スマートフォンの位置変化の影響を考慮した両足ジェスチャ認識手法

田村 柾優紀† 中村 聡史‡

†明治大学先端数理科学研究科 〒164-8525 東京都中野区中野 4-21-1
‡明治大学総合数理学部 〒164-8525 東京都中野区中野 4-21-1
E-mail: †mogamusa31@gmail.com, ‡satoshi@snakamura.org

あらまし 我々はこれまで目の前の電子機器を操作する際に手や声による操作ができない状況を考慮し，片方のポケットにスマートフォンを入れた状態で両足のジェスチャを認識する手法を提案してきた．しかし，プロトタイプシステムを使用した後に回答してもらった体感認識精度が，データセットを用いて算出した認識精度から悪化するという問題があった．そこで，我々はその問題を解決するために認識までにかかる時間を短縮し，位置変化が与える影響を考慮した手法の検討を行う．具体的には，認識時間の短縮では，望ましいシステム応答時間内のセンシングデータを用いてジェスチャを認識する．使用するフレームを制限することで，認識精度が低下することが考えられるため，ジェスチャを開始する直前のジェスチャデータを使用することで認識精度が向上するかを検証した．その結果，予備動作のセンシングデータを利用することで認識精度が向上する傾向にあった．位置変化が与える影響を考慮した認識では，ポケット内のスマートフォンの傾きを特徴量として使用することで，位置変化による影響を低減することが出来るかを検証する．

キーワード ジェスチャ認識，両足，予備動作，位置変化，デバイス操作，機械学習

1. はじめに

2019 年現在では，目の前にあるスマートフォンや，タブレット，テレビ，照明などのデバイスを操作する方法として，主に手による操作と，音声認識を用いた声による操作が考えられる．しかし，一度に操作をするわけではないが，同じ操作が何回も必要となる際には音声認識での操作は手間である．また音声認識は，タブレット端末に楽譜を表示し楽器を演奏する状況や，子供を抱っこして寝かしつけている際にデバイスを操作したい状況，電車に乗っている際にデバイスを操作したい状況など，声を発することが困難な状況で音声認識によるデバイス操作が向かない状況が存在する．

手による操作と音声認識による操作を用いずにデバイス操作を行う方法の一つとして，足によるジェスチャを用いたデバイス操作の研究が様々行われている[1][2]．これらの研究は，事前に中敷や靴下にセンサを取り付ける必要があるため，導入に手間がかかる．これに対し，新たにセンサを取り付けずに，身近なデバイスであるスマートフォンを用いて，足のジェスチャ認識を行う研究には，Scott らの研究[5]があげられる．その研究では，ポケット内のスマートフォンを用いて，スマートフォンを入れた側の足によるジェスチャ認識精度を検証している．しかし，前ポケットでの認識精度は 60.2%であり，認識精度は高くない．また，スマートフォンは日常において，様々な角度や向きでポケット内に入れられることがある．そのため，ジェスチャのデータセット構築時とは違う角度や向きでポケットに入れられた際には，センサの軸の向きが変化し，認識精度に影響が出ることが考えられるが，その影響についての検証がされていない．さらに同研究では，スマートフォンを入れたポケット側の片足のみによるジェスチャ認識であったが，UI 設計において逆方向に同じ操作をすることで，対となる操作が出来る UI 設計が多く行われている．つまり，左足と右足で同じジェスチャを行い，対となる操作が実現できれば，より直感的な操作が可能になると考えられる．

そこで，我々はポケット内のスマートフォンを利用した両足ジェスチャ認識を実現するための手法を提案する．これまでに我々が行ってきた研究[4]では，両足ジェスチャのデータセット構築を行い，スマートフォンのマイクや加速度，角速度のセンサを組み合わせた時の両足ジェスチャの認識精度を検証した．その結果，個人のデータを使って認識させた場合には両足のジェスチャ認識精度を表す F 値が 0.93 であった．しかし，使用実験では，プロトタイプシステムを使用した後に，体感認識精度を回答してもらったが，認識精度の評価実験の結果と比較をすると，認識精度を低く感じる実験協力者が多いという問題が判明した．

体感認識精度が期待したほどの精度ではなかった要因としては，3 つの要因が考えられる．1 つ目の要因は，これまでの研究[4]で提案したジェスチャ認識手法によるジェスチャ認識の精度が低いことである．また，2 つ目の要因は，システムが認識するまでにかかる時間が長かったことである．そして，3 つ目の要因は，

ポケット内におけるスマートフォンの位置変化である．

そこで，本研究では，1 つ目の要因に対しては特徴量の選定を行い，認識精度の改善を図る．次に 2 つ目の要因に対しては，認識までにかかる時間を短縮するため，望ましいシステムの応答時間以内にジェスチャを認識するための手法を検討する．さらに，3 つ目の要因に対しては，まずポケット内のスマートフォンの位置変化による認識精度への影響を明らかにし，認識精度への影響を低減する手法の検討を行う．

