	11	0.988	0.963
4	25	2	0.997	0.993
5	64	21	0.951	0.951
6	50	0	0.998	0.999

4.4.3 実験 1：バス停別乗降者数推定

データセット

新たに 3 運行分の動画データを追加し、より広範なデータセットで精度検証を行う。動画データの選定条件は 3 章同様に以下である。矩形アノテーションは行わず、各バス停での乗降人数のみを目視で計測した。データセット名を運行データ 3, 4, 5 とし、詳細を表 4.7 に示す。

- 極端に密集しすぎない中程度の人数（30 人前後）
- 暗すぎない日中の時間帯
- 同じ制服を着た学生が大量に利用する通学路でない

Table 4.7: データセットの詳細

	アノテーション	利用者数	乗車口 動画数	降車口 動画数
運行データ 1	矩形あり	25	6	12
運行データ 2	矩形あり	32	13	15
運行データ 3	人数のみ	25	9	8
運行データ 4	人数のみ	26	7	14
運行データ 5	人数のみ	22	11	4

実験目的

提案手法による個人識別と乗降判定の精度検証を行う。

実験設定

予備実験 1 の結果から、人物検出器は YOLOR を使用し、MSCOCO で学習させ、Person クラスの確信度の閾値を 0.4 に設定する。予備実験 2 の結果から、人物追跡は MPNTrack を使用し、条件 2, 4 のとおり乗車口と降車口でモデルを分けて運行データ 2 で学習させる。提案手法の実装に関して、乗車口では処理①～⑤を全て実装し、降車口では②, ④, ⑤, を実装する。降車口ではカメラがフロントガラス上部あたりに設置されており、体の一部が見切れるほど利用者が近づくことがないため①は実装していない。また、降車口では手すりのように全身とオクルージョンする車内オブジェクトがないため、③のような実装もしていない。評価用データセットとして、運行データ 1, 3, 4, 5 を用いる。評価指標としては以下で定義される誤差率を用いる。以下の式は乗車口のため発生交通量 tg_{β} で計算しており、降車口では集中交通量 ta となる。

$$\text{誤差率} = \frac{1}{B} \sum_{\beta=1}^B \frac{|\hat{tg}_{\beta} - tg_{\beta}|}{tg_{\beta}}$$

実験結果

実験結果を表 4.8 に示す。運行データ 1, 3, 4, 5 に対してシステムを適用した結果、乗車口では誤差率が第 3 章システムの 41.7% から 13.3% に減少した。降車口では 29.5% から 5.24% に減少した。

Table 4.8: バス停別乗降者数の誤差率

	乗車口	降車口
第3章システム	41.7%	29.5%
提案システム	13.3%	5.24%

4.4.4 実験2：アブレーションスタディ

実験設定

提案手法の①～⑤の各処理が有効に機能しているかどうかを検証するため、アブレーションスタディを行う。提案手法を全て実装した場合を With All, そこから各処理を抜いた場合を w/o (without) とする。評価指標として新たに ID スイッチ (IDSW) と MOTA も計算した。ID スイッチは Ground Truth の軌跡に対して、推定した ID が変化した回数, MOTA は以下で表される式で, どちらも人物追跡の分野では一般に用いられる指標である。本研究の目的がバス停別乗降者数の推定であることから, 誤差率を中心的な指標として扱う。

$$MOTA = 1 - \frac{FN + FP + IDSW}{GT}$$

実験結果

実験結果を表 4.9, 4.10 に示す。誤差率は乗車口, 降車口ともにすべて実装した With All が最も誤差率が低く, どちらも 8.33% となった。

Table 4.9: 乗車口における各条件の効果比較

	誤差率	IDSW	MOTA
with All	8.33%	1	0.557
w/o ①	27.1%	1	0.536
w/o ②	39.6%	5	0.395
w/o ③	13.9%	2	0.434
w/o ④	9.72%	2	0.559
w/o ⑤	43.1%	3	0.538

4.4.5 実験3：乗車区間別乗降者数推定

実験目的

提案手法により推定精度が向上した乗降人物を用いた, 同一判定の精度検証を行う。

Table 4.10: 降車口における各条件の効果比較

	誤差率	IDSW	MOTA
with All	8.33%	0	0.512
w/o ②	8.33%	1	0.446
w/o ④	17.4%	2	0.498
w/o ⑤	13.9%	3	0.521

実験設定

実験1で推定した人物軌跡を用いて、乗車区間別乗降者数を推定する。乗降人物の同一判定に関しては第3章と同じ設定で実験を行う。メトリックラーニングモデルにはBoTを使用し、DukeMTMCデータセットで学習させた。割当アルゴリズムにはハンガリアン法を使用する。評価指標はRecall, Precision, F-measureを使用する。

実験結果

実験結果を表4.11に示す。Recall0.77, Precision0.83, F-measure0.80となった。データセットを追加したため単純な比較にはならないが、第3章の実験におけるRecall0.75, Precision0.72, F-measure0.74よりも高い精度となった。

Table 4.11: 第4章手法を実装した場合の乗車区間別乗降者数推定結果

合計人数	正解ペア数	Recall	Precision	F-measure
98	75	0.77	0.83	0.80

4.5 まとめ

本章では、BAIRD-ODの個人識別における推定精度を向上させるための手法を提案した。実用的なシステムの制約下で、人物検出と人物遺跡のそれぞれに対して候補手法を検討し、以下の処理を導入した。

