

# 一時遮蔽が生じる空間における映像と電波強度を併用した人物位置推定\*

牧村 悠司\* 林 大智\* 木挽 哲星\* 酒本 将希\* 西山 正志\*

Estimation of the person's position using video sequences and radio wave strengths in a space with temporary shielding

Yuji MAKIMURA, Daichi HAYASHI, Teppei KOBIKI, Shouki SAKEMOTO, and Masashi NISHIYAMA

The video sequence acquired from the camera can be used to estimate a person's position accurately. We need to pay attention to the fact that the camera does not estimate the person's position when the camera's field of view is unavoidably blocked by temporary shielding. In recent years, several methods have been proposed to estimate the person's position using radio waves, which emit a wavelength longer than that of visible light and are unaffected by shielding. However, collecting training samples consisting of pairs of the radio wave strength and the person's position is difficult because the radio wave strength acquired from wireless devices cannot be intuitively annotated by humans like the video sequences acquired from the camera. We propose a method for estimating the person's position, even temporarily shielded, by combining the camera and the wireless devices. Specifically, our method automatically collects training samples of the radio wave strength and the person's position during no shielding. Our method estimates the person's position using a regression model with only the radio wave strength acquired from the wireless devices when shielding occurs. The experimental results show that the estimation error of the person's position was  $15.9 \pm 5.2$  cm when one person existed at the same time in a poster panel area with temporary shielding and  $18.7 \pm 6.6$  cm when two persons existed at the same time in that area.

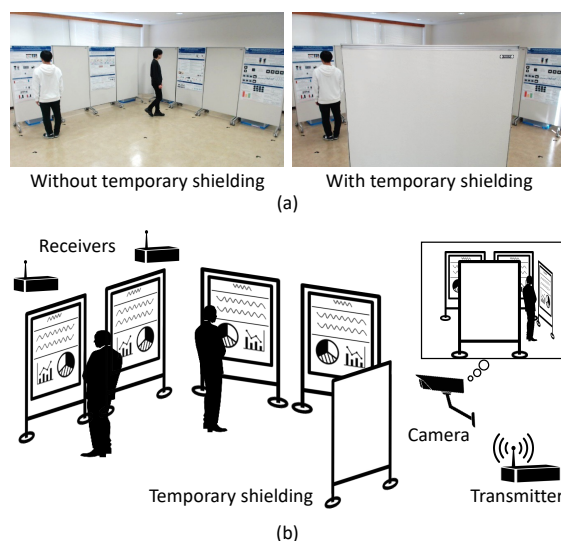
**Key words:** Person, Position, Video sequence, Radio wave strength, Temporary shielding

## 1. はじめに

様々なセンサを用いることで、公共の空間内において人物が存在する位置を推定する技術が求められている。人物位置を推定するためのセンサとして、高い空間分解能で広い範囲を見渡すことができるカメラが一般的である。カメラを用いて映像中の人物位置を推定するため、近年では深層学習を用いた物体検出手法<sup>1,2,3,4)</sup>が広く普及している。これらの手法の発展により、カメラを用いて手軽に高精度な人物位置推定を行うことが可能である。

カメラを敷設する場合、その視野内で遮蔽による人物の隠れが発生しないようレイアウトするが、空間内で物体が一時的に設置されることで、カメラの視野がやむを得ず塞がれる場合が時々発生することに気を付けなければならない。例えば、学会や展示会などのポスター会場では、開催者が案内板を急遽追加するといった目的でパネルを設置することで、カメラ視野に一時的な遮蔽が発生することがある。その例を図1(a)に示す。この場合、遮蔽された奥側に存在する人物の位置を、カメラから取得された映像では推定できなくなることを考えなければならない。

ここで、カメラ視野内に一時的な遮蔽が発生する場合でも、人物位置を安定に推定できるセンサがあるかどうかについて考える。近年、可視光よりも長い波長で遮蔽物を透過できる電波を発する無線機をセンサとして利用する手法が盛んに研究されている。例えば、受信信号強度 (RSSI: Received Signal Strength Indicator) の時系列信号を無線機から取得し、人物位置を推定



**Fig. 1** We develop a method for estimating the person's position stably even when people are temporarily shielded in the field of view of the camera. In (a), the camera cannot estimate the person's position due to temporary shielding. As shown in (b), our method estimates the person's position using video sequences acquired from the camera and radio wave strengths acquired from wireless devices.

する手法<sup>5,6)</sup>や人数を推定する手法<sup>7)</sup>が提案されている。また、無線機のチャンネル状態情報 (CSI: Channel State Information) を用いて、人物姿勢を推定する手法<sup>8)</sup>やジェスチャを認識する手法<sup>9)</sup>が提案されている。さらに近年では、人物姿勢に加えて体型も同時に推定する手法<sup>10,11,12)</sup>が提案されている。これらの既存手法の結果が示すように、無線機は、送信機と受信機との間に遮蔽物が存在したとしても、人物に関する電波の時系列

\* 原稿受付 令和\*年\*月\*日

\* 掲載決定 令和\*年\*月\*日

\* 鳥取大学大学院工学研究科 (鳥取市湖山町南 4 丁目 101)

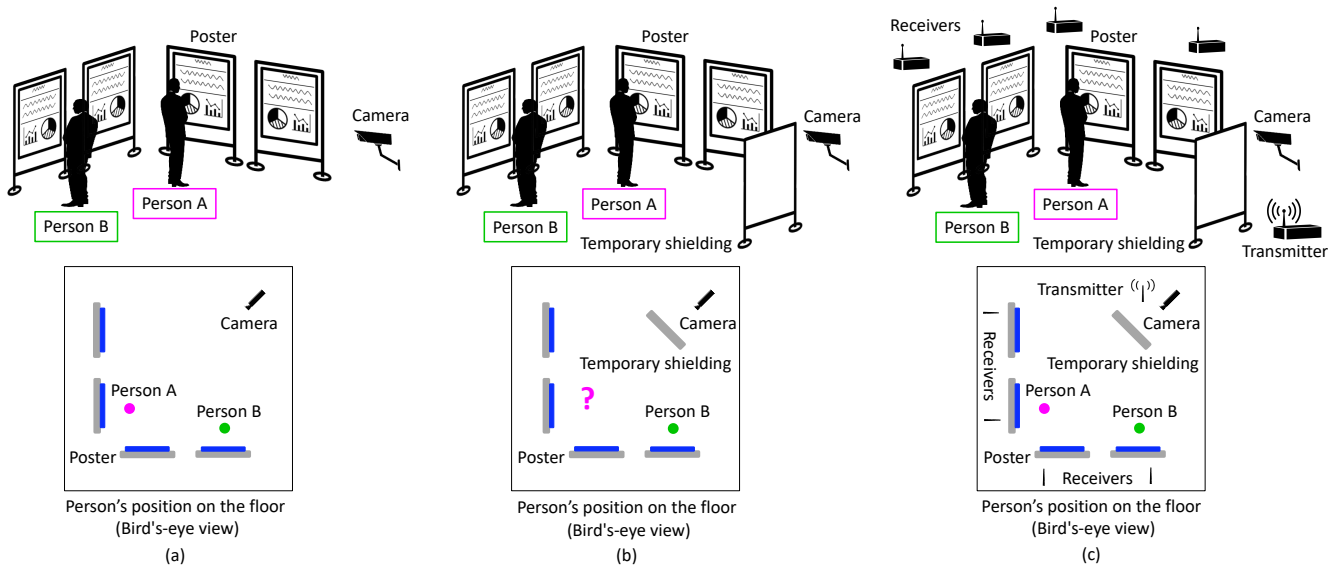


Fig. 2 Estimation of the person's position in a public poster panel area with/without temporarily shielding.

