ジェスチャの早期認識・予測ならびにそれらの高精度化のためのネットワークモデルに関する検討

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

† 九州大学大学院システム情報科学府　〒 812–8981 福岡市東区箱崎 6–10–1
E-mail: †mori@human.is.kyushu-u.ac.jp

あらまし　本論文ではジェスチャの早期認識，及びその認識結果に基づいた動作予測について検討する．早期認識とはジェスチャ入力の初期段階における識別を可能とするものである．一方，動作予測とはジェスチャ動作者の数フレーム後の姿勢を推定するものである．さらに，本研究ではジェスチャ間の関係によって構築されるネットワークに対して上記の手法を適用することを検討する．ネットワークを用いることで，早期認識および動作予測がもつ原理的な限界による影響を抑えることが可能となる．本研究では比較的単純なアプローチによって上記手法の基本的な性質を明らかにする．最後に，これらの実験結果を踏まえ，ヒューマンインタフェースの開発など，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，ジェスチャネットワーク，モーションプリミティブ

Early Recognition and Prediction of Gestures with Network Model

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

† Graduate School of Information Science and Electrical Engineering, Kyushu University　6–10–1 Hakozaki, Higashi-ku, Fukuoka-shi, Fukuoka, 812–8581 Japan
E-mail: †mori@human.is.kyushu-u.ac.jp

Abstract　This paper concerns two topics on gesture recognition. The first topic is early recognition of gestures: the recognition result of a gesture is provided at the beginning part of the gesture. The second topic is motion prediction: the subsequent posture of the person who makes a gesture is predicted by using the result of early recognition. In addition to them, this paper concerns a network model constructed for inproving the performance of early recognition and motion prediction. The effectiveness of these methods was shown by experimental results.

Key words　gesture recognition, early recognition, motion prediction, dynamic programming, gesture network, motion primitive

1. はじめに

本論文では，ジェスチャの早期認識，及びその認識結果に基づいた動作予測について検討する．早期認識とはジェスチャ入力の初期段階における識別を可能とする手法である．一方，動作予測とは動作者の数フレーム後の姿勢を推定する手法である．さらに本研究では，これらの高精度化のためにネットワークモデルの利用ついて検討する．

ジェスチャの早期認識 [1] は，ジェスチャに基づいた効率的なマンマシンインタフェースの構築に有用である．例えば，早期認識によって認識結果をジェスチャの開始時点付近で確定できれば，残りのジェスチャは不要となる．つまり，この時点で動作者はジェスチャを止めても良いことになり，省力化に有効である．こうした早期認識に関する研究は，音声認識の分野には見られるが [2]，ジェスチャ認識の分野では検討されていなかった．すなわち，従来法のほとんどがジェスチャの入力が完了した時点で認識結果を出力するものであった．

動作予測は，プロアクティブ（先回り）ヒューマンインタフェースを実現する上で必要となる技術である．先回りとは，動作者がジェスチャを終える前にシステムが次の行動をとるこ

とであり，自然な対話の実現に有効な処理である．これに加え，動作予測は人間の行動に対して何かを応答するようなシステムの遅れ補償に有用である．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，動作予測によって数フレーム後の姿勢パラメータを求め，それをヒューマノイドに渡しておけば，遅れを最小化することができる．なお，本論文では早期認識の結果により動作予測が可能であることを示す．

このように，ジェスチャの早期認識及び動作予測は有用な技術であるが，これらの手法には本質的な限界が存在する．後述するように，認識対象となるジェスチャの中に共通した動作で開始するジェスチャが複数存在した場合，これらのジェスチャの冒頭部分では早期認識を行うことができない．また，動作予測は早期認識の結果に基づいているため，このような場合には適切な予測を行うことができなくなってしまう．

この問題に可能な限り対処するため，本研究では認識・予測の際にネットワークモデルを用いることを検討する．すなわち，ジェスチャ間の関係からネットワークを構築し，このネットワークに対して上記の早期認識および動作予測を適用する．ジェスチャネットワークを用いることで，ジェスチャ間の共通部分を考慮した認識・予測を行うことができるようになり，本質的な限界による影響を抑えることができる．

本研究では，これらの手法の有効性を確認するために，認識対象として 18 種のジェスチャを想定した実験を行った．

2. 早 期 認 識

2.1 連続 DP による従来の認識手法

本節では，連続 DP を用いた従来の認識手法の基本的な考え方について説明する．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，すなわちスポッティング認識を実現可能な手法である [3]．また，フレームと同期して処理が出来る点も特徴の一つである．このため，連続 DP は実時間ジェスチャ認識に非常に適している．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 [4]∼ [10]．

以下では，システムにあらかじめ登録されている標準ジェスチャパターンを特徴ベクトルの時系列 $\boldsymbol{R}_{c,1},\ldots,\boldsymbol{R}_{c,t},\ldots,\boldsymbol{R}_{c,T_c}$ 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 $\boldsymbol{R}_{c,t}$ は，フレーム t での動作状態を表すベクトルである．標準パターンの場合と同様に，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 $\boldsymbol{I}_1,\boldsymbol{I}_2,\ldots,\boldsymbol{I}_\tau,\ldots$ で表す．ここで τ は現在の入力フレームを表す．

従来法は以下の漸化式を各フレーム τ で計算する (図 1)．

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

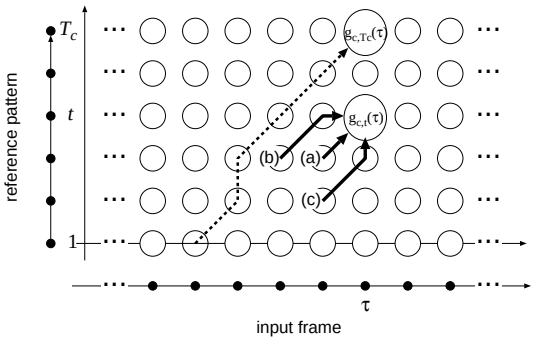

図 1　連続 DP による従来のジェスチャ認識.

