プロアクティブヒューマンインターフェースの研究
－第5報 ジェスチャネットワークの利用による予測駆動の高精度化－

○森明慧　内田誠一　倉爪亮　谷口倫一郎　長谷川勉　迫江博昭

Study on Proactive Human Interface
-Application of Gesture Network for High Accurate Predictive Control-

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

Abstract— This paper examines a gesture prediction method that the subsequent posture of a person who makes a gesture is predicted. This method is useful to realize an embodied proactive human-machine interface, which can react to user's action before its end. The gesture prediction method is based on early recognition that the recognition result of a gesture is provided at the beginning part of the gesture. An experiment was performed to evaluate the accuracy of the prediction.

Key Words: Proactive System, Active Human Interface, Humanoid, Predictive Control, Gesture Recognition, Gesture Network

1. はじめに

本研究で実現しようとしているプロアクティブ（先回り）ヒューマンインターフェースとは，システムの使用者の行動意図の推定・予測に基づき，使用者が行動を終える前に次の行動に備えることが可能なインタフェースである．プロアクティブヒューマンインターフェースには次の利点がある．

- 今後の行動が予測できた時点で，その後の詳細な指示が不要になる (省力化)
- 通信混雑やハードウェア制約などによって生じる遅れを補償できる (遅延補償)

特に，このインターフェースの形態としてヒューマノイド (“PICO-2”[1]) を用いることを検討している．ヒューマノイドという実体を伴ったインターフェースとすることで，表現能力が向上し，上述の利点がさらに活かされるものと期待される．

本インターフェースの基幹技術である行動意図の推定・予測のために，ジェスチャの早期認識に基づく動作予測を検討してきた [2]．早期認識とはジェスチャ入力の初期段階において，そのジェスチャが何であるかを認識する手法である．例えば，両腕を挙げ始めた段階で，そのジェスチャが「万歳 (“hurrah”)」であると認識する手法である．この早期認識が実現すれば，動作予測も可能になる．先の例で言えば，両腕を挙げ始めた時点で “hurrah” だとわかれば，続いてその両腕を高く挙げると予測できる．

本論文は，この早期認識および動作予測をより高精度化するための検討に関するものである．具体的には，ジェスチャネットワークと呼ぶ，複数ジェスチャの表現モデルを提案し，それを用いて，現時点で早期認識が可能であるか，またどの程度先まで予測が可能であるかの判定を試みる．以下，2 章ではこのジェスチャネットワークについて述べる．続く 3 章では，このジェスチャネットワーク上での早期認識ならびに動作予測について述べる．

本手法の有効性を確認するために，認識対象として 18 種類のジェスチャを想定して実験を行う．具体的には，これらのジェスチャについて手動でジェスチャネットワークを構築し，連続入力に対する早期認識，及び動作予測の実験を行う．特に，動作予測による遅れ補償の効果を検証するために，実際に人工的に遅れを発生させた環境でヒューマノイドを駆動して，本手法を用いなかった場合との比較実験を行う．この結果については，4 章で述べる．

2. ジェスチャネットワーク

Fig.1 は，「万歳 (“hurrah”)」，「ばいばい (“bye”)」，「指さし (“point”)」という 3 種のジェスチャを特徴空間内の軌跡として表現したものである．“hurrah” は，両手を頭上まで挙げた後に再び元の位置まで降ろすジェスチャである．“bye” は，片手を一度肩の高さまで挙げ，左右に何度か振り，最後に肩の位置から元の位置まで降ろすジェスチャである．“point” は，片手を一度肩の高さまで挙げ，前後に何度か振り，最後に肩の位置から元の位置まで降ろすジェスチャである．すなわち，“bye” と “point” はジェスチャの開始部分と終了部

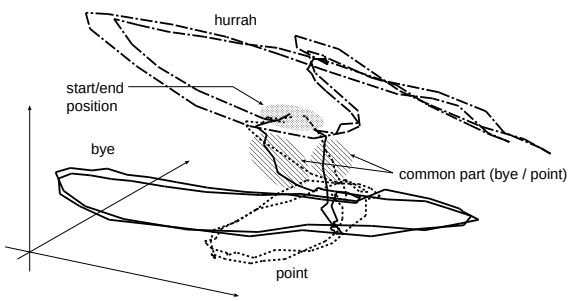

Fig.1 異なるカテゴリに属する 3 種類のジェスチャの軌道.