信号を取得できる利点がある。特に無線機の一つである Wi-Fi は幅広く普及しており、手軽に利用しやすい利点もある。

ただし、無線機をセンサとして利用する場合、特に上記で述べた既存手法は統計的な学習に基づくため、大量の訓練サンプルをどのように収集するかを考える必要がある。例えば、人物位置推定の既存手法<sup>5,7)</sup>の場合、無線機で取得された電波の時系列信号に加えて、その空間内の何処に人物が存在するかを表す教師信号が必要になる。ただし、無線機から取得された電波信号は、カメラから取得された映像のように、空間が可視化されていないため、人間が目で見ても直感的に理解できない。このため、空間内で人物が存在する位置を表す教師信号をどのように与えるかが難しくなる。例えば、無線機の RSSI を利用して受信機を保持する人物の立ち位置を推定することを想定した公開データセット RSSI Dataset for Indoor Localization Fingerprinting<sup>13)</sup>では、人物位置に相当する受信機が空間中のどこに位置していたかを、データ収集時に全て手作業で記録している。人物位置推定において、無線機をセンサとして利用するためには、電波強度と人物位置とのペアからなる訓練サンプルの収集をどのように実施するかを考えなければならない課題がある。

そこで本論文では、一時的な遮蔽が生じる空間でも人物位置を安定に推定するため、カメラと無線機とを併用することで、電波強度と人物位置とのペアからなる訓練サンプルを遮蔽がない時に自動で収集し、いざ遮蔽が生じたとしても電波強度のみで人物位置を推定できる手法について述べる。提案手法では、図 1(b) に示す通り、カメラに加えて、電波の送信機と受信機とを空間に設定する。なお人物は、スマートフォンなどの電波を送受信する機器を全く保持していなくてよいとする。まず本論文では評価実験のため、カメラと無線機とを用いて、遮蔽がない場合と遮蔽がある場合について、複数人が見学する Poster Panel Area (PPA) データセットを構築する。遮蔽物を透過する無線の周波数帯域は様々に存在するが、本論文では広く普及している Wi-Fi の 5 GHz 帯に着目し、手軽に取得可能な電波信号である RSSI を使用する。提案手法では、一時的な遮蔽が存在しない通常時に、RSSI 時系列信号と人物位置の教師

信号とのペアからなる訓練サンプルを収集する。遮蔽がない場合、カメラから取得された映像を用いて、人物位置を高精度に推定することができる。カメラ視野内に一時的な遮蔽が発生した場合、訓練サンプルを用いて学習された回帰モデルを用いることで、無線機から取得された RSSI 時系列信号のみに基づき人物位置の推定を行う。実験では、一時的な遮蔽が発生した場合、カメラ視野の端で人物を捉えることができるが、その視野の中央付近に遮蔽物が設置され、カメラから取得された映像のみでは人物位置を推定できないという前提条件を設けた。実験結果より、上記前提条件の下において、RSSI 時系列信号から推定された人物位置の誤差は、同時に見学する人数が 1 名の場合に  $15.9 \pm 5.2$  cm、同時に見学する人数が 2 名の場合に  $18.7 \pm 6.6$  cm であった。

## 2. Poster Panel Area (PPA) データセットの構築

### 2.1 想定する空間

人物位置を推定する手法を応用する空間として、成果発表のためのポスター会場、新商品を宣伝するための見本市会場、絵画や彫刻などの美術品を鑑賞するための展示会場などが挙げられる。以下では、ポスター会場の空間について、詳細を掘り下げていく。この空間に来場した人物は、展示されている多数のポスターを見学することを目的としている。来場した人物は、興味のあるポスターの前で立ち止まりその中身を読んだり、ポスターを一瞥して興味がなければ立ち去ることがある。なお、ポスター会場での一般的なセッションでは、それぞれのポスターに説明員が存在しているが、ここでは休憩時間などポスターは張られているが、説明員は不在の時間帯を想定して議論を進める。この想定で人物位置推定を行うことは、説明員不在時のポスターの視聴率計測の応用に繋がると考えられる。以下では、この想定のパネル会場で、人物位置を推定することを考えていく。

人物が多数集まる公共の空間にはカメラが設置されていることが多く、これらのカメラを用いて、来場する人物の位置を推定することが一般的である。ポスター会場におけるカメラを用いた人物位置推定の例を図 2(a) に示す。カメラ視野の中に存

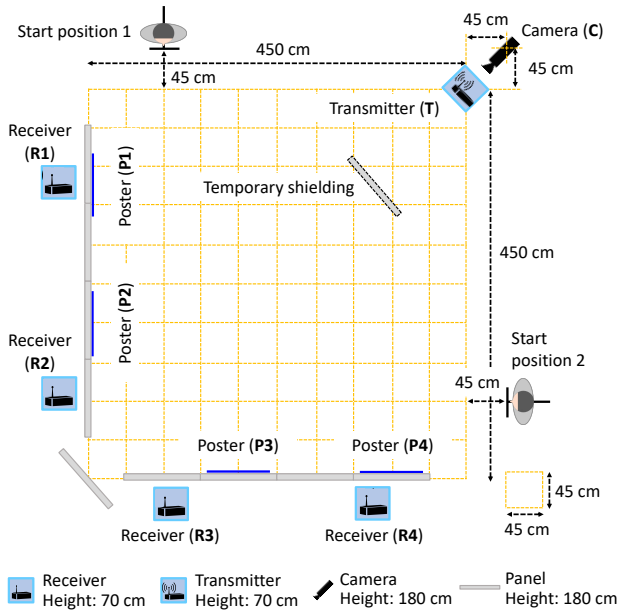
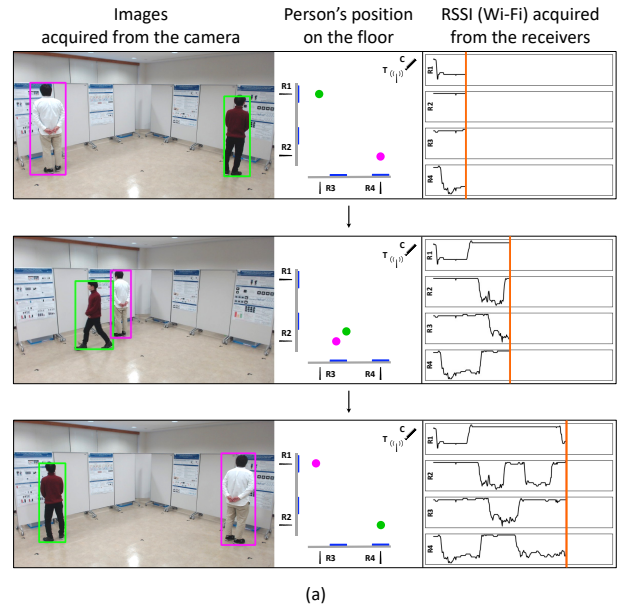


Fig. 3 Experimental setting of the PPA dataset.