- 不自然な検出のフィルタリング
- 軌跡の結合
- 短い軌跡の除去

追加のデータセットを作成し、バス停別乗降者数推定の精度検証の実験を行った結果、第3章のシステムよりも推定精度が向上し、乗車口では誤差率13.3%、降車口では5.24%となった。また、アブレーションスタディを行い、各処理が有効に

機能していることを確認した。さらに、乗車区間別乗降者数推定の精度検証の実験を行った結果、Recall0.77, Precision0.83, F-measure0.80 となった。

第5章 乗降判定の精度向上

5.1 はじめに

本章の目的は、第3章で提案したバス乗客 OD 推定システム BAIRD-OD における③乗降判定の精度を向上させることである。本章の位置づけは図 5.1 のとおりである。ルールベースの手法は機器費用や専門的なチューニング工数の面で低コストであるが、精度に限界がある。本章では必要最低限の機械学習を導入することで、精度向上を図る。

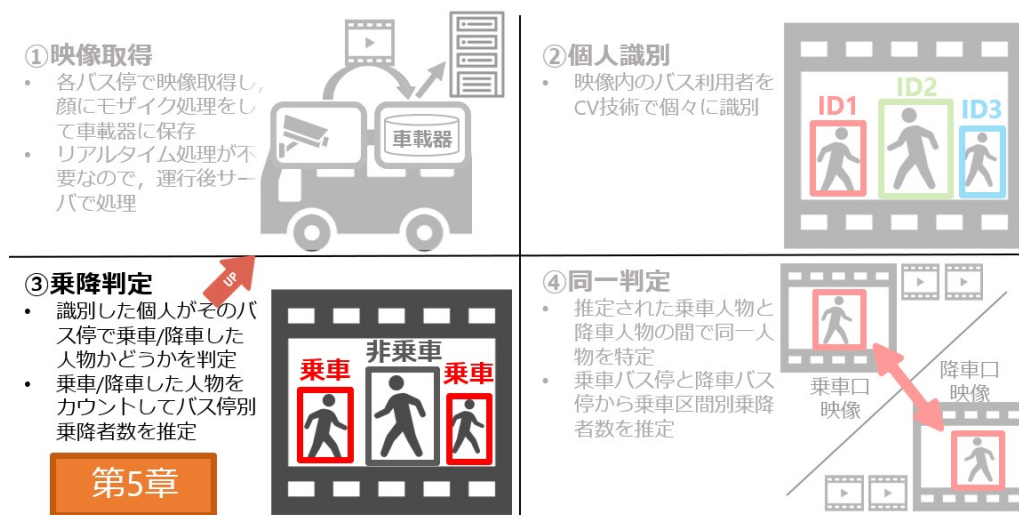


Figure 5.1: 本論文における第5章の位置づけ

本章では、4.2 節では、乗降判定における課題と提案手法について述べる。4.3 節では、提案手法の有効性を実験により検証し、4.4 節では、本章のまとめを示す。

5.2 乗降判定における精度向上手法の検討

本節では、乗降判定において発生している課題の分析と、それに対処するための提案手法について述べる。

5.2.1 乗降判定における課題

第4章までのシステムでは人物軌跡に含まれる矩形の中心座標と路線バスの扉範囲との位置関係から乗車人物および降車人物の判定を行っていた。しかし、乗車口ではバス後方から前方に高速で移動する人物が画像左から右に移動することで乗車人物と判定されたり、降車口では手すりの近くに立っていたり両替をしに来た利用者が降車人物と判定されるといった、誤判定が発生していた。

具体的な例を図5.2に示す。青い矩形で囲まれた人物は個人識別で推定された軌跡の最初のフレームである。この矩形は第4章で実装してフィルタリングでは弾かれず、中心座標が扉範囲に入っているため、乗車人物として判定される。しかし、この人物はバス停で降りるためにバス後方から前方に移動している人物であり、乗車人物ではない。ルールベースの乗降判定は簡便な実装である一方で、このような既に乗車している人物を乗車判定してしまうことがある。