ただし，$g_{c,1}(\tau)=3d_{c,1}(\tau)$ とする．ここで $d_{c,t}(\tau)$ は入力パターンのフレーム τ を標準パターン c のフレーム t に対応させた場合の局所距離 $||\boldsymbol{I}_\tau-\boldsymbol{R}_{c,t}||$ である．(1) 式中の (a)∼(c) は図 1 の経路 (a)∼(c) に対応する．(1) 式に従って累積距離 $g_{c,t}(\tau)$ を全ての c,t で計算することによって，フレーム τ での認識結果 c^* が次のように得られる．

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

式 (1) は入力パターンのフレーム τ 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．

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になる．すなわち，入力パターンをあらかじめセグメンテーションしておく必要が無い．始点 $\tau'(1\leqq\tau'<\tau)$ は，バックトラック処理を別途に行うことで明示的に求めることができる．

多くの場合，この従来法による認識は適切に機能するが，早期認識の手法としては不十分である．従来法ではそれぞれの標準パターン全体との類似区間を入力パターンから探し出す (図 2(a))．このため従来法は，原理的にはジェスチャ全体が完全に入力されるまで認識結果を出力できない．より詳細には，図 1 の破線で示されたマッチング経路からも予想されるように，フレーム τ' において発生したマッチング経路は，最短でも $\tau'+T_c/2$ にならないと $t=T_c$ に到達しない．すなわち，ジェスチャ開始後，認識結果を得るまでには必ず一定時間の遅れが発生する．

2.2 早期認識の原理

提案する認識手法では，入力中のジェスチャがまだ完全に動作を終了していなくても結果を出力することができる．提案手法の基本的な考え方を図 2(b) に示す．また，以下でその具体的な内容を述べる．

早期認識のために，提案手法では従来の連続 DP に対して簡単な変更を行う．具体的には，フレーム τ において (2) の変わりに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，冒頭部の部分パターン $\boldsymbol{R}_{c,1},\boldsymbol{R}_{c,2},\ldots,\boldsymbol{R}_{c,t}(t\leqq T_c)$ 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に

