

2019年度

筑波大学情報学群情報科学類

卒業研究論文

題目

手首内側に低解像度赤外線センサを搭載した  
腕時計型端末によるハンドジェスチャ認識

主専攻 知能情報メディア主専攻

著者 大和 優輝

指導教員 高橋 伸 志築 文太郎

## 要旨

スマートウォッチ等のウェアラブルデバイスの普及に伴い、手首にセンサやカメラを装着し、手指の認識や周囲、行動を認識する研究が行われている。しかし、視覚ベースの先行研究では、手首から大きく放り出した形状を持つことや、周囲の人へのプライバシー侵害の恐れを持つことといった問題点が存在する。また、加速度や筋電位センサを用いた研究においては、手指の豊かな表現を捉えきれないことがある。本研究では、手首内側に低解像度赤外線センサアレイを搭載した腕時計型端末によるハンドジェスチャ認識を提案する。低解像度赤外線センサアレイを利用することにより、周囲の人へのプライバシー侵害の恐れをなくしながら、ハンドジェスチャ認識が可能となる。また、手首内側にセンサを配置することにより、腕時計型端末のフォームファクタを損なうことなく、多様なハンドジェスチャ認識によるインタラクションを可能とする。利用可能なハンドジェスチャとして、静的なハンドポーズ、動的なハンドジェスチャ、指先の動き、および手の位置関係がある。本手法により、端末を装着した手の反対の手の動きを自然に認識することが可能となる。さらには、スマートウォッチの画面を見るなどの端末を装着した側の掌を下に向ける動作の拡張が可能となる。赤外線センサアレイ、腕時計型マイクロコントローラからなるハードウェアと、ハードウェアによって取得したセンサデータを基に機械学習を行い、ハンドジェスチャを認識、アプリケーションへ適用するソフトウェアからなるプロトタイプを実装した。ハンドジェスチャ認識精度を評価する複数の実験を行い、プロトタイプの性能を評価した。本提案手法を適用できる3つのアプリケーションの提案とともに、今後の課題を示した。

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	背景 . . . . .	1
1.2	目的とアプローチ . . . . .	1
1.3	本研究の貢献 . . . . .	2
1.4	本論文の構成 . . . . .	2
<b>第 2 章</b>	<b>関連研究</b>	<b>3</b>
2.1	ウェアラブルデバイスにおけるハンドジェスチャ認識 . . . . .	3
2.2	赤外線センサを利用した研究 . . . . .	4
2.3	センサの位置 . . . . .	4
2.4	本研究の位置付け . . . . .	5
<b>第 3 章</b>	<b>手首内側の低解像度赤外線センサアレイ によるハンドジェスチャ認識</b>	<b>6</b>
3.1	概要 . . . . .	6
3.2	多様なハンドジェスチャ入力 . . . . .	7
<b>第 4 章</b>	<b>プロトタイプ</b>	<b>9</b>
4.1	センシングデバイス . . . . .	10
4.1.1	AMG8833 . . . . .	10
4.1.2	M5StickC . . . . .	10
	IMU . . . . .	11
4.1.3	実装 . . . . .	11
4.2	認識ソフトウェア . . . . .	12
4.2.1	処理の流れ . . . . .	12
4.2.2	ポーズ・ジェスチャ分類 . . . . .	13
4.2.3	指先の動きのトラッキング . . . . .	13
4.2.4	手の位置の推定 . . . . .	13
4.2.5	Android アプリケーション . . . . .	14
<b>第 5 章</b>	<b>実験</b>	<b>15</b>
5.1	予備実験 . . . . .	15

5.1.1	予備実験 1：距離による取得温度データの差異	15
5.1.2	予備実験 2：手のポーズ認識	16
	目的	16
	手順	16
	結果	17
5.1.3	予備実験 3：手のジェスチャ認識	17
	目的	17
	手順	17
	結果	18
5.2	実験 1：ハンドジェスチャ・ポーズ分類精度評価実験	19
5.2.1	概要	19
5.2.2	手順	21
5.2.3	結果	21
	セッション内精度	22
	leave-one-session-out 交差検証	23
	leave-one-user-out 交差検証	24
	加速度データの利用	24
5.3	実験 2：データ拡張による汎化性能評価実験	25
5.3.1	概要	25
5.3.2	手順	26
5.3.3	結果	27
5.4	実験 3：既存データセットを用いた認識精度評価実験	28
5.4.1	概要	28
5.4.2	手順	29
5.4.3	結果	29
5.5	実験 4：他の入力モダリティにおける精度評価実験	30
	指先の動きのトラッキング	30
5.5.1	概要	30
5.5.2	手順	30
5.5.3	結果	30
	手の位置の推定	30
5.5.4	概要	30
5.5.5	手順	31
5.5.6	結果	31
<b>第 6 章</b>	<b>アプリケーション</b>	<b>32</b>
6.1	スマートウォッチのジェスチャ入力拡張	32
6.2	指先トラッキングを利用した手書きメモアプリケーション	32
6.3	手とセンサの相対距離を用いた連続値操作	32

第 7 章 結論	34
謝辭	35
參考文獻	36

# 目次

3.1	ハンドジェスチャ認識イメージ	6
3.2	指先の動き利用概要図	8
3.3	手の位置利用概要図	8
4.1	プロトタイプの概要	9
4.2	AMG8833 (Conta サーモグラフィ)	10
4.3	M5StickC	10
4.4	実装したハードウェア	11
4.5	装着している様子	12
4.6	認識ソフトウェアの処理の流れ	13
4.7	データ収集時の Android アプリケーション例	14
5.1	予備実験 1 の実験の様子	15
5.2	距離と温度データの可視化	16
5.3	対象となるハンドポーズ (上:3 クラス, 下:5 クラス)	17
5.4	予備実験 2-1: 主成分分析結果可視化	18
5.5	予備実験 2-1: 混同行列	18
5.6	予備実験 2-2: 主成分分析結果可視化	18
5.7	予備実験 2-2: 混同行列	18
5.8	予備実験 3: 主成分分析結果可視化	19
5.9	予備実験 3: 混同行列	19
5.10	ポーズセット	20
5.11	ジェスチャセット	20
5.12	混同行列: ポーズ	22
5.13	混同行列: ジェスチャ	22
5.14	LOSOVCV 混同行列: ポーズ	23
5.15	LOSOVCV 混同行列: ジェスチャ	23
5.16	LOUOCV 混同行列 (ポーズ: P1)	25
5.17	LOUOCV 混同行列 (ポーズ: P2)	25
5.18	LOUOCV 混同行列 (ポーズ: P3)	25
5.19	LOUOCV 混同行列 (ポーズ: P4)	25

5.20	LOUOCV 混同行列 (ポーズ:P5)	26
5.21	LOUOCV 混同行列 (ジェスチャ:P1)	27
5.22	LOUOCV 混同行列 (ジェスチャ:P2)	27
5.23	LOUOCV 混同行列 (ジェスチャ:P3)	27
5.24	LOUOCV 混同行列 (ジェスチャ:P4)	27
5.25	LOUOCV 混同行列 (ジェスチャ:P5)	28
5.26	データ拡張 LOUOCV 混同行列	28
5.27	混同行列 (EgoGestureDataset による学習)	29
5.28	各交点ごとの誤差の平均 (cm)	31
6.1	スマートウォッチアプリケーション例	33
6.2	メモアプリケーション例	33

# 表目次

5.1	サポートベクタマシンハイパパラメータ候補 . . . . .	20
5.2	ランダムフォレストハイパパラメータ候補 . . . . .	20
5.3	勾配ブースティングハイパパラメータ候補 . . . . .	21
5.4	セッション内の精度結果 . . . . .	22
5.5	leave-one-session-out 交差検証結果 . . . . .	23
5.6	leave-one-user-out 交差検証結果 . . . . .	24
5.7	加速度を利用したセッション内の精度結果 . . . . .	24
5.8	画像変換範囲 . . . . .	26
5.9	画像変換範囲 . . . . .	30
5.10	画像変換範囲 . . . . .	31

# 第1章

## 序論

### 1.1 背景

スマートウォッチや活動量計などの手首装着型端末が普及しており、有力な操作手法としてスクリーンへのタッチが挙げられるが、手首装着型端末に搭載されるスクリーンは小型であるため、入力や操作はユーザにとって負担となり得る。いくつかの端末にはボタンインタフェースが存在するが、スクリーンへのタッチと比較して入力語彙は限られたままである。一方、ハンドジェスチャによるインタラクションは、端末の特性に影響されずに豊かな表現力を持つため、特にウェアラブル端末に対して有望なアプローチである。そのため、スマートウォッチに搭載されたセンサを用いてハンドジェスチャを認識することが提案されてきた [1]。さらには、手首に外部のセンサやカメラを配置し、手 [2, 3] や指 [4, 5]、行動 [6, 7] を認識する研究も行われている。

手指によるジェスチャは豊かな表現性を持つため、視覚ベースのセンサやカメラを用いる手法が多く存在する [8, 9]。しかしながら、高解像度の動画像を用いる認識には、個人の特定につながるなどプライバシー侵害の恐れがある。特にウェアラブルデバイスにおいては、センサの取得可能範囲が装着しているユーザだけに留まらないため、周囲の人への配慮が必要となる。また、いくつかの視覚ベースの研究 [8, 9, 7] は、センサを取り付けた手首から同じ側の手指の認識を行う。これらのデバイスは、センサの視野範囲に手を配置する必要があり、手首から大きく突き出した形状となっているために、日常生活における動作を阻害する恐れがある。すなわち、スマートウォッチといった腕時計型端末のフォームファクタを損なわないことは、ハンドジェスチャ認識を可能とする上で重要な要素となる。

### 1.2 目的とアプローチ

本研究の目的は、プライバシー侵害を軽減しながら、腕時計型端末の手首内側に装着したセンサによりハンドジェスチャ認識を可能とし、多様なインタラクションを実現することである。そのため、低解像度の赤外線センサアレイを手首内側に装着し、センサから取得したデータを機械学習にかけて手指を認識する手法を提案する。

赤外線センサは、自動ドアや自動照明、セキュリティカメラといった人体検知にしばしば用いられており、日常生活の至る所に存在しているが、ほとんどの人はその赤外線センサによりプライバシーが侵害されていると感じないだろう。赤外線センサは、周囲の人へのプライバシー侵害の問題を解決することができるが単眼のセンサを用いた場合、手指の豊かな動きを捉えることは難しい。そこで、赤外線センサを  $8 \times 8$  のアレイ状に並べたセンサを利用することにより、手指の細かな動きを認識し、多様なインタラクションを可能とする。

また、手首内側にセンサを配置することにより、腕時計型端末のフォームファクタを損なうことなく、多様なハンドジェスチャを認識することを可能とする。この配置により、スマートウォッチの画面を隠すことなくハンドジェスチャ入力や指先トラッキング入力が可能となり、スマートウォッチのセンサと組み合わせることにより片方の手首に装着したデバイスにて両手の動きをインタラクションに利用することが可能となる。

### 1.3 本研究の貢献

本研究の貢献を以下に述べる。

- 低解像度赤外線センサがハンドジェスチャ認識可能であることを確かめた。
- 開発した提案手法のプロトタイプによって、腕時計型端末の手首内側に装着したセンサによりハンドジェスチャ認識を可能とすることを確かめた。
- 手首内側にセンサを配置することにより、多様なハンドジェスチャ入力を可能とすることを提案し、実現性を示した。

### 1.4 本論文の構成

第1章では、本研究の背景と目的を述べた。第2章では、関連研究と本研究の位置付けを述べる。第3章では、提案手法について述べる。第4章では、作成したプロトタイプについて述べる。第5章では、作成したプロトタイプの性能を評価するために行った実験について述べる。第6章では、提案した手法のアプリケーション例を述べる。第7章では、本研究の結論を述べる。

## 第2章

# 関連研究

### 2.1 ウェアラブルデバイスにおけるハンドジェスチャ認識

ウェアラブルデバイスにおいて、手のポーズやジェスチャを認識する手法は多数存在する。一般的なアプローチの一つとして、視覚に基づく手法が存在する [8, 10, 11, 12, 9, 13]. 腕や手首にカメラや光学センサを取り付け、ハンドポーズやハンドジェスチャ認識を行っている。Chen ら [11] は、手首の外側にひりだす形にて、RGB カメラを装着し、カメラを装着している側の手の 10 種類の American Sign Language(ASL) ハンドポーズの認識を行った。Kim ら [8] は、赤外線レーザーラインプロジェクタを用いハンドポーズの認識、Yeo ら [9] は、深度カメラとして LeapMotion を用い、ハンドポーズおよび指のタップアクションの認識を行っている。Sridhar ら [13] の研究において、スマートウォッチ等のウェアラブルデバイスに埋め込まれた深度センサを使用し、手首上方および、皮膚上の指先のトラッキングを検出した。これらの手法は、ウェアラブルデバイスに視覚ベースのセンサを取り付けることにて実現しているが、自分自身だけでなく周囲の映像を記録してしまうため、日常生活において周囲の人へのプライバシー侵害の懸念がある。

別のアプローチとして、様々なセンサ、もしくはセンシング手法を用いるものがある。グローブ型のデバイスは、高い精度にて手の形状と指の屈曲状況を認識することを可能とする [14]. しかしながら、日常生活において、手全体を覆ってしまうグローブ型のデバイスを身につけ続けることはユーザに煩わしさを与える可能性がある。そのほかにも、圧力 [15], 音波 [16, 17], 加速度 [1, 6], 筋電位 [18], 心拍 [19], 生体音響 [20], 電気インピーダンス法 [21] を用いたものが存在する。Dementyev ら [15] は、圧力センサを手首に装着し、腱の動きを検出することにより、5つの指のピンチジェスチャを分類した。

Irvantchi ら [16] は、手首に小さなトランスデューサアレイを配置し、超音波ビームフォーミングにより 6 クラスのハンドジェスチャ認識を行っている。Wang ら [17] は、高周波 (60GHz) の短距離レーダ (Google の Soli センサ\*1) に焦点をあて、11 個の指の細かい動きを含む、動的なジェスチャの認識を行っている。これらの研究は、専用の高度なセンサもしくは、日常動作を阻害する恐れのある手首から飛び出した形状のセンサを装着する必要がある。

Wen ら [1] は、市販のスマートウォッチに内蔵されている加速度センサとジャイロセンサを利用し、指をつまむ、叩く、擦るといった 5 つの動的ジェスチャの検出を行った。上記研究は、スマートウォッチを利用することにより、ユーザに追加のデバイスや日常生活を妨げるセンサを装着させずにハンドジェスチャの検出を可能としている。しかしながら、加速度用いた場合、センサを手首に装着した性質上、指先細かなの動きといった手の豊