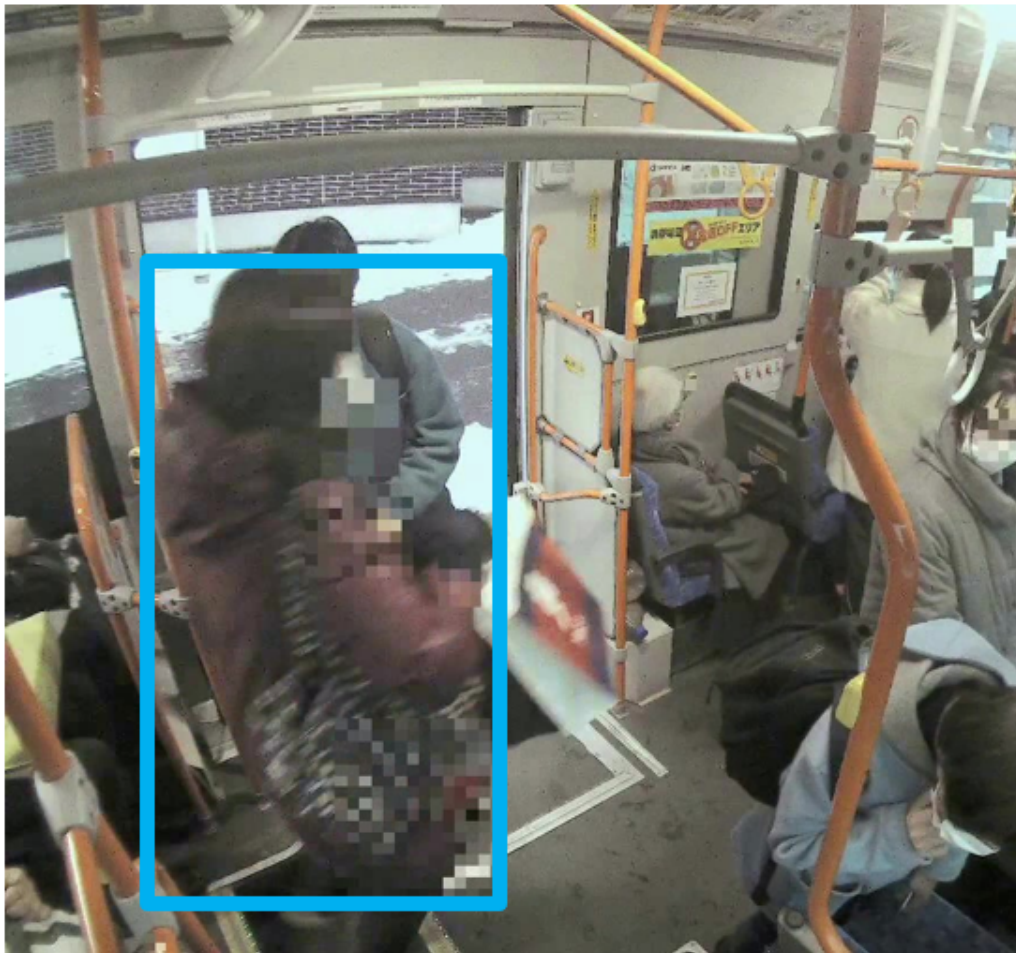


Figure 5.2: 誤って乗車人物と判定される例

乗車口では利用者がもともとバス車内にいたのかバス車外にいたのかを正しく判定することが重要であり、つまり推定された軌跡の最初のフレームのバス内外判定の精度を向上させることが乗降判定精度を向上させることにつながると判断し

た．降車口ではバス前方にいる人物がそのまま降車したかどうかを正しく判定することが重要であり，つまり推定された軌跡の最後のフレームのバス内外判定の精度を向上させることが乗降判定精度を向上させることにつながると判断した．

5.2.2 提案手法

本項では，ルールベースで行っていた乗降判定に対して DNN を導入する提案手法と，その際に検討した他手法候補との比較について述べる．

DNN によるバス内外判定 ルールベースで行っていた乗降判定に機械学習を導入し，DNN モデル [71] によるバス内外判定を行う．DNN モデルは人物矩形がバス車内かバス車外かを二値分類する．入力には矩形の中心座標 x, y ，幅 w ，高さ h から構成される幾何的特徴と，動画の総フレーム数および矩形が出現したフレーム番号から構成される時間的特徴である．出力はバス車内を示す `inside` クラスとバス車外を示す `outside` クラスの事後確率である．乗降判定の手順としては，乗車口では人物軌跡の 1 フレーム目が `outside` クラスと判定された人物を乗車人物とし，降車口では人物軌跡の最終フレームが `outside` クラスと判定された人物を降車人物とする．

他手法候補との比較検討 本研究における乗降判定は，個人識別により推定した各軌跡に対し，「当該バス停で乗車または降車した対象人物か否か」を判定する二値分類問題として定式化できる．入力は，前述の幾何的・時間的特徴から構成される比較的次元の低い特徴量であり，これに対しては汎用的な機械学習手法としてロジスティック回帰，サポートベクトルマシン (SVM)，決定木ベースの手法 (ランダムフォレストや勾配ブースティング決定木) など候補となる．

ロジスティック回帰はモデル構造が単純で学習も高速であり，各特徴量の重みから寄与度を解釈しやすいという利点がある．一方で決定境界が本質的に線形であるため，扉付近を通過するだけの人物と実際に乗降する人物のように，複数特徴の非線形な組み合わせによってのみ分離できるクラスに対しては表現力が不足する．また，車内の混雑度やカメラ揺動といった環境要因により特徴分布がバス停ごと・便ごとに変動する状況では，線形モデルの仮定が崩れやすいという問題がある．

ランダムフォレストや勾配ブースティング決定木は，非線形な決定境界や特徴間の相互作用を自動的に表現できるため，ロジスティック回帰より高い表現力を持つ．また，特徴量のスケールにあまり敏感でないことや，変数重要度の算出による解釈性の高さも利点である．しかし木ベースのモデルは学習データの分割構造に強く依存するため，学習時に現れなかった稀な軌跡パターンや，路線変更などによる分布外データに対する汎化性能に注意が必要である．さらに，本研究では将来的に画像特徴との統合や他の深層学習モジュールとの結合を見据えており，同一フ

レームワーク上で扱いにくい点も運用上の制約となる。

SVM はカーネル関数を用いることで高い識別性能を示すことが知られており、マージン最大化に基づくロバストな決定境界が期待できる。一方でサンプル数が増えると学習計算量とメモリ使用量が急増しやすく、本研究のように停留所単位で継続的にデータが蓄積されていく応用では学習の再実行コストが問題となる可能性がある。また確率的出力を得るためには別途キャリブレーションが必要であり、しきい値調整やルールベースとの組み合わせをシンプルに保ちたいという要件とは相性が良くない。

