

3次元点群を用いたマイクロ行動認識手法の提案

三嶋 祐輝^{1,a)} 松井 智一¹ 松田 裕貴^{1,2} 諏訪 博彦^{1,2} 安本 慶一^{1,2}

概要:近年、宅内環境のスマート化に伴い、宅内で生活する人の行動認識に関する研究ならびにシステム開発が盛んに行われている。宅内の生活行動を認識することにより、居住者の生活・行動支援への応用が期待できる。従来の行動認識手法として、動画像を使ったものはプライバシーの侵害が懸念され、ウェアラブルセンサを使ったものは常に身につける必要がある上、バッテリーの定期的な充電が必要であり、振動センサを使ったものは、認識できない行動があるといった問題があった。また、従来の行動認識では、料理や入浴といったマクロな行動を対象としており、正確な生活・行動支援への応用には、より詳細な行動であるマイクロ行動（例えば、料理を構成する食材を持つ、切るなど）の認識が必要であると考えられる。そこで本研究では、これらの問題点を解決する手法として、3次元点群を用いたマイクロ行動認識手法を提案する。3次元点群は対象物を3次元座標の点の集合で表現したものであり、動画像よりも低プライバシーで高精度な認識が期待される上に、非接触でのセンシングが可能であり、行動の種類に関係ない認識を行うことができると考える。提案手法で解決すべき課題は、対象とするマイクロ行動の定義と3次元点群からのマイクロ行動の認識である。本研究では一般家庭の料理行動の中に現れる、“切る”や“持つ”、“置く”といった行動をマイクロ行動として定義する。一般家庭に設置した Azure Kinect によって得られる3次元点群に、時系列3次元点群を入力データとして分類を行うことが可能な機械学習モデルである P4Transformer を適用することでマイクロ行動認識を行う。実験の結果、Leave-One-Session-Out 交差検証を用いた実験の結果、8クラスのマイクロ行動を平均0.743の精度で分類可能であることが分かった。

キーワード: 3次元点群, マイクロ行動, 行動認識, 機械学習

1. はじめに

近年、宅内環境のスマート化によって宅内で生活する人の行動認識に関する研究ならびにシステム開発が盛んに行われている。宅内の生活行動を認識することにより、居住者の生活習慣に応じた家電の自動制御 [1] [2] や高齢者の生活補助といった行動支援への応用が期待できる。家電の自動制御では居住者の行動に応じて、風呂の湯を沸かす、エアコンの電源をオンにする、寝室の電気を消灯するといったことが可能になる。これによって、本来であれば居住者が行わなければならない家電の操作を自ら行う必要がなくなり、結果として居住者の生活の質向上に加えて、効率的な省エネ家電制御の実現が期待される。高齢者の生活補助では、高齢者の宅内行動を認識することで、転倒などの異状な行動を検知し、非常時の迅速な対応が可能となる。総務省の調査によると、平成19年に65歳以上の割合が全人

口の内21%を超える超高齢化社会となった。令和5年版高齢者社会白書によると、65歳以上人口は3,624万人となり、総人口に占める割合は29.0%となっている [3]。これらを踏まえると、今後の日本の高齢化社会において、高齢者の精神的負担の軽減や生活の質を向上させるという観点から、高齢者の生活を補助することは年々重要度が増して行くことが考えられる。ここで居住者の日常生活行動をより正確に認識するためには、料理や掃除、洗濯といったマクロな行動を構成するより詳細な行動であるマイクロ行動を認識することが重要であると考えられる。

行動認識の手法としてカメラから得られる動画像を用いるものがある [4]。この手法では認識のために動画像の情報を用いるためプライバシーの侵害が懸念される。また、人物以外の物体によりオクルージョンが発生するという問題がある。ウェアラブルデバイスを用いる行動認識手法では、加速度センサを用いるものが主流となっており、足や腕、胸に装着するものが存在する [5] [6]。しかしこれにはウェアラブルセンサを常に身につけなければならない点やバッテリーの定期的な充電が必要であるといった問題があり、持続的な使用が難しいといったデメリットがある。振

¹ 奈良先端科学技術大学院大学

Nara Institute of Science and Technology

² 理化学研究所革新知能統合研究センター (AIP)

