

LK_013

指先の動きを利用した仮想ボタン入力インタフェースの検討 A Virtual Push-button Interface Using Fingertip Movement

森本 一広[†] 宮島 千代美[†] 伊藤 克亘[‡] 武田 一哉[†]
Kazuhiro Morimoto Chiyomi Miyajima Katsunobu Itou Kazuya Takeda

1. はじめに

人と人とのコミュニケーションにおいて、音声などの言語的な手段と同様に、ジェスチャなどの非言語的な手段の果たす役割は大きい。例えば、会話の中で身振りや手振りをすることで、相手にわかり易く情報を伝えることができ、より円滑なコミュニケーションが可能となる。また、人と機械とのコミュニケーションにおいても、情報伝達をより豊かなものにするために、人間の動作や行動の認識に関する技術が必要不可欠なものとして注目されている。以上のことから、人間にとって自然で使いやすい、ジェスチャを用いたマンマシンインタフェースを実現することが求められている。

関連する先行研究として、塚田らは人差し指の曲げ伸ばし情報を用いて携帯情報機器や情報家電機器の操作を実現するモバイル指向のデバイスを提案し、主観評価実験において、利用者から「直感的な操作で扱いやすい」、「魅力的である」といった高い評価を得ている [1]。また、Rekimoto による Gesture Wrist は、手首に 2 軸加速度センサと静電検出装置 (送信電極と受信電極) が搭載された腕時計型機器を装着し、簡単なジェスチャ入力を試みる研究である [2]。手首以外に機器を装着する必要がなく、数種類のジェスチャを認識が可能となっている。

本研究では、ジェスチャの中でも指先の動きに注目した入力インタフェースの実現を目的とする。指先で空中に閉図形を描くことで入力領域が確保され、指先でその領域内部を仮想的に押すことにより入力が可能となるインタフェースを考えた。このインタフェースが実現すれば、様々な状況下での入力操作が安易になると想像される。例えば、車内運転中に機器の操作をする場合、リモコンや機器自体が離れていても、指先で簡単なジェスチャをすることで、リモコンに手を伸ばすことや、機器の操作ボタンの位置を目で確認することなく機器操作が行えるため、運転の安全性も高まる。本研究では、このようなジェスチャ入力インタフェースの一例として、空中で四角形を描き、その領域内を 2×2 の 4 つのボタンに見立てて、ボタンを仮想的に押すことで入力が行えるインタフェースを考えた。ここでは、1~4 の数字からなる 3 桁~5 桁の数字を人差し指で入力するインタフェースを想定し、人差し指の先に装着した三次元位置センサを用いて数字を入力する際の指先の動きの三次元データを収録し、入力数字列の認識を行った。認識方法として 2 種類の手法を考えた。一つ目は、指先の速度の符号の変化点から推定した四角形の領域と、数字を押した位置との関係から入力数字を認識する方法。二つ目は、描かれた四角形の位置と大きさの情報を利用して位置データの正規化を行い、四角形の描画やボタン入力の指先の動作を隠

れマルコフモデル (HMM) を用いてモデル化し、入力数字の認識をする方法である。

2. 三次元空間インタフェース

2.1 押すジェスチャによる入力

従来の家電器具の操作方法を参考にして、入力インタフェースを考えた。例えば、電源のオン・オフ操作をする際にはスイッチを押すという使い方が一般的である。本研究では以上のような、押すという一般的な機器操作方法を利用することで、誰でも簡単に操作でき、指先のジェスチャを用いることでいつでもどこでも扱える入力インタフェースの構築を目指す。

2.2 2×2 領域数字入力システム

前節を踏まえて、本研究で考えた入力インタフェースは以下の通りである。

- 空中に指先で四角形を描き、入力可能領域を確保する。
- 四角形の内側が 2×2 の領域の分割されていて、それぞれが 1~4 の数字に対応していると想像する。
- 指先で入力したい数字列に対応する領域を順番に押す。
- 入力可能領域から離れた位置に指を移動することで終了の合図とする。

上記のインタフェースは、操作が直感的でわかりやすい。また、四角形を描くことで入力開始の合図となると共に、押す領域をほぼ決定できる。領域を決定することにより、指先の位置情報の正規化を行うことも可能となる。データの正規化については 3 節で述べる。

