

加速度センサを用いた複合的動作の認識手法

村 尾 和 哉^{†1} 寺 田 努^{†1,†2}

加速度センサを用いた行動認識技術に関する研究が数多く提案されている．それらの研究において取り扱われている行動の多くは「座っている」や「歩いている」「携帯電話を把持している」「ボールを投げている」など 1 種類の動詞で表現可能な単体の動作である．しかしながら，現実では複数の行動や状態から成る複合的な動作も発生する．例えば，手で行うジェスチャ動作は立っている状態だけでなく，歩いたり座ったりしている最中にも行われる．このような複合的動作のもっとも単純な認識手法としては，考えられるすべての複合的動作のデータを採取して学習する方法が考えられるが，その組合せは膨大になる．本研究では単体の動作のデータのみの学習から複合的動作を認識する手法を提案する．評価より，提案手法は 85%の精度で複合的動作を認識し，これはすべてのデータを学習した場合と同等の性能であることを確認した．

A Recognition Method using Accelerometers for Combined Activities

KAZUYA MURAO^{†1} and TSUTOMU TERADA^{†1,†2}

Many activity recognition systems using accelerometers have been proposed. Activities that have been recognized are “single” activities which can be expressed with one verb, such as sitting, walking, holding a mobile phone, and throwing a ball. In actual, however, “combined” activities including more than two kinds of state and movement are often taken place. Focusing on hand gestures, they are performed not only while standing, but also while walking and sitting. Though the simplest way to recognize such combined activities is to construct the recognition models for all the possible combinations of the activities, the number of combinations becomes immense. In this paper, we propose a recognition method for combined activities by learning single activities only. Evaluation results confirmed that our proposed method achieved 0.85 of accuracy, which is comparable to the method that had learned all the combined activities.

1. はじめに

近年，マイクロエレクトロニクス技術の発展によるコンピュータの小型化や軽量化により，コンピュータを常時身に付けて生活するウェアラブルコンピューティングに注目が集まっている．ウェアラブルコンピューティングの発展に伴い，加速度センサやジャイロセンサ，筋電計¹⁴⁾ や心電計¹²⁾，GSR (皮膚電気反射)¹⁰⁾，あるいは独自のデバイス⁵⁾ といったさまざまな装着型センサを用いてユーザの状態を認識し，状態に依存した適切なサービスを提供するシステム (コンテキストウェアシステム) が提案されている^{9),10),13)}．例として，LifeMinder¹⁰⁾ は日常生活での行動を温度センサ，GSR センサ，加速度センサ，光電脈波センサ，地磁気センサ，ジャイロセンサを用いて認識し，生活習慣の改善などのアドバイスをを行う．具体的には，腕時計型センサを用いて食事や歩行，仕事などの装着者の行動を認識し，運動不足や過労を警告する．位置・動作情報を検出するセンサとして加速度センサのほかにカメラや GPS，ジャイロ，地磁気センサなどがあげられるが，これらのセンサは装着性や精度の低さ，動きと静止時の方向が同時に取得不可能といった問題点をもつ．一方，加速度センサは装着部位の動きを検出可能であり，静止時でも地球の重力を検出することによりその方向が取得可能であるため使い勝手が良い．また，精度および分解能が高く，小型で容易に装着可能である．このように既存センサの中では加速度センサが行動認識システムの構築に適している．

一般的な行動認識の流れは，まず認識したい動作を定義し，そのデータを採取し，採取したデータの正解の行動を付与し，認識モデルを構築する．そのため，認識結果は付与した正解の行動，つまりあらかじめ定義した動作に限られる．例えば，「歩きながら携帯電話を把持する」という動作を認識したい場合，その状態のデータを採取し，そのようにラベル付けを行わなければならない，仮にこの行動を学習していない場合は単に「歩いている」や「携帯電話を把持している」という結果が得られることになる．

ここで，本研究では「グローバル動作」と「ローカル動作」の 2 種類の人間の動作を定義する．グローバル動作とはからだ全体の動作であり「歩く」や「立つ」などが挙げられる．グローバル動作は同時に 2 種類以上は発生しない．対して，ローカル動作はからだの一部分

^{†1} 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

^{†2} 科学技術振興機構さきがけ

PRESTO, Japan Science and Technology Agency

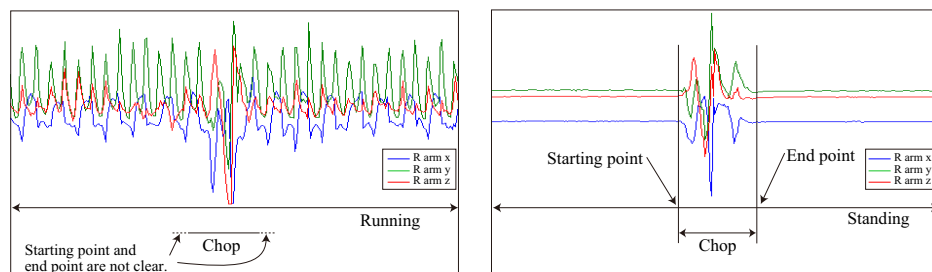


図 1 走っている最中のチョップの加速度波形 (左) と立っている最中のチョップの加速度波形 (右)
Fig. 1 Acceleration of a chop gesture while running (left) and while standing (right).