2. 関連研究

情報機器における操作性を拡張するためにジェスチャを利用した操作を検討している研究は様々行われている．

CommandBoard[5]では，スマートフォンで主に使われているソフトキーボードでは，文字色の変更や，太文字に変更などの書式の変更を行いたい場合には手間がかかってしまう点に着目した．その手間を解決するために，ジェスチャによるコマンド入力で書式の変更ができる手法を提案し，入力時間，選択時間，入力ミス率の全てにおいて既存の手法より優れていることを明らかにした．石原らの研究[6]では，手に持った状態の携帯端末から取得した加速度の値に対して DPMatching を用いることにより手軽かつ高精度の個人認証を可能とする 3D 動作認証を提案し，そのシステムを 1 か月以上にわたって動作を行った場合，本人拒否率を 10%以下にできることを確かめた．

これらの研究は情報機器における手を使った UI 操作を拡張するための方法として，ジェスチャ入力を検討している．一方で，我々の研究における目的は UI 操作を拡張することではなく，手による操作が向かない状況における操作を実現する手法の提案を目的としている点で異なる．

3. データセットの再構築

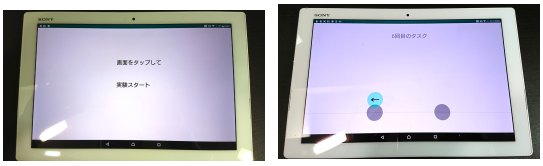
これまでの研究[4]で行った使用実験のアンケート結果から，使用時における認識精度や認識までにかかる時間に問題点があることが明らかとなった．これまでの研究では，認識精度を最重要の項目に位置付けていたため，閾値を超えた時点の前後のフレームを用いてジェスチャ認識を行っており，ジェスチャ認識に使用する時間は，閾値を超えてから約 500[ms]前後経過したときの時間までのセンシングデータを使用していた．その結果，少なくともジェスチャを開始した時間から，実際に操作が反映されるまで約 500[ms]前後は経過していた．また，8 人中 7 人の実験協力者が認識までにかかる時間が遅いと感じていることが明らかとなったため，500[ms]前後かかる認識時間はユーザの要求するレベルに達していなかったと考えられる．ここで，ある動作をしてから反映されるまでの最適な時間を検証している研究の一つに Miller らの研究[7]があげられる．Miller らの研究[7]では，キーボードを押してから 100[ms]から 200[ms]以内がシステム応答時間の限界であることを実験により明らかにしている．この研究から，ジェスチャを行ってから認識するまでにかかる時間は最低でも 200[ms]以内に認識させる必要があると考えられる．

そのため，ジェスチャ認識に使用する時間の短縮を検討することにより，ポケット内のスマートフォンを利用した両足ジェスチャ認識手法の改善を行う．具体的には，あるユーザがジェスチャを開始した時間から，どのくらいの時間のセンシングデータを使用することで，ジェスチャを正しく認識することができるかを検証する．そこで，ジェスチャ開始タイミングから，少しずつフレーム数を増やした際のそれぞれの使用フレームにおけるジェスチャ認識の精度を検証する．しかし，これまでの我々の研究[4]で行ったデータセット構築では，図 1 のように待機画面では，待機画面からタスク実行画面に切り変わるまでの時間を 1 秒単位の整数で表示していた．そのため，実験協力者はタスク画面に切り替わるタイミングを正確に推測することが難しく，実験協力者がジェスチャを開始するタイミングが，システムにより指示したジェスチャ開始のタイミングからずれていたことが考えられる．つまり，これまでのデータセットを用いずに，システムにより指示したジェスチャ開始タイミングと，実験協力者のジェスチャのタイミングが正確に一致したデータセットを構築する必要がある．

そこで，本章ではまず正確にタイミングが一致するデータセット構築手法について検討し，データセットの再構築を行う．

(a)待機画面　(b)タスク提示画面

図 1　これまでの研究におけるタスク提示方法



(a)実験開始前　(b)タスク提示画面

図 2 データセット構築システム

3.1. データセット構築用システム

ここで，人によるジェスチャを開始するタイミングのぶれをなくすための方法として，我々が着目したのは，音楽ゲームをプレイしている際の UI である．音楽ゲームの UI ではリズムに合わせて流れてくるノーツがタップ位置に完全に重なったタイミングでタップする UI が実装されている．そこで我々は音楽ゲームにおける UI を参考にし，図 2 のようにタブレットでジェスチャの開始タイミングを，実験協力者が正確に把握できるように統制を図った．ノーツとは音楽ゲームなどで用いられるリズムに合わせて流れる音符である．

これまでに行ったデータセット構築では，スマートフォンで収集していたデータのセンシング周波数は約 30[Hz]であったが，サンプリング周波数を約 50[Hz]に変更することにより，ジェスチャ時の細かい動きを収集することができると考えられる．そこで，サンプリング周波数は約 50[Hz]とした．センシングした情報は，これまでの我々の研究[4]をもとに 3 軸の加速度$[m/s^2]$, 3 軸の角速度$[rad/s]$, マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧[db]と，最大音圧の周波数[Hz]を使用した．

