社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－ＨＩ－122 (14)
2007－SLP－ 65 (14)
2007／2／10

ジェスチャインタフェースのための動作軌跡信号の統計的分割と認識

森本 一広†, 宮島 千代美†, 北岡 教英†, 伊藤 克亘‡, 武田 一哉†

† 名古屋大学大学院 情報科学研究科 〒464-8603 名古屋市千種区不老町
‡ 法政大学 情報科学部 〒184-8584 東京都小金井市梶野町 3-7-2
†{morimoto,miyajima,kitaoka,takeda}@sp.m.is.nagoya-u.ac.jp, ‡itou@k.hosei.ac.jp

あらまし 本研究では，ジェスチャの中でも特に指先の動きに注目したジェスチャインタフェースの実現を目指し，その一例として，空中に指先で閉図形を描くことで入力領域を確保し，その内部をボタンの並びと見なして，それらを仮想的に押すことで入力を行うインタフェースを考えた．三次元位置センサを用いて指先の動作軌跡を収録し，主成分分析および曲率・速度を用いた動作軌跡の単純動作への分割と，アフィン変換を用いたジェスチャ動作方向と大きさに対する正規化を行った．HMM による連続数字入力の認識実験を行った結果，正規化前の認識率 60.0% に対して，正規化後では 91.3% の認識率が得られた．

キーワード ジェスチャインタフェース，三次元位置センサ，主成分分析，アフィン変換，隠れマルコフモデル

Statistical Segmentation and Recognition of Fingertip Trajectories for a Gesture Interface

Kazuhiro MORIMOTO†, Chiyomi MIYAJIMA†, Norihide KITAOKA†,
Katsunobu ITOU‡, and Kazuya TAKEDA†

† Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8603, JAPAN
‡ Faculty of Computer and Information Sciences, Hosei University
3-7-2, Kajino-cho, Koganei, Tokyo 184-8584, JAPAN
†{morimoto,miyajima,kitaoka,takeda}@sp.m.is.nagoya-u.ac.jp, ‡itou@k.hosei.ac.jp

Abstract This paper presents a virtual push-button interface created by drawing a shape or line in the air with a fingertip. As an example of such a gesture-based interface, we developed a four-button interface for entering multi-digit numbers by making pushing gestures within an invisible 2x2 button matrix inside a square drawn by the user. Trajectories of the fingertip movements entering randomly chosen multi-digit numbers are captured with a 3D position sensor mounted on the tip of the forefinger. We propose a statistical segmentation method for the trajectory of movements and a normalization method, associated with direction and size of gestures. The performance of the proposed method is evaluated in HMM-based gestures recognition. Recognition rate of 60.0% was improved to 91.3% after applying the normalization method.

Keywords Gesture interface, 3D position sensor, Principal component analysis, Affine transformation, Hidden Markov model

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の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を実現するために動作軌跡信号の統計的処理方法について検討する．指先の動きであれば,「描く」,「押す」といった直感的なジェスチャが可能であることから指先の動きに注目した．現在のジェスチャ認識の課題の一つとして，動作の中でどの区間が何か意図を持って行ったジェスチャ区間であるのかを自動検出することが挙げられる．そのために，例えば手にタッチグローブを装着し，人差し指と中指を接着させてジェスチャを行った区間のみを認識対象とするといった研究 [3] も行われているが，グローブを装着しなければならないといった煩わしさが存在する．本研究では，将来的には付け爪程度の無線位置センサや加速度センサなどが開発され，そのような小型のセンサを用いることでそういった煩わしさが大幅に軽減できるのではないかと考えている．

2 仮想ボタン入力インタフェース

本研究では指先の動きに注目し，指先のジェスチャとして一般的な「描く」という動作や「押す」という動作を含むジェスチャ入力を考えた．ここでは，ユーザが空中に四角形を描き，内部を 2×2 領域の 4 つのボタンとみなして仮想的にボタンを押すことで連続数字を入力する次のようなインタフェースを考えた．