RIKEN Center for Advanced Intelligence Project (AIP)

a) mishima.yuki.mu6@is.naist.jp

動センサを利用した行動認識手法も多数提案されており、Sunら [7] は、レーザードップラー振動センサを天井のある一点に設置することで、壁や天井の微小な振動の測定を行い、振動データを学習データとして、機械学習モデルにより解析を行うことによって、“ドライヤーを使う”や“シャワーを浴びる”などの、18種類の行動を認識するシステムを提案している。Bondeら [8] は、卓上に振動センサを設置し、周囲の構造物の振動から、“歩く”や“タイピング”、“話す”などの行動の識別を行なった。振動センサを用いた手法は、カメラなどのビジョンベースの手法に比べて、オクルージョンが発生しないという利点がある。一方で、物体と接触しない行動は原理的に分類することが困難であるという問題がある。また、これらの研究はマイクロ行動の認識には至っていない。そこで本研究では、RGB-Depthカメラによってセンシングされる3次元点群を用いてプライバシーに配慮した行動認識を目指す。3次元点群は、対象物を3次元座標の点の集合で表したものである。各点は空間内の位置座標と必要に応じて色や法線ベクトルなどの情報を保持することができるため、人物と物体との位置関係を3次元的に扱うことができる。そのため物体を考慮した処理を行うことで従来の動画像による行動認識の手法に比べて、より精度の高い行動認識が期待される。

本システムは、一般家庭のキッチンに構築する。4台のAzure Kinectを用いて3次元点群を収集し、従来認識が難しかった料理行動におけるマイクロ行動認識を行う。マイクロ行動認識によって、マクロな行動をより詳細な粒度で理解することができる。これにより、最終的には一連のマイクロ行動からマクロな行動を認識することを目指す。

提案手法では、一般家庭の料理行動の中に現れる“切る”や“持つ”、“置く”といった16種類の詳細な行動をマイクロ行動として定義した。また、一般家庭のキッチンに設置したRGB-DカメラであるAzure Kinectを用いて料理行動におけるマイクロ行動のデータを収集した。収集したデータを3次元点群に変換し、時系列3次元点群を入力データとして分類を行うことが可能な機械学習モデルであるP4Transformer [9]に学習させることで、マイクロ行動認識モデルの構築を行なった。

実験の結果、Leave-One-Session-Out 交差検証によって料理における8種類マイクロ行動のいずれの状態であるかを平均0.743の精度で分類可能であることを示した。

2. 関連研究

本章では、宅内行動における非接触なマクロ行動認識手法およびマイクロ行動認識に関する研究について概観する。また、3次元点群を用いた行動認識に関する既存研究について述べる。

2.1 非接触センシングによるマクロ宅内行動認識

Cookらが提案したCASASは、光量センサ、気温センサ、モーションセンサ、ドアセンサなどから取得したデータを用いて、10種類のマクロ行動を平均で0.581の精度で認識することに成功している [10]。

Sasakiらは、家電の利用状況から行動認識を行う手法を提案し、“食事”や“洗濯”、“読書”といった9種類のマクロ行動を平均で0.675の精度で認識している [11]。

Mastuiらは、SALONと呼ばれるPIRセンサ、気温・湿度・騒音などが計測可能な環境センサ、ドアセンサ、環境発電によって駆動するアノテーション用ボタン、データ収集・解析用サーバからなるシステムを提案した。提案システムを用いることで“入浴”、“食事”、“調理”、“睡眠”の4種類の行動認識において、再現率で0.723、F値で0.407の精度を達成している [12]。

これらの研究では、宅内における一般的な日常行動の認識を対象としているが、適切な生活補助や家電の自動制御のためには、より詳細な行動であるマイクロ行動の認識が必要であると考えられる。

2.2 マイクロ行動認識に関する研究

Michibataら [13] は、CNNを用いて一人称視点での動画情報から調理作業におけるマイクロ行動認識を行なった。調理作業は“調理器具を取り出す”、“食材を取り出す”、“洗う”などの行動を対象としている。