3.2. データセット構築時の手順

実験協力者には，最初にかかとの上げ下げのジェスチャの仕方を著者自身が実演し，教示した．その次に，ジェスチャを開始するタイミングについては図 2 の(b)の灰色の部分と，水色のノーツの部分が完全に重なったタイミングでジェスチャを開始するように著者自身が実演し，実験協力者にジェスチャの開始タイミングを教示した．実験協力者が実験について理解したことを確認した後，スマートフォンとタブレット用のそれぞれのデータセット構築システムを用いて，かかとの上げ下げのジェスチャを 200 回してもらった．ここで，スマートフォンを入れるポケットと向きは，右足の前ポケットで，マイク部分が下になり，ディスプレイが足に触れない向きになるように実験の統制を図った．また，実験協力者が約 5000[ms]に 1 回の間隔でジェスチャを行うように，ノーツが約 5000[ms]に一回の間隔で灰色の部分と完全に重なるようにシステムで制御した．ジェスチャを行う時間間隔の設定は，前のジェスチャが完全に終了しており，次のジェスチャに影響を与えない間隔となるように設定している．約 5000[ms]という時間は，筆者自身が実際にジェスチャを行った際にかかった時間をもとに設定した．ジェスチャの合間の姿勢については，特に指示をしなかった．

なお，このデータセット構築では 18〜25 歳の大学生 15 人の実験協力者に 200 回のかかとの上げ下げのジェスチャを行ってもらった．15 人の実験協力者の中で利き足が右足の実験協力者が 14 人であり，左足の実験協力者が 1 人だった．また，200 回のジェスチャの内訳は，ジェスチャの種類による偏りを防ぐため，右足によるジェスチャを 100 回，左足によるジェスチャを 100 回とした．

4. 特徴量の再選定

これまでの我々の研究[4]では，SVM を用いた認識精度が最も高い認識精度手法であったが，使用実験のアンケート結果から，その認識精度もユーザが期待しているほどの認識精度ではないことが明らかとなった．そこで，本研究では，近年機械学習で使用されている分類器である Random Forest を用いた際の認識精度と，SVM を用いた際の認識精度の比較を行い，Random Forest による認識の有用性についての検討を行う．

4.1. 特徴量の検討

これまでの研究では，最大値と最小値のみを用いてジェスチャ認識を行っていため，波形の最大値と最小値の情報のみしか用いていなかった．しかし，波形の特徴をさらに活かすことにより，両足ジェスチャの認識精度改善が期待される．例えば，波形の平均値はその時間帯にどのくらいの時間をかけて動いたのかという情報を取得することができる．また波形の中央値も同様の情報を取得することができると考えられる．平均値と中央値は同様の情報を取得することができるが，ジェスチャ中に極端に大きな動きを短時間にした場合には，平均値と中央値に違いが出ると考えられる．

そこで最大値と最小値以外にも，平均値と中央値を使って，ポケット内のスマートフォンを用いた両足ジェスチャを認識する．

4.2. データの抽出

本章では第 3 章で構築したデータセットを使用して認識精度の検証を行う．そこで，まず特徴ベクトルを生成するための区間を決定し，ジェスチャ時と待機時におけるセンシングデータの抽出を行う．これまでに我々が行った研究では，センシング周波数は 30Hz であり，特徴ベクトルは閾値を超えた時間より前の 9 フレーム分(約 300[ms])と，後の 16 フレーム分(約 540[ms])の計 25 フレームを使用して特徴ベクトルを生成し，ジェスチャを認識していた．そこで本研究でも閾値を超えた時間よりも前の 15 フレーム(約 300[ms])のセンシングデータと，後の 27 フレーム(約 540[ms])の計 42 フレームのセンシングデータを使用する．

右足と左足によるジェスチャデータの抽出では，まずタスク開始時間に一番近いセンシングデータから 100 フレーム分(約 2000[ms]分)のデータを抽出する．そして，その 100 フレーム分のデータを基に，著者が設定した閾値を超えた最初のフレームを特定し，その前後のフレームを抽出した．また，待機状態のセンシ

ングデータの抽出では，ジェスチャを開始するタイミングより前の時間帯のセンシングデータを使用し，ジェスチャ認識に用いるフレーム数に合わせて待機時のフレーム数を選択する．そこで，ジェスチャ開始タイミングから 102 フレーム前（約 2000[ms]前）から 61 フレーム前(約 1250[ms]前)までの計 42 フレーム分のセンシングデータを抽出した．

4.3. 性能評価

ジェスチャ認識の性能評価では，最大値・最小値により生成した特徴ベクトルとそれぞれの分類器の組み合わせにおける認識精度と，最大値・最小値・平均値・中央値により生成した特徴ベクトルとそれぞれの分類器における認識精度を比較することにより，ジェスチャ認識手法を改善するための認識手法を検討する．