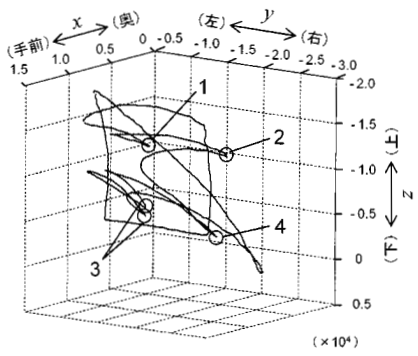

図 1: 収録された動作軌跡信号の例
(5 桁数字列：12433)

- 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 領域に分割されていて，それぞれが 1〜4 の数字に対応 (左上が 1, 右上が 2, 左下が 3, 右下が 4) 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．

このインタフェースでは，四角形描写動作を動作開始時に行うことで入力開始の合図となるとともに，ジェスチャ範囲を推定することが容易になる．後に述べる正規化手法はこの情報を用いた手法である．

指先の動きは三次元位置センサで収録した．収録した動作軌跡信号の例を図 1 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 ジェスチャ動作の単純動作への分割

本研究では，ジェスチャを直線や曲線といった単純な動きの連続であると考えることにする．例えば四角形を一筆書きする動作は，1. 上から下，2. 左から右，3. 下から上，4. 右から左といった 4 つの直線的な動きの連続と見なすことができる．すなわち複雑な動作であっても単純な動きの連続として捉えることで，より単純にデータを扱うことができると考えられる．そこでジェスチャ動作を単純動作へと分割する方法を考える．ジェスチャの単純動作への分割は次の流れで行う．

- 三次元データ系列に対して逐次 PCA を行い，頂点が一つ含まれると考えられる区間に分割する．
- 分割したデータ (二次元) に対して曲率データを求め，頂点の候補 $V_C^{(n)}$ を決定する．
- 三次元データに対して速度データを求め，頂点の候補 $V_D^{(n)}$ を決定する．

- 頂点候補 $V_C^{(n)}$ と $V_D^{(n)}$ とを統合し，頂点を決定する．

3.1 PCA による区間分割

収録した三次元データの時系列 $\mathbf{r}(t_s),\ldots,\mathbf{r}(t_e)$ に対して主成分分析 (PCA) を行う．

そして第二主成分の寄与率に閾値を設定し，その閾値を越えるまで終点 $\mathbf{r}(t_e)$ を延長し，越える直後のデータを $\mathbf{r}(t_s)$ として同様の操作を繰り返す．以上の操作を行って分割した例を図 2 に示す．× 印で分割された各区間のデータ系列には，ある単純動作 (直線) とその次の単純動作 (直線) との境目 (頂点) が 1 つ存在すると考えられる．

3.2 曲率による頂点検出

次に，前節で求めた区間に対して曲率を計算する [4]．ここで，時刻 t における曲率 $C(t)$ は，時刻 t での二次元座標値 (第一・二主成分軸上) $\mathbf{r}'(t)$ と 2 ポイント前と後の座標値 $\mathbf{r}'(t-2),\mathbf{r}'(t+2)$ の三点を通る円の半径の逆数 $(1/R)$ として求められる．

図 2 のデータに対して曲率を計算した結果を図 3 に示す．各区間において曲率が最大となる点 (○印) が頂点 $V_C^{(n)}$ であると考えられる．

なお，前節で分割されたデータが円弧のような頂点を持たない部分である可能性もあるため，その区間内での曲率の平均値 $\tilde{C}$ と最大曲率 C_{max} とを比較して，

$$C_{max} > \tilde{C}\cdot\alpha \tag{1}$$

の条件を満たさない場合にはその区間に頂点は存在しないとすることにした．本研究では $\alpha = 2.0$ とした．

3.3 速度による頂点検出

以上の頂点検出と同時に速度データによる頂点検出を行う．まず三次元座標値に対して，速度データ

$$\Delta\mathbf{r}(t) = \sqrt{\Delta x^2(t)+\Delta y^2(t)+\Delta z^2(t)} \tag{2}$$

