

修士論文

可視光映像と遠赤外線映像の融合による 気象変化の影響を最小化した歩行者数推定手法の提案

福田 匠

奈良先端科学技術大学院大学

先端科学技術研究科

情報理工学プログラム

主指導教員: 藤川 和利 教授

(情報科学領域)

令和5年1月31日提出

本論文は奈良先端科学技術大学院大学先端科学技術研究科に
修士(工学) 授与の要件として提出した修士論文である。

福田 匠

審査委員：

藤川 和利 教授	(主指導教員, 情報科学領域)
笠原 正治 教授	(副指導教員, 情報科学領域)
安本 慶一 教授	(副指導教員, 情報科学領域)
新井 イスマイル 准教授	(副指導教員, 情報科学領域)

可視光映像と遠赤外線映像の融合による 気象変化の影響を最小化した歩行者数推定手法の提案*

福田 匠

内容梗概

都市計画の中で行われる歩道の整備ではどのような人がどのように歩道を利用しているかを調査する必要がある。歩道の利用状況の指標となる歩行者数の計測では、可視光カメラで撮影した映像をもとに歩行者数を推定することが一般的である。しかし、可視光映像を用いた歩行者数推定は雨天時や夜間のような照度が低い状況で精度が低下することが考えられる。遠赤外線カメラは、照度に関係なく歩行者や周辺の温度を撮影することができるため、遠赤外線映像を用いることでこのような状況でも高精度に、歩行者数を推定できる可能性がある。可視光映像と遠赤外線映像のそれぞれを単独で用いて、歩行者数を推定する予備実験を実施した結果、予想に反して、遠赤外線映像を用いた場合の推定精度より可視光映像を用いた場合の推定精度が高くなった。一方で、可視光映像を用いた場合に数えることができなかった歩行者を、遠赤外線映像を用いた場合は数えることができた場合があった。そこで本研究では、可視光映像と遠赤外線映像を組み合わせることで、可視光映像のみを利用するよりも高精度に歩行者数を推定することを目指す。歩行者数の推定には膨大な数の学習データが必要であるが、歩行者数推定地点ごとにモデルを作成するのは非常にコストがかかるため、事前学習済みの物体検出モデルを利用することが求められる。可視光映像と遠赤外線映像を組み合わせる手法として、単純に2つの映像を入力しようとする場合、事前学習済みの物体検出モデルが利用できない。提案手法では可視光映像と遠赤外線映像の特徴を同時に利用して歩行者数を推定し、事前学習済みの物体検出モデルを利用可

*奈良先端科学技術大学院大学 先端科学技術研究科 修士論文, 令和5年1月31日.

能にするために Densefuse を用いた映像の融合を行う。評価の結果，提案手法は可視光映像のみを利用する場合に比べて最大 1.5%，遠赤外線映像のみを利用する場合に比べて最大 11.1%の歩行者推定誤差を削減できた。

キーワード

スマートシティ，画像処理，歩行者追跡，遠赤外線映像

Estimation of time- and weather-tolerant pedestrian counts using fused visible and far-infrared video*

Takumi Fukuda

Abstract

Developing sidewalks in urban planning requires investigating sidewalk usage. Visible video is commonly used to estimate the head count using the sidewalk. However, the accuracy of head count estimation using visible video is expected to decrease at night and in the rain. This is because pedestrians are not lit. A far-infrared camera can capture the temperature of pedestrians and the vicinity despite lighting conditions, and the far-infrared video can accurately estimate the head count. A preliminary experiment was conducted to estimate the head count using both visible and far-infrared video alone. The results were contrary to expectations, and the estimation accuracy using visible-light images was higher than that using far-infrared images. On the other hand, pedestrians who could not be counted using visible images were sometimes counted using far-infrared images. Therefore, this study aims to estimate the head count more accurately than using only visible video by combining visual and far-infrared videos. Since estimating the head count requires a huge amount of training data and creating a model for each count point is very costly, it is necessary to use a pre-trained object detection model. The pre-trained object detection model does not allow

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, January 31, 2022.

the input of visible and far-infrared video together. The proposed method uses Densefuse to fuse the visible and far-infrared video features.

Fused video can use pre-trained object detection models. The evaluation results showed that the proposed method reduced the error rate up to 1.5% compared with using only visible video. In addition, the head count error rate was reduced up to 1.5% compared with using only far-infrared video.

Keywords:

Smart city, Image processing, Pedestrian tracking, Far-infrared video

目次

1. はじめに	1
2. 歩行者数推定及び人物検出・追跡の関連研究	3
2.1 様々な歩行者数推定手法	3
2.2 カメラ映像を用いた歩行者数推定のための基礎技術	5
2.2.1 物体検出手法	5
2.2.2 物体追跡手法	10
2.3 カメラ映像を用いた歩行者数推定に関する研究と課題	12
3. 本研究で利用するデータセット	16
3.1 撮影機材	16
3.1.1 撮影用カメラ	16
3.1.2 制御用コンピュータ	19
3.1.3 デジタルサイネージ	19
3.2 撮影環境	24
3.3 データセット	25
4. 予備実験	28
4.1 可視光映像を用いた場合に推定精度が下がる要因の仮説	29
4.1.1 環境要因	29
4.1.2 人的要因	30
4.2 遠赤外線映像を用いた場合に推定精度が下がる要因の仮説	30
4.3 物体検出器の転移学習	32
4.4 物体追跡アルゴリズムの追跡パラメータ	33
4.5 歩行者数推定誤差率の計算方法	34
4.6 可視光映像を用いた場合に推定精度が下がる要因の検証	34
4.7 遠赤外線映像を用いた場合に推定精度が下がる要因の検証	41
4.8 時間と天気ごとの誤差率	44
4.9 可視光映像と遠赤外線映像の補完関係	46

5. 可視光映像と遠赤外線映像を融合した映像による歩行者数推定手法	48
6. 提案手法の評価	52
6.1 可視光映像と融合映像を用いた推定結果の比較	53
6.2 遠赤外線映像と融合映像を用いた推定結果の比較	53
6.3 3つの映像を用いた場合の時間と天気ごとの誤差率	58
7. 考察	61
7.1 気温 24°C 未満の場合	61
7.2 気温 24°C-30°C の場合	62
7.3 気温 30°C を超過する場合	62
7.4 今後の展望	63
8. 終わりに	65
謝辞	66
参考文献	67
付録	71
A. 撮影のアナウンスに利用した準備物	71

目 次

1	カメラ映像を用いた歩行者数推定の方法	6
2	YOLO によるグリッド毎の物体検出	8
3	YOLO の出力例	8
4	bounding box が示す値	9
5	correct box と predict box のずれと IoU の計算	11
6	MOTA を構成する追跡要素：FP, FN, IDSW	13
7	物体検出器 (YOLO) と追跡アルゴリズム (BYTE) の出力例	14
8	実験機材の設置概要図	17
9	可視光カメラと遠赤外線カメラをそれぞれ雲台に固定した様子	18
10	撮影映像を送信する流れ	19
11	サイネージ上部に設置する箱のモデル	21
12	GAT が透過する波長帯と透過率	22
13	デジタルサイネージの様子	23
14	三宮山手バス停の周辺環境	24
15	撮影環境と撮影可能エリアの様子	25
16	作成したデータセットの一例	27
17	可視光映像と遠赤外線映像を利用した歩行者数推定手法の概要図	28
18	遠赤外線映像を利用した場合の温度差による影響の想定図	31
19	遠赤外線映像のアノテーションの様子	32
20	予備実験で用いる推定値と誤差率を加えた結果の一例	35
21	可視光映像を利用した場合の昼間と夜間の誤差率の箱ひげ図	36
22	可視光映像をそれぞれの時間と天気を利用した例	37
23	可視光映像を利用した場合の天気ごとの昼間と夜間の誤差率の箱ひげ図	38
24	自転車の追跡失敗例	39
25	夜間の雨天時における推定失敗例	40
26	遠赤外線映像を利用した場合の気温と誤差率のグラフ	42
27	遠赤外線映像を3つの気温で利用した例	43

28	気温ごとの遠赤外線映像本数	45
29	夜間の雨天時における検出結果例	47
30	気温 27°C における検出結果例	47
31	融合映像を利用した歩行者数推定手法の概要図	49
32	DenseFuse のアーキテクチャ	50
33	可視光映像と遠赤外線映像を元にした Densefuse の出力結果例 . .	51
34	本実験で用いるデータ	52
35	可視光映像と融合映像を利用した場合の昼間と夜間の誤差率の箱 ひげ図	54
36	昼間・雨天時の可視映像	55
37	昼間・雨天時の融合映像	56
38	可視光映像と融合映像を利用した場合の比較例	57
39	可視光映像と融合映像の時間ごとの誤差率	59
40	遠赤外線映像と融合映像を利用した場合の比較例	60
41	気温 30°C 超過時の映像 (2868–3586 フレーム間)	64
42	バス停に設置したステッカー	71

表 目 次

1	歩行者数推定手法の比較: 高評価 ○ > △ > × 低評価	3
2	MS COCO のデータセットのカテゴリ番号とカテゴリの対応付け の一例	7
3	カメラの仕様	17
4	撮影用窓に利用するシートの比較: 高評価 ◎ > ○ > △ > × 低評価	22
5	気象庁のデータと検証用データセットで用いる “Weather” の対応表	26
6	転移学習実行時のパラメータ	33
7	物体追跡に利用したパラメータ	34
8	歩行者数推定の誤差率	44
9	可視光映像と遠赤外線映像, 融合映像を利用した場合の誤差率 . .	58

10	気温 24°C 未満の場合の誤差率	61
11	気温 24°C–30°C の場合の誤差率	62
12	気温 30°C を超過する場合の誤差率	63

1. はじめに

現在の都市計画では経済成長や環境保全・居住性の向上のために歩道の整備が重要視されている。例として吉村らは歩行者天国化が周辺小売店の収益増加に影響を及ぼすことを示した [1]。Blaga らは歩行空間の整備が環境や文化・経済など幅広いメリットを得ることができ、都市計画にも積極的に組み込むべきだという主張をしている [2]。また、Shuhana らは住みやすい空間と持続可能な環境を促進するために、居住性の構成要素の一部として歩行空間の必要性を述べている [3]。

歩道の整備に必要な歩道の利用状況調査の一つとして歩行者の計測が行われている。歩行者の計測には、可視光映像から人物を検出し追跡することが一般的である。歩道の利用状況調査を行った報告 [4] では、可視光カメラが採用された CCTV¹カメラを利用して朝の時間帯に歩行者数推定を行った結果、±5%の精度を達成したカメラは全体の1%であった。本研究では歩道の利用状況調査における、歩行者数推定精度が低い課題として下記の2点に着目した。

1. 歩行空間を高解像度に撮影するカメラが少ない点

歩行者を高解像度に撮影できない場合は、歩行者の検出精度が低下する。検出精度の低下により歩行者の追跡漏れが発生する可能性がある。

2. 雨天時や夜間のような照度が低い状況での歩行者数推定が未調査である点

可視光カメラを用いた歩行者数推定の特性上、撮影範囲に十分な照度がなければ歩行者を検出できない。また、道路上のカメラを利用しているため雨天による傘の影響により歩行者を撮影できなかった可能性がある。

1つ目の課題を解決するためには、歩道を高精度に撮影した映像の利用が必要である。2つ目の課題を解決するためには、照度に関係なく歩行者や周辺の温度を撮影した遠赤外線映像を利用することで、歩行者数を推定可能だと考える。

本研究でははじめに、可視光映像と遠赤外線映像を利用した場合の歩行者数推定精度を調査・計算し、それぞれの映像を利用する場合の推定精度低下要因を調査した。次に、調査結果を元に可視光映像を利用した歩行者数推定の欠点を補える

¹closed-circuit television: ケーブルで結ばれた道路上カメラと交通管制室などのモニター間だけの閉じた回路のテレビジョン

ように遠赤外線映像を利用する手法を提案した。歩行者数の推定には人物検出に必要な膨大な数の学習データが必要であり、歩行者数推定地点ごとにモデルを作成する必要がある。都市計画における運用上、推定地点ごとにモデルを作成するのは非常に計算コストがかかるため、事前学習済みの物体検出モデルの利用が求められる。事前学習済みの物体検出モデルは、1400万枚以上の画像データを持つImageNET [5]のような学習データを利用しているため、低い計算コストで高い検出性能を発揮できる。しかし、単純に可視光映像と遠赤外線映像の2つを入力として歩行者数推定を行おうとした場合に、事前学習済みの物体検出モデルを利用できない。そこで本研究では可視光映像と遠赤外線映像の両方の特徴を持つ映像を作成し、1つの映像を入力にすることで事前学習済みの物体検出モデルを利用可能な手法を提案することで、歩行者数推定の精度向上を目指す。

本研究による貢献を以下にまとめる。

- 可視光映像を用いた歩行者数推定は時間や天気、歩行者の服装の影響により推定精度が下がることを確認した。
- 遠赤外線映像を用いた歩行者数推定は特定の気温で推定精度が下がることを確認した。
- 可視光映像と遠赤外線映像を融合させた映像を、歩行者推定に用いることで、上記2つの推定精度の減少を抑制した。

本論文の構成は次の通りである。2章では歩行者数推定手法の比較を行い、カメラ映像を用いた歩行者数推定のための基礎技術について述べたあとに、課題を整理する。3章では実験機材と実験環境、歩行者数推定精度に利用したデータセットについて述べる。4章では事前実験として、可視光映像と遠赤外線映像を利用した場合の歩行者数推定精度低下要因を調査する。5章では提案手法となる、可視光映像と遠赤外線映像の融合手法とそれに伴う実験のフローを述べる。6章では実験の結果を示し、結果を考察する。最後に、7章では本論文のまとめを述べる。

2. 歩行者数推定及び人物検出・追跡の関連研究

本章でははじめに無線タグや Wi-Fi を用いた歩行者数推定手法を比較し，カメラ映像を利用する理由を述べる．次にカメラ映像を利用した歩行者数推定に必要な基礎技術を紹介する．最後にカメラ映像を用いた歩行者数推定に関する研究を紹介し，課題を整理する．

2.1 様々な歩行者数推定手法

本節では無線タグ，Wi-Fi&BLE，レーザスキャニング，3D LiDAR，カメラ映像の5種類の歩行者数推定手法を，精度，移動方向，設置コスト，計算コスト，計測対象の変更という5つの観点から比較する．移動方向と計測対象の変更は，歩道の利用状況をより詳細に分析する際に，拡張性がある手法か比較するために比較の観点としている．計測対象の変更は，歩行者だけではなく自転車やペット等を推定可能かを示す．プライバシーは，計測の際にプライバシーを侵害する可能性があるかを示す．

表 1: 歩行者数推定手法の比較: 高評価 ○ > △ > × 低評価

計測に利用するもの	無線タグ	Wi-Fi & BLE	レーザ	3D LiDAR	カメラ映像
精度	○	△	△	○	○
移動方向	○	×	○	○	○
設置コスト	×	○	○	○	○
計算コスト	○	○	○	×	○
計測対象の変更	○	×	×	○	○
プライバシー	○	○	○	○	×

無線タグを用いた歩行者数推定手法は，歩行者に RFID²タグを持たせ，RFIDリーダーに近づいた時に ID を読み取ることで歩行者をカウントする．この手法は，歩行者全員が無線タグを所持していれば高精度に歩行者数を推定可能であり，RFIDリーダーの複数設置によりタグの移動方向も推定可能である．さらに無線

²Radio Frequency Identification

タグを用いた手法は計算コストが低く、すべての無線タグと計測対象を紐付けることができれば、様々な移動物体にも適用可能である。しかし設置コストの面では、現実的に歩行者全員へのタグの配布は不可能であることから、歩道の歩行者数推定にこの手法を利用することは難しい。

多くの人がスマートフォン等のデバイスを所持している現在、それらのデバイスが定期的に発する Bluetooth のビーコンと Wi-Fi プローブ要求の取得により歩行者推定を行う手法 [6] では、設置コストはパケット収集デバイスのみである。さらにパケットのキャプチャとその解析という計算コストで済み、混雑度の推定に多く用いられる。しかし、デバイスを持たない歩行者も存在するため、高精度に歩行者数を推定することは困難である。また、歩行者の持つ端末が発する無線の RSSI や MAC アドレスを用いるが、現行のスマートフォンは MAC アドレスランダム化によってデバイス情報を隠匿するため移動方向の推定も困難である。

レーザのスキヤニングを用いる手法 [7] では、地面から 20cm ほどの高さでレーザによるスキャンを行うことで歩行者数を推定する。スキャンのサンプリング回数を増加させることによって移動方向の推定が可能である。設置コストの面では、スキャンに必要なデバイスを設置するだけであり、計算コストも低い。しかし大勢の人が通るときなど足元が他の人と重なる場合は歩行者数推定精度が低下する。さらに、スキャン範囲の通行人数を推定するのみであり、通行した物体が歩行者かどうかを判断できないため、自転車やペット等の計測対象を変更できない。

