

複数 Wi-Fi 機器の CSI を用いた行動推定手法の評価

寺本 京祐¹ 斎藤 隆仁² 池田 大造³ 太田 賢² 西村 雅史¹ 水野 忠則⁴
峰野 博史¹

概要: 近年のスマートホーム化や IoT などの発展によって, 日常行動のセンシングが注目されている. 行動推定手法として人が映らないことによるプライバシー保護やコスト削減が図れる Wi-Fi CSI ベースの手法が研究されている. 先行研究においては, 屋内における簡単な人の動きを CSI で推定可能であることが報告されている. しかし, 先行研究では一度に推定可能な行動数は限られており, 収集可能な行動範囲についても非常に狭いという課題がある. そこで, 本研究では 2 台の CSI 受信機から日常生活で想定される行動を推定することを目標とする. 提案手法では, 2 台の CSI 受信機を用いて CSI を収集し, ノイズ除去を実施する. また, 推定精度を上げるため, ノイズ除去による信号の歪みが抑えられるスペクトログラムによる画像化を行う. 使用する学習モデルは画像認識分野で使用されている CNN と近年画像認識分野で注目されている ViT を用いて性能比較を行った. 結果, 人の有無を推定する基礎実験において, CNN モデルの交差検証の精度が 92% となった.

Evaluation of a Behavior Estimation Method Using CSI for Multiple Wi-Fi Devices

KYOSUKE TERAMOTO¹ TAKATO SAITO² DAIZO IKEDA³ KEN OHTA²
MASAFUMI NISHIMURA¹ TADANORI MIZUNO⁴ HIROSHI MINENO¹

1. はじめに

近年の AI 技術の発展などにより, 人間の行動を推定するセンシング技術に注目が集まっている. 従来の人間の行動推定手法として, カメラを用いたものや, ウェアラブルセンサを用いたものがある. しかし, カメラを用いる方法では, 人が映っている画像情報を使用することになることからプライバシー侵害の恐れがあり, 人を映せない暗闇では使用できないという問題点がある. また, ウェアラブルセンサを用いた方法については, 被験者の身体にセンサを取り付ける必要があり, 実用的なシステムとして好ましくない. 一方, スマートフォンの普及や IoT 機器の増加により, Wi-Fi の需要が増加している. 近年では, プライバシーや装着の必要性などの問題を解決でき, 一般家庭にも広く普及している Wi-Fi を用いたセンシング技術が注目されている. Wi-Fi 無線を用いたセンシング技術では, Wi-Fi チャンネル状態情報 (Channel State Information:CSI) を用いた手法がある.

CSI は Wi-Fi 通信における物理層で用いられる, 送受信機間の電波路の状態を表す情報である [1]. CSI は送信された電波の反射・回折などのマルチパスの影響による振幅と位相の変化を複素数の絶対値と偏角で表される. Wi-Fi の

規格である 802.11a/g/n では, 直交周波数分解多重 (Orthogonal Frequency Division Multiplexing:OFDM) により異なる周波数を直交させる. CSI では OFDM により一度に複数の情報伝達が可能となり, 異なる直交する周波数を持つサブキャリアごとに CSI が収集可能となる. 周波数が異なると電波の減衰の仕方や反射の仕方が異なり, 各送受信アンテナ間における伝搬路によって電波が送られる. そのため, 各受信アンテナ間のサブキャリアごとに異なった振幅変化と位相変化が起こる. X_i と Y_i をそれぞれ i 番目のサブキャリアで得られる M_T 次元の送信ベクトルと M_R の受信ベクトルとすると式 1 の関係が得られる.

$$Y_i = H_i X_i + N_i \quad (1)$$

ここで, N_i は M_R 次元のノイズベクトルであり, H_i は i 番目のサブキャリアで得られる, $M_T \times M_R$ 次元の行列であり, 式 2 のような多次元行列として CSI を得る.

$$H_i \begin{pmatrix} h_{11} & \cdots & h_{1M_T} \\ \vdots & \ddots & \vdots \\ M_T 1 & \cdots & M_T M_T \end{pmatrix} \quad (2)$$

CSI ではサブキャリアごとのデータを多次元行列の形で保持できることから, カメラやセンサ単体では認識が難しい物理的な環境変化をより詳細に分析しやすくなる. また, 振幅や位相の変化は明るさに関係なく発生するため, カメ

1 静岡大学情報学部
Faculty of Informatics, Shizuoka University
2 株式会社 NTT ドコモ
NTT DOCOMO, INC.