の動作であり、「モノを投げる」や「モノを把持する」などである。ローカル動作は同一の箇所でない限り、2 種類以上が同時発生し得る。

複合的な動作を認識するもっとも単純な手法は、複合的動作のデータを採取して認識モデルを作成することであるが、グローバル動作とローカル動作の組合せの数は膨大になる。例えば「立つ」「座る」「歩く」「走る」「自転車に乗る」の 5 種類のグローバル動作に対して 10 種類の手のジェスチャのローカル動作を考える場合、50 種類の動作のデータを採取しなければならない。ジェスチャが 1 種類増えるにつれて 5 通りのデータが必要となり、開発者にとっては骨の折れる仕事であり、一般利用者にとっては極めて困難な作業である。ここに足のジェスチャが加われば、可能な組合せはグローバル動作 × 手のローカル動作 × 足のローカル動作となる。

複合的動作の認識の問題は学習データの準備に時間を要する点だけでなく、以下のような点が挙げられる。図 1 に走っている最中にチョップした際と立っている最中にチョップした際の手に装着した加速度センサの波形を示す。図から立っている最中のジェスチャの開始点と終了点は正確に指し示すことができる。一方、走っている最中のジェスチャに関しては、ジェスチャが行われた大まかな箇所は判断できるが、正確な開始点と終了点を指し示すことができず、採取したデータへの正解の動作情報の付与が困難である。加えて、手以外の箇所は同じ動作を行っているため、体全体で見た場合、両者の判別性能が低下する。ここで、このような複合的な動作を認識する手法として、単体の動作のみを学習してセンサごとに個別に動作認識を行い、認識結果を統合するアプローチが考えられるが、図 1 に示すように、走っている際の足の振動が手に伝わるため、走っている最中のジェスチャは立ちながらのものとは異なる。また、走りながら行うジェスチャは立っている時ほど自由に行えない。

このような動作の違いによって誤認識が発生する可能性がある。筆者らの先行研究において、歩くなどのグローバル動作中のジェスチャ認識手法を提案している⁷⁾。しかし、この研究で取り扱われているジェスチャはグローバル動作からスムーズに切り替わっており、複合的なものではなく単体であり、複合的動作の効率的な認識手法は筆者らの知る限り報告されていない。

本研究では、グローバル動作および立位でのローカル動作である単体の動作のみの学習による複合的動作の認識手法を提案する。提案システムは加速度センサの値の変動および周期性から各部位が静止しているか、周期的な動作をしているか、ジェスチャ動作をしているかを判断する。その後、各部位の状態に応じて動作を認識してグローバル動作とローカル動作を出力し、最終的に両者を統合する。提案システムを用いることで「歩く」「走る」「投げる」「携帯を持って見る」の動作を学習するだけで「歩きながら投げる」や「走りながら携帯を見る」という複合的動作を認識できる。

2. 関連研究

行動認識に関する 10 の研究が文献 1) に紹介されているが、それらの研究で扱われている行動の多くが歩行や姿勢などの単体の動作である。例として、文献 11) では 22 種類のセンサを用いて Lying, Rowing, Running, Nordic walking, Bike, Walking, Sitting, Standing を認識している。他の研究では、マイクや加速度センサなど 7 種類のセンサを搭載した MSB(Multi-modal sensor board) と呼ばれる端末を用いて Sitting, Standing, Walking, Walking up stairs, Walking down stairs, Riding elevator down, Riding elevator up, Brushing teeth を認識している⁶⁾。これらの研究で認識対象となっている動作はいずれも体全体を用いたグローバル動作であり、動作中に別のローカル動作を行うことは想定していない。

一方、体の一部だけを動かすローカルな動作に着目している研究もいくつか存在する。Graeme らはビデオ映像の解析のみによるアノテーションは困難であるため、手首に装着した加速度センサを用いたジェスチャ認識によるビデオアノテーション手法を提案している²⁾。HMM を用いてジェスチャを認識しており「チョップ」「ひじ打ち」「パンチ」の 3 種類のカンフーの型を 30 回行い 29 回認識に成功している。Holger らは「ボタンを押す」や「飲み物を飲む」「受話器を上げる」などの日常の 10 種類の動作を装着型センサを用いて認識している³⁾。この研究では Sliding-window and bottom-up (SWAB)⁴⁾ を用いてリアルタイムでセグメント化した時系列データの特徴量変換し、HMM で分類している。ローカル動作

であるジェスチャは、ある決められた姿勢で行われており、それ以外の状況は想定していない。例えば、上記の研究で「ボタンを押す」というジェスチャがあるが、正確には「立ちながらボタンを押す」であり、「座りながらボタンを押したり」「歩きながらボタンを押す」動作には対応出来ない。

このように、行動認識システムの認識対象となっている行動の大半が単体の動作であり、複雑な複合的動作は一つ一つ定義する必要がある。しかし、複合的動作の種類はグローバル動作とローカル動作の積となり、学習用データの採取に膨大な時間を要する。10 種類のグローバル動作と 10 種類の 10 秒程度のローカル動作からなる複合的な動作を各 5 回ずつ採取する場合、5,000 秒を要する。これは実際に動作を行っている時間であるため、動作間のインターバルや休憩を含んだ場合 3~4 時間となる。さらに長時間に渡って何百回もジェスチャを行うことで蓄積する被験者が疲労も、正常時と異なる動作を誘発する可能性があるため無視できない。