これらに対して、本研究で採用する小規模な DNN は、非線形な決定境界を柔軟に学習できると同時に、既存の人物検出・再識別モジュールと同じ深層学習フレームワーク上で実装・運用できるという利点を持つ。これにより、推論パイプラインの統一や将来的な end-to-end 統合、画像特徴との連結といった拡張が容易になる。一方で内部表現の解釈は容易ではないが、第 3 章で定義した扉範囲や軌跡パターンに基づくルールベース処理と併用することで、挙動の把握可能性を一定程度確保している。以上より、本研究では DNN によるバス内外判定を乗降判定器として採用し、他方式はベースラインおよび参考手法として位置付ける。

5.3 提案手法の精度検証

本節では、提案手法の精度を検証するために新たに作成したデータセットと、精度検証の実験について示す。

5.3.1 データセット

DNN モデルを学習させるため、人物矩形とバス内外のクラスアノテーションを行ってデータセットを作成する。新たに 2 路線の動画を選定し、運行データ 6,7 とする。データセットの詳細を表 5.1 に示す。バス内外のクラスアノテーションの例を図 5.3 に示す。左の画像が乗車口、右の画像が降車口で、緑の矩形がバス車内を示す 'inside' クラス、赤の矩形がバス車外を示す 'outside' クラスである。

Table 5.1: 第 5 章で用いるデータセットの概要

データ名	発車時刻	利用者数	総画像数
運行データ 6	9:24 発	31 人	2,262 枚
運行データ 7	16:40 発	18 人	4,680 枚

5.3.2 実験 1：物体検出器の精度確認

実験目的

YOLOv10 より新しく開発された物体検出モデル YOLOv10[72] の精度検証を行う。

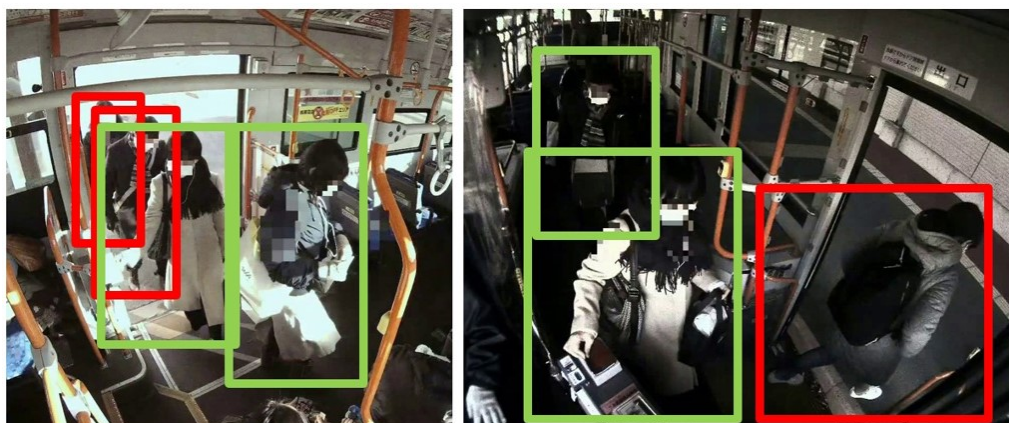


Figure 5.3: バス内外のアノテーション（緑：inside, 赤：outside）

実験設定

YOLOv10 の中で最もパラメータ数が多く検出精度の高い YOLOv10-X を使用する。学習は MSCOCO で行い、評価は第 3 章同様、運行データ 1, 2 の全矩形で評価する。評価指標としては物体検出で最も一般的な AP を使用する。

実験結果

実験結果を表 5.2 に示す。YOLOv10 では閾値を 0.7 に設定したときが最も精度が高く、AP0.403 となった。また、降車口においては利用者がカメラに近く画角が頭上からになった場合に検出精度が低下する傾向にあった。

Table 5.2: YOLOv10 の検出精度

閾値	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
YOLOv10	0.259	0.316	0.360	0.390	0.368	0.403	0.368	0.184

5.3.3 実験 2：DNN モデルの学習

実験目的

本節では本タスクにおける最適な DNN モデルのパラメータを探索する。探索するパラメータは層数と活性化関数である。

実験設定

乗車口と降車口ではカメラの設置位置や撮影方向、ドア位置が異なるため、同一のモデルで推定を行うことは適切でない。本研究では、乗車口側および降車口側に対してそれぞれ独立に、バス内外判定用の DNN による二値分類モデルを構築する。

モデル構造の探索にあたっては、隠れ層数を 3, 5, 7 層とし、活性化関数を ReLU[73], sigmoid[74], tanh[75] の 3 種類から選択した組合せに対してグリッドサーチを行う。評価には、バス内外クラスアノテーションを付与した運行データ 6, 7 を用いて 5 分割交差検証を実施する。各分割で得られた正解率を平均し、その試行における「層数×活性化関数」組合せの正解率とする。さらに、乱数シードを変更して同じ手順を 10 回繰り返し、得られた 10 個の正解率を平均することで、当該「層数×活性化関数」組合せに対する最終的な正解率を算出する。

実験結果

実験結果を表 5.3, 5.4 に示す。乗車口では 5 層×sigmoid の場合が最も精度が高く、正解率は 95.2%となった。降車口では 3 層×tanh が最も精度が高く、正解率は 97.4%となった。

Table 5.3: 乗車口における DNN モデルのパラメータのグリッドサーチ結果

	3 層	5 層	7 層
ReLU	94.0%	93.4%	93.2%
sigmoid	94.9%	95.2%	71.6%
tanh	93.8%	91.1%	92.6%

Table 5.4: 降車口における DNN モデルのパラメータのグリッドサーチ結果

	3 層	5 層	7 層
ReLU	96.5%	96.6%	96.3%
sigmoid	93.2%	89.0%	68.5%
tanh	97.4%	97.2%	97.2%