測域センサ (3D LiDAR) は、3次元の点群データとして空間をキャプチャし、歩行者を検出する [8]。歩行者の形状を把握するため、高精度な歩行者数推定が可能であり、撮影範囲内の移動方向を推定可能である。加えて、計測対象を変更する際にはペットや自転車などのデータを学習することにより、対応可能である。その一方で撮影範囲全体の3次元点群データを扱うため、推定に必要な計算量が大きい欠点がある。

カメラを用いた歩行者数推定 [9] の場合は、撮影した映像をもとに歩行者を検出・追跡するため、Wi-Fi や BLE、レーザスキヤニングを用いた場合よりも高精度に歩行者数を推定可能であり、一定のフレームレートがあれば移動方向を推定可能である。設置コストはカメラを設置するのみである。計算量の面では 3D

LiDAR よりも計算量を削減できる．計測対象を変更する場合は，新たな対象の学習データを用意し転移学習させることで対応可能である．しかし，これまでに述べた手法と異なり，プライバシーを侵害する恐れがあるため，個人情報保護のガイドラインを遵守しなければならない．

このように歩行者数推定にカメラ映像を利用することは精度，移動方向の検出，設置コスト，計算コスト，計測対象の変更に利点があるため，本研究ではカメラ映像を用いる．ここまでに比較した結果を表 1 に示す．

2.2 カメラ映像を用いた歩行者数推定のための基礎技術

カメラ映像を用いた歩行者数推定は一般的に図 1 のような物体検出・追跡という手順によって実行される．物体検出とは映像の各フレームごとに，物体の種類とその種類である信頼度を推測し，位置を特定する技術である．物体追跡とは検出された物体を各フレームごとに同じ物体であると紐付ける技術である．物体ごとに固有の ID が付与されるため，ID の数とその映像における歩行者数となる．本節では本研究で利用する歩行者数推定に必要な基礎技術として，物体検出手法と物体追跡手法を紹介する．

2.2.1 物体検出手法

従来の深層学習を用いた物体検出アルゴリズムである R-CNN [10] は学習に時間を要し，メモリの消費量が大きいという課題が存在した．このような問題を解決するために Fast R-CNN [11] や Faster R-CNN [12] が開発された．しかし，背景の誤検出が多いことや検出時間が長いことが課題であった．YOLO [13] は Joseph らが提案した代表的な 1 段階物体認識アルゴリズムである．YOLO は図 2 に示すように画像全体をグリッドに分割し，各領域ごとに物体の種類と位置を求めることによって背景の誤検出の問題を解決し検出時間の短縮に成功した．映像を用いた場合，YOLO による物体検出の出力は図 3 に示すように映像のフレーム数，検出した物体のカテゴリ，カテゴリの信頼度，bounding box (x, y, w, h) の各値によって構成される．検出カテゴリは一般的に，表 2 に示すような，MS COCO [14]

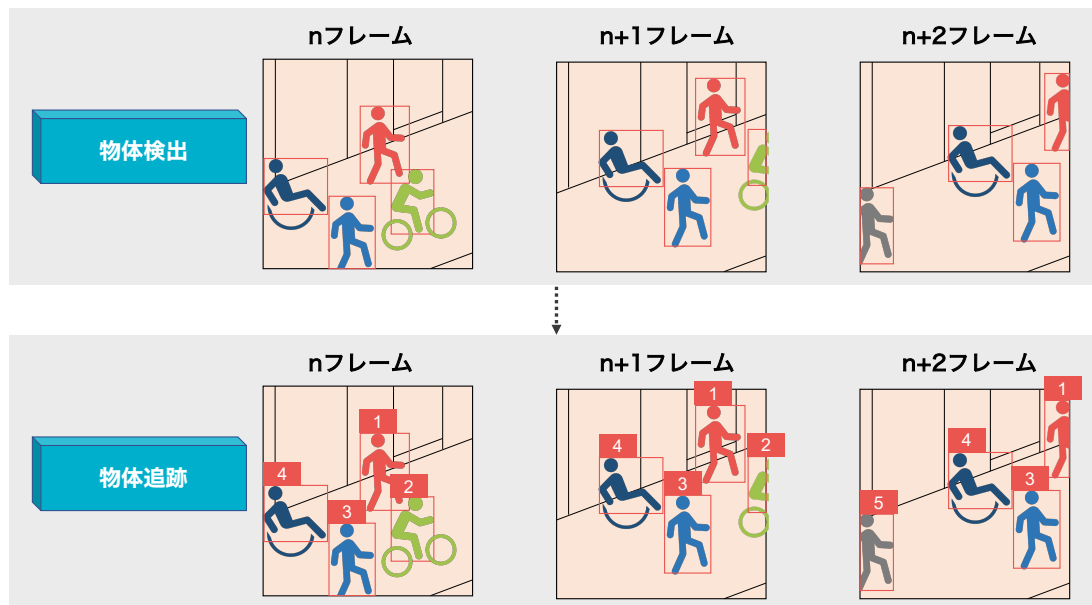


図 1: カメラ映像を用いた歩行者数推定の方法

のデータセットで定義したものが利用されている。bounding box とは、図 4 に示すようにフレーム内の対象物体が、どの位置に存在するかを示す (x, y, w, h) の値である。 (x, y) は画像の左上を始点とするボックスの中心座標、 (w, h) はそれぞれボックスの横幅と縦幅を示す。

表 2: MS COCO のデータセットのカテゴリ番号とカテゴリの対応付けの一例

カテゴリ番号	カテゴリ
1	person
2	bicycle
3	car
4	motorcycle
5	airplane
6	bus
7	train
8	truck
9	boat
10	traffic light
11	fire hydrant
12	street sign
13	stop sign
14	parking meter
15	bench
16	bird
17	cat
18	dog
19	horse
20	sheep

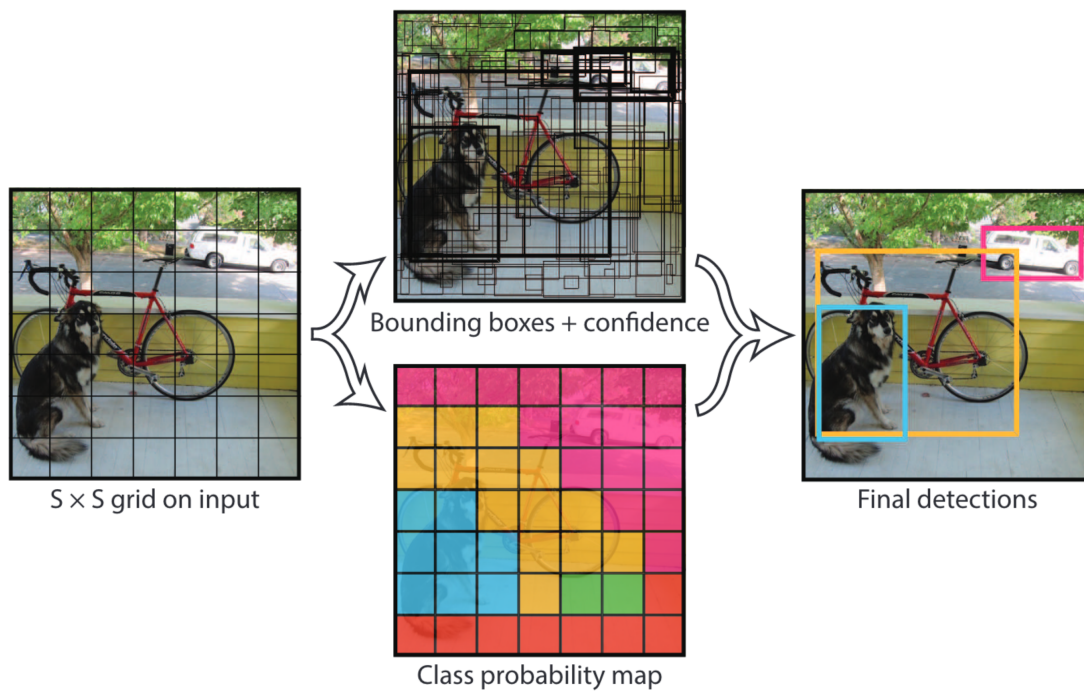


図 2: YOLO によるグリッド毎の物体検出 (文献 [13] より引用)

Frame	Category	Confidence	x	y	w	h
1	1	0.87	60	240	50	200
1	1	0.75	150	280	30	180
2	1	0.85	65	230	40	190
2	1	0.71	160	300	28	170
3	1	0.88	68	220	37	185
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮

図 3: YOLO の出力例

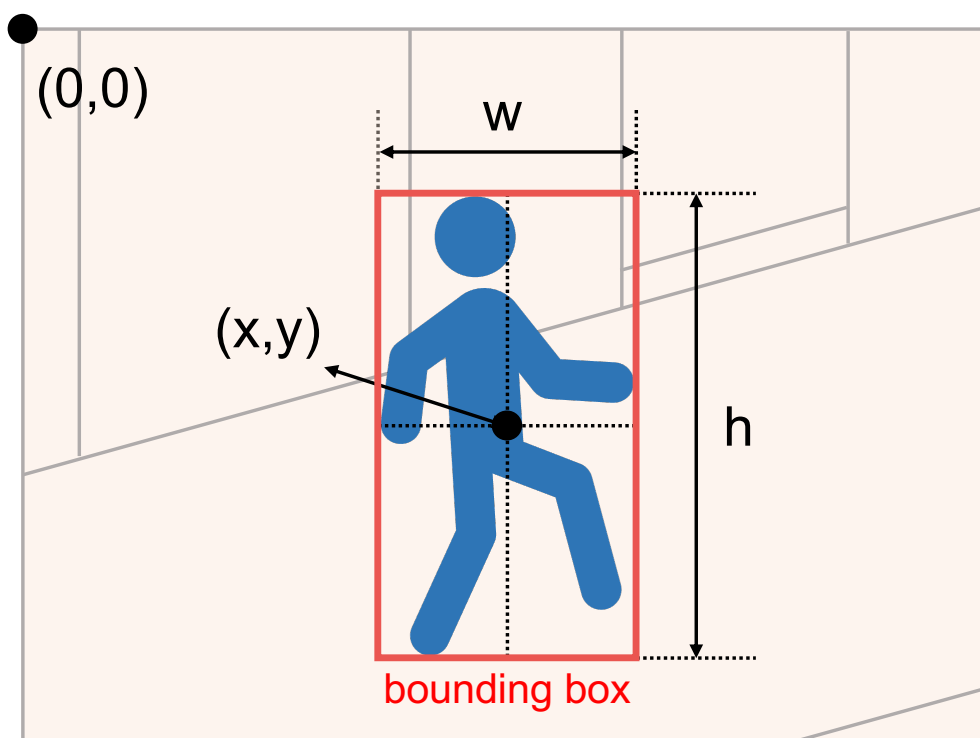


図 4: bounding box が示す値

2.2.2 物体追跡手法

物体追跡手法は一般的に2つの技術的要素によって構成される。

- 物体検出映像のフレーム単位で特定の物体が確認された場合に位置や種類、個数を特定する。
- 物体追跡検出した物体を映像フレーム間で関連付け固有のIDと紐付ける。

物体追跡手法の1つであるSORT [15]では、検出器が出力した bounding box を元に各フレーム間の物体を紐付ける。SORT アルゴリズムでは、bounding box の値を使って式 (1) のような物体のモデルを作成し、次のフレームにおける物体の位置を推定する。

$$\mathbf{X} = [x, y, s, r, \dot{u}, \dot{v}, \dot{s}]^T \quad (1)$$

ここで、 x , y , s , r はそれぞれ推定した bounding box の x 座標, y 座標, 面積, アスペクト比を示す。 \dot{u} , \dot{v} , \dot{s} はそれぞれカルマンフィルタ [16] を使って速度成分を最適に計算される。このモデルの利用により、過去のフレームの動きから次のフレームにおける物体の位置を推定可能となる。推定結果は次のフレームにおける物体検出結果の bounding box との Intersection over Union (IoU) の計算に用いられる。IoU は検出結果と推定結果の重なりを示す指標であり、式 (2) のように計算する。Overlap area は検出結果と推定結果の bounding box の重複面積であり、Total area は2つの bounding box の合計面積である (図5)。

$$\text{IoU} = \frac{\text{Overlap area}}{\text{Total area}} \quad (2)$$

しかし、SORT は高速に動作することを目的としたアルゴリズムのため、オクルージョン³の発生時にはカルマンフィルタの推定外の場合、追跡不可能である。

SORT をベースに提案された DeepSORT [17] は、YOLO などの物体検出器が出力した bounding box と box 内の画像の特徴量を元に追跡を行うため、オクルージョンに強いアルゴリズムである。しかし、検出結果のうち信頼度の低い物体を

³手前にある物体が後ろにある追跡対象を隠す状態のこと

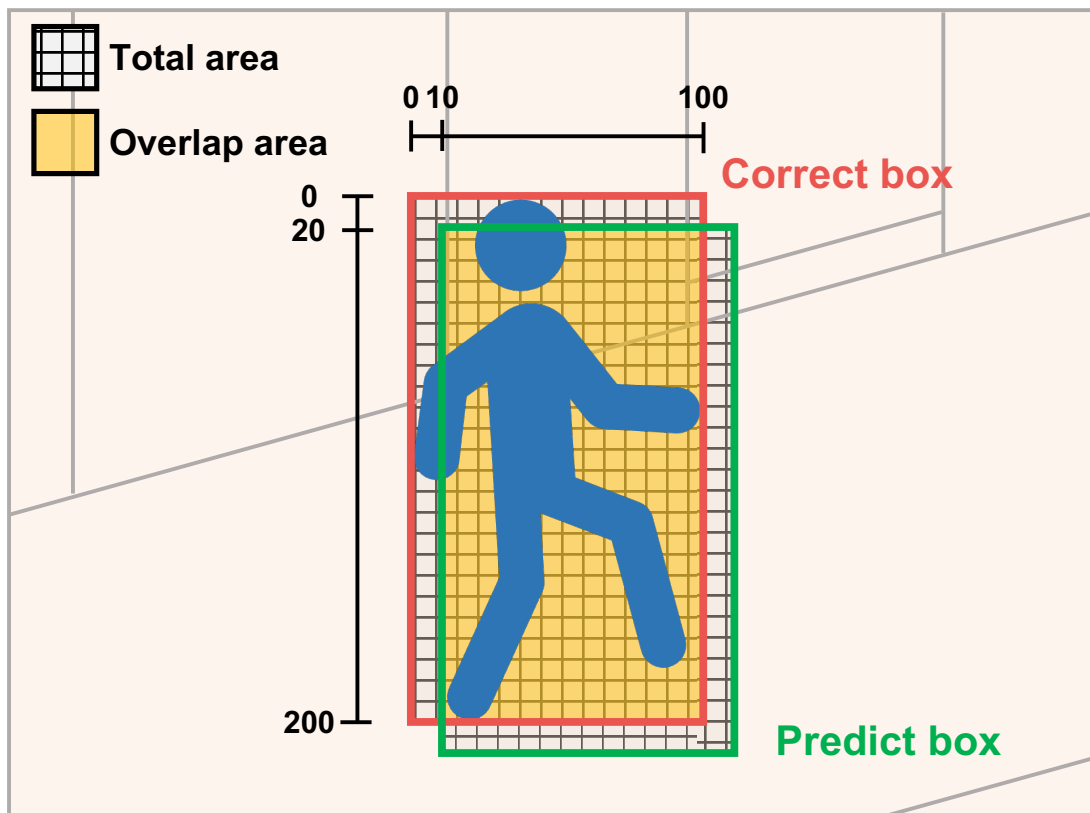


図 5: correct box と predict box のずれと IoU の計算

排除することによって、誤追跡の発生を抑制しているため、小さい物体や照度が低い状況など、物体が確認しづらい時に追跡できない問題があった。

そこで Zhang らは信頼度の低い物体も追跡対象にするようなアルゴリズムとして、BYTE⁴ [18] を提案した。BYTE の手法では物体検出器の検出結果にしきい値を設定することで、高信頼度の検出結果と低信頼度の検出結果に分け、それぞれで異なるアルゴリズムを適用する。このような手法を用いることによって Multi-Object Tracking on MOT17 の分野における MOTA の評価指標で性能は3位となった (2023年1月現在)。

MOTA は式 (3) で表現され、正解データの bounding box と推定モデルの bounding box が近いほど、更に推定モデルのトラックが正解データと一致しているほど高くなる指標である。この数値が高いと高精度に追跡できたといえる。

$$\text{MOTA} = 1 - \sum_t \left(\frac{\text{FP}_t + \text{FN}_t + \text{IDSW}_t}{g_t} \right) \quad (3)$$