3. 提案手法

本節では提案する複合的動作の認識手法について述べる。処理の流れを図 2 に示す。提案手法は行動のカテゴリ分類フェーズ、行動認識フェーズ、行動統合フェーズの 3 段階で構成されている。カテゴリ分類フェーズではセンサが装着されている各部位の動作状態を静止、運動、ジェスチャの 3 種類に分類する。静止は「立つ」や「座る」などのポーズを取っている状態、運動は「歩く」や「走る」など定常的な動作が繰り返されている状態、ジェスチャは「投げる」や「回す」などの一回限りの動作である。次に、行動認識フェーズでは分類されたカテゴリの動作のみを学習した認識器を用いて各部位の動作を認識する。カテゴリ分類フェーズおよび行動認識フェーズの詳細な手順は図 3 に示す。最後に行動統合フェーズでは各部位から出力された動作情報を統合して 1 つの複合的動作とする。以下に、各フェーズについて詳しく説明する。ただし、本研究ではユーザの右手、左手、腰、右足、左足の 5 か所に 3 軸加速度センサを装着している環境を想定する。

3.1 カテゴリ分類フェーズ

カテゴリ分類フェーズではセンシングデータ取得時にデータに変化があるかを判定する。現在時刻を $t = T$ とすると、時刻 $t = T$ から過去 20 サンプル (1 秒) のセンシングデータの移動平均 $\bar{x}(t)$ と現在のセンシングデータ $x(t)$ の差の絶対値が閾値 ε を超えた場合に、装着者に動きがあると判定し、それ以外は姿勢と判断する。

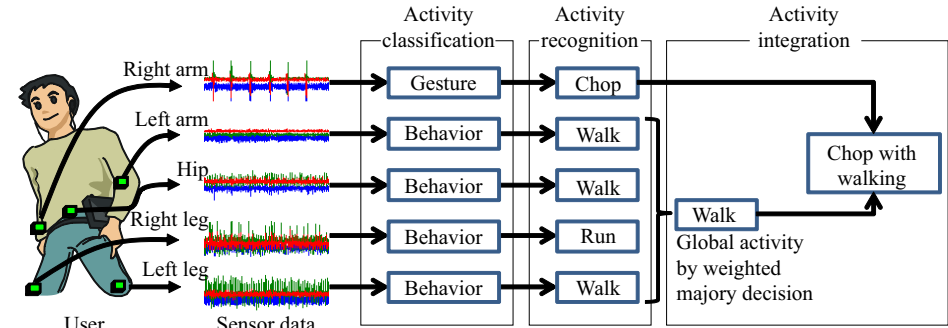


図 2 認識フロー
Fig. 2 Recognition flow.

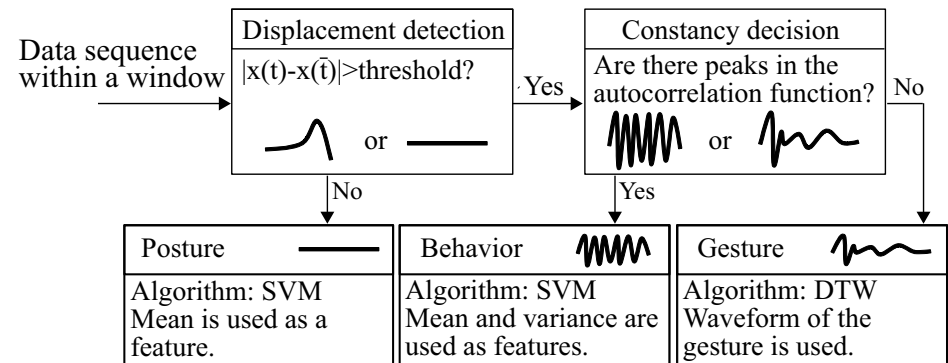


図 3 動作分類と動作認識の手順
Fig. 3 Procedure of activity classification and activity recognition.

$$\begin{aligned} \text{if } |x(T) - \bar{x}(T)| > \varepsilon &\Rightarrow \text{Behavior or Gesture} \\ \text{otherwise} &\Rightarrow \text{Posture} \end{aligned} \quad (1)$$

この移動平均 $\bar{x}(t) \pm \varepsilon$ の領域を Epsilon Tube と呼ぶ。走りながら手は動かさない場合、手を姿勢と分類したいが、足の振動が手に伝搬するため、本研究では ε を $\max\{std(\log(x))/3, 200\}$ と設定した。ただし、 $std(\log(x))$ は両足のセンサの 1 秒間の標準偏差であり、走るなどの動作を中は大きな値を示し、手が姿勢と判定されやすくしている。また、運動中でも現在

のデータ $x(T)$ が一時的に Epsilon Tube 内に存在する場合があるため、連続 0.25 秒以上 Epsilon Tube 内にとどまらない限り運動と判断する．さらに、静止中でも接触などの衝撃によって現在のデータ $x(T)$ が一時的に Epsilon Tube 外に出る場合があるが、一瞬のセンシングデータの変動は装着者が意図しないものであると判断し、1.6 秒以内の乖離は無視する．これらの値は予備実験の結果から設定した．