3 日本電信電話株式会社
NIPPON TELEGRAPH AND TELEPHONE CORPORATION
4 愛知工業大学情報科学部
Faculty of Information Science, Aichi Institute of Technology

ラのような暗闇では人の動きを認識できないといった問題は発生しない。ただし、Wi-Fi 信号を用いた行動推定では、人が受信機のある部屋とは別の部屋にいる場合や、人と受信機間の距離が長い場合において、人の行動によって引き起こされる信号の変動が小さくなる。CSI の信号対雑音比 (Signal-to-Noise Ratio:SNR) が小さくなってしまふことでノイズと行動による変位の区別がつきにくくなり、結果として推定結果が悪くなるという課題も生じる[2]。

本研究では CSI ベースの行動推定手法として、想定する環境内に 2 台の CSI 受信機を設置し、それぞれの CSI データをスペクトログラムの画像データに変換した後、深層学習による推定を行う。これにより、高精度に人の動きを推定する手法を提案する。

2. 関連研究

2.1 RSSI を用いた手法

受信電波強度 (Received Signal Strength Indicator:RSSI) は、Wi-Fi 通信におけるデータリンク層で計測される電波強度を示す値のことである。RSSI は複数の経路で受信される電波を組み合わせるため、出力は単一の振幅情報となる。

RSSI を用いた研究は、2012 年に発表された B. Marazovac らによる人間の存在検知手法がある[3]。無線通信を常時実施する部屋を用意して部屋の中に人が入ることによって受信信号強度が変化することを検知し、システムに連携している家電製品の制御を実施している。

RSSI を用いた行動推定では、カメラや加速度センサでの課題を解決することができるが、RSSI の出力は受信信号の強度のみである。よって、電波の強度が変化する複数の要因があった場合に正確に推定することが難しいという課題がある[4]。

2.2 CSI を用いた手法

CSI を用いた研究は、2013 年に発表された J. Xiao らによる Pilot[5]がある。ある地点での CSI を Fingerprints としてあらかじめ記録しておくことで、相関の高い地点を実際の位置と推定するものである。日常生活における行動推定まではされていないものの、CSI を用いて人の行動推定を行える可能性を示している。

また、2015 年に発表された CARM[6]では、CSI の動態と人の移動速度の相関関係を定量化した CSI 速度モデルならびに、異なる人体部位の移動速度と人間活動の相関関係を定量化した CSI 活動モデルの 2 つを用いた機械学習によって、歩く、座るといった日常行動推定を約 96%の精度で実現している。

K. Ali らによる WiKey[7]では、フィルタ処理をした CSI によってキーボードで入力した文字を高精度に推定している。キーボードへ入力する動作は非常に小さいものである

が、微小な動きであっても受信機や送信機の位置関係及びフィルタ処理などの前処理を工夫することで、CSI から高精度に行動が推定可能であることが示された。

その後、時系列を考慮した学習として LSTM を用いた CSI ベースの行動推定手法が発表されている。S. Yousefi らは、歩く、走るといった基本的な行動に対し、LSTM による行動推定が最も高精度であることを示した[8]。

H. Lee らは、歩く、トイレに入る、寝る、食事をするといった一般住宅における基本的な日常行動に対し、複数の受信機を設置して CSI を取得し、CNN で特徴量を抽出、LSTM で学習することで高精度な行動推定を実現している[9]。CNN で特徴量を入力する際に、スペクトログラムの画像に変換している。H. Lee らの研究から CSI で推定を行う際にスペクトログラム画像を有効活用することで高精度な識別が可能であることが示された。

田中らの手法では、立つ、座るなど 5 つの基本的な生活行動に対し、CSI と動画を timestamp で同期させて収集し、動画による自動正解ラベルによって CSI を LSTM で学習させ、約 75%の精度で行動推定を可能にした[10]。これまででは、CSI の行動推定モデルを作成するための学習データは手動で付けるものが多かったため、田中らの手法で、学習データの自動生成の可能性が示された。

2.3 画像による推定手法

Takashi Nakamura らの手法[11]では、CSI データをそのまま使用するのではなく、CSI からスペクトログラムの画像を生成し、CNN (ResNet34) を用いて人の転倒検知を実施した。スペクトログラムの画像にすることで CSI での推定における課題の一つであったノイズ除去時の信号の歪みによる動きの捉えにくさを抑えられることが示された。

画像認識の分野において近年注目されている学習モデルとして Vision Transformer (ViT) がある[12]。ViT は、画像認識分野で一般的に使用される CNN を一切使わずに、これまで自然言語処理で主に使用されてきた Transformer を使った学習モデルである。ViT では、従来の CNN 型モデルを超える性能を出す事例もある[13]。

2.4 関連研究まとめ

既存研究では、加速度センサや CSI などのセンシング技術から人の歩く、座るといった基本的な行動を、フィルタ処理やスペクトログラムによる画像化などを用いて高精度に推定できることを示した。そこで本研究では、フィルタ処理によるノイズ除去を実施しても動きの正確さを捉えられやすいスペクトログラムによる画像化に注目をして、画像認識分野で広く使用されている CNN を用いて人の行動推定を実施した。また、近年画像認識分野で性能を上げている ViT についても行動推定モデルを作成し、CNN との性能比較を行う。