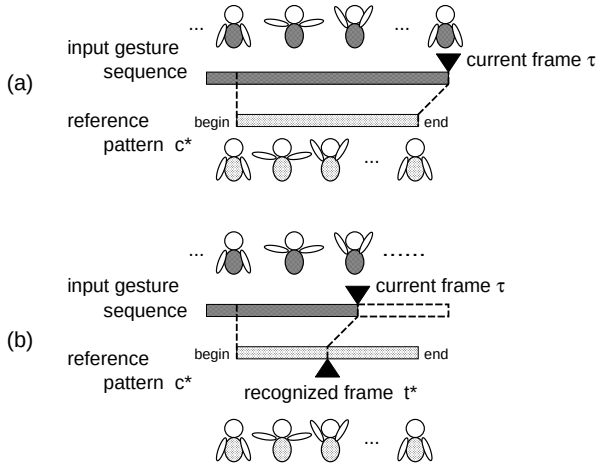

図 2　(a) 従来のジェスチャ認識手法．(b) 提案する早期認識手法．

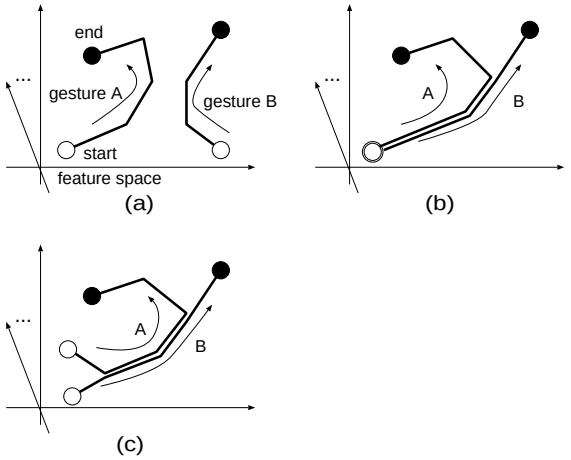

図 3　2 つのジェスチャ間の関係．(a) 共通部分なし．(b) 冒頭部分共通．(c) 中間部分共通．

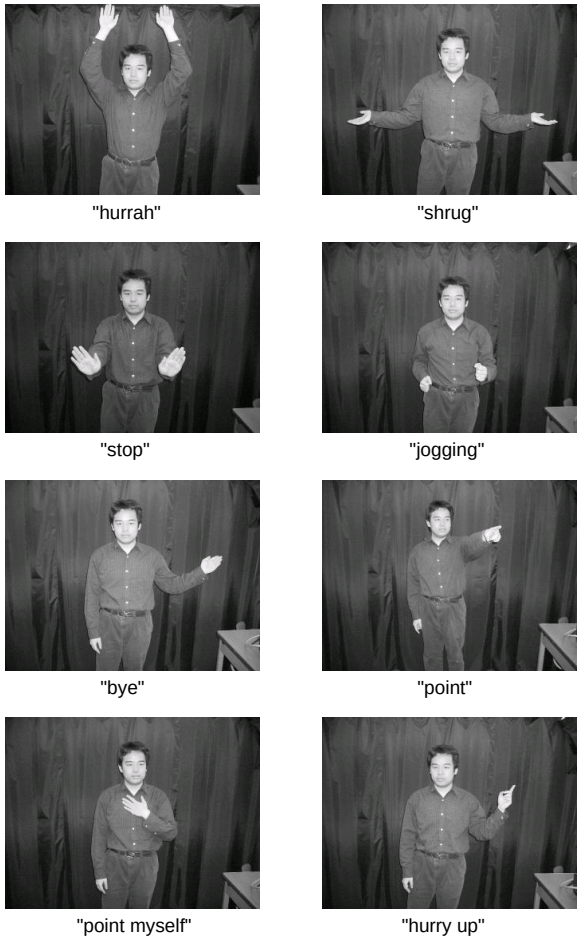

図 4　実験で想定したジェスチャの一部．上の 4 つは両手を使ったジェスチャ．下の 4 つは片手のみを使ったジェスチャ．

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 τ 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．

早期認識の結果を安定させるために，識別規則 (3) に $t \geqq t_{\min}$ という制約を課す．この制約は非常に短い区間でのマッチングによる結果を排除する．この制約を用いた場合，ジェスチャが入力されてから最初の $t_{\min}/2$ フレームまでは認識結果を出力することができない．

2.3　早期認識の限界

早期認識には本質的な限界がある．以下では，この限界について図 3 を用いて議論する．図 3 は 2 つのジェスチャA 及び 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，それらの関係はおよそ同図 (a)∼(c) の 3 通りに大別される．これら以外にもいくつかの特別な場合 (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省略する．

図 3(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，識別規則 (3) により現フレーム τ の入力が (c^*, t^*) に対応するとわかれば，どちらの軌道上にあるかがわかる．すなわち，入力されたジェスチャが t^* の値によらず A，B のいずれであるかを直ちに区別できる．このように図 3 (a) は早期認識が容易な場合である．

図 3(b) は，ジェスチャA と B の冒頭部分が共通の動作だった場合である．この場合，現フレーム τ の入力が (c^*, t^*) であるという結果が得られても，t^* がこの共通部分であれば，入力されたジェスチャが正しく c^* であるとは限らない．これは，ジェスチャ A の標準パターンの冒頭 t_a フレーム $\boldsymbol{R}_{A,1}, \boldsymbol{R}_{A,2}, \ldots, \boldsymbol{R}_{A,t_a}$ とジェスチャB の冒頭 t_b フレーム $\boldsymbol{R}_{B,1}, \boldsymbol{R}_{B,2}, \ldots, \boldsymbol{R}_{B,t_b}$ が類似している場合には，入力パターンの区間 $[\tau', \tau]$ がジェスチャ A の冒頭区間 $[1, t_a]$ に対応するとき，同時にそれがジェスチャ B の冒頭区間 $[1, t_b]$ にも対応するからである．つまり，仮に c^* =A であっても，本当は c =B であるという可能性が半分ある．このように，ジェスチャの冒頭部に曖昧性があると，早期認識は不可能である．

図 3(c) は，ジェスチャA とジェスチャB の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，識別規則 (3) 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ジェスチャA もジェスチャB も同じなので，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

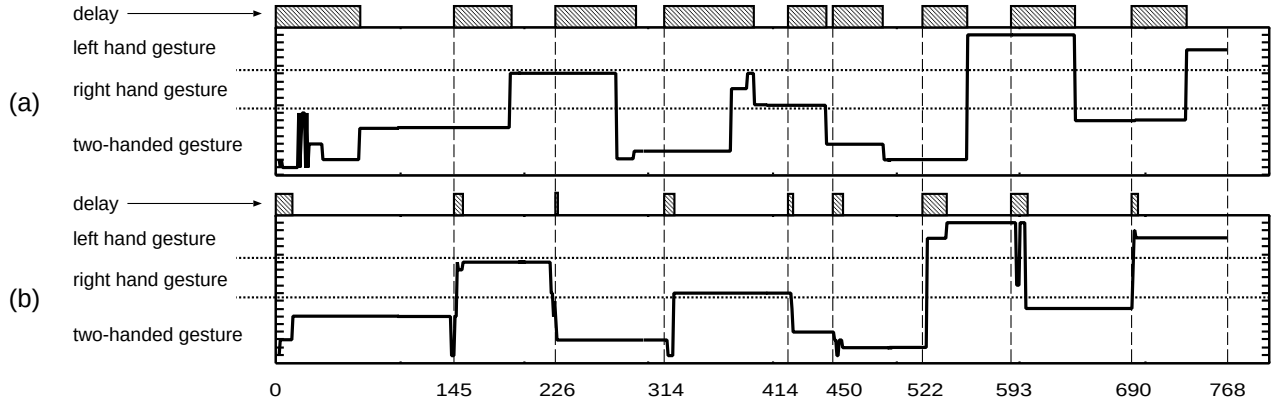

図 5　ジェスチャ認識結果の一部．(a) 連続 DP を用いた従来の認識法．(b) 本手法．実線は認識結果の推移を示す．グラフ上部の斜線部は正しい認識結果が得られるまでの時間を示す．