データに変動がある場合は定常性の有無を判定することで、運動かジェスチャに分類する．「歩く」や「走る」、「自転車に乗る」、「階段昇降」などの行動はある一定のデータの繰返しからなる．定常性検出フェーズでは、自己相関関数 (ACF:Auto-CorrelationFunction) を用いて装着者の動作内の繰返しを検出し、運動かジェスチャかの判断をする．入力信号に繰返しがある部分で ACF の値が大きくなる．つまり、ACF に周期的なピークが現れることは、入力信号も周期的であり、装着者が周期的な運動をしていることを意味する．離散値入力信号 $x(t)$ のずれ τ における ACF は次式で定義される．

$$R_{xx}(\tau) = \sum_{t=0}^{N-1} x(t)x(t-\tau) \quad (2)$$

ただし、 N は ACF を計算するウィンドウサイズで、本論文では 64 サンプル (3.2 秒) とした． N が小さすぎると繰返しをとらえることができず、大きすぎると計算負荷が大きくなるためこの値とした．さらに、自己相関は、ずれ $\tau = 0$ において最大値を示すため、次式に従い $[-1, +1]$ に正規化した．

$$R'_{xx}(\tau) = R_{xx}(\tau)/R_{xx}(0) \quad (3)$$

自己相関の計算後、定常性の有無を判断する必要がある．図 4 に「歩いて、チョップをして、歩く」動作の加速度波形および歩行とチョップの各自己相関関数を示す．歩行時は一歩に対応する波形が繰返されているが、チョップは 1 回きりの動作であり繰返しは見られない．ただし、本研究では 1 回きりの動作の繰返しはその動作をしている状態と見なすため運動と定義する．図より定常性を有する行動の自己相関には明らかなピークが現れているが、定常性を有さない行動の自己相関には高いピークは現れていない．本研究では、第一のピーク $R'_{xx}(n)(n > 0)$ の高さが $\alpha \cdot (1 - n/N)$ を超える場合に定常性ありと判断する．

$$\begin{aligned} \text{if } R'_{xx}(n) &\geq \alpha \cdot (1 - n/N) &\Rightarrow & \text{Behavior} \\ \text{otherwise} &&\Rightarrow & \text{Gesture} \end{aligned} \quad (4)$$

ここで、ずれ τ の増加に伴い自己相関は線形的に減少し、正弦波などの完全な周期波形の

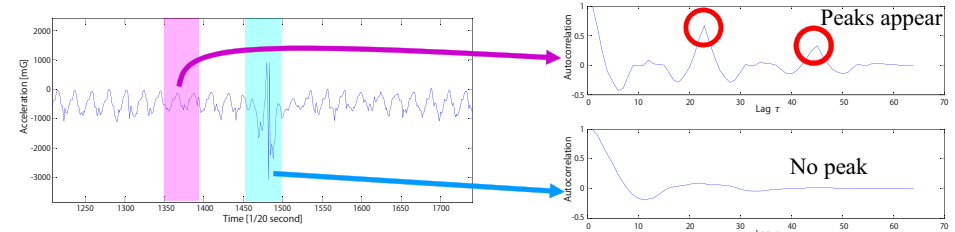


図 4 歩行中にチョップをした加速度波形 (左) および歩く (左上) とチョップ (右下) の自己相関関数
Fig. 4 Accelerations of chop gesture while walking (left) and autocorrelation of walking (upper right) and chop (lower right).

ピークの高さは $1 - n/N$ となる．ただし、 $\alpha(0 \leq \alpha \leq 1)$ は係数であり、本論文では 0.6 とした．この値は運動中にジェスチャを行うデータに対して α を 0.5 から 1.0 まで 0.1 刻みで定常性判定を行った結果から決定した． α が大きすぎる場合、定常性が現れにくくなり、運動時をジェスチャと誤判定する割合が増加する．逆に α が小さすぎる場合、定常性が現れやすくなり、ジェスチャが検出されない割合が増加する．

3.2 行動認識フェーズ

行動認識フェーズではカテゴリ分類フェーズの分類結果に応じて適切な手法で動作を認識する．カテゴリが姿勢の場合は一定長のウィンドウ内のデータの平均値を特徴量として、姿勢のみを学習した SVM⁽¹⁵⁾ で認識する．姿勢を取っている状態でのセンシングデータの分散値はほぼ 0 であるため、平均値のみを用いる．運動中に体の一部分が姿勢を取っている場合も、その部位の向きが識別に影響を与えるため分散値はノイズであり、使用しない．カテゴリが運動の場合はウィンドウ内の平均値と分散値を特徴量として、運動のみを学習した SVM で認識する．カテゴリがジェスチャの場合はウィンドウ内のデータを切り取り、ジェスチャのみを学習した DTW⁽⁸⁾ で認識する．このように認識することで、走りながら携帯電話を把持している動作でも、手は姿勢と判定されれば姿勢の中から動作が選択され、「走る」と認識されることは無い．以下に、姿勢と運動を認識するためのサポートベクタマシン (SVM)、およびジェスチャを認識するためのダイナミックタイムワーピング (DTW) について述べる．

3.2.1 サポートベクタマシン

SVM は現存する認識器の中で最も性能が良いとされるものの一つである⁽¹⁵⁾．学習データセット $(x_1, y_1), (x_2, y_2), \dots, (x_J, y_J)$ を 2 クラスに分類すると想定する．ここで、 $x_i \in R^N$

