

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

田中悠貴[†] 石坂拓海[†] 斎藤隆仁[‡] 池田大造[‡] 峰野博史[‡]

[†]静岡大学情報学部 [‡]株式会社NTT ドコモ

1. はじめに

近年, IoT や AI の発展によりセンシング技術に注目が集まっている. 既存のセンシング技術として, カメラや加速度センサなどが用いられてきたが, プライバシーの問題やデバイス着用による使用の煩わしさが課題である. これらの課題を解決するセンシング技術として, Wi-Fi チャネル状態情報(Channel State Information: CSI)を用いたセンシングが注目されている. 昨今では Wi-Fi 通信環境は生活基盤として浸透しているため, 新たに設置するコストが抑えられる点においても有用と言える.

CSIを用いた研究は, Deep Learning-Basedの手法が増加しており, 飛躍的な認識精度の向上を実現している. しかし, Deep Learning-Basedの手法は大量の学習データセットを必要とし, 多くの場合が手動によってラベル付けされているため, データセット作成に労力を要するという課題がある. また, データ収集の際には, 識別する行動ごとにデータを収集するため, データ収集にも多くの時間を要してしまう.

本研究では, SoTA (State of the Art)な動画認識技術で検出された行動を正解ラベルとし, CSIへ自動でラベル付けすることで, 正解ラベル付けや行動ごとにデータ収集する労力削減を図りつつ, これまで認識できていなかった複雑な状況の行動も認識できる CSI ベース行動認識手法を提案する.

2. 関連研究

従来の身近なセンシング技術として, 加速度センサを搭載したウェアラブル端末等がある. Ling B.らは, 5つの小型二軸加速度計を身体の異なる部位に装着し, 歩くや座るなどの身体活動を検出する加速度やエネルギー, 周波数領域エントロピーの相対性を使ったアルゴリズムを開発し, 約84%の精度で識別可能にした[1]. 一方, CSIを用いたデバイスフリーな CARM[2]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 速度モデルと, 異なる人体部位の移動速度と特定の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, 座るなどの日常活動識別を約96%の精度で実現している. また昨今では, Deep Learning-Basedの手法が増加しており, Siamak らによる長期短期記憶(Long Short Term Memory: LSTM)を用いた行動認識手法[3]では, 歩く, 座るなどの日常活動識別で約90%の精度を達成している. 時系列情報を保持できる LSTMを採用することで, これまで必要であった CSI の複雑な前処理なしで日常活動識別精度の高精度化を実現している. しかし, Deep Learning-Basedの手法は, 大量のデータセットを必要とし, データ収集や正解ラベル付けの労力が多いだけでなく対象行動も限定されていると言える.

Investigating of CSI-based Behavior Recognition with Correct Annotation Using Video

Yuki Tanaka[†], Takumi Ishisaka[†], Takato Saito[‡], Taizo Ikeda[‡], Hiroshi Mineno^{‡*}

[†]Faculty of Informatics, Shizuoka University

[‡]NTT DOCOMO, Inc.

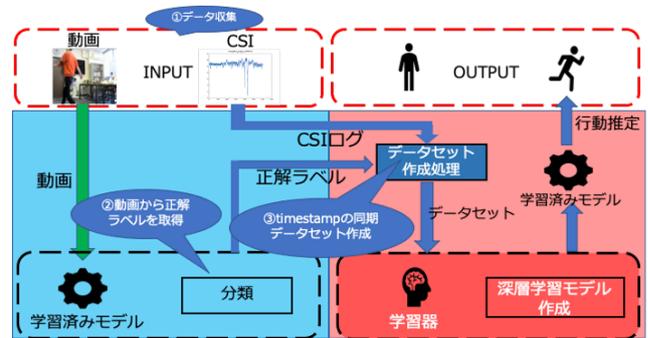


図1 提案手法の概要

3. 動画による正解ラベルを用いた CSI ベース行動認識

Deep learningによる動画認識を用いて検出された行動を正解ラベルとし, CSIへ自動でラベル付けすることで, 正解ラベル付けや行動ごとにデータ収集する労力削減を図りつつ, これまで認識できていなかった複雑な状況の行動も認識できる CSI ベース行動認識手法を提案する. 図1に提案手法の概要を示す. また, CSI収集時は多くの場合, 識別する行動ごとに CSI データを収集するが, 本研究では, 識別する対象行動を個別に収集せずとも, 一連の行動として CSI データを収集できるようにし, データ収集の労力削減を実現する.