今回使用する特徴量はこれまでの研究[4]と同じく表 1 の特徴を選定した．分類器については，近年機械学習で使用されている Random Forest と，これまでの研究[4]で使用した SVM を用いる．今回使用した SVM と Random Forest の認識に使用したライブラリは，scikit-learn である．SVM と Random Forest におけるパラメータはそれぞれデフォルトのパラメータを使用した．本研究では，生成した特徴ベクトルにおいて，センサの値の大きさによる偏りが認識精度に与える影響を考慮し,その特徴ベクトルの値を標準化した.また，第 3 章で行ったデータセット構築では両足ジェスチャそれぞれの 100 回の合計 200 回のジェスチャを行ってもらっており，本研究ではジェスチャが行われる前の時間帯を待機状態として収集しているため，待機状態のデータが 200 回分存在する．しかし，200 回分の待機状態を全て使った場合には，ジェスチャのデータ数と待機状態のデータ数に偏りが生じてしまい，待機状態での認識精度がジェスチャのデータよりも認識精度へ与える影響が大きくなってしまうため，待機状態のデータが 100 回分になるようにアンダーサンプリングを行った．さらに，認識精度の算出の際には，300 回分のデータ（左足によるジェスチャデータ 100 回分，右足によるジェスチャデータ 100 回分，待機状態のデータ 100 回分）から，8 割のデータ（240 回分のデータ）を用いて学習を行い，残りの 2 割のデータ（60 回分のデータ）を用いて認識精度の算出を行った．

4.4. 実験結果

表 2 が各クラスにおける Precision（適合率）の平均値，Recall（再現率）の平均値，それらの平均値をもとに算出した F-measure(F 値)の表である.表 2 の Previous と書かれた列が最大値・最小値で生成した特徴ベクトルを用いた時の認識精度を求めたものであり，New と書かれた列が最大値・最小値・平均値・中央値で生成した特徴ベクトルを用いた時の認識精度である．

表 2 の結果から，Random Forest と SVM のそれぞれの手法で，従来の特徴量で生成した特徴ベクトルを用いた際の認識精度を，新たに選定した特徴量で生成した特徴ベクトルを用いた際の認識精度が上回っている．

本章では，ジェスチャの認識精度を改善することを目的として，近年機械学習で用いられている Random Forest による認識手法を検討した．その結果としては，表 2 から分かるように，従来の特徴量により生成した特徴ベクトルを用いた際と，新たに選定した特徴量により生成した特徴ベクトルを用いた際の認識精度は，Random Forest を用いた際の認識精度の方が高かったが，認識精度の差がわずかであるため，特徴量による差は見受けられなかった．しかし，センシング周波数と特徴量，分類器を変更したことにより，これまでの研究[4]における最も高い認識精度であった F 値 0.93 から認識精度が改善された．

5. 予備動作を用いた認識手法の提案

3 章でも述べた通り，認識にかかる時間に問題があった.そこで,本章では最低でも 200[ms]以内に両足ジェスチャの認識を可能とする手法の検討を行う．

5.1. 予備動作を用いた認識手法

そこで，ジェスチャ認識に使用する時間を短縮する手法を検討する.200[ms]以内にジェスチャを認識するためには，ユーザがジェスチャを開始した時間から 200[ms]以降のデータを使うことができないため,少ない情報からジェスチャの認識を行うことになる．そのため，約 500[ms]を用いて認識させていたこれまでの手法と比べると，ジェスチャの認識精度が低下することが考えられる．ここで，200[ms]以降のジェスチャデータは使用できないが，ジェスチャを始める直前に無意識に行われる構えの動作である予備動作のセンシングデータを利用することで，予備動作を用いない時より認識精度が高くなることが考えられる．その結果として，望ましい認識時間以内に，高い精度でジェスチ

表 1 ジェスチャ認識に使用した特徴一覧

特徴
X 軸・Y 軸・Z 軸の加速度値
X 軸・Y 軸・Z 軸の角速度値
マイクで得た最大音圧の周波数
マイクで得た最大音圧

表 2 それぞれの認識手法における認識精度
（Previous:従来の特徴量, New:新たに選定した特徴量）

	SVM		RandomForest	
	Previous	New	Previous	New
Precision	0.9820	0.9838	0.9848	0.9863
Recall	0.9811	0.9833	0.9844	0.9856
F-measure	0.9816	0.9836	0.9846	0.9859

ャを認識したい場合には，予備動作のセンシングデータを利用することで，認識精度の悪化を防ぎながら認識時間を短縮することができると考えられる．

5.2. 性能評価

予備動作のセンシングデータを用いた認識手法の性能評価では，予備動作なしで生成した特徴ベクトルより，予備動作を用いて生成した特徴ベクトルを用いた場合の方が，ジェスチャの認識精度が高くなるかを検証する.今回使用するセンサは第 4 章と同じであり，使用する特徴量は第 4 章の結果から最大値と最小値，平均値，中央値の 4 種類である．分類器については，第 4 章で用いた Random Forest と，SVM を使用し，特徴ベクトルの値に対して標準化の処理を行っている．