および $y_i \in \{-1, +1\}$ は特徴量ベクトルおよびクラスラベルである．この 2 クラスを超平面 $w * x_i + b$ で分割可能であると仮定し，データの分散に関して事前知識が得られていないものとする．この 2 クラスを分割する超平面と超平面に最も近いデータとの距離をマージンと呼び，マージンが最大になるときに識別超平面は最適となる．最適超平面の w および b は次式で表される最小化問題を解くことで得られる．

$$\min \frac{1}{2} \|w\|^2 \quad \text{subject to } y_i(w * x_i + b) \geq 1, \forall i = 1, \dots, n \quad (5)$$

上式に対し Lagrange の未定乗数法を用いると次式の識別関数が得られる．

$$f(x) = \text{sign} \left(\sum_{i=1}^n \lambda_i y_i x_i * x + b \right) \quad (6)$$

ここで，学習データの大部分の Lagrange 未定乗数 λ_i が 0 となり， $\lambda_i > 0$ となる x_i のみで識別関数が決定され，このような x_i をサポートベクトルと呼ぶ．また，分離不可能な場合，Lagrange 未定乗数を $0 \leq \lambda_i \leq C (i = 1, \dots, n)$ に修正する．ここで C は誤識別に対するペナルティである．この処理をソフトマージンと呼び，サポートベクタマシンの性能が良い理由の一つである．

非線形識別および線形識別におけるいくつかの方式を実装し予備実験を行ったところ， $C = 50000$ の線形 SVM が最も良い性能を示したため評価は線形 SVM で行った．本来 SVM は 2 クラス分類器であるが，1 クラスとそれ以外のクラスを認識する SVM を全てのクラスに対して構築することで N クラスを分類する SVM を実現した．

3.2.2 ダイナミックタイムワーピング

時系列データは科学や医学，経済学，工学など様々な分野で取り扱われており，これらの分野ではデータマイニングのために時系列データの類似度の計算が求められてきた．類似度計算の最も簡単な手法はユークリッド距離であるが，時間的歪みの影響を受けやすく，比較するデータのサンプル数が同じでなければならないなどの欠点がある．DTW はユークリッド距離のもつ欠点を克服した時系列データの類似度計算手法である．DTW の特徴として，時間的非線形圧縮距離を計算することで時系列のスケールや部分的な速度に違いがあるデータ間やサンプル数の異なるデータの間の類似度が求められる．たとえば，「空中に円を描く」データが 2 つある場合，片方が早く回し，もう一方が遅く回したとしても DTW は両者が類似していると判定する．仮に，円の一部分のみを早く描いたとしても非線形伸縮距離を求めるため，両者は高い類似度を示す．DTW と同様に時系列データの認識を行うアルゴリズム

として HMM が挙げられる．両者の違いは DTW は学習モデルをテンプレートとして保持するのに対し，HMM では遷移確率やシンボル出力確率として保持する点である．そのため，DTW は学習したデータに対して強く，HMM は未学習のデータに対して強いという特徴をもつ．しかし，HMM のモデル構築には DTW と比較して多くのデータサンプルが必要であり，センサを用いた行動認識において各ジェスチャのデータを大量に採取することは困難であると考え本論文では DTW を採用する．

具体的なアルゴリズムを以下に示す．長さ m の学習データ (テンプレート) $X = (x_1, \dots, x_m)$ と長さ n の認識対象のデータ $Y = (y_1, \dots, y_n)$ を比較する際， $m \times n$ 行列 $d(x_i, y_j) = (x_i - y_j)^2$ を定義する．次に， X および Y のインデックスのペアから成るワーピングパス $W = (w_1, \dots, w_k)$ を探索する．ワーピングパスは以下の 3 つの条件を満足する．

- 境界条件

$$w_1 = (1, 1), w_k = (m, n)$$

- 連続性

$$w_k = (a, b), w_{k-1} = (a', b') \Rightarrow a - a' \leq 1 \wedge b - b' \leq 1$$

- 単調性

$$w_k = (a, b), w_{k-1} = (a', b') \Rightarrow a - a' \geq 0 \wedge b - b' \geq 0$$

距離最小のパスを探索するために，以下の手順を適用する．

- (1) Initialization:

$$DTW(0, 0) = 0$$

$$DTW(i, 0) = \infty \text{ for } i = 1, \dots, m$$

$$DTW(0, j) = \infty \text{ for } j = 1, \dots, n$$

- (2) Do for $i = 1, 2, \dots, m$:

Do for $j = 1, 2, \dots, n$:

$$DTW(i, j) = d(x_i, y_j) + \min \begin{cases} DTW(i-1, j-1) \\ DTW(i-1, j) \\ DTW(i, j-1) \end{cases}$$

- (3) Output:

Return $DTW(m, n)$

上式によって得られたコスト $DTW(m, n)$ が X と Y の距離となる．コンテキストを認

識する際は、波形と正解ジェスチャのラベルの組である複数のテンプレートを保持しておく。入力データとすべてのテンプレートとの距離を計算し、最短距離のテンプレートに付与されているジェスチャラベルを結果として出力する。