---

\*1 <https://atap.google.com/soli/>

な表現を検出できない。

Zhang ら [21] は、電気インピーダンス法と呼ばれる電気伝導性の差を利用し内部構造を予測する手法を手首に適用し、手首の内部構造を検出することにより手形状の分類を行った。Saponas ら [18] は、肘付近にリストバンド型の筋電位センサを取り付けることにより、筋活動電位信号を取得し、大小様々な物体の把持の姿勢分類を行った。Laput ら [20] は、スマートウォッチを用いて生体音響を取得し、フリック、クラップ、タップと行ったハンドジェスチャの分類を行った。Zhang ら [19] の別の研究として、スマートウォッチに搭載されている光学式心拍センサを利用し、フォトプレチスモグラフィ信号を分析することにより、10種のハンドジェスチャを認識することを可能としている。これらの手法は、人体情報を用いるため、視覚ベースの手法に比べてジェスチャ認識の汎化性能が低下しやすい傾向がある。

## 2.2 赤外線センサを利用した研究

赤外線センサをインタラクションに利用した先行研究がある [22, 3, 23]。Gong ら [22] は、指から反射される熱赤外線信号に基づいた指先ジェスチャ認識技術“Pyro”を提案した。赤外線焦電センサを1つ利用し、プロトタイプの作成を行い、6種類の親指先のジェスチャの分類を可能とした。この研究において、様々な光状態、手の温度、背景の動きに対する“Pyro”の堅牢性が報告されており、ウェアラブルデバイスに赤外線センサを統合し、インタフェースとすることへの実現可能性を示している。Kim ら [3] は、ユーザの手首に装着し、デバイスのハンドジェスチャ制御を可能とするモバイルワイヤレスデバイス“The Gesture Watch”を提案した。“The Gesture Watch”は、ジェスチャ取得用に4つの赤外線センサを上面に、トリガ用に1つの赤外線センサを側面に配置し、それぞれのセンサから得られるデータを用いることにより、手のスワイプジェスチャを中心として、10種類のハンドジェスチャの分類を可能としている。Wojtczuk ら [23] は、16の赤外線焦電センサを4×4の正方形に配置し、PIRセンサアレイとすることにより、上下左右方向へのスワイプという単純なジェスチャの認識を可能とした。

## 2.3 センサの位置

ウェアラブル環境にて、視覚ベースのハンドジェスチャを行うには、身体にセンサを装着する必要がある。センサを取り付ける場所は、多様であり、例えば、頭に装着する研究 [24, 25, 26]、胸に装着する研究 [27]、耳に装着する研究 [28]、手首に装着する研究 [8, 10, 11, 12, 9, 13, 3] がある。

頭部に装着する研究 [24, 25, 26] は、AR, VR, MR 環境下といったヘッドマウントディスプレイ (HMD) を用いる状況にて、用いられている。センサを HMD に統合する形にて実装することにより違和感ない形にて実現されている。

胸に装着する研究 [27] は、胸部にジンバルを土台とした LeapMotion を配置し、センサ自体の方向を再配向しながら、ハンドトラッキングを行った。

耳に装着する研究 [28] は、掌へ投影を行うレーザプロジェクタと RGB カメラを耳に取り付け、投影されている映像の操作にハンドジェスチャ認識が用いられている。

手首に装着する研究は多数存在し、認識対象により取り付ける向きと位置が異なる。装着側の手指を認識する場合、手首から指先の方向へカメラやセンサを取り付ける必要がある [8, 9, 10, 12]。反対に、装着していない手指を認識する場合、手の甲側から垂直方向の向きに取り付けられている [13, 3]。これらは、向きが外側を向いており、周囲の映像を意図せず取得する可能性が存在する。

## 2.4 本研究の位置付け

本研究では、手首内側に低解像度赤外線センサを搭載した腕時計型端末によるハンドジェスチャ認識手法を提案する。提案手法は、手首内側に  $8 \times 8$  の低解像度赤外線センサアレイを装着し、装着した手とは反対の手指の認識を行う。

提案手法により、視覚ベースながらも周囲の人へのプライバシー侵害の恐れのないインタラクションが可能となる。そして、Wojtczuk ら [23] の研究と類似しているが、多様なハンドジェスチャ認識、ハンドポーズ認識、指先トラッキングを可能とし、ウェアラブルデバイスに適用している点で異なる。

## 第3章

# 手首内側の低解像度赤外線センサアレイによるハンドジェスチャ認識

### 3.1 概要

本研究では，手首内側に低解像度赤外線センサアレイを搭載した腕時計型端末によるハンドジェスチャ認識を提案する．

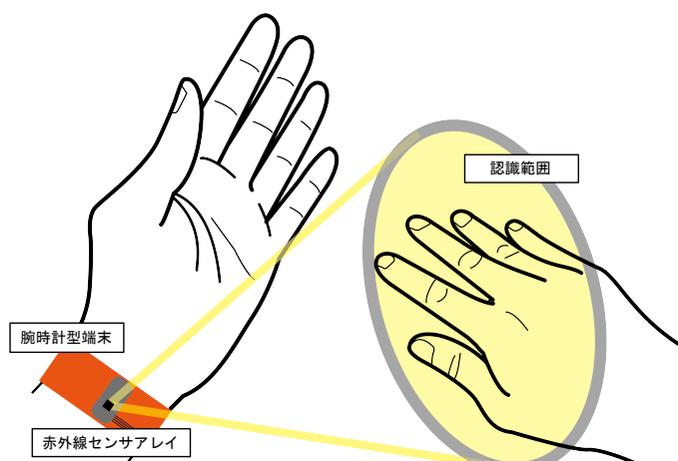


図 3.1 ハンドジェスチャ認識イメージ

図 3.1 に示すように，手首内側に赤外線センサアレイを配置し，反対側の手指の認識を行う．認識には，多くの関連研究 [8, 9, 23] において手指の認識への有効性が示されている機械学習を利用する．手首内側に赤外線センサアレイを配置することにより，腕時計型端末のフォームファクタを壊すことなく，多様なハンドジェスチャ入力を可能とする．本手法により，端末を装着した手の反対の手指の動きを自然に認識することが可能となる．

本提案手法は，スマートウォッチ等の腕時計型端末に統合されることを想定しており，元のデバイスのフォームファクタを損なわないことから，ユーザの日々の生活を阻害することはないと考えられる．また，手首内側がセンサの取得範囲であることから，スマートウォッチへのジェスチャ入力拡張時においても画面を隠すことなく，入力を行うことが可能となる．

赤外線センサは，公共の場においても多く用いられており，利用時にプライバシーの侵害の恐れが少ない．そこ

で、赤外線センサをウェアラブルデバイスの入力モダリティ拡張に用いることにした。単眼の赤外線センサによる単純なハンドジェスチャ認識の実現性については、関連 [22] に示されていたが、手指の豊かな表現力を捉えるには不十分であると考え。関連研究 [3, 23] において、赤外線センサを複数個利用し、より細かな動きの認識を可能としていたことを踏まえ、本提案手法において、赤外線センサアレイをウェアラブルデバイスに統合している。これにより、周囲へのプライバシー侵害の危険性を低減しながら、手指の豊かな動きをインタラクションに利用することが可能となる。

また、スマートウォッチや活動量計に搭載されている加速度センサを用いることにより姿勢推定が可能となり、赤外線アレイセンサの向いている方向を認識に用いることが可能となる。さらには、関連研究 [1] から、加速度センサから動的ジェスチャを認識することが可能であることがわかっている。赤外線アレイセンサと組み合わせることにより、赤外線アレイセンサの対象となるインタフェースを装着していない手指の認識だけでなく、インタフェースを装着している手の動きをインタラクションに用いることが可能であることが考えられる。これにより、人々の生活に溶け込んでいる腕時計やスマートウォッチのように片手に装着したインタフェースから、両手の手指による多様なポーズやジェスチャを認識とすることが可能となるであろう。

## 3.2 多様なハンドジェスチャ入力

本提案手法が持つ多様なハンドジェスチャ入力を以下に示す。

- ハンドポーズ・ハンドジェスチャ

赤外線センサアレイを用いて、手指の動きを捉えることにより、動的なハンドジェスチャや、静的なハンドポーズの分類が可能となる。教師あり学習により、予め用意したハンドジェスチャセット・ハンドポーズセットの分類を行う。分類結果を用いることにより、スマートウォッチへのコマンド入力やメニュー選択、モード切り替えといった利用方法がある。

- 指先の動き

アレイ状の赤外線センサを利用することにより、取得したデータを画像として捉えることが可能である。そして、画像内の指の位置を認識することにより、指先の位置に応じたインタラクションが可能となる(図 3.2)。指先の位置を連続的に取得することによる手書き文字入力や絵をかくといった利用方法が考えられる。

- 手の位置

赤外線センサは、深度センサに用いられておりセンサと対象の物体との距離によって出力する値に差異が生まれる。その差異については、予備実験にて調査を行った。また、一般的にカメラ等の視覚ベースのセンサは、対象の物体が近づくにつれ、取得した画像内の対象の物体の占める画素の割合が増加する。上記性質を利用することにより、センサと手の相対距離を利用した連続値の操作を行うことができる可能性がある。図 3.3 に示されるように、センサと手の距離を変化させることにより、音量調節やズームイン/ズームアウトといった連続値を操作するアプリケーションにも利用可能であると考えられる。

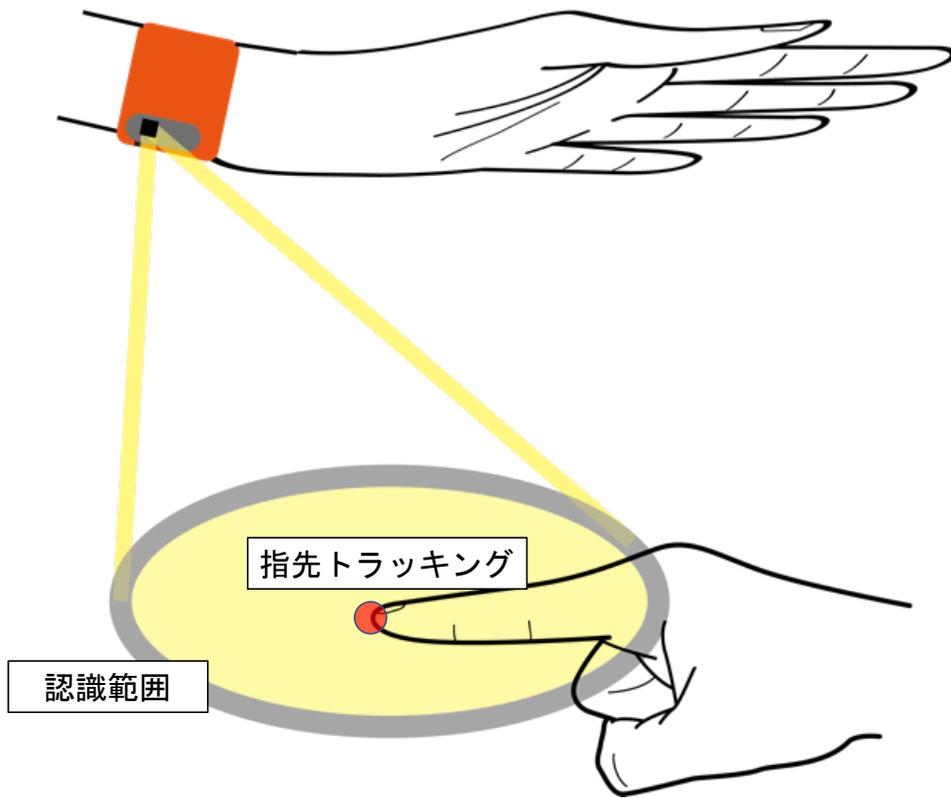


図 3.2 指先の動き利用概要図



図 3.3 手の位置利用概要図

## 第4章

# プロトタイプ

本章では、提案手法のプロトタイプ実装について述べる。プロトタイプは、赤外線センサアレイ、マイクロコントローラからなるセンシングデバイスと、センサからデータを取得し、前処理、手・指の認識を行う Android スマートフォン・ラップトップコンピュータ上にて動作する認識ソフトウェアからなる。図 4.1 にプロトタイプの概要を示す。



図 4.1 プロトタイプの概要

プロトタイプにおける認識の流れは以下のようにになっている。

1. マイクロコントローラと I<sup>2</sup>C 接続した赤外線センサアレイからセンサの値を取得する。
2. Bluetooth シリアル通信により、認識ソフトウェアが動作する Android スマートフォンあるいはラップトップコンピュータへデータを送信する。
3. データを基に、認識、トラッキングを行う。

## 4.1 センシングデバイス

プロトタイプ of センシングデバイス実装のために、赤外線アレイセンサとして Panasonic 社の AMG8833<sup>\*1</sup> (図 4.2), マイクロコントローラとして, M5Stack 社の M5StickC<sup>\*2</sup> (図 4.3) を採用した. 実装には, AMG8833 が搭載された評価基板, Conta サーモグラフィ<sup>\*3</sup>を用いている.

### 4.1.1 AMG8833

Panasonic 社の AMG8833 は,  $8 \times 8$  (64 画素) の赤外線アレイセンサである. 単眼焦電センサとは異なり, 2次元の低解像度赤外線画像を取得が可能であり, 内蔵されている演算 IC により, エリアの温度分布を出力する. AMG8833 のサイズは,  $8.0 \text{ mm} \times 11.6 \text{ mm} \times 4.3 \text{ mm}$  であり, 非常に小さいものとなっている. 特性として, 取得可能温度範囲は  $0^\circ\text{C} - 80^\circ\text{C}$ , 温度精度は  $\pm 2.5^\circ\text{C}$ , 分解能は  $0.0625^\circ\text{C}$ , 検知距離は 5 m 以内, 視野角は  $60^\circ$ , フレームレートは毎秒 10 フレーム, 消費電力は Typ.4.5 mA (収集時), Typ.0.8 mA (待機時) である. このセンサの対象となるものは人の手指であり, 人の体温が  $35^\circ\text{C} - 37^\circ\text{C}$  程度であることから, このセンサを手指の認識に用いることが可能であると考え. マイクロコントローラとの接続には, I<sup>2</sup>C 接続を用いている.