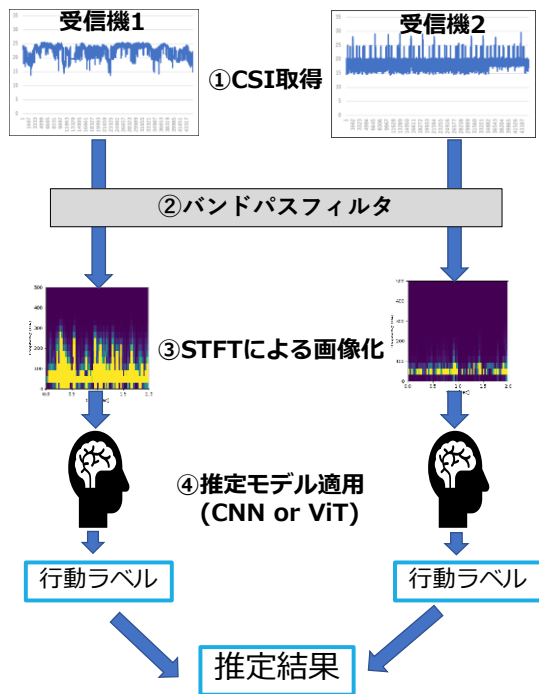
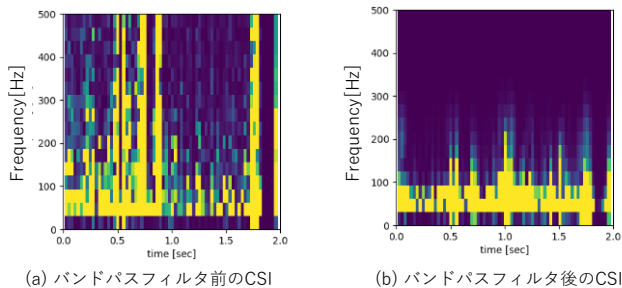


図 1 提案手法の推定フェーズの概要



(a) バンドパスフィルタ前のCSI (b) バンドパスフィルタ後のCSI

図 2 フィルタ適用前の CSI とフィルタ適用後の CSI

3. 提案手法

3.1 概要

提案手法は、学習フェーズと推定フェーズで構成される。図 1 に提案手法における推定フェーズの流れを示す。Wi-Fi 送信機（アクセスポイント）1 台と、異なる場所に配置された受信機 2 台を利用する。各受信機は、アクセスポイントから送信された Wi-Fi 信号を同時に収集する。CSI は位相データと振幅データの両方が同時に収集可能であるが、本研究では位相データのみを用いて推定を行うこととする。学習フェーズでは、収集した CSI へバンドパスフィルタ処理によるノイズ除去を行い、その後ノイズ除去後の CSI データをスペクトログラムに変換する。生成したスペクトログラムの画像データから受信機ごとに CNN あるいは ViT で行動推定モデルの作成を行う。正解ラベルのアノテーション

は、[10]で実現される動画による自動正解ラベルを想定するが、本稿では議論を容易にするため手動で正確に実施する。推定フェーズでは、学習フェーズ同様に 2 つの受信機から同時に CSI を収集し、バンドパスフィルタを用いてノイズ除去処理を行った後、スペクトログラムに変換する。スペクトログラム化した画像データを基にして作成した受信機ごとの学習モデルで行動推定を実施する。最後に受信機ごとに出力される行動ラベルに対し、学習時 F 値を用いて比較することで最終的な推定結果を決定する。

3.2 CSI の前処理

CSI は $M_T \times M_R \times N$ 次元の複素数の時系列データとして得られる。各時刻の各次元において CSI の位相値を計算し、 $M_T \times M_R \times N$ 次元の位相の時系列データを求める。

まず、得られた時系列データから、バンドパスフィルタを用いてノイズの除去を行う。先行研究[14]によると多くの人間活動の周波数帯域は 300Hz 以下であるとしており、低周波と高周波の両成分を除去するために通過帯域を 5Hz-100 Hz と設定している。本研究では高周波域である 300Hz 未満を通過帯域として設定し、ノイズ除去を実施した。

フィルタ処理を行った後、サブキャリアごとに短時間フーリエ変換（short-term Fourier transform:STFT）を行った。2000ms のウィンドウサイズで STFT を行い、スペクトルを算出した。生成したスペクトログラム画像は、 128×128 の画像サイズにリサイズして学習モデルに入力する。図 2 にバンドパスフィルタ適用前の CSI のスペクトログラムと、バンドパスフィルタ適用後の CSI のスペクトログラムを示す。フィルタ適用前にあった高周波域のデータがフィルタ適用後には除去ができていことが分かる。