5.3.4 実験 3：バス停別乗降者数推定

実験目的

DNN モデルを用いた乗降判定による、バス停別乗降者数推定の精度を検証する。

実験設定

物体検出器、物体追跡器、不適データセグメントの処理に関しては第 4 章と同様である。DNN モデルは実験 2 の結果から、乗車口では 5 層で活性化関数に sigmoid を採用したモデル、降車口では 3 層で活性化関数に tanh を採用したモデルを用いる。学習データには運行データ 6, 7, 検証データには運行データ 1, 3, 4, 5 を使用し、評価指標には誤差率を用いる。

実験結果

実験結果を表 5.5, 5.6 に示す. 乗車口では DNN モデルと YOLOv10 を組み合わせた手法が最も精度が高く, 誤差率 11.7%となった. 降車口では DNN モデルと YOLOR を組み合わせた手法が最も精度が高く, 誤差率 0.803%となった.

Table 5.5: 乗車口におけるバス停別乗降者数推定の結果

人物検出器	ルールベース		DNN モデル	
	YOLOR	YOLOv10	YOLOR	YOLOv10
誤差率	14.4%	19.7%	20.1%	11.7%

Table 5.6: 降車口におけるバス停別乗降者数推定の結果

人物検出器	ルールベース		DNN モデル	
	YOLOR	YOLOv10	YOLOR	YOLOv10
誤差率	6.59%	36.7%	0.803%	37.0%

5.3.5 実験 4: 乗車区間別乗降者数推定

実験目的

提案手法により推定精度が向上した乗降人物を用いた, 同一判定の精度検証を行う.

実験設定

実験 3 で推定した人物軌跡を用いて, 乗車区間別乗降者数を推定する. 乗降人物の同一判定に関しては第 3 章, 第 4 章と同じ設定で実験を行う. メトリックラーニングモデルには BoT を使用し, DukeMTMC データセットで学習させた. 割当アルゴリズムにはハンガリアン法を使用する. 評価指標は Recall, Precision, F-measure を使用する

実験結果

実験結果を表 5.7 に示す. Recall0.79, Precision0.93, F-measure0.85 となり, 第 4 章までの既存手法を超える精度となった.

Table 5.7: 第 5 章手法を実装した場合の乗車区間別乗降者数推定結果

	正解ペア数	Recall	Precision	F-measure
第 4 章システム	75	0.77	0.83	0.80
提案システム	77	0.79	0.93	0.85

5.4 まとめ

本章では、BAIRD-OD の乗降判定における推定精度を向上させるための手法を提案した。DNN モデルによる矩形のバス内外判定を導入し、人物矩形の矩形情報と時間情報からバス内外を判定できることを確認した。DNN モデルの層数と活性化関数というパラメータについて、本タスクにおける最適な値の探索を行った。追加のデータセットを作成し、バス停別乗降者数推定の精度検証の実験を行った結果、第4章のシステムよりも推定精度が向上し、乗車口では誤差率 11.7%、降車口では 0.803% となり、ともにルールベースの手法を超える精度を達成した。乗車区間別乗降者数推定の精度検証の実験を行った結果、第4章のシステムよりも推定精度が向上し、Recall0.79, Precision0.93, F-measure0.85 となった。

第6章 結論

本論文では、路線バス車内に設置された既設カメラの映像から利用者の乗降行動を解析し、路線バス事業者が運行計画の見直しに用いることのできるバス OD データを、低コストかつ実装容易な方法で推定することを目的として研究を行った。バス事業者を取り巻く赤字経営や運転者不足といった社会的背景、車内カメラ映像を用いるに伴うプライバシー保護上の制約、さらに照度変動・オクルージョン・車体揺動など車内環境特有の要因に起因する技術的課題を踏まえ、個人認証レベルの精度を追求するのではなく、運行計画に足る整合的な精度と導入・運用コストのバランスを重視した。本論文全体を通じて、バス OD 推定システム BAIRD-OD の全体設計と要素技術の選定・改良、および実路線データを用いた評価を行い、車載カメラ映像を用いた OD データ収集の実現可能性と限界を明らかにした。

第3章では、実際に札幌市内を運行する路線バスにカメラを設置して冬季の乗降映像を収集し、得られた動画の特徴と Computer Vision タスクとしての難しさを整理した。白飛び・黒つぶれ、強いオクルージョン、暗色かつ類似した服装、マスク着用などの要因が人物検出・追跡・同一判定を困難にしていることを定性的に示した上で、動画取得、個人識別、乗降判定、同一判定の4処理からなるバス OD 推定システム BAIRD-OD を設計し、各処理の入出力と、バス停別の発生交通量・集中交通量および区間別の分布交通量への変換方法を形式的に定義した。さらに、選定した2便の動画に対して人物矩形と ID の完全アノテーションを行い、システムの有効性を検証した。その結果、バス停別乗降者数については乗車口・降車口ともに実測にほぼ一致する人数を推定できること、乗車区間別乗降者数についてもハンガリアン法を用いた軌跡間マッチングにより18ペアの正解を得て、F 値 0.74 程度の精度で OD を推定できることを示し、BAIRD-OD の基本的な有効性を確認した。

第4章では、BAIRD-OD を他路線の動画に適用した際に個人識別精度が十分に得られなかった問題に対し、人物検出・追跡の観点から精度向上手法を検討・提案した。まず、ID スイッチと軌跡断片化という二つの代表的な誤りパターンを整理し、End-to-End 型の人物追跡モデルと Tracking-by-Detection (TbD) + ヒューリスティクス の利点・欠点を比較した。そのうえで、本研究のように利用者の移動経路やカメラ設置位置などドメイン知識が豊富で、計算資源にも制約があるバス車