3. データ収録

3.1 三次元位置収録装置

人差し指の位置情報を計測するために、Ascension Technology 社の Flock of Birds (FOB) を使用した。FOB は磁気変換技術を用いて 3 次元空間内の位置を測定する装置であり、トランスミッタ (磁界発生源) を原点にして、手首などに取り付けられたレシーバの 3 次元位置座標値 (x, y, z)、およびオイラー角 (方位角 ψ , 仰角 θ , 回転角 ϕ) の合計 6 次元の情報をリアルタイムに計測できる。

3.2 収録方法

収録は以下の流れで行った。

- (1) オペレータが被験者に数字列を口頭で伝える。
- (2) 被験者が数字列を覚えたのを確認し、開始の合図をする。

[†]名古屋大学大学院情報科学研究科

[‡]法政大学情報科学部

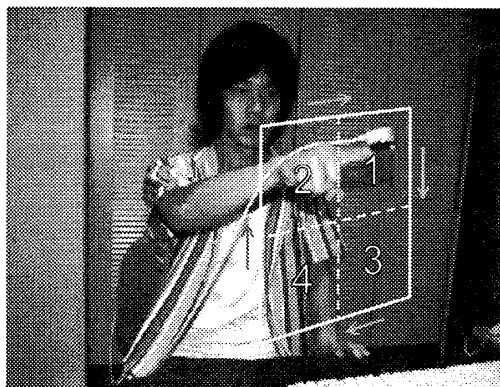


図 1: 収録時の様子

- (3) 被験者は指定位置 (机の上) から手を離し, 上体正面に四角形を描く.
- (4) 四角形の内側が 2×2 の領域に分割されていて, それぞれが 1~4 の数字に対応していると想像する.
- (5) 人差し指で数字列に対応する領域を順番に押す.
- (6) 手を元の位置に戻す.

手を指定位置から離してから戻すまでの一連の動作 (上記 (3)~(6)) を 1 回とし, 4 名の被験者に対して 1 人当たり 60 回の収録を行った. 入力対象の数字列は 1, 2, 3, 4 から構成された 3 桁~5 桁とし, 最終的にデータ中に現れる 4 数字の出現頻度がほぼ等しくなるように数字をランダムに選択した. 収録時は被験者は椅子に座り, 正面に四角を描くようにした. また, 原点となるトランスミッタは正面左側へ設置し, 図 1 のようにレシーバは右手人差し指先端部にしっかりと固定した. なお, 本実験ではレシーバのサイズが比較的大きな有線のセンサを用いているが, 実用化を考えた場合, 将来的には付け爪程度のサイズの無線のセンサを用いることが出来ればと考えている.

四角形内部での 1~4 の数字の配置 (数字の並び) は全ての被験者で等しくなるようにした. 四角形の描き方や数字の押し方はあらかじめ被験者に例を示した. 収録した 3 次元データの軌跡の例を図 2 に示す. y - z 平面に四角形を描き, x 軸のマイナス方向に数字列 12433 を入力している様子が確認できる.

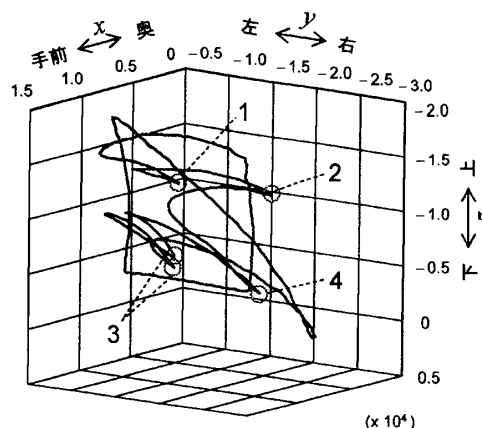
3.3 データ補正

3.3.1 データ補間の必要性

データのサンプリング周波数は約 100Hz で不等間隔である. 本研究では動作パターンをモデル化する際にサンプリング間隔を一定にする必要があるため, 区分的三次エルミート補間によってサンプリング間隔を 10ms に統一した.