Mizumotoら [14] は、深度カメラKinectV2を用いて、調理作業中のマイクロ行動認識を行なっている。深度カメラより得られる深度データをもとに算出した骨格情報から、調理者のキッチンにおける相対位置ならびに胴体と上肢の位置関係と角度を特徴量としている。これらの特徴量をランダムフォレストの学習用データに用いることで、行動認識を行なっている。クロスバリデーションを用いて、“picking up a seasoning”、“cutting”、“mixing”などの5種類の調理作業を78%の精度で分類可能とした。

マイクロ行動を認識するためには、人とオブジェクトを3次元的に扱うことができる3次元点群の活用が望ましいと考えられる。

2.3 3次元点群を用いた行動認識

深度カメラやLiDARセンサの小型化、低価格化に伴い、3次元点群を用いた研究が盛んに行われている。

Liuら [15] は、時系列3次元点群を学習するための新しいニューラルネットワークアーキテクチャであるMeteorNetを提案した。MeteorNetを用いて行動認識や、セマンティックセグメンテーション、シーンフロー推定など、様々な3D認識タスクでベンチマークを行なっている。

Fanら [9] は、Point 4D ConvolutionとTransformerから構成されるP4Transformerを提案した。P4Transformer

表 1: 料理行動におけるマイクロ行動群

マイクロ行動	行動内容	マイクロ行動	行動内容
開ける	引き出しや電子レンジ等を開ける動作	切る	包丁で食材を切る動作
閉める	引き出しや電子レンジ等を閉める動作	注ぐ	水や調味料を入れる動作
持つ	調理器具や食材等を持つ動作	電源を押す	IH やガスコンロの電源をオンオフする動作
置く	調理器具や食材等を置く動作	混ぜる	鍋やボウル等をかき混ぜる動作
剥く	ピーラーを使って皮を剥く動作	撒く	調味料を食材にかける動作
蛇口を触る	シンクの蛇口を触る動作	潰す	マッシャーを使って食材を潰す動作
洗う	シンクで食器や食材等を洗う動作	入れる	鍋やボウル等に食材や調味料を入れる動作
拭く	シンク周りを台拭きで拭く動作	掬う	お玉やスプーン等を使って掬い上げる動作



図 1: マイクロ行動データ収集システム構成図

は、時系列 3 次元点群から時空相関を捉えることができる。P4Transformer を用いて、行動認識、シーンフロー推定といった 3D 認識タスクでベンチマークを行った。行動認識のタスクでは、深度マップシーケンスの公開データセットである MSR Action3D Dataset [16] から 3 次元点群を構築し、時系列 3 次元点群を P4Transformer に学習をさせた。その結果、20 クラスの分類において最大で 90% の認識精度を得ている。また、3 次元行動認識のための膨大なデータセットである NTU RGB+D 60 [17] とその拡張版である NTU RGB+D 120 [18] でも同様なベンチマークを行い、最大で 96.4% の認識精度を得ている。このことから、P4Transformer は時系列 3 次元点群を用いた行動認識タスクにおいて有用であると考えられる。しかし、MSR Action3D Dataset は“jogging”, “tennis serve”, “pickup&throw” といった運動動作を中心としたデータセットとなっている。また、NTU RGB+D 60 と NTU RGB+D 120 は、“drink water”, “eat meal”, “brush teeth” といった様々な日常行動に関するものから構成されている。しかし、データ収集の環境が実際に人が生活しているような宅内環境ではないことに加えて、マイクロ行動に対する行動認識は行われていない。

そこで本研究では、Fan らの提案した P4Transformer を用いて、時系列 3 次元点群データからマイクロ行動の認識を行う。

3. 提案システム

本章では、3 次元点群を用いたマイクロ行動認識システムについて述べる。RGB-Depth カメラからセンシングされる人物の 3 次元点群を分析し、マイクロ行動認識を行うことを目的とする。

3.1 提案システムの概要

一般家庭のキッチンに設置した Microsoft 社製の RGB-Depth カメラである Azure Kinect を用いて、料理行動におけるマイクロ行動データを収集する。Azure Kinect はキッチンの四隅に配置されており、これによってあらゆる角度から網羅的にデータ収集を行うことができる。本システムでは、ユーザは料理開始時と終了時にスマートフォンを用いてデータ収集の開始と終了の操作を行うことができる。Azure Kinect と制御用アプリケーションが導入された PC から構成される本システムを図 1 に示す。

