

二つの簡易ジェスチャの連続入力によるスマートフォン操作の検討

A Study of Operation for Smartphone by Continuous Input of Two Simple Gestures

網屋 友彰[†] 勝間 亮[†]
Tomoaki Amiya Ryo Katsuma

概要

近年、スマートフォンをはじめコンピュータが小型化している。同時に、ユーザが入力に用いるボタンが減少し小型化することで、ユーザ入力における操作性が低下している。その中で、スマートフォンの加速度センサを利用して、スマートフォンを把持したユーザの手の動きを入力とするジェスチャ入力が新たな入力手法として期待されている。ジェスチャ入力とは、加速度センサが搭載されている機器であれば本体の小型化の影響を受けず入力でき、画面も見ることがないという利点がある。しかし、加速度からジェスチャを認識する際の誤認識が多いことが問題となっている。本稿では、歩行中のスマートフォンの音楽プレイヤーの操作を想定し、誤認識に耐性のある入力として、操作の開始もしくは内容を決定するトリガージェスチャと、実際の入力操作とするメインジェスチャの2つのジェスチャを組み合わせを用いた入力手法を提案する。

1. 導入

近年、スマートフォンをはじめコンピュータが小型化しており、容易に持ち運びできたり身につけたりできる利便性が向上している。その反面、ボタン数の減少や操作画面の縮小により、ユーザ入力における操作性が低下している。現在普及しているスマートフォンでは、操作する際にボタンのように触覚で判別できないため、小さな画面を見ながらタッチパネルを操作しなければならず、周囲への注意が散漫になってしまう。最近ではウォーキングやジョギングのときにスマートフォンの音楽プレイヤーで音楽を楽しむだけでなく、リズムを取ってペースを維持するため等に音データが利用されているが、歩きながらのスマートフォンのタッチパネル操作が危険であるとして社会的な問題となっている。消防庁によると、平成25年から平成29年までの5年間で歩きスマホ等に係る事故により、199人が救急搬送されている[1]。これは救急搬送まで及んだ例であるので、実際はより多くの事故が発生していると考えられる。よってスマートフォンの入力として、本体の小型化に左右されないこと、歩行中は画面を見ずに操作できることが求められている。

スマートフォンには、多くのセンサが搭載されている。中でも加速度センサは、スマートフォンを把持したユーザの手の動きを認識することができる。このときの手の動きをジェスチャといい、ジェスチャによる入力ことをジェスチャ入力という。ジェスチャ入力は、スマートフォンだけでなくゲーム機などでも利用されており、実際に手を動かすという入力では、直観的な操作を可能にしている[2]。またジェスチャ入力は、加速度センサが搭載されている機器であれば本体の小型化の影響を受けず入力でき、画面も見ることがないので、新たな入力手法として期待されている。特に音楽プレイヤーのような必要な入力の種類が少ないアプリケーションでは、ユーザの記憶負担もかからずジェスチャ入力が有効である。

本稿では、歩行中のスマートフォンの音楽プレイヤーの操作を想定し、誤認識に耐性のある入力として、ジェスチャを組み合わせることを提案する。これは、スマートフォンを振るなどの簡単なジェスチャを2つ組み合わせるもので、1つ目を操作の開始もしくは内容を決定するトリガージェスチャ、2つ目を実際の入力操作とするメインジェスチャとする。また、ジェスチャを認識する基本的な方法は、入力となるジェスチャの加速度の時間変化を事前に測定しておき、実際に入力したときの加速度が事前データと近い場合、それを認識したと出力するというものである。このときの事前データと入力データとの類似度の計算に動的時間伸縮法(DTW:Dynamic Time Warping)を用いる。

2. 関連研究

ジェスチャ入力では、誤認識が多いことが問題となっている。ジェスチャの認識精度が悪くなる主な理由は2つ挙げられる。1つ目は加速度データに誤差が多いこと、2つ目はジェスチャの入力時間が毎回同じではないことである。

