

ユーザ非依存なジェスチャ認識のための 動作の実施時間と勢いの影響

岩崎正裕[†] 藤波香織[‡]

[†] 東京農工大学大学院 工学府情報工学専攻

[‡] 東京農工大学大学院 工学研究院 先端情報科学部門

1 はじめに

ジェスチャ認識において、認識を行うユーザ以外から得られたデータを学習データとして用いる場合がある。ユーザ非依存な認識とはこのような場合における高精度での認識のことである。ユーザ非依存な認識が実現すれば、新たなユーザの学習データが不必要となる。その結果、ユーザに対する学習データ取得の負担は軽減される。

多くのジェスチャ認識手法がユーザ非依存な条件で精度低下するのは、学習データの特徴がユーザ間で異なることが影響していると考えられる。このユーザ間の差として、動作の実施時間と勢いの変化が挙げられる。実施時間の増減により時系列データのサンプル数は変化し、波形は時系列方向へ伸縮する。また、勢いの強弱により、例えば加速度センサのデータは振幅が増減する。そこで、動作の実施時間と勢いの変化がデータに与える影響に対して正規化を行うことで、学習データ間の個人差を吸収できると考えた。

以上の議論を元に、本研究では加速度センサを用いたジェスチャ認識手法の高精度化を目的とする。その中でも特に、動作の実施時間と勢いの変化を考慮したユーザ非依存な認識に対して着目する。

2 提案手法

2.1 手法概要

まず、本研究では腕の動きに着目したジェスチャ動作を対象とする。また、ジェスチャ認識に用いられるデータは大きく分けて2種類あり、慣性センサの手首への装着や把持により取得されるデータと、環境側に設置したカメラから得られる画像である。カメラを用いる場合、ジェスチャ動作時にカメラの前に存在する必要があることがユーザに対して制約となる。自由なインタラクションが必要とされる場合を考え、本稿で

は手首に取り付けた加速度センサを用いる方法をとる。

多くのジェスチャ認識に関する研究では DTW や HMM が用いられている^[1]。一方、Akl, et al. は圧縮センシング技術に着目した認識手法を提案した^[1]。そして、Akl, et al. はユーザ非依存な認識において、彼らが提案した手法が DTW や HMM より精度が高いことを示した^[1]。そこで我々の提案手法はその手法を基本とする。本研究ではこの手法に加えて、ユーザ間や同一ユーザにおいても動作の反復時に変化すると考えられる実施時間と勢いについて正規化を行う。この処理により、個人差に対して頑健になり、他人の学習データを用いても高精度で認識が実現できると考えた。

2.2 Akl, et al.^[1] の手法概要

まず、ランダム射影によって圧縮された学習データが各列に格納された行列 R と、同様に圧縮された入力データを1列に格納した行列 y を用いて、 $y = Rx$ という式を考える。この時、 R の1行目に格納されたデータが y と同一であった場合、 x は $\{1, 0, 0, \dots, 0\}^T$ と表現できる。このように、各学習データをそれぞれ定数倍し、その和で入力データを表現する。この際に、複数得られる定数倍の組み合わせから、各次元の値の絶対値の和が最小となる解が x となる。この係数 x を、入力データと各学習データの類似度として利用する。

x については、式(1)に示す L1 ノルム最適化問題を解くことで得る。L1 ノルム最適化問題はフリーライブラリである `lp_solve`^[2] を用いて解く。そして、次節で述べるように、提案手法では学習データと入力データに対して正規化を実施する。

$$\arg \min \|x\|_1 \text{ subject to } y = Rx \quad (1)$$

2.3 動作の実施時間の変化に対する正規化

動作の実施時間が増減すると、加速度センサのデータ長も伴って伸縮する。そのため、同じ軌跡をたどった動作でも各データの周波数特性は異なり、データを比較する際に影響することが考えられる。そこで、最もデータ長が短いデータに合わせてデータ長を縮めることで実施時間の変化に対する正規化を行う。

データ長はデータのダウンサンプリングによって

An impact of motion duration and acceleration toward user-independent gesture recognition system

Masahiro IWASAKI[†] Kaori FUJINAMI[‡]

^{†,‡}Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology

縮める。まず、ダウンサンプリングされるデータのデータ長を A 、入力データと全ての学習データの中で最も短いデータ長を B とする。そして、データの $i = 0, 1, 2, \dots, A-1$ 番目の要素のうち、式 (2) で表現される要素のみを利用することでデータ長を縮める。

$$i = [(A/B) \times n + 0.5] \quad (n = 0, 1, 2, \dots, B-1) \quad (2)$$

また、この処理の前に指数移動平均を用いたローパスフィルタをかける。これは加速度センサの計測誤差などによって得られる高周波ノイズと、ダウンサンプリング時における折り返しノイズを考慮するためである。