がそのまま残ることになる．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 3(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もつかによって，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調べるために，実験を行った．本実験では 18 種類のジェスチャを想定した．図 4 にその一部のジェスチャについて，特徴的なフレームを示す．このうち 8 種は両手を使うジェスチャ，5 種は片手のみを使うジェスチャで，片手のジェスチャについては右手で動作を行ったものと左手で動作を行ったものとを別の種類のジェスチャとして区別している．両手を使うジェスチャ8 種のうち 4 種は，冒頭部が共通の動作になっている．また片手のみを使うジェスチャ 5 種は全て，冒頭部が共通の動作になっている．すなわち，18 種のジェスチャのうち 14 種には図 3 (b) のように冒頭部における曖昧性が存在する．いずれのジェスチャも開始・終了時には手を下げた状態である．

本実験では，成人男性一人が各ジェスチャを 6 回ずつ行ったもの (計 108 パターン) を使用した．これら 108 パターンの平均フレーム長は約 89 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及び左手先の 3 次元位置からなる 6 次元特徴ベクトルである．この位置特徴は，(i) まずユーザの前方におかれた 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．

標準パターンとして各ジェスチャにつき 3 パターンずつ (計 54 パターン) を使用した．認識対象とした入力データは，残りの 54 パターンから 9 パターンをランダムに抜き取り，それらをランダムに並べることで得た疑似的な連続ジェスチャパターンである．この入力データをパターンの重複が無いように 6 つ作成し，早期認識を行った際の認識までに要したフレーム数を計測する．認識時のパラメータ $t_{\min}$ は 5 とした．

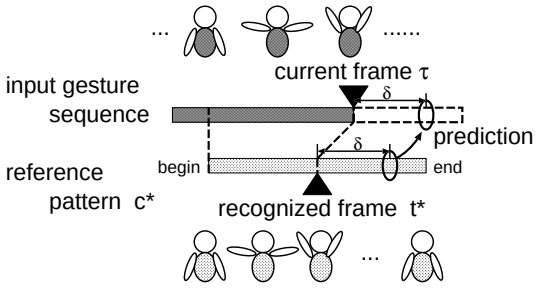

図 6　早期認識に基づく動作予測手法の原理．

実験の結果の一部を図 5 に示す．従来法では，認識に平均 43.4 フレーム，最大 83 フレームを要していた．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とになる．これは，2.1 節で述べた従来の連続 DP を用いた認識手法の性質による遅れである．これに対して本手法を用いた場合，平均 7.8 フレームで認識結果を得ることができた．また，認識結果が得られるまでの遅れは最大で 26 フレームであった．このことから，本手法を用いることで認識に要する時間を短縮できたと言える．

一方，ジェスチャの曖昧性が早期認識の性能を低下させることも確認された．一例を示すと，図 5(b) の 522 フレーム目付近では,「急げ (左手)」というジェスチャが入力されているが,「胸に手を当てる (左手)」という冒頭部の動作が共通する他のジェスチャを認識結果として出力してしまっている．このように，認識対象となるジェスチャ間に曖昧性がある状況では，早期認識を行うことは本質的に不可能であるという 2.3 節での考察が実験的に検証された．

3.　動 作 予 測

3.1　動作予測の原理

前節の早期認識手法を用いることで，比較的簡単にジェスチャ動作者の数フレーム後の姿勢を予測することができる．早期認識の結果，現在の入力がどの標準パターンのどのフレームに該当するのか，という情報が得られる．このとき，現在より δ フレーム先では，認識結果の標準パターンでも δ フレームだけ進むと考えられる．この標準パターンの該当フレームの特徴

ベクトルを予測値とする，というのが基本的な考え方である．具体的には，現在の入力フレーム τ が標準パターン c^* のフレーム t^* に対応すると仮定した場合，δ フレーム後の動作者の姿勢 $\boldsymbol{I}_{\tau+\delta}$ は次式で予測される (図 6)．

$$\widehat{\boldsymbol{I}}_{\tau+\delta} = \boldsymbol{R}_{c^*,t^*+\delta} \tag{4}$$

この単純な予測方法は，入力ジェスチャと標準パターンのジェスチャが同じ速さの動作であるという仮定に基づいている．より一般的な場合，すなわち動作速度が異なる場合への拡張は今後の課題である．

1 章で述べたように，この予測手法はヒューマンインタフェースシステムの遅れ補償に利用することができる．つまり，仮にシステムが動作者の行動に対してなんらかの応答を返すのに δ フレーム必要だったとしても，システムへの入力として $\boldsymbol{I}_\tau$ を与える代わりに，予測された姿勢 $\widehat{\boldsymbol{I}}_{\tau+\delta}$ を与えることで，この遅れを補償できる．

3.2 動作予測の限界

以下では (4) 式による動作予測の限界，すなわちどれだけ δ を大きくできるか，について図 3 に示したジェスチャ間の関係を基に考察する．

図 3(a), (c) の場合は，ジェスチャの最後まで正しい予測結果を出力することができる．これらの場合はジェスチャ間に曖昧性が無く，どのフレームにおいても早期認識の結果に信頼をおくことができる．これはつまり，現時点以降のジェスチャ軌道を確定できるということを意味する．従って，もし識別規則 (3) によって (c^*, t^*) を得ることができれば，その後の動作は $\boldsymbol{R}_{c^*,t^*}, \boldsymbol{R}_{c^*,t^*+1}, \ldots, \boldsymbol{R}_{c^*,T_{c^*}}$ として予測することができる．

図 3(b) の場合は，ジェスチャ間に曖昧性が存在するため，前述の通りジェスチャ冒頭部において早期認識を行うことが本質的にできない．この場合のジェスチャ冒頭部における予測は，$t^* + \delta$ が共通部分を越えるまでは成功するが，それ以降では予測結果が正しいかどうかは不明となる．このため，あらかじめジェスチャ間の関係を調べておき，どこまでが共通部分であるかを知っておく必要がある．