5.3. 実験結果

表 3 が Random Forest と SVM において，ジェスチャ開始タイミングより前のフレームと，ジェスチャ開始タイミングのフレームと，それ以降のフレームを含んだフレームを用いて特徴ベクトルを生成し，認識精度を算出した F-measure（F 値）である．今回使用した特徴量は，最大値，最小値，平均値，中央値であるため，使用フレームの総数が 2 フレーム未満の場合は，最大値と最小値が一致してしまうため，認識精度の算出は行わなかった．表 3 の赤い背景で表示されている箇所は，ジェスチャ開始タイミングのフレームと，それ以降のフレームを含む使用フレーム数が同一の特徴ベクトルの中で，最も認識精度が高いことを表している．

本研究では望ましいシステムの応答時間（200[ms]）以内でジェスチャを認識することを目材しているため，ユーザがジェスチャをし始めてから最低でも 10 フレーム（約 200[ms])以内にジェスチャを認識する必要がある．そのため，その条件を満たした中で最も高い精度だったのは，ジェスチャをする前の 10 フレームと，ジェスチャをし始めたフレームとそれ以降のフレームを含む 10 フレームの組み合わせが最も高い認識精度であった．またジェスチャをし始めたフレームと，それ以降のフレームを含む 10 フレーム以外の組み合わせにおいても，全ての組み合わせにおいて，ジェスチャを開始する前のフレームを含むことにより認識精度が向上した．

表 3　Random Forest と SVM における F 値
（縦軸：前フレーム使用数，横軸：後フレーム使用数）

		0	1	2	9	10	15	20
0	RF			0.758	0.870	0.881	0.925	0.970
	SVM			0.739	0.862	0.877	0.931	0.956
1	RF		0.737	0.753	0.883	0.888	0.939	0.972
	SVM		0.723	0.745	0.864	0.879	0.938	0.956
2	RF	0.707	0.730	0.771	0.876	0.885	0.929	0.961
	SVM	0.702	0.716	0.750	0.869	0.889	0.936	0.957
3	RF	0.715	0.762	0.776	0.875	0.871	0.941	0.965
	SVM	0.711	0.734	0.755	0.874	0.887	0.940	0.956
4	RF	0.744	0.744	0.792	0.884	0.893	0.933	0.973
	SVM	0.727	0.734	0.760	0.871	0.879	0.933	0.960
5	RF	0.728	0.745	0.782	0.869	0.891	0.943	0.970
	SVM	0.734	0.739	0.758	0.865	0.878	0.934	0.960
6	RF	0.735	0.747	0.769	0.887	0.891	0.935	0.969
	SVM	0.736	0.756	0.761	0.860	0.882	0.937	0.962
7	RF	0.709	0.762	0.783	0.892	0.888	0.939	0.967
	SVM	0.745	0.751	0.764	0.863	0.878	0.936	0.960
8	RF	0.718	0.759	0.775	0.886	0.879	0.936	0.970
	SVM	0.754	0.757	0.772	0.862	0.880	0.935	0.960
9	RF	0.740	0.753	0.765	0.878	0.894	0.939	0.967
	SVM	0.734	0.751	0.771	0.868	0.877	0.935	0.959
10	RF	0.725	0.757	0.773	0.878	0.897	0.937	0.962
	SVM	0.754	0.747	0.772	0.870	0.882	0.935	0.959

6. 位置変化による影響調査実験

6.1. 実験目的

これまでに我々が行ってきた研究[4]では，スマートフォンを入れる向きが変わった際の認識精度の考慮を行わずに，特定の向きにおける認識精度を算出していた．しかし，スマートフォンを入れる際には，特定のポケットや向き，位置にスマートフォンを入れる状況は稀である．そのため，ポケットにスマートフォンを入れる際に，データセット構築時とはスマートフォンを入れたポケットや向き，位置を変えて入れた場合には，データセット構築時とは加速度や角速度などの軸の正方向の向きが変化するなどにより，センサの情報が変化し,認識精度に影響を及ぼすことが考えられる．そのため，ポケット内でのスマートフォンの位置変化による認識精度への影響について検証する．

第 3 章のデータセット構築では，ポケットにスマートフォンを入れる際の向きを,マイク部分が下になり，ディスプレイ部分が足に触れない向きになるように実験の統制を図っていた．しかし，本章ではポケット内でのスマートフォンの位置変化による認識精度への影響を調査する必要があるため，新たにデータセット構築を行う．

6.2. データセット構築

データセット構築では，20 歳～25 歳の大学生 7 人の実験協力者に，両足によるかかとの上げ下げのジェスチャを，第 3 章で提案したデータセット構築用システムを用いて行ってもらった．本実験でポケットにスマートフォンを入れる際のスマートフォンの向きのパターンは以下の 4 パターンとなる．