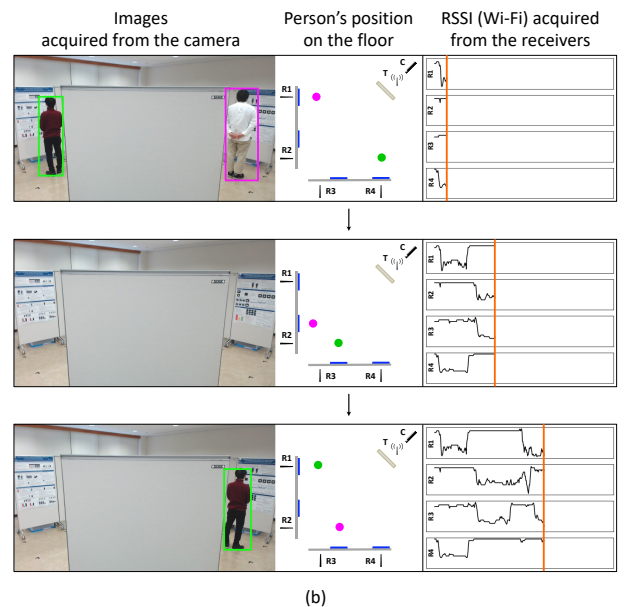
在する人物については、それらの位置を精度よく推定することができる。ただし、1. 節で述べたように、カメラを遮蔽する物体が一時的に設置されることが時々ある。例えば、初日にポスター会場が案内が悪く迷う人物が多かった場合、次の日に開催者が案内版を急遽設置し、カメラの視野を一時的に塞ぐことがある。カメラの前に一時的な遮蔽が存在する場合、カメラから人物が見えなくなるため、人物位置を推定することができない。その遮蔽の例を図 2(b) に示す。この例では、カメラ視野に入る人物 B の位置を推定できるが、遮蔽で視野に入らない人物 A の位置を推定できない。このように一時的な遮蔽が発生する空間でも安定して人物位置を推定するため、本論文では、無線機から取得される RSSI を、カメラから取得される映像と併用することを考えている。その併用の例を図 2(c) に示す。遮蔽されている人物 A の位置を、遮蔽前に学習された帰帰モデルを用いて、無線機の RSSI から推定する。なお、ポスター会場など人物が集まる空間には、カメラだけでなく、来場者がネットワークを利用するための Wi-Fi 無線機が設置されていることが多い。カメラと無線機との併用による人物位置推定は、将来的な応用の可能性をもつと考える。

## 2.2 セッティング

ポスター会場を想定した PPA データセットを構築するためのセッティングを図 3 に示す。本論文では、ポスターを 4 枚準備し、それらを張り付けるパネルを 9 枚設置した。それぞれのポスターは、図中の P1, P2, P3, P4 の位置に設置されたパネルへ張り付けた。ポスターを張り付ける各パネルは、高さが 180 cm で幅が 90 cm で厚みが 4 cm とした。カメラの前に設置される一時的な遮蔽物として、パネル 1 枚をさらに追加した。次に、センサとして、カメラを 1 台、Wi-Fi 受信機を 4 台、Wi-Fi 送信機を 1 台設置した。カメラの位置は、この空間を見渡せるよう図 3 右上に設置されており、その高さは 180 cm とした。Wi-Fi 送信機をカメラ下部に配置し、その高さは 70 cm とした。それぞれの Wi-Fi 受信機は、図中の R1, R2, R3, R4 の位置に存在し、カメラから見てパネルの背後に高さ 70 cm で設置した。



(a)



(b)

Fig. 4 Examples of the sequences contained in the PPA dataset.

人物位置の推定精度を評価するため、一時的な遮蔽が存在する場合の人物位置の正解信号が必要である。ただし図 3 の環境では、遮蔽時に、ポスター P2 や P3 の付近に存在する人物の位置をカメラから獲得することができない。本論文では、正解信号付与のための補助カメラをさらに別の位置へ設置した。この補助カメラは、一時的な遮蔽の有無にかかわらず、全ての人物が視野に入り、それらの位置を推定できる配置とした。なお、この補助カメラで取得された人物位置は、精度を評価する際に、予測時のテストサンプルの正解信号としてのみ扱い、学習時の訓練サンプルには含めないようにした。

## 2.3 映像と電波強度と人物位置とからなるシーケンス

構築された PPA データセットに含まれるシーケンスの例を図 4 に示す。左から順に、カメラで撮影された画像、人物位置を示す俯瞰図、Wi-Fi 受信機の RSSI 時系列信号のグラフである。カメラで撮影された画像中の人物の枠の色と俯瞰図の人物

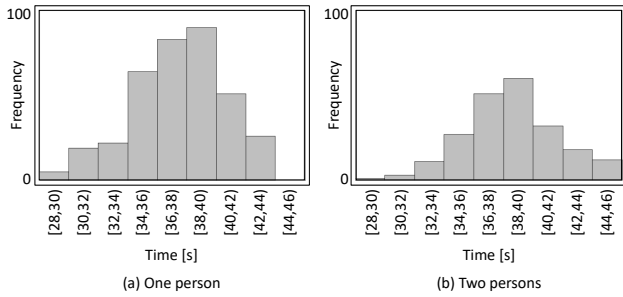


Fig. 5 Histograms of the time length of the sequences.

位置の色とが対応している。俯瞰図では、空間を真上から見ており、床面座標系における人物位置を表す。RSSI 時系列信号のグラフでは、横軸が時刻、縦軸が Wi-Fi 受信機から取得された RSSI の値を表す。Wi-Fi 受信機ごとに RSSI の値が出力されており、上から順に R1, R2, R3, R4 の受信機に該当する。なお各人物は、スマートフォンなどの電波を送受信する機器を保持しないこととした。

PPA データセットでは、空間内で同時に見学している人数が、1 名の場合と 2 名の場合とで、それぞれシーケンスを収集した。合計 576 個のシーケンスを収集しており、その内訳を以下に示す。同時見学の人数が 1 名の場合、(実験協力者) 15 人 × (歩行パターン) 12 種類 × (遮蔽有無) 2 種類 = 360 個のシーケンスを収集した。同時見学の人数が 2 名の場合、(実験協力者ペア) 9 組 × (歩行パターン) 12 種類 × (遮蔽有無) 2 種類 = 216 個のシーケンスを収集した。次に、同時見学の人数が 1 名の場合の歩行パターンについて以下で説明する。歩行経路として、図 3 左上に示した開始位置 1 から入りポスター P1 から P4 の順に見て出る経路と、図中右下に示した開始位置 2 から入りポスター P4 から P1 の順に見て出る経路との 2 種類を準備した。また、ポスターの見方として、4 枚のポスターの中からランダムに選ばれた 2 枚のポスターを注視して、選ばれなかった残り 2 枚のポスターを一瞥するように、6 種類を準備した。なお、注視の場合はポスターに記載されているタイトルと目的とを理解しながらじっくりと読むこととし、一瞥の場合はタイトルを読んだ後に流し見することにした。以上より、同時見学する人数が 1 名の場合は (歩行経路) 2 種類 × (見方) 6 種類 = 12 種類の歩行パターンとした。なお、同時見学の人数が 2 名の場合、上記のポスターの見方が人物ごとに 6 種類のため、その 2 乗の 36 種類となるが、ここでは各人物が見方をランダムに選択することで、歩行パターンを 12 種類とした。また、歩行経路の開始位置として、2 名が重ならないように、それぞれ開始位置 1 と開始位置 2 とをランダムに割り振った。