3.3 動作統合フェーズ

体の各部位の動作が出力されるため、それらを統合して複合的な 1 つの動作を出力する必要がある。しかしながら、複数のセンサを装着している場合、各部位の認識結果がすべて一致するとは限らないため、それらの結果を統合する手法を述べる。本研究ではローカル動作は右手のみで発生すると想定しており、右手の動作認識結果が手の姿勢あるいは手のジェスチャである場合はそれをローカル動作とし、そうでなければローカル動作は無しとする。次に、グローバル動作を各部位の認識結果の重み付き多数決で決定する。重みは学習時に学習データを分類した際の Recall を用いる。各部位は 1 票の投票権をもち、各部位が認識した結果に 1 票を投じる。その際、各部位における投票した行動の Recall を重みとして乗じる。例えば、腰のセンサが「歩く」と認識し、腰のセンサの「歩く」の Recall が 0.85 であれば、歩くに 0.85 票が入る。このようにして各部位からの投票が最も多かった認識結果をグローバル動作とする。この投票において、右手が既にローカル動作である場合は右手は投票に参加しない。最終的に得られたグローバル動作とローカル動作を統合して 1 つの複合的動作を出力する。

4. 評価

本節では提案手法および比較手法について複合的動作の識別性能の評価を行う。

4.1 評価環境

評価では右手首、左手首、腰、右足首、左足首の 5 か所に 3 軸無線加速度センサ¹⁶⁾を装着した 22 歳～28 歳の被験者 3 名から採取したデータを用いて、オフラインで評価した。サンプリング周波数は 20Hz である。採取した行動は 4 種類の姿勢（座る、立つ、横になる、片膝立ち）、5 種類の運動（歩く、走る、自転車、階段を上る、階段を下りる）、5 種類のジェスチャ（チョップ、ボールを投げる、パンチ、時計回りの円を描く、反時計回りの円を描く）、2 種類の手のパーズ（携帯電話を持って見る、手を挙げる）である。また複合的動作の学習および評価用として、座る、立つ、歩く、走る、自転車の動作を行いながら 5 種類のジェスチャと 2 種類の手のパーズを各組合せで 10 回ずつ行った。収集したデータは手作業で正解のラベル付けを行った。

評価では SVM のみ、DTW のみ、提案手法の 3 種類の認識手法を用いて複合的動作の認

識を行う。前者 2 手法は単純にすべての複合的動作を学習する比較手法である。すべての動作を学習しているため、目標とすべき認識精度となる。後者は動作のカテゴリに応じて SVM と DTW を使い分けて体の部位ごとに動作を認識してから統合する提案手法である。比較手法で用いる認識器はすべての複合的動作を学習しているが、提案手法で用いる認識器はカテゴリの単体の動作しか学習していない。複合的動作の学習に用いたデータは採取したデータの 20% で、残りの 80% を評価用データとした。

4.2 結果と考察

複合的動作を 3 種類の認識手法で認識した場合の Recall および Precision を表 1 に示す。ローカル動作の列の“Null”はグローバル動作のみを行っていることを意味する。表より、SVM の結果が低く、特に運動中のジェスチャ動作の認識精度はほぼ 0 である。これは特徴量は姿勢の向きや運動の強度に関する情報はもっているが、動き方に関する情報をもたないためである。つまり、運動中の各動作および運動のみの動作を姿勢の向きや運動強度のみで判別することは困難であり、すべてが運動のみと誤認識されている。立たちながらのジェスチャおよび座りながらのジェスチャの認識精度が低いことも同じ理由である。加えて、座るのみの動作が正しく認識されていない。これは座りながらジェスチャをする際に被験者らが手が肘掛けに当たらないように体を少し後ろに反らし、学習していた座る動作と異なったためと考えられ、さらに、座りながらのジェスチャ動作は座りながらのジェスチャ動作と比較して遅いため分散が小さいため、動作の影響が小さくなり、体全体として座っているだけの状態でも何かしらのジェスチャを行っているとして誤認識された。走りながらの手のパーズはすべて走るのみと誤認識されている。これは手の向きの違いが走る振動のぶれに吸収されたためと考えられる。これらの結果より、類似した多種の複合的動作を特徴量空間に落として認識することは困難であるといえる。

これに対して DTW はすべての動作に対して高い性能を示している。しかし、注目すべき点はすべての複合的動作を学習した DTW に匹敵する性能を提案手法が示していることである。DTW は 9 種のグローバル動作と 7 種のローカル動作の組合せ 63 通りすべてを学習しているのに対して提案手法はその和である 16 通りしか学習していない。結果から得られた提案手法の欠点として、走りながら携帯電話を把持する動作と走りながら手を挙げる動作の認識がある。これは、走っている際の足の振動が大きく、カテゴリ分類フェーズにおいて走りながらの手のパーズが姿勢に分類されなかったためである。カテゴリ分類フェーズにおける Epsilon Tube の ε を足の振動を考慮して動的に設定したが、歩行時の振動を基準に設定したが、走行時の振動が手に伝搬する割合が大きかったことが原因である。ただし、追加

表 1 認識結果の Recall と Precision
Table 1 Recall and precision of recognition.