1つ目の、加速度データに誤差が多いことについて、加速度センサで取得する値は、高いサンプリングレートであっても誤差を多く含んでいる。ジェスチャ認識の場合、事前データと入力データのどちらも誤差を多く含んでいるため、誤認識が多く発生する。村尾らの研究では、より正確な値取得のため、加速度センサを9個使用した[2]。このとき27種類のジェスチャを認識した際の認識

[†] 大阪府立大学, Osaka Prefecture University, Sakai, Osaka
599-8531, Japan

精度は平均 62.7%であることが示されている。本稿では、スマートフォンの加速度センサを用いることを想定しているので、加速度センサの数を増やすことはできない。1つの加速度センサのみで誤差を含む入力精度を上げる必要がある。歩行中の場合、さらに誤検知が多くなることが予想される。

2つ目の、ジェスチャの入力時間が毎回同じではないことについて、同じジェスチャであっても時間や状況によって動作の速さが変わってしまう。そうすると、取得される入力データのサンプル数や加速度が違ったものになるので、事前データとの類似度を求めることが難しくなる。そこで、そのようなデータ間の類似度を計算するアルゴリズムとして DTW がある [3][4]。DTW では、2つの時連続データにおいて、時間を伸縮させることで、サンプル数や加速度の違いを解消し類似度を計算することができる。

3. 問題設定

本稿のねらいは、スマートフォンの加速度センサで取得した値からジェスチャを認識し、それを音楽プレイヤーの入力操作とすることである。本章では、ジェスチャ認識を考えるうえで想定する条件について設定する。

まず、入力として想定される n 個のジェスチャを $gesture_i$ ($i = 1, 2, 3, \dots, n$) とし、事前に測定した $gesture_i$ の加速度データを $G_i = (g_1, \dots, g_{T_i})$ とする。 T_i は事前データ G_i のサンプル数である。そして、実際に入力される加速度データを $A = (\dots, a_t, \dots)$ とする。 a_t は時刻 t で取得された加速度データである。実際、加速度データ g, a は x, y, z の3軸について取得されるため、3つのパラメータを持つベクトルとなる。1つのサンプルにおいて、事前データ g および入力データ a のそれぞれのパラメータを (a_x, a_y, a_z) および (g_x, g_y, g_z) とする。

ジェスチャの認識では、入力データ A と事前データ G_i との類似度が計算される。本稿では、類似度は DTW によって計算され距離 $D(A, G_i)$ として評価される。実際、 A のデータは永続的に取得し続けるので、どこかのデータを抜き取る必要がある。抜き取るデータのサンプル数を、トリガージェスチャでは 2β 、メインジェスチャでは γ とする。距離 D は値が小さいほど2つのデータがより類似していることを示す。よって、 $\min\{D(A, G_i), i = 1, 2, 3, \dots, n\}$ となる $gesture_*$ が出力されることとなる。

本稿では、音楽プレイヤーの操作に焦点を当てることにする。操作に最低限必要な5つの入力を表1に示す。

表 1: 音楽プレイヤーの操作に必要な入力

再生・一時停止
次の曲へ送る
前の曲へ戻す
音量を上げる
音量を下げる

4. 提案手法

4.1 ジェスチャと操作の割り当て

複雑なジェスチャはユーザの負担になってしまうので、それぞれのジェスチャは簡単であるほうが好ましい。そこでまず、組み合わせるジェスチャについて考える。スマートフォンに搭載されている加速度センサは3軸加速度センサである(図1)。各軸の方向の成分に沿った運動をするように、スマートフォンを振るジェスチャを用いる。表2に各軸に沿ったジェスチャを示す。