PPA データセットに含まれる各シーケンスの時間長の違いについて、その分布を可視化した結果を図 5 に示す。図中 (a) は同時見学する人数が 1 名の場合であり、(b) は同時見学する人数が 2 名の場合である。各グラフの縦軸はシーケンスの出現頻度であり、横軸は時間長のビンである。各シーケンスの時間長の違いは、ポスターをじっくりと読む時間や流し見する時間が、人物の間で異なるため生じていた。なお、シーケンス時間長の平均と標準偏差は、1 名の場合で  $36.7 \pm 3.0$  秒、2 名の場合で  $38.0 \pm 3.1$  秒であった。

Table 1 Averages (dBm) of RSSI with the presence of a standing person and the presence of a temporary shielding

(a) W/O temporary shielding		
Receiver	No standing person	Standing person
R1	$-22.0 \pm 1.0$	$-32.4 \pm 2.6$
R2	$-26.7 \pm 0.6$	$-36.7 \pm 2.3$
R3	$-26.0 \pm 0.7$	$-37.1 \pm 2.1$
R4	$-24.7 \pm 1.0$	$-36.0 \pm 1.7$

(b) W/ temporary shielding		
Receiver	No standing person	Standing person
R1	$-24.3 \pm 0.9$	$-34.8 \pm 1.7$
R2	$-27.7 \pm 1.0$	$-37.4 \pm 2.4$
R3	$-28.2 \pm 0.9$	$-37.8 \pm 1.8$
R4	$-25.7 \pm 1.0$	$-37.1 \pm 2.1$

## 2.4 RSSI の振る舞い

本論文のセッティングにおいて、人物や遮蔽物の有無により RSSI の値がどのように変化するかを調査した。その結果を表 1 に示す。表中 (a) は一時的な遮蔽がない条件であり、(b) は一時的な遮蔽がある条件とした。さらに (a) と (b) の条件の中で、各受信機の手前に、人物が存在しない場合と存在する場合との条件を設けた。各条件において、600 ミリ秒間の RSSI 時系列信号を 15 名分について計測し平均と標準偏差とを求めた。人物が存在しない時に比べて、人物が存在する時に RSSI の値が大きく低下した。また、遮蔽物が存在しない時に比べて、遮蔽物が存在する時に RSSI の値が僅かに低下した。Wi-Fi 受信機の R1 から R4 は同じ型番であるが、それらの設置位置が異なるため、各受信機で RSSI の値が違っていた。ただし、どの受信機においても、遮蔽物の有無と比べて、人物の有無は、RSSI の減衰へ大きく影響を与えていると言える。この調査結果を踏まえて図 4 を見ると、各人物が存在する位置付近に設置されている受信機において、RSSI の値が大きく低下していることが分かる。なお、図 4 右のグラフの縦軸は RSSI の値 (単位は dBm) を表し、横軸は時刻 (単位は秒) を表す。橙色の縦線は現在時刻を指し示しており、カメラで撮影された画像の時刻と同期している。

## 2.5 データセット構築で使用した機器

PPA データセットを構築するために利用した機器の詳細について述べる。無線機として、図 3 の Wi-Fi 送信機 (T) の位置に設置したアクセスポイント (Buffalo 社, WXR-5700AX7S) を利用した。また、Wi-Fi 受信機 (R1 から R4) の位置に設置した筐体にアンテナ (Buffalo 社, WI-U2-433DHP) を取り付けた。それぞれの受信機の筐体は Raspberry Pi 4 とし、OS に Kali Linux 2020.1 を用いた。送信機 (T) と受信機 (R1 から R4) との間の周波数帯域は IEEE 802.11n の 5 GHz に設定した。送信機 (T) において、ビームフォーミングをオフにし、また、周波数帯のチャンネルを固定した。受信機 (R1 から R4) で Wi-Fi の RSSI 時系列信号を取得するため、Aircrack-ng を用いて Monitor モードに設定した。送信機 (T) から 102 ミリ秒間隔で発せられるビーコンを用いて、RSSI 時系列信号を受信機 (R1 から R4) で記録した。

映像を取得するためのカメラとして、USB カメラ (Logicool

社, C920n) を用い, Ubuntu 18.04 の v4l-utils ライブラリを用いた. カメラのフレームレートを 30 fps とし, 解像度を 1920 × 1080 画素とした. RSSI 時系列信号とカメラ映像との間で時間を同期させるため, NTP サーバを利用した. 各シーケンスを取得した後, カメラと無線機との間でサンプリングレートを揃えるため, RSSI 時系列信号について, 9.8 Hz から 30 Hz に線形補間を用いてアップサンプリングした.

### 3. 推定手法

#### 3.1 概要

本論文では, 1. 節で述べたように, 一時的な遮蔽が発生した時, 無線機から取得された RSSI 時系列信号のみを入力とする回帰モデルを用いて人物位置を推定していく. この回帰モデルを学習するため, RSSI 時系列信号と人物位置とのペアからなる訓練サンプルが必要となる. 推定手法では, 遮蔽がない場合において, 回帰モデルのための訓練サンプルを自動で収集し, 遮蔽がある場合において, 先に収集された訓練サンプルから学習された回帰モデルを適用する. 具体的には, 遮蔽がない場合, カメラを用いて人物位置を教師信号として取得し, それと同時に無線機を用いて RSSI 時系列信号も取得する. これにより, 人物位置の教師信号を与えることが目視では難しい RSSI 時系列信号に対して, その訓練サンプル収集を自動で処理できるようになる. 一方, 遮蔽がある場合, RSSI 時系列信号に前処理を施し, 一般的な機械学習に基づく回帰モデルを適用する. これにより, 一時的な遮蔽がやむを得ず発生し, カメラから取得された映像において人物位置を推定できなくなったとしても, 無線機の RSSI 時系列信号を用いて円滑に人物位置を推定できるようになる.

#### 3.2 流れ

推定手法の詳細な流れについて説明する. 遮蔽がない場合, 図 6 上に示す流れにより, 人物位置と RSSI 時系列信号とのペアからなる訓練サンプルを学習時に収集する. 無線機の RSSI 時系列信号は, カメラ映像と比べて空間分解能が低く, 人物の有無による空間内の密な変化を取得することが難しい. 推定手法では, 現在時刻  $t$  から過去時刻  $t - T$  までの短時系列信号を, RSSI 時系列信号からサンプリングし, 訓練サンプルとして参照することで, 人物位置の推定精度の向上を狙っていく. 時刻  $t$  において,  $N$  個の Wi-Fi 受信機 R1 から Rn で取得された RSSI の値をそれぞれ  $r_{1,t}$  から  $r_{n,t}$  とすると, 訓練時に使用する RSSI の短時系列信号  $\mathbf{R}_t$  は次式で表される.

$$\mathbf{R}_t = \begin{pmatrix} r_{1,t} & \cdots & r_{1,t-T} \\ \vdots & \ddots & \vdots \\ r_{n,t} & \cdots & r_{n,t-T} \end{pmatrix} \quad (1)$$

また, カメラ映像から獲得された人物位置を, 床面座標系における  $\mathbf{p}_t = (x_t, y_t)$  とする. 以下では, 1 個の訓練サンプルを  $\langle \mathbf{R}_t, \mathbf{p}_t \rangle$  とする. RSSI 時系列信号には受信機ごとの特性の違いなどのノイズが含まれるため, 前処理を適用しノイズを除去する. 訓練サンプルが一定数ほど収集されると, 機械学習に基づく回帰モデルを学習する. 学習時にはサイズが  $N \times (T + 1)$  の行列  $\mathbf{R}_t$  を  $N \cdot (T + 1)$  次元のベクトルに変換する. なお, 人物位置の獲得方法について 3.3 節で述べ, 前処理の方法を 3.4 節で述べ, 回帰モデルについて 4.3 節で述べる.