奈良先端科学技術大学院大学の承認(倫理審査番号:2019-I-9-2)を得て、提案システムを、一般から募集した 5 家庭のキッチンに設置し、各家庭 9 料理分の合計 45 回分の料理行動のデータを収集した。

本稿では、収集したデータのうち 1 家庭の 4 料理分のデータを扱う。収集したデータを時系列 3 次元点群に変換し、変換したものを P4Transformer に学習させることで、マイクロ行動認識モデルの構築を行う。学習済みモデルに時系列 3 次元点群を入力することで、マイクロ行動の推定を行う。

3.2 マイクロ行動の定義

本研究におけるマイクロ行動は、料理や食事、洗濯といったマクロ行動の中に現れる詳細な行動とする。本稿ではマクロ行動の中でも料理行動におけるマイクロ行動の認識を行う。実際の一般家庭の料理行動において目視で確認された単一の動作からなるマイクロ行動を対象とする。より具体的には、料理行動中に現れる“切る”や“持つ”、“置く”といった行動をマイクロ行動として定義した。定義したマイクロ行動と各行動の内容を表 1 に示す。

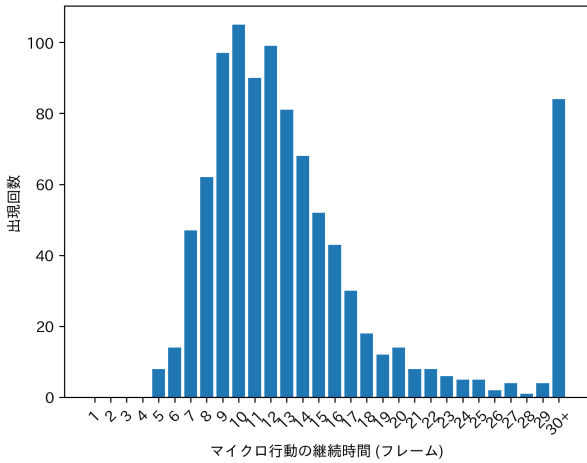


図 2: マイクロ行動の継続時間の分布

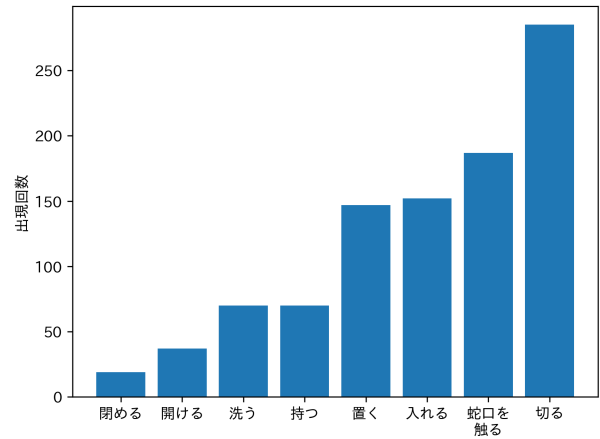


図 3: マイクロ行動の分布

3.3 マイクロ行動認識モデルの構築

Azure Kinect は RGB に加えてカメラからの距離情報を 15fps で記録しているため、1/15 秒ごとにカラーマップと深度マップが生成される。深度マップから人物の 3 次元点群に変換する流れを以下に示す。

- (1) 深度マップを Python のオープンソースライブラリである Open3D [19] を用いて 3 次元点群に変換する。
- (2) Point Cloud Library (PCL) [20] に実装されている Octree Point Cloud Change Detector を用いて、深度マップ変換後の 3 次元点群を背景データと前景データに分割する。
- (3) 前景データは背景データに存在しない人物とオブジェクトの 3 次元点群から構成されているため、密度ベースのクラスタリングアルゴリズムである DBSCAN [21] を用いて 3 次元点群のクラスタリングを行う。
- (4) (3) で算出されたクラスターのうち、最も点群の数の多いものを人物の 3 次元点群として抽出する。