また，まだ入力されていないジェスチャを前もって予測することは 2 ジェスチャの生起相関性を想定していない本論文の立場では本質的に不可能である．つまり，あるジェスチャが終了する直前に，次に来るジェスチャを予測することはできない．

3.3 動作予測実験

式 (4) で示される単純な動作予測が正しく機能するかを確かめるために，簡単な実験を行った．図 7 は式 (4) に基づく予測結果を示している．図 7(a) は早期認識で正しい認識結果が得られたフレーム (具体的には，図 5(b) の $\tau = 20$) における予測値 $\widehat{\boldsymbol{I}}_{\tau+\delta}$ の 1 要素を表している．動作速度の違いにより，時間方向に関して伸縮があるものの，図に示した部分において予測値は入力とほぼ同じ軌道を通ることがわかる．

一方，図 7(b) は「急げ (左手)」という入力を「胸に手を当てる (左手)」というジェスチャに誤認識してしまったフレーム (具体的には図 5(b) の $\tau = 527$) における予測値を表している．この 2 つのジェスチャの冒頭部は「左手を肩の高さまで上げる」という共通した動作である．すなわち，このフレームにおける誤認識は 2.3 節で述べた曖昧性によるものである．このときの予測は，冒頭の共通部分においてはある程度成功しているが，それ以降では大きく外れている．ただし，こうした曖昧性による誤認識があっても予測はある程度可能である．この点については次章で詳しく述べる．

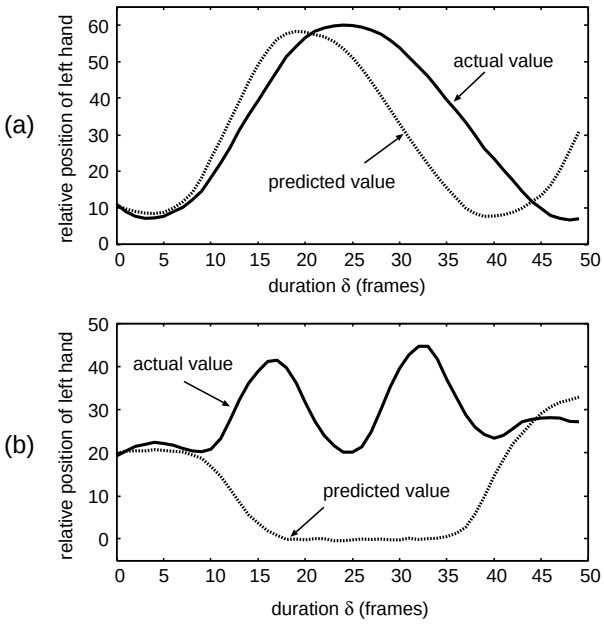

図 7　予測結果．横軸は予測幅 δ．(a) 正しく認識できたフレームにおける予測．(b) 共通動作．

4. ジェスチャネットワーク

あらかじめ認識対象となるジェスチャ間の関係を調べておき，どこまでが共通部分であるかを知っておけば，現時点において予測が可能な範囲を知ることができる．すなわち，ジェスチャ間の曖昧性による影響を抑えることが可能になる．そこで，次節で述べるように図 3 に示されるような軌道モデルを認識対象となるジェスチャ全てに対して構築することでジェスチャ間の関係を明らかにし，ここで得られる情報を早期認識および動作予測に適用することを検討する．

4.1 ジェスチャネットワークの定義

ジェスチャ間の関係を調べるために，複数のジェスチャを特徴空間内の軌跡として考えると，図 8 のようなネットワーク構造になる．図 9 及び図 10 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状の軌道が構築可能であることを示唆している．図 9 において，異なるジェスチャの非共通部分は異なった軌道を示しており，共通部分は類似した軌道を示している．また，図 10 をみると，同じカテゴリのジェスチャは多少のゆらぎがあるにせよ，およそ同じ軌道を描いている．そこで，類似した軌道をまとめてひとつのエッジとしたネットワークを構築する．これにより，現時点での入力がこのネットワーク上のどの位置に相当するかを同定することで，動作予測が可能な範囲を定めることができるようになる．

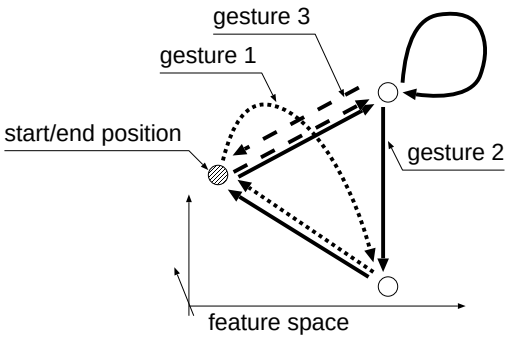

図 8　複数のジェスチャの軌道の模式図.