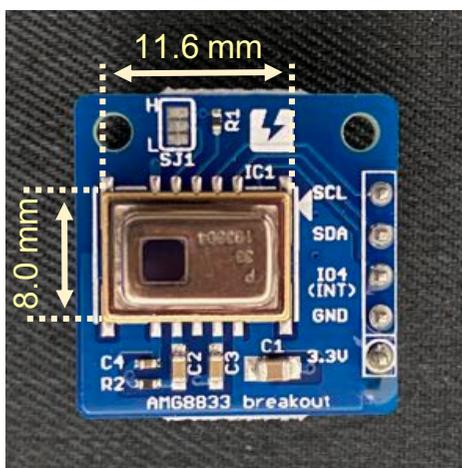


図 4.2 AMG8833 (Conta サーモグラフィ)

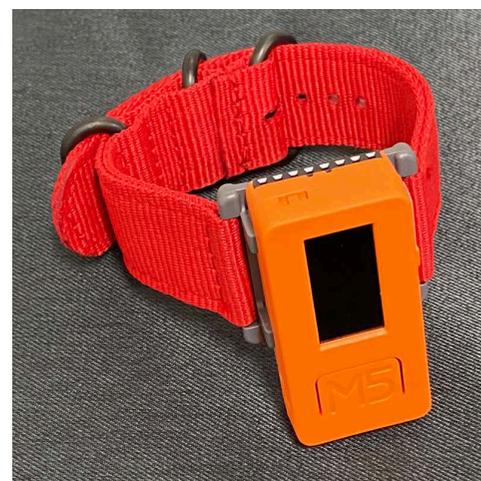


図 4.3 M5StickC

### 4.1.2 M5StickC

M5Stack 社の M5StickC は, ESP32 開発用ボードである. 搭載されているマイクロコントローラは, ESP32-PICO-D4 であり, 240 MHz デュアルコア, 520KB SRAM が内蔵されている. Wi-Fi モジュール, Bluetooth モジュール, IMU(MPU6886) が統合されており, 手首に装着できる腕時計マウンタ, GPIO ピンが搭載されているため, 今回のプロトタイプ実装に最適であると考え採用した.

\*1 <https://industrial.panasonic.com/jp/products/sensors/built-in-sensors/grid-eye>

\*2 <https://m5stack.com/collections/m5-core/products/stick-c>

\*3 <https://www.switch-science.com/catalog/3395/>

## IMU

M5StickC に統合されている IMU, MPU6886 は, 6 軸 IMU センサであり, 3 軸のジャイロスコープと 3 軸の加速度センサを含む. 6 軸 IMU により姿勢推定が可能となり, 赤外線アレイセンサの向いている方向を認識に用いることが可能となる.

### 4.1.3 実装

M5StickC を腕時計形状にて利用し, AMG8833 を I<sup>2</sup>C 接続する. M5StickC によって, AMG8833 を制御し, AMG8833 から得られる 8 × 8 の計 64 点の温度データを Bluetooth シリアル通信を用いて, Android スマートフォンもしくはラップトップコンピュータに送信する. この時, M5StickC に搭載されている MPU6886 から得られる 6 軸の加速度データも同時に送信を行う. センサが取得できる速度の規定値が 10 frame/秒 であることから, マイクロコントローラにて 1 フレーム分のデータ取得毎に 54 ミリ秒の待ち時間を挟みデータの送信を行うことにより, データ収集時に 1 フレームあたり約 100 ミリ秒, つまり 10 frame/秒 となるように調整した.

M5StickC に備わっているバッテリーが 80mAh であり, AMG8833 へ定格の 3.3V 出力にて連続的に動作させた時, 理論上 1.14 分間データ収集可能可能である. 実験においてバッテリー切れの可能性が存在したため, データ収集時には 1A 出力のモバイルバッテリーを接続している. 今回採用した M5StickC は概念実装用のマイクロコントローラであり, 今後別途スマートウォッチ等への統合を考えている.

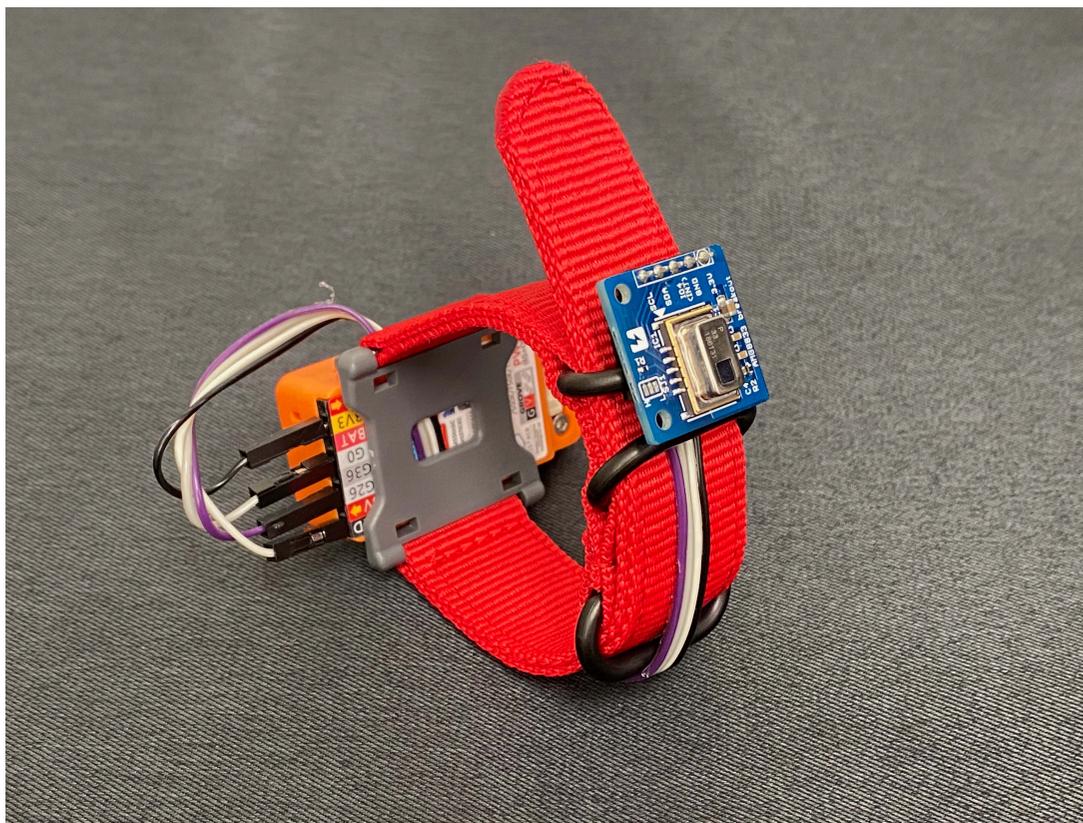


図 4.4 実装したハードウェア

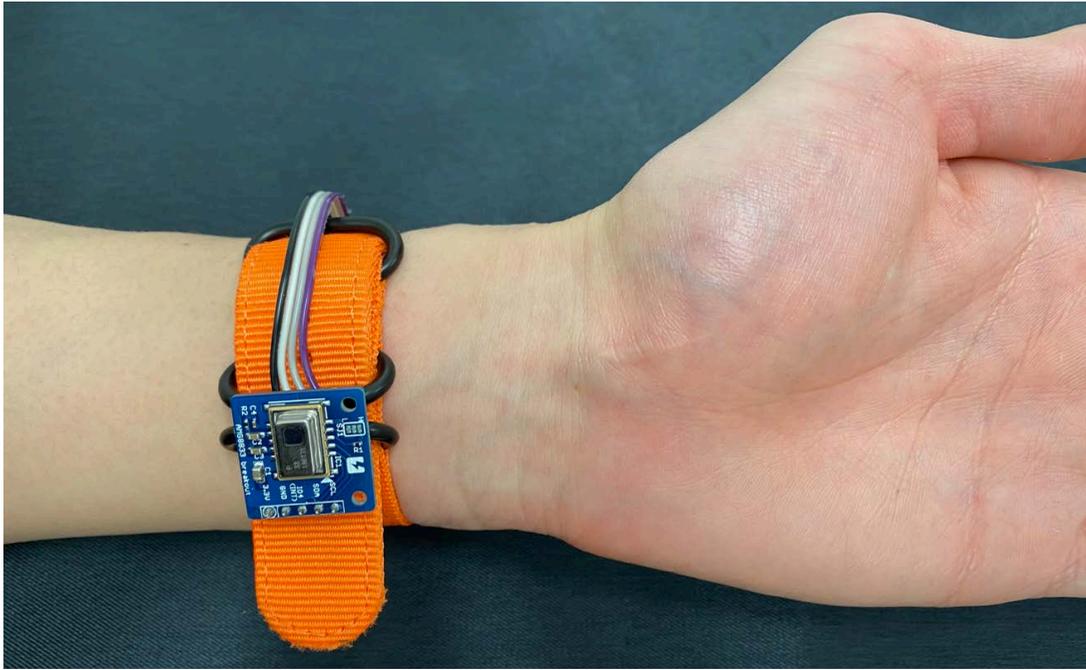


図 4.5 装着している様子

## 4.2 認識ソフトウェア

本節では、認識ソフトウェアの処理の流れ、認識器（ポーズ・ジェスチャ分類、指先の動きのトラッキング、手の位置の推定）の構築について述べる。以下に述べる認識器は Python によって、Android アプリケーションは Java/Kotlin によって開発した。機械学習用のライブラリとして、scikit-learn<sup>\*4</sup>、データ処理用のライブラリとして、SciPy<sup>\*5</sup>、シリアル通信用のライブラリとして、PySerial<sup>\*6</sup>を使用した。

### 4.2.1 処理の流れ

認識ソフトウェアの処理の流れを図 4.6 に示す。はじめに、マイクロコントローラを通し、Bluetooth シリアル通信によって、収集されたデータを受信する。受信したデータは温度データと加速度データであり、人の体温、外気温はユーザおよび環境ごとに異なるため、前処理として温度データを 0.0 から 1.0 の範囲へ正規化を行い認識器に渡す。

アプリケーションごとに処理が異なるが、ポーズ・ジェスチャ分類を行う場合は、分類器にデータを渡し、各ポーズもしくはジェスチャの分類を行う。指先の動きのトラッキングを行う場合は、指先検出器にデータを渡し、指先の位置を算出する。手の位置の推定を行う場合は、回帰分析器にデータを渡し、手とセンサの相対距離を算出する。それぞれの結果をアプリケーションに用いることによりインタラクションを可能とする。

---

\*4 <http://scikit-learn.org>

\*5 <https://www.scipy.org>

\*6 <https://pythonhosted.org/pyserial>

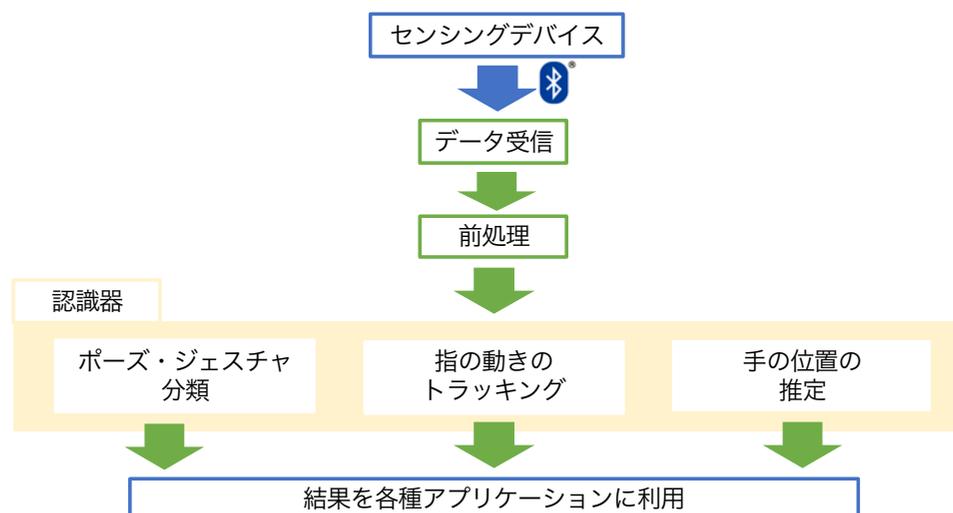


図 4.6 認識ソフトウェアの処理の流れ

#### 4.2.2 ポーズ・ジェスチャ分類

ハンドポーズ・ジェスチャの分類を行う分類器を構築した。ハンドポーズ・ジェスチャの分類は、機械学習の教師あり学習によって行う。機械学習のモデルは、実験により決定する。各モデルごとに、ハイパパラメータを最適化するために、収集した学習データに対して交差検証を用いたグリッドサーチを行う。グリッドサーチは、指定されたハイパパラメータの総当たりの組み合わせのモデルを構築し、最も性能の良いパラメータの組を決定する。ハイパパラメータの最適化後、最も精度の良かったモデルを分類に用いる。

#### 4.2.3 指先の動きのトラッキング

指先の位置の推定を行う回帰分析器を構築した。本実装においては、人差し指を立てた状態にて、その人差し指の指先の位置を推定する。機械学習のモデルは、サポートベクタマシン (SVM)、ランダムフォレストを利用する。分類器と同様にグリッドサーチによりハイパパラメータの最適化を行う。

#### 4.2.4 手の位置の推定

手とセンサの相対距離の回帰分析を行う回帰分析器を構築した。赤外線センサは、深度センサにも用いられていることから、距離により取得するデータに変化が生まれる。後述の実験の章に詳細は記述するが、予備実験を行い、センサと手の距離による取得温度データの差異を検証した。その結果、距離と温度に負の相関が存在し、距離をインタラクションに用いることが可能であることが確認された。さらに、手とセンサを近づけることにより、画像内における手の領域が変化することも推定に利用可能である。手とセンサの相対距離の回帰分析を行い、距離の推定を行う。

機械学習のモデルは、サポートベクタマシン (SVM)、ランダムフォレストを利用する。分類器と同様にグリッドサーチによりハイパパラメータの最適化を行う。

## 4.2.5 Android アプリケーション

ウェアラブルデバイスへの統合を考え、Android アプリケーションを作成した。マイクロコントローラとの通信は、Bluetooth シリアル通信を用いている。

図 4.7 は、データ収集時の Android アプリケーションを示している。画面内には、8 × 8 画素の温度情報を表すヒートマップ、6 軸加速度センサから得られる加速度情報とジャイロセンサ情報、取得するインスタンス数とフレーム数の表示、データ取得時の取得済みインスタンス数とフレーム数を示す進捗バーがある。ヒートマップは、低い温度ほど青色、高くなるにつれ、緑色、黄色、赤色へ変化していく。