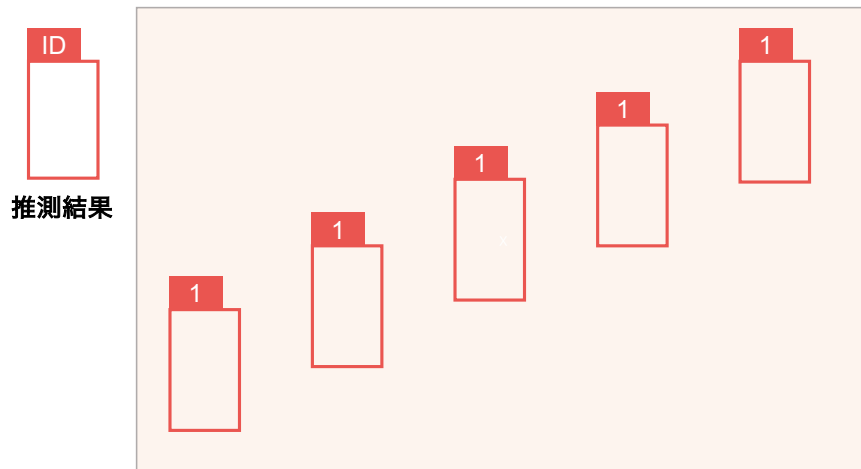
ここで t は t 番目のフレーム、 g_t は映像フレーム内の対象物体の数、 FP_t はフレーム内に対象物体がない場合に追跡した回数 (図 6(a))、 FN_t は対象物体が存在するにも関わらず追跡できなかった回数 (図 6(b))、 IDSW_t は誤った ID にスイッチした回数 (図 6(c)) を示す。

このような評価で高い性能を示す BYTE の出力は、図 7 に示すように物体検出器の出力に ID を追加する。本研究では、ID の数を数えることによって歩行者数を推定する。

2.3 カメラ映像を用いた歩行者数推定に関する研究と課題

物体検出と物体追跡の精度向上によりカメラを用いた歩行者数推定の精度が向上している。実際に、Qiu らは建物の出入り口に設置した可視光カメラを利用し、YOLOv5 [19] 用いた検出器と DeepSORT を用いることで建物に出入りする歩行者数推定を高精度化した [20]。また、Ramana らは、屋内外の群衆の上から撮影

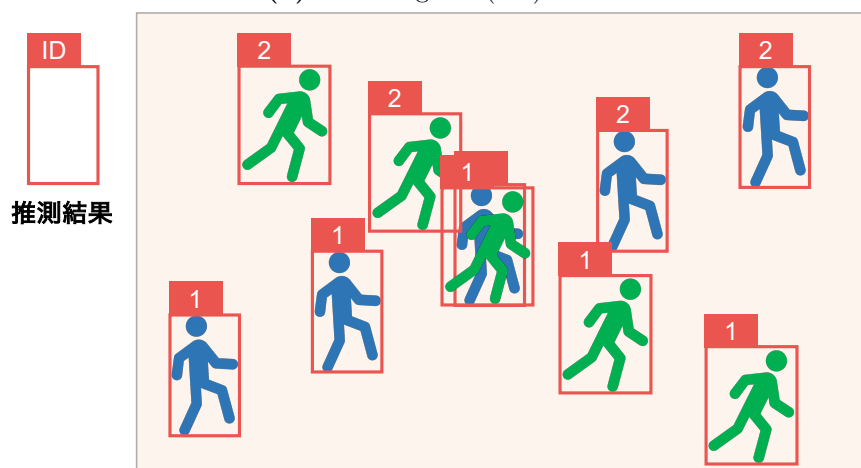
⁴<https://github.com/ifzhang/ByteTrack>



(a) False Positive (FP) の例



(b) False Negative(FN) の例



(c) Id Switch (IDSW) の例

図 6: MOTA を構成する追跡要素 : FP, FN, IDSW

<u>Frame</u>	<u>Category</u>	<u>Confidence</u>	<u>x</u>	<u>y</u>	<u>w</u>	<u>h</u>	<u>ID</u>
1	1	0.87	60	240	50	200	1
1	1	0.75	150	280	30	180	2
2	1	0.85	65	230	40	190	1
2	1	0.71	160	300	28	170	2
3	1	0.88	68	220	37	185	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

図 7: 物体検出器 (YOLO) と追跡アルゴリズム (BYTE) の出力例

した可視光映像を用いて、歩行者頭部の特徴量活用により歩行者数推定を高精度化した [21].

一方、遠赤外線映像を用いた場合は、人物検出精度の向上を目的とする研究が多く、調査した限りでは歩行者数を推定する研究例は少ない。Li らは、YOLOv5 を元にした低解像度かつ小さな物体も検出可能な検出器の作成により、赤外線画像を用いた歩行者検出精度を向上させた [22]。また、可視光映像が利用できない環境として煙が充満するような火災現場に着目した Tsai らは、要救助者の居場所を推定するために赤外線映像を利用することの有効性を検証した [23]。Shen らは、歩行者の垂直方向上に設置した赤外線映像を用いて歩行者を検出し、新たな追跡アルゴリズムを提案することにより、歩行者数や歩行者の移動方向を推定した [24].

従来の可視光映像を用いた歩行者数推定では、照度の高い環境の映像が利用されている。しかし、夜間のような照度の低い環境では、歩行者を検出・追跡できず、推定精度は低下すると考えられる。照度の低い環境での歩行者数推定を高精度に行うためには、照度に関係なく歩行者数を推定できる遠赤外線映像の利用が必要である。しかし、歩行者と背景の温度が近い場合は、歩行者を検出・追跡す

ることができず、推定精度は低下すると考えられる。そこで本研究の予備実験では、可視光映像と遠赤外線映像を利用した場合の推定精度低下要因を調査する。調査した要因への対策として、可視光映像と遠赤外線映像を組み合わせる手法によって、それぞれの映像のみを利用する場合よりも高精度に歩行者数推定が可能になると予想する。

3. 本研究で利用するデータセット

本章では、撮影のための機材と撮影環境を示した後に、本研究で用いるデータセットについて述べる。本研究では、可視光映像と遠赤外線映像の2つの映像から歩行者数推定を実施することを想定している。しかし、公開されている範囲では歩道を対象に撮影された可視光映像と遠赤外線映像は存在しない。そのため、可視光カメラと遠赤外線カメラで歩道を撮影し、歩行者数推定に利用するデータセットを作成した。

なお、撮影の際には実験機材にアナウンス用のステッカー（付録 図 42）を貼り付けており、オプアウト方式で撮影を行っている。撮影した映像におけるプライバシーポリシーは、研究室の Web ページ⁵に記載している。

3.1 撮影機材

本研究で利用する歩道の映像は、図 8 に示すようにデジタルサイネージ上部に設置した、可視光カメラと遠赤外線カメラで撮影する。更に1年中歩道を撮影することができるように、撮影した映像はネットワークを通して、学内に送信できるようにした。本節では、撮影の際に利用した実験機材を紹介する。

3.1.1 撮影用カメラ

本撮影では、可視光カメラ、遠赤外線カメラにそれぞれ、3MP M12 HD 2.8-12mm 1/2.5 IR 1:1.4 CCTV カメラレンズマニュアルズーム用セキュリティカメラ⁶と BOSON 640, 32° (HFOV) 14mm⁷を利用した。それぞれのカメラの仕様を表 3 に示す。

これら2つのカメラは図 9 のように水平方向を調整した上で雲台に固定する。雲台はカメラの撮影方向を上下左右に変更できるため、撮影範囲を容易に変更できるようになる。

⁵<https://inet-lab.naist.jp/announce/>

⁶<https://www.ebay.com/itm/313027731968>

⁷<https://www.flir.jp/products/boson/?model=20640A032-6IARX>

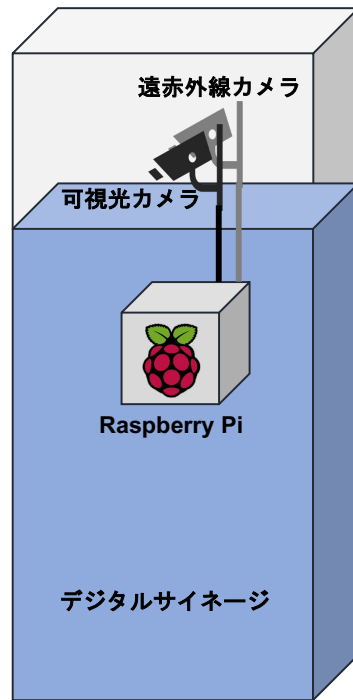


図 8: 実験機材の設置概要図

表 3: カメラの仕様

	可視光カメラ	遠赤外線カメラ
解像度 (pixel)	600×800	514×640
フレームレート (fps)	10	10
水平方向視野角 (°)	31-108	32
インターフェース	USB2.0	USB2.0



図 9: 可視光カメラと遠赤外線カメラをそれぞれ雲台に固定した様子

3.1.2 制御用コンピュータ

使用したコンピュータは Raspberry Pi 4 Model B⁸である。可視光カメラと遠赤外線カメラは本コンピュータに接続されており、同一フレームレートで撮影している。2つのカメラは Raspberry Pi の USB バスパワーによって稼働しているため、許容アンペア数である 1100mA を超えないようにする必要がある。そこでコンピュータのネットワーク接続には、セルフパワーのマルチキャリア M2M ルータ (IO-DATA 社製 UD-LT1)⁹を利用する。このルータには 4G LTE 通信網に常時接続されており、この回線を通して映像を奈良先端科学技術大学院大学内のサーバに一定間隔で転送する。図 10 には撮影映像を送信する流れを示す。Raspberry Pi は撮影した映像を、SSH 経由で大学内に設置されたサーバに転送する。

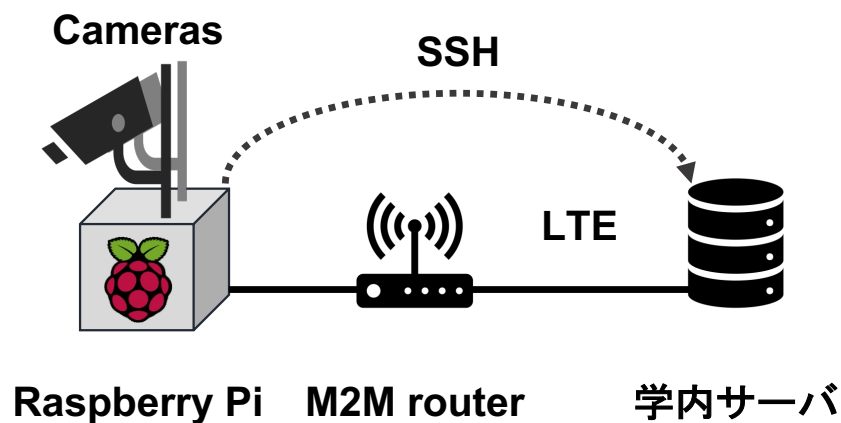


図 10: 撮影映像を送信する流れ

3.1.3 デジタルサイネージ

可視光カメラと遠赤外線カメラ、コンピュータはみなと観光バス¹⁰所有の兵庫県神戸市にある三宮山手バス停のデジタルサイネージ上部に設置する。デジタル

⁸<https://www.raspberrypi.com/products/raspberry-pi-4-model-b/specifications/>

⁹<https://www.iodata.jp/product/lan/appliance/ud-lt1/spec.htm>

¹⁰<https://www.kobe-minato.co.jp/>

サイネージとは公共空間や交通機関などあらゆる場所で、ディスプレイなどの電子的な表示機器を使って情報を発信するメディアの総称である。本撮影で使用するデジタルサイネージは歩道に面した位置に設置されており、サイネージ上部にカメラを設置することによって、成人の一般的な身長より高い約2mの位置から撮影できる。更にコンピュータに対して安定的に電源を供給可能であるため、本実験のような歩道の撮影に適した設備であると言える。本実験ではデジタルサイネージの上部にカメラを設置するために、金属箱を準備した。金属箱の3Dモデルを図11に示す。撮影用窓の大きさは110mm×250mmとなっており、雲台の場所は撮影用窓を開ける面から65mm離れた位置とした。また、底面2つの穴はカメラの配線の考慮および夏場に箱内の高温下を防ぐためにデジタルサイネージ内部に直結できるようにしている。

撮影の際は窓の方向にカメラを向ける必要があり、撮影用窓の穴が開口状態だと粉塵や降水による機器損傷の危険性があるため、可視光と遠赤外線を透過するシートをつけなければならない。本撮影では双方の要件を満たすシートとして旭化成エンジニアリング社製遠赤外線透過透明シート GAT¹¹を利用する。GATは表4のような特徴を持っている。表4にあるようにGATはアクリルや石英ガラス、ポリエチレンと比べて可視光、遠赤外線の両方を透過し、軽量で加工が容易である。図12には、GATが透過する波長帯と透過率を示す。本実験で用いる遠赤外線カメラが集光する遠赤外線の波長は、緑の範囲に示す12 μ mであり、透過率は50%程度であった。

最後に図13には実験機材を格納した後のデジタルサイネージを示す。デジタルサイネージ上部に設置された金属箱の穴にはGATを貼り付け、箱の中から可視光カメラと遠赤外線カメラで撮影を行うことができるようにした。

¹¹<https://www.asahi-kasei.co.jp/advance/jp/gat/index.html>

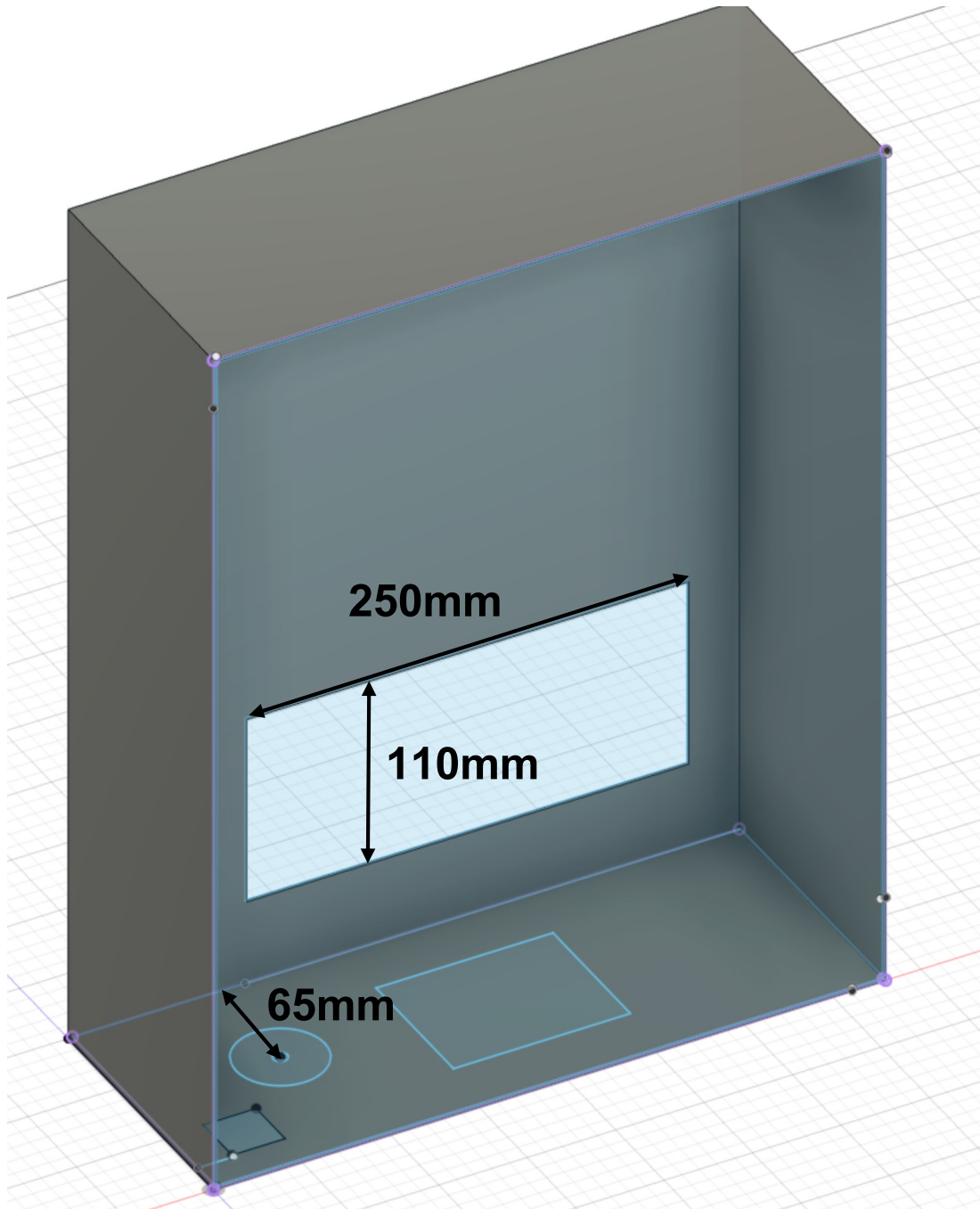


図 11: サイネージ上部に設置する箱のモデル

表 4: 撮影用窓に利用するシートの比較: 高評価 ◎ > ○ > △ > × 低評価

素材	アクリル	石英ガラス	ポリエチレン	GAT
可視光透過率	◎	◎	△	○
赤外線透過率	×	△ (遠赤外線は×)	○	○
コスト	◎	×	◎	○
比重	△	×	◎	◎
加工の容易さ	○	×	◎	○