データは Azure Kinect によって逐次的に収集されることから、人物の行動は時系列 3 次元点群として表現される。この時系列 3 次元点群を P4Transformer の学習データに用いることで、マイクロ行動認識のモデル構築を行う。

4. マイクロ行動認識に関する実験

4.1 実験概要

Azure Kinect によってセンシングされる人物の時系列 3 次元点群を学習データとし、P4Transformer を用いて行動の分類および評価を行う。図 1 に示されている 4 台のカメラの中でも、今回はシンクの左側に設置されている 1 台から得られるデータを用いた。

本実験では、表 1 にある料理行動に関するマイクロ行動群を対象とする。被験者 1 名による 4 回の料理行動のそれぞれについて 4 台の Azure Kinect によって 15 fps で記録

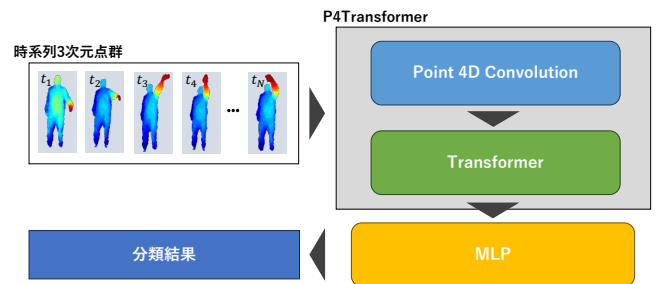


図 4: P4Transformer のアーキテクチャ図

し、合計 4 セッション（1 回の料理行動が 1 セッションに対応）のデータを作成した。セッション全体を通して現れるマイクロ行動の長さの分布を図 2 に示す。1 回のマイクロ行動の長さは最小で 5 フレーム、最大で 710 フレームであった。

4.2 分析手法

マイクロ行動認識モデル学習の前処理として、記録されたデータが“切る”や“持つ”など、どのマイクロ行動を行なっているかをラベリングした。本データは 15fps で記録しているため 1/15 秒ごとにカラーマップと深度マップが生成される。そこで、カラーマップをもとに目視で表 1 にあるマイクロ行動のいずれに該当するかをラベリングした。4 セッションにおけるラベル数の合計は 1322 個であった。しかし、4 セッションのうち料理の種類によっては存在しないマイクロ行動がある。例えばカレーを作る際には、じゃがいもを“切る”やカレー鍋をお玉で“混ぜる”といったマイクロ行動が現れるが、何かしらの食材を“潰す”といったマイクロ行動は現れない。そのため、本実験では 4 セッションに共通して存在するマイクロ行動を抽出することで、対象行動を抽出することとした。抽出後のマイクロ行動は 8 クラスであった。

データサンプルの分布を図 3 に示す。全データサンプル数は 967 点である。最も少ないデータサンプル数は“閉める”で 19 点であり、最も多いデータサンプル数は“切る”

表 2: Leave-One-Session-Out 交差検証によるマイクロ行動ごとの F 値

セッション	入れる	開ける	閉める	持つ	置く	蛇口を触る	洗う	切る
1	0.750	0.813	0.429	0.304	0.510	1.0	1.0	0.995
2	0.537	0.828	0.400	0.267	0.370	0.904	0.732	0.969
3	0.460	0.588	0.286	0.391	0.195	0.952	0.931	0.688
4	0.667	0.857	0.0	0.222	0.0	0.878	0.857	0.980
平均	0.604	0.772	0.279	0.296	0.269	0.934	0.88	0.908

表 3: 各セッションにおける Accuracy と F 値

セッション	Accuracy	F 値
1	0.841	0.725
2	0.702	0.626
3	0.567	0.562
4	0.863	0.558
平均	0.743	0.601

で 285 点である。

1 台の Azure Kinect からの視点におけるマイクロ行動のラベルが付与された時系列 3 次元点群を P4Transformer に学習させ、マイクロ行動の分類を行った。P4Transformer のアーキテクチャ構造を図 4 に示す。P4Transformer は、Point 4D Convolution と Transformer から構成されるモデルである。Point 4D Convolution では、入力された時系列 3 次元点群を時空間ごとに特徴ベクトルにエンコードすることで、局所特徴量を得る。局所特徴量を Transformer に入力し、複数フレーム間を考慮した大域特徴量へ変換する。MLP に大域特徴量を入力して、クラス分類を行う。学習時のハイパーパラメータは、入力時のフレーム数をオーバーラップのない 4 フレームとし、それ以外は Fan らの MSR Action3D Dataset を用いたクラス分類に関する実験 [9] と同じにした。