を求める．但し，$\Delta x(t),\Delta y(t),\Delta z(t)$ は各方向での速度で，次式のように計算する．

$$\Delta x(t) = \frac{\sum_{k=-K}^{K} k x(t+k)}{\sum_{k=-K}^{K} k^2} \tag{3}$$

ここで K は速度データを計算する範囲であり，本研究では $K = 2(40\text{ms})$ とした．

例として図 4 に四角形を描いた際の速度データを示す．これを見ると，頂点付近を描く速さが辺の中心付近などを描く速さよりも遅くなっていることが分かる (○で囲んだ部分)．そこで，この速度データに対してその区間内での速度の平均値 $\tilde{\Delta\mathbf{r}}$ と最小速度 $\Delta\mathbf{r}_{min}$ とを比較し，

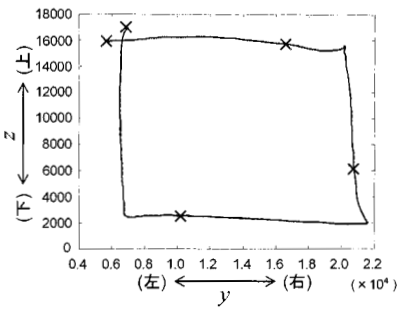

図 2: PCA による区間分割

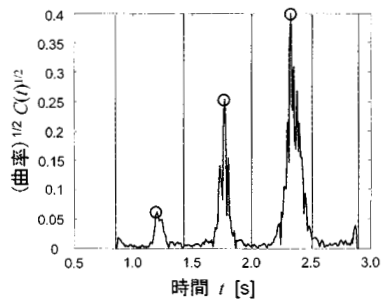

図 3: 四角形描写時の曲率データ $C(t)(=1/R)$

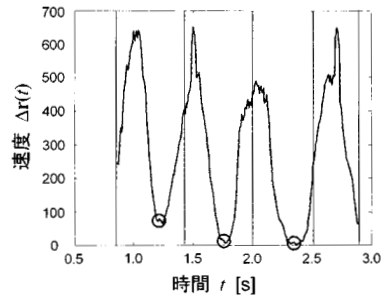

図 4: 四角形描写時の速度データ $\Delta\mathbf{r}(t)$

$$\Delta\mathbf{r}_{min} < \tilde{\Delta\mathbf{r}}\cdot\beta \tag{4}$$

の条件を満たした場合には，最小速度となる点が頂点 $V_D^{(n)}$ であると考えた．逆に条件満たさない場合にはその区間に頂点は存在しないと考えた．本研究では $\beta = 0.5$ とした．

3.4 頂点候補データの統合

最後に，求められた頂点候補 $V_C^{(n)}$ と $V_D^{(n)}$ を統合する．一般的な傾向として，頂点付近は他の部分と比べ曲率が大きく速度が遅いので，$V_C^{(n)}$ と $V_D^{(n)}$ が共に検出される可能性が高い．しかしそれぞれの方法で検出した点は互いに若干ずれている場合が多い．そのため，ある頂点に対して $V_C^{(n)}$ と $V_D^{(n)}$ が共に検出され，検出時刻の差が 0.1 秒以内である場合には，速度

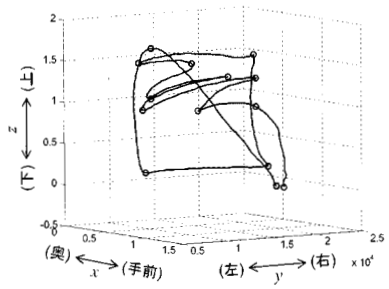

図 5: ジェスチャ動作の分割例 (被験者 A)

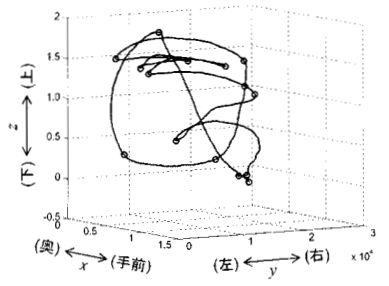

図 6: ジェスチャ動作の分割例 (被験者 B)