3.3 行動推定モデル作成

行動推定のモデルとして、本手法では 2 層畳み込みの CNN を用いたものと ViT の 2 つのモデルを使用して、双方の性能の比較を行う。学習時の Epoch 数は CNN と ViT で共通で、30Epoch で行うこととした。

図 3 に CNN の学習モデル概要について示す。 128×128 のスペクトログラム画像を入力データとして、2 層の畳み込み層とプーリング層を通り、全結合層と過学習を防ぐためのドロップアウト層を通った後、softmax 関数でクラス分類を行う。畳み込み層及び全結合層における活性化関数には relu 関数を使用している。また、ドロップアウト層のドロップ率は 0.5 としている。

ViT については、[11]のモデルを使用している。ViT では自然言語処理における単語のように画像データをパッチとして入力することで画像認識を可能にしている。本研究ではバッチサイズは 16×16 に設定し学習モデルの作成を実施した。

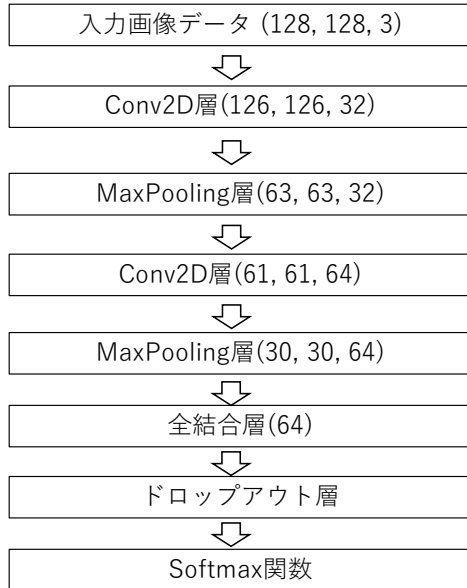


図 3 CNN のモデル概要

3.4 行動推定結果

推定フェーズにおいて、作成した行動推定モデルを適用し、受信機ごとに行動ラベルを取得する。出力された 2 つの行動ラベルに対し、以下のルールに基づいて推定結果を決定する。

- **2 つの行動ラベルが同一の場合**
推定された行動ラベルをそのまま推定結果とする。
- **片方が無人のラベル、片方が行動有のラベルの場合**
行動有のラベルを推定結果とする。無人のラベルを考慮しない理由としては、人が行動していたとしても片方の受信機に変化が検出できない場合があるためである。例えば受信機 1 と受信機 2 が異なる部屋に配置されている時に、受信機 1 の設置されている部屋に人が入っていても受信機 2 の部屋では人がいないため無人となると考えられる。よって、片方が無人であっても実際には行動している場合があるため、行動有のラベルを結果として出力する。
- **共に行動有のラベルでラベルが異なる場合**
学習時 F 値を基に推定結果を決定する。例えば、受信機 1 で学習時 F 値 0.7 の入室ラベルとなり、受信機 2 では学習時 F 値 0.8 の退室ラベルとなった場合、学習時 F 値が高かった受信機 2 の退室ラベルを最終的な推定結果とする。ただし、無人状態のラベルのみ例外として、受信機 1 と受信機 2 がともに無人状態となった場合のみ無人状態を推定結果とする。

4. 基礎実験 1

4.1 実験概要

基礎実験 1 では、CSI のスペクトログラム画像から無人と有人の検知が可能であるかを調査した。データ収集においては、将来の一般住宅での防犯や独居高齢者の生活支援システムへの応用を想定し、浜松市環境共生住宅実験施設（住まいのオープンラボ）にてデータを取得した。住宅環境内では他の Wi-Fi 通信機器による通信はなく、実験環境内において他の人物がいない状況を想定する。図 4 に基礎実験 1 の実験環境を示す。受信機を 2 台設置しているが、基礎実験 1 では台所での行動は行わないことから受信機 2 のデータは基礎実験 1 では使用しないこととする。ただし、今後の実験での条件を合わせる目的として受信機 1 と同じ動作をするように設定している。受信機 1 は居間に、受信機 2 は台所に設置し、アクセスポイントはほぼ中央の食堂に設置した。また、学習時の正解ラベルを正確に付けるために、2 台のカメラを行動の確認用に設置した。アクセスポイントとして Buffalo WSR-2533DHPL を使用し、受信機は Intel5300 無線 LAN モジュールを挿入したノート PC を使用した。CSI は、Linux CSI802.11n ツール[15]を用いて、受信機で収集する。具体的には、受信機から Wi-Fi で ping を 1ms 間隔でアクセスポイントへ送信することで CSI を収集した。