4.3 実験結果および考察

各マイクロ行動データに対して、Leave-One-Session-Out 交差検証を行い、マイクロ行動の分類精度を確認した。この手法では、1 料理を 1 セッションとし、1 つのセッションを精度検証用のテストデータ、それ以外を学習データとする。これを全てのセッションがテストとなるように繰り返し、その平均精度を求めた。各セッションにおける混同行列を図 5 に、マイクロ行動ごとの分類精度 F 値を表 2 に示す。また、表 3 に各セッションにおける Accuracy と F 値のマクロ平均を示す。表 3 にある通り、平均で 0.743 の精度で 8 クラスのマイクロ行動の分類を行うことができた。

表 2 では、“蛇口に触る”や“洗う”、“切る”といったマイクロ行動の分類精度は 90% 近くの精度となっており、高精度に分類されていることがわかる。しかし、“閉める”や“持つ”、“置く”といったマイクロ行動の分類精度は 30% を下回っている。“閉める”に関しては、図 2 にある通り

データのサンプル数が少ないことから、正しく学習が行われていないことが考えられる。

“持つ”に関しては、オブジェクトの大きさや形状によって動作の内容が異なることが原因であると考えられる。玉ねぎを“持つ”とフライパンを“持つ”は、本稿では同じマイクロ行動のラベルが割り当てられているが実際の動作の内容としては異なる。

“置く”に関しては、図 5c の混同行列より大半が“入れる”、“切る”に分類されている。これは、学習データのサンプル数が少ないことに加えて、セッション 3 では被験者が右手で包丁を持った状態でニンニクの皮を剥き、まな板の上に“置く”といったマイクロ行動が多く行われていたことが原因であると考えられる。

これらのことを踏まえると提案システムにオブジェクト認識を組み込むことで、より認識精度の改善が見込めると考える。マイクロ行動は対象とするオブジェクトの大きさや形状によってその動作内容が異なる場合があり、その際にはオブジェクトを考慮したマイクロ行動のラベリングが適切であると考えられる。

5. まとめ

本稿では、一般家庭の料理行動におけるマイクロ行動を定義し、Azure Kinect によってセンシングされる時系列 3 次元点群を用いたマイクロ行動認識手法を提案した。一般家庭のキッチンで取得した料理行動の時系列 3 次元点群を P4Transformer に学習させることで、マイクロ行動の分類を行った結果、8 種類のマイクロ行動を平均で 0.743 の精度で分類可能であることを示した。

今回は 1 台の RGB-D カメラから取得されたデータを用いてマイクロ行動を認識したが、今後は複数台のカメラから得られたデータを用いてアンサンブル学習を行うことで、遮蔽物によるオクルージョンにロバストなマイクロ行動認識モデルの構築を行う。その上で、複数台のカメラのデータを用いる時とそうでない時との分類精度を比較することで、コスト面の評価も行う。さらに、人物のみではなくオブジェクトの 3 次元点群を扱うことで、人のオブジェクトに対するインタラクションを考慮したより高精度な行動認識を目指す。