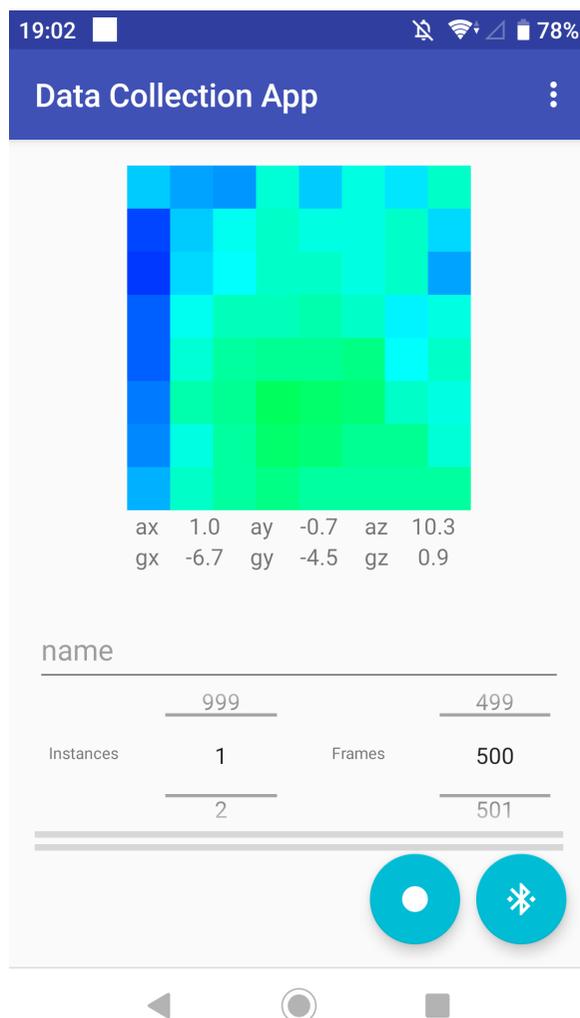


図 4.7 データ収集時の Android アプリケーション例

## 第 5 章

# 実験

### 5.1 予備実験

赤外線センサアレイがインタラクションに利用可能であることを検証するために予備実験を行った。

#### 5.1.1 予備実験 1：距離による取得温度データの差異

センサと対象の物体の距離の違いによる取得温度データの差異を検証することを目的とした予備実験を行った。被験者は 1 名（著者）、23 才、利き手は右である。実験は、屋内（明るい蛍光灯下）にて行われた。実験の手順は、図 5.1 のように、メジャーを軸として、センサと手の距離を、[0 cm, 5 cm, 10 cm, 15 cm, 20 cm, 30 cm, 40 cm] の位置へ手を開いた状態で推移させる。



図 5.1 予備実験 1 の実験の様子

各距離におけるデータを収集した。各距離 30 フレーム計測している。取得したデータの形式は、

$$(8 \times 8)_{pixels} \times 30_{frames} \times 7_{classes} \quad (5.1)$$

である。

全ての画像において、 $8 \times 8$  画素のうち、左から 4 つめ、上から 5 つめの画素に常に手の領域が存在したため、この点を本実験結果のデータとして用いた。データの可視化を行ったところ、図 5.2 となった。距離と温度データの相関係数を計算したところ、 $-0.967$  であった。

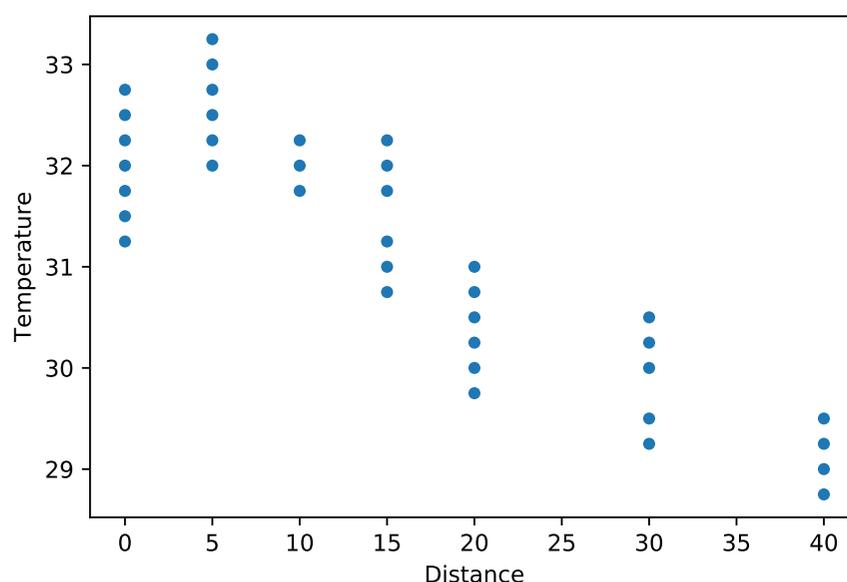


図 5.2 距離と温度データの可視化

結果から、センサと物体との距離は、取得する温度データに対して、強い負の相関を持つことがわかった。強い負の相関を持つということは、取得データが距離に強い影響を与えることを示しており、センサと物体の距離をインタラクションに利用できる可能性を示している。

### 5.1.2 予備実験 2：手のポーズ認識

#### 目的

赤外線センサアレイを用いて手指の認識が可能であるか検証するための予備実験として、ハンドポーズの分類が可能であるかを調査する。

#### 手順

対象とするハンドポーズは、ポーズは「グー」「チョキ」「パー」の 3 クラスと、より細かなポーズとして指を 1 本から 5 本たてた 5 クラスとした (図 5.3)。被験者は 1 名 (著者)、23 才、利き手は右である。実験は、屋内 (明るい蛍光灯下) にて行われた。実験器具として、作成したプロトタイプを利用する。データの取得方法は、センサから 10 cm 程度離れた位置にて、プロトタイプを装着している手とは反対の手を用いてハンドポーズを行い、1

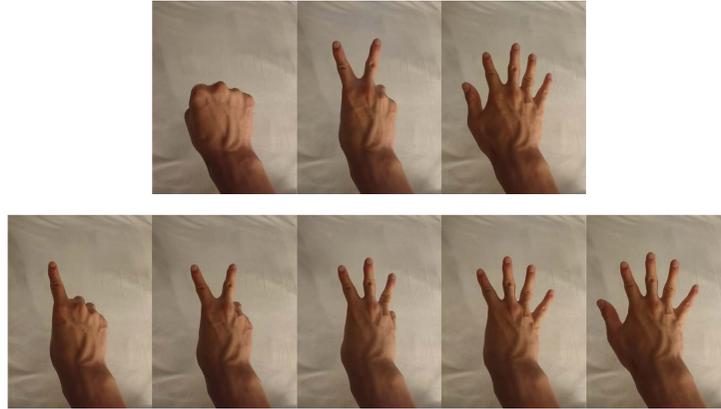


図 5.3 対象となるハンドポーズ (上: 3 クラス, 下: 5 クラス)

クラスあたり 100 フレームを取得する。取得したデータの形状は、

$$(8 \times 8)_{pixels} \times 100_{frames} \times (3 \text{ or } 5)_{classes} \quad (5.2)$$

である。

分類器として、Linear SVC を利用した。実装は Python を用いて行い、ライブラリは scikit-learn を利用している。本予備実験において、ハイパパラメータの最適化は行っていない。

## 結果

5-fold 交差検証を用いて、分類精度を検証したところ、3 クラス分類において平均 91.1%、5 クラス分類において平均 92.0% となった。各クラスの主成分分析の結果を図 5.4、図 5.6、分類結果の混同行列を図 5.5、図 5.7 に示す。「パー」に関しては誤認識がなく、「グー」「チョキ」の誤認識があった。指の本数の分類に関しても、近い本数の時に誤認識が存在した。これらから、1 クラスあたり 100 フレーム程度のデータでも高い分類精度を出すことができるが、画素が少ないため手の細かい変化の検知に少し不安があることがわかった。

### 5.1.3 予備実験 3: 手のジェスチャ認識

#### 目的

赤外線センサアレイを用いて手指の認識が可能であるか検証するための予備実験として、簡単なハンドジェスチャの分類が可能であるかを調査する。

#### 手順

対象とするハンドジェスチャは、上下左右 4 方向へのフリックジェスチャである。被験者は 1 名 (著者)、23 才、利き手は右である。実験は、屋内 (明るい蛍光灯下) にて行われた。実験器具として、作成したプロトタイプを利用する。データの取得方法は、センサから 10 cm 程度離れた位置にて、プロトタイプを装着している手とは

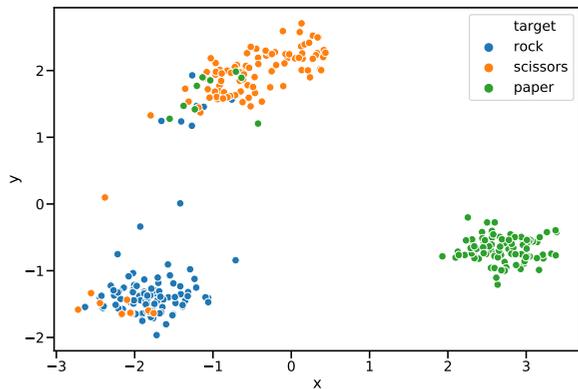


図 5.4 予備実験 2-1：主成分分析結果可視化

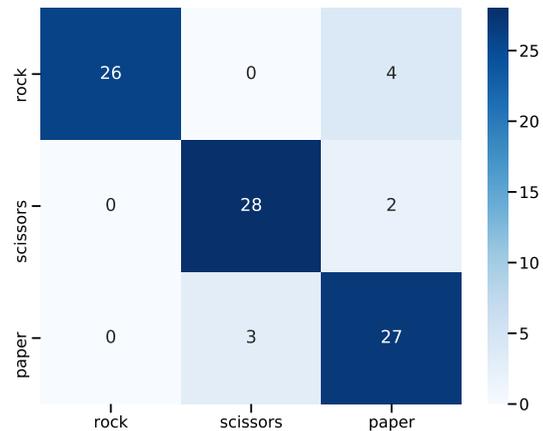


図 5.5 予備実験 2-1：混同行列

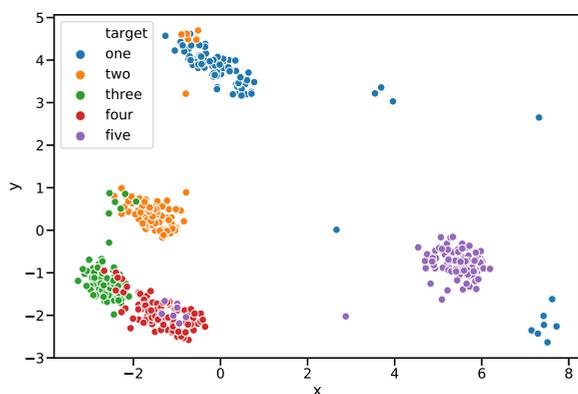


図 5.6 予備実験 2-2：主成分分析結果可視化

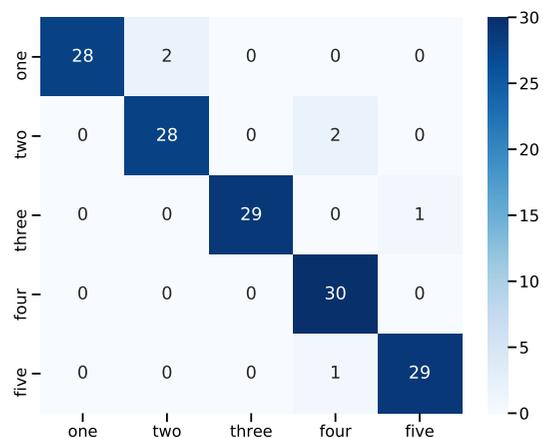


図 5.7 予備実験 2-2：混同行列

反対の手を用いてハンドジェスチャを行う。ハンドジェスチャの実行時間が約 1 秒程度と計測されたことから、1 ジェスチャを 10 フレームの連続画像とし、1 クラスあたり 30 インスタンス取得した。取得したデータの形状は、

$$(8 \times 8)_{pixels} \times 10_{frames} \times 30_{instances} \times 4_{classes} \quad (5.3)$$

である。

分類器として、Linear SVC を利用した。実装は Python を用いて行い、ライブラリは scikit-learn を利用している。本予備実験において、ハイパラメータの最適化は行っていない。

## 結果

5-fold 交差検証を用いて、分類精度を検証したところ、4 クラス分類において平均 83.3% となった。各クラスの主成分分析の結果を図 5.8、分類結果の混同行列を図 5.9 に示す。フリックジェスチャにおいて、動作として似

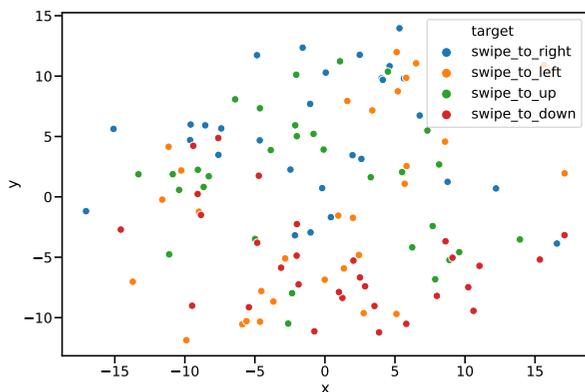


図 5.8 予備実験 3：主成分分析結果可視化

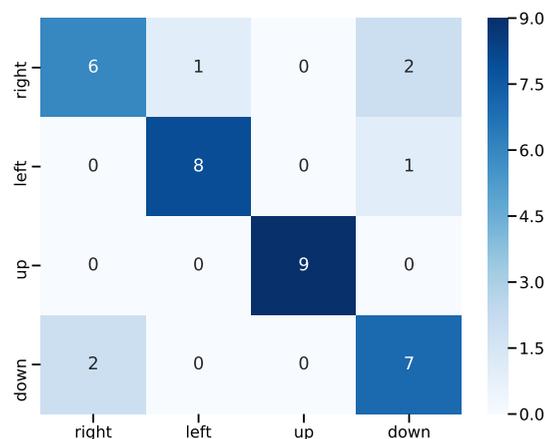


図 5.9 予備実験 3：混同行列

ている「右と下」「左と上」に誤認識が見られたが、得られた精度は低解像度赤外線アレイセンサがハンドジェスチャを認識することの実現可能性を示している。

予備実験 2, 3 において、分類器に線形分類モデルを採用した。しかしながら、可視化結果を参照すると、ポーズ認識においては線形分類可能な結果となっているがジェスチャ分類においては線形分類が困難な分布となっていることが見て取れる。そのため、分類器のモデルを変更することにより精度を向上させることができると考えられる。

## 5.2 実験 1：ハンドジェスチャ・ポーズ分類精度評価実験

### 5.2.1 概要

ハンドジェスチャ・ハンドポーズの識別精度を評価するための実験を行った。被験者は、5 名 (P1 から P5)、全員が男であり、平均年齢は、22.2 才、全員が右利きであった。実験は、屋内 (明るい蛍光灯下) にて行われた。識別精度を評価するためにジェスチャセットおよびポーズセットを用意した (図 5.11, 図 5.10)。

これらのターゲットセットは、EgoGestureDataset[29] から選択したものである。EgoGestureDataset とは、ハンドジェスチャ認識用のマルチモーダル大規模データセットであり、2081 の RGB-D ビデオ、24161 のジェスチャサンプル、および 50 名の異なる被験者からの 2953224 のフレームが含まれている。