3.3.2 データ正規化

四角を描く位置やサイズは被験者によって異なり, 収録の度にも変化する. それらのデータを同じ尺度で表す

図 2: 収録された 3 次元位置データの例
(5 桁数字列: 12433)

ために, データの正規化を行った. 描いた四角形はほぼ長方形でかつ y - z 平面上に描かれ, 押す動作は x 軸方向の動きであると仮定し, y - z 平面は四角形を描いている間の y, z 座標の最小・最大値で正規化し, x 方向 (手前・奥方向) はボタンを押している間の x 座標の最小・最大値で正規化した.

4. 四角形と押す位置の関係による認識 (手法 1)

まず, 四角形の領域と押す位置の関係による認識を行った. この手法は以下の流れで行う.

- 1 四角形を描いた開始時刻 T_{BEG} と終了時刻 T_{END} を求める.
- 2 各数字を押した時刻 t_n ($n = 1, 2, \dots, N$) を求める.
- 3 $T_{BEG} \leq t \leq T_{END}$ における (y, z) の軌跡から描いた四角形の領域を決定する.
- 4 四角形の領域と時刻 t_n における座標 (y_n, z_n) の位置関係から押した数字列を判定する.

各時刻の決定には位置情報に関する一次動的特徴量 (Δ) を利用した. 押す瞬間は Δx の値が負から正に変化するため, $\Delta x = 0$ の時刻が押した瞬間とみなせる. 四角形の描画の開始・終了時刻の決定も同様に $\Delta y, \Delta z$ の値を用いており, 四角形描画の開始時は, 指先を指定位置から四角形の開始地点 (左上) へ移動し, 四角を描き始める時で, Δz の値が負から正に変化するため, $\Delta z = 0$ の時刻が開始時刻とみなせる. また, 四角形描画の終了時は, 四角形を描き終えて, 押す動作へと進む時で, Δy の値が正から負に変化するため, $\Delta y = 0$ の時刻が終了時刻とみなせる.

4.1 実験方法

上で述べた手法に関して連続数字認識実験を行った. 実験条件を表 1 に示す. 表 1 における領域拡大率とは, 実際に描いた四角形の 1 辺の長さを M 倍して, その範囲を入力可能領域として許容するということである. これは, 実際に入力を行う際には描いた四角形は見えてい

表 1: 実験条件 1

標準化周波数	100Hz
領域拡大率 M	1.0, 1.1, 1.2, 1.3, 1.4 倍
特徴量	y, z $\Delta x, \Delta y, \Delta z$
データ数	240 (60回×4名)

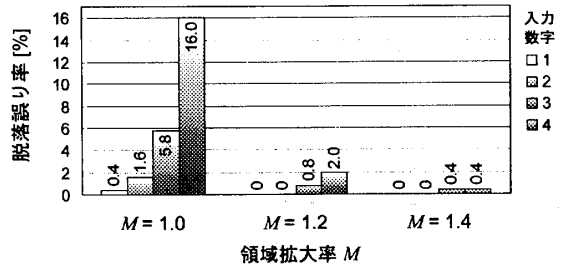


図 4: 領域拡大率変化による数字別脱落誤り率

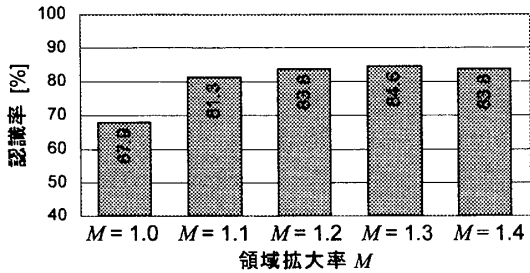


図 3: 領域拡大率変化における認識率

ないため、実際よりも広い範囲で操作を行うかもしれないという考えに基づいている。

4.2 実験結果

図 3 に領域拡大率変化における認識率を示す。拡大率を上げるにより認識率の上昇がみられ、最大で 84.6% の認識率が得られた。しかし、拡大しすぎると、認識率は低下した。これは終了位置に戻す動作が押す動作として誤認識されたためである。また、図 4 に領域拡大率変化による数字別脱落誤り率を示す。範囲を広げることにより各数字において脱落誤り率が大幅に減少していることが分かる。さらに上の領域 (1 と 2) よりも下の領域 (3 と 4) を押す場合の脱落誤りが多く、特に右下の 4 の領域が脱落誤りが多いことが分かる。