基礎実験 1 における収集行動を表 1 に示す。実験環境内において食堂のドアから入室し、居間にある椅子に座り、立った後に食堂のドアから出るという動作を 15 回実施してデータを収集した。

基礎実験 1 では無人と有人であるかの識別のみを目的としているため、入室、座る、立つ、退室の 4 つのラベルを有人ラベルとしてまとめて学習を行うこととした。収集したデータを 2000ms 間隔でスペクトログラム化した際の合計の数は、無人と有人それぞれで 1637 個となった。学習時には 8 割をトレーニング用、2 割を交差検証用として分割して学習を行った。

4.2 実験結果・考察

図 5 (a), (b) に基礎実験 1 における交差検証用における CNN と ViT での混同行列を示す。トレーニングデータの Accuracy は CNN が 0.989, ViT が 0.499, 交差検証での Accuracy は CNN が 0.921, ViT が 0.500 となった。

CNN では、トレーニングデータでの精度がおよそ 99% であり、学習に関しては問題なく進んでいることが分かる。交差検証用データについては、約 92% となっておりトレーニングデータと比較するとやや下がっている結果となっているが、混同行列を見ると無人データの正答率は 99% 以上であり、有人データについても 90% 以上で正しく推定できていることから CNN での無人と有人の推定は可能であると考えられる。

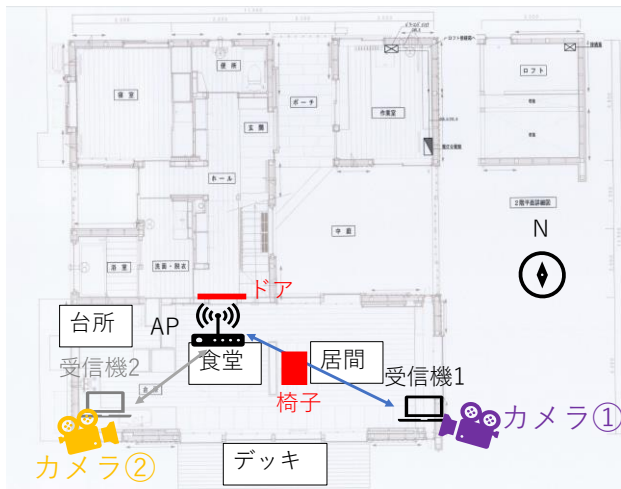


図 4 基礎実験 1 の実験環境

表 1 基礎実験 1 の収集行動

行動名	内容
無人	どの部屋にも人が居ない状態
入室	食堂のドアから入る
座る	居間の椅子に座る
立つ	居間の椅子から立つ
退室	椅子から立った後、食堂のドアから出る

ViT では、トレーニングデータの精度がおおよそ 50%であり、本実験では 2 クラス分類であることから学習が進んでいないことが考えられる。また、混同行列を見ると無人データと有人データを全て有人データとして推定してしまっている。ViT において学習が上手くいかなかった理由の一つとして、学習データ量の不足が考えられる。[13]では、ViT が従来の CNN 型モデルと比較して同等かそれ以上の結果を出す事例があるが、事例では ViT に大量のデータで事前学習を行っている。使用しているデータセットも ImageNet や CIFAR-100 のような大量のデータがあるものであるため、本研究で使用した 3274 個の画像のみでは ViT がある[16]。スペクトログラムにおいてはどの時間のどの周波数が変化したのが重要であるため、正確な位置情報の性能を出すにはデータ不足であった可能性があると考えられる。また、ViT はオブジェクト検出やセグメンテーションといったタスクにおいては 16×16 などの小さいパッチから正確な位置情報の特定ができず検出が難しいという課題を特定しにくいというのが ViT での学習が進まなかった理由の一つなのではないかと考える。

今後 ViT で高精度な推定をスペクトログラムで行うことを考える場合、事前に RGB 画像などを大量に準備して ViT に事前学習をさせることが必要であると考えられる。本研究で

	無人	有人
無人	99.5%	0.5%
有人	9.6%	90.4%

(a)CNNでの混同行列

	無人	有人
無人	0.0%	100.0%
有人	0.0%	100.0%

(b)ViTでの混同行列

図 5 基礎実験 1 での実験結果

はパッチサイズを 16×16 に設定していたが、事前学習をした ViT に適用するパッチサイズを 32×32 などより大きなサイズにすることでスペクトログラムから位置情報を推定しやすくなる可能性もあると考える。また、ViT におけるオブジェクト検出やセグメンテーションにおける課題点を改善した Swin Transformer[16]などを適用することも改善策として挙げられる。

