

## 動作の早期認識およびその予測への応用に関する検討

内田, 誠一  
九州大学システム情報科学研究院知能システム学部門

森, 明慧  
九州大学システム情報科学研究院知能システム学部門

倉爪, 亮  
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎  
九州大学システム情報科学研究院知能システム学部門

他

<https://hdl.handle.net/2324/5960>

---

出版情報 : 信学技報パターン認識・メディア理解 (PRMU2004-94), pp. 7-12, 2004-11  
バージョン :  
権利関係 :

## 動作の早期認識およびその予測への応用に関する検討

内田 誠一<sup>†</sup> 森 明慧<sup>††</sup> 倉爪 亮<sup>†</sup> 谷口倫一郎<sup>†</sup> 長谷川 勉<sup>†</sup>  
迫江 博昭<sup>†</sup>

<sup>†</sup>九州大学大学院システム情報科学研究所 〒812-8581 福岡市東区箱崎 6-10-1

<sup>††</sup>九州大学電気情報工学科 〒812-8581 福岡市東区箱崎 6-10-1

E-mail: †uchida@is.kyushu-u.ac.jp

**あらまし** 本報告では、ジェスチャの早期認識、およびその認識処理を利用した動作予測の2点について検討する。早期認識とはジェスチャの終了を待たずになるべく早い段階でそれが何のジェスチャであるかを識別するものである。一方、動作予測とはユーザ(ジェスチャ行為者)の数フレーム後の状態を推定するものである。本報告では比較的単純なアプローチにより以上2つの課題を検討しながら、それらの原理的な限界を明らかにする。さらに、そうした限界を踏まえた上で、プロアクティブヒューマンインタフェースの開発等、今後さらに検討を進めて行くための適切な方針について展望する。

**キーワード** ジェスチャ認識, 早期認識, 動作予測, DP

## Early Recognition and Prediction of Gestures for Proactive Human-Machine Interface

Seiichi UCHIDA<sup>†</sup>, Akihiro MORI<sup>††</sup>, Ryo KURAZUME<sup>†</sup>, Rin-ichiro TANIGUCHI<sup>†</sup>,  
Tsutomu HASEGAWA<sup>†</sup>, and Hiroaki SAKOE<sup>†</sup>

<sup>†</sup> Faculty of Information Science and Electrical Engineering, Kyushu University  
6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, Fukuoka, 812-8581 Japan

<sup>††</sup> Department of Electrical Engineering & Computer Science, Kyushu University  
6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, Fukuoka, 812-8581 Japan

E-mail: †uchida@is.kyushu-u.ac.jp

**Abstract** This paper concerns two topics on gesture recognition. The first topic is early recognition for providing the recognition result of a gesture before the gesture is completed. The second topic is motion prediction for guessing the subsequent posture of the person who makes a gesture. Both topics are mutually related and linked to the realization of proactive human-machine interface. For each of those two topics, a simple technique is developed and examined to reveal its limitation. Possible directions to deal with the limitation are also discussed as the future work on those topics.

**Key words** gesture recognition, early recognition, motion prediction, dynamic programming

### 1. ま え が き

本報告では、(i) ジェスチャの早期認識、および (ii) その認識処理を利用した動作予測、の2点について検討する。早期認識とはジェスチャの終了を待たずになるべく早い段階でそれが何のジェスチャであるかを識別するものである。例えば、両手が上がり始めた段階でそれがジェスチャパターン「万歳」の冒頭部であると認識するような処理である。一方、動作予測とは

ユーザ(ジェスチャ行為者)の数フレーム後の状態を推定するものである。

第一の検討課題であるジェスチャの早期認識には例えば以下のような利用価値がある。

- 省力化… 認識結果を早期確定した時点で残りのジェスチャは不要になる。
- 先回り処理の実現… 認識結果に応じて他のシステムを制御する場合、そのシステムの制御を早期に開始できる。例え

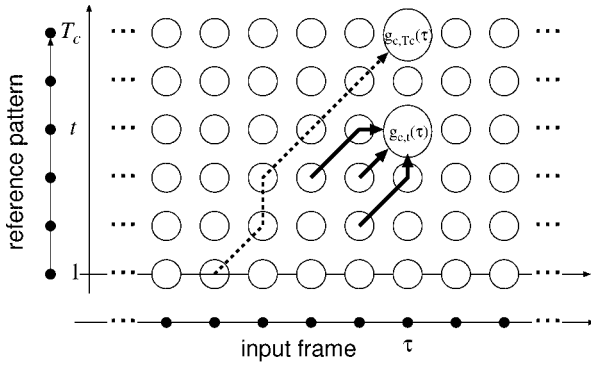


図1 従来の連続 DP によるジェスチャ認識

ばジェスチャによりロボットを制御する場合、ロボットを早めに動作すべき状態に進めておくことができる。

ジェスチャの早期認識についてはこれまであまり検討されていなかった。すなわち、従来のジェスチャ認識法のほとんどすべては、ジェスチャが終了した時点で認識結果を出力するものであった。先の「万歳」の例で言えば、両手が上がり、次いで元の位置まで下ろした状態に至って初めて認識結果が出力されることになる。

