

ネットワークモデルを用いたジェスチャの早期認識・予測に関する検討

九州大学 森明慧, 内田誠一, 倉爪亮, 谷口倫一郎, 長谷川勉, 迫江博昭

Early Recognition and Prediction of Gesture by Using Network Model

Akihiro MORI, Seiichi UCHIDA, Ryo KURAZUME, Rin-ichiro TANIGUCHI,

Tsutomu HASEGAWA and Hiroaki SAKOE, Kyushu Univ.

Abstract: This paper considers an early recognition and prediction method of gestures. Early recognition is the method to provide recognition results before input gestures are completed. Motion prediction is the method to predict the subsequent posture of the performer by using early recognition. In addition to them, this paper considers for improving the performance of these methods by using gesture network.

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. これらは, 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の要素技術である¹⁾. 特に, 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も有用である.

さらに本研究ではジェスチャ間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も検討する. このジェスチャネットワークを用いることで, 早期認識・予測をより高精度に行うことが可能になる.

本研究では実際にヒューマノイド PICO-2²⁾ を駆動して実験を行い, 上記手法について考察する.

2. 早期認識と動作予測

2.1 ジェスチャの早期認識手法

ジェスチャの早期認識は, 連続 DP を用いた手法^{3, 4)} に変更を加えることで実現できる. 本節では, まず従来の連続 DP を用いた認識手法について説明し, 続いて早期認識のための変更について説明する.

以下では, 標準ジェスチャパターン R_c を特徴ベクトルの時系列 $R_{c,1}, \dots, R_{c,t}, \dots, R_{c,T_c}$ 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 $R_{c,t}$ 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) I を特徴ベクトルの時系列 $I_1, I_2, \dots, I_\tau, \dots$ で表す. ここで τ は現在の入力フレームを表す.

連続 DP では各フレーム τ において以下の漸化式をすべての c, t について計算する.

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

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 $\|I_\tau - R_{c,t}\|$ である. フレーム τ での認識結果 c^* は次式で得られる.

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

式 (1)(2) 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. ただし, この手法は原理的にジェスチャ全体が完全に入力されるまでは認識結果を出力できない.

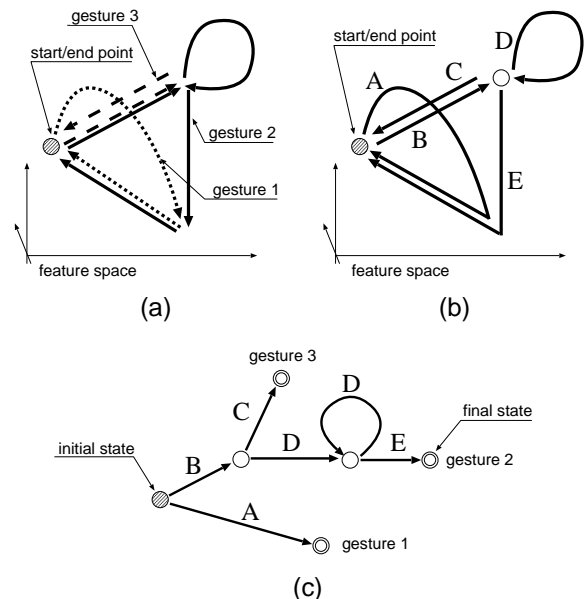


Fig.1 (a) 複数のジェスチャ軌道の模式図. (b) 冒頭の共通部分をまとめたもの. (c) ジェスチャネットワーク. 終了状態から開始状態の遷移 (ナル遷移) は省略.

本手法では, 早期認識のために従来の連続 DP による手法に変更を加える. 具体的には, フレーム τ において式 (2) の代わりに次の識別規則を用いる.

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

式 (2) との違いは, 冒頭部の部分パターンも評価の対象となる点である. 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.

2.2 動作予測の原理

前節の早期認識手法を用いることで, 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, 現在の入力フレームが標準パターン c^* のフレーム t^* に該当する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. すなわち, δ フレーム後の動作者の姿勢の予測値 $\hat{I}_{\tau+\delta}$ を次式で与える.