内環境においては、TbD を基盤としつつ検出矩形のフィルタリングや軌跡結合などの後処理を組み合わせるアプローチが現実的であることを示した。具体的には、扉周辺以外の矩形や極端に小さい矩形の除去、短軌跡の削除、空間的一貫性に基づく軌跡の結合といった処理を導入し、人物追跡器 MPNTrack をバス車内データで再学習することで、ID スイッチと軌跡断片化の双方を抑制した。実運行 4 便に対する評価の結果、これらの処理を全て導入した場合には、バス停別乗降者数の誤差率が乗車口で 41.7 % から 13.3 %、降車口で 29.5 % から 5.24 % へと大きく低減し、アブレーションスタディでも各処理が誤差率や ID スイッチ数の削減に寄与していることを確認した。

第 5 章では、BAIRD-OD における乗降判定処理の精度向上を目的として、ルールベース判定の限界を明らかにし、深層学習を用いた乗降判定器を提案した。まず、扉の開閉タイミングと矩形の位置・移動方向に基づく既存のルールベース手法では、光条件の変動や乗客の立ち振る舞いの多様性、短い停車時間などにより誤判定が生じやすいことを分析した。そこで、人物検出器 YOLOv10 により抽出した人物矩形系列から特徴量を構成し、「乗車」「降車」「不適人物」を多クラス分類する DNN モデルを構築し、アノテーション済みデータを用いて学習した。検出精度の比較実験により、YOLOv10 がバス車内データに対して十分な人物検出性能を有することを確認した上で、提案 DNN の乗降判定精度を評価した結果、乗車口で 95.2 %、降車口で 97.4 % と高い正解率を達成した。さらに、BAIRD-OD 全体に組み込んでバス停別乗降者数を推定したところ、降車口ではルールベースに比べて誤差率を 6.59 % から 0.803 % へと大きく削減し、乗車口についても検出器と判定器の組み合わせとして YOLOv10 + DNN が最も誤差率の低い構成となることを示した。これにより、乗降判定モジュールの学習ベース化がシステム全体の精度向上に有効であることを明らかにした。

以上のことから、本論文では、路線バス車内の既設カメラ映像を用いたバス OD データ収集について、第一に、実運用上の制約条件に整合したシステム設計とモジュール構成 (BAIRD-OD) を提示し、動画取得から個人識別・乗降判定・同一判定を経て発生交通量・集中交通量・分布交通量を推定するまでの一連の処理を形式的に定式化した。第二に、人物検出・追跡および乗降判定に対して、バス車内というドメインの特性を踏まえたフィルタリングや軌跡結合、DNN による乗降分類といった具体的な精度向上手法を提案し、アブレーションスタディを通じてそれぞれの効果を定量的に示した。第三に、実際の路線バス映像から構築したアノテーション付きデータセットを用いて、バス停別・区間別の乗降者数がどの程度の精度で推定可能かを示し、個人認証レベルの精度を要求しないという前提の下で、運行計画の見直しに耐えうるレベルの OD データが得られることを実証した。これらの成果は、カメラ映像を用いた OD データ収集の研究領域において、実環境を想定した包括的な評価事例として学術的・実務的な意義を持つ。

今後の課題としては、まず要素技術レベルでのさらなる精度向上が挙げられる。本論文では人物検出・追跡に対して TbD + ヒューリスティクスを基盤とした構成を採用したが、第 3 章および第 4 章の要素技術の検討で挙げたように、頭部検出や骨格推定、Transformer ベースの End-to-End 型追跡モデル、より高性能な Re-ID モデル、深度センサや複数カメラなどのマルチモーダル化といった候補手法も存在する。今後は、これらの手法をバス車内というドメインの制約とコスト構造を踏まえて部分的に取り入れ、提案したフィルタリングや軌跡結合処理と組み合わせることで、ID スイッチや軌跡断片化を一層抑制し、OD 推定精度の向上を図る必要がある。乗降判定についても、本論文で構築した DNN モデルを出発点として、長期依存関係を考慮する時系列モデルやシーン全体のコンテキストを利用するモデルへの拡張、少量のラベル付きデータと大量の未ラベルデータを併用する半教師あり学習などを検討することで、さらなる汎化性能の向上が期待できる。社会実装の観点からは、複数路線・長期にわたる運行データに BAIRD-OD を適用し、ダイヤ改正や減便・増便といった施策と組み合わせた効果検証を行うこと、交通系 IC カードデータなど他のデータソースとの連携によって推定結果の検証と補完を行うこと、プライバシー保護の観点から撮影・保存・解析の運用ルールやガイドラインを具体化することが重要である。本論文で構築した BAIRD-OD とその精度向上手法は、路線バスに限らず、多様な移動体内空間における行動計測と OD データ獲得の基盤技術として発展する可能性を持っており、今後の研究と社会実装を通じて、持続可能な地域公共交通の維持・改善に貢献することが期待される。

謝辞

本研究を進めるに当たり、様々なご指導を頂きました北海道大学大学院情報科学研究院調和系工学研究室の川村秀憲教授，山下倫央准教授，横山想一郎助教に深謝申し上げます。

実験にご協力いただいた株式会社シーズ・ラボ，株式会社じょうてつの関係者の皆様に感謝いたします。

小学校，中学校，高校，大学，サークル，アルバイト，ゼミ，研究室など，ここまで支えてくれた多くの友人皆様に感謝いたします。特に MATCH BOX の皆様，大変お世話になりました。