この点をふまえ、本報告では連続 DP [1] を早期認識用に拡張したジェスチャ認識法を提案する。連続 DP は、DP による時間軸非線形マッチング処理を基本とした手法であるが、開始時刻と終了時刻を自由化したことによりスポッティング認識が可能になっている。この性質は、実時間ジェスチャ認識においても有用である。実際、連続 DP に基づくジェスチャ認識手法は数多く提案されている [2]~[9]。ただし前述のように、これまでに早期認識について特に注意した手法は提案されていない。

第二の検討課題である動作予測とは、前述のように、現時刻の数フレーム後にはユーザがどのような動作状況 (例えば「いまから 5 フレーム後には『万歳』の両手を上げきった状態になる」など) にあるかを予測するものである。動作予測は、人間の行動に対して何か応答するようなシステムの遅れ補償に有用である。例えば、ヒューマノイドに人間の動作を模倣させるシステムでは、ユーザの現在の姿勢パラメータをそのままヒューマノイドに与えたとしても、瞬時に目標姿勢に達することはできない。従って模倣に遅れが生じることになる。これに対し、数フレーム後の姿勢パラメータを予測によって求め、それを常に渡しておけば、遅れを最小化できる。本報告で提案する動作予測手法は、前述の早期認識の結果を利用するものである。従って、動作予測と早期認識の間には密接な関係がある。

本報告の 2 つの検討課題、すなわち早期認識および動作予測のいずれも、筆者らが検討しているプロアクティブ (先回り) ヒューマンインタフェース (4.3 節で説明) を実現する上で必要となる要素技術である。本報告ではまず、比較的単純なアプローチにより以上 2 つの課題を検討しながら、それらの原理的な限界を明らかにする。さらに、そうした限界を踏まえた上で、今後プロアクティブヒューマンインタフェースを実現して行くための適切な方針について展望する。

## 2. ジェスチャの早期認識

### 2.1 連続 DP による従来のジェスチャ認識

本節では、連続 DP [1] による従来のジェスチャ認識手法 [2]~[9] の基本的な考え方を説明する。連続 DP は、連続的に入力される時系列パターンの中に、標準パターンと類似した区間を見出す処理、いわゆるスポッティング認識を実現可能な手法である。またフレーム同期処理ができる点も特長の一つである。

以下では、システムにあらかじめ登録されてる標準ジェスチャパターンを特徴ベクトルの時系列  $\mathbf{R}_{c,1}, \dots, \mathbf{R}_{c,t}, \dots, \mathbf{R}_{c,T_c}$  で表す。ここで  $c$  はジェスチャの種類を表す添字である。また各特徴ベクトル  $\mathbf{R}_{c,t}$  は、フレーム  $t$  での動作状態 (例えば、エッジ画像 [2]、濃淡画像 [4]、特定部位の回転角 [9]、基本動作ラベル [8]) を表すベクトルである。標準パターンと同様に、認識対象とする連続的なジェスチャパターン (入力パターン) を特徴ベクトルの時系列  $\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_\tau$  で表す。ここで  $\tau$  は現在の時刻を表すものとする。

従来の連続 DP による認識では、次の漸化式を各入力フレーム  $\tau$  で計算する (図 1)。

$$g_{c,t}(\tau) = \min \begin{cases} g_{c,t-1}(\tau-1) + 3d_{c,t}(\tau) \\ g_{c,t-1}(\tau-2) + 2d_{c,t}(\tau-1) + d_{c,t}(\tau) \\ g_{c,t-2}(\tau-1) + 3d_{c,t-1}(\tau) + 3d_{c,t}(\tau) \end{cases} \quad (1)$$

ここで  $d_{c,t}(\tau)$  は入力パターンのフレーム  $\mathbf{I}_\tau$  を標準パターンのフレーム  $\mathbf{R}_{c,t}$  に対応させた場合の局所距離  $\|\mathbf{I}_\tau - \mathbf{R}_{c,t}\|$  である。漸化式 (1) に従って累積距離  $g_{c,t}(\tau)$  を各  $\tau$  のすべての  $c, t$  について計算することで、フレーム  $\tau$  での認識結果  $c^*$  が次のように得られる。

$$c^* = \underset{c}{\operatorname{argmin}} g_{c,T_c}(\tau) \quad (2)$$

以上により入力パターンの区間  $[\tau', \tau]$  が標準パターン  $c$  に対応したというスポッティング認識が可能となる。始点  $\tau' (1 \leq \tau' < \tau)$  は、バックトラック処理を別途行うことで明示的に求められる。漸化式 (1) は入力パターンのフレーム  $\tau$  に同期して計算できるので、認識結果  $c^*$  を時々刻々と出力できる。