ポーズセットは 7 種類の静的な手のポーズにて構成されている。“Number 1”から“Number 5”までは、指の立っている本数を表しており、“Number 0”は掌を丸めてまるを作る形とし、“Fist”は拳を握りしめた形状である。ジェスチャセットは 7 種類の動的な手のジェスチャにて構成されている。“Hold Hand”はセンサに掌を向けた状態にて静止する動作であり、“Push”はセンサに掌を向けた状態にてセンサに向けて掌を近づける動作である。“Phone Call”は親指と小指を立てた状態にて手首を小刻みに回す動作である。“Swipe ○○”は、人差し指と中指を立てた状態にて、センサの向きの垂直平面方向にて上下左右へと手を動かす動作である。

本実験において、機械学習のモデルは、サポートベクタマシン (SVM)、ランダムフォレスト (RF)、勾配ブースティング (GB) を使用した。本実験において、利用したそれぞれのモデルにおけるハイパラメータの候補を

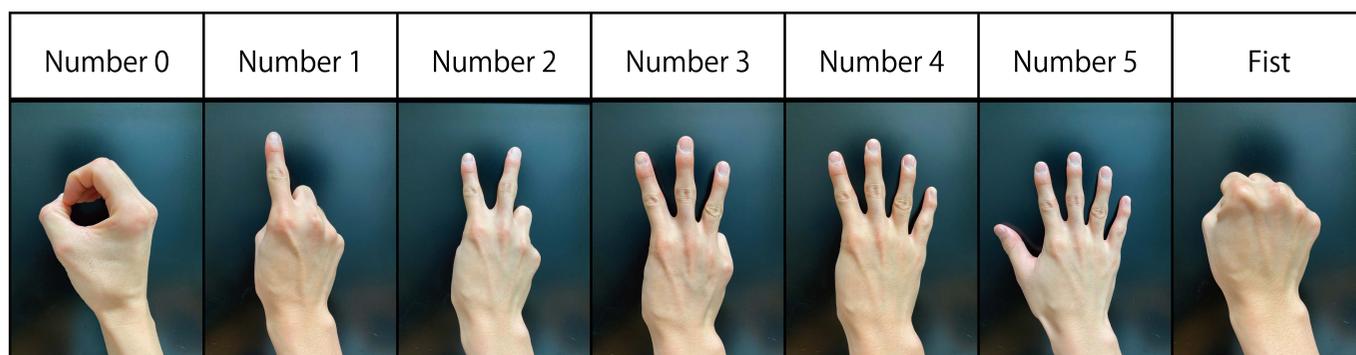


図 5.10 ポーズセット

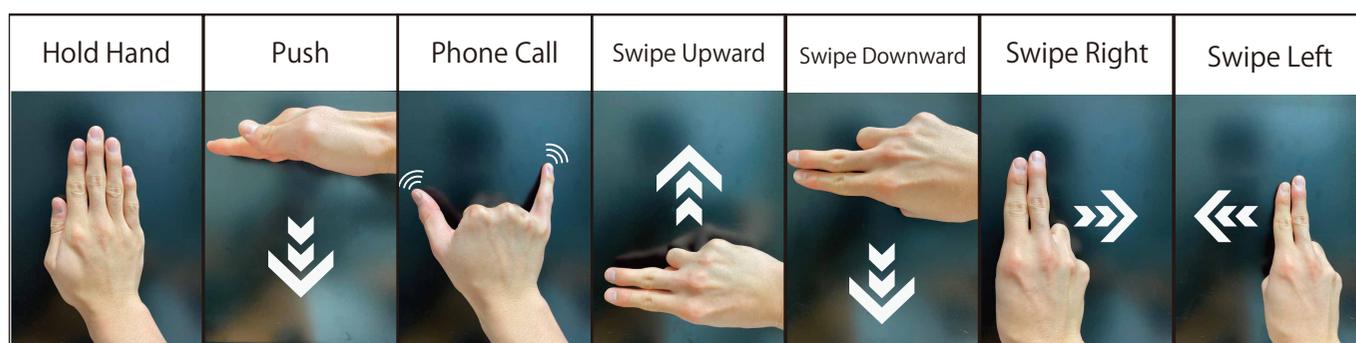


図 5.11 ジェスチャセット

表 5.1, 5.2, 5.3 に示す. 今回用いたパラメータは提案手法の実現可能性を確認するためのものであり, 再度調整することにより精度の向上が見られる可能性が存在する.

表 5.1 サポートベクタマシンハイパパラメータ候補

パラメータ	候補
C	0.1, 1, 10, 100, 1000
kernel	rbf, linear, poly, sigmoid
kernel 関数のパラメータ	$\theta = 1, 2, 3, 4, 5, 6, \gamma = 0.01, 0.12, 0.23, 0.34, 0.45, 0.56, 0.67, 0.78, 0.89, 1.00$

表 5.2 ランダムフォレストハイパパラメータ候補

パラメータ	候補
決定木の数	10, 20, 30, 40, 50, 60, 70, 80, 90
決定木の深さ	1,2,3,4,5,5,6,7,8,9,10,,11,12,13,14,15,16,17,18,19
分割の指標	gini, entropy

各カーネル関数は, 以下のように表され, それぞれの引数は表 5.1, 5.2, 5.3 に従う.

表 5.3 勾配ブースティングハイパパラメータ候補

パラメータ	候補
決定木の数	10, 20, 30, 40, 50, 60, 70, 80, 90
学習率	0.05,0.06,0.07, 0.08, 0.09,0.10,0.11,0.12,0.13,0.14,0.15,0.16,0.17,0.18,0.19,0.20

- 線形関数 (linear)

$$K(\mathbf{u}, \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} \quad (5.4)$$

- 放射基底関数 (rbf)

$$K(\mathbf{u}, \mathbf{v}) = \exp(-\gamma \|\mathbf{u} - \mathbf{v}\|^2) \quad (5.5)$$

- 多項式関数 (poly)

$$K(\mathbf{u}, \mathbf{v}) = (\alpha + \mathbf{u} \cdot \mathbf{v})^d \quad (5.6)$$

- シグモイド関数 (sigmoid)

$$K(\mathbf{u}, \mathbf{v}) = \tanh(\gamma \mathbf{u} \cdot \mathbf{v} + \theta) \quad (5.7)$$

## 5.2.2 手順

被験者には、以下のタスクを行うように指示した。

1. 被験者は、利き手とは反対の手首にセンシングデバイスを装着する。
2. データ収集
  - ポーズ
    - (a) 被験者は、指示されたポーズをセンサに向けて 50 秒 (500 フレーム) 行う。
    - (b) 実験者は、ポーズ実行時はセンサの範囲内において手を上下左右に動かすように指示する。
  - ジェスチャ
    - (a) 被験者は、指示されたジェスチャをセンサに向けて 5 秒 (30 フレーム) × 10 回 (インスタンス) 実行する。
3. 上記を 1 セッションとし、ポーズセットは各 5 セッション、ジェスチャセットは 3 セッション行う。

5 人の被験者がタスクを終了した時、

- ポーズ

$$500_{frames} \times 7_{classes} \times 5_{sessions} \times 5_{persons} \quad (5.8)$$

- ジェスチャ

$$30_{frames} \times 10_{instances} \times 7_{classes} \times 3_{sessions} \times 5_{persons} \quad (5.9)$$

のデータを取得した。

## 5.2.3 結果

取得したデータを用いて、検証を行った。

## セッション内精度

センシングデバイスが初めて装着され、学習が終了した時の性能を調査するために、それぞれのセッション内において、データの8割を訓練データとし、残りの2割をテストデータとした。取得したデータのうち温度データのみを利用し分類を行った。分類器のハイパパラメータは、訓練データについて、5-fold 交差検証を用いたグリッドサーチを用いたことにより決定された。それぞれのユーザ、対象セットにおける、平均識別精度を表5.4に示す。

表 5.4 セッション内の精度結果

ターゲットセット	P1	P2	P3	P4	P5	平均
ポーズ	99.99%	100.0%	98.70%	100.0%	99.34%	99.61% (SE=0.23)
ジェスチャ	97.09%	96.30%	94.19%	97.76%	92.50%	95.57% (SE=0.87)

代表して、P1 によるポーズ分類の混同行列、ジェスチャ分類の混同行列を図 5.12、図 5.13 に示す。

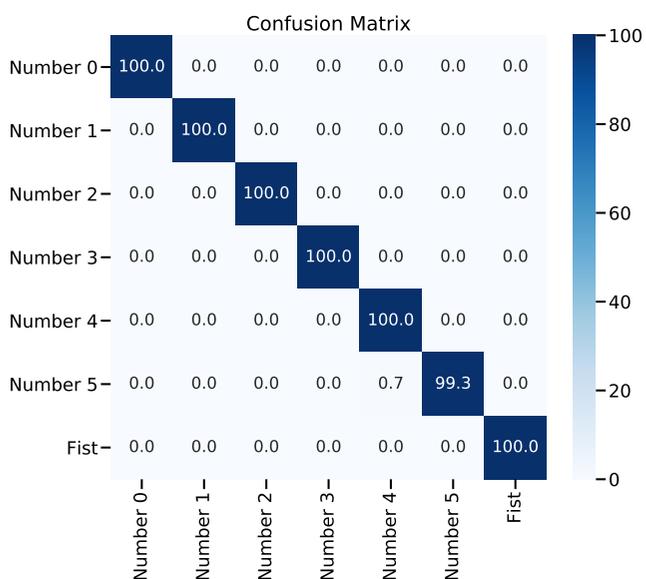


図 5.12 混同行列：ポーズ

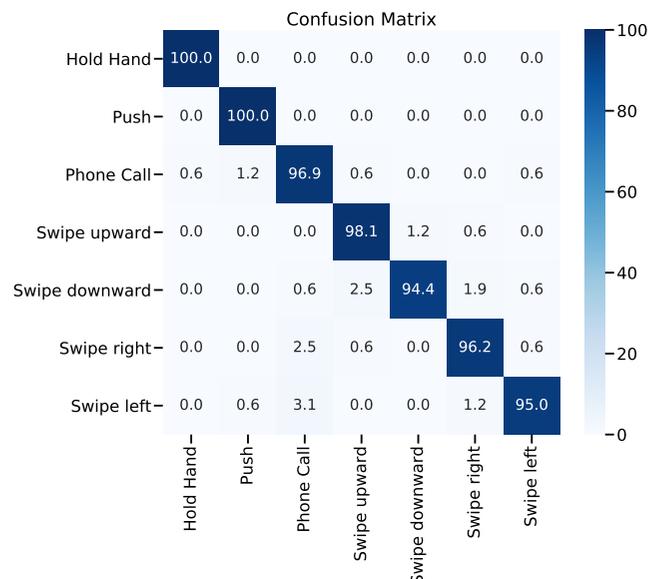


図 5.13 混同行列：ジェスチャ

ポーズ分類においては、ほとんどエラーは見られなかった。ジェスチャ分類については、4方向へのスワイプジェスチャにおいてエラーが多く見られた。収集したデータを確認したところ、スワイプの方向が斜めになっており、他の方向へと混同してしまうことや、ある方向へのフリックジェスチャを行ったのち手を元の位置に戻してしまい、逆方向のジェスチャを行ってしまっていることがわかった。これらは、データ収集時の指示を的確に行うことや方向のガイドを作成することにより改善できると考えられる。

今回の分類器の実装において、ジェスチャデータをそのまま用いており、ジェスチャを行っていないフレームも学習に用いていた。そのため、分類を行う前に、手の検出もしくはジェスチャ検出を行い、検出されているフレームのみを利用することによりさらなる精度の向上が見込まれる。

### leave-one-session-out 交差検証

ウェアラブルデバイスにおける認識の重要な課題として、再装着後も安定した性能を維持することがある [16]. 作成したプロトタイプの再装着後の性能の定価を評価するために各被験者に対して、3つ (5つ) のセッションのうち一つをテストデータ、残り4つを訓練データとし、全ての組み合わせの平均を算出する leave-one-session-out 交差検証 (LOSOCV) を行った. 各被験者の平均精度を表 5.5 に示す. 全ての被験者の平均精度は、ポーズ分類において 72.57% (SE=2.02), ジェスチャ分類において 75.39% (SE=3.50) であった. 混同行列を図 5.14, 図 5.15 に示す.

表 5.5 leave-one-session-out 交差検証結果

ターゲットセット	P1	P2	P3	P4	P5	平均
ポーズ	77.87%	71.56%	75.37 %	66.02 %	71.91 %	72.57% (SE=2.02)
ジェスチャ	63.99%	72.03%	80.07%	76.43%	84.41%	75.39% (SE=3.50)

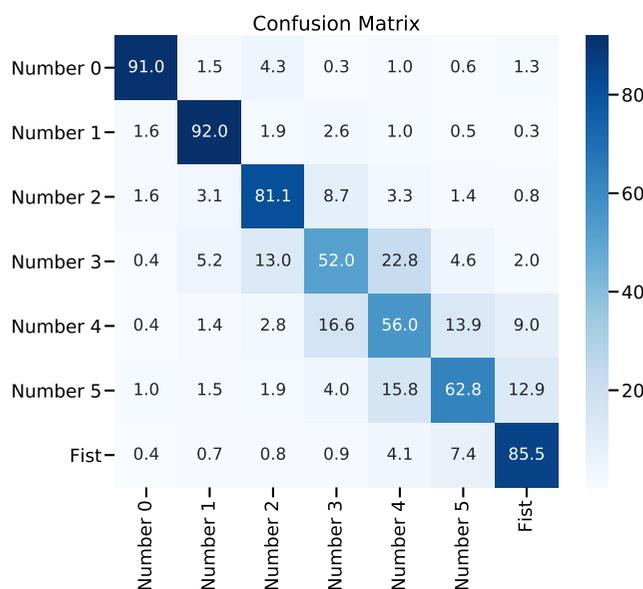


図 5.14 LOSOCV 混同行列：ポーズ

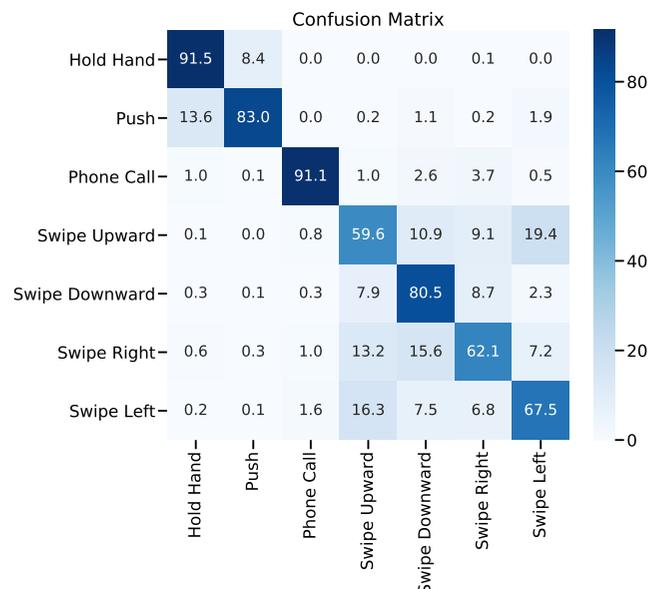


図 5.15 LOSOCV 混同行列：ジェスチャ

一部ポーズ・ジェスチャについて、精度が低くなっている要因としては、実行のたびに、ポーズ・ジェスチャを行う高さや方法が多少変化するためと考えられる. また、センシングデバイスを再装着後のセンサの位置が他のセッションとずれることがあり、分類結果に影響していると考えられる.