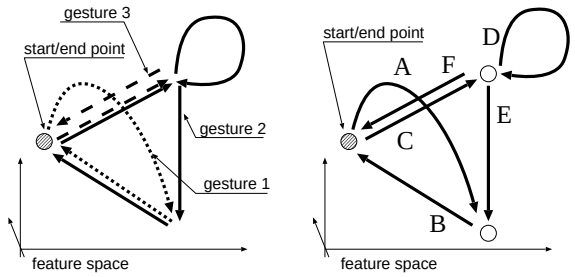

Fig.2 (a) 複数のジェスチャ軌道の模式図．(b) ジェスチャネットワークとそのエッジとしてのモーションプリミティブ．

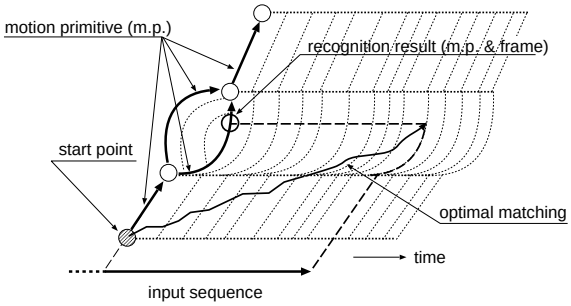

Fig.3 ジェスチャネットワークを用いた早期認識．

分に共通した動作をもっていることになる．**Fig.1** の斜線部は，この共通部分を示している．同図より，それぞれのジェスチャの共通部分は類似した軌道を示し，非共通部分は異なった軌道を示すということがわかる．このようなジェスチャの軌道間の関係を模式的に表すと，一般に **Fig.2**(a) のようなものになる．この図では，ジェスチャ2 と 3 の開始部分と，ジェスチャ1 と 2 の終了部分がそれぞれ共通していることになる．

本論文では，このような複数のジェスチャ軌道間の関係を表現する手段として，ジェスチャネットワークを提案する．これは，ジェスチャの共通部をまとめることで生成されるものである．先の **Fig.2**(a) の例からは，**Fig.2**(b) のようなジェスチャネットワークが構築される．このジェスチャネットワークの意図推定における有用性については，次章で述べる．

以下ではジェスチャの共通／非共通部分をそれぞれモーションプリミティブとして定義する．すなわち，ジェスチャネットワークの各エッジがモーションプリミティブとなる．こうして定義されたモーションプリミティブは，従来のような 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 [3]~[6] 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という特徴をもつ．後述のように，本手法ではこれらモーションプリミティブを単位とした認識処理を行う．

3. 動作予測と早期認識

3·1 動作予測の原理

本手法の基本的なアイディアは，「現在の入力がジェスチャネットワーク上のどの位置に相当するかがわかれば，これからの入力がどのような軌道となるか予測できる」という性質に基づく．**Fig.2**(b) を使って例を示すと，現在の入力がモーションプリミティブ E 上のどこかにあるとすれば，今後の入力はそのまま E の軌道に沿ってなされ，さらに右下にあるノードを経由してモーションプリミティブ B の軌道を通ると予測される．

また，どの程度先の入力まで予測できるか，ということについてもジェスチャネットワークを参照することで知ることができる．すなわち，現在の入力があるモーションプリミティブ上にあるときには，少なくともそのモーションプリミティブの末端までは予測が可能である．さらに，現在のモーションプリミティブから遷移可能なモーションプリミティブがただ一つであれば，予測可能な範囲はそのモーションプリミティブの末端にまで延びる．これを繰りかえすことで，厳密な予測限界を定めることができる．前出の例では，モーションプリミティブ E 上の点については B の末端が予測限界となる．