基礎実験 1 では、無人と有人の 2 クラス分類を行い、CNN については無人と有人の推定が CSI のスペクトログラムから可能であるということが分かった。よって、基礎実験 2 においては、人間の行動をより細かく推定することを目標として基礎実験 1 で高精度であった CNN を用いて基礎実験 1 で用いた環境と同様の一般住宅において想定される行動について推定する学習モデルを作成、評価を行う。

5. 基礎実験 2

5.1 実験概要

基礎実験 2 では、基礎実験 1 をベースとして居間での行動に焦点を当てて実験を実施した。図 6 に実験環境を示す。ここで、基礎実験 2 では、基礎実験 1 では使用していなかった台所に設置している受信機 2 も推定モデルを作成する。その他の計測条件は、基礎実験 1 と同様である。基礎実験 2 では無人状態を含む全 10 種類の行動について分類を行う。表 2 に基礎実験 2 で実施した行動一覧を示す。椅子で静止、食事、床で寝転がるについては、居間に入室してから椅子に座った後、約 1 分間食事等の各行動を実施してから居間を出る、という一連の行動を 15 回ずつ実施した。デッキの出入りについては、居間からデッキに出入りする行動を 1 分おきに計 15 回実施した。基礎実験 2 で収集したデータは受信機 1、2 それぞれで 25248 個となった。なお、基礎実験 2 では、受信機 1 と受信機 2 のそれぞれで CNN モデルを作成し、評価を行う。

5.2 実験結果・考察

図 7 (a), (b) に基礎実験 2 における交差検証用データの CNN モデルでの混同行列を示す。トレーニングデータの Accuracy は受信機 1 が、受信機 2 が 0.250、交差検証の

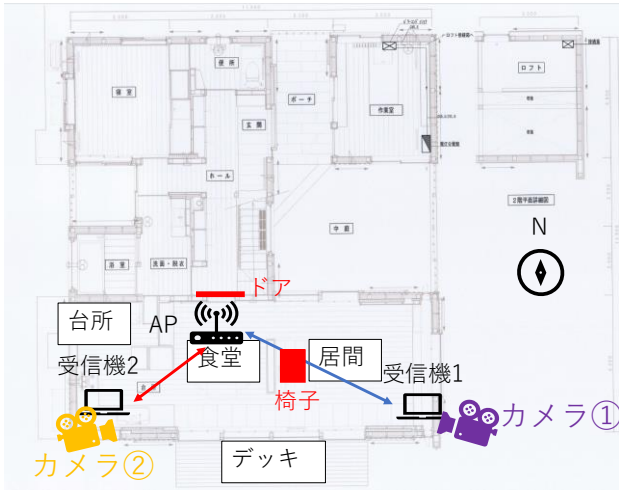


図 6 基礎実験 2 の実験環境

表 2 基礎実験 2 の収集行動

行動名	内容
無人	どの部屋にも人が居ない状態
入室	食堂のドアから入る
座る	居間の椅子に座る
立つ	居間の椅子から立つ
退室	食堂のドアから出る
椅子で静止	居間の椅子に座ったまま静止
食事をする	居間に椅子に座り、食事を取る
パソコンをする	居間の椅子に座り、パソコンを触る
床で転がる	居間の床で寝転がる
デスク	居間からデスクに出入り

	1.無人	2.入室	3.座る	4.立つ	5.退室	6.椅子で静止	7.食事	8.パソコン	9.床で転がる	10.デスク
1.無人	93.9%	0.0%	0.0%	0.0%	0.0%	0.0%	1.2%	0.0%	3.1%	1.8%
2.入室	1.1%	79.2%	0.0%	0.0%	6.4%	0.0%	5.6%	1.7%	4.2%	1.9%
3.座る	3.0%	64.0%	0.0%	0.0%	5.5%	0.0%	7.5%	0.0%	5.0%	15.0%
4.立つ	12.6%	66.5%	0.0%	0.0%	3.3%	0.0%	9.9%	0.0%	3.3%	4.4%
5.退室	11.1%	54.9%	0.0%	0.0%	10.1%	0.0%	7.4%	0.0%	6.4%	10.1%
6.椅子で静止	58.2%	14.5%	0.0%	0.0%	0.3%	0.0%	8.5%	0.5%	15.8%	2.2%
7.食事	36.5%	24.6%	0.0%	0.0%	0.7%	0.0%	15.4%	1.9%	16.4%	4.5%
8.パソコン	27.2%	31.0%	0.0%	0.0%	0.0%	0.0%	14.6%	2.8%	23.7%	0.6%
9.床で転がる	44.7%	20.0%	0.0%	0.0%	0.0%	0.0%	13.0%	2.6%	15.0%	4.6%
10.デスク	33.4%	26.4%	0.0%	0.0%	3.1%	0.0%	7.3%	0.5%	7.5%	21.8%