データによる頂点 $V_D^{(n)}$ を採用することにした．なぜなら，曲率よりも速度によって得られた頂点候補の方が経験的にみて信頼性が高かったためである．また片方の候補のみが得られた場合には，そのまま頂点として採用することにした．

3.5 単純動作への分割結果

本分割手法を用いてジェスチャ動作に対して分割を行った結果を図 5 と図 6 に示す．各図において○で示された部分が検出した頂点である．図 5 は比較的正確にジェスチャ動作を行った被験者の例で，直線部や曲線部に分けることができていると考えられる．しかし，曲線の途中で分割されてしまう部分も存在した．また，図 6 は比較的大雑把にジェスチャ動作を行った被験者の例である．この被験者は四角形の頂点付近でも速度の減少があまり見られず，速度データのみを使った場合には四角形を分割することができない．しかし，曲率データを併用することにより，四角形部分も辺部分毎に分割することができた．なお，前例と同様に，曲線部などで不必要な分割が生じる場合があった．

4 動作方向と大きさに対する正規化

人が頭の中で想像している動作の軌跡と実際に動作した軌跡とではずれが生じると考えられる．これは，人の描画動作にはもともとぶれがあるため，また，手指・腕などの身体的な構造で可動範囲に制約があるためと考えられる．更に磁気センサで収録される三次元データが周囲の環境の影響によって軸が歪む可能性も考えられる．また人によっても描く大きさや動作方向，描く正確さなどの違いが存在している．そこで，この二つの軌跡の対応を関数を用いて表現する (線形変換を行う) ことにより，正規化を行うことを考える．

4.1 PCA による描写平面の検出

3 節の方法で分割した中から四角形を描いている部分のみを取り出し，その区間内のデータ $\mathbf{r}_i = (x_i, y_i, z_i)'$ (i: 正整数) に対して PCA を行う．人は一般的に，三次元空間のある平面上に図形を描くと考えられるので，PCA を行って得られた主成分のうち，第一・第二主成分軸によって表される平面がその描写平面に対応していると考えられる．本インタフェースで考えるとその平面は動作者の正面に描かれるので，この x'-y' 平面検出を行うことによって方向の正規化を行うことができる．また第三主成分はその平面に直交した向き (z' 軸) となり，本インタフェースで考えると，押す動きを最もよく表す軸と考えることができる．

数式で表すと，元の軌跡 $\mathbf{r}_i = (x_i, y_i, z_i)'$ と方向正規化後の軌跡 $\mathbf{r}'_i = (x'_i, y'_i, z'_i)'$ との関係は，

$$\begin{bmatrix} x'_i \\ y'_i \\ z'_i \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \tag{5}$$

となる．ここで，a_1, a_2, a_3 は第一主成分，b_1, b_2, b_3 は第二主成分，c_1, c_2, c_3 は第三主成分を表す．

4.2 アフィン変換による大きさの正規化

4.1 節の方法で求めた x'-y' 平面上の四角形の大きさと形の歪み，及びボタン押下動作の位置に対して正規化を行うために次式に示すアフィン変換を用いて線形変換を行った．

$$\begin{bmatrix} x''_i \\ y''_i \\ z''_i \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & 0 \\ d_{21} & d_{22} & 0 \\ 0 & 0 & \frac{1}{\sigma_{z'}} \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \\ z'_i \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ -\frac{\mu_{z'}}{\sigma_{z'}} \end{bmatrix} \tag{6}$$

但し，$\mu_{z'}$ 及び $\sigma_{z'}$ はそれぞれボタン押下時の z' 軸データの平均と標準偏差を表す．ここで係数 $d_{11} \sim d_{22}, e_1, e_2$ は，変換後の座標と対応する理想座標との二乗誤差が最小となるように決定した．なお，押す方向 z''_i の幅に関しては第三主成分軸をそのまま用いた．

4.3 ジェスチャ動作方向を考慮した正規化手法

4.1 節および 4.2 節の手法を組み合わせることにより，元の軌跡 $\mathbf{r}_i = (x_i, y_i, z_i)'$ に対して方向・大きさの正規化とともにデータの修正を行った正規化後の軌