参考文献

- [1] 国土交通省. 「片利用共通接続システム」の構築に関する方向性（概要）. Technical report, 国土交通省, 3 2017. 別紙資料.
- [2] 国土交通省総合政策局. 参考資料 3 利用者アンケート調査結果. Technical report, 国土交通省, 2022.
- [3] 国土交通省. 交通系 ic カードの普及・利便性拡大に向けた検討会 とりまとめ. <https://www.mlit.go.jp/common/001097000.pdf>, 7 2015. 参照日: 2025-11-21.
- [4] 国土交通省. 月例経済報告関係資料（交通系 IC カードに関する資料）. Technical report, 2004. 図 7 「長崎県バス各社における導入費用」.
- [5] バス用ドライブレコーダ車両搭載のガイドライン. バス部会・資材部会 審議.
- [6] 小田浩幸, 児玉陽太, 小岩弘樹. 地域公共交通事業における収益性と効率性に関する調査研究Ⅱ. 国土交通政策研究 第 138 号, 国土交通省 国土交通政策研究所, 3 2017. 本編 PDF: kkk138_1.pdf（約 99p）, 資料編 PDF: kkk138_2.pdf（約 199p）.
- [7] 交通系 IC カードの普及・利便性拡大に向けた検討会. 交通系 ic カードの普及・利便性拡大に向けた検討会 とりまとめ. Technical report, 国土交通省, 7 2015.
- [8] 国土交通省 都市・地域整備局 都市計画課都市計画調査室. 都市・地域総合交通戦略及び特定の交通課題に対応した都市交通計画検討のための実態調査・分析の手引き. PDF, 7 2010. PT 調査・道路交通センサス OD 等の活用手法を解説.
- [9] 国土交通省. 国土交通白書 2021 第 2 章 第 1 節「公共交通への深刻な影響」, 2021. p.35 – 36 に「2019 年時点で全国のパス事業者の約 7 割が赤字、地方圏では約 9 割が赤字」との記述あり。

- [10] 国土交通省. 令和7年版 国土交通白書 概要, 2025. p.4 に「バスの運転手数は2030年に必要人員全体の28%（約3.6万人）不足」等の将来推計を記載。資料出所：日本バス協会。
- [11] 国土交通省. 国土交通白書 2020 第3節「地域における移動手段を確保するために」－1 現状や将来予測に基づく課題, 2020. 自動車運転の職業の有効求人倍率は全職業平均の2倍以上、道路旅客運送業（バス・タクシー）は55歳以上が61.2%と高齢化—今後の担い手不足が懸念される旨を記載。
- [12] 国土交通省関東運輸局. 関東運輸局管内におけるバス運転者不足問題を踏まえた地域公共交通の確保維持に関する調査, 3 2025. 自治体アンケート：2021年4月～2024年12月に減便・廃止あり65.6%、2025年1月以降に減便・廃止予定あり17%等。改善基準告示改正以降（2024年4月～12月）に減便・廃止の回答が最多。
- [13] 交通分野におけるデータ連携の高度化に向けた検討会 参考資料3：利用者アンケート調査結果. Technical report, 国土交通省 総合政策局, 3 2022. 調査期間：2022-03-07～2022-03-11、n=1457。事務局：国土交通省総合政策局モビリティサービス推進課。
- [14] 交通系 IC カードの普及・利便性拡大に向けた検討会. 交通系 ic カードの普及・利便性拡大に向けた検討会 とりまとめ. Technical report, 国土交通省, 7 2015.
- [15] Pedram Saeidizand. Global bus survey, 5 2019.
- [16] 辰巳浩, 堤香代子, 吉城秀治, 末次真梨. 路線バスのローカル性に関する基礎的研究. Poster (PDF), 2014.
- [17] 用語集. <https://www.kkr.mlit.go.jp/plan/pt/data/glossary.html>. パーソントリップ調査で使用する用語の説明. (参照 2025-11-21).
- [18] 兵庫県まちづくり部都市計画課. パーソントリップ調査. https://web.pref.hyogo.lg.jp/ks21/wd21_000000018.html, 2024. 閲覧日: 2025年11月21日.
- [19] 国土交通省都市局都市計画課都市計画調査室. パーソントリップ調査. https://www.mlit.go.jp/toshi/tosiko/toshi_tosiko_tk_000031.html, 2024. 閲覧日: 2025年11月21日.
- [20] 堺市建築都市局交通部交通政策課. パーソントリップ調査に関する用語の説明. <https://www.city.sakai.lg.jp/shisei/toshi/kotsuseisaku/>

- kento/kanrencyosa/persontrip/yogosetsumei.html, 2012. 閲覧日: 2025 年 11 月 21 日.
- [21] 塩土圭介. 今更聞けない!?移動手段ってどんな種類があるの? [一部更新]. <https://kotsutorisetsu.com/20201025-2/>, 2020. 地域公共交通のトリセツ (取組説明書). 初出 2020 年 10 月 25 日, 2022 年 3 月 16 日一部更新. (最終閲覧日: 2025-11-21) .
- [22] 伊豆の国市. 広報いずのくに. <https://www.city.izunokuni.shizuoka.jp/hisyo/shisei/kouhou/kouhou/documents/2210012.pdf>, n.d. 伊豆の国市の市政広報誌 PDF (最終閲覧日: 2025-11-21) .
- [23] 国土交通省都市局都市計画調査室. パーソントリップ調査におけるビッグデータの活用について. https://www.soumu.go.jp/main_content/000588038.pdf, 2018. 閲覧日: 2025 年 11 月 21 日.
- [24] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.
- [25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, 2017.
- [26] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [27] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
- [28] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273–297, 1995.
- [29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [30] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

- [31] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [32] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. IEEE, 2009.
- [33] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, *NIPS*, pp. 91–99, 2015.
- [34] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection, 2015. cite arxiv:1506.02640.
- [35] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott E. Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *ECCV (1)*, Vol. 9905 of *Lecture Notes in Computer Science*, pp. 21–37. Springer, 2016.
- [36] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.
- [37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241. Springer, 2015.
- [38] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.
- [39] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1302–1310, 2017.
- [40] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation, 2019.