Activity		Proposal		SVM		DTW	
Global	Local	Recall	Precision	Recall	Precision	Recall	Precision
Stand	Chop	1.000	0.917	0.925	0.633	1.000	0.950
	Throw	1.000	1.000	0.839	0.942	0.996	1.000
	Punch	1.000	1.000	0.874	0.989	0.964	0.996
	Clockwise	1.000	1.000	0.397	0.426	1.000	1.000
	Anticlockwise	1.000	0.882	0.938	0.661	1.000	0.920
	Hold mobile	0.996	-	1.000	-	1.000	-
	Raise hand	1.000	-	0.996	-	1.000	-
	Null	0.999	-	0.919	-	0.993	-
Sit	Chop	1.000	1.000	0.736	0.866	0.993	0.908
	Throw	0.955	1.000	0.840	0.296	0.842	0.983
	Punch	1.000	1.000	0.880	0.680	0.920	0.625
	Clockwise	1.000	1.000	0.831	0.503	0.805	0.914
	Anticlockwise	1.000	0.955	0.855	0.473	1.000	0.471
	Hold mobile	1.000	-	1.000	-	1.000	-
	Raise hand	1.000	-	1.000	-	1.000	-
	Null	1.000	-	0.000	-	0.522	-
Walk	Chop	1.000	1.000	0.000	0.000	0.994	1.000
	Throw	1.000	1.000	0.000	0.000	0.946	1.000
	Punch	0.929	0.833	0.000	0.000	0.983	1.000
	Clockwise	0.944	0.944	0.000	0.000	0.993	0.995
	Anticlockwise	1.000	0.944	0.000	0.000	0.990	0.990
	Hold mobile	0.705	-	0.975	-	1.000	-
	Raise hand	0.442	-	0.956	-	0.999	-
	Null	0.988	-	0.956	-	0.537	-
Run	Chop	0.917	0.826	0.000	0.000	0.986	1.000
	Throw	1.000	0.975	0.000	0.000	0.894	0.997
	Punch	0.429	0.429	0.000	0.000	0.966	1.000
	Clockwise	0.500	0.500	0.000	0.000	0.978	0.916
	Anticlockwise	0.000	0.000	0.000	0.000	0.975	0.942
	Hold mobile	0.000	-	0.000	-	0.886	-
	Raise hand	0.000	-	0.000	-	0.734	-
	Null	0.981	-	0.875	-	0.309	-
Bike	Chop	0.900	1.000	0.000	0.000	0.729	0.956
	Throw	0.913	0.917	0.500	0.500	0.872	0.843
	Punch	1.000	0.742	0.000	0.000	0.930	0.711
	Clockwise	0.622	1.000	0.000	0.000	0.780	0.908
	Anticlockwise	0.500	0.450	0.000	0.000	0.937	0.686
	Hold mobile	1.000	-	1.000	-	1.000	-
	Raise hand	0.999	-	1.000	-	1.000	-
	Null	0.991	-	0.997	-	0.878	-

実験から手が正しく姿勢と分類されていれば正しく携帯電話を把持すると手を上げるとして認識されていたため、線形の閾値ではなく非線形の閾値を設定することで解決できると考えられる。

これらの結果から、提案手法が単体の動作のみの学習で複合的動作を認識できることを確

認した。提案手法によって学習データの収集やラベル付けに要する時間を削減できるだけでなく、学習するデータの種類が減ることで学習および認識に要する時間も削減できる。

4.3 複数の部位のローカル動作の認識

最後に、本研究ではローカル動作を右手のみに限って評価を行ったが、左手や足など複数の部位がローカル動作を行う場合について言及する。複数の部位がローカル動作を行う場合でも提案手法の流れは変わらないが、右手のジェスチャの学習データで左手の同じジェスチャを認識することはできないため、すべての部位のジェスチャおよび姿勢のデータを採取する必要がある。

ローカル動作とグローバル動作の組合せは以下の 6 通りが考えられる。

- (1) ローカル動作はすべて姿勢，グローバル動作は姿勢
例) 右手は携帯電話を把持し，左手を挙げて，立っている
- (2) ローカル動作はすべて姿勢，グローバル動作は運動
例) 右手は携帯電話を把持し，左手を挙げて，歩いている
- (3) ローカル動作はすべてジェスチャ，グローバル動作は姿勢
例) 右手はパンチし，左手はチョップし，立っている
- (4) ローカル動作はすべてジェスチャ，グローバル動作は運動
例) 右手はパンチし，左手はチョップし，歩いている
- (5) ローカル動作は姿勢とジェスチャ，グローバル動作は姿勢
例) 右手はパンチし，左手を挙げて，立っている
- (6) ローカル動作は姿勢とジェスチャ，グローバル動作は運動
例) 右手はパンチし，左手を挙げて，歩いている

(1) および (2) のローカル動作がすべて姿勢，つまりどの部位もジェスチャを行っていない場合は，姿勢をとっている部位は姿勢に分類され，姿勢のみを学習した SVM で各部位を認識する。右手を認識する SVM は「携帯電話を把持して立っている」と認識し，左手の SVM は「手を挙げて立っている」と認識する。残りの部位は「立っている」や「歩いている」と認識されるため，正しく複合的動作を認識できる。一方，(3)～(6) のローカル動作にジェスチャが含まれる場合は，ジェスチャを行っていない部位に振動が伝わることで，その部位がジェスチャを行っていると誤認識される。ローカル動作を行う部位を右手のみに制限している場合は，カテゴリ分類フェーズにおいて右手以外の部位のカテゴリをジェスチャに割り当てないことで右手以外の誤認識を防いでいる。例えば「立ちながら右手を回した」場合，右手の振動が左手に伝わり，その振動が大きいと左手はジェスチャに分類される。(3)