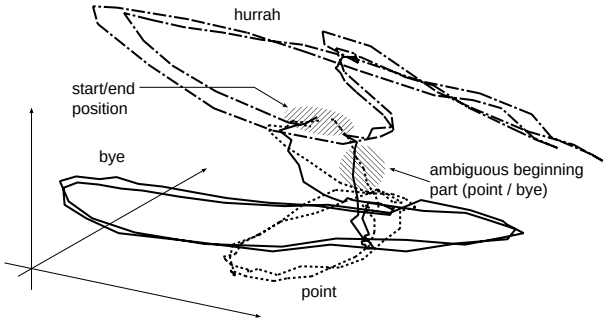

図 9　異なるカテゴリに属する 3 種類のジェスチャの軌道．全てのジェスチャは両手を下げた状態から開始し，終了する．図中央付近がその状態に相当する．

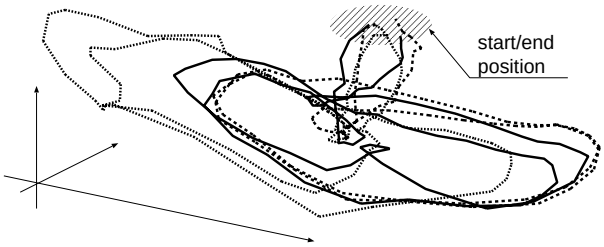

図 10　同じカテゴリ「ばいばい」に属する 3 つのジェスチャの軌道.

4.2　モーションプリミティブ

モーションプリミティブ (基本動作) はこれまでにも様々に定義されてきた．大崎ら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 [11]．澤田ら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 [12]．Sanger 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 [13]．Fod らは以上のアプローチを混合した方法を提案している [14]．これらに対し，向井らは自己組織化ネットワークをジェスチャ動画像に適用し，そのトポロジーを解析することによりモーションプリミティブを抽出する手法を提案している [15] [16].

本研究では，図 8 に示すジェスチャ軌道のエッジ間を，新たにモーションプリミティブとして定義することにする．すなわち，各ジェスチャを共通／非共通部分に分割することでモーションプリミティブを得ることができる．こうして定義されたモーションプリミティブは，次のような性質を持つ.

- 1 つのモーションプリミティブは少なくとも動作予測可能な範囲を表している．ここで「少なくとも」というのは，実際にはモーションプリミティブの範囲を越えて予測可能な場合があるためである．例えば図 11 は図 8 を以上の定義に従ってモーションプリミティブ表現したものである．ここで，現在のジェスチャがモーションプリミティブ E 上にあったとすると，その予測限界は E の末端ではなく B の末端となる．このように，ジェスチャ確定後の軌道であっても，他のジェスチャと共通の軌道があればそれらをまとめてモーションプリミティブとしているために，モーションプリミティブだけでは厳密な予測限界は得られないが，ジェスチャネットワークを参照することで，これを知ることができる.
- 1 つのジェスチャの物理的性質により事前確定されるものではなく，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.

このようにして得られたモーションプリミティブに対して早期認識および動作予測を適用することで，ジェスチャ間の曖昧性による影響を抑えることができる.

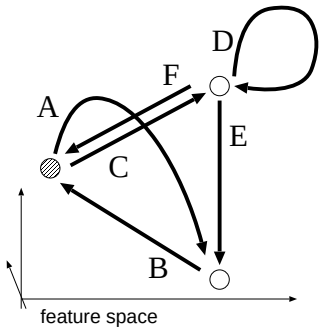

図 11　ジェスチャネットワークと，そのエッジとしてのモーションプリミティブ.

4.3　ジェスチャネットワークに基づいたモーションプリミティブの早期認識の手法

4.1 節で述べたように，現在の入力がジェスチャネットワーク上でどの位置に相当するかを同定することで，動作予測が可能な範囲を定めることができるようになる．この同定は，ネットワーク状に繋がったモーションプリミティブに対して早期認識を行うことに相当する．以下では，その手法について述べる.

基本的な考え方は，ジェスチャネットワーク上でのモーションプリミティブ間の関係を用いて早期認識を行うというものである．いま，モーションプリミティブ c に対して，その直前に出現するモーションプリミティブの集合を p_c で表すことにする．図**??**の例で言えば，$c = \mathrm{B}$ のとき $p_c = \{\mathrm{A}, \mathrm{E}\}$ である．認識の際には，c と $c' \in p_c$ を接続ながら累積距離を計算していけば良い (図 12)．これは，(1) 式における $t = 1$ での累積距離 $g_{c,1}$ を次式 (5) で置き換えることで実現できる.

$$g_{c,1}(\tau) = \begin{cases} g_{c',T_{c'}}(\tau - 1) + 3d_{c,1}(\tau) \\ g_{c',T_{c'}}(\tau - 2) + 2d_{c,1}(\tau - 1) + d_{c,1}(\tau) \\ g_{c',T_{c'}-1}(\tau - 1) + 3d_{c',T_{c'}}(\tau) + 3d_{c,1}(\tau) \end{cases} \quad (5)$$

ただし，c' の選択基準は後述する．また，$p_c = \phi$ のとき，すなわちそのモーションプリミティブが何かのジェスチャの始めに

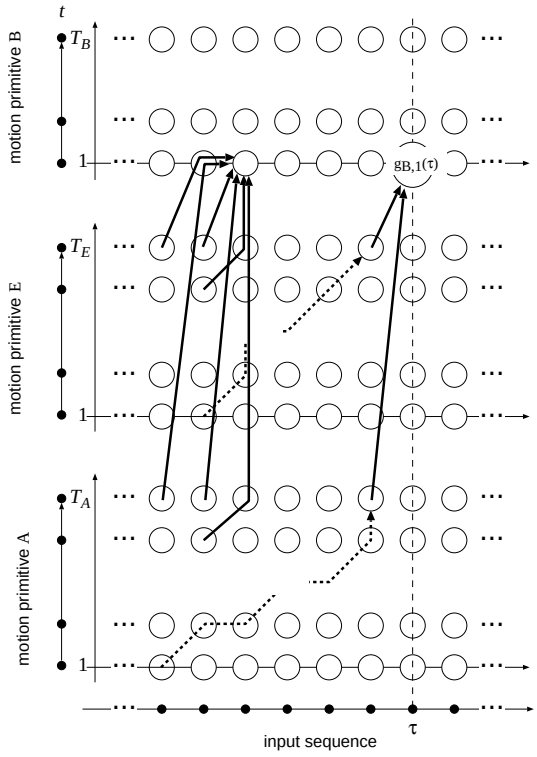

図 12　DP におけるモーションプリミティブの接続.