GAT-10

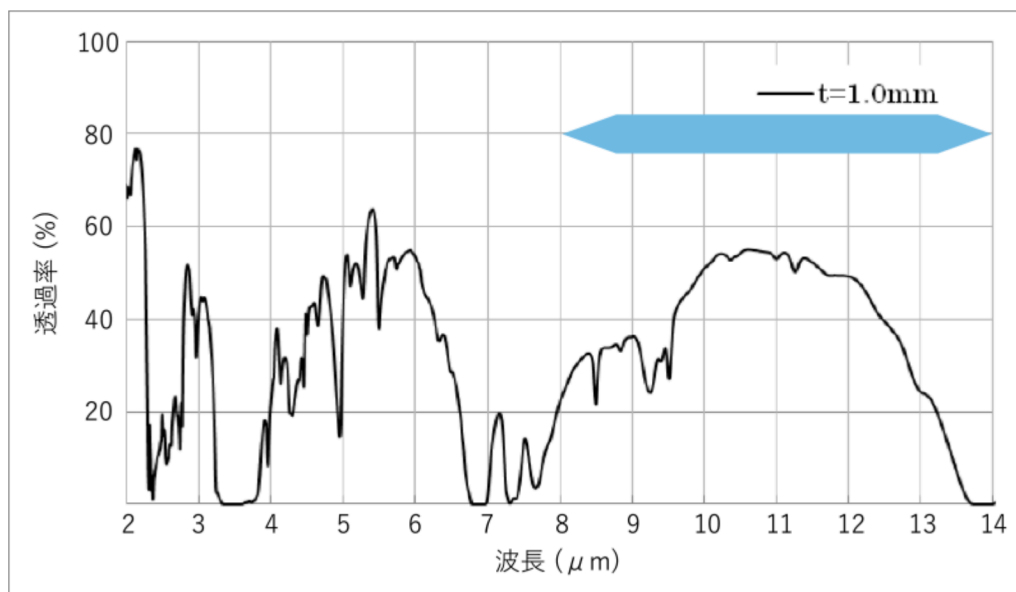


図 12: GAT が透過する波長帯と透過率 (旭化成アドバンス株式会社 Web ページより引用: <https://www.asahi-kasei.co.jp/advance/jp/gat/index.html>)



図 13: デジタルサイネージの様子

3.2 撮影環境

図 14 には撮影場所の周辺地図を示す。三宮山手バス停は1日の乗降客数がおおよそ18万人と兵庫県で最も多い¹²三ノ宮駅の近くに位置するバス停である。周囲には飲食店やコンビニ等が立ち並んでおり、朝には出勤・通学する人、昼には食事に出かける人、夜には買い物に行く人、のように1日を通して数多くの歩行者が通行する。



© OpenStreetMap, <https://www.openstreetmap.org/copyright>

図 14: 三宮山手バス停の周辺環境

¹²<https://opendata-web.site/station/28/>

三宮山手バス停に設置したカメラの撮影範囲とその周辺を図15に示す。歩道の道幅はおよそ5mであり、最大で7人程度が横一列に並んで歩行できる程度だった。デジタルサイネージの近辺には街灯や木があるため撮影画角に入らないように画角の調整を行った。夜間は街灯が点灯し歩行者は安全に移動できる環境であった。

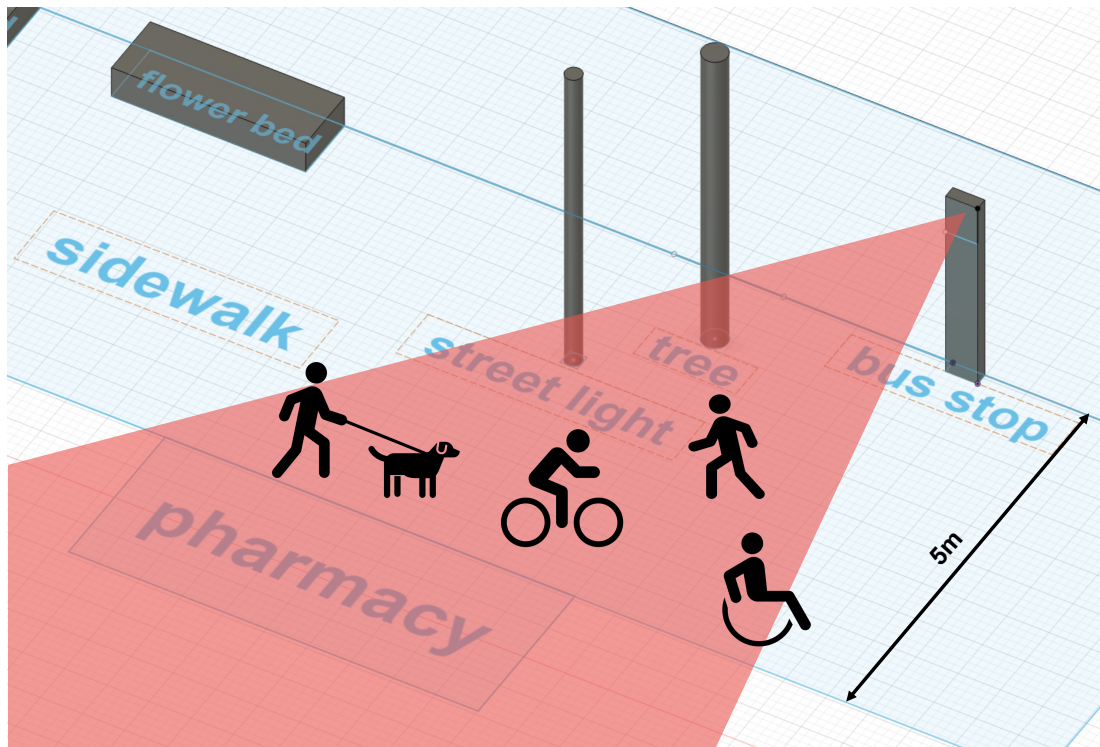


図 15: 撮影環境と撮影可能エリアの様子

3.3 データセット

本研究では、可視光映像と遠赤外線映像を利用した場合に、時間と天気の変化が歩行者数推定に及ぼした影響を検証する。検証に必要な可視光映像と遠赤外線映像の撮影は、2022年3月24日から2023年1月26日の期間に実施し、撮影した映像は10分単位の映像として保存した。本研究では歩行者数推定に対する照度や天気の変化による影響に着目しているため、すべての映像を昼間の晴天時、昼間の雨天時、夜間の晴天時、夜間の雨天時という4パターンで分類した。デー

タセットとしては、各パターンの映像の本数が同じになるように、可視光映像と遠赤外線映像それぞれについて各パターンに該当する映像を14本ずつ利用した。それぞれの映像ごとに、撮影開始時間や正解歩行者数、撮影時の天気情報があるため、それらを集約したデータセットを作成した。図16はデータセットの一例を示す。“Date”カラムには映像の撮影の開始日時，“Correct”カラムは、可視光映像の目視確認により歩行者数をカウントした正解歩行者数の結果を格納している。なお，“Correct”のカウント対象は、歩行者や自転車・車椅子に乗っている人とし、抱えられた幼児はカウント外とした。“Time”カラムは撮影日時から算出しており、8:30–16:40を昼間、19:30–22:45を夜間として定義した。なお、本研究では定義外の時間の映像は利用していない。“Video type”は撮影した映像の種類を示しており、可視光映像である Visible、遠赤外線映像である Far-infrared の2種類である。“Weather”と“Temperature”の結果は気象庁のデータ¹³を利用する。気象庁のデータは、雲量や雨の強度等によって分類されているため，“Weather”の入力の際は表5のように分類した。気象データ取得の場所は、撮影場所である兵庫県神戸市としており，“Date”カラムの情報を元にデータを結合した。

表 5: 気象庁のデータと検証用データセットで用いる“Weather”の対応表

項番	天気	Weather
2	晴	Sunny
4	曇	Sunny
10	しゅう雨または止み間のある雨	Rainy
15	着氷性の雨	Rainy

¹³<https://www.data.jma.go.jp/obd/stats/etrn/index.php>

<u>Date</u>	<u>Correct</u>	<u>Time</u>	<u>Video type</u>	<u>Weather</u>	<u>Temperature(°C)</u>
2022-03-24 11:55	109	Daytime	Visible	Sunny	13
2022-04-21 16:39	58	Daytime	Visible	Rainy	17
2022-04-26 20:52	67	Night	Visible	Rainy	20
2022-04-28 21:53	42	Night	Visible	Sunny	20
⋮	⋮	⋮	⋮	⋮	⋮
2022-03-24 11:55	109	Daytime	Far-infrared	Sunny	13
2022-04-21 16:39	58	Daytime	Far-infrared	Rainy	17
2022-04-26 20:52	67	Night	Far-infrared	Rainy	20
2022-04-28 21:53	42	Night	Far-infrared	Sunny	20
⋮	⋮	⋮	⋮	⋮	⋮

図 16: 作成したデータセットの一例

4. 予備実験

本章では、可視光映像と遠赤外線映像を利用した場合の、歩行者数推定における推定精度の低下要因を検証するために実施した予備実験について述べる。双方の映像を用いた場合の、推定精度減少要因の仮説をそれぞれ4.1節、4.2節に示す。

仮説の検証には、図17のような歩行者数推定手法を用いる。本研究における歩行者数推定は、歩行者検出と歩行者追跡の順で実施する。歩行者検出に利用する検出器作成方法と、歩行者追跡に利用するパラメータの説明をそれぞれ4.3節、4.4節に示す。

BYTEの出力であるIDの数を歩行者数推定結果とし、目視確認の正解データをもとに誤差率を計算する。

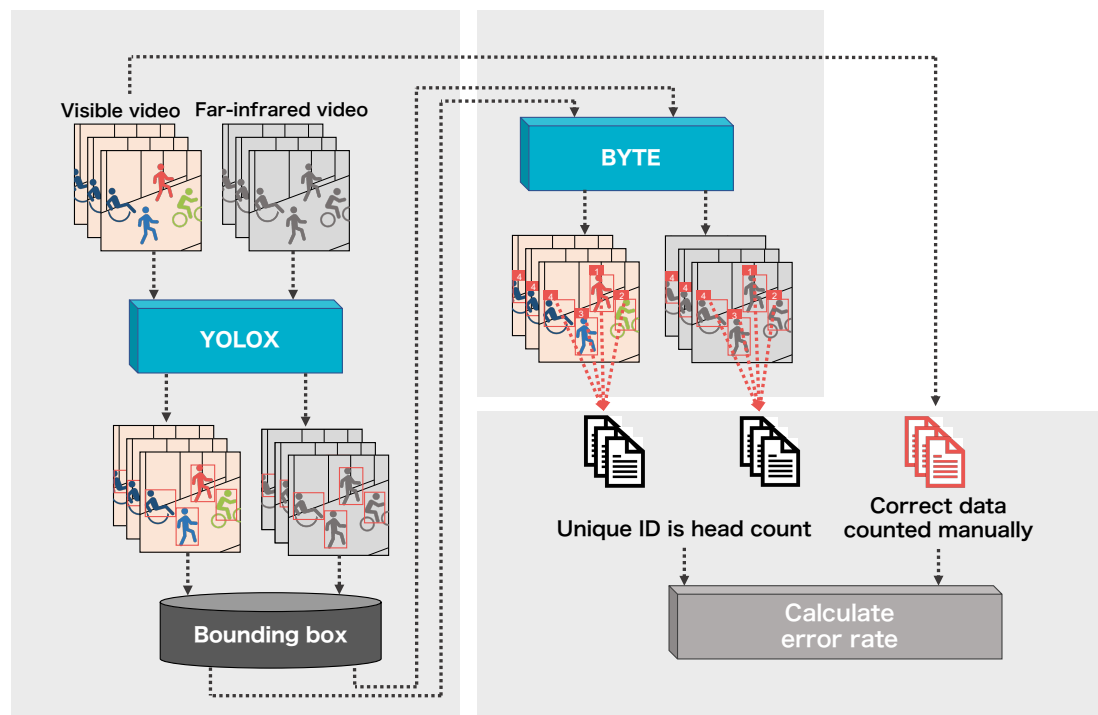


図 17: 可視光映像と遠赤外線映像を利用した歩行者数推定手法の概要図

4.1 可視光映像を用いた場合に推定精度が下がる要因の仮説

可視光映像を用いた場合に、歩行者数の推定精度が下がる要因は環境要因と人的要因に大別されると考えた。本節で述べる内訳は下記の箇条書きとしてまとめることができる。なお、4.1.1項から示す仮説中では、対象物体が光源ではない場合を想定する。

- 環境要因
 - － 時間：昼・夜・逆光時(太陽光によるフレア)
 - － 天気：晴れ・くもり・雨
- 人的要因
 - － 着用物：衣服と背景の同化・マスクの着用・雨具の使用
 - － 行動：歩行・走行時・自転車移動・車椅子移動

4.1.1 環境要因

環境要因を時間的要因・天气的要因の2種類に分けて考える。本項では、それぞれの要因が歩行者数推定に及ぼす影響を述べる。

時間的要因 可視光映像を用いた歩行者数推定が低下する時間的要因は2つある。1つ目は夜間に代表される照度の低さである。1章でも述べたように歩行者に対して光が当たらなければ、可視光映像による歩行者検出ができず、推定精度は低下する可能性がある。2つ目は太陽光のフレアに代表される逆光である。時間帯によっては、太陽光などの強い光がレンズに入射し、画像全体がぼやける可能性がある。このような現象はフレアと呼ばれ、ぼやけた画像では歩行者の検出ができず、推定精度が低下する可能性がある。しかし、逆光が発生した映像を撮影していないため、本研究では検証外の仮説とする。

天氣的要因 可視光映像の認識精度が低下する天氣的要因には雨や雪などの降水が関係すると考えられる。降水発生時にはカメラ・レンズ全面に雨粒が付着し、乱反射した光により歩行者数を推定しづらい映像になる可能性がある。

4.1.2 人的要因

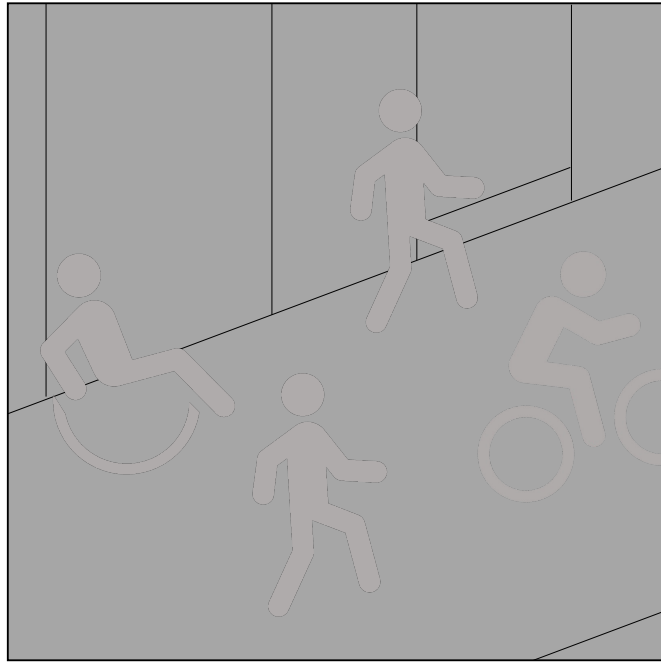
人的要因を着用物的要因と行動的要素の2種類に分けて考える。本項では、それぞれの要因が歩行者数推定に及ぼす影響を述べる。