図 1 の破線で示したマッチング経路からも予想されるように、従来法では原理的にジェスチャ開始後  $T_c/2$  フレーム以上経たないと認識結果は出力されない (時刻  $\tau'$  において発生したパスは、 $\tau' + T_c/2$  にならないと  $t = T_c$  に到達しない)。すなわち必ず一定時間の遅れが発生する。このように、従来の連続 DP 法はスポッティング機能とフレーム同期処理により実時間処理に適した枠組みではあるものの、早期認識法としては不十分であると言える。

### 2.2 早期認識のための変更

本手法では、早期認識のために (2) を変更した次の識別規則を用いる。

$$(c^*, t^*) = \underset{c,t}{\operatorname{argmin}} (g_{c,t}(\tau)/t) \quad (3)$$

識別規則 (2) との違いは、標準パターンの冒頭部の部分パター

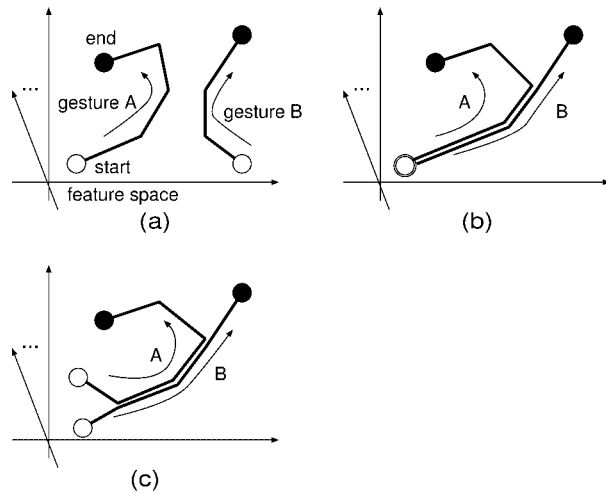


図2 2つのジェスチャ間の関係. (a) 共通部分なし. (b) 冒頭部分共通. (c) 中間部分共通.

ン  $R_{c,1}, R_{c,2}, \dots, R_{c,t}$  ( $t \leq T_c$ ) も評価の対象となっている点である. 各部分パターンはその長さ  $t$  によって正規化された後, 入力パターンと比較されていることになる. この識別規則により, 現フレーム  $\tau$  がジェスチャ  $c^*$  の第  $t^*$  フレームに相当するという結果を出力できる.

ところで, この識別規則中の最小値選択においてあまりに小さな  $t$  を許すと, 非常に短い区間でのマッチングの結果を信頼することになり, 結果が不安定になる場合がある. この実行上の問題のために, 識別規則 (2) の最小値選択において  $t \geq t_{\min}$  という制約を設ける. この場合, ジェスチャ開始後  $t_{\min}/2$  フレーム以下での早期認識は原理的に不可能となる.

文献 [10] では, Reference Interval-free 連続 DP という, 標準パターンの任意の部分区間と入力パターンの任意の部分区間のマッチング法が提案されている. さらに西村ら [6] は, 過去の影響を指数関数的に低減させるディスカウントの概念を取り入れることで, この部分対部分のマッチングをより効率的に行なう手法を提案している. 本手法はこれら部分対部分マッチングの特殊な場合 (標準パターンのマッチング区間の始点が必ず  $t = 1$  となるように制限したもの) と考えることができる. また, 文献 [11], [12] では, 他の早期認識のアプローチとして, 通常の連続 DP に加え, サポートと呼ばれる曖昧性の有無を判定するための厳しい基準を利用する方法を提案している.

### 2.3 早期認識の限界

以下では, 以上の早期認識手法の性能限界について, 図2を用いて議論する. 同図は2つのジェスチャAおよびBの標準パターンを特徴空間における軌跡として表したもので, それらの関係はおおよそ同図 (a)-(c) の3通りに大別される. これら以外にも幾つかの特別な場合 (例えばBがAに完全に包含された場合やBがAの途中から始まるような場合) も考えられるが, それらについては (a)-(c) に関する議論から説明できるので略す.