しかしながら、Number 0 や Number 1, Fist, Phone Call, Hold Hand といったポーズ・ジェスチャについては、90%, Number 2, Push, Swipe Downward についても 80% を超える精度を示しており、ポーズやジェスチャによりセッション間の差が少ないものが存在することを示している.

## leave-one-user-out 交差検証

ウェアラブルデバイスにおける認識の別の重要な課題として、ユーザごとに学習や調整なしに機能することがある [16]. この可能性を調査するために 5 人のうち一人の被験者のデータをテストデータとし、他の 4 人のデータを訓練データとし、全ての組み合わせの平均をとる leave-one-user-out 交差検証 (LOUOCV) を行った. それぞれの被験者をテストデータとした時の精度を表 5.6 に、混同行列を図 5.16, 図 5.17, 図 5.18, 図 5.19, 図 5.20, 図 5.21, 図 5.22, 図 5.23, 図 5.24, 図 5.25 に示す. 被験者全体の平均精度は、ポーズ分類において 64.69% (SE=7.15), ジェスチャ分類において 51.21% (SE=2.38) であった.

表 5.6 leave-one-user-out 交差検証結果

ターゲットセット/テストデータ	P1	P2	P3	P4	P5	平均
ポーズ	89.76%	51.57%	51.57%	60.36%	70.19%	64.69% (SE=7.15)
ジェスチャ	53.80%	47.15%	59.41%	48.34%	47.35%	51.21% (SE=2.38)

精度が低くなっている要因としては、ユーザにより手の形状が異なり、ポーズ・ジェスチャの実行方法が異なることを示唆している. しかしながら、Number 0 や Number 1, Fist, Phone Call といったポーズ・ジェスチャについては、ユーザによって 90% を超える精度を示しており、ポーズやジェスチャによりユーザ間の差が少ないものが存在することを示している. それらのジェスチャを探し、検証することによりユーザ間においても堅牢なポーズセット・ジェスチャセットを構築することが可能となると考えられる.

## 加速度データの利用

取得したデータのうち温度データと加速度データ両方を利用し分類を行った. 各ユーザのセッション内精度を検証した. 分類器のハイパパラメータは、訓練データについて、5-fold 交差検証を用いたグリッドサーチを用いたことにより決定された. それぞれのターゲットセットにおける、平均識別精度を表 5.7 に示す.

表 5.7 加速度を利用したセッション内の精度結果

ターゲットセット	P1	P2	P3	P4	P5	平均
ポーズ	99.68%	100.0%	97.01%	100.0%	99.35%	99.21% (SE=0.56)
ジェスチャ	97.29%	98.80%	96.29%	96.53%	94.31%	96.64% (SE=0.73)

温度データのみを利用した時と、温度データと加速度データ両方を利用した時を比較すると、ジェスチャ認識については精度の向上が見られた. ジェスチャは動的であり、加速度が取得できる動きに関する情報が学習に有効であったためと考えられる.

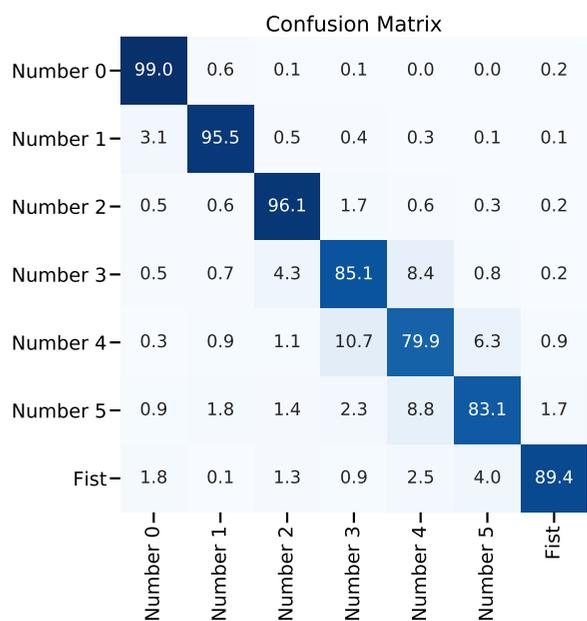


図 5.16 LOUCV 混同行列 (ポーズ : P1)

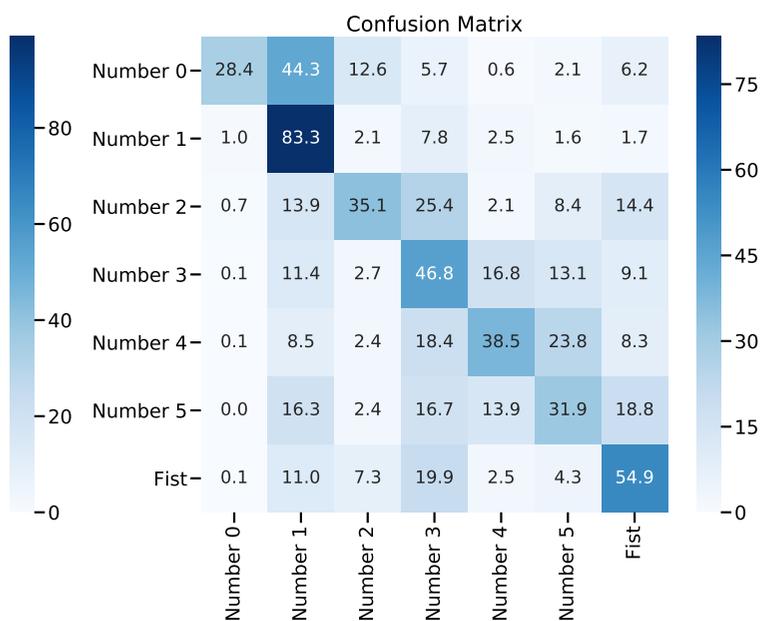


図 5.17 LOUCV 混同行列 (ポーズ : P2)

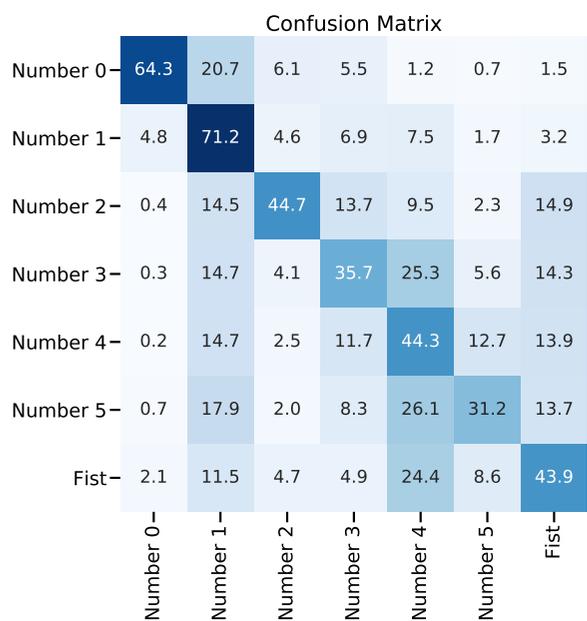


図 5.18 LOUCV 混同行列 (ポーズ : P3)

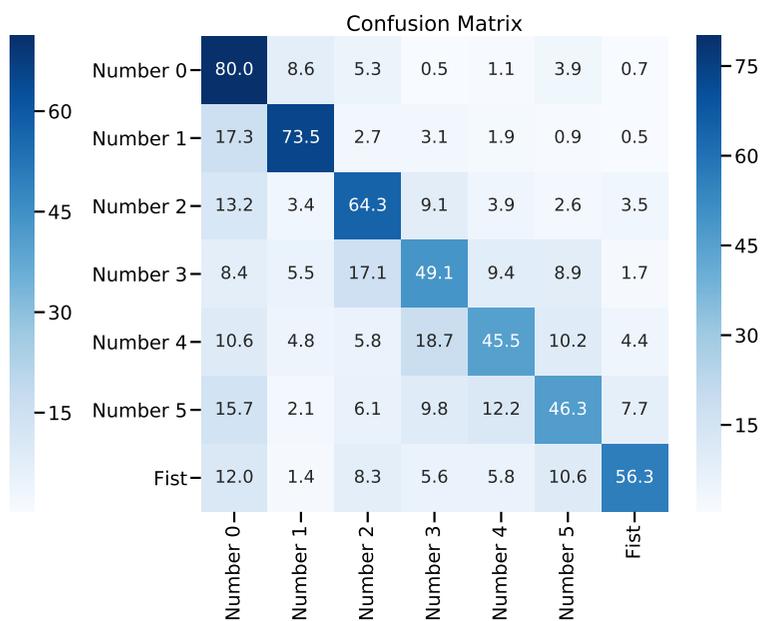


図 5.19 LOUCV 混同行列 (ポーズ : P4)

## 5.3 実験 2 : データ拡張による汎化性能評価実験

### 5.3.1 概要

機械学習を利用した認識には大きな課題がある。それは、大量の学習データ収集の必要性である。しかしながら、ハンドジェスチャ、ハンドポーズの大量の学習データ収集には、実行者に多大な労力と時間を要求する。学習データを十分に収集できなかった場合、分類精度の低下や過学習を引き起こす可能性が存在する。そこで、学

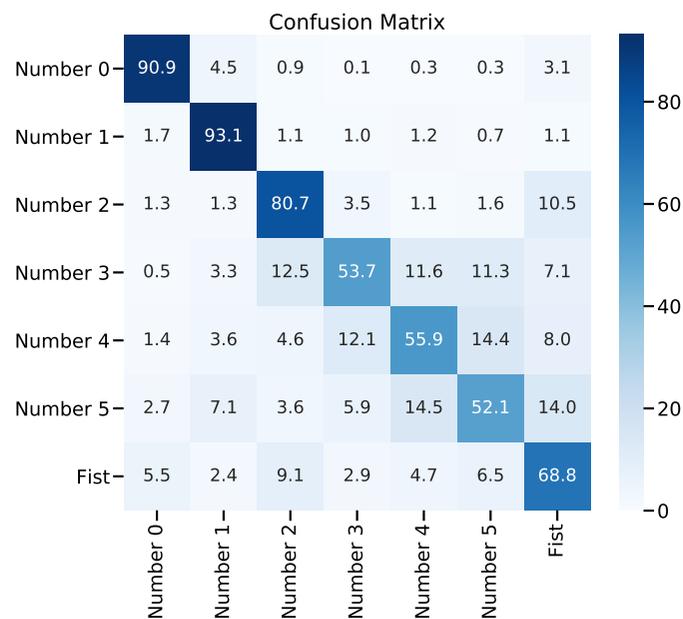


図 5.20 LOUCV 混同行列 (ポーズ : P5)

習データの拡張を行い，データ数を増やすことにより少ない学習データでも汎化性能を高めることができるのではないかと考えた。

### 5.3.2 手順

赤外線センサアレイにより取得したデータは， $8 \times 8$  画素の画像として捉えることが可能である．画像データの拡張手法として，画像の反転，拡大縮小，上下左右方向へのシフトといった画像変換が用いられる．

本実験において，画像の拡大縮小，上下左右方向へのシフト，回転を行い，実験 1 において収集した P1 のポーズデータ ( $500_{frames} \times 7_{classes} \times 5_{sessions} = 17500_{frames}$ ) をセッションごとに 10 倍の  $175000_{frames}$  へ拡張を行い，拡張後のデータを用いて，leave-one-session-out 交差検証を行った．画像の拡張には，Keras ライブラリ<sup>\*1</sup>を利用し，画像の拡大縮小，上下左右方向へのシフト，回転の範囲は，表 5.8 に示す．

表 5.8 画像変換範囲

変換方法	範囲
拡大縮小	$\times 1.0 \pm 0.4$
シフト	$\pm 20\%$
回転	$\pm 10^\circ$

\*1 <https://keras.io/>

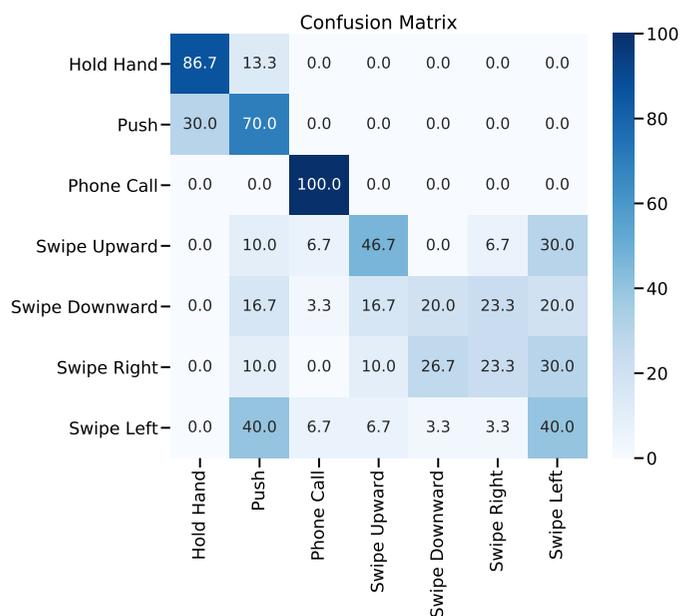


図 5.21 LOUOCV 混同行列 (ジェスチャ: P1)

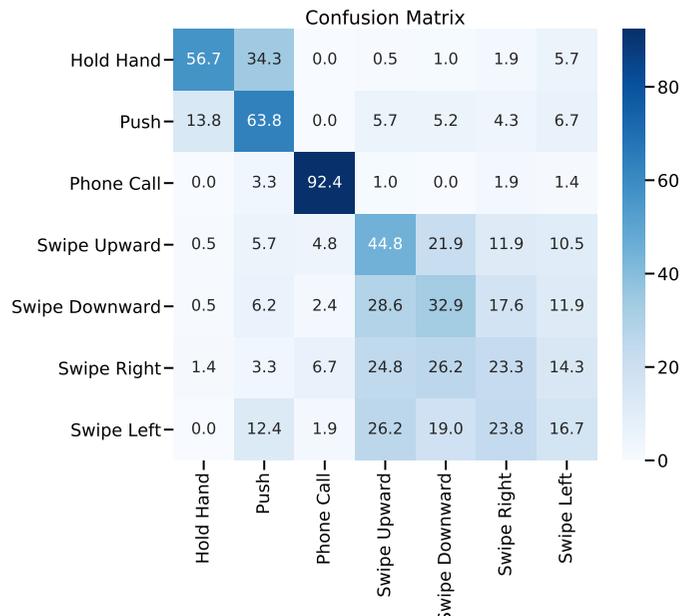


図 5.22 LOUOCV 混同行列 (ジェスチャ: P2)