また、被験者とセッション数を増やすことで、データセットの汎用性を高めると同時に、Leave-One-Person-Out

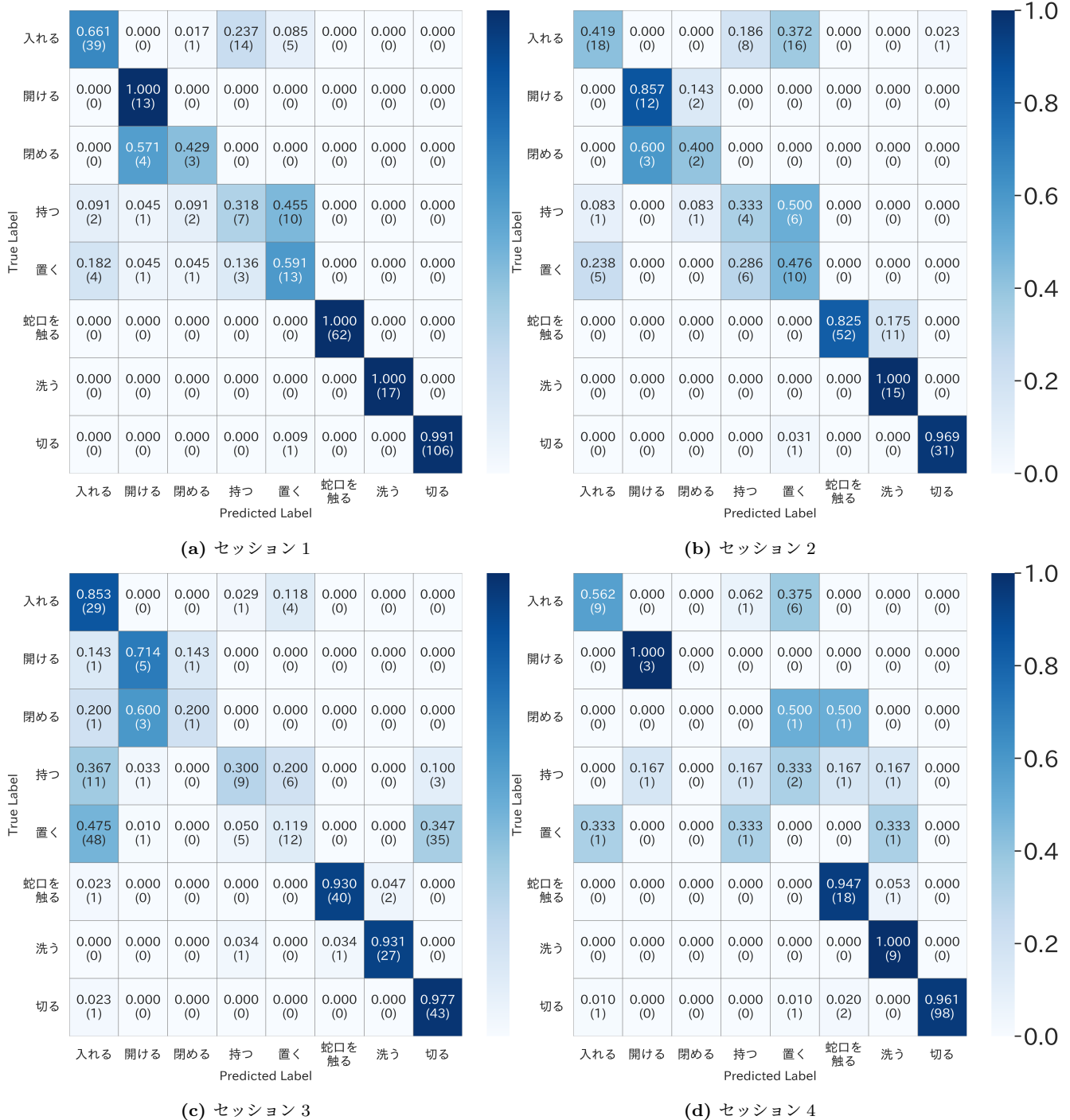


図 5: Leave-One-Session-Out 交差検証の結果に対する混同行列

交差検証を行う。これにより、環境や人に依存しない汎用的なマイクロ行動の認識が可能になると考えている。

謝辞

本研究の成果は NICT 課題 222C03 「ウイルス等感染症対策に資する情報通信技術の研究開発」の支援によるものです。

参考文献

- [1] John Krumm Brian Meyers Michael Hazas Stephen Hodges James Scott, A.J. Bernheim Brush and Nicolas Villar. *Controlling Home Heating Using Occupancy Prediction*. Proc. UbiComp 2011 (2011).
- [2] David Irwin Prashant Shenoy Sean Barker, Aditya Mishra and Jeannie Albrecht. *SmartCap: Flattening peak electricity demand in smart homes*. Proc. Percom 2012. pp.67-75.
- [3] 総務省. 令和 5 年版高齢社会白書. https://www8.cao.go.jp/kourei/whitepaper/w-2023/zenbun/05pdf_index.html, 2023.