- 図2(a)は, ジェスチャAとBが全く共通部分を持たない場合である. この場合, 現入力フレーム  $\tau$  の特徴ベクトル  $I_\tau$  について (3) 式により  $(c^*, t^*)$  が得られれば,  $c^* \in \{A, B\}$  の軌

道にいたことが直ちにわかり, 早期認識できる. このように図2(a)は早期認識が容易な場合である.

- 図2(b)は, 早期認識に限界がある場合である. 具体的には, ジェスチャAとBが類似した冒頭部分を持つ場合であり, より簡単に言えば, ジェスチャAとBが最初は全く見分けがつかない状況である. この場合,  $(c^*, t^*)$  がこの共通部分にあると, 早期認識は原理的に不可能である. より正確に論ずるために, ジェスチャAの標準パターンの冒頭  $t_a$  フレーム  $R_{A,1}, R_{A,2}, \dots, R_{A,t_a}$  とジェスチャBの冒頭  $t_b$  フレーム  $R_{B,1}, R_{B,2}, \dots, R_{B,t_b}$  が類似しているとする. このとき, ジェスチャAの冒頭区間  $[1, t_a]$  が入力パターンの区間  $[\tau', \tau]$  に対応しているとすれば, ジェスチャBの冒頭区間  $[1, t_b]$  もやはり同一区間  $[\tau', \tau]$  に対応しているはずである. よって  $g_{A,t_a}(\tau)/t_a \sim g_{B,t_b}(\tau)/t_b$  が任意の  $\tau$  において成り立つことになり, もし  $I_\tau$  がジェスチャAの  $t_a$  フレームだとしても, それをジェスチャBの  $t_b$  フレームと誤認識する可能性が半分はある.

- 図2(c)は, 中間部分が共通している場合である. この場合は, 共通部分があっても, ジェスチャの早期確定が可能である. 入力がジェスチャAだったと仮定する. このとき, 冒頭部の非共通部分では  $g_{A,t}(\tau) < g_{B,t}(\tau)$  となり, (3) 式で与えられる  $c^*$  はAとなる. 共通部分だけのマッチング距離はAにBも同じであるから, 結局共通部分の累積距離には, 非共通部分の累積距離の差がそのまま残る. 従って  $c^*$  はAとなる. このように識別結果は冒頭からAのままであり, 原理的には図2(a)と同様の早期認識が可能である.

以上のように, どのジェスチャとどのジェスチャがどのように共通部分を持つかによってどの時点で早期認識が可能になるかが決まる. 従って, こうしたジェスチャの共通部分をあらかじめ把握しておくことが重要になってくる. 4.1節ではその把握手段について展望する.

### 2.4 早期認識実験

以上で述べた早期認識法の基本的性能を観察するために, 予備的な実験を行った.

#### 2.4.1 認識タスクと実験試料

想定したジェスチャは「さようなら (bye)」「万歳 (banzai)」「指差し (point)」の3種である. 「さようなら」は右手を顔付近の高さに上げ, 左右に2回振り, 元の位置に戻す動作, 「万歳」は両手の上げ下げを2回繰り返す動作, 「指差し」は右手を顔付近の高さに上げ, 前後に2回振った後, 元の位置に戻す動作である. このように「さようなら」と「指差し」の前半, 顔付近の高さまで右手を上げる部分は同じ動作となっている (すなわち図2(b)の関係にある). なお, いずれのジェスチャも開始・終了時には両手を下げた状態としているが, 本実験ではこの事実を積極的に使っていない.

成人男性一人に以上3種のジェスチャをそれぞれ10回繰り返してもらい, 合計30パターンを取得した. これら30パターンの平均フレーム数は約83フレームであった. 各フレームの特徴ベクトルは, 顔の位置を基準とした右手先および左手先の3次元位置特徴6次元, およびそれらの時間差分 (速度) 特徴6次元からなる計12次元特徴ベクトルである. この位置特徴