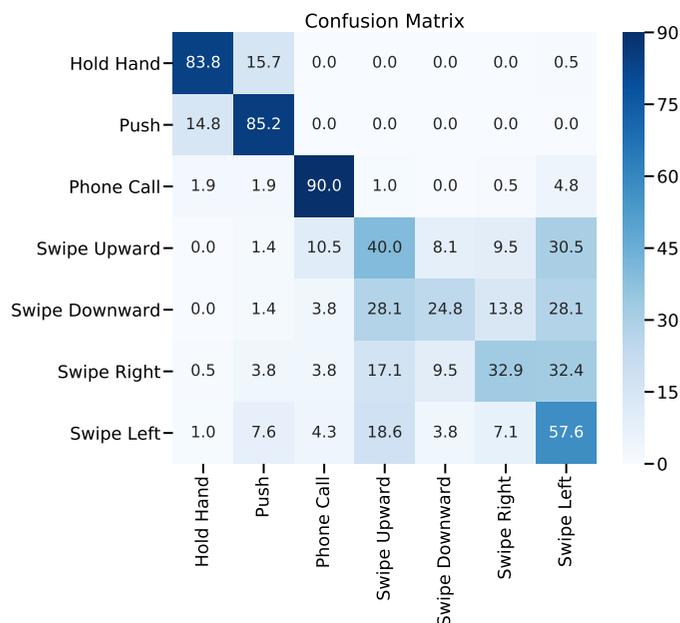


図 5.23 LOUOCV 混同行列 (ジェスチャ: P3)

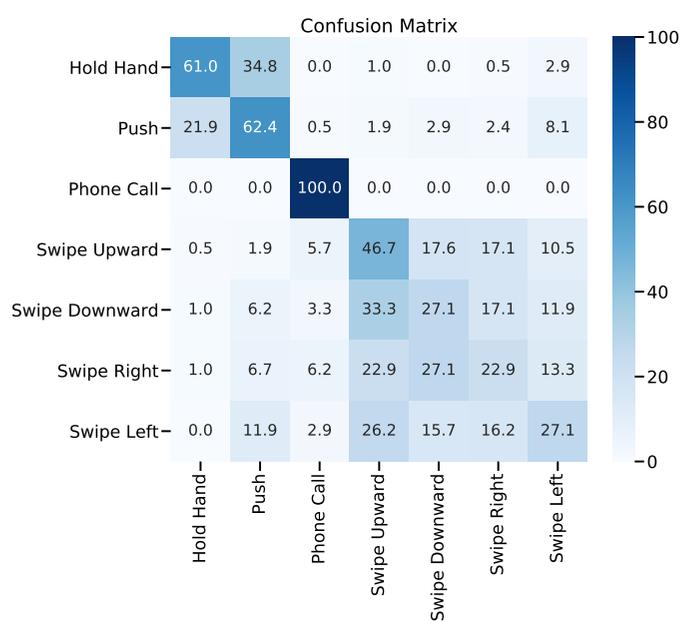


図 5.24 LOUOCV 混同行列 (ジェスチャ: P4)

### 5.3.3 結果

データ拡張を行ったデータを用いた分類において、85.49%の結果を得た。拡張を行わない手法と比較すると、12.92%高い。混同行列を図 5.26 に示す。leave-one-session-out 交差検証において、精度が低くなっていた Number 3, Number 4, Number 5 について顕著に分類結果が上昇していることが見てとれる。セッション間において、ユーザのポーズの行う位置が変化する、プロトタイプ再装着後にセンサの位置がずれるといった要因に対して、画像の拡大縮小、上下左右方向へのシフト、回転を行うデータ拡張手法が有効であることが示された。

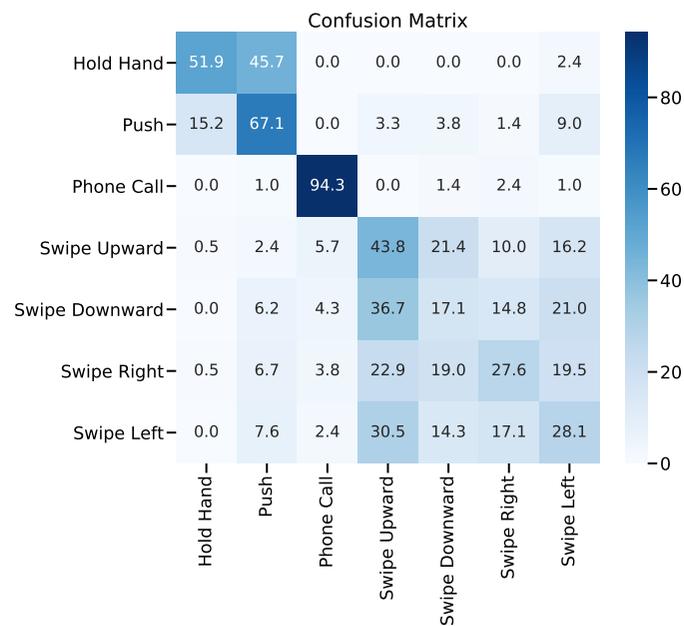


図 5.25 LOUOCV 混同行列 (ジェスチャ: P5)

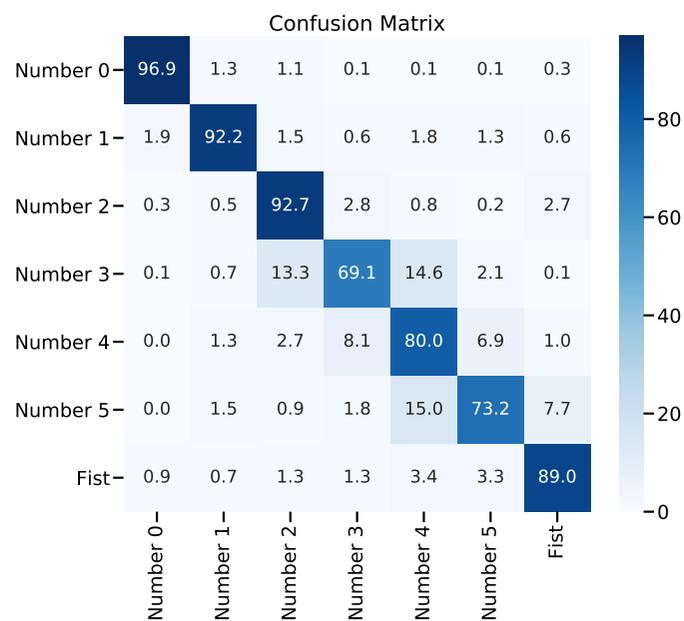


図 5.26 データ拡張 LOUOCV 混同行列

## 5.4 実験 3: 既存データセットを用いた認識精度評価実験

### 5.4.1 概要

実験 2 において、述べた問題点を解決するための別のアプローチとして、既存の公開データセットを用いて学習を行うことがある。既存の公開データセットを用いることにより、ユーザが面倒かつ時間を要するデー

タ収集を行うことなく、本提案手法を利用することができると考えられる。この可能性を検証するために、EgoGestureDataset[29] のデータを学習データとし、実験 1 において収集したデータをテストデータとして精度の評価を行った。

### 5.4.2 手順

EgoGestureDataset の対応するラベルを持つ深度画像を読み込み、それぞれの画像を  $8 \times 8$  画素へ中央の切り抜き、低解像度化を行い利用する。また、データセット内のデータは、白黒画像によって表現されており、取得したデータと値の範囲が異なるため、正規化を行っている。変換後のデータを基にモデルの学習およびハイパラメータの最適化を行い、実験 1 にて取得したデータのうち P1 のデータを用いて、各セッションごとにテストデータとして検証を行い、その平均値を算出した。EgoGestureDataset のデータの形状は、

$$(640 \times 480)_{pixels} \times 300_{frames} \times 7_{classes} \quad (5.10)$$

であり、変換後は

$$64_{pixels} \times 300_{frames} \times 7_{classes} \quad (5.11)$$

となる。

### 5.4.3 結果

検証の結果、精度は 24.36% となった。この時の混同行列を図 5.27 に示す。

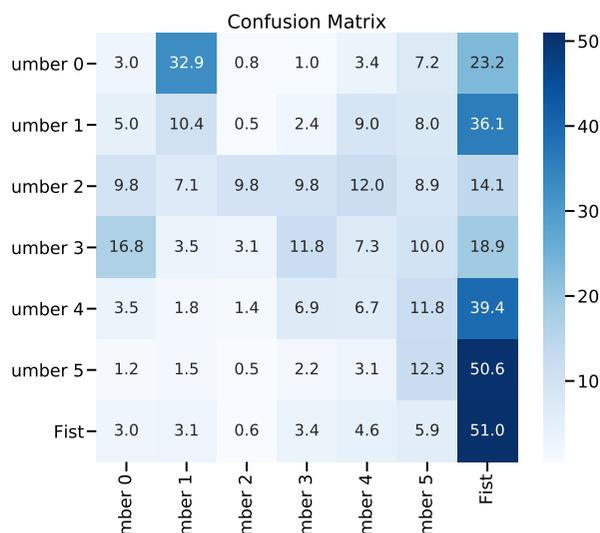


図 5.27 混同行列 (EgoGestureDataset による学習)

この結果は、今回の実装において、EgoGestureDataset のデータをポーズ分類に用いることが困難であることを示している。要因として、EgoGestureDataset の画像の取得方法の違いが挙げられる。EgoGestureDataset のデータは、頭の位置から手の形状を撮影したものであり、本提案手法における手首内側からの取得とは異なる。そのため、画像内における手の占める範囲や、角度が異なり分類に影響を与えたと考えられる。手の占める範囲が小

さいものは、低解像度化により指部分が潰れてしまい判別不能となってしまうものが存在した。今後、分類前に画像内における手の範囲の切り出し等を取り入れることにより EgoGestureDataset のデータを利用することができる可能性がある。

## 5.5 実験 4：他の入力モダリティにおける精度評価実験

### 指先の動きのトラッキング

#### 5.5.1 概要

指先の動きのトラッキングに用いる指先検出器の精度を評価することを目的とした実験を行った。被験者は 1 名（著者）、23 才、利き手は右である。実験は、屋内（明るい蛍光灯下）にて行われた。

#### 5.5.2 手順

人差し指を立てた状態にて、その指先の位置を推定する。テーブルから 10 cm 離れた位置にセンサを配置した時、センサの取得範囲は、およそ 10 cm × 10 cm となる。11 cm × 11 cm の方眼紙をテーブル上に配置し、真上から見たとき方眼紙の中央がセンサの位置となるように調整する。その後、各 1 cm 間隔の交点に指先を置き、センサ値を取得する。この時、それぞれの交点ごとに 25 フレーム取得した。取得したデータの形状は

$$100_{points} \times 25_{frames} \quad (5.12)$$

である。

#### 5.5.3 結果

それぞれのモデルにおける誤差の絶対値の平均を表 5.9 各交点ごとの誤差を図 5.28 に示す。ランダムフォレスト

表 5.9 画像変換範囲

モデル	誤差平均
SVM	0.74 cm
ランダムフォレスト	0.15 cm

トにおいて、最小の誤差 0.15 cm を達成した。図 5.28 により、誤差に偏りがあることが確認された。

センサ取得範囲の端においては、画像内における手の領域が非常に小さくなり、指先の位置が不明瞭になったために、精度が非常に低くなったと考えられる。この結果は、本手法において、指先トラッキングをインタラクションに用いることの実現可能性を示している。今後、ノイズ除去のフィルタリング処理、深層学習の導入を行うことにより精度の向上を目指す。

### 手の位置の推定

#### 5.5.4 概要

手とセンサの距離の回帰分析を行う回帰分析器の性能を評価することを目的とする精度評価実験を行った。

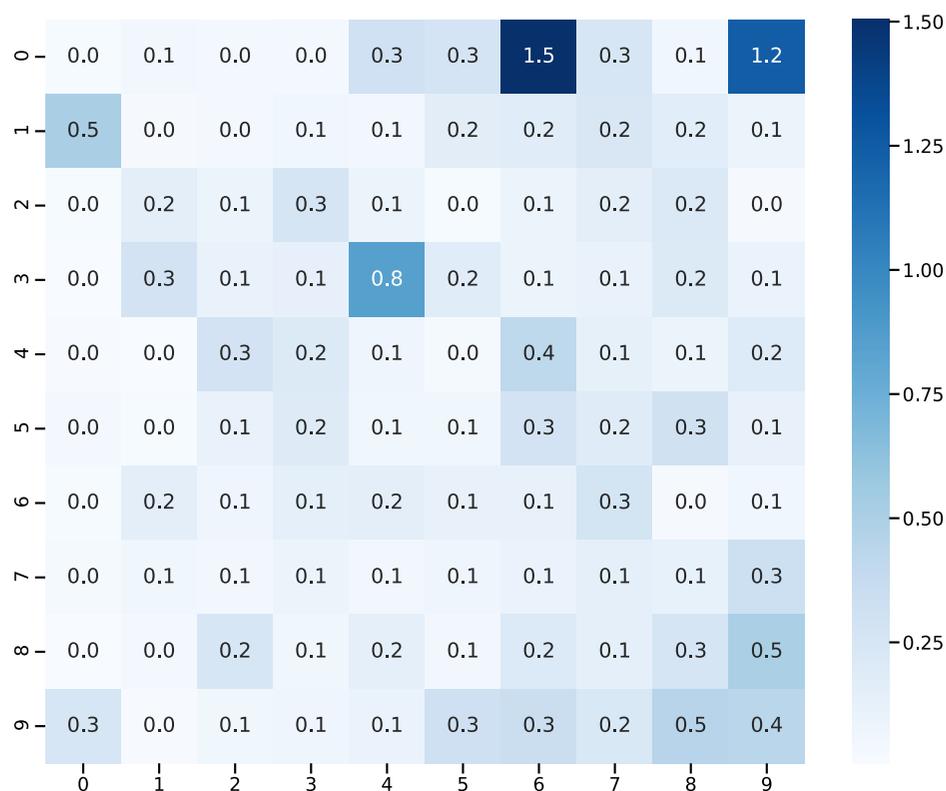


図 5.28 各交点ごとの誤差の平均 (cm)

### 5.5.5 手順

予備実験 1 にて、収集したデータを利用する。収集したデータを学習データ 8 割、テストデータ 2 割に分割し、その学習データを用いて、回帰分析器の学習およびハイパラメータの最適化を行う。その後テストデータをもちいて、誤差の評価を行った。