着用物的要因 着用物的要因では、3つの要因で推定精度が減少すると考える。1つ目は衣服と背景の同化である。森の中の迷彩服に代表されるような、背景と同色の服を着ている場合は歩行者の特徴を捉えられず、推定精度は低下すると考えられる。2つ目はマスク等の着用である。歩行者の特徴である顔が隠れる場合には、歩行者の特徴を捉えられず、推定精度は低下すると考えられる。3つ目は雨具等の使用である。天氣的要因にもつながるが、降水発生時には傘やレインコートなどの雨具の利用により、歩行者の一部が隠される。加えて傘の利用は、街灯などの照明の遮断にもつながるため、さらに推定精度が低下する可能性がある。

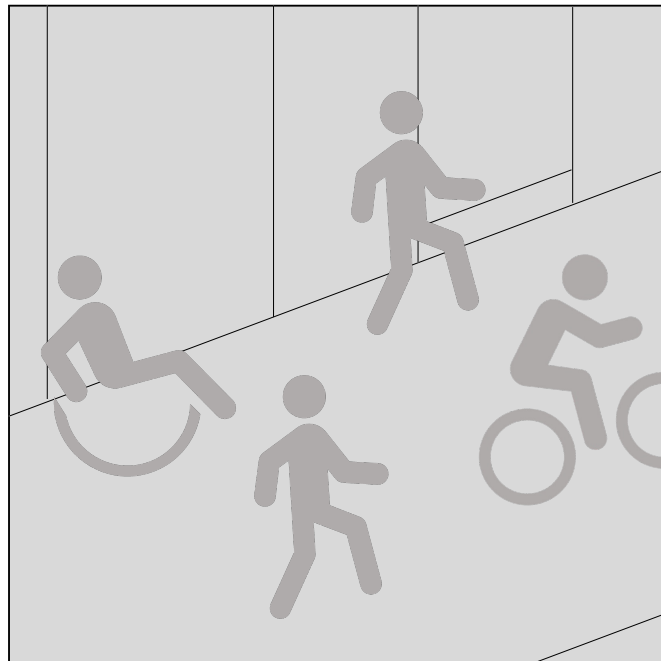
行動的要素 行動的要素は人々の移動方式に関わる点である。走っている人や自転車で移動している人も含めて推定する場合、映像のフレームレートが小さいと、歩行者を追跡できない可能性がある。

4.2 遠赤外線映像を用いた場合に推定精度が下がる要因の仮説

遠赤外線画像の推定精度が下がる要因は歩行者と背景の温度が一致する場合と考える。このような状況の場合は図 18(a) のような映像が生成され、図 18(b) に示すような温度差が大きい場合に比べて、歩行者の輪郭や温度変化を捉えられず、推定精度が下がると考える。



(a) 背景の温度と歩行者の温度差が小さい場合



(b) 背景の温度と歩行者の温度差が大きい場合

図 18: 遠赤外線映像を利用した場合の温度差による影響の想定図

4.3 物体検出器の転移学習

本研究における物体検出にはYOLOX [25] を利用した。YOLOX はCVPR2021のAutonomous Driving WorkshopのStreaming Perception Challenge [26] で最も高いAverage Precision (AP) を獲得した物体検出アルゴリズムである。YOLOXにはMS COCOで学習された事前学習済みモデルが利用可能であり、このモデルを利用することによって、高精度に歩行者検出が可能である。また、事前学習済みモデルを利用することによって、設置場所ごとに検出器を学習する必要がなくなり、計算コストを削減できる。しかしYOLOXの事前学習済みモデルでは、遠赤外線映像内の歩行者を検出することができないため、事前学習済みモデルに対して遠赤外線映像の転移学習を実施した。学習の際には遠赤外線画像と、映像の各フレームで歩行者を囲んだアノテーション結果が必要となる。そこで3月に撮影された、歩行者数推定に利用しない遠赤外線映像の内、約8000枚の画像に対してFastLabel¹⁴を利用し、アノテーションを実施した。アノテーションの結果は学習用画像とテスト用画像をそれぞれ7:3の割合で分割した。アノテーションの様子を図19に示す。

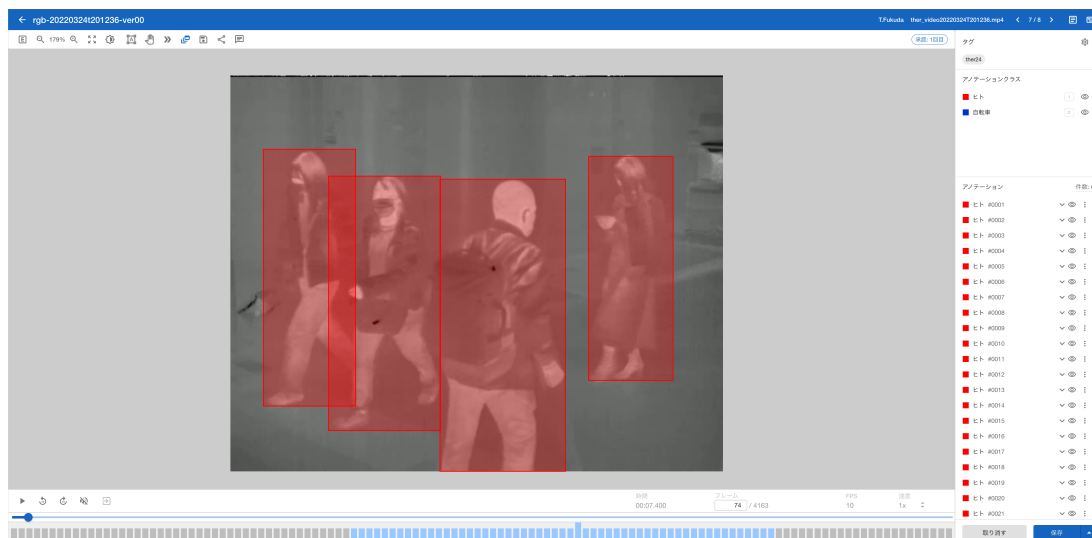


図 19: 遠赤外線映像のアノテーションの様子

¹⁴<https://fastlabel.ai/>

アノテーションした結果をもとに、事前学習済みモデルに対して転移学習したパラメータを表6に示す。ベースモデルとは事前学習済みモデルを示し、公開されている中で最も mAP が高い事前学習済みモデル¹⁵を利用した。バッチサイズは、1回の学習に利用する画像の枚数である。今回は学習に使用するマシンで利用可能なサイズに設定している。最大エポック数とは学習の際にすべての画像を学習した回数（エポック数）の上限値を決定する。これによって300エポックの学習の中で最も汎化性能の高いモデルを利用することができる。カテゴリ数は、“person” だけを検出することを目的とするために1に設定した。

表 6: 転移学習実行時のパラメータ

パラメータ	値
学習用画像枚数	約 5600 枚
テスト用画像枚数	約 2400 枚
ベースモデル	yolox_x.pth
バッチサイズ	64
最大エポック数	300
カテゴリ数	1

4.4 物体追跡アルゴリズムの追跡パラメータ

物体追跡アルゴリズムには BYTE を利用した。BYTE による追跡の際に利用したパラメータを表7に示す。

追跡カテゴリでは歩行者のみを対象にするため、“person” を指定する。信頼度の最小値とは、追跡に使用する bounding box の信頼度の最小値を示す。IoU しきい値は、物体検出器の bounding box と BYTE の追跡結果の bounding box の IoU を計算した際に、同一 ID を割り振ることを決定するしきい値である。信頼度のしきい値は、物体検出器の bounding box のカテゴリ信頼度のしきい値を決定する。

¹⁵<https://github.com/Megvii-BaseDetection/YOLOX>

表 7: 物体追跡に利用したパラメータ

パラメータ	値
追跡カテゴリ	“person”
信頼度の最小値	0.25
IoU のしきい値	0.8
信頼度のしきい値	0.7

4.5 歩行者数推定誤差率の計算方法

本実験では各映像の歩行者数の正解値と歩行者検出・追跡による推定結果より、推定の誤差率を計算する。歩行者数推定結果には、各映像ごとに歩行者に割り当てられた固有の ID の数を用いる。誤差率の計算方法を式 (4) に示す。

$$\text{error_rate} = \frac{\text{Predict} - \text{Correct}}{\text{Correct}} \times 100 \quad (4)$$

ここで Correct は歩行者数の正解値、Predict は提案手法が推定した人数である。error_rate は、符号付きの数値であり、+3%の誤差であれば、100 人通った際に 103 人と推定していることを示す。

各映像における歩行者数推定値と計算した誤差率をもとに、図 20 のように、検証用データセットを更新する。

4.6 可視光映像を用いた場合に推定精度が下がる要因の検証

本節では環境要因の変化が、可視光映像を用いた歩行者数推定に与える影響の検証結果を述べる。人的要因に関しては、エラーが発生した場合の推定映像を確認し、考察する。なお、環境要因の中でも時間的要因の 1 つである逆光の様子は、本実験で用いる映像では確認できなかったため、この仮説は検証外とする。

まず環境要因の内、時間的要因が、可視光映像を用いた歩行者数推定精度をどう変化させるか検証する。図 21 は、可視光映像を用いた場合の誤差率 “Error rate” と昼間か夜間かを示す撮影時間 “Time” の箱ひげ図である。誤差率がプラス方向に高いと歩行者を過剰に数えていることを示し、誤差率がマイナス方向に

Date	Correct	Time	Video type	Weather	Temperature [°C]	Predict	Error rate [%]
2022-03-24 11:55	109	Daytime	Visible	Sunny	13	103	-5.5
2022-04-21 16:39	58	Daytime	Visible	Rainy	17	57	-1.7
2022-04-26 20:52	67	Night	Visible	Rainy	20	59	-11.9
2022-04-28 21:53	42	Night	Visible	Sunny	20	40	-4.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2022-03-24 11:55	109	Daytime	Far-infrared	Sunny	13	98	-10.1
2022-04-21 16:39	58	Daytime	Far-infrared	Rainy	17	56	-3.4
2022-04-26 20:52	67	Night	Far-infrared	Rainy	20	58	-13.4
2022-04-28 21:53	42	Night	Far-infrared	Sunny	20	40	-4.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

図 20: 予備実験で用いる推定値と誤差率を加えた結果の一例

大きいと歩行者を見逃していることを示す。中央のオレンジのエリアは誤差率が±5%の範囲である。図 21 より、可視光映像を用いた場合は、昼間よりも夜間の方がマイナス方向に誤差率が大きいことがわかる。さらに、夜間の方が誤差率の分散も大きく、昼間に比べて夜間の方が歩行者の見逃し回数が増えると考えられる。そこで図 22 には、時間ごとの追跡映像の様子を示す。図 22(a) は昼間の晴天時に追跡が成功している例である。このように、昼間の映像を利用する場合は、画像内の歩行者を検出し追跡によって、歩行者ごとに ID を割り当てていることが多かった。一方、図 22(b) のような夜間の場合は、歩行者に対して十分な光が当たらず、歩行者として検出・追跡されないことが多かった。

次に環境要因の内、気象的要因が可視光映像を用いた歩行者数推定精度をどう変化させるかを検証する。図 23 は、可視光映像を用いた場合の誤差率 “Error rate” と撮影時間 “Time” の箱ひげ図を天気ごとに分割した。グラフの見方は図 21 と同様である。図 23(a) は晴天時、図 23(b) は雨天時の結果であり、雨天時の方が誤差率が大きくなると考えていたが、天気による差は見られなかった。これは夜間の晴天時に夜間の雨天時に見られない、自転車上の人物の追跡漏れが発生していることが、1つの要因だと考える。図 24 には、夜間に自転車移動をしている人物を撮影した様子を示す。この映像例からもわかるように、夜間の自転車上の人物はブレて撮影され、検出・追跡できないことが多かった。夜間はシャッター

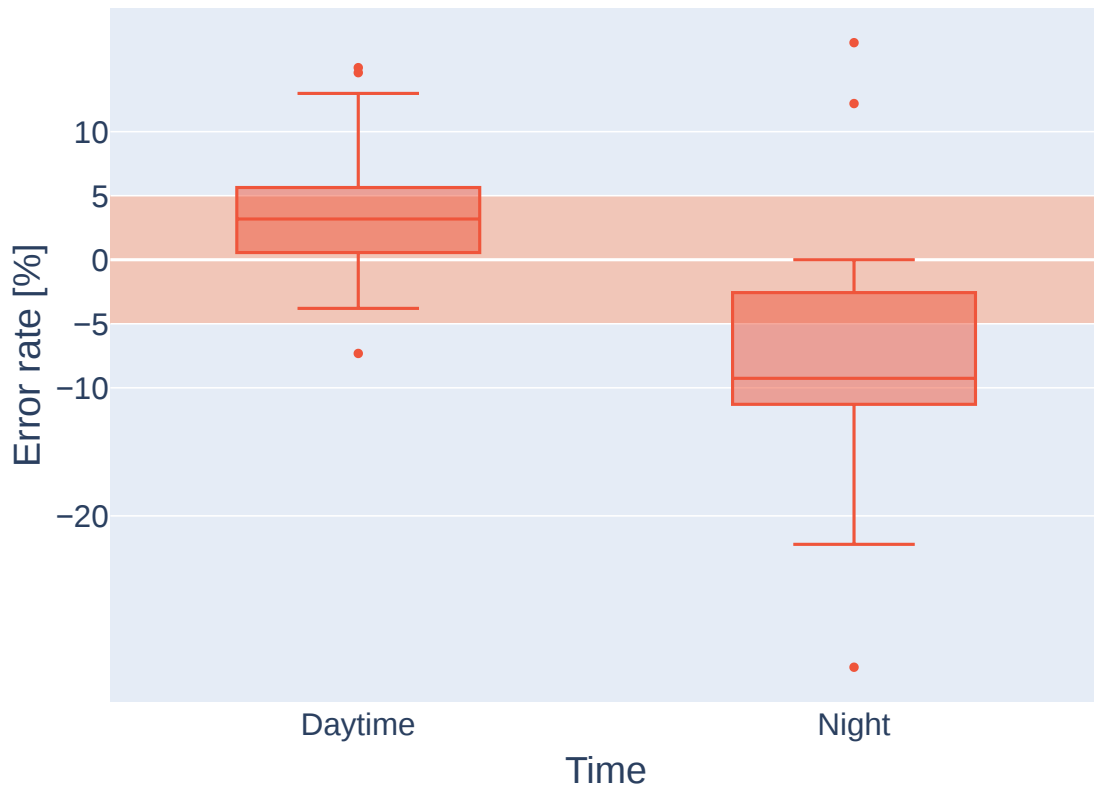


図 21: 可視光映像を利用した場合の昼間と夜間の誤差率の箱ひげ図

スピードが低下することによって、ブレた画像が生成されたと考える。晴天時は雨天時よりも自転車を利用する人が多くなると予想され、自転車の見逃し数が多いことによって、雨天時のほうが誤差率が大きくなるという想定とは異なる結果となった。

最後に人的要因が可視光映像を用いた歩行者数推定精度をどう変化させるかを検証する。人的要因では、図 22(b) のように、夜間に黒い服装を着用している事によって、検出・追跡に失敗している例がある。また、図 25 の用に黒い服装でなくとも、雨天時における雨具の利用によって歩行者の特徴である、頭部や体の一部が隠れる場合により推定失敗につながる可能性がある。このように可視光映像を用いた場合は、環境要因と人的要因が組み合わさることによって、推定精度が減少すると考えられる。

以上のような可視光映像を用いた予備実験により、4.1 節で示した仮説を検証

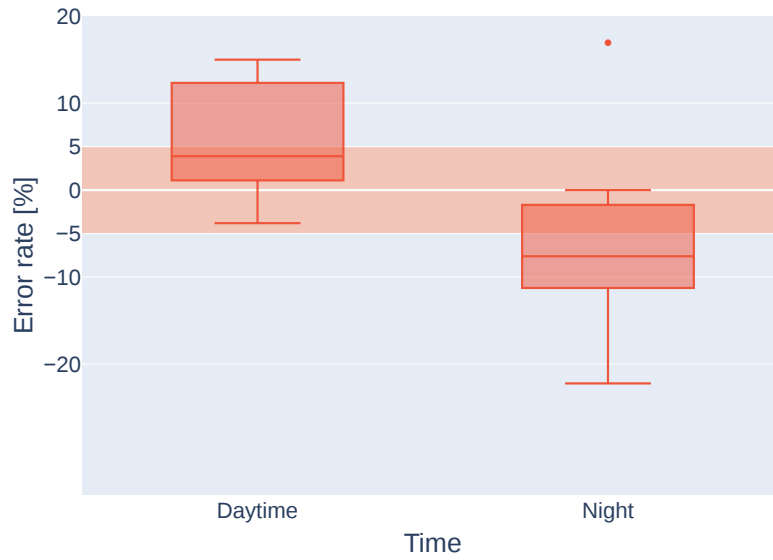


(a) 昼間 + 晴天時

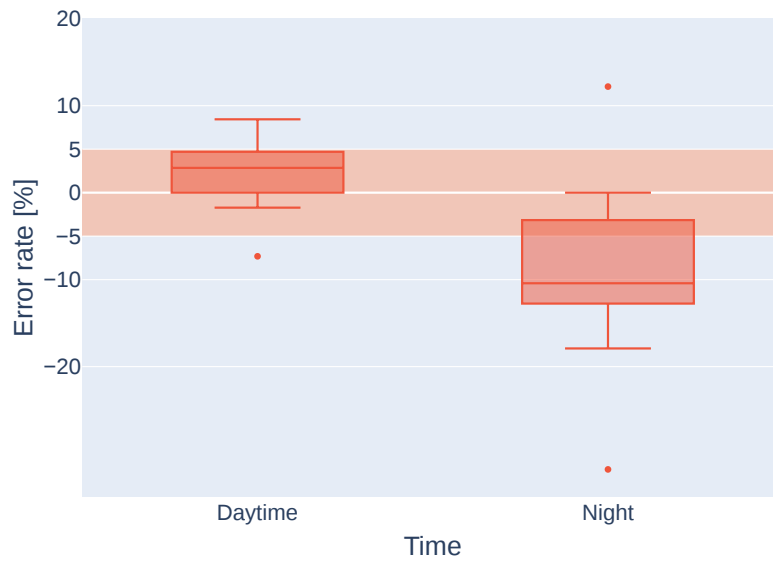


(b) 夜間 + 雨天時

図 22: 可視光映像をそれぞれの時間と天気を利用した例



(a) 晴天時



(b) 雨天時

図 23: 可視光映像を利用した場合の天気ごとの昼間と夜間の誤差率の箱ひげ図



(a) 1852 フレーム目



(b) 1853 フレーム目



(c) 1854 フレーム目

図 24: 自転車の追跡失敗例



図 25: 夜間の雨天時における推定失敗例

した.