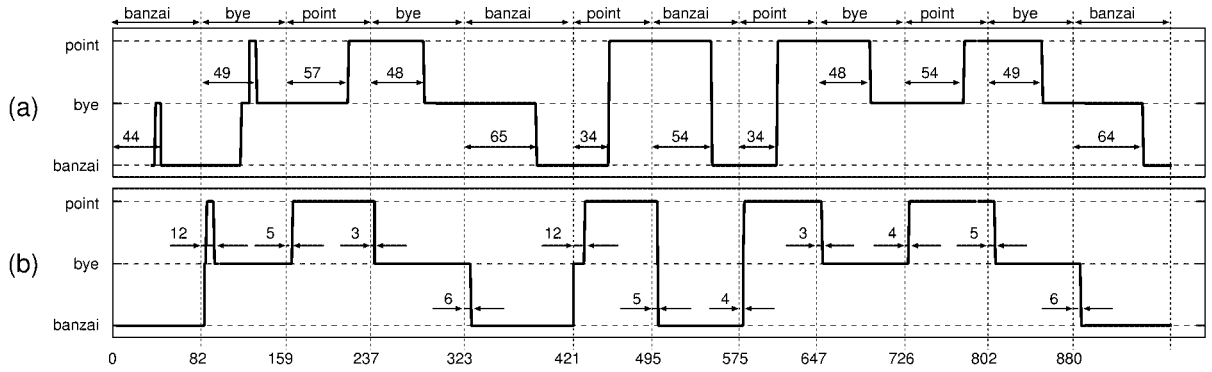


図3 ジェスチャ認識結果. (a) 従来の連続 DP 法. (b) 本手法. 図上部には正解ジェスチャとその範囲を示した. また矢印上の数字は正しい認識結果が得られるまでに要したフレーム数.

は, (i) まずユーザの前方に置かれた 2 台の IEEE1394 カメラ (Sony 製 DFW-X700, 15 フレーム/秒) により距離画像をステレオ計測し, (ii) 次に肌色検出により両手と顔部分を同定することで自動取得したものである.

#### 2.4.2 認識結果

図 3 は従来法および本手法による認識結果である. 認識対象とした入力データは, 前述の 30 パターンから 12 パターン (各ジェスチャ 4 パターンずつ) をランダムに抜き取り, それらをランダムに並べることによって得た, 擬似的な連続ジェスチャパターンである. 残りの 18 パターンは標準パターンとして使用した. また認識時のパラメータ  $t_{\min}$  は 5 とした.

図 3(a) にあるように, 従来法では認識結果が得られるまでに 34~65 フレーム, すなわち標準パターン長の半分程度の遅れが出ている. これは 2.1 節で述べた従来の連続 DP の性質に因る遅れである. 一方, 同図 (b) に示す本手法の結果を見ると, 認識結果が得られるまでの遅れは最大でも 12 フレームであり, 従来法に比べて大幅に遅れを短縮できていることがわかる. この結果から, 早期認識のための識別基準 (3) が有効に機能することが示唆された.

次に 2.3 節で述べたジェスチャの曖昧性が早期認識に与える影響について観察する. 2.4.1 節で述べたように, 「指差し」と「さようなら」のジェスチャの冒頭部分は共に右手を顔の高さまで上げる動作であり, 図 2(b) の状況に相当する. 従って, 冒頭の共通動作段階 (10~15 フレーム) では安定して正確な早期認識を行うことは原理的に不可能である. 図 3(b) の結果にはこの曖昧性の影響が現れている. 具体的には, 他に比べて認識遅れが出ている場合を見ると, 第 1 回目の「さようなら」(12 フレーム) と, 第 2 回目の「指差し」(12 フレーム) であった. 一方, 他と共通部分の少ない「万歳」については遅れは常にその半分以下であった. 以上から, 図 2(b) のような曖昧性がある状況では, 早期認識が誤り易いという 2.3 節での考察が実験的に検証された.

### 3. 早期認識に基づく動作予測

本節では, まず 3.1 節において, 前述の早期認識アルゴリズムを利用した, 単純な動作予測法について述べる. 動作予測と

は, 1. 節で述べたように, 現時刻の数フレーム後にはユーザがどのような動作状況にあるかを予測し, 特徴ベクトルとして出力するものである. 次に 3.2 節において, この単純な動作予測法の限界について考察する. 最後に 3.3 節において, この考察を検証する.

#### 3.1 動作予測の原理

識別規則 (3) により, 現フレーム  $\tau$  がジェスチャ  $c^*$  の  $t^*$  フレームに相当すると同定された, すなわち,  $I_\tau$  が  $R_{c^*, t^*}$  に対応すると同定されたとする. このとき,  $\delta$  フレーム後には  $R_{c^*, t^* + \delta}$  に似た動作が現れると予測できる. すなわち,  $I_{\tau + \delta}$  の予測値を  $\hat{I}_{\tau + \delta}$  とすると,

$$\hat{I}_{\tau + \delta} = R_{c^*, t^* + \delta} \quad (4)$$

となる. この単純な予測式では, 標準パターンと入力パターンのジェスチャの速度が等しいということが暗に仮定されている. 速度変化の補償は今後の課題とする.

#### 3.2 予測限界と早期認識との関係

本節では, (4) 式による予測の限界, 換言すれば (4) 式において  $\delta$  をどの程度大きくできるかについて考察する.