(a) 受信機1の混同行列

	1.無人	2.入室	3.座る	4.立つ	5.退室	6.椅子で静止	7.食事	8.パソコン	9.床で転がる	10.デスク
1.無人	43.7%	15.6%	0.0%	0.0%	4.6%	0.1%	0.0%	10.6%	10.6%	14.8%
2.入室	11.2%	79.5%	0.0%	0.0%	1.4%	0.0%	0.0%	0.3%	0.3%	7.3%
3.座る	25.9%	40.8%	0.0%	0.0%	9.7%	1.0%	0.0%	2.9%	2.9%	16.8%
4.立つ	18.9%	37.5%	0.0%	0.0%	15.3%	0.4%	0.0%	4.0%	4.0%	20.0%
5.退室	13.9%	75.3%	0.0%	0.0%	0.9%	0.5%	0.0%	1.4%	1.4%	6.6%
6.椅子で静止	19.8%	7.1%	0.0%	0.0%	28.0%	0.2%	0.0%	6.3%	6.3%	32.5%
7.食事	26.0%	7.3%	0.0%	0.0%	15.4%	0.9%	0.0%	15.9%	15.9%	18.5%
8.パソコン	19.4%	3.3%	0.0%	0.0%	14.5%	1.1%	0.2%	22.5%	22.5%	16.5%
9.床で転がる	39.9%	5.8%	0.0%	0.0%	4.7%	1.6%	0.0%	16.7%	16.7%	14.6%
10.デスク	33.3%	23.7%	0.0%	0.0%	2.5%	0.0%	0.0%	4.0%	4.0%	32.4%

(b) 受信機2の混同行列

図 7 基礎実験 2 の実験結果

Accuracy は受信機 1 が、受信機 2 が 0.276 となった。受信機 1、受信機 2 ともに Accuracy が 0.25~0.27 とかなり低い結果となっている。受信機 1 の混同行列を見ると、無人ラベルに椅子で静止、食事などの行動ラベル、入室ラベルに入室、座る、立つ、退室のラベルが特に偏っている傾向にあるとわかる。受信機 2 の混同行列についても、偏り方の度合いに違いがあるが、おおよそ同じ傾向がみられる。10クラス分類において上手くいかなかった要因は、大きく分けて2つあると考える。一つ目は、スペクトログラム画像自体が推定しづらいものであることだと考える。図 8(a)~(d) に1.無人、7.食事中、2.入室、5.退室のスペクトログラム画像を示す。画像を見ると、2.入室、5.退室、7.食事中においては、全体的に傾向が似ていることがわかる。特に 0Hz から 100Hz にかけての低周波域については、1.無人状態でも

変化が起きており、低周波域のみを見て行動の区別をつけるのは難しいと考えられる。高周波域については人の行動によって変化が起きると考えられるが、高周波域が変動するタイミングというのは人の動き方、動くタイミングによって大きく異なることが考えられる。本研究では、スペクトログラムは CSI データを 2 秒間隔、オーバーラップはなしで生成している。このときに、画像の 0 秒地点が行動の開始とは限らないという問題点がある。例えば、入室の動作が 8 秒間あった場合、合計で 4 個のスペクトログラム画像が生成される。よって、同じ行動ラベルのスペクトログラムでも変化が大きいタイミングと変化が小さいタイミングが存在する。よって、変化が大きいタイミングでは入室、変化が小さいタイミングでは無人と誤推定が起きているのではないかと考える。また、椅子で静止、食事、パソコン

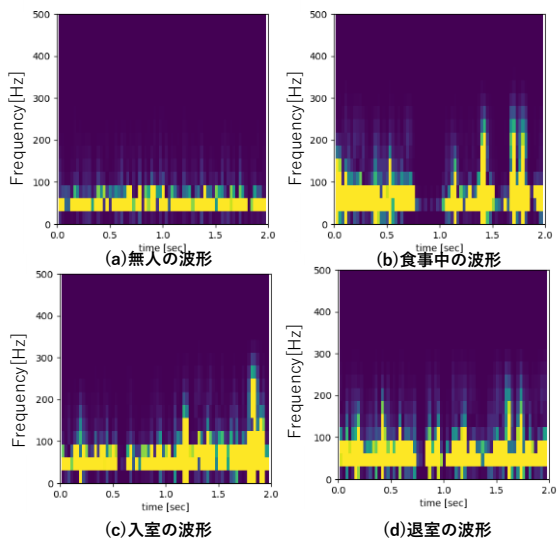


図 8 基礎実験 2 でのスペクトログラム画像

などの座ったままの行動では無人状態が推定される割合が入室，座る，立つ，退室と比べて多い．よって，スペクトログラムのような変化量を示す画像では人があまり動かない行動についての推定は難しいと考える．