パターン①：　マイク部分が下になり，ディスプレイ部分が足に触れない向き

パターン②：　マイク部分が下になり，ディスプレイ部分が足に当たる向き

パターン③：　マイク部分が上になり，ディスプレイ部分が足に触れない向き

パターン④： マイク部分が上になり，ディスプレイ部分が足に当たる向き

そのため，本データセット構築では右足と左足の 2 パターン，スマートフォンをポケットに入れる際の 4 パターンのすべての組み合わせ（計 8 パターン）のそれぞれで 40 回ずつ両足のかかとの上げ下げのジェスチャを行った際の合計 320 回のジェスチャを 1 セットとした．そして，実験協力者には 2 セットの合計 640 回かかとの上げ下げのジェスチャをしてもらい，データセットを構築した．

6.3. 影響調査

位置変化による認識精度への影響調査では，ポケット内のスマートフォンの向きや位置が認識精度へ与える影響について検証する．本章では，まずポケット内でのスマートフォンの向きが与える認識精度への影響を調査するため，まずスマートフォンを入れるポケットと向きが，右足の前ポケットで，マイク部分が下になり，ディスプレイが足に触れない向きでのジェスチャデータを基に，ほかの向きでの認識精度の算出を行う．次に，右足すべての向きで学習した分類器を用いた際の認識精度の算出を行う．そして，右足と左足の全ての向きで学習した分類器を用いた際の認識精度の算出を行う．

認識精度の算出に用いる分類器と前フレームと後フレームの組み合わせは前の章を参考に，分類器には Random Forest を用いて，前フレームを 10 フレーム使用し，後フレームを 10 フレーム使用する．Random Forest におけるパラメータはそれぞれデフォルトのパラメータを使用し，特徴ベクトルの値は標準化した．また，第 5 章と同じように，ジェスチャのデータ数と待機状態のデータ数の偏りを考慮し，同じデータ数となるようにアンダーサンプリングを行った．ここで，学習に使用するパターンの認識精度の算出の際には，120 回分のデータ（左足によるジェスチャデータ 40 回分，右足によるジェスチャデータ 40 回分，待機状態のデータ 40 回分）から，8 割のデータ（96 回分のデータ）を用いて学習を行い，残りの 2 割のデータ（24 回分のデータ）を用いて認識精度の算出を行った．学習に使用しないパターンの認識精度の算出ではアンダーサンプリングは行わずに 120 回分（右足 40 回分，左足 40 回分，待機 40 回分）のデータを全てテストデータとして用いて認識精度の算出を行った．本章で使用するセンサは第 4 章と同じく加速度，角速度，マイクセンサにより収集した音声データに対してフーリエ級数展開を行って得た最大音圧と，最大音圧の周波数を使用した．また今回使用する特徴量は第 4 章の結果から，最大値と最小値，平均値，中央値を用いる．

6.4. 実験結果･考察

まず，表 4 がスマートフォンを入れるポケットと位置と向きが，右足の前ポケットのパターン①のみで学習したモデルを用いた時，右足の前ポケットの 4 パターンで学習させたモデルを用いた時，右足と左足の 4 パターンで学習させたモデルを用いた時に，それぞれの向きと位置におけるテストデータを分類した際の認識精度を表した表である．

表 4(a)の結果から，向きが一致しない場合には，認識精度が悪化した．つまり，ポケット内でのスマホの向きが大きく変わった場合には認識精度が悪化する可能性が示唆されている．また右足のポケットに入っているか左足のポケットに入っているかというポケットの違いが認識精度に影響していた．表 4(b)は右足のポケットにおける全ての向きのパターン（4 パターン）で学習したモデルを用いてそれぞれの組み合わせのテストデータを用いて認識精度を算出した表であるが，

表 4　ポケット内のスマートフォンの位置と向きによる認識精度への影響

(a)　右足のパターン①のみで学習

ポケット	向き	Precision	Recall	F-measure
右足	①	0.9196	0.9107	0.9151
	②	0.2882	0.3417	0.3127
	③	0.5372	0.5405	0.5388
	④	0.4226	0.4488	0.4353
左足	①	0.5921	0.5536	0.5722
	②	0.3949	0.4286	0.4110
	③	0.4557	0.4845	0.4697
	④	0.5248	0.5071	0.5158

(b)　右足の 4 パターンで学習

ポケット	向き	Precision	Recall	F-measure
右足	①	0.8742	0.8631	0.8686
	②	0.8641	0.8512	0.8576
	③	0.9223	0.9167	0.9195
	④	0.8886	0.8810	0.8848
左足	①	0.4675	0.4702	0.4689
	②	0.4459	0.4369	0.4413
	③	0.4117	0.3774	0.3938
	④	0.4580	0.4655	0.4617