- [41] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4489–4497, 2015.
- [42] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 27, pp. 568–576, 2014.
- [43] Gedas Bertasius, Heng Wang, and Lorenzo Torresani. Is space-time attention all you need for video understanding? In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021. TimeSformer.
- [44] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1487–1495, June 2019.
- [45] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric, 2017.
- [46] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. Fairmot: On the fairness of detection and re-identification in multiple object tracking. *arXiv preprint arXiv:2004.01888*, 2020. accessed 2025-11-21.
- [47] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fangjian Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *European Conference on Computer Vision (ECCV)*, 2022.
- [48] V. A. Sindagi and V. M. Patel. A survey of recent advances in crowd counting and density estimation using deep learning. *Pattern Recognition Letters*, Vol. 107, pp. 3–13, 2018.
- [49] 遊馬山田, 聡仁廣森, 弘純山口, 輝夫東野. 路線バスにおけるカメラ画像を用いた OD 計測システムの提案. 研究報告高度交通システムとスマートコミュニティ, pp. 1–8, 2019.
- [50] R. Sutopo, J. M.-Y. Lim, V. M. Baskaran, K. Wong, M. Tistarelli, and H. F. Liau. Appearance-based passenger counting in cluttered scenes with lateral movement compensation. *Neural Computing and Applications*, Vol. 33, No. 23, pp. 16119–16135, 2021.

- [51] F. Li, F. Yang, H. Liang, and W. Yang. Automatic passenger counting system for bus based on RGB-D video. In *Proc. 2nd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2016)*, pp. 209–214. Atlantis Press, 2017.
- [52] Davide Calì, Philipp Matthes, Kristina Huchtemann, Ralf Streblow, and Dirk Müller. CO₂ based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. *Energy and Buildings*, Vol. 86, pp. 395–405, 2015.
- [53] Jin Liu, Yilin Wang, Yan Chen, Xinyu Yang, Xi Chen, and Jinsong Han. WiFi sensing with channel state information: A survey. *IEEE Communications Surveys & Tutorials*, Vol. 22, No. 3, pp. 1963–1988, 2020.
- [54] 裕太秋山, 司工藤. 顔認証を用いた自然な動作での入退室管理. 情報処理学会第 82 回全国大会講演論文集, pp. 599–600, 2020.
- [55] 早耶佳良永, 敦大坪, 大和橋本, 法道廣重, 直之鶴田. プライバシーを考慮した安価でポータブルな入退室者カウント画像処理システムの開発. 情報処理学会第 80 回全国大会講演論文集, pp. 523–523, 2018.
- [56] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. You only learn one representation: Unified network for multiple tasks. *arXiv preprint arXiv:2105.04206*, 2021.
- [57] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015.
- [58] Guillem Brasó and Laura Leal-Taixé. Learning a neural solver for multiple object tracking. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6247–6257, June 2020.
- [59] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé. Mot20: A benchmark for multi object tracking in crowded scenes, March 2020.
- [60] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016.

- [61] Harold W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, Vol. 2, No. 1-2, pp. 83–97, 1955.
- [62] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, Vol. 5, No. 1, pp. 32–38, 1957.
- [63] K. Kadirvel and K. Balamurugan. Method for solving hungarian assignment problems using triangular and trapezoidal fuzzy number. *International Journal of Engineering Research and Applications*, Vol. 2, No. 5, pp. 399–403, 2012.
- [64] Jenny Seidenschwarz, Guillem Brasó, Víctor Castro Serrano, Ismail Elezi, and Laura Leal-Taixé. Simple cues lead to a strong multi-object tracker. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13813–13823, June 2023.
- [65] Leandro Di Bella, Yangxintong Lyu, Bruno Cornelis, and Adrian Munteanu. Hybridtrack: A hybrid approach for robust multi-object tracking. *arXiv preprint arXiv:2501.01275*, 2025.
- [66] Pavel Tokmakov, Jie Li, Wolfram Burgard, and Adrien Gaidon. Learning to track with object permanence. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10860–10869, October 2021.
- [67] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. *arXiv preprint arXiv:2110.06864*, 2021.
- [68] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*, 2022.
- [69] Glenn Jocher. Ultralytics yolov5, 2020.
- [70] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. YOLO by Ultralytics, 1 2023.
- [71] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, Vol. 323, No. 6088, pp. 533–536, 1986.

- [72] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*, 2024.
- [73] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Vol. 15 of *Proceedings of Machine Learning Research*, pp. 315–323. PMLR, 2011.
- [74] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, NY, 2006.
- [75] Yann LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. Efficient backprop. In Genevieve B. Orr and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade*, Vol. 1524 of *Lecture Notes in Computer Science*, pp. 9–50. Springer, Berlin, Heidelberg, 1998.