まず早期認識が可能な時点について考える. すなわち, 図 2 (a), (c) の場合や同図 (b) の非共通部分の場合である. 早期認識ができるということは, 現時点以降のジェスチャ軌道を確定できることを意味し, 従ってジェスチャ終了まで動作を予測できることを意味する. よって, (3) 式により  $(c^*, t^*)$  が求まれば, 以降の動作は  $R_{c^*, t^*}, R_{c^*, t^* + 1}, \dots, R_{c^*, T_c^*}$  として予測できる.

次に早期認識できない場合, すなわち図 2 (b) の冒頭の共通部分の場合を考える. この共通部分では現在のジェスチャ  $c$  を確定できないものの, 今後しばらくは共通部分の軌道を通っていくことだけは確定している. 従って, 早期認識できない場合であっても, 共通部分の範囲内での予測は可能である. 例えば, 現時点が「さようなら」と「指差し」の冒頭部であるとすれば, どちらのジェスチャであるかは確定できない (= 早期認識はできない) が, 「顔付近の高さまで手を上げる」ということは確定できる (= 予測できる).

#### 3.3 動作予測実験

図 4 は, (4) 式による動作予測結果である. 同図 (a) は早期

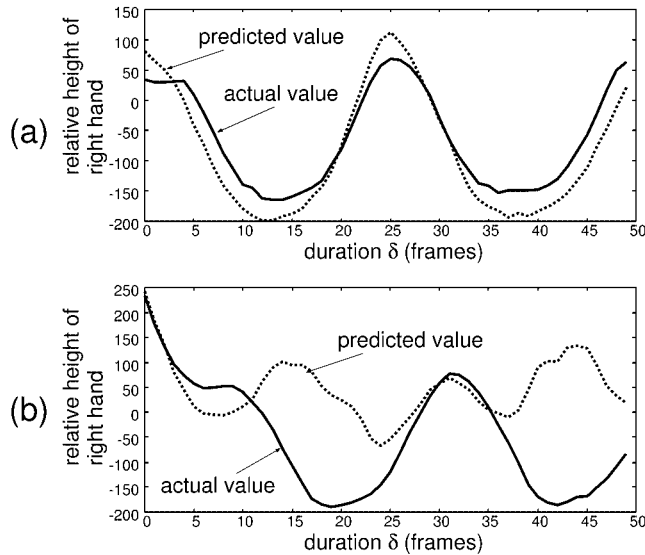


図4 右手高さの予測結果. 横軸は予測幅.

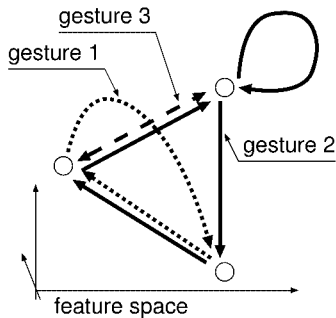


図5 ジェスチャネットワーク

認識できている時点(図3(b)の170フレーム目)において、予測幅 $\delta$ を変えながら得られた予測値 $\hat{I}_{\tau+\delta}$ である。ただし、同図では12次元特徴ベクトル $\hat{I}_{\tau+\delta}$ のうち、右手の高さ成分だけを示している。このように、50フレーム程度先まで良好な予測ができています。一方、同図(b)は「さようなら」と「指差し」の冒頭の共通部において早期認識を誤っている時点(図3(b)の425フレーム目)における同様の結果である。このように誤認識であっても共通部が終わるまで( $\delta < 10$ )は予測できているが、それ以降は予測が大きく外れている。

## 4. 今後の課題

### 4.1 ジェスチャネットワークの構築

以上では、比較的簡単な手法により早期認識ならびに動作予測を実装し、またそれらの限界について見てきた。いずれにおいても、ジェスチャ軌道間の関係からその原理的な限界が決まった。従って、早期認識および動作予測を今後システムティックなものにするためには、まずは図2のような軌道モデルを構築することが肝要と思われる。

複数のジェスチャを考えると、図2は図5のようなネットワーク構造になる。連続DPによるスポットニング認識を行うと同時に、現時点がこのネットワーク上のどこに相当するかを同定しておけば、早期認識の可否ならびに行動予測可能範囲を

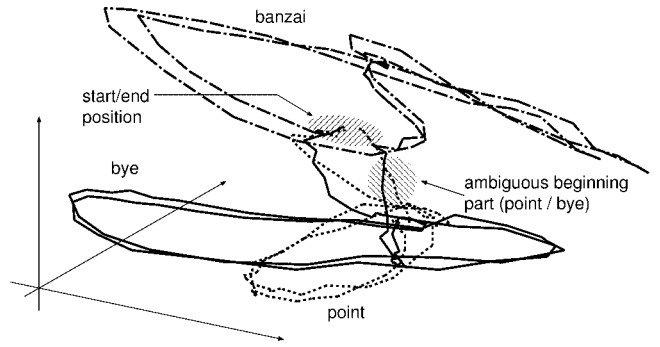


図6 異なるカテゴリに属する3種類のジェスチャの軌跡. 本来の12次元特徴空間を3次元部分空間に投影して表示している. すべてのジェスチャは両手を下げた状態から開始し、終了する. 図中央付近がその状態に相当.