本手法では, 時系列データである CSI に対して動画解析結果を用いて自動ラベリングを実現するため, 行動認識タスクにおいて高い予測性能を持ち時系列情報も学習可能な 3D ResNet[4]を用いることとする. もちろん, より高性能な動画認識技術が研究開発されれば, それを採用すればよい.

また, timestampと動画解析の出力結果である出力ラベルを用いて CSI にラベル付けするため, データ収集時にカメラと Wi-Fi 受信機間で時刻同期させる必要がある. そのため, NTP(Network Time Protocol)を用いて収集時に時刻同期する. ただし, NTPを使用しても1s程のラグが生じるため, 自動ラベル付けプログラム中で, CSI の timestampと動画解析結果で出力された timestampを同期させる. また, CSI の timestampは, Wi-Fi 受信機に搭載された NIC に依存するため実時間に変換しておく必要がある.

CSI ベースの行動認識で使用する学習器は, LSTMを使用することとした. CSIを用いた関連研究[4]によると, 隠れマルコフモデル(Hidden Markov Model: HMM), ランダムフォレスト, LSTMで性能比較を行っており, LSTMが最も日常活動識別性能が高い結果を得ていることから, 本研究でも再帰的ニューラルネットワーク(Recurrent Neural Network: RNN)において主流な LSTMを採用した.

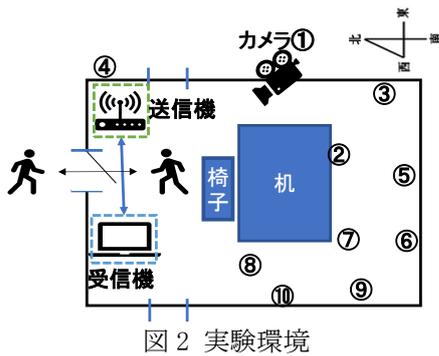


図2 実験環境

4. 基礎実験と結果

評価実験として、CSI への自動ラベル付け用動画解析モデル作成のための実験 1 と、CSI ベース行動認識モデル作成用のデータ収集を行う実験 2 を実施した。対象となる行動は、行動なし (class 0)、部屋への入室 (class 1)、椅子へ座る (class 2)、椅子から立つ (class 3)、部屋からの退室 (class 4)、の 5 種類とした。図 2 に実験環境を示す。壁は一部鉄筋コンクリートで、ドアや窓は締め切った状態で実験を行った。

実験 1 では、図 2 に示す①から⑩の 10 箇所に Web カメラを設置し、5 種類の行動データの収集を行った。収集した行動データは、各箇所につき 5 種類の行動×10 試行の 50 個のデータとなる。実験 2 では、図 2 のように部屋内のドア周辺に Wi-Fi 送受信機を設置した。無線通信はフレネルゾーンという楕円形の無線通信路が存在し、送受信機間の直線を中心とした楕円体を通じて送受信される。そのため、フレネルゾーンに人が入り、行動することで CSI に顕著な変化が表れると考え、送受信機を設置した。CSI 取得には、Intel 5300[5]を搭載した受信機用ノート PC、送信機として無線 LAN ルーター (Buffalo WSR-2533DHPL) を使用し、受信機から送信機へ ping を 1ms 間隔で送信することで CSI を収集した。ここで、Intel 5300 の timestamp は、1MHz クロックの下位 32 ビットのため、約 4300 秒ごとにラップする。そのため、CSI の timestamp を 10^6 倍したものを 1 秒単位として扱う。