次に, 遮蔽がある場合, 図 6 下に示す流れにより, RSSI 時系列信号のみを用いて人物位置を推定する. 時刻  $\hat{t}$  において,

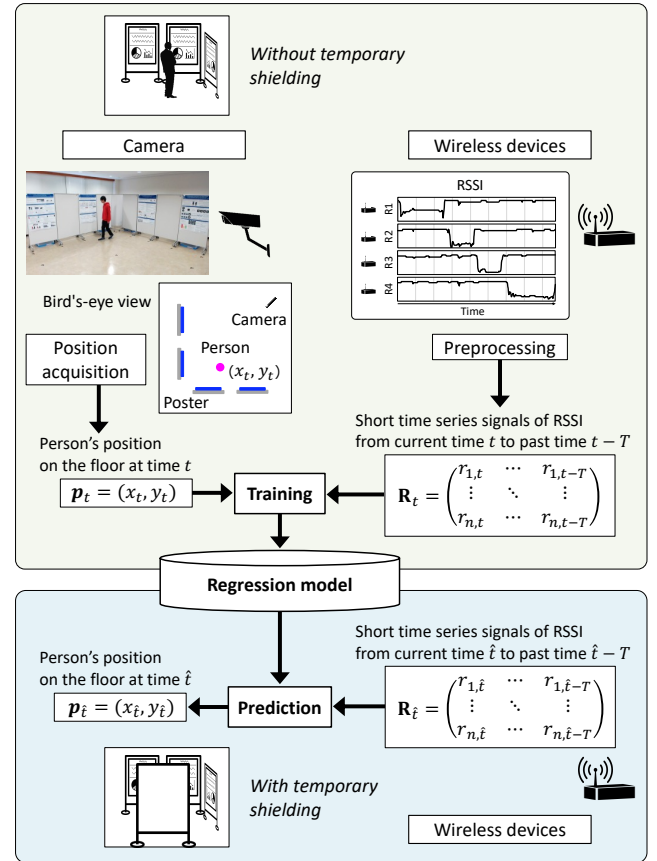


Fig. 6 Overview of our estimation method.

予測時に回帰モデルへ入力される RSSI の短時系列信号  $\mathbf{R}_{\hat{t}}$  を, 学習時の訓練サンプルと同様に次式で表す.

$$\mathbf{R}_{\hat{t}} = \begin{pmatrix} r_{1,\hat{t}} & \cdots & r_{1,\hat{t}-T} \\ \vdots & \ddots & \vdots \\ r_{n,\hat{t}} & \cdots & r_{n,\hat{t}-T} \end{pmatrix} \quad (2)$$

予測時の前処理も, 学習時と同様に短時系列信号  $\mathbf{R}_{\hat{t}}$  へ適用する. なお予測時に, 回帰モデルから出力される時刻  $\hat{t}$  の人物位置を  $\mathbf{p}_{\hat{t}} = (x_{\hat{t}}, y_{\hat{t}})$  とする.

ここまでは空間内で同時に見学している人数が 1 名の場合で説明したが, 2 名以上になる場合について説明する. 推定手法では, 時刻  $t$  の人物位置  $\mathbf{p}_t$  の要素へ, 各人物の位置を単純に追加していく. 例えば, 同時見学の人数が 2 名の場合は,  $\mathbf{p}_t = (x_{1t}, y_{1t}, x_{2t}, y_{2t})$  とする. この  $\mathbf{p}_t$  を用いる場合, 遮蔽内に存在する人数に応じて回帰モデルを切り替える必要がある. 本論文では, 1. 節で述べたように, 遮蔽物はカメラ視野の中央付近に発生する前提条件を設け, カメラ視野の端では人物が見えており, 遮蔽内に入った人数はカウントできていると仮定する.

#### 3.3 遮蔽がない場合における訓練サンプル収集のための人物位置の獲得

遮蔽がない場合において, カメラから取得された映像を用いて, 訓練サンプルを収集するための人物位置を獲得する方法を述べる. 本論文では, カメラから取得された時刻  $t$  の画像において, Mask R-CNN<sup>1)</sup> を用いて人物領域マスクを推定する. この人物領域マスクから, 足元重心の位置推定<sup>14)</sup>を用いて, 床面座標系での人物位置を取得する. 足元重心を推定するため,

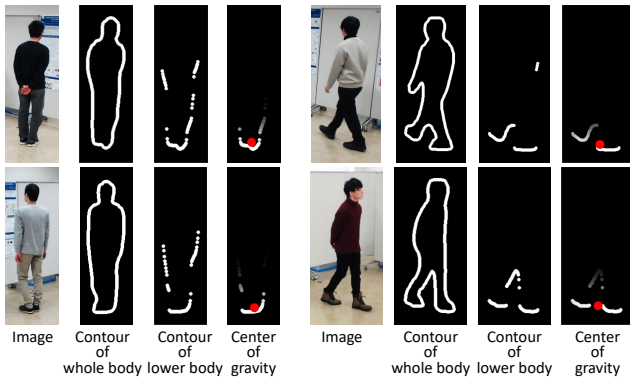


Fig. 7 Examples of the estimation of the position of the center of gravity of the feet.

人物領域マスクから全身輪郭を求め、脚部に該当する領域下部の輪郭のみを用いて加重平均を行い、重心の位置を計算する。その例を図7に示す。得られた画像座標系での足元重心の位置を、ホモグラフィ行列を用いて射影変換することで、空間の床面座標系における時刻  $t$  の人物位置  $\mathbf{p}_t$  を決定する。なお、ホモグラフィ行列を求めるため、床面の特徴点を事前に手作業で教えることとする。

足元重心から推定された人物位置には、歩行時の足の開閉に起因するノイズが含まれることがある。このノイズを除去するため、カメラ映像において時間方向の人物追跡を適用することで、各人物について位置の時系列信号を求める。得られた位置の時系列信号に対して加重移動平均フィルタを適用する。なお歩行時に、ある人物が別の人物の一部を隠すことが発生するため、人物領域から算出されたカラーヒストグラムを用いることで、時刻間の人物領域の対応付けを行い追跡処理を実施する。

### 3.4 RSSI 時系列信号に適用する前処理

Wi-Fi 受信機から取得された RSSI 時系列信号へ適用する前処理について述べる。  $N$  個の受信機で時刻  $t$  に取得された RSSI の値  $r_{1,t}$  から  $r_{n,t}$  において、受信機の配置や伝送路の空間特性に違いがあるため、人物が不在の時でも、受信機ごとにその値が異なることがよく発生する。推定手法では、人物不在時の RSSI の値を受信機の間で揃えるため、受信機ごとに人物不在時のオフセットを求め、取得された全ての  $r_{1,t}$  から  $r_{n,t}$  において、受信機ごとにオフセットを差し引くことにする。また、取得された RSSI の値において、複数の人物が存在する時に、ノイズのために電波の減衰度合いが受信機ごとに異なることがある。推定手法では、人物存在時に減衰した RSSI の値を揃えるため、  $r_{1,t}$  から  $r_{n,t}$  において閾値  $\tilde{r}$  以下の値をとる場合、ノイズにより電波強度が大きく低下しているとし、それぞれ  $\tilde{r}$  に置き換えることにする。