4.7 遠赤外線映像を用いた場合に推定精度が下がる要因の検証

本節では周囲の温度状況が、遠赤外線映像を用いた歩行者数推定に与える影響の検証結果を述べる。図26は、遠赤外線映像を利用した場合の気温と誤差率の関係性を示す。中央のオレンジのエリアは誤差率が $\pm 5\%$ の範囲である。この図より、気温が 20°C を下回る場合に、誤差率が小さくなっていることがわかる。図27(a)には、気温 11°C における遠赤外線映像を用いた検出・追跡の結果を示す。遠赤外線映像は、撮影した物体の温度が高いほど白く、温度が低いほど黒く写る。気温が 20°C よりも低い場合は、地面の温度が歩行者の表面温度よりも十分に低く、背景と歩行者の温度差によって歩行者の輪郭が明瞭になることで、歩行者追跡が容易な環境になっていると考えられる。

また、気温が 35°C 付近でも誤差率が小さい結果がある。これは、地面の温度が歩行者の表面温度よりも十分に高くなり、背景との温度差によって歩行者の輪郭が明瞭になったためであると考えられる。実際、図27(c)に示すように、気温が 35°C 付近の歩道では、地面の方が白く、歩行者のほうは黒く撮影され、歩行者の輪郭が明瞭になるため、歩行者追跡が容易になったと考えられる。

しかし、気温が 24°C – 30°C 付近の場合はマイナス方向への誤差率が大きくなっていることがわかる。これは、地面の温度が歩行者の表面温度と近くなり、地面と歩行者の温度差が少なくなることで、歩行者と背景を識別できなくなってしまったためであると考えられる。例として、図27(b)に示す、気温が 27°C の場合、地面の温度と歩行者の温度が近く、輪郭が不明瞭となっている。そのため、歩行者の検出・追跡に失敗し、誤差率が大きくなったと考えられることができる。

以上のような予備実験によって、4.2節で示した仮説を検証し、推定精度が低下する気温はおおよそ 24°C – 30°C の範囲であると考えられる。

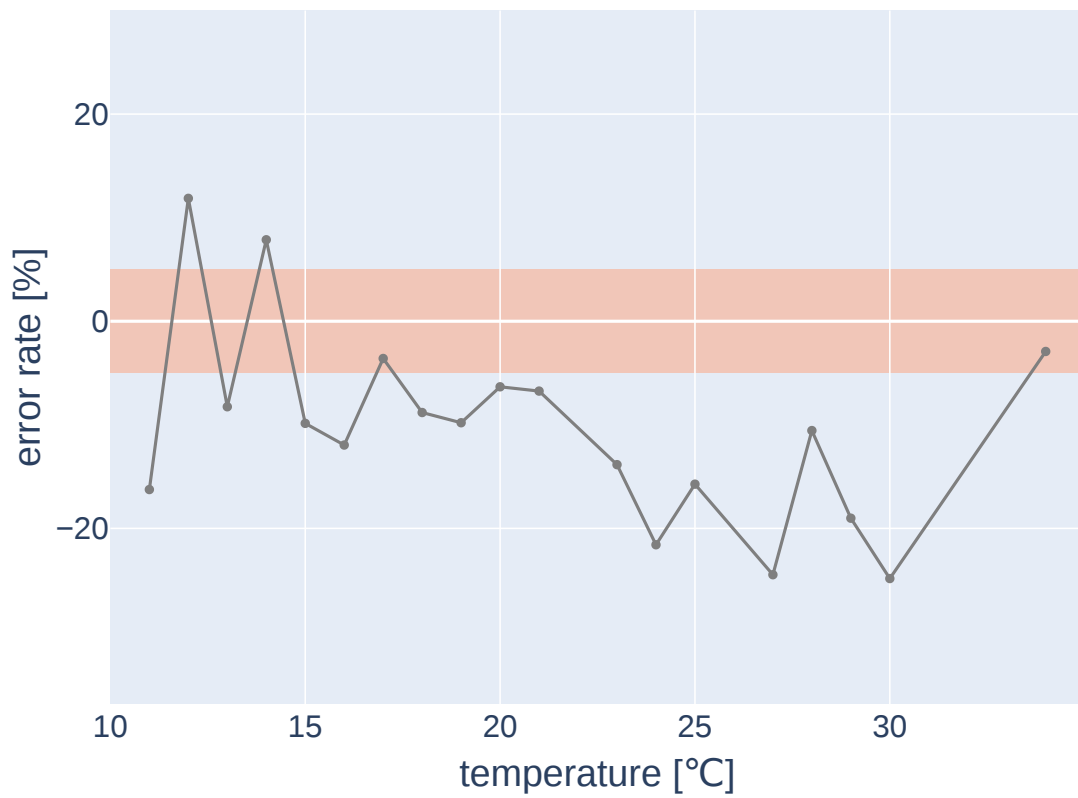


図 26: 遠赤外線映像を利用した場合の気温と誤差率のグラフ



(a) 11°C



(b) 27°C



(c) 34°C

図 27: 遠赤外線映像を3つの気温で利用した例

4.8 時間と天気ごとの誤差率

表8は時間や天気の各4パターンにおける、可視光映像と遠赤外線映像を用いた歩行者数推定の誤差率を示す。各パターンの誤差率の計算方法を式5に示す。ここで、 N は、各パターンで利用した動画の本数を示し、今回は14本の動画を利用しているため、14である。 $Predict_i$ と $Correct_i$ はそれぞれ i 本目の動画における予測人数と正解人数である。 e は各パターンでの誤差率を示し、14本の動画内に映る歩行者数の合計値と動画から推定された人数の合計値から計算される。

表 8: 歩行者数推定の誤差率

	昼間		夜間	
	晴天時	雨天時	晴天時	雨天時
可視光映像	4.9% (+)	3.1% (+)	7.0% (-)	7.6% (-)
遠赤外線映像	7.3% (-)	16.2% (-)	14.3% (-)	14.0% (-)

$$e = \frac{100 \sum_i^N Predict_i - \sum_i^N Correct_i}{\sum_i^N Correct_i} \quad (5)$$

この結果より、昼間に可視光映像を利用した場合は5%未満の誤差率となっており、歩道の利用状況調査 [4] で使用された目標値を達成している。対して夜間では撮影場所付近に街灯があるにも関わらず、5%以上の誤差率となっていた。天気に着目すると、雨天時は晴天時に比べて見逃しの回数が増加する可能性がある。そのため、昼間の場合は晴天時に比べて誤差率が低下し、夜間の場合は誤差率が大きくなったと考えられる。

次に遠赤外線映像を利用した場合は、可視光映像に比べて誤差率が高い。誤差率が高いのは、遠赤外線映像の推定精度が低下する気温の映像を、多く利用している可能性が考えられる。そこで、図28には、4パターンにそれぞれ14本、合計56本の映像の気温ごとの内訳を示す。このグラフから分かるように、遠赤外線映像を利用する場合に推定精度が低下する、 24°C – 30°C 付近の映像があり、特に 24°C – 25°C の動画は20本存在していた。半分近くが推定精度が低下するような気温であったため、平均した場合に誤差率が大きくなった可能性がある。

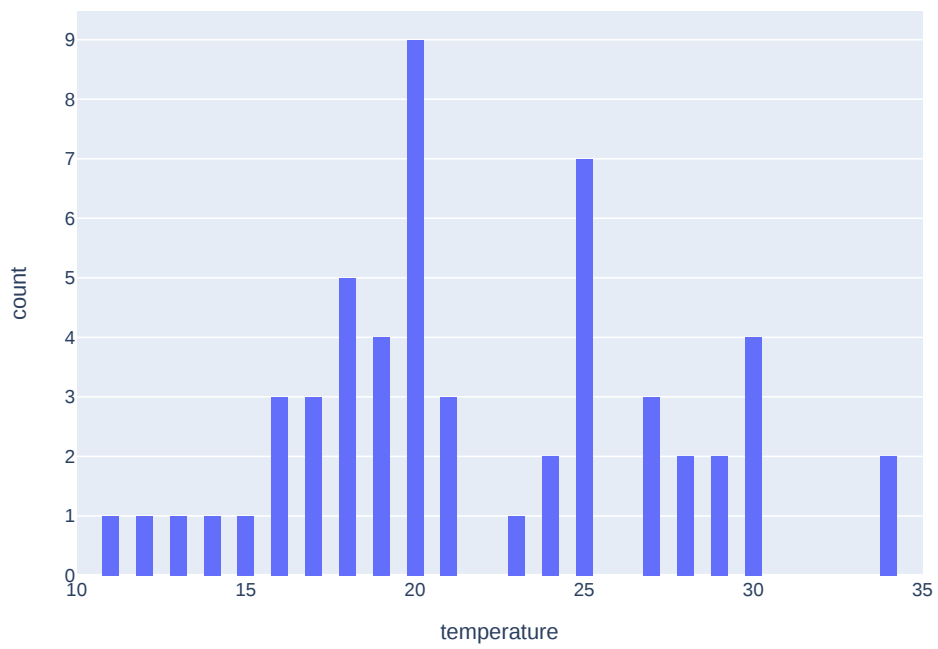


図 28: 気温ごとの遠赤外線映像本数

4.9 可視光映像と遠赤外線映像の補完関係

予備実験において、可視光映像を利用する場合は、照度が低い場合に推定精度が減少することがわかった。また、遠赤外線映像を利用する場合は気温 24°C – 30°C 付近で推定精度が減少することがわかった。そこで、一方の映像で検出・追跡できなかつた場合に、もう一方の映像を利用した場合の結果を調査する。図29は、可視光映像で検出・追跡できなかつた場合を示す。図29(a)は図22(b)で示した画像と同じ時間に撮影した遠赤外線画像である。撮影時は気温 19°C であり、可視光映像では検出・追跡できていない歩行者が、図29(b)に示す遠赤外線映像を利用することで検出・推定可能になっている。このような状況は可視光映像の推定精度が低下する、夜間の雨天時の場合に多く確認できた。また、天気の違いによる推定結果として、遠赤外線は傘を透過する場合がある。そのため、雨天時にはより有効に遠赤外線映像が利用できると考えられる。

図30は図27(b)で示した画像と同じ時間に撮影した、可視光映像と遠赤外線映像を示す。撮影時は気温 27°C である。遠赤外線映像では、背景と歩行者の表面温度の差がないため、検出・追跡に失敗しているが、図30(b)に示す可視光映像では、2人の歩行者を検出・追跡可能である。このような状況は遠赤外線映像の推定精度が低下する、 24°C – 30°C 付近で多く確認できた。

このように可視光映像では物体検出できないが、遠赤外線映像では物体検出可能な場面がある。反対に遠赤外線映像で検出できない場合には可視光映像を利用することで検出可能になる場面もある。そのため、1日中歩行者数推定を行う場合は、あらゆる時間や気温に対応できるように、2つの映像の特徴を利用することが必要だと考える。



(a) 可視光映像

(b) 遠赤外線映像

図 29: 夜間の雨天時における検出結果例



(a) 可視光映像

(b) 遠赤外線映像

図 30: 気温 27°C における検出結果例

5. 可視光映像と遠赤外線映像を融合した映像による歩行者数推定手法

4章では、可視光映像と遠赤外線映像を用いる場合に、歩行者数推定精度が低下する要因を調査した。可視光映像は夜間のような照度が低い場合に推定精度が低下し、遠赤外線映像は気温およそ 24°C–30°C の範囲で推定精度が低下する。結論として1日中歩行者数推定を行う場合は、あらゆる時間や気温に対応できるように、2つの映像の特徴を利用する必要があることを述べた。2つの映像の特徴を利用する1つの手法として、双方の映像を入力可能な物体検出器を作成し、検出器の出力を追跡アルゴリズムに入力する手法がある。この手法では2つの映像を入力可能な物体検出器の学習のために、可視光映像と遠赤外線映像のペアデータセットが大量に必要となる。しかし、可視光映像と遠赤外線映像のペアデータセットは公開されていない。また、歩行者数推定地点ごとの新たなモデル作成は、都市計画上非常に大きな計算コストがかかる欠点がある。

そこで本研究では、1つの映像を入力とする事前学習済みモデルを利用して歩行者数推定を行うために、可視光映像と遠赤外線映像を融合する手法を提案する。事前学習済みモデルの利用によって、学習に必要な大量の画像および、計算資源が必要なく、計算コストを削減可能である。図 31 に提案手法の実験フローを示す。可視光映像と遠赤外線映像を DenseFuse¹⁶ [27] を用いて融合させ、1つの融合映像を作成する。融合映像は予備実験と同様に、YOLOX を用いて歩行者検出を行った後、BYTE で歩行者に ID を付与する。付与した ID は歩行者数推定に利用する。

本提案手法で用いる映像融合手法は DenseFuse を利用する。DenseFuse のアーキテクチャを図 32 に示す。エンコーダは C1 と Dense block から構成され、入力画像から特徴の抽出を行う。デコーダは4つの畳み込み層から構成され、最終的に融合された画像を再構成する。

図 33 は DenseFuse によって融合する前の画像と後の画像である。これらの画像には3名の歩行者が写っている。可視光映像である図 33(a) では、左側の歩行