5. HMM を用いた認識 (手法 2)

次に隠れマルコフモデル (HMM) を用いた認識実験を行った。HMM は統計的モデル化手法として音声・画像認識の分野でその有効性が示されている [3][4]。

本実験では、各動作 (指定位置から四角を描く直前、四角を描く、1 を押す、2 を押す、3 を押す、4 を押す、押す動作後から指定位置に戻す) の合計 7 種類のモデルを作成した。HMM の構造は、初期状態と終了状態を持ち、スキップの遷移を認めない left-to-right 型とした。

5.1 実験方法

予備実験より、位置センサで収録した 6 次元の値 ($x, y, z, \psi, \theta, \phi$) のうち、三次元座標値 (x, y, z) を用いた。また、それらの一次・二次動的特徴量 ($\Delta xyz, \Delta^2 xyz$) を算出し、計 9 次元の特徴量を用いて認識実験を行った。動的特徴量は窓幅 40ms の線形回帰係数として求めた。認識実験には HTK[4] を用い、各動作の HMM のパラ

表 2: 実験条件 2

標準化周波数	100Hz
HMM の状態数	5, 10, 15, 20
特徴量	x, y, z $\Delta x, \Delta y, \Delta z$ $\Delta^2 x, \Delta^2 y, \Delta^2 z$ (上記計 9 次元の組合せ)
学習用データ	180 (60回×3名)
評価用データ	60 (60回×1名) (クロスバリデーション)

メータは、セグメンタル K-means で初期化し、Baum-Welch アルゴリズムによって再推定を行った。実験条件を表 2 に示す。認識では、全 4 名分のデータの内 3 名分を学習用、残りの 1 名分を評価用とし、それらの組み合わせを替えて 4 回分の平均を結果とするクロスバリデーションによる評価を行った。

5.2 実験結果

図 5 に状態数と特徴量変化による認識率を示す。認識率は、一連の動作が全て正しい場合を正解としたものである。各動作の HMM の状態数を 10 かつ特徴量を $x, y, z, \Delta x, \Delta y, \Delta z$ の 6 次元とした場合に最も高く、95.4% の認識率が得られた。この結果は、手法 1 の実験結果よりも 10 ポイント以上高い結果となった。図 5 より、特徴量 x, y, z のみを用いた場合において、状態数を増やすことで認識率に上昇が見られたが、特徴量に一次、二次動的特徴量を用いた場合においては、状態数を増やしても認識率に大きな上昇はみられないことが分かる。また、動的特徴量に加えることで、少ない状態数においても高い認識率が維持されていることが分かる。一次動的特徴量 (Δ) は指先の向きや速さを表し、特徴量として重要であることが分かる。加速度である二次動的特徴量 (Δ^2) を加えた場合には、認識率は 6 次元の場合と同程度であり、特徴量として加えても大きな効果はないと考えられる。

6. 考察

手法 1 の実験結果から、描いた四角形の範囲よりも押すときには広い範囲で動作を行っているということがわ

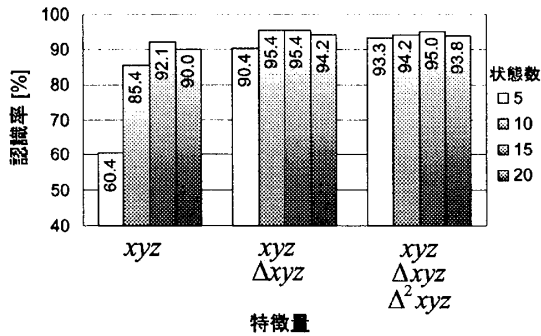


図 5: 状態数・特徴量変化による認識率

かる。その原因として考えられるのは、今回の手法では押した位置の座標を描いた四角形の平面 ($y-z$ 平面) 上に射影しているため、ボタンを押す動作が四角形に対して手前上から奥下への動きであるとする、下側の数字を押した場合に入力可能領域から外れて脱落してしまったと考えられる。