(c)　右足と左足の 4 パターンで学習

ポケット	向き	Precision	Recall	F-measure
右足	①	0.7903	0.7798	0.7850
	②	0.8213	0.8036	0.8124
	③	0.8403	0.8393	0.8398
	④	0.8370	0.8155	0.8261
左足	①	0.8774	0.8452	0.8610
	②	0.8459	0.8333	0.8396
	③	0.8517	0.8333	0.8424
	④	0.8428	0.8155	0.8289

右足で学習したモデルは反対の足のポケットにスマートフォンを同じ向きで入れた場合にも認識精度が悪化していることが分かる．また，4 パターンで学習させたモデルを用いて，スマートフォンを入れている足のパターン①の認識精度を算出したが，1 パターンのみで学習させたモデルで算出した認識精度から認識精度が低下していることが分かる．つまり，スマートフォンのポケット内での向きと位置によって認識精度が悪化することが明らかとなった．

7. 位置変化に適応した提案手法

7.1. 提案手法

6 章での結果から位置変化が認識精度へ悪影響を与えていることが明らかとなった．そのため，位置変化による認識精度への影響を軽減する必要がある．そこで，ポケット内でのスマートフォンの状態を表す特徴量を用いることで，スマートフォンの位置変化による認識精度への影響を軽減する．ここでは，これまでに用いた特徴量以外にも，地磁気と加速度センサから算出されるスマートフォンの傾きの特徴量を利用することで，位置変化に適応した認識手法を提案する．

7.2. 性能評価

位置変化に適応した認識手法を用いることで，ポケット内のスマートフォンの向きや位置が認識精度へ与える影響を軽減することができるかを検証する．具体的には，これまでに用いていた特徴量以外に傾きの特徴量を用いることで，認識精度へ与える影響がどのように変化するのかを調査する．

7.3. 実験結果

表 5 は，スマートフォンを入れるポケットと位置と向きが，右足の前ポケットのパターン①のみで学習したモデルを用いた時，右足の前ポケットの 4 パターンで学習させたモデルを用いた時，右足と左足の 4 パターンで学習させたモデルを用いた時のそれぞれの向きと位置におけるテストデータを位置変化に適応した認

表 5　ポケット内のスマートフォンの位置と向きによる認識精度への影響（傾きの特徴を使用）

(a) 右足のパターン①のみで学習

ポケット	向き	Precision	Recall	F-measure
右足	①	0.9263	0.9226	0.9245
	②	0.3378	0.3560	0.3467
	③	0.6217	0.5548	0.5863
	④	0.5549	0.5048	0.5287
左足	①	0.5429	0.5560	0.5494
	②	0.3901	0.3857	0.3879
	③	0.4417	0.4381	0.4399
	④	0.5934	0.5857	0.5895

(b) 右足の 4 パターンで学習

ポケット	向き	Precision	Recall	F-measure
右足	①	0.8999	0.8929	0.8964
	②	0.8861	0.8810	0.8835
	③	0.9170	0.9107	0.9138
	④	0.8983	0.8929	0.8955
左足	①	0.4339	0.4440	0.4389
	②	0.3839	0.3845	0.3842
	③	0.4479	0.4381	0.4429
	④	0.4452	0.4417	0.4434

(c) 右足と左足の 4 パターンで学習

ポケット	向き	Precision	Recall	F-measure
右足	①	0.8890	0.8750	0.8819
	②	0.8508	0.8452	0.8480
	③	0.8762	0.8631	0.8696
	④	0.8762	0.8631	0.8696
左足	①	0.9091	0.9048	0.9069
	②	0.8957	0.8869	0.8913
	③	0.8630	0.8571	0.8601
	④	0.9221	0.9107	0.9164

表 6　Random Forest の認識精度への特徴量の寄与度[%]（1P：右足のパターン①のみで学習，4P：右足の 4 パターンで学習，8P：右足と左足の 4 パターンで学習）

特徴量			学習パターン数 1P	4P	8P	特徴量			学習パターン数 1P	4P	8P
加速度	X軸	最小	3.8	3.5	3.5	角速度	X軸	最小	5.7	5.2	4.4
		最大	1.4	3.7	3.6			最大	3.4	5.6	4.4
		平均	9.7	2.8	2.7			平均	4.7	4.5	2.5
		中央	5.7	1.9	2.1			中央	1.9	1.9	1.8
	Y軸	最小	2.8	4.2	2.5		Y軸	最小	2.9	4.6	4.3
		最大	5.1	4.5	2.5			最大	2.0	4.2	4.1
		平均	4.1	3.8	2.0			平均	1.5	1.6	2.3
		中央	1.6	2.6	1.8			中央	1.3	1.0	1.9
	Z軸	最小	4.5	4.7	4.3		Z軸	最小	3.0	4.6	5.4
		最大	4.2	4.3	4.3			最大	4.9	5.7	6.1
		平均	2.6	3.1	2.7			平均	4.1	3.5	4.5
		中央	3.2	1.9	1.8			中央	2.3	2.2	2.7
傾き	X軸	最小	0.8	1.6	2.0	マイク	周波数	最小	0.0	0.3	0.3
		最大	1.9	1.3	2.2			最大	0.1	0.1	0.2
		平均	1.3	1.5	2.0			平均	0.3	0.6	0.7
		中央	0.7	1.2	2.1			中央	0.2	0.2	0.4
	Y軸	最小	2.9	1.4	2.5		最大音圧	最小	0.2	0.4	0.5
		最大	0.9	1.5	2.1			最大	0.5	0.5	0.6
		平均	2.2	1.2	2.4			平均	0.6	0.7	0.9
		中央	0.6	1.3	2.4			中央	0.4	0.5	0.6