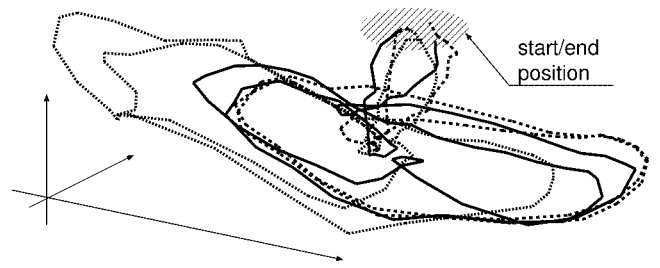


図7 同じカテゴリ「さようなら」に属する3つのジェスチャの軌跡

定めることができるようになると思われる。図6および図7に示した実際のジェスチャ軌道は、こうしたネットワークが生成可能であることを示唆している。すなわち、図6を見ると、異なるジェスチャの非共通部分は異なる軌道を示しており、一方、共通部分の軌道は類似している。また図7を見ると、同じカテゴリのジェスチャは、多少のゆらぎはあるにしろ、およそ同じ軌道を描いている。従って、今後は、類似軌道を内包するように(例えば確率的に)膨らませた有向辺を持つジェスチャネットワークを構築すればよいと言える。

### 4.2 モーションプリミティブとの関係

モーションプリミティブ(基本動作)はこれまでも様々な定義されてきた。大崎ら[8]は速度変化の大きな時点を基準としてジェスチャを分割し、得られたセグメントをDP距離に基づいてクラスタリングすることでモーションプリミティブを得ている。澤田ら[13]も同様に加速度を用いて手話を基本動作に分解する方法を提案している。Sanger[14]はジェスチャ時系列の最適基底としてモーションプリミティブを定義している。Fodら[15]は以上のアプローチを混合した方法を提案している。

これに対し、図5のジェスチャネットワークのエッジ間を新たに定義されたモーションプリミティブと考えることができる。こうして定義されたモーションプリミティブには、従来のモーションプリミティブにはない、次のような性質がある。

- 1つのモーションプリミティブは動作予測可能な範囲を表している。
- モーションプリミティブの時系列によりジェスチャを表現できる。
- 対象とするジェスチャの集合に応じて動的に構成される。

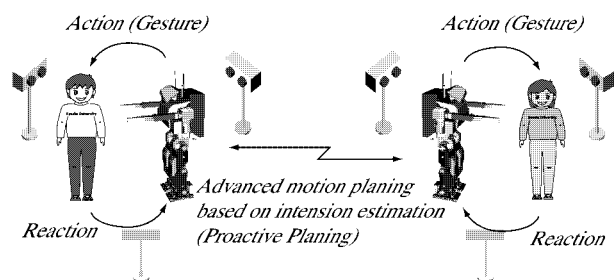


図8 ヒューマノイドを用いたプロアクティブヒューマンインタフェースによる遠隔地コミュニケーション

(従来のモーションプリミティブは加速度の大きさや一定時間長といった絶対的な基準によって決まる)

こうして定義されるモーションプリミティブの有用性については今後明らかにする必要があるが、有用性の基準がジェスチャ記述能力と独立性(非冗長性)であるとすれば、両基準を十分に満足するものと期待できる。

#### 4.3 プロアクティブヒューマンインタフェースの実現

以上の早期認識法および動作予測法は、筆者らが検討しているプロアクティブヒューマンインタフェースの実現のための要素技術である。このインタフェースは次の二つの性質を有する。

- (1) 実体を伴い、人間に対して物理的な働きかけが可能
- (2) 行動意図を推定・予測して先回りする機能を具備

これら2性質はいずれも、様々な人にとって使い易いコンピュータシステムの枠組みの提供することを目的としている。性質1はインタフェースに身体性[16]を持たせることで実世界との乖離が少ない状況での情報授受を目指すものである。一方、性質2の予測に基づく先回り機能はこのインタフェースの独創的な点の1つであり、本報告で議論した早期認識ならびに動作予測が利用できる。

図8はこうしたプロアクティブヒューマンインタフェースの一実現形態である。これは円滑な遠隔コミュニケーションを目指したインタフェースである。ヒューマノイドがインタフェース本体であり、一方のユーザの行動(ジェスチャ)を模倣によりもう一方のユーザに伝達する。このインタフェース系における、早期認識および動作予測の役割は以下の通りである。