¹⁶https://github.com/hli1221/imagefusion_densefuse

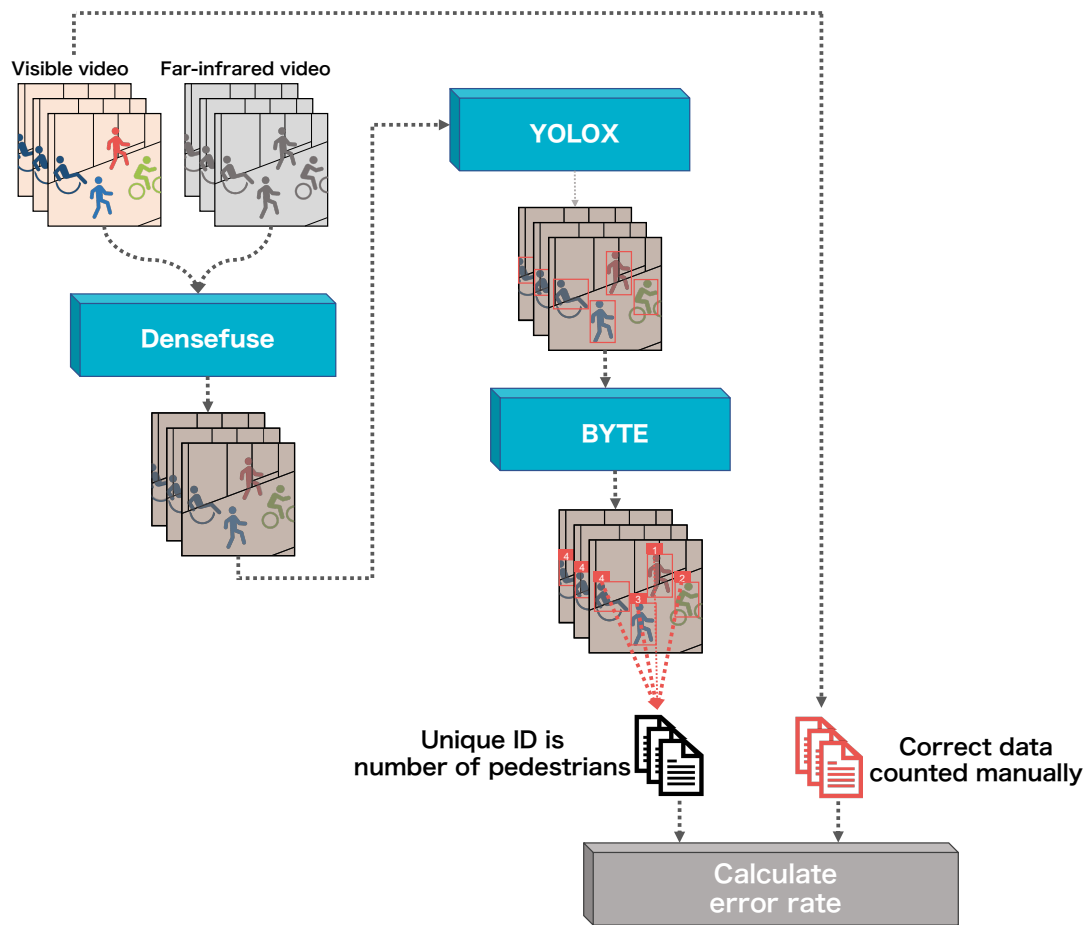


図 31: 融合映像を利用した歩行者数推定手法の概要図

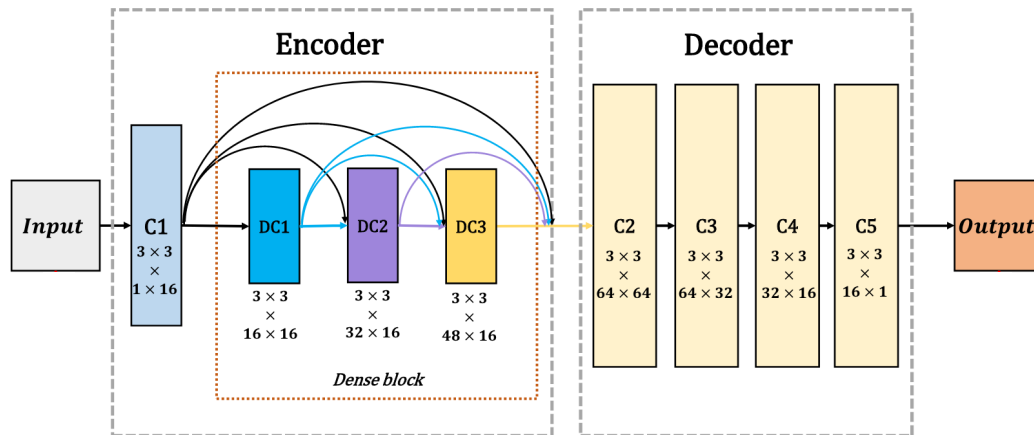


図 32: DenseFuse のアーキテクチャ (文献 [27] より引用)

者 2 人が黒い服装を着用しているため、1 名と推定される可能性がある。さらに、右側の歩行者も黒い服装であること、雨天による傘で頭部が隠れ、歩行者としての特徴が少ないことから検出・追跡されない可能性がある。一方、図 33(b) に示す遠赤外線映像の場合は、左側の 2 人を別の歩行者として推定できる可能性が高い。右側の歩行者についても、背景である地面との温度差によって歩行者を推定できる可能性が高い。

図 33(c) には DenseFuse を用いて、可視光映像と遠赤外線映像の特徴を融合した結果を示す。融合映像は可視光映像に対して遠赤外線映像の情報が付与されているため、可視光映像による検出が困難な右側の歩行者の輪郭を取得できる。



(a) 可視光映像



(b) 遠赤外線映像



(c) 融合映像

図 33: 可視光映像と遠赤外線映像を元にした Densefuse の出力結果例

6. 提案手法の評価

本章では、可視光映像や遠赤外線映像のみを利用した場合と、融合映像を利用した場合のそれぞれの歩行者数推定結果を確認し、融合映像の有効性を検証する。融合映像を用いた場合の歩行者数推定精度を調査するために、予備実験同様に推定値と誤差率を加え、検証用データセットを更新する。図 34 は予備実験の検証用データセットに、融合映像を利用した場合の推定結果と誤差率を追加したデータセットの一例を示す。それぞれのカラムは予備実験と同様であり、撮影した映像の種類を示す“Video type”には、可視光映像である Visible、遠赤外線映像である Far-infrared に加えて融合映像である Fused を追加した。

Date	Correct	Time	Video type	Weather	Temperature [°C]	Predict	Error rate [%]
2022-03-24 11:55	109	Daytime	Visible	Sunny	13	103	-5.5
2022-04-21 16:39	58	Daytime	Visible	Rainy	17	57	-1.7
2022-04-26 20:52	67	Night	Visible	Rainy	20	59	-11.9
2022-04-28 21:53	42	Night	Visible	Sunny	20	40	-4.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2022-03-24 11:55	109	Daytime	Far-infrared	Sunny	13	98	-10.1
2022-04-21 16:39	58	Daytime	Far-infrared	Rainy	17	56	-3.4
2022-04-26 20:52	67	Night	Far-infrared	Rainy	20	58	-13.4
2022-04-28 21:53	42	Night	Far-infrared	Sunny	20	40	-4.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2022-03-24 11:55	109	Daytime	Fused	Sunny	13	101	-7.3
2022-04-21 16:39	58	Daytime	Fused	Rainy	17	55	-1.7
2022-04-26 20:52	67	Night	Fused	Rainy	20	60	-10.4
2022-04-28 21:53	42	Night	Fused	Sunny	20	38	-4.8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

図 34: 本実験で用いるデータ

6.1 可視光映像と融合映像を用いた推定結果の比較

図 35 は、可視光映像と融合映像それぞれを用いた場合の誤差率 “Error rate” と昼間か夜間かを示す撮影時間 “Time” の箱ひげ図である。昼間の場合、融合映像を用いた歩行者数推定は、可視光映像の結果よりも分散が大きく、推定精度は低下しているように考えられる。そこで、可視光映像で同じ ID を追跡できているにも関わらず、融合映像で同一歩行者に別の ID が割り当てられている場面を調査した。可視光映像で正しく追跡できている様子を図 36、同じフレームにおいて融合映像で別 ID が割り振られている様子を図 37 に示す。融合映像は可視光映像に対して遠赤外線映像の情報を融合しているため、可視光映像よりも全体的にぼやけた色となる。ぼやけた映像によって、追跡の過程で他の歩行者であると推定された可能性がある。

一方で夜間の場合は、可視光映像よりも最大値・最小値の幅が小さく、マイナスの外れ値が少ないことから、可視光映像のみを利用する場合に比べて歩行者の見逃しを抑制していることが考えられる。図 38 には、夜間の雨天時における可視光映像と融合映像の例を示す。図 38(a) の可視光映像の場合は、歩行者を検出・追跡できていないのに対し、図 38(b) の場合は歩行者の検出・追跡が可能となっている。

6.2 遠赤外線映像と融合映像を用いた推定結果の比較

図 39 は、遠赤外線映像を利用した場合と融合映像を利用した場合の気温と誤差率を示す。図 39(a) は予備実験の結果と同じく、およそ 24°C–30°C の気温範囲で誤差率がマイナス方向に大きくなっている。一方で図 39(b) に示す融合映像を利用した場合は、気温が 24°C–30°C 付近の範囲内でも誤差率の増加を抑制し、気温の影響を受けづらくなっていると考えられる。融合映像が、気温の影響を受けづらくなったと考えられる様子を図 40 に示す。図 40(a) は 27°C における遠赤外線映像の追跡例である。4.7 節では、背景との温度差が小さいことで、追跡に失敗していることを述べた。図 40(b) は融合映像を用いた場合の同じフレームを示す。融合映像利用により、遠赤外線映像では追跡できなかった歩行者を追跡でき

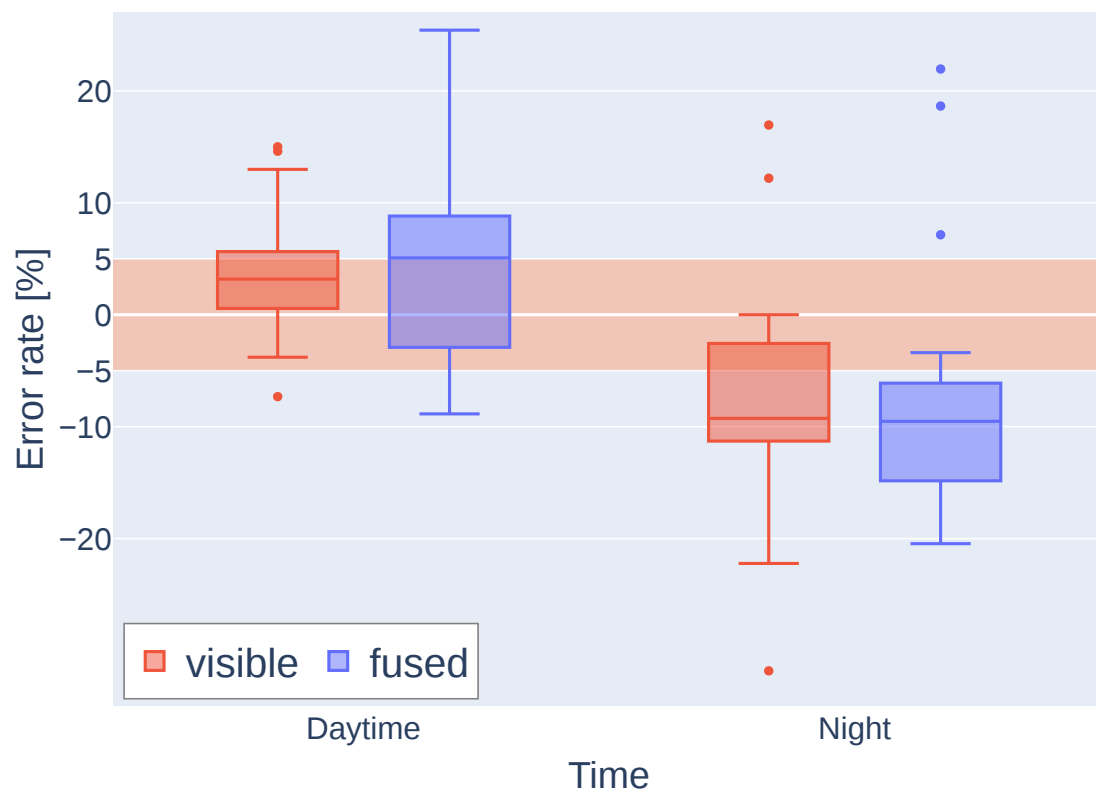
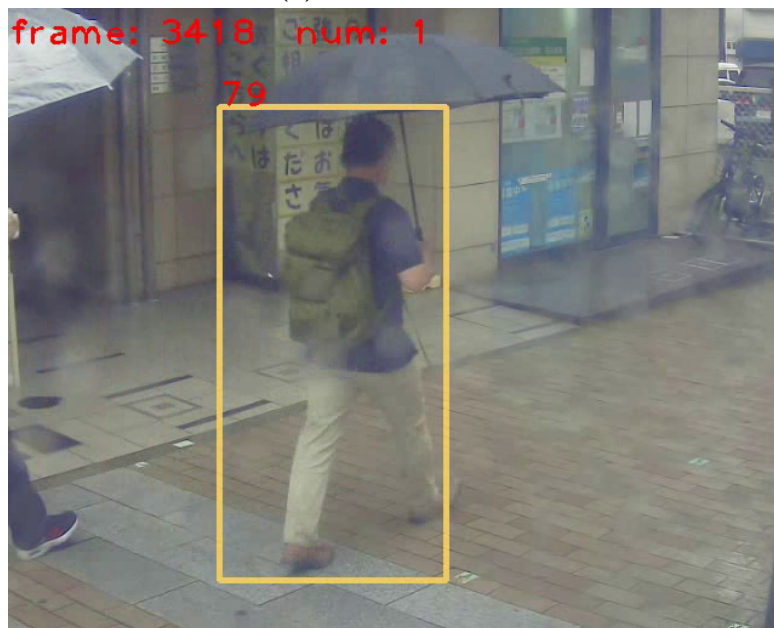


図 35: 可視光映像と融合映像を利用した場合の昼間と夜間の誤差率の箱ひげ図



(a) 3413 フレーム目



(b) 3418 フレーム目

図 36: 昼間・雨天時の可視映像



(a) 3413 フレーム目



(b) 3418 フレーム目

図 37: 昼間・雨天時の融合映像



(a) 可視光映像



(b) 融合映像

図 38: 可視光映像と融合映像を利用した場合の比較例

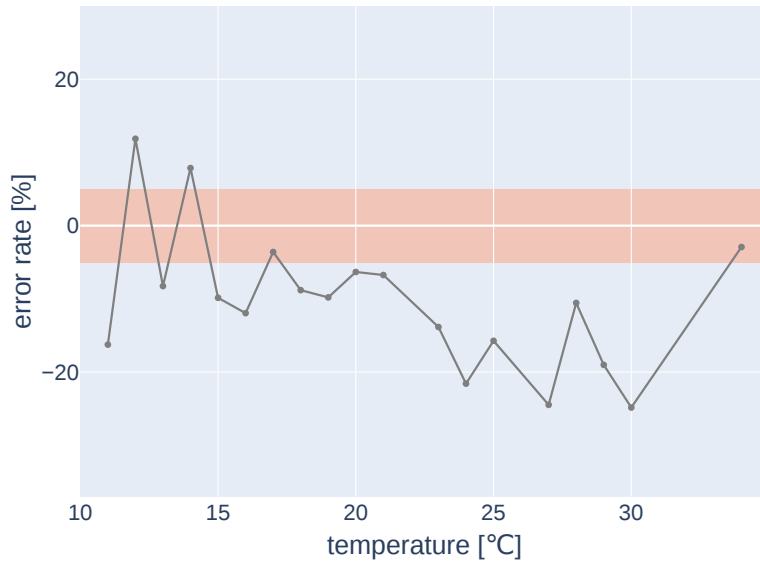
ている。これは遠赤外線映像に対して可視光映像の情報を追加しているためだと考えられる。

6.3 3つの映像を用いた場合の時間と天気ごとの誤差率

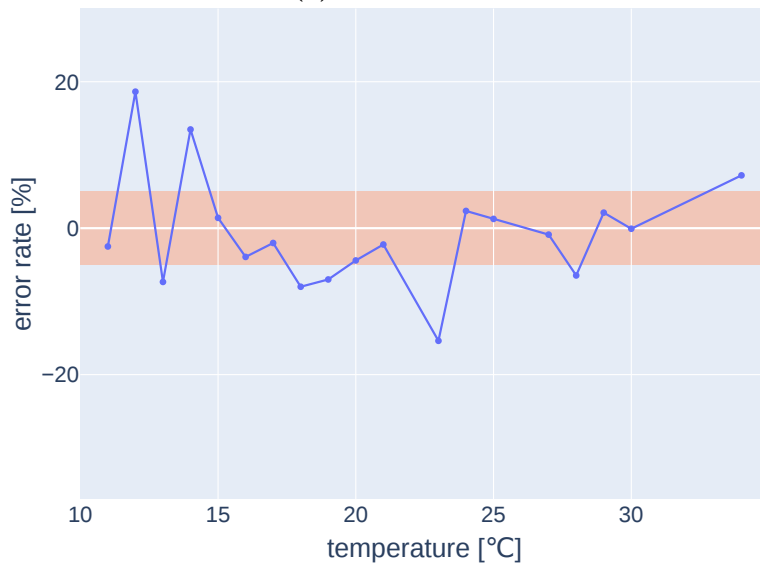
表9は、時間や天気の各4パターンにおける、可視光映像と遠赤外線映像、融合映像を用いた歩行者数推定の誤差率を示す。可視光映像と融合映像を比較すると、昼間の雨天時では可視光映像を利用した方が推定精度が高かったのに対して、その他の条件では、融合映像を利用するほうが推定精度が高いことがわかる。そして、融合映像を用いた歩行者数推定は、可視光映像を利用する場合に比べて最大1.5%の誤差率削減に貢献した。遠赤外線映像のみを利用した場合と、融合映像を利用した場合の結果を比較すると、すべての時間、天気のパターンにおいて融合映像を利用した方が誤差率が削減されており、最大で11.1%の誤差率削減となった。