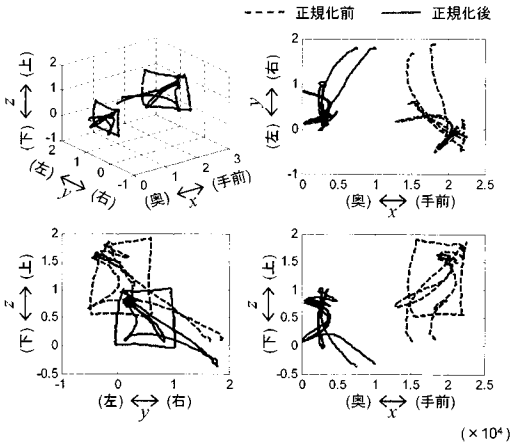

図 7: 方向・大きさの正規化前後の軌跡

跡 $\mathbf{r}''_i = (x''_i, y''_i, z''_i)'$ を得ることが出来る．

$$\begin{bmatrix} x''_i \\ y''_i \\ z''_i \end{bmatrix} = \begin{bmatrix} d_{11} & d_{12} & 0 \\ d_{21} & d_{22} & 0 \\ 0 & 0 & \frac{1}{\sigma_{z'}} \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ -\frac{\mu_{z'}}{\sigma_{z'}} \end{bmatrix} \tag{7}$$

ここで，x'' 軸は左右方向，y'' 軸は上下方向，z'' 軸は奥行き方向に対応する．上記の正規化を行う前後の軌跡の例を図 7 に示す．正面とは異なる方向に向かって行った動作の軌跡が，正面方向を向いて行った動作の軌跡へと変換できているのが分かる．また，四角形の形も線形的な歪みが改善され，大きさも $y\,[0,10000], z\,[0,10000]$ となるように変換された．但し，非線形な歪みの正規化についても今後考える必要がある．

5 HMM を用いた連続数字入力認識

上記の正規化手法の有効性を確認するために，隠れマルコフモデル (HMM) を用いた連続数字入力の認識実験を行った．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 [5][6].

5.1 データ収録

収録には三次元位置計測装置 (Ascension Technology 社の Flock of Birds) を使用した．磁気変換技術を用いて 3 次元空間内の位置情報 (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) を測定することができる．

被験者は右手人差し指先端部にセンサを装着し (図 8)，2 節で述べた一連の動作を各被験者に対して 60 回 (評価用 1 名は 80 回) 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 (四角形内部での 1〜4 の数字の配置は全ての被験者で等しい) 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．

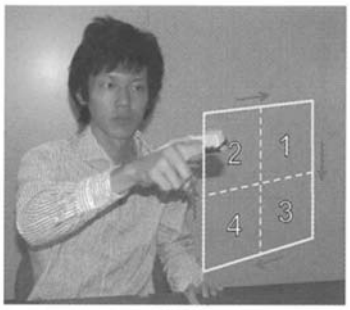

図 8: 収録時の様子

表 1: 実験条件

標本化周波数	100Hz
HMM の状態数	15, 20, 25, 30, 35
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	240 回 (60 回× 4 名)
評価用データ	80 回 (右・左・正面・正面上・左上 × 各 16 回 ×1 名

5.2 実験条件

予備実験より，位置センサで収録した 6 次元の値 (x,y,z,ψ,θ,ϕ) 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 $(\Delta x, \Delta y, \Delta z, \Delta^2 x, \Delta^2 y, \Delta^2 z)$ 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[6] を用いて各動作 (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．各動作の HMM のパラメータは，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 1 に示す．