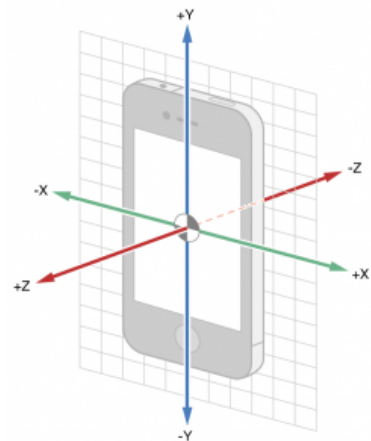


図 1: スマートフォンの加速度センサ

表 2: 各軸におけるジェスチャ

軸	正 (+) 方向	負 (-) 方向
x 軸	右に振る	左に振る
y 軸	上に振る	下に振る
z 軸	手前に振る	奥に振る

ここで、2つのジェスチャを選ぶ組み合わせは 6×6 の36通り存在する。そのうち、音楽プレイヤーの操作に必要な5つの入りに割り当てる。トリガージェスチャの役割に関して、2種類の方式が考えられる。1つ目のパターンは、全てのメインジェスチャに対して同一のトリガージェスチャを用いる組み合わせのみを選ぶ割り当て方で、パターンAとする。2つ目のパターンは、複数の

トリガージェスチャを用いる組み合わせを選ぶ割り当て方で、パターンBとする。パターンAはトリガージェスチャが入力開始の合図の意味を持ち、メインジェスチャが実際の入力内容を指している。パターンBでは、複数あるトリガージェスチャのうち1つを入力した時点で既に入力内容が決まっており、メインジェスチャはその度合いを決定する。各パターンの組み合わせを表3、4に示す。組み合わせは入力内容とジェスチャが直感的に結びつきやすいように決定した。表中のカッコ内は主に反応する軸の方向を示す。

表 3: パターン A

入力	トリガー	メイン
再生・一時停止	手前に振る (z+)	奥に振る (z-)
次の曲へ送る	手前に振る (z+)	右に振る (x+)
前の曲へ戻す	手前に振る (z+)	左に振る (x-)
音量を上げる	手前に振る (z+)	上に振る (y+)
音量を下げる	手前に振る (z+)	下に振る (y-)

表 4: パターン B

入力	トリガー	メイン
再生・一時停止	上に振る (y+)	手前に振る (z+)
次の曲へ送る	手前に振る (z+)	右に振る (x+)
前の曲へ戻す	手前に振る (z+)	左に振る (x-)
音量を上げる	右に振る (x+)	上に振る (y+)
音量を下げる	右に振る (x+)	下に振る (y-)

4.2 ジェスチャの認識方法

4.2.1 システム構造

入力される加速度データを A に対して、以下のシステム構造によって、ジェスチャ入力を認識する。

- (1) トリガージェスチャの入力開始認識
- (2) トリガージェスチャの種類の判別
- (3) メインジェスチャの入力開始認識
- (4) メインジェスチャの種類の判別

4.2.2 ジェスチャの入力開始認識アルゴリズム

実際の入力データ A は永続的に取得し続けるので、ジェスチャ入力が始まったことを認識する必要がある。認識

に必要な数値として、ある時刻の総合加速度 $|a|$ を以下のように定める。

$$|a| = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (1)$$

まず、トリガージェスチャの入力開始認識を考える。現在時刻を t として、過去 β サンプルの総合加速度の平均値をとる。 $|a_t|$ がその平均値より α 大きい値を超えた場合に、トリガージェスチャの入力がなされたとする。よって、トリガージェスチャの入力開始を認識する条件は以下になる。

$$|a_t| > \frac{1}{\beta} \sum_{i=t-\beta}^{t-1} |a_i| + \alpha \quad (2)$$

その後トリガージェスチャの種類が判別された後、メインジェスチャの入力開始認識を行う。ここで、時刻を β サンプル分進める。そこから γ サンプルの間に再び $|a_t|$ が $\frac{1}{\beta} \sum_{i=t-\beta}^{t-1} |a_i| + \alpha$ を超えた場合に、メインジェスチャの入力がなされたとする (図 2)。

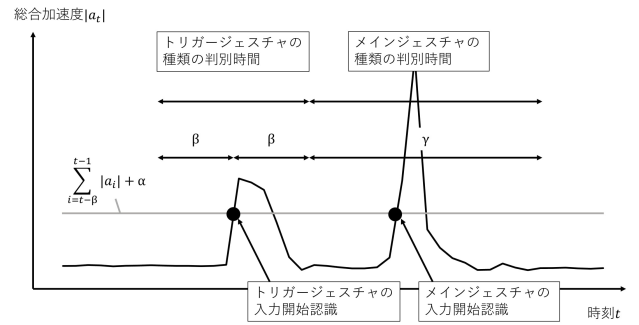


図 2: ジェスチャ認識の時間経過

4.2.3 ジェスチャの種類の判別アルゴリズム

入力データと事前データとの距離を算出することで、ジェスチャの種類を判別する。入力データは、トリガージェスチャでは入力開始を認識した時刻の前後 β サンプル、メインジェスチャではトリガージェスチャの種類の判別が終わってからの γ サンプルを用いる。