2つ目の要因としては，CNN モデルの簡潔さにあると考える．本研究では 2 層畳み込みの CNN モデルを用いて行動推定モデルを作成したが，先行研究[11]において使用していたモデルは，34 層の畳み込み層がある ResNet34 を使用していた．CNN の層を増やすと計算量が増えるというデメリットもあるが，より高精度を目指すことを考えるとさらに層を増やしたモデルを適用することも改善策として良いと考える．

6. おわりに

本研究では，CSI からスペクトログラムの画像を生成し，CNN や ViT に入力としていれることで，有人，無人の判定や日常生活において想定される行動を対象として高精度で推定可能かについての基礎検討を実施した．有人と無人を推定する基礎実験 1 では，CNN において Validation データが 92% の精度で推定可能であることを確認した．日常生活で想定される行動を推定する基礎実験 2 においては，精度が 0.25~0.27 と改善の必要がある結果となった．

今後は，日常行動における推定精度向上のために，CNN 等のモデルの改善や適切なデータセットの選択，前処理を行っていくとともに，CSI での学習における課題である正解ラベル付けに関して半教師あり学習の適用を検討を進めていく．

謝辞

本研究の一部は，JST 創発的研究支援事業（JPMJFR201B）の支援を受けたものである．また，実験環境の支援をいただいた浜松市住まいの研究会と浜松市環境部環境政策課の方々に感謝の意を表する．

参考文献

- [1] IEEE Computer Society: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification (2012).
- [2] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, Shengjie Li, :RT-fall: a real-time and contact less fall detection system with commodity WiFi devices, IEEETrans. Mob.Computing, vol.16, pp.511-526 (2017).
- [3] Bojan Mrazovac, Milan Z. Bjelica, Dragan Kukulj, Branislav Todorović, Nikola Teslic, A Human Detection Method for Residential Smart Energy Systems Based on Zigbee RSSI Changes, IEEE Transactions on Consumer Electronics, Vol.58, No.3, pp.819-824 (2012).
- [4] Zixiang Ma, Bang Wu, Stefan Poslad, A WiFi RSSI ranking fingerprint positioning system and its application to indoor activities of daily living recognition, International Journal of Distributed Sensor Networks 2019, Vol. 15, No.4 (2019).
- [5] Jiang Xiao, Kaishun Wu, Youweh Yi, Lu Wang, Lionel M. Ni, Pilot: Passive Device-Free Indoor Localization Using Channel State Information, IEEE 33rd International Conference on Distributed Computing Systems, pp.236-245(2013).
- [6] Wei Wang, Alex X. Liu, Muhammad Shahzad, Kang Ling, Sanglu Lu, :Understanding and Modeling of WiFi Signal based Human Activity Recognition, Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp.65-76 (2015).
- [7] Kamran Ali, Alex X. Liu, Wei Wang, Muhammad Shahzad, :Keystroke Recognition Using WiFi Signals, Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp.90-102 (2015).
- [8] Siamak Yousefi, Hirokazu Narui, Sankalp Dayal, Stefan Ermon, Shahrokh Valae, A Survey on Behavior Recognition Using WiFi Channel State Information, IEEE Communications Magazine, Vol.55, pp.98-104 (2017).
- [9] Hoonyong Lee, Changbum R Ahn, Nakjung Choi, :Fine-grained occupant activity monitoring with Wi-Fi channel state information: Practical implementation of multiple receiver settings, Advanced Engineering Informatics, Vol.46 (2020).
- [10] 田中悠貴, 石坂拓海, 斎藤隆仁, 池田大造, 峰野博史, “動画による正解ラベルを用いた CSI ベース行動認識

の検討,” 第 83 回情報処理学会全国大会(2021).

- [11] Takashi Nakamura, Mondher Bouazizi, Kohei Yamamoto, Tomoaki Ohtsuki, Wi-Fi-CSI-based Fall Detection by Spectrogram Analysis with CNN, 2020 IEEE Global Communications Conference(2020).
- [12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby, : An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, ICLR2021 (2021)
- [13] Han K., Wang Y., Chen H., Chen X., Guo J., Liu Z., Tang Y., Xiao A., Xu C., Xu Y., Yang Z., Zhang Y., Tao D.: A Survey on Visual Transformer (2021).
- [14] Shi, C., Liu, J., Liu, H., and Chen, Y., Smart User Authentication through Actuation of Daily Activities Leveraging WiFi-enabled IoT. In Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing (2017).
- [15] Daniel Halperin, Wenjin Hu, Anmol Sheth, David Wetherall, Tool release: gathering 802.11n traces with channel state information, ACM SIGCOMM Computer Communication Review, Vol.41, pp.53–53(2011).
- [16] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, : Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, ICCV 2021(2021).