## 4. 実験

### 4.1 実験条件

提案手法の有効性を確認するため、2.節で述べた PPA データセットを用いて人物位置の推定精度を評価した。学習時に、図4(a)に示した遮蔽がない場合の人物位置と RSSI 時系列信号とのシーケンスを用いて、図6上に示した流れで訓練サンプルを収集した。また予測時に、図4(b)に示した遮蔽がある場合の RSSI 時系列信号のみのシーケンスを用いて、図6下に示した流れで人物位置を推定した。なお提案手法では、3.2節で述べ

たように、空間内で同時に見学する人数に応じて回帰モデルを切り替える必要がある。PPA データセットでは、遮蔽がある場合でもカメラ視野の両端では人物を捉えることができると前提条件を置いた。予測時に人物がカメラ視野へ入る時刻においては、3.3節で述べた方法を用いて映像から人物位置を算出し、遮蔽された空間に入る人数をカウントすることで予測時の回帰モデルを切り替えた。

提案手法について、実験で用いたパラメータを述べる。RSSI 時系列信号から、3.2節で述べた短時系列信号  $\mathbf{R}_t, \mathbf{R}_f$  を生成する時に参照する過去時刻を  $T = 90$ 、Wi-Fi 受信機の個数を  $N = 4$  とした。各シーケンスにおいて、ポスターを見学する時の速さが人物ごとに異なるため、収集できる訓練サンプルの個数はシーケンス毎に異なり、平均で 473.1 個、最大で 974 個、最小で 226 個であった。学習時の遮蔽がない場合において、3.3節で述べた位置の時系列信号へ適用する加重移動平均フィルタでは、現在時刻から過去時刻に向かって 45 サンプルを用いて平滑化した。RSSI 時系列信号に適用する前処理において、3.4節で述べた閾値を  $\tilde{r} = -35$  dBm とした。

### 4.2 誤差の算出方法

人物位置の推定精度の評価指標として、各時刻において推定された人物位置と正解信号の人物位置との間の誤差を用いた。具体的には、それらの位置の間でユークリッド距離を求めた。同時見学する人数が 1 名の場合、予測時に出力される人物位置  $\mathbf{p}_f$  は 1 名分のみであるため、誤差を単純に算出できる。ただし、同時見学する人数が 2 名の場合、誤差の計算時に注意が必要である。これは、推定された人物位置  $\mathbf{p}_f$  に 2 名分が含まれており、また、正解信号の人物位置にも 2 名分が含まれており、どの予測値と正解値とのペアから誤差を算出するかを決める必要があるためである。本論文では、推定された人物位置と正解信号の人物位置との間で、全てのペアで誤差を求め、人物が重ならないように誤差が小さい方から 2 名分のペアを選択し、そのペアに対する誤差を算出した。なお、誤差を計算する際、遮蔽により視野が塞がれており、カメラから取得された映像では人物位置を推定できない時刻のみを対象とした。具体的には、同時見学する人数が 1 名の場合においては、その人物が遮蔽物に隠れている時刻を対象とし、同時見学する人数が 2 名の場合においては、双方ともが遮蔽物に隠れている時刻を対象とした。

### 4.3 回帰モデル

機械学習に基づく回帰モデルとして以下を用いることで、人物位置の推定精度を比較した。

- **LR:** 線形回帰 (Linear regression) の一つである Ridge 回帰を用いた。ここでは、正則化項の係数を  $\alpha = 1$  とした。
- **PLS:** 部分的最小二乗回帰 (Partial least squares regression) を用いた。ここでは、成分数を 10 とした。
- **MLP:** 多層パーセプトロン (Multilayer perceptron) 回帰を用いた。ここでは、線形層を 3 層設け、それぞれの層において活性化関数の ReLU を適用した。学習率を 0.001 とし、Adaptive moment で最適化した。
- **RF:** ランダムフォレスト (Random forest) 回帰を用いた。ここでは、決定木の数を 100 とし、決定木の最大深度を 13 とした。
- **XGB:** XGBoost<sup>15)</sup> 回帰を用いた。ここでは、決定木の数を 100 とし、決定木の最大深度を 13 とし、学習率を 0.1

**Table 2** Error (cm) of the person's position estimation when *one person* existed at the same time in the temporary shielding area.

Regression	W/O preprocessing	W/ preprocessing
LR	27.0 ± 7.8	26.6 ± 7.6
PLS	27.2 ± 7.9	26.8 ± 7.6
MLP	18.5 ± 6.7	16.2 ± 4.7
RF	20.4 ± 10.2	12.4 ± 3.8
XGB	20.5 ± 9.2	13.1 ± 3.8
kNN	17.9 ± 5.9	15.9 ± 5.2

**Table 3** Error (cm) of the person's position estimation when *two persons* existed at the same time in the temporary shielding area.

Regression	W/O preprocessing	W/ preprocessing
LR	61.7 ± 11.2	64.1 ± 10.1
PLS	61.6 ± 11.0	64.0 ± 10.0
MLP	43.0 ± 11.3	38.1 ± 8.6
RF	48.2 ± 14.5	36.6 ± 11.9
XGB	51.1 ± 18.3	40.5 ± 11.8
kNN	22.0 ± 9.6	18.7 ± 6.6

とした。

- **kNN:**  $k$  近傍法 ( $k$ -nearest neighbor algorithm) 回帰を用いた。ここでは、近傍サンプル数を  $k = 5$  とし、重み付き平均を出力した。KD-Tree 構造を採用し、そのリーフサイズを 75 とした。距離尺度はマンハッタン距離とした。

誤差を算出する際、同時に見学する人数が 1 人の場合では Leave-one-participant-out を適用した。具体的には、2. 節で述べた PPA データセットに含まれる実験協力者 15 名の内、14 名の遮蔽なしのシーケンスを訓練サンプルとし、残り 1 名の遮蔽ありのシーケンスをテストサンプルとした。これを全 15 名について繰り返し、平均誤差を算出した。また、同時に見学する人数が 2 人の場合では Leave-one-pair-out を適用した。具体的には、PPA データセットに含まれる実験協力者ペア 9 組の内、8 組の遮蔽なしのシーケンスを訓練サンプルとし、残り 1 組の遮蔽ありのシーケンスをテストサンプルとした。これを全 9 組について繰り返し、平均誤差を算出した。

#### 4.4 結果

同時見学する人数が 1 人の場合における人物位置の推定誤差を表 2 に示す。表中の数値は、各回帰モデルにおける推定誤差の平均と標準偏差とを表す。まず、4.3 節で述べた回帰モデルに着目すると、多層パーセプトロン (MLP) とランダムフォレスト (RF) と XGBoost (XGB) と  $k$  近傍法 ( $k$ NN) は、Ridge 回帰 (LR) と部分的最小二乗回帰 (PLS) と比べて、誤差を減少させていた。また、3.4 節で述べた前処理に着目すると、前処理を適用した場合は、前処理を適用しない場合と比べて、誤差の改善が見られた。RSSI 時系列信号から推定された人物位置の誤差は、同時に見学する人数が 1 人の場合、RF と前処理とを用いた提案手法で  $12.4 \pm 3.8$  cm であった。日本全国の平均足長は左足で 24.3 cm、右足で 24.2 cm と調査結果<sup>\*1</sup>が報告されて