しか来ない場合は，$g_{c,1}(\tau) = 3d_{c,1}(\tau)$ とする．早期認識の識別規則には式 (3) を用いる．

集合 p_c に複数のモーションプリミティブ c' が含まれる場合，どのモーションプリミティブを c に接続するかを各フレーム毎に選ばなくてはならない．このときの選択の基準として，直前のフレームでの累積距離を用いる．具体的には，フレーム τ における c' を次式で定義する．

$$c' = \underset{c \in p_c}{\operatorname{argmin}}\, g_{c,T_c}(\tau - 1) \tag{6}$$

これにより，ジェスチャネットワークに基づくモーションプリミティブの早期認識を実現できる．

4.4　モーションプリミティブの早期認識実験

実際にジェスチャネットワークを用いたモーションプリミティブの早期認識実験を行った．本実験では，2.4 節で想定した 18 種のジェスチャに対して，目視で共通／非共通部分を判定することでジェスチャネットワークを構築し (図 13)，26 種のモーションプリミティブを得た．こうして得たモーションプリミティブを標準パターンとして，2.4 節の実験と同じ入力に対し 4.3 節の手法で早期認識を行った．このとき入力されたジェスチャは合計 153 個のモーションプリミティブによって構成されていた．

このときの実験結果の一部を図 14 に示す．正しい認識結果を出力するまでに要した時間は平均 4.1 フレームであった．モーションプリミティブの平均フレーム長が 29.0 フレームであることから，モーションプリミティブに対する早期認識は有効に機能しているといえる．

モーションプリミティブに対する早期認識が成功しているフレームでは，現在の入力がネットワーク上のどの位置にあるかを同定することができる．つまり，そのフレームにおける予測が可能な範囲を知ることができる．図 15 は．図 5(b) の $\tau = 527$，すなわち 3.3 節で述べた誤認識フレームにおける予測可能範囲を示している．図 15 より，予測可能な範囲においては予測が成功していることがわかる．これにより，ジェスチャ単位で誤認識を起こした場合であっても，その誤認識がジェスチャ間の曖昧性に基づくものであれば，予測可能な範囲を知ることによって，その範囲での予測結果を信頼できるようになることを確認できた．

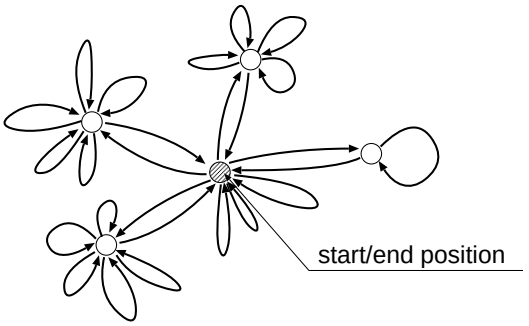

図 13　18 ジェスチャから構築されたネットワーク.

5. ま と め

本論文では，まずジェスチャの早期認識と動作予測の手法について述べた．ジェスチャの早期認識とは，ジェスチャの入力が開始されて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．本研究では，連続 DP を用いた従来の手法に若干の変更を加えた識別規則を用いることで，この早期認識を実現した．また，その原理的な限界を考察し，実験により早期認識の有効性を示した．

動作予測とは，早期認識の結果に基づいて動作者の数フレーム後の姿勢を予測するものである．これは，マンーマシンインタラクティブシステムの遅れ補償に利用することが出来る．本研究では簡単な規則によって，この動作予測を実現した．また，この方法を用いた場合の限界について述べ，実験で動作予測の効果と問題点を確認した．

さらに本研究ではジェスチャネットワークを構築し，それに対して上記手法を適用した実験も行っている．ジェスチャネットワークは認識対象とするジェスチャが互いにどのような共通部分を持つかという関係を示す．このジェスチャネットワークを用いることで，動作予測可能な範囲を知ることができる．また，本論文ではこのジェスチャネットワークからモーションプリミティブを定義することを提案した．このモーションプリミティブに対する早期認識と動作予測の実験を行い，本手法の有効性を示した．

今後の課題としては，以下のものが挙げられる．

- **ジェスチャネットワークの自動構築**

4.4 節の実験では手動でジェスチャネットワークを構築していたが，これを自動化することが今後の課題の一つとして挙げられる．4.2 節で述べたように，このジェスチャネットワークは対

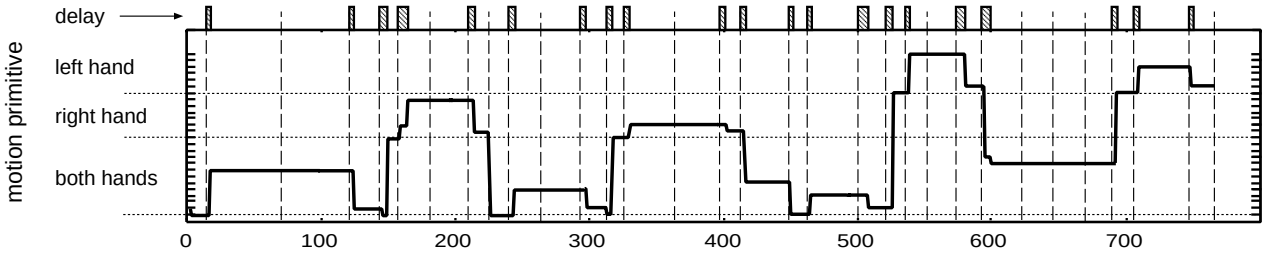

図 14　ジェスチャネットワークを用いた早期認識の実験結果の一部.