CSI は床や壁、人の動きによる電波損失や反射・回折等のマルチパスの影響によって生じる電波の振幅と位相の変化を複素数の絶対値と偏角で表したものであり、1 パケットごとに送信機アンテナ数、受信機アンテナ数、サブキャリア数の積の次元の行列として収集される。今回使用する送受信機のアンテナ数はそれぞれ 2 本、サブキャリア数は 30 であるため、 $2 \times 2 \times 30$ 次元の行列が 1 パケットに含まれる。そこで、評価モデルで使用する説明変数として、図 2 の環境で収集した CSI の絶対値、つまり CSI 振幅データの $2 \times 2 \times 30$ の 120 次元を用いた。また、CSI 収集回数は、計 20 回実施した。既存研究における CSI 収集方法の多くは、対象とする行動ごとに収集しているが、本研究はデータ収集の労力削減も目的としているため、5 種類の動作を一連の動作として一回で収集している。

3D ResNet を用いて自動ラベリング用の動画認識モデルを構築した際の推論結果を図 3 に示す。上記行動について、平均約 90% の精度で正解ラベリング可能なことを確認した。図 4 に、本結果を用いて CSI に対しラベル付を行って学習させた CSI ベース行動認識モデルの結果を示す。上記行動について、AUC (Area Under the Curve) が 0.7 であった。全体的に性能がそれほど高くないため、自動ラ

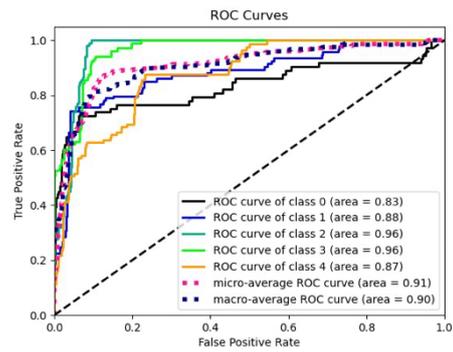


図3 自動ラベリング用モデルの評価結果

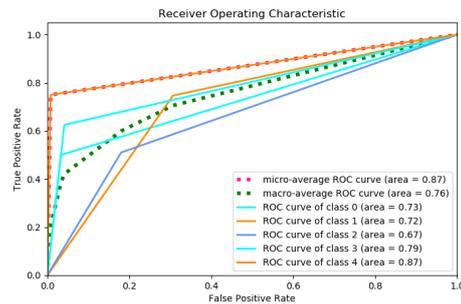


図4 CSI ベース行動認識モデルの評価結果

ベリング用の動画認識モデルの精度向上に加え、CSI 変化の特徴と各行動をより適切に表現する特徴量エンジニアリングが重要であると考える。

5. おわりに

SoTA な動画認識技術で検出された行動を正解ラベルとし、CSI へ自動でラベル付けすることで、正解ラベル付けや行動ごとにデータ収集する労力削減を図りつつ、これまで認識できていなかった複雑な状況の行動も認識できる CSI ベース行動認識手法を検討した。基礎実験の結果、動画認識モデルの精度向上と、CSI 位相情報といった説明変数の追加や、CSI 変化と各行動をより表現する適切な特徴量エンジニアリングで性能向上を目指す。また、送受信器や部屋にあるモノの位置変更など、より実生活に近い環境条件で実験を行い本手法の有効性実証を進める

謝辞

本研究の一部は、東北大学電気通信研究所における共同プロジェクト研究の支援によって行われた。

参考文献

- [1] Ling B. et al.: Activity Recognition from User-Annotated Acceleration Data, Pervasive (2004).
- [2] W. Wang et al.: Understanding and Modeling of WiFi Signal based Human Activity Recognition, MobiCom, pp.65-76 (2015).
- [3] S. Yousefi, et al.: A Survey on Behavior Recognition Using WiFi Channel State Information, IEEE Commun. Mag., Vol.55, Issue.10, pp.98-104 (2017).
- [4] K. Hara, et al.: Towards Good Practice for Action Recognition with Spatiotemporal 3D Convolutions, ICPR, pp.2516-2521 (2018).
- [5] Halperin, et al.: Tool Release: Gathering 802.11N Traces with Channel State Information, ACM SIGCOMM CCR, Vol.41, No.1, p53 (2011).