3·2 予測限界を越えた部分の処理

予測限界を越えて，認識に基づいた適切な予測を行うことは原理的にできない．しかし，実際には予測可能／不可能にかかわらず，一定時間先の予測値が要求される場合が多いと考えられる．従って，予測限界を越えた部分であっても，何らかの予測値を出力するこ

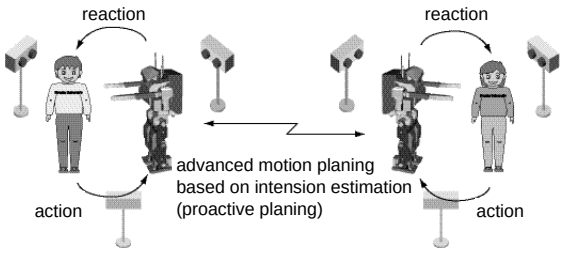

Fig.4 ヒューマノイドを用いたプロアクティブヒューマンインターフェースによる遠隔地コミュニケーション．

Fig.5 実験で想定したジェスチャの一部．

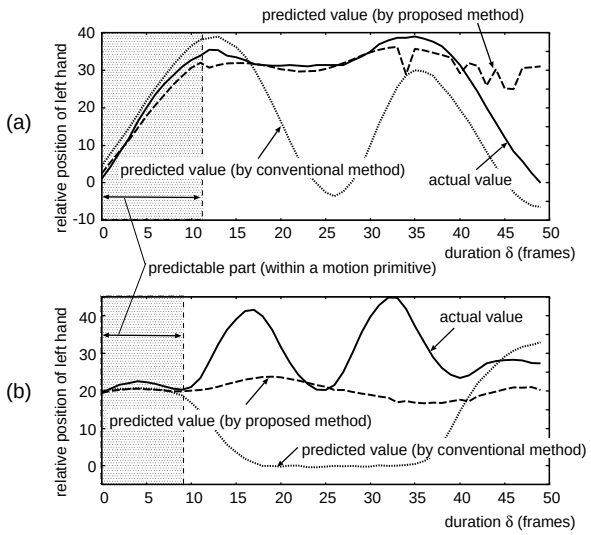

Fig.6 動作予測実験の結果.

とが必要になる.

この問題に対処するため，予測限界を越えた部分においては，現在の状態から遷移可能なモーションプリミティブすべての軌道を合成したものを予測に利用することにする．本来予測結果となるべきモーションプリミティブは，現在の状態から遷移可能なモーションプリミティブのうちのいずれかなので，これらすべての軌道を合成することで，予測値に正しい値を反映させることができると考えられる．具体的には，遷移先の各モーションプリミティブにおいて「一定時間先」に相当するフレームの特徴ベクトルを調べ，それらの平均値を予測値とする．この手法を用いることで，本来ならば予測が不可能な場合であっても，妥当な予測結果が得られるものと期待される.

3·3　早期認識

現在の入力がジェスチャネットワーク上のどの位置に相当するかを知るために早期認識を用いる．通常のジェスチャ認識では，ジェスチャの入力が完了した時点で認識結果を出力していた．早期認識は，前報 [2] で述べたように，ジェスチャの冒頭部における部分的なマッチングを評価対象とすることで，ジェスチャの完了を待たずに認識結果を確定する手法である．本論文では，標準パターンとしてモーションプリミティブを用いて早期認識を行う [7]．このとき，モーションプリミティブのどの部分とマッチングされたかによって，現在の入力がジェスチャネットワーク上のどの位置に相当するかを知ることができる．**Fig.3** にその概要を示す.

4.　実験

前章で述べた，ジェスチャネットワークに基づいた早期認識手法の基本的な性能を調べるために実験を行った．本実験では，プロアクティブヒューマンインターフェースの実現例として **Fig.4** のような遠隔地コミュニケーションを想定した．この場合，信号の伝送時やヒューマノイドのハードウェア制約などによる遅延が発生する．実験では，こうした遅延を早期認識・動作予測によって補償する．以下，4·1 節では，簡単な実験

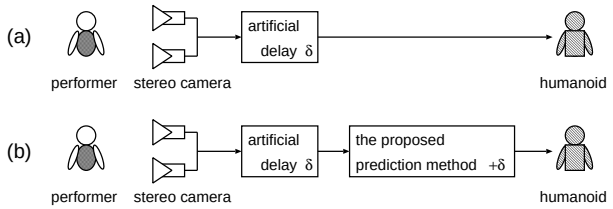

Fig.7 ヒューマノイドの予測駆動実験の概要．(a) 予測駆動による遅れ補償無し．(b) 予測駆動による遅れ補償あり.