- ユーザがジェスチャを途中で止めても、早期認識の結果を用いてヒューマノイドに引き続き動作を継続させることができる。これによりユーザがジェスチャ全体を提示する必要がなくなる。
- 動作予測の結果に従ってヒューマノイド側の動き制御を早期に開始させる。この処理により、回線混雑もしくはヒューマノイドのハードウェア的限界による動作表現の時間遅れを低減できる。

## 5. まとめ

本報告では、ジェスチャの早期認識と動作予測について述べた。この早期認識とは、ジェスチャの開始から認識結果が出力されるまでの遅れを極力小さくすることに相当する。従来より用いられている連続DP法に若干の変更を加えたアルゴリズム

を用い、早期認識がどの程度可能かを調査し、またその原理的な限界についても考察した。

今後は、まず4.1節で述べたジェスチャネットワークの生成法を確立する。次いでそれを用いて早期認識を行い、さらに4.3節のプロアクティブヒューマンインタフェースの駆動実験を行いたいと考えている。またこれらの検討と並行して、4.2節のモーションプリミティブについても考察していきたい。

## 謝辞

本研究の一部は総務省戦略的情報通信研究開発推進制度の支援を受けた。

## 文献

- [1] 岡, “連続 DP を用いた連続単語認識,” 日本音響学会音声研究会資料, S78-20, 1978.
- [2] 高橋, 関, 小島, 岡, “ジェスチャー動画像のスポッティング認識,” 信学論, vol. J77-DII, no. 8, pp. 1552-1561, 1994.
- [3] 佐川, 酒匂, 大平, 崎山, 阿部, “圧縮連続 DP 照合を用いた手話認識方式,” 信学論, vol. J77-DII, no. 4, pp. 753-763, 1994.
- [4] 太田, 潮崎, 新井, “動的計画法に基づくマッチングによる運動認識,” 精密工学会誌, vol. 63, no. 6, pp. 812-818, 1997.
- [5] 西村, 向井, 野崎, 岡, “低解像度特徴を用いた複数人物によるジェスチャの単一動画像からのスポッティング認識,” 信学論, vol. J80-DII, no. 6, pp. 1563-1570, 1997.
- [6] 西村, 古川, 向井, 岡, “時系列パターン検索のための重み減衰型 Reference Interval-Free 連続 DP について,” 信学論, vol. J81-DII, no. 3, pp. 472-482, 1998.
- [7] 西村, 野崎, 向井, 岡, “連続 DP への非単調性導入によるジェスチャ動画像からの戸惑い動作のスポッティング認識,” 信学論, vol. J81-DII, no. 1, pp. 18-26, 1998.
- [8] 大崎, 嶋田, 上原, “速度に基づく切り出しとクラスタリングによる基本動作の抽出,” 人工知能学会誌, vol. 15, no. 5, pp. 878-886, 2000.
- [9] 呉, 木戸, 塩山, “ジェスチャ認識のための連続 DP の改良,” システム制御情報学会論文誌, vol. 14, no. 6, pp. 283-290, 2001.
- [10] 伊藤, 木山, 小島, 関, 岡, “時系列標準パターン上の任意区間によるスポッティングのための Reference Interval-free 連続 DP (RIFCDP),” 信学論, vol. J79-D-II, no. 9, pp. 1474-1483, 1996.
- [11] 倉爪, 内田, 長谷川, 谷口, “プロアクティブヒューマンインタフェースの研究 — 第1報 人間型アクティブインタフェースの開発 —,” 日本機械学会ロボティクス・メカトロニクス講演会, 1A1-H-76, 2004.
- [12] 内田, 倉爪, 谷口, 長谷川, 迫江, “プロアクティブヒューマンインタフェースのためのジェスチャの早期認識に関する検討,” 第3回情報科学技術フォーラム, 6N-7, 2004.
- [13] 澤田, 橋本, 松島, “運動特徴と形状特徴に基づいたジェスチャー認識と手話認識への応用,” 情報処理学会論文誌, vol. 39, no. 5, pp. 1325-1333, 1998.
- [14] T.D. Sanger, “Optimal movement primitives,” Advances in Neural Information Processing Systems, vol. 7, pp. 1023-1030, 1995.
- [15] A. Fod, M.J. Mataric, and O.C. Jenkins, “Automated deviation of primitives for movement classification,” Autonomous Robots, vol. 12, no. 1, pp. 39-54, 2002.
- [16] 岡田, 三嶋, 佐々木 (編), 身体性とコンピュータ, 共立出版, 2001.