おり、MLP、RF、XGB もしくは  $k$ NN を用いた提案手法の推定誤差は、その平均足長と比べて小さく、一時遮蔽がやむを得ず発生した時の人物位置推定の可能性を示したと考える。

次に、同時見学する人数が 2 人の場合における人物位置の推定誤差を表 3 に示す。表中の数値は、各回帰モデルにおける推定誤差の平均と標準偏差とを表す。どの回帰モデルであっても、先に述べた人数が 1 人の場合と比べて、誤差が増加していた。ただし、 $k$ NN のみで誤差の増加は僅かであり、LR と PLS と MLP と RF と XGB とで誤差の増加が著しかった。人数が 1 人の場合と同様に、人数が 2 人の場合でも、前処理を適用することで、MLP と RF と XGB と  $k$ NN とにおいて誤差の改善が見られた。RSSI 時系列信号から推定された人物位置の誤差は、同時に見学する人数が 2 人の場合、 $k$ NN と前処理とを用いた提案手法で  $18.7 \pm 6.6$  cm であった。ここでも提案手法の推定誤差は平均足長と比べて小さく、人物位置推定の可能性を示したと考える。

以上の結果より、提案手法では、同時見学する人数が 1 人の場合と 2 人の場合との両方で誤差が小さかった  $k$  近傍法を適用することがよいと言える。また、RSSI 時系列信号に前処理を適用することもよいと言える。 $k$  近傍法と前処理とを適用した時の人物位置の推定誤差は、同時に見学する人数が 1 人の場合に  $15.9 \pm 5.2$  cm、同時に見学する人数が 2 人の場合に  $18.7 \pm 6.6$  cm であった。カメラと無線機とを併用することで、電波強度と人物位置とのペアからなる訓練サンプルを遮蔽がない時に収集し、一時的な遮蔽が生じたとしても電波強度のみで人物位置を推定できる可能性があることを確認した。

一時的な遮蔽が存在する場合において、推定された人物位置の可視化結果を図 8 に示す。図中の左側では、正解信号の人物位置を示し、右側では、提案手法により推定された人物位置を示している。図中では図 4 の中央で示した俯瞰図に人物位置を時系列でプロットした。丸が各人物の現在時刻の位置を表しており、丸に付随する細線が過去時刻の位置を表している。点線枠内は、一時的な遮蔽でカメラから人物を捉えられない空間である。ここでは同時に歩行する人物が 2 人の場合とした。図 8(a) と (b) とのシーケンスでは、一時的な遮蔽が生じたとしても、人物位置を精度よく推定できていることが分かる。ただし、提案手法は、各時刻の人物位置を推定する機能のみを有しており、時間方向に人物を追跡する機能を有していない。このため図 8(c) のように、シーケンスによっては人物位置が時間方向で入れ替わる例も見られた。各人物について時間方向に安定して追跡できる機能を開発することが今後の課題として挙げられる。

#### 4.5 提案手法のパラメータに関する評価

短時系列信号  $\mathbf{R}_t$  や  $\mathbf{R}_i$  をサンプリングする際に用いる過去時刻を制御するパラメータ  $T$  について評価した。ここでは  $T = 0, 30, 60, 90, 120, 150$  とした。なお  $T = 0$  は過去時刻を使用せず現在時刻のみを使用することを指す。回帰モデルとして、同時見学する人数が 1 人の場合で RF、同時見学する人数が 2 人の場合で  $k$ NN を用いた。上記以外の実験条件を 4.1 節と同じにした。 $T$  を変更したときの人物位置の推定誤差の平均と標準偏差とを図 9(a) と (b) に示す。図中の誤差バーは標準偏差を表す。 $T$  が大きくなるにつれて推定誤差が減少していき、同時見学する人数が 1 人の場合の (a) では  $T = 60$  から推定誤差がほぼ減少しなくなった。この傾向は、(b) の 2 人の場合で

<sup>\*1</sup> <https://corp.zozo.com/news/20200611-10522/>

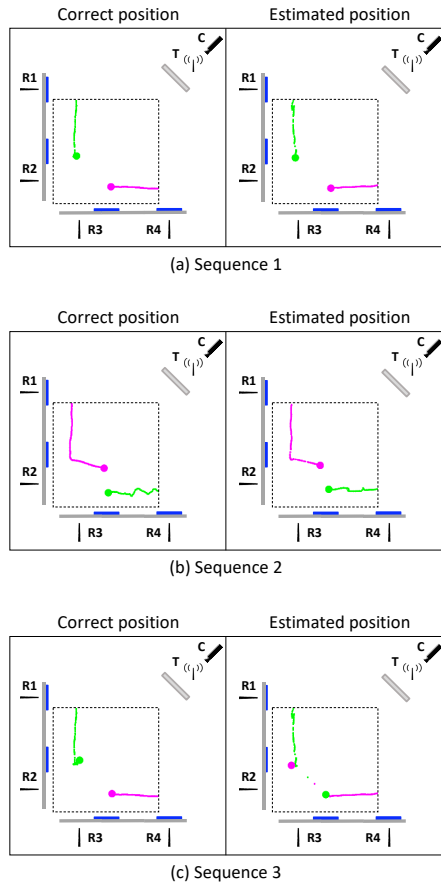


Fig. 8 Visualization of the positions estimated by our method and the correct positions.

$T = 90$  から見られた。本論文では、推定誤差が安定して小さくなり、かつ、 $T$  になるべく小さい方が好ましいと考え  $T = 90$  を採用した。この場合、3 秒間の短時系列信号を回帰モデルで用いるため、人物位置の推定結果に、時間方向のずれが含まれる問題に気を付けなければならない。本論文では、ポスター会場を想定しており、ポスターを眺めながら歩く人物の移動速度は遅く、また、ポスターを読むために立ち止まっている時間が長いと仮定できるため、 $T = 90$  でも時間方向のずれは許容できると考えている。ただし、人物が走る場合や早歩きする場合に提案手法は対応できないため、今後のさらなる手法改良が必要である。

次に、Wi-Fi 受信機の個数であるパラメータ  $N$  について評価した。ここでは  $N = 4, 3, 2, 1$  とした。  $N = 4$  の時、図 3 で示した受信機 R1 から R4 の 4 台で試行した。  $N = 3$  の時、1 台を除いた 3 台の組み合わせを全て試行しており、  $N = 2$  や  $N = 1$  の時も同様に除いて試行した。回帰モデルとして、同時見学する人数が 1 名の場合で RF、同時見学する人数が 2 名の場合で  $kNN$  を用いた。上記以外の実験条件を 4.1 節と同じにした。  $N$  を変更したときの人物位置の推定誤差の平均と標準偏差とを図 9(c) と (d) に示す。  $N = 4$  で推定誤差が最も小さく、  $N = 3$  から  $N = 1$  になるにつれ推定誤差が大きく増加した。この傾向は、(c) の 1 名の場合と (d) の 2 名の場合とで同様に見られた。提案手法は、Wi-Fi 受信機の個数が多いほど推定誤差が低下すると言えるが、その設置コストの課題があるため、今後の手法改良が必要である。

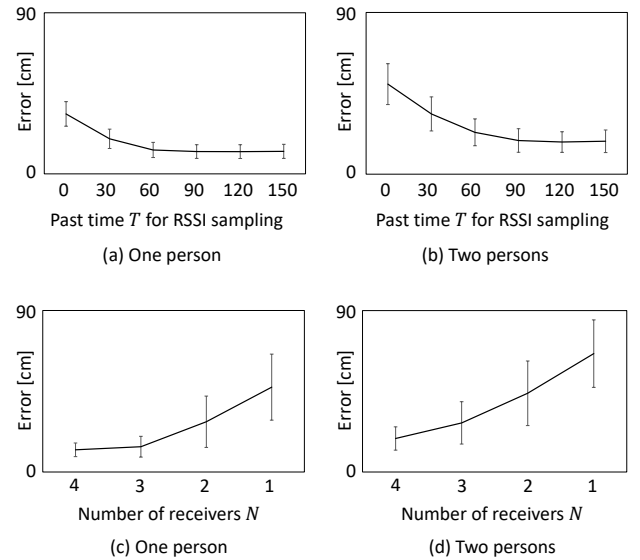


Fig. 9 Estimation errors when changing the parameters  $T$  and  $N$ .