表 9: 可視光映像と遠赤外線映像、融合映像を利用した場合の誤差率

	昼間		夜間	
	晴天時	雨天時	晴天時	雨天時
可視光映像	4.9% (+)	3.1% (+)	7.0% (-)	7.6% (-)
遠赤外線映像	7.3% (-)	16.2% (-)	14.3% (-)	14.0% (-)
融合映像	3.4% (+)	5.1% (+)	6.7% (-)	6.6% (-)



(a) 遠赤外線映像



(b) 融合映像

図 39: 可視光映像と融合映像の時間ごとの誤差率



(a) 27°C における遠赤外線映像



(b) 27°C における融合映像

図 40: 遠赤外線映像と融合映像を利用した場合の比較例

7. 考察

6章では、融合映像を用いた歩行者数推定が、可視光映像や遠赤外線映像のみを利用する場合よりも推定精度が向上するかを評価した。しかし、遠赤外線映像を利用する場合には気温によって評価結果が異なるため、本章では、気温ごとに各映像を利用した歩行者数推定精度を調査する。その後、今後の展望について述べる。

7.1 気温 24°C 未満の場合

気温が 24°C 未満の場合の誤差率を表 10 に示す。なお、使用した映像は各パターン 7 本ずつである。この表より、昼間は可視光映像と融合映像を利用した方が良いと考えることが出来る。また、夜間の晴天時には可視光映像、夜間の雨天時には融合映像の方が誤差率が低い。遠赤外線映像の誤差率に着目すると、全ての天気を平均した表 9 の結果よりも誤差率が低くなっていることがわかる。さらに夜間の雨天時における可視光映像の誤差率と、遠赤外線映像を利用した場合の誤差率の方を比較すると、遠赤外線映像を利用するほうが誤差率が低い。これは、夜間の雨天時には可視光映像を利用する場合に、雨具の影響や背景との同化が発生しやすい場面で遠赤外線映像が有効にはたらくことを示していると考えられる。

表 10: 気温 24°C 未満の場合の誤差率

	昼間		夜間	
	晴天時	雨天時	晴天時	雨天時
可視光映像	2.1% (+)	1.2% (+)	1.0% (-)	10.7% (-)
遠赤外線映像	7.2% (-)	10.7% (-)	1.8% (-)	7.5% (-)
融合映像	1.0% (-)	2.0% (+)	4.0% (-)	6.5% (-)

7.2 気温 24°C–30°C の場合

気温が 24°C–30°C の場合の誤差率を表 11 に示す。なお、使用した映像は各パターン 5 本ずつである。この表より、昼間は可視光映像を利用するほうが誤差率が低い。夜間の場合、晴天時は融合映像、雨天時は可視光映像の方が誤差率が低い。遠赤外線映像に着目すると、各 4 パターンにおいて表 9 の場合に比べて誤差率が高くなっている。これは、予備実験通り、歩行者の検出漏れを発生させる気温が理由であると考えられる。また、融合映像に着目した場合も同様に表 9 の場合に比べて誤差率が高くなっている。これは、気温 24°C–30°C において遠赤外線映像が一定温度の映像しか撮影できず、融合映像に悪影響を与えたと考えられることが出来る。

表 11: 気温 24°C–30°C の場合の誤差率

	昼間		夜間	
	晴天時	雨天時	晴天時	雨天時
可視光映像	6.6% (+)	4.3% (+)	12.7% (-)	10.7% (-)
遠赤外線映像	9.6% (-)	16.5% (-)	18.2% (-)	36.9% (-)
融合映像	9.9% (-)	7.6% (+)	11.3% (-)	16.4% (-)

7.3 気温 30°C を超過する場合

気温が 30°C を超過する場合の誤差率を表 12 に示す。本研究において取り扱った映像は、気温 30°C を超過する映像が 2 本のみであり、それは昼間の晴天時であった。この表より、気温 30°C を超す場合は、遠赤外線映像が最も誤差率が低い結果となった。これは予備実験でも示したように、背景温度が歩行者の表面温度よりも十分に高いことで、歩行者の検出と追跡が容易になったためであると考えられる。可視光映像に着目すると、他の気温の場合に比べて最も高い誤差率となっている。そこで、図 41 には気温 30°C 超過時の可視光映像の様子と、遠赤外線映像の様子を示す。それぞれの映像の 2868–3586 フレーム目には、日陰で待つ人物が存在し、その人物の手前を歩行者が通ることによって ID の再割り当てが発生

している。図 41(a) の可視光映像の場合には合計 6 回、ID の再割り当てが発生しており、この影響によって大きく誤差率が増加したと考えることが出来る。一方、図 41(b) の会陰赤外線映像の場合には合計 2 回、ID の再割り当てとなっていた。

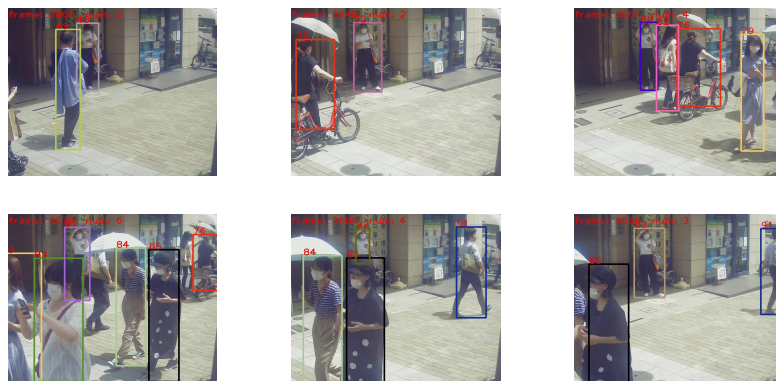
以上のような結果より可視光映像と遠赤外線映像、融合映像を用いて歩行者数の推定を行う場合には、時間帯や天候、気温によって映像を使い分けることにより高精度に歩行者数を推定出来ると考える。

表 12: 気温 30°C を超過する場合の誤差率

	昼間		夜間	
	晴天時	雨天時	晴天時	雨天時
可視光映像	12.7% (+)	-	-	-
遠赤外線映像	2.7% (-)	-	-	-
融合映像	7.3% (+)	-	-	-

7.4 今後の展望

本稿では、可視光映像と遠赤外線映像の融合映像を歩行者数推定に利用することで、推定精度が向上することを示した。本研究を通して今後の展望は 2 つある。1 つ目は、融合映像の転移学習である。本研究において、遠赤外線映像を転移学習させたモデルを利用することによって歩行者の検出を実施した。そのため、融合映像に最適化されたモデルではなく、融合映像を用いた歩行者検出精度が低かった可能性がある。そこで、融合映像を転移学習させたモデルを利用することにより、融合映像を用いた場合の歩行者検出精度が向上し、歩行者数推定精度も向上すると考える。また映像を切り替える際には、それぞれの映像を転移学習させたモデルを利用することによって、さらなる精度改善を見込むことが出来る。2 つ目は、移動方向の推定や歩行者以外への追跡対象を拡張した時の評価である。本研究では、歩道の利用状況として歩行者数に着目したが、カメラ映像は、歩行者の移動方向や自転車・ペット等歩行者以外も追跡できる。そのため、移動方向推定時の評価や、追跡対象を変更した場合の評価が必要であると考えられる。



(a) 可視光映像



(b) 遠赤外線映像

図 41: 気温 30°C 超過時の映像 (2868-3586 フレーム間)

8. 終わりに

本研究では、時間や天気による照度の変化に着目して、歩道の利用状況調査におけるカメラ映像を用いた歩行者数推定精度を改善することに取り組んだ。現在の歩道利用状況調査では、可視光映像が利用されており、雨天時や夜間のような照度が低い状況では、可視光映像中の歩行者を検出することが困難である。そのため、従来の可視光映像を用いた歩行者数推定の精度が低下すると考えられる。一方、夜間における物体検出には、照度に関係なく物体の温度を撮影可能な遠赤外線映像が利用される。歩行者数を推定するために必要な映像は、バス停上部に設置した可視光カメラと遠赤外線カメラを用いて収集した。予備実験では、収集した可視光映像と遠赤外線映像を用いて歩行者数を推定し、時間や天気ごとに仮説を立て、推定精度低下要因を検証した。予備実験の結果、可視光映像は照度が低い夜間の環境において、推定精度が低下することを確認した。また、遠赤外線映像を用いた場合は、気温が24°C–30°C付近の場合に推定精度が低下すると考察した。そして、双方の映像を利用した場合の推定精度が低下する要因が明らかになったため、双方の特徴を利用して、あらゆる時間や天気において歩行者数推定をより高精度にするための手法を提案した。提案手法では、DenseFuseによって可視光映像と遠赤外線映像を融合した映像を利用し、歩行者数を推定する。提案手法のように2つの映像を1つの映像にまとめることで、MS COCOなどの大量の画像を事前に学習したモデルが利用可能となる。Densefuseによる融合映像を用いた歩行者数推定の結果、可視光映像、遠赤外線映像それぞれを利用した場合に比べて最大1.5%、11.1%の推定精度を向上した。そして、時間や天気、気温の状況によって最も誤差率が小さくなる映像の種類を調査した。今後の展望として、可視光映像と遠赤外線映像、融合映像のそれぞれの特徴を活かし、時間帯や気温、天気情報によって利用する映像を変更することができるようなアルゴリズムの設計が挙げられる。

謝辞

本研究を進めるにあたり、既存手法との関連性や、新規性についてご指導をいただいた、主指導教員である本学情報基盤システム学研究室の藤川和利教授に深く感謝いたします。副指導教員であり、研究の方向性についての的確な意見をいただきました、本学大規模システム管理研究室の笠原正治教授に心より感謝いたします。副指導教員であり、研究結果の考察について貴重なご意見をいただきました、本学ユビキタスコンピューティングシステム研究室の安本慶一教授に心より感謝いたします。副指導教員であり、本学情報基盤システム学研究室の新井イスマイル准教授には、実験方法の議論や結果の考察方法など細部にわたるご指導をいただきました。実証実験では、カメラの設置に関わる倫理審査等の手続きをしていただき、円滑な実験へのご協力をいただきました。ここに深く感謝いたします。本学情報基盤システム学研究室の垣内正年助教には、実証実験時の学内システムの利用や遠隔地からの安全なデータ転送における技術的なご指導をいただきました。ここに深く感謝いたします。本学情報基盤システム学研究室の遠藤新助教には、研究の軸となる研究目的の明確化や最適な手法の検討に際し、細部まで論理的なご指導をいただきました。また、日頃から研究の進み具合を気にかけていただきましたことにも、心より感謝申し上げます。そして様々な面から研究活動を支援してくださいました、本学総合情報基盤センターの辻元理恵女史に心より感謝いたします。実証実験を行うにあたり、みなと観光バス株式会社の皆様には、サイネージ・金属箱の加工や実験用ステッカーの貼付など、多くのサポートをしていただきました。提案手法の評価に必要なデータと実証実験環境を提供していただき誠にありがとうございました。本研究の趣旨を理解し快く協力していただいた、兵庫県神戸市三宮をご利用の調査対象者の皆様に心から感謝いたします。そして2年間の博士前期課程をともに過ごし、技術的・精神的にサポートしてくれた本学情報基盤システム学研究室同期の丹羽君、松永君、山村君、並びに本学情報基盤システム学研究室の学生の皆様に感謝申し上げます。最後に、経済面や生活面において多大な支援を頂いた家族に心より感謝いたします。

福田 匠

参考文献

- [1] Y. Yoshimura, Y. Kumakoshi, Y. Fan, S. Milardo, H. Koizumi, P. Santi, J. M. Arias, S. Zheng, and C. Ratti, “Street pedestrianization in urban districts: Economic impacts in spanish cities,” *Cities*, vol. 120, p. 103468, Jan. 2022.
- [2] O. Blaga, “Pedestrian zones as important urban strategies in redeveloping the community - case study: Alba iulia borough park,” *Transylvanian Review of Administrative Sciences*, vol. 9, no. 38, pp. 5–22, Jan. 2013.
- [3] S. Shamsuddin, N. R. A. Hassan, and S. F. I. Bilyamin, “Walkable environment in increasing the liveability of a city,” *Procedia - Social and Behavioral Sciences*, vol. 50, pp. 167–178, 2012.
- [4] 国土交通省道路局 企画課 道路経済調査室, 国土交通省 国土技術政策総合研究所 道路研究室, “CCTV カメラ (AI 解析) の精度に関する報告”, <https://www.mlit.go.jp/road/ir/ir-council/ict/pdf05/02.pdf>, Sep. 2022.
- [5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Aug. 2009, pp. 248–255.
- [6] L. Schauer, M. Werner, and P. Marcus, “Estimating crowd densities and pedestrian flows using Wi-Fi and Bluetooth,” in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ser. MOBIQUITOUS '14. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014, pp. 171–177.
- [7] H. Zhao, and R. Shibasaki, “A novel system for tracking pedestrians using multiple single-row laser-range scanners,” *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35, no. 2, pp. 283–291, Feb. 2005.

- [8] C. Benedek, “3d people surveillance on range data sequences of a rotating lidar,” *Pattern Recognition Letters*, vol. 50, pp. 149–158, Dec. 2014.
- [9] X. Zhao, E. Delleandrea, and L. Chen, “A people counting system based on face detection and tracking in a video,” in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, Oct. 2009, pp. 67–72.
- [10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014, pp. 580–587.
- [11] R. Girshick, “Fast R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1440–1448.
- [12] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems*, vol. 28. Curran Associates, Inc., June 2015, pp. 91–99.
- [13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 779–788.
- [14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and L. Zitnick, “Microsoft COCO: Common objects in context,” in *ECCV. European Conference on Computer Vision*, Sep. 2014.
- [15] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and real-time tracking,” in *2016 IEEE International Conference on Image Processing (ICIP)*, Aug. 2016, pp. 3464–3468.

- [16] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1960.
- [17] Z. Cao, H. Yang, J. Zhao, S. Guo, and L. Li, “Attention fusion for one-stage multispectral pedestrian detection,” *Sensors*, vol. 21, no. 12, June 2021.
- [18] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, “Bytetrack: Multi-object tracking by associating every detection box,” in *Computer Vision – ECCV 2022*. Cham: Springer Nature Switzerland, Oct. 2022, pp. 1–21.
- [19] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, K. Michael, TaoXie, J. Fang, imyhxy, Lorna, Z. Yifu, C. Wong, A. V, D. Montes, Z. Wang, C. Fati, J. Nadar, Laughing, UnglvKitDe, V. Sonck, tkianai, yxNONG, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, “YOLOv5 SOTA: Realtime instance segmentation,” Nov. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.7347926>
- [20] X. Qiu, X. Sun, Y. Chen, and X. Wang, “Pedestrian detection and counting method based on yolov5+deepsort,” in *4th International Symposium on Power Electronics and Control Engineering (ISPECE 2021)*, vol. 12080, International Society for Optics and Photonics. SPIE, Nov. 2021.
- [21] R. Sundararaman, C. De Almeida Braga, E. Marchand, and J. Pettre, “Tracking pedestrian heads in dense crowd,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 3865–3875.
- [22] S. Li, Y. Li, Y. Li, M. Li, and X. Xu, “YOLO-FIRI: Improved yolov5 for infrared image object detection,” *IEEE Access*, vol. 9, pp. 141 861–141 875, Oct. 2021.
- [23] P. F. Tsai, C. H. Liao, and S. M. Yuan, “Using deep learning with thermal

- imaging for human detection in heavy smoke scenarios,” *Sensors*, vol. 22, no. 14, July 2022.
- [24] G. Shen, L. Zhu, J. Lou, S. Shen, Z. Liu, and L. Tang, “Infrared multi-pedestrian tracking in vertical view via siamese convolution network,” *IEEE Access*, vol. 7, pp. 42 718–42 725, Jan. 2019.
- [25] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “YOLOX: Exceeding YOLO series in 2021,” July 2021, <https://github.com/Megvii-BaseDetection/YOLOX>.
- [26] S. Zhang, L. Song, S. Liu, Z. Ge, Z. Li, X. He, and J. Sun, “Workshop on autonomous driving at CVPR 2021: Technical report for streaming perception challenge,” July 2021.
- [27] H. Li, and X.-J. Wu, “Densefuse: A fusion approach to infrared and visible images,” *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, Dec. 2019.

付録

A. 撮影のアナウンスに利用した準備物



学術研究用カメラ 撮影中

Camera for academic research in operation

研究題目:「遠赤外線カメラを活用した環境変化耐性のある
交通・人流計測システムの開発」

実施主体:奈良先端科学技術大学院大学
情報基盤システム学研究室

本研究への問い合わせ窓口
連絡先:✉ inet-info@is.naist.jp
☎ 0743-72-5151

プライバシーポリシー:
<https://inet-lab.naist.jp/announce/>

研究概要



図 42: バス停に設置したステッカー