および (4) のようにすべてのローカル動作がジェスチャである場合は、左手がジェスチャに分類されても問題ない。振動がノイズとして影響し、誤認識する可能性はあるが、正しくジェスチャを認識しようとする。しかし (5) および (6) のようにポーズを取っている手や歩いている際の足がジェスチャに分類された場合は、強制的に学習したジェスチャのいずれかが認識結果となる。これに対処するには、認識結果とする正解のデータの DTW の距離に閾値を設けて、正解データとの距離が遠い認識結果は出力しないとする対応が考えられる。ただし、誤認識結果が抑制される一方で正しい認識結果が破棄されるというトレードオフの関係にあるため閾値の設定に注意する必要がある。そのほかの対処法として、右手でジェスチャを行っている際の各部位のデータを各部位のジェスチャを認識する DTW の学習データに含める方法が挙げられるが、左手への振動のみから右手のジェスチャの判別や、歩行中の足に伝わった振動から手のジェスチャの判別は困難であると考えられる。

5. ま と め

本研究では単体の動作を学習するだけでそれらの組合せから成る複合的動作を認識するシステムを提案した。提案システムではセンサを装着している体の各部位を姿勢、運動、ジェスチャに分類した後、それらの分類に応じた適切な手法で認識し、各部位の認識結果を重み付き多数決によって統合して複合的動作を出力する。

今後は、複合的動作を行っている波形から単体の動作の抽出や分離および単体の動作の波形から複合的動作の波形の生成など、より汎用的なアプローチで複合的動作の認識システムを提案する予定である。

謝辞 本研究の一部は、科学技術振興機構戦略的創造研究推進事業（さきがけ）および文部科学省科学研究費補助金基盤研究（A）(20240009, 23240010) によるものである。ここに記して謝意を表す。

参 考 文 献

- 1) Bao, L. and Intille, S. S.: Activity Recognition from User-annotated Acceleration Data, *Intl. Conference on Pervasive Computing (Pervasive 2004)*, pp.1–17 (2004).
- 2) Chambers, G. S., Venkatesh, S., West, G. A. W., and Bui, H. H.: Hierarchical Recognition of Intentional Human Gestures for Sports Video Annotation, *Intl. Conference on Pattern Recognition (ICPR '02)*, pp.1082–1085 (2002).
- 3) Junker, H., Amft, O., Lukowicz, P., and Tröster G.: Gesture Spotting with Body-worn Inertial Sensors to Detect User Activities, *Pattern Recognition*, pp.2010–2024

- (2008).
- 4) Keogh, E., Chu, S., Hart, D., and Pazzani, M.: An Online Algorithm for Segmenting Time Series, *Intl. Conference on Data Mining (ICDM 2001)*, pp.289–296 (2001).
- 5) Laerhoven, K. V. and Gellersen, H. W.: Spine versus Porcupine: a Study in Distributed Wearable Activity Recognition, *Intl. Symposium on Wearable Computers (ISWC 2004)*, pp.142–149 (2004).
- 6) Lester, J., Choudhury, T., and Borriello, G.: A Practical Approach to Recognize Physical Activities, *Intl. Conference on Pervasive Computing (Pervasive 2006)*, pp. 1–16 (2006).
- 7) Murao, K. and Terada, T.: A Motion Recognition Method by Constancy-Decision, *Intl. Symposium on Wearable Computers (ISWC 2010)*, pp.69–72 (2010).
- 8) Myers, C. S. and Rabiner, L. R.: A Comparative Study of Several Dynamic Time-warping Algorithms for Connected Word Recognition, *The Bell System Technical Journal*, Vol.60, pp.1389–1409 (1981).
- 9) Naya, F., Ohmura, R., Takayanagi, F., Noma, H., and Kogure, K.: Workers' Routine Activity Recognition using Body Movement and Location Information, *Intl. Symposium on Wearable Computers (ISWC 2006)*, pp.105–108 (2006).
- 10) Ouchi, K., Suzuki, T., and Doi, M.: LifeMinder: A wearable Healthcare Support System Using User's Context, *Intl. Workshop on Smart Appliances and Wearable Computing (IWSAWC 2002)*, pp.791–792 (2002).
- 11) Pärkkä, J., Ermes, M., Korpipää, P., Mäntylä, J., Peltola, J., and Korhonen, I.: Activity Classification Using Realistic Data From Wearable Sensors, *Trans. on Information Technology in Biomedicine*, Vol.10, pp.119–128 (2006).
- 12) Shen, C. L., Kao, T., Huang, C. T., and Lee, J. H.: Wearable Band Using a Fabric-Based Sensor for Exercise ECG Monitoring, *Intl. Symposium on Wearable Computers (ISWC 2006)*, pp.143–144 (2006).
- 13) Stiefmeier, T., Ogris, G., Junker, H., Lukowicz, P., and Tröster, G.: Combining Motion Sensors and Ultrasonic Hands Tracking for Continuous Activity Recognition in a Maintenance Scenario, *Intl. Symposium on Wearable Computers (ISWC 2006)*, pp.97–104 (2006).
- 14) Toda, M., Akita, J., Sakurazawa, S., Yanagihara, K., Kunita, M., and Iwata, K.: Wearable Biomedical Monitoring System Using TextileNet, *Intl. Symposium on Wearable Computers (ISWC 2006)*, pp.119–120 (2006).
- 15) Vapnik, V.: *The Nature of Statistical Learning Theory* (1995).
- 16) Wireless Technologies Inc.: <http://www.wireless-t.jp/>.