#### 4.6 遮蔽物についての考察

本論文では、一時的な遮蔽物として、ポスター会場で使用される一般的なパネルを使用した。その大きさは 2.2 節で述べたとおり、高さが 180 cm で幅が 90 cm で厚みが 4 cm であった。A0 サイズで印刷した紙のポスターを押しピンで張り付けることができるものとした。使用したパネルに型番が記載されていないため、中心材の詳細は分からないが、合板やロールコアやハニカムコアなどの一般的な素材であると想定される。このようなパネルが遮蔽物であれば、2.4 節で示したように、RSSI の値は極端に衰弱しないため、無線機の電波から人物位置を推定できると言える。また、プロジェクタスクリーンなどの布地の遮蔽物も、無線機の電波取得で問題ないことを確認している。ただし、水槽などの水分を含むものやアルミなどの金属が遮蔽物であれば、RSSI の値は著しく減衰するため、無線機の電波から人物位置を推定できない限界がある。以上より、一時的な遮蔽が生じたとしても Wi-Fi の電波を送受信できる環境であれば、提案手法を適用できると考える。

#### 4.7 設置条件についての考察

本論文では、Wi-Fi の受信機と送信機、および、ポスターのパネルについて、それらの設置条件を図 3 とした。この設置条件を決めるため、事前に予備実験を行うことで、受信機と送信機との間における RSSI の特性を調査した。予備実験の結果より、送信機から受信機までを結ぶ直線上に人物が立ち入ると、RSSI の値が大きく減衰することを確認した。また、受信機と送信機の高さが人物の身長より高くなると、RSSI の値に変化が表れ難くなることを確認した。このため、ポスターが貼られたパネルの正面手前 90 cm で、ポスターを読みながら立ち止まる人物の腰元を主に捉えることを狙い、1 台の送信機と 4 台の受信機とを設置した。本論文で用いた実験環境において、RSSI の値は、電波の直進性から主な影響を受けており、電波の回折性や屈折性からそれほど影響を受けていなかったと考えられる。このことより、RSSI を用いて人物位置を高精度に推定するためには、多数の受信機を適切な位置に設置しなければならない限界があると言える。今後は Wi-Fi の CSI やミリ波レーダーなど、より豊富な情報を少数台で取得できる無線機を利用



した手法改良が必要である。

## 5. おわりに

本論文では、一時的な遮蔽が生じる空間でも人物位置を推定するため、カメラと無線機とを併用することで、遮蔽がない場合には RSSI 時系列信号と人物位置との訓練サンプルを収集し、遮蔽がある場合は RSSI 時系列信号のみから回帰モデルを用いて人物位置を推定する手法について述べた。評価実験のため、空間内にカメラと無線機とを設置し、遮蔽がない場合と遮蔽がある場合について、1名もしくは2名が同時に見学するシーケンスで構成される PPA データセットを構築した。遮蔽がない場合、カメラから取得された映像を用いて人物位置を獲得し、RSSI 時系列信号と併せて訓練サンプルを収集した。遮蔽がある場合、訓練サンプルを用いて学習された回帰モデルを用いることで、RSSI 時系列信号のみから人物位置を推定した。本論文では、カメラ視野の端で人物を捉えることができるが、その視野の中央付近に遮蔽物が設置される前提条件を設けた。この前提条件の下で一時的な遮蔽が発生した場合、人物位置の推定誤差は、同時に見学する人数が1名の場合に  $15.9 \pm 5.2$  cm、同時に見学する人数が2名の場合に  $18.7 \pm 6.6$  cm であることを実験で確認した。

今後の課題として、遮蔽された空間内で同時に見学する人数に依存しない人物位置の推定手法の開発が挙げられる。また、人数が増加した場合の評価や、空間内での歩き方が変化した場合の評価などが必要である。さらには、各時刻の人物位置だけでなく、時間方向に追跡することで移動軌跡を推定することが挙げられる。

## 謝辞

本研究を進めるにあたり、貴重なご助言やご意見をくださいました鳥取大学工学部の岩井儀雄先生と井上路子先生に感謝の意を表す。

## 参考文献

- 1) K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017.
- 2) Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff,

- and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of 15th European Conference on Computer Vision, Part VII*, pp. 833–851, 2018.
- 3) C. Wang, A. Bochkovskiy, and H. Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464–7475, 2023.
- 4) S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 28, pp. 91–99, 2015.
- 5) Moustafa Youssef, Matthew Mah, and Ashok Agrawala. Challenges: device-free passive localization for wireless environments. In *Proceedings of the 13th annual ACM International Conference on Mobile Computing and Networking*, pp. 222–229, 2007.
- 6) Apidet Booranawong, Nattha Jindapetch, and Hiroshi Saito. Adaptive filtering methods for rssi signals in a device-free human detection and tracking system. *IEEE Systems Journal*, Vol. 13, No. 3, pp. 2998–3009, 2019.
- 7) 倉聖美, 白石陽, 山口弘純ほか. Wi-Fi 受信電波強度の変化を利用した通路の歩行者数推定手法. 情報処理学会論文誌, Vol. 59, No. 2, pp. 496–507, 2018.
- 8) Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics*, Vol. 34, No. 6, pp. 1–13, 2015.
- 9) Zengshan Tian, Jiacheng Wang, Xiaolong Yang, and Mu Zhou. Wicatch: A Wi-Fi based hand gesture recognition system. *IEEE Access*, Vol. 6, pp. 16911–16923, 2018.
- 10) Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. mmMesh: towards 3D real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 269–282, 2021.
- 11) Fernando De la Torre Jiaqi Geng, Dong Huang. Densepose from WiFi. In *CoRR/2301.00250*, pp. 1–13, 2022.
- 12) Yichao Wang, Yili Ren, Yingying Chen, and Jie Yang. Wi-Mesh: A WiFi vision-based approach for 3D human mesh construction. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pp. 362–376, 2023.
- 13) Sebastian Sadowski, Petros Spachos, and Konstantinos N. Plataniotis. Memoryless techniques and wireless technologies for indoor localization with the internet of things. *IEEE Internet of Things Journal*, Vol. 7, No. 11, pp. 10996–11005, 2020.
- 14) 我那覇航, 尾崎匠, 井上路子, 西山正志. カメラで撮影された歩行中の人物グループから生成されたインタラクティブ動画を聞いた会話の活発さ認識. 精密工学会誌, Vol. 90, No. 2, pp. 233–244, 2024.
- 15) Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.