また、HMM に基づく認識においては、状態数が 10 で、特徴量として 1 次動的特徴量 (Δ) を加えた場合に、95.4% の認識率が得られ、手法 1 の実験結果よりも 10 ポイント以上認識率が向上していることから、有効な認識手法であると言える。手法 1 の実験との差が一番顕著に現れたのは表 3 に示すように上の領域から下の領域への置換誤り率であった。この結果から、HMM による認識では置換誤りが大幅に減少していることが分かる。これは、HMM によるモデル化は入力可能領域からのずれも学習データから得られたガウス分布の分散でカバーされており、領域の拡大を明示的に行う必要がないことや、押す位置だけでなく移動方向などの情報も含まれていることから、置換誤りが減少し認識率が上昇したと考えられる。

人は、自分が描いた四角形的位置と実際に入力する際の四角形的位置に誤差があると考えられる。図 6 のように、平面自体が多少傾き、押す動作は水平ではなく手前上から奥下へと押す傾向にある。その傾向は上の領域を下側の領域として誤ってしまう置換誤りが多く存在していることから分かる。さらに押すジェスチャをした際に、上側の部分は比較的奥側で動作を行っているのに対し、下側の部分は手前側で動作を行っている傾向が見られた。HMM でモデル化された動作は、上側の領域での動作は奥側、下側の領域での動作は手前で行うモデルとなっていると考えられ、上側領域と下側領域での奥行き情報に差があったことから、認識率が高い結果となったと考えられる。

7. まとめと今後の課題

本研究では、誰にでも扱いやすい入力インタフェースの構築を目指して、指先で空中に四角形を描き、その領域内を 2×2 の領域に分割、1~4 の数字に対応した 4 つのボタンに見立てて仮想的に押すことで入力が行えるインタフェースを考えた。その一連の動作を 3 次元位置

表 3: 上領域から下領域への置換誤り率の比較

手法 1 ($M = 1.3$)	9.0 %
手法 2 (状態数 15・特徴量 6 次元)	1.6 %

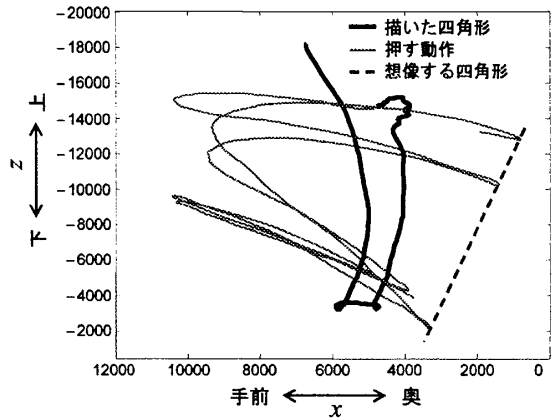


図 6: 描いた四角形と想像した四角形的位置関係

センサによって収録し、HMM に基づく認識実験を行った結果、状態数 10、特徴量 6 次元 ($x, y, z, \Delta x, \Delta y, \Delta z$) の場合に認識率は最大となり、95.4% の認識率が得られた。また認識誤りには傾向があり、上下関係にある領域の置換誤り率が他の誤りと比べて高かった。実際に、被験者が押すジェスチャをした場合に、上側の部分は比較的奥側で動作を行っているのに対し、下側の部分は手前側で動作を行っている傾向が見られた。

今後は、以上のような人間の動作の特性を考慮して置換誤りを減らすための手法を考える必要がある。また、本研究は動作に対する制約が多かったために認識率が比較的高かった可能性もある。従って、さまざまなジェスチャの種類・動作速度・位置・方向を考慮した調査が必要である。また、正規化の方法として、拡大・縮小・移動のみでなく回転や非線形な歪みに対応した正規化についても考える必要がある。どのような位置・方向での動作も今回のように正面での動作として見なせるように正規化が行えれば、より高い認識率が得られると考えられる。

参考文献

- [1] 塚田浩二, 安村通晃, “Ubi-Finger: モバイル指向ジェスチャ入力デバイスの研究,” 情報処理学会誌, Vol.43, No.12, pp.3675-3684, Dec. 2002.
- [2] J. Rekimoto, “GestureWrist and GesturePad: Unobtrusive Wearable Interaction Devices,” Proc. of 5th International Symposium on Wearable Computers, Oct. 2001.
- [3] 鹿野清宏, 伊藤克巨, 河原達也, 武田一哉, 山本幹雄, 音声認識システム, オーム社, 2001.
- [4] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, “The HTK Book,” Microsoft Co.