2.4 動作の勢いの変化に対する正規化

動作時の勢いが増減すると、加速度センサのデータ波形における振幅は増減する。振幅の増減は、学習データの定数倍で入力データを表現する際に影響することが考えられる。そのため、振幅の正規化が必要であると考へた。この時、各軸に対して別々に正規化を行うと、手の揺れなどによって生じる微弱な波形の振幅と、動作によって生じる波形の振幅を同様に正規化してしまう。この問題は、各軸の値をまとめることで対処する。

振幅に対する正規化として、まず加速度センサの各軸の値をまとめる。つまり、データ長 B のデータ 3 軸分を 1 列にまとめ、 $B \times 3$ の長さの列に、 X 軸、 Y 軸、 Z 軸の順番に格納する。そして、3 軸をまとめたデータに対して平均値を求め、全ての値から平均値を減算することで平均値を 0 にする。その後、その最大値と最小値を求め、全てのデータを最大値 1、最小値 -1 に線形的に正規化する。

3 評価実験

多様なジェスチャにおける提案手法の頑健性を評価するため、以下の 3 種類のデータセットを用いた。

- A. 空中に図形などを描画する 18 動作 $\times$ 4 人 [3]
- B. 空中に図形などを描画する 8 動作 $\times$ 8 人 [4]
- C. 携帯電話の向きを変えるなどの 27 動作 $\times$ 7 人 [5]

また、認識手法は以下の 4 種類で試行した。

- 1. Akl, et al. [1] と同様な手法
- 2. 1 に加えて 2.3 節の処理を加えた手法
- 3. 1 に加えて 2.4 節の処理を加えた手法
- 4. 1 に加えて 2.3 節と 2.4 節の処理を加えた手法

各データセットにおいて、一人分の各データをテストデータとし、他の人の全データを学習データとして利用する。これを全被験者が一回ずつテストデータとなるよう試行を繰り返した。また、分類精度に対する F 値の被験者間の平均値を評価に用いる。 F 値は式 (2) によって求める。 F 値により適合率と再現率の調和平均を求めることで総合的な精度評価が可能である。

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (3)$$

3.1 結果

図 1 は各手法における各データセット利用時の実験結果で、 F 値の平均、最大値、最小値を示している。まず、各条件における手法 1 の結果をみると、利用するジェスチャによって認識精度が異なることがわかった。次に、提案手法の有効性について、データセット A と C に対しては手法 2 < 手法 3 であるが、データセット B に対しては手法 2 > 手法 3 であった。また、全てのデータセットに対して、平均値と最小値において手法 4 が最も高かった。このことから、2.3 節や 2.4 節のどちらかの正規化のみを用いるのではなく、どちらも組み合わせることが有効であることが示唆された。

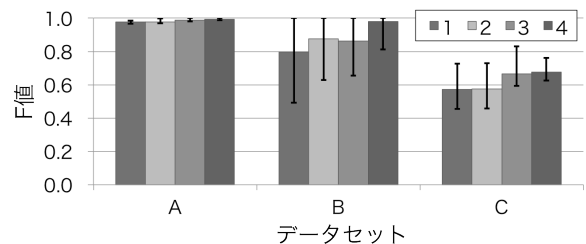


図 1: 実験結果

4 結論

本研究では、加速度センサを用いたジェスチャ認識手法の高精度化を目的とした。そして、その中でも特に個人差を考慮したユーザ非依存な認識に対して着目した。実験の結果から、動作の実施時間と勢いの変化を考慮することが、ユーザ非依存な認識においての精度向上に有効であることがわかった。提案手法はジェスチャ区間が切りだされていることを前提としているため、今後はジェスチャ切り出しの実装と評価を行う。

謝辞

本研究は科学研究費補助金基盤研究 (A)23240014 の助成を受けたものである。

参考文献

- [1] A. Akl, et al.: Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, and compressive sensing. In *Proc. of ICASSP2010*, pp. 2270–2273, 2010.
- [2] Free Software Foundation. lp_solve reference guide. [http://lpsolve.sourceforge.net/5.0/\(2014-1-14 アクセス\)](http://lpsolve.sourceforge.net/5.0/(2014-1-14 アクセス)).
- [3] A. Akl: Wireless and Internet Research Lab. [http://www.wirlab.utoronto.ca/wirlab/Members/ahmadakl84/ahmad-akl-masc-candidate\(2014-1-14 アクセス\)](http://www.wirlab.utoronto.ca/wirlab/Members/ahmadakl84/ahmad-akl-masc-candidate(2014-1-14 アクセス)).
- [4] Z. Wang: uWave. [http://www.owl.net/~zw3/projects/_uWave.html\(2014-1-14 アクセス\)](http://www.owl.net/~zw3/projects/_uWave.html(2014-1-14 アクセス)).
- [5] T. Terada: Project: Constructing mobile phone-based context recognition systems for real world use. [http://ubi.eeddept.kobe-u.ac.jp/research/special/mobilerecognition_e.html\(2014-1-14 アクセス\)](http://ubi.eeddept.kobe-u.ac.jp/research/special/mobilerecognition_e.html(2014-1-14 アクセス)).