### 5.5.6 結果

それぞれのモデルにおける誤差の絶対値の平均を表 5.10

表 5.10 画像変換範囲

モデル	誤差平均
SVM	1.821 cm
ランダムフォレスト	2.052 cm

SVM において、最小の誤差 1.821 cm を達成した。利用アプリケーションとして、連続値を手の推移によって操作することを挙げており、今回の結果は十分な精度に達していると考えられる。

## 第6章

# アプリケーション

本章では、提案手法を用いたアプリケーション例について述べる。

### 6.1 スマートウォッチのジェスチャ入力拡張

スマートウォッチ等のウェアラブルデバイスは、小型であることが求められ、搭載されている画面が小さくなることからタッチによる入力およびその操作はユーザに負担がかかる。これは、指よりも小さなオブジェクトは押し分けが困難であるというファットフィンガー問題 [30] として、提起されている。そこで、タッチ入力の代換として本提案手法のハンドポーズ・ハンドジェスチャによる入力を行うアプリケーションを紹介する。スマートウォッチの操作として、画面のフリックによる画面の遷移または移動が存在し、代替ジェスチャとして、手の空中フリックジェスチャを取り入れる。また、選択には手をセンサに近づける“Push”動作を取り入れる。これにより、画面を隠すことなくスマートウォッチの操作が可能となる。さらには、“Call”ジェスチャを行うことにより、電話モードを起動し、0から5本の指のポーズを認識することにより、番号の入力を行い、短縮ダイヤルに電話をかけることも可能となる。利用イメージを図6.1に示す。

### 6.2 指先トラッキングを利用した手書きメモアプリケーション

スマートウォッチのアプリケーションとして、指先の動きのトラッキングを利用した手書きメモアプリケーションを紹介する。前述のようにスマートウォッチに搭載されている画面は小さく文字を入力するには不十分である。そこで、赤外線センサアレイによる指先トラッキングを利用し、手書き入力を行うことを考えた。利用イメージを図6.2に示す。赤外線センサアレイによるトラッキング範囲は、スマートウォッチ画面よりも大きく、ユーザに負担を強いることがない。また、ウェアラブルデバイスに適用していることにより、いつでもどこでも、特別なデバイスを取り出すことなく、手書きのメモが利用可能となる。

### 6.3 手とセンサの相対距離を用いた連続値操作

様々なアプリケーションにおいて、連続値を操作する動作は行われている。例えば、音量操作や音楽の再生位置の調整、ズームインアウト等がある。今回は、音楽再生に着目したアプリケーションを紹介する。音量操作と再生位置の調整それぞれにジェスチャを割り当て、その後の手のセンサの相対距離に応じて対応する値を変化させる。これにより、ユーザはスマートウォッチの画面を用いて入力することや、別のデバイスのある場所に行き

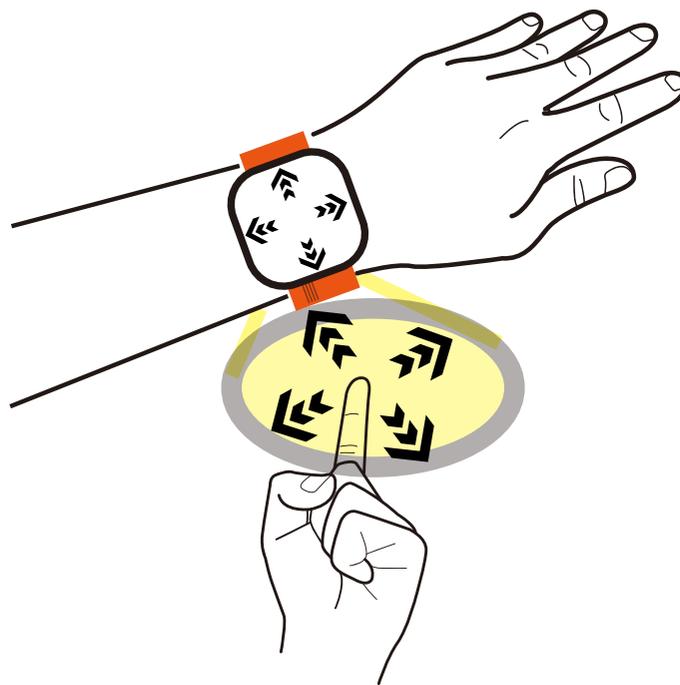


図 6.1 スマートウォッチアプリケーション例

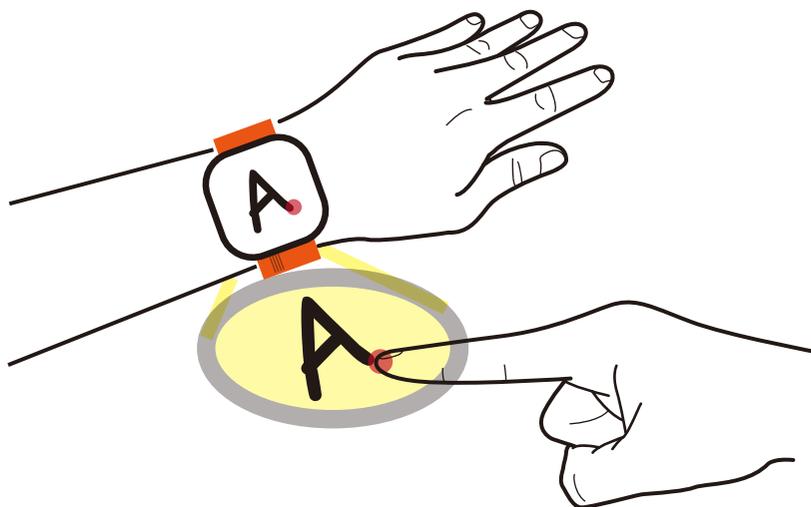


図 6.2 メモアプリケーション例

操作する必要がなくなり，ユーザは煩わしい動作から解放される。

## 第7章

# 結論

本研究では、低解像度の赤外線センサアレイを手首内側に装着し、センサから取得したデータを機械学習にかけて手指を認識する手法を提案した。本提案手法により、プライバシー侵害を軽減しながら、腕時計型端末の手首内側に装着したセンサによりハンドジェスチャ認識を可能とし、多様なインタラクションを実現することが可能となる。赤外線センサアレイおよびマイクロコントローラである M5StickC を使用し、プロトタイプの実装を行った。また、Python により、ジェスチャ・ポーズ分類・指先トラッキング・センサと手の距離の回帰分析を行う認識ソフトウェアを開発した。プロトタイプの性能を評価するために被験者実験を行った。ジェスチャ・ポーズ分類において、セッション内では 95.29% (ジェスチャ), 99.29% (ポーズ), セッション間では 75.39% (ジェスチャ), 72.57% (ポーズ) の精度を達成し、指先トラッキングにおいて 0.15 cm の誤差, 回帰分析において 1.821 cm の誤差となった。また、機械学習のデータ拡張手法を用いることにより、ポーズ分類において、セッション間精度を 85.49% へと向上させた。アプリケーション例として、ジェスチャ・ポーズ分類によるスマートウォッチの操作拡張と、指先トラッキングを用いた手書きメモアプリケーション、センサと手の距離による連続値操作の適用を説明した。今回利用したセンサは温度データを取得可能であり、温度データを直接利用したインタラクションについても考えていきたい。

今後、本提案手法を適用したアプリケーションの使用感を調査する。今回作成したプロトタイプは、概念実証のために手首に装着したマイクロコントローラによって制御を行っていた。本手法は、スマートウォッチ等の腕時計型ウェアラブルデバイスへの統合を想定しており、今後スマートウォッチを用いたプロトタイプを作成したいと考えている。ハンドジェスチャ・ハンドポーズ認識において、先行研究 [9, 1, 17] では深層学習が多く用いられており、分類器を深層学習を用いたモデルに変更することによりさらなる精度の向上が見込まれる。また、今回加速度と温度データ両方を利用しており、異なるデータごとに別の分類器を用いて分類を行い両方の結果を統合して認識結果とする手法も存在する [9] ため、考慮に入れていきたい。

# 謝辞

本論文の執筆にあたり，指導教員である高橋伸准教授，志築文太郎准教授には熱心なご指導を頂き，深く感謝を申し上げます．特に高橋伸准教授には研究を進める上で，方針や内容について多くのご指導を賜りました．心よりお礼申し上げます．また，インタラクティブプログラミング研究室の皆様には，研究および研究生活に関する様々な助言を頂きました．特に UBIQUITOUS チームの皆様にはチームゼミをはじめ，論文執筆時の添削など多くの支援を頂き，深く感謝を申し上げます．最後に，研究生活を支えてくださった家族，友人に感謝を申し上げます．

## 参考文献

- [1] Hongyi Wen, Julian Ramos Rojas, and Anind K. Dey. Serendipity: Finger gesture recognition using an off-the-shelf smartwatch. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 3847–3851, New York, NY, USA, 2016. Association for Computing Machinery.
- [2] Yichen Li, Tianxing Li, Ruchir A. Patel, Xing-Dong Yang, and Xia Zhou. Self-powered gesture recognition with ambient light. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, pp. 595–608, New York, NY, USA, 2018. Association for Computing Machinery.
- [3] J. Kim, J. He, K. Lyons, and T. Starner. The gesture watch: A wireless contact-free gesture based wrist interface. In *2007 11th IEEE International Symposium on Wearable Computers*, pp. 15–22, Oct 2007.
- [4] Chao Xu, Parth H. Pathak, and Prasant Mohapatra. Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile '15, pp. 9–14, New York, NY, USA, 2015. Association for Computing Machinery.
- [5] Robert Xiao, Teng Cao, Ning Guo, Jun Zhuo, Yang Zhang, and Chris Harrison. Lumiwatch: On-arm projected graphics and touch input. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, New York, NY, USA, 2018. Association for Computing Machinery.
- [6] Gierad Laput and Chris Harrison. Sensing fine-grained hand activity with smartwatches. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [7] K. Ohnishi, A. Kanehira, A. Kanezaki, and T. Harada. Recognizing activities of daily living with a wrist-mounted camera. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3103–3111, June 2016.
- [8] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pp. 167–176, New York, NY, USA, 2012. Association for Computing Machinery.
- [9] Hui-Shyong Yeo, Erwin Wu, Juyoung Lee, Aaron Quigley, and Hideki Koike. Opisthenar: Hand poses and finger tapping recognition by observing back of hand using embedded wrist camera. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, UIST '19, pp. 963–971, New York, NY, USA, 2019. Association for Computing Machinery.
- [10] Manuel Prätorius, Dimitar Valkov, Ulrich Burgbacher, and Klaus Hinrichs. Digitap: An eyes-free vr/ar symbolic input device. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, VRST

- '14, pp. 9–18, New York, NY, USA, 2014. Association for Computing Machinery.
- [11] Feiyu Chen, Jia Deng, Zhibo Pang, Majid Baghaei Nejad, Huayong Yang, and Geng Yang. Finger angle-based hand gesture recognition for smart infrastructure using wearable wrist-worn camera. *Applied Sciences*, Vol. 8, No. 3, 2018.
- [12] Andrew Vardy, John Robinson, and Li-Te Cheng. The wristcam as input device. In *Proceedings of the 3rd IEEE International Symposium on Wearable Computers*, ISWC '99, p. 199, USA, 1999. IEEE Computer Society.
- [13] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. Watchsense: On- and above-skin input sensing through a wearable depth sensor. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 3891–3902, New York, NY, USA, 2017. Association for Computing Machinery.
- [14] Oliver Glauser, Shihao Wu, Daniele Panozzo, Otmar Hilliges, and Olga Sorkine-Hornung. Interactive hand pose estimation using a stretch-sensing soft glove. *ACM Trans. Graph.*, Vol. 38, No. 4, July 2019.
- [15] Artem Dementyev and Joseph A. Paradiso. Wristflex: Low-power gesture input with wrist-worn pressure sensors. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pp. 161–166, New York, NY, USA, 2014. Association for Computing Machinery.
- [16] Yasha Iravantchi, Mayank Goel, and Chris Harrison. Beamband: Hand gesture sensing with ultrasonic beam-forming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [17] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pp. 851–860, New York, NY, USA, 2016. Association for Computing Machinery.
- [18] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pp. 167–176, New York, NY, USA, 2009. Association for Computing Machinery.
- [19] Yu Zhang, Tao Gu, Chu Luo, Vassilis Kostakos, and Aruna Seneviratne. Findroidhr: Smartwatch gesture input with optical heartrate monitor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 2, No. 1, March 2018.
- [20] Gierad Laput, Robert Xiao, and Chris Harrison. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pp. 321–333, New York, NY, USA, 2016. Association for Computing Machinery.
- [21] Yang Zhang and Chris Harrison. Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, pp. 167–173, New York, NY, USA, 2015. Association for Computing Machinery.
- [22] Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. Pyro: Thumb-tip gesture recognition using pyro-electric infrared sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, pp. 553–563, New York, NY, USA, 2017. Association for Computing Machinery.
- [23] Piotr Wojtczuk, Alistair Armitage, David Binnie, and Timothy Chamberlain. Recognition of simple gestures

- using a pir sensor array. *Sensors & transducers*, Vol. 14, pp. 83–94, 2011.
- [24] Andrew Clark and Deshendran Moodley. A system for a hand gesture-manipulated virtual reality environment. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists*, SAICSIT '16, New York, NY, USA, 2016. Association for Computing Machinery.
- [25] Hui Liang, Junsong Yuan, Daniel Thalmann, and Nadia Magnenat Thalmann. Ar in hand: Egocentric palm pose tracking and gesture recognition for augmented reality applications. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pp. 743–744, New York, NY, USA, 2015. Association for Computing Machinery.
- [26] Theophilus Teo, Gun A. Lee, Mark Billingham, and Matt Adcock. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, OzCHI '18, pp. 406–410, New York, NY, USA, 2018. Association for Computing Machinery.
- [27] Pascal Chiu, Kazuki Takashima, Kazuyuki Fujita, and Yoshifumi Kitamura. Pursuit sensing: Extending hand tracking space in mobile vr applications. In *Symposium on Spatial User Interaction*, SUI '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [28] Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. Brainy hand: An ear-worn hand gesture interaction device. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pp. 4255–4260, New York, NY, USA, 2009. Association for Computing Machinery.
- [29] Y. Zhang, C. Cao, J. Cheng, and H. Lu. Egogesture: A new dataset and benchmark for egocentric hand gesture recognition. *IEEE Transactions on Multimedia*, Vol. 20, No. 5, pp. 1038–1050, May 2018.
- [30] Katie A. Siek, Yvonne Rogers, and Kay H. Connelly. Fat finger worries: How older and younger users physically interact with pdas. In *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*, INTERACT'05, pp. 267–280, Berlin, Heidelberg, 2005. Springer-Verlag.