識手法を用いて分類した際の認識精度を表した表である．背景が赤い箇所は傾きの特徴量を使用した提案手法を用いることで認識精度が向上した組み合わせを表している．表 6 が傾きの特徴量が両足ジェスチャの認識にどのくらい寄与しているかを表した表である．

表 4(a)が傾きを用いなかった際，表 5(a)が傾きの特徴量を用いた際の，右足のポケットにパターン①の向きでスマートフォンを入れたジェスチャデータで学習させた時に，それぞれのポケット内のスマートフォンの位置と向きにおける認識精度への影響を調査した表である．その結果，傾きの特徴量を用いることで両足ジェスチャの認識精度が改善していた．また，表 4(b)が傾きの特徴量を用いずに，表 5(b)が傾きの特徴量を用いた際の，右足の前ポケットに入れる際の 4 パターン全ての向きのトレーニングデータを使用して学習させたモデルを用いて，それぞれの組み合わせのテストデータを分類した際の認識精度を表した表である．その結果，1 パターンで学習させた場合と同様に傾きの特徴量を用いることで認識精度が改善される結果となった．

8. 考察

本研究では，予備動作をセンシング可能なデータセット構築システムを実装し,予備動作を用いることで,認識精度が向上した．この知見を生かすことで他のジェスチャ認識の研究においても，予備動作を用いることで，認識精度改善や認識時間短縮をすることができると期待される．また本研究では，スマートフォンの向きや位置の変化が認識精度に与える影響を明らかにし，スマートフォンの傾きの特徴量を用いることにより，認識精度への影響を軽減できることを明らかにした．この知見は，本研究のように位置が固定されないデバイスを用いたジェスチャ認識の研究において，位置変化による認識精度への影響を軽減することができると期待される．

9. おわりに

本研究では，これまでに我々が行ってきたポケット内のスマートフォンを用いた両足ジェスチャによる研究において明らかになった問題について述べ，それぞれの問題を解決するための手法を提案した．

まず認識手法の改善では，特徴量や分類器，センシング周波数を再検討し，結果としてジェスチャの認識精度が改善された．

次に，ジェスチャの認識時間を短縮するために，ジェスチャを開始したタイミング以降の使用フレーム数を減らした場合には，ジェスチャの認識精度が悪化した．そのため，ジェスチャ中の動作以外にも，ジェスチャを開始する直前に行われる構えの動作である予備動作を用いて認識することで，認識時間を短縮したことによる認識精度への影響を軽減する手法の提案を行い，その手法の有用性を明らかにした．

さらに，ポケット内でのスマートフォンの位置変化については，まずポケット内のスマートフォンの位置変化による認識精度への影響を明らかにし，ポケット内のスマートフォンの位置変化による認識精度への影響を低減する手法の提案を行い，その手法の有用性を明らかにした．

参 考 文 献

[1] Matthies, D.J.C., Roumen, T.: CapSoles: who is walking on what kind of floor?, In Proc MobileHCI ’17, p.9(2017).

[2] Fukahori, K., Sakamoto, D,. Igarashi, T.: Exploring Subtle Foot Plantar-based Gestures with Sock-placed Pressure Sensors, In Proc CHI’15, p.10(2015).

[3] Scott, J., Dearman, D., Yatani and K. Truong, N.K.: Sensing Foot Gestures from the Pocket, In Proc. of UIST ’10, pp.199-208(2010).

[4] 田村柾優紀, 中村聡史: ポケット内のスマートフォンによる両足ジェスチャ認識手法の提案と分析, 研究報告グループウェアとネットワークサービス, p.8(2017).

[5] lvina, J., CarlaF.Griggio, C. F., Bi X. and Mackey W. E.: CommandBoard: Creating a General-Purpose Command Gesture Input Space for Soft Keyboards, In Proc of UIST’17, pp.17-28(2017).

[6] 石原進, 太田雅敏, 行方エリキ, 水野忠則. 端末自体の動きを用いた携帯端末向け個人認証. 情報処理学会論文誌, 2005 vol. 46, no. 12, p.2997-3007(2005).

[7] Miller, R. B.: Response Time in Man-computer Conversational Transactions, Fall Joint Computer Conferences, pp.267-277(1968).