によりジェスチャネットワークを用いた動作予測の基本的な性能を評価し，4·2 節では実際に 1 秒の遅延を人工的に発生させてヒューマノイドの予測駆動実験を行った結果について述べる.

4·1　ジェスチャネットワークに基づく動作予測実験

本実験では 18 種類のジェスチャを想定した．**Fig.5** 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．これらのうち 14 種類は，その冒頭に共通部を持っている.

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．これら 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，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，連続入力に対して早期認識・動作予測の実験を行った.

実験結果の一部を **Fig.6** に示す．**Fig.6**(a) は，「胸に手を当てる (“point myself”)」(**Fig.5**) というジェスチャの冒頭部における左手の高さの予測結果である．また，**Fig.6**(b) は，「急げ (“hurry up”)」(**Fig.5**) というジェスチャの冒頭部における左手の横方向座標の予測結果である．これらのジェスチャは冒頭部に他のジェスチャとの共通部分を持っている．この共通部分，すなわち予測可能な範囲は，ジェスチャネットワークを用いることで知ることができる (**Fig.6** の網掛け部)．実験結果より，予測可能な範囲においては予測が成功していることが確かめられた.

次に予測不可能な部分 (**Fig.6** の網掛け部以降) について考察する．ジェスチャ単位で早期認識・動作予測を行う従来の手法では，予測限界を越えた範囲では予測を大きく外している．これは，ジェスチャ冒頭部の曖昧性のために誤認識されたジェスチャの軌道を元に予測を行ったためである．これに比べ，本手法を用いた予測結果は，予測限界を越えた範囲では予測を外しているものの，従来法に比べると真値との誤差は小さくなっている．これは，3·2 節で述べたように，現在の状態から遷移可能なモーションプリミティブすべての軌道を合成したものを予測値として用いていることによる.

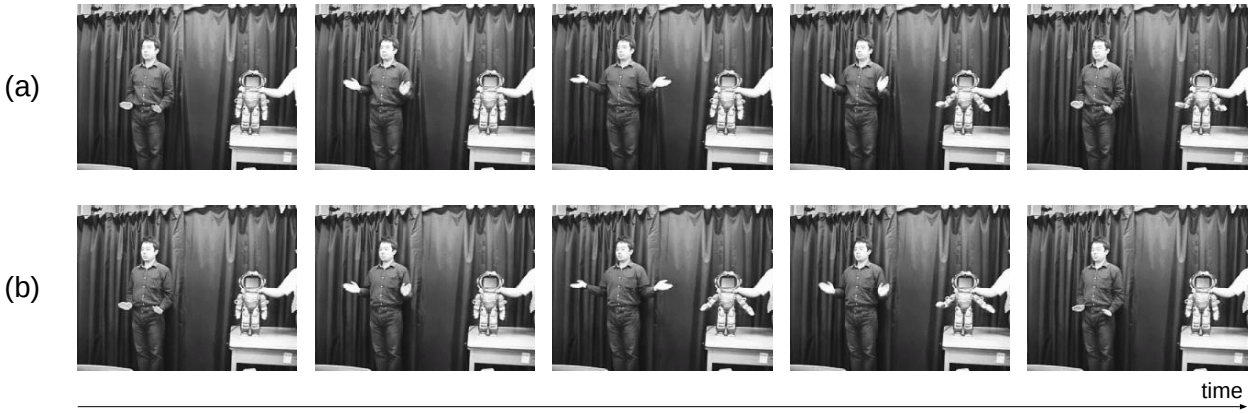

Fig.8 1 秒の遅れがある環境でのヒューマノイドの予測駆動実験の結果．(a) 予測駆動による遅れ補償無し．(b) 予測駆動による 1 秒の遅れ補償の結果.