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

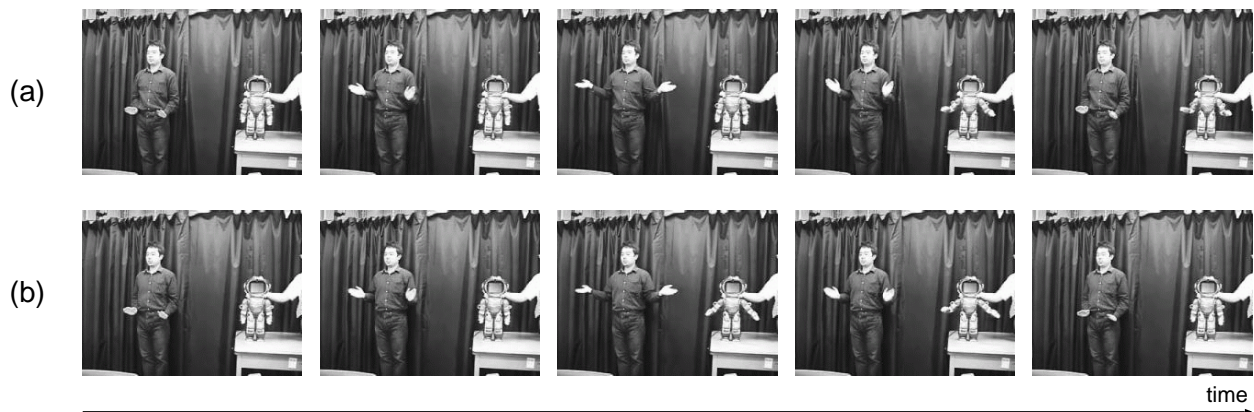


Fig.2 ヒューマノイドの予測駆動実験の結果。(a) 予測による遅れ補償無し。(b) 予測による1秒の遅れ補償の結果。

1. で述べたように、この予測手法は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。

3. ジェスチャネットワーク

互いに共通した動作で始まるジェスチャが認識対象に含まれている場合、現在の入力がこの共通部分にある時点では、早期認識によって信頼性のある認識結果を得ることは原理的に不可能である。そこで、本研究ではジェスチャネットワークを用いることで、現在位置で早期認識が可能であるか否かを判定する。以下では、その具体的な内容について説明する。

ジェスチャネットワークは冒頭の共通部分をまとめることで構築される。例えば、Fig.1 (a) は部分的に共通した動作をもつ複数のジェスチャの軌道を模式的に示したものである。このうち、冒頭における共通部分のみをまとめたものが Fig.1 (b) である。これからネットワークを構築すると Fig.1 (c) となる。

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同定することで、早期認識が可能かどうか判断できる。すなわち、現在の位置が冒頭の共通部分に相当するならば、この時点での早期認識は不可能だとわかる。また、ジェスチャネットワークを用いることで、現在位置からどれだけ先までの予測が可能であるかを知ることができる。これは、ネットワークの各辺が動作予測可能な範囲を示しているからである。本論文では、このネットワークの各辺を動作プリミティブとして定義する。

現在の入力がネットワーク上のどの位置に相当するかの同定は、動作プリミティブを単位とした早期認識を行うことで実現できる。具体的には、プリミティブ c の直前に現れるプリミティブの集合を P_c として、次式で $g_{c,1}(\tau)$ を計算する。

$$g_{c,1}(\tau) = 3d_{c,1}(\tau) + \min_{c' \in P_c} g_{c',T_{c'}}(\tau - 1) \quad (5)$$

$t \neq 1$ については (1) 式をそのまま用いることで、ネットワークの連結性を考慮しながら累積距離 $g_{c,t}(\tau)$ を計算できる。こうして得られた累積距離に対して式 (3) の識別規則を用いることで、ネットワーク上での同定を行うことができる。

4. 実験

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。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す、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”)」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.2 のようになった。Fig.2(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.2(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.2(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.2(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.2(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.2(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。

5. まとめ

本論文では、ジェスチャネットワークを用いた早期認識及び動作予測法について検討した。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。ヒューマノイドの駆動実験を行った結果、本手法の有効性を示すことができた。

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) 森, 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 -,” 第23回日本ロボット学会学術講演会講演予稿集, Sep. 2005.
- 2) 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.
- 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トティング認識,” 電子情報通信学会論文誌, vol. J77-DII, no. 8, pp. 1552-1561, Aug. 1994.