- [4] Oscar T.-C. Chen, Ching-Han Tsai, Hung Ha Manh, and Wei-Chih Lai. *Activity recognition using a panoramic camera for homecare*. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2017. 1-6.
- [5] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. Physical human activity recognition using wearable sensors. *Sensors*, Vol. 15, No. 12, pp. 31314–31338, 2015.
- [6] Serkan Balli, Ensar Arif Sağbaş, and Musa Peker. Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. *Measurement and Control*, Vol. 52, No. 1-2, pp. 37–45, 2019.
- [7] Wei Sun, Tuochao Chen, Jiayi Zheng, Zhenyu Lei, Lucy Wang, Benjamin Steeper, Peng He, Matthew Dressa, Feng Tian, and Cheng Zhang. *VibroSense: Recognizing Home Activities by Deep Learning Subtle Vibrations on an Interior Surface of a House from a Single Point Using Laser Doppler Vibrometry*, Vol. No.21. September 2020.
- [8] Amelie Bonde, Shijia Pan, Mostafa Mirshekari, Carlos Ruiz, Hae Young Noh, and Pei Zhang. *OAC: Overlapping Office Activity Classification through IoT-Sensed Structural Vibration*. 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI), 2020. pp.216-222.
- [9] Hehe Fan, Yi Yang, and Mohan Kankanhalli. Point 4D Transformer Networks for Spatio-Temporal Modeling in Point Cloud Videos. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*, 2021.
- [10] Diane J. Cook, Aaron S. Crandall, Brian L. Thomas, and Narayanan C. Krishnan. CASAS: A Smart Home in a Box. *Computer*, Vol. 46, No. 7, pp. 62–69, 2013.
- [11] Wataru Sasaki, Masashi Fujiwara, Hirohiko Suwa, Manato Fujimoto, Yutaka Arakawa, Aki Kimura, Tomoko Miki, and Keiichi Yasumoto. Daily Living Activity Recognition through Time Series Analysis of Data Obtained with ECHONET Lite Appliances and Motion Sensors. *JSAI Technical Report, Type 2 SIG*, Vol. 2018, No. SAI-031, p. 05, 03 2018.
- [12] Tomokazu Matsui, Kosei Onishi, Shinya Misaki, Manato Fujimoto, Hirohiko Suwa, and Keiichi Yasumoto. SALON: Simplified Sensing System for Activity of Daily Living in Ordinary Home. *Sensors*, Vol. 20, No. 17, 2020.
- [13] Shinya Michibata, Katsufumi Inoue, Michifumi Yoshioka, and Atsushi Hashimoto. *Cooking Activity Recognition in Egocentric Videos with a Hand Mask Image Branch in the Multi-Stream CNN*. CEA '20. Association for Computing Machinery, 2020. pp.1–6.
- [14] Teruhiro Mizumoto, Alberto Fornaser, Hirohiko Suwa, Keiichi Yasumoto, and Mariolino De Cecco. *Kinect-Based Micro-Behavior Sensing System for Learning the Smart Assistance with Human Subjects Inside Their Homes*. 2018 Workshop on Metrology for Industry 4.0 and IoT, 2018. pp.1-6.
- [15] Xingyu Liu, Mengyuan Yan, and Jeannette Bohg. *MeteorNet: Deep Learning on Dynamic 3D Point Cloud Sequences*. IEEE/CVF International Conference on Computer Vision (ICCV), 2019. 9245-9254.
- [16] Zhengyou Zhang Wanqing Li and Zicheng Liu. *Action recognition based on a bag of 3d points*. In CVPR Workshops, 2010.
- [17] Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. NTU RGB+D: A Large Scale Dataset for 3D Human Activity Analysis, 2016.
- [18] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C. Kot. NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, No. 10, pp. 2684–2701, 2020.
- [19] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3d: A modern library for 3d data processing, 2018.
- [20] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *2011 IEEE International Conference on Robotics and Automation*, pp. 1–4, 2011.
- [21] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96*, p. 226–231. AAAI Press, 1996.