以上の結果より，ジェスチャネットワークを用いることでより高精度に動作予測を行うことができることを確認できたと言える．

4·2　ヒューマノイドの予測駆動実験

続いて，遠隔地コミュニケーションを想定したヒューマノイドの予測駆動実験を行った．**Fig.7** は，この実験の概要を簡単に示したものである．実験の手順としては，まず人間の動作をステレオカメラを使って取得し，得られた入力データに対して人工的に 1 秒 (15 フレーム) の遅れを加える．一方では，この遅れを加えた入力データをそのままヒューマノイドへの入力として用いる．従って，ヒューマノイドは人間に比べ 15 フレーム遅れて動作する．他方では，遅れを加えたデータに対して動作予測を行って 1 秒 (15 フレーム) の遅れを補償し，この遅れ補償後のデータをヒューマノイドへの入力として用いる．この場合，予測が適切に行われれば，ヒューマノイドは人間と同期した動作を行うはずである．それぞれの入力でヒューマノイドを動かし，遅れ補償を行わなかった場合と遅れ補償を行った場合とで生ずる両者の遅れの差を比較した．

ジェスチャ「肩すくめ (“shrug”)」(**Fig.5**) を行った際の，実験中の動作者とヒューマノイドの動きは **Fig.8** のようになった．**Fig.8**(a) は，予測駆動による遅れ補償を行わなかった場合である．**Fig.8**(b) は，予測駆動による遅れ補償を行った場合である．動作のピークにあたる 3 枚目の画像に注目すると，**Fig.8**(a) の場合はまだヒューマノイドが動作を開始していないのに対し，**Fig.8**(b) の場合はヒューマノイドの姿勢が動作者の姿勢にほぼ追いついていることがわかる．また，動作が完了した 5 枚目の画像に注目すると，**Fig.8**(a) の場合はまだヒューマノイドが動作を完了しておらず，遅れをそのまま反映していることがわかる．一方，**Fig.8**(b) の場合は動作者とヒューマノイドが同時に動作を完了している．このように，遅れ補償を行わなかった場合は，動作がヒューマノイドに反映されるまでに常に 1 秒の遅れが存在するのに対し，遅れ補償を行った場合は各ジェスチャの後半部分で動作者と同期した動きをすることができた．これにより，ヒューマノイドの予測駆動の実現性を確かめることができた．

5.　まとめ

本論文では，ジェスチャネットワークを用いた高精度な早期認識及び動作予測法を提案した．ジェスチャネットワークを用いることで，現時点で早期認識が可能であるか，またどの程度先まで予測が可能であるかを判定できるようになる．さらに，予測限界を越えた部分においても，ジェスチャネットワークを用いることで妥当な値を予測結果として出力することができるようになる．

これらの手法を検証するために実験を行った結果，早期認識の有効性と動作予測による遅れ補償の効果を確認することができた．特に，ヒューマノイドの予測駆動実験では，本研究の目標であるプロアクティブヒューマンインターフェースをある程度実現できており，将来への展望を示すことができたといえる．今後の課題としては，実際に遠隔地コミュニケーションの実験を行い，さらなる検証を加えていくことが挙げられる．

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．

参考文献

[1] 大政，倉爪，内田，谷口，長谷川，“プロアクティブヒューマンインタフェースの研究－第 4 報 2 次元距離場を用いた人間動作の計測と再現－,” 第 23 回日本ロボット学会学術講演会講演予稿集, Sep. 2005.
[2] 倉爪，内田，長谷川，谷口，“プロアクティブヒューマンインターフェースの研究－第 3 報 予測駆動型アクティブインターフェース実験－,” 第 22 回日本ロボット学会学術講演会講演予稿集, Sep. 2004.
[3] 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.
[4] 澤田，橘本，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.
[5] T.D. Sanger, “Optimal movement primitives,” Advances in Neural Infomation Processing Systems, vol. 7, pp. 1023-1030, 1995.
[6] 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.
[7] 森，内田，倉爪，谷口，長谷川，迫江，“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,” 画像の認識理解シンポジウム (MIRU2005), IS3-106, Jul. 2005