入力データ a と事前データ g_i の距離として、ユークリッド距離 d を以下のように定義する。

$$d(a, g) = \sqrt{(g_x - a_x)^2 + (g_y - a_y)^2 + (g_z - a_z)^2} \quad (3)$$

次に DTW を用いた距離の算出方法を示す [5]。行列 $dtw[N][T_i]$ を考える。 N は入力データのサンプル数であり、トリガージェスチャでは $2\beta + 1$ 、メインジェスチャでは γ である。 dtw の要素 $dtw[i][j]$ には以下のように

定義される DTW 距離と呼ばれるものを格納する。動的計画法により $dtw[0][0]$ から $dtw[N-1][T_i-1]$ までのコストを順に求め、 $dtw[N-1][T_i-1]$ のコストが最小になるパスを求める。このときのコストはユークリッド距離 d であり、 $dtw[N-1][T_i-1]$ の値が類似度として用いる距離 $D(A, G_i)$ である。

$$dtw[0][0] = 0 \quad (4)$$

$$dtw[i][0] = \infty \text{ for } i = 1, 2, 3, \dots, N-1 \quad (5)$$

$$dtw[0][j] = \infty \text{ for } j = 1, 2, 3, \dots, T_i-1 \quad (6)$$

Do for $i = 1, 2, 3, \dots, N-1$

Do for $j = 1, 2, 3, \dots, T_i-1$

$$dtw[i][j] = d(a_i, g_j) + \min \begin{cases} dtw[i-1][j], \\ dtw[i-1][j], \\ dtw[i-1][j-1], \end{cases} \quad (7)$$

$$D(A, G_i) = dtw[N-1][T_i-1] \quad (8)$$

以上の方法ですべての事前データ $G_i = (g_1, \dots, g_{T_i})$ について入力データ A との距離を求め、 $\min\{D(A, G_i), i = 1, 2, 3, \dots, n\}$ となる $gesture_*$ を出力する。

5. まとめ

本稿では、スマートフォンの加速度センサを用いたジェスチャ入力において、ジェスチャの組み合わせを音楽プレイヤーの操作の入力に割り当て、DTW を利用したジェスチャの認識方法を提案した。今後は、歩行を伴う実験によって、 α, β, γ の値の評価や、各パターンにおける具体的な認識精度を算出し比較したい。

参考文献

- [1] 消防庁: “歩きスマホ等に係る事故に注意!”, <http://www.tfd.metro.tokyo.jp/lfe/topics/201602/mobile.html>.
- [2] 村尾和哉, 寺田努, 矢野愛, 松倉隆一, 西尾章治郎: “センサ内蔵型モバイル機器を用いたジェスチャ認識に関する一考察”, 研究報告ユビキタスコンピューティングシステム (UBI), pp. 1-8, (2010).
- [3] Ryo Izuta, Kazuya Murao, Tsutomu Terada, and Masahiko Tsukamoto: “Early gesture recognition method with an accelerometer”, *International Journal of Pervasive Computing and Communications*, Vol. 11, No. 3, pp. 270-287, (2015).

- [4] C. S. Myers and L. R. Rabiner: “A comparative study of several dynamic time-warping algorithms for connected-word recognition,” *The Bell System Technical Journal*, Volume 60, pp. 1389-1409, (1981).
- [5] Kazuya Murao, Gaku Yoshida, Tsutomu Terada, and Masahiko Tsukamoto: “Determining a Number of Training Data for Gesture Recognition Considering Decay in Gesture Movements”, *Information and Media Technologies*, Volume 10, Issue 3, pp. 449-458, (2015).