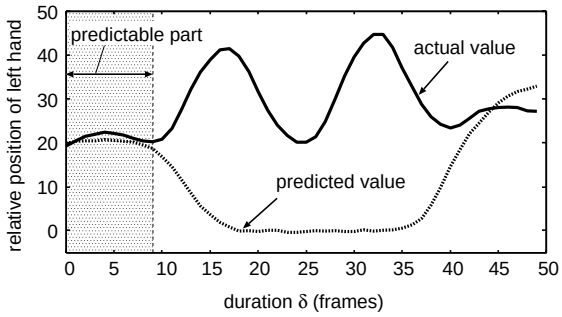

図 15　ジェスチャネットワークを用いて図 7(b) の場合について推定された動作予測可能な範囲.

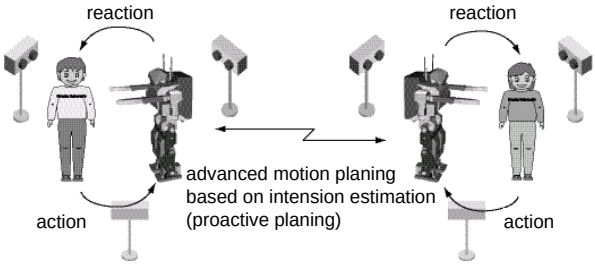

図 16　ヒューマノイドを用いたプロアクティブヒューマンインタフェースによる遠隔地コミュニケーション.

象とするジェスチャの集合に応じて動的に構成される．このため，ネットワークの構築の自動化は必要不可欠であるといえる．

- **予測限界を越えた予測**

原理的に予測限界を越えて予測を行うことはできない．しかし，例えば遅れ補償に本手法を適用する場合では，遅れの大きさに応じて予測幅をとらねばならず，予測限界を越えた予測が必要になることがある．この問題に対処するため，認識時のコスト比や動作の出現確率などに基づいて複数のモーションプリミティブを合成した軌道を予測結果とするなどといった手法について検討するよていである．

- **プロアクティブヒューマンインタフェースの実現**

本論文で検討した早期認識法及び動作予測法は，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の実現のための要素技術である．このインタフェースは，実態を伴っているために人間に対して物理的な働きかけが可能である．また，使用者の行動意図を推定・予測して先回りする機能を具備している．これらの性質はいずれも，さまざまな人にとって使いやすいコンピュータシステムの枠組みを提供することを目的としている．

図 16 に示すのは，このインタフェースを遠隔地コミュニケーションに応用した例である．このシステムにおいては，早期認識の結果が出力されてしまえば，動作者はジェスチャを完全に入力する必要が無いという特徴を持つ．さらに，通信混雑やヒューマノイドのハードウェア制約による遅れを補償することが出来るため，遅延の無い円滑なコミュニケーションを実現できると考えられる．

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 (SCOPE) の支援を受けた.

文　献

[1] 内田，森，倉爪，谷口，長谷川，迫江，“動作の早期認識およびその予測への応用に関する検討,” 信学技報, PRMU2004-94, Nov. 2004.
[2] O. Scharenborg, L. t. Bosch, L. Boves, “ ‘Eary recognition’ of words in continuous speech,” IEEE workshop of Automatic Speech Recognition and Understanding, pp. 61-66, 2003.
[3] 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, pp. 145-152, Jun. 1978.
[4] 高橋，関，小島，岡，“ジェスチャーの動画像のスポッティング認識,” 電子情報通信学会論文誌, vol. J77-DII, no. 8, pp. 1552-1561, Aug. 1994.
[5] 佐川，酒匂，大平，崎山，安部，“圧縮連続 DP 照合を用いた手話認識方式,” 電子情報通信学会論文誌, vol. J77-DII, no. 4, pp. 753-763, Apr. 1994.
[6] 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.
[7] 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,” 電子情報通信学会論文誌, vol. J80-DII, no. 6, pp. 1563-1570, 1997.
[8] 西村，古川，向井，岡，“時系列パターン検索の為の重み減衰型 Reference Interval-Free 連続 DP について,” 電子情報通信学会論文誌, vol. J81-DII, no. 3, pp. 472-482, Mar. 1998.
[9] 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,” 電子情報通信学会論文誌, vol. J81-DII, no. 1, pp. 18-26, Jan. 1998.
[10] 下野，佐藤，北澤，“特徴空間中の部分グラフ間距離の高速計算による実時間行動識別,” 信学技報，PRMU2004-197, Feb. 2005.
[11] 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.
[12] 澤田，橋本，松嶋，“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.
[13] T.D. Sanger, “Optimal movement primitives,” Advances in Neural Infomation Processing Systems, vol. 7, pp. 1023-1030, 1995.
[14] 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.
[15] 向井，西村，遠藤，岡，“ジェスチャ動画像の自己組織化ネットワークによるモデル化と要素動作の自動抽出,” 信学技報, PRMU97-128, Oct. 1997.
[16] 矢部，西村，向井，岡，“ジェスチャ動画像と意味記述単語系列とのネットワーク構造対応に基づくジェスチャ認識,” 信学技報, PRMU99-40, Jul. 1999.