認識では，正面方向に収録した 4 名分のデータを学習用，様々な方向に向いて収録した 1 名分のデータを評価用とした．

5.3 実験結果

状態数と特徴量変化による連続数字入力認識実験において正規化を行わない場合の結果を図 9 に，正規化後の結果を図 10 に示す．なお，3 桁 〜5 桁の数

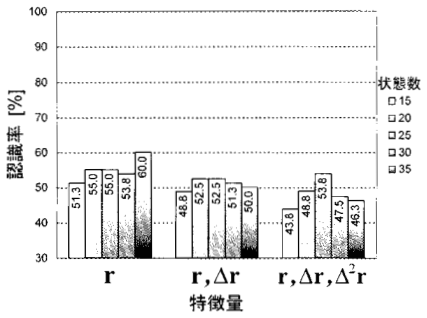

図 9: 状態数と特徴量変化による連続数字入力認識率 (正規化前)

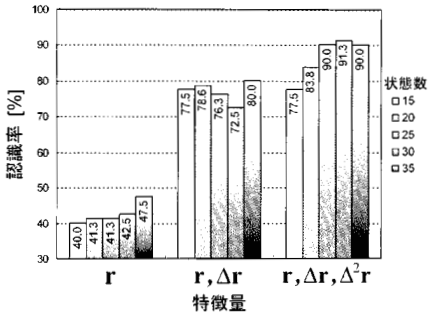

図 10: 状態数と特徴量変換による連続数字入力認識率 (正規化後)

字入力全てが正しく認識された場合に正解とした．正規化前で最大 60% の認識率であったのに対し，正規化後では最大 91.3% と約 30 ポイントの認識率向上が見られた．特に，正規前は四角形描写動作を押す動作として認識する置換誤りが非常に多く見られたが，正規化によって大きく改善された．なお，本実験では被験者が正面を向いてジェスチャを行っているため，正規前のデータでもある程度の認識ができたが，被験者の向きが不定である場合は認識は困難である．しかし本手法を用いればどのような方向に向かって行ったジェスチャも認識できる．また図 10 の結果から，特徴量として動的特徴量 Δ, Δ^2 を加えることにより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，二次動的特徴量 (Δ^2) はその加速度を表すと考えられ，ジェスチャの認識においてこれらの特徴量は重要であるということが分かる．

また，HMM によるモデル化では，データを統計的に扱う必要があるため，方向や大きさの正規化を行うことは非常に重要であり，本正規化手法の有効性が示された．本正規化手法は，数字入力ジェスチャに対してのみでなく，一般的なジェスチャ認識に対しても応用が可能であると考えられる．但し，今回の実験では，必ずジェスチャの始めに四角形を描き，その後に数字を入力するという一連の動作が，一つのデータとして独立に収録された条件で認識を行っているため，今後複数のジェスチャが含まれるデータや無意味な動作も含まれるデータに対しても用いることができるように改善をする必要がある．

6 まとめと今後の課題

本研究ではジェスチャの中でも特に指先の動きに注目したジェスチャインタフェースの実現を目指し，その一例として，空中に指先で閉図形を描くことで入力領域を確保し，その内部をボタンの並びと見なして，それらを仮想的に押すことで入力を行う 2×2 領域インタフェースを考えた．そしてその動作を三次元位置センサを用いて指先の動作軌跡として収録した．収録された動作軌跡を統計的に単純動作へ分割し，分割して得られた四角形描写間のデータを用いることによる動作方向と大きさの正規化を行った．HMM による連続数字入力の認識実験を行った結果，正規化前の認識率 60.0% に対して正規化後で 91.3% の認識率が得られ，本正規化手法の有効性が示された．

今後の課題としては，曲線部分の分割に対する誤りを軽減するなど，分割手法の精度を改善することが挙げられる．また，今回は正規化にアフィン変換を用いたが，非線形変換を用いて正規化を行うことも考えられる．

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] I.C. Kim, S.I. Chien, “Analysis of 3D Hand Trajectory Gestures Using Stroke-Based Composite Hidden Markov Models,” Applied Intelligence, vol.15, pp.131–143, 2001.

[4] T.M. Sezgin, T.Stahovich, and R.Davis, “Sketch Based Interfaces: Early Processing for Sketch Understanding,” Perceptive User Interfaces Workshop, 2001.

[5] 鹿野清宏，伊藤克亘，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.

[6] S. Young, D. Kershaw, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, *The HTK Book*, Microsoft Co.