

固有ジェスチャを用いた動作予測モデルの高精度化

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

Motion Prediction based on Early Recognition and Eigen-gestures

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

Abstract: This paper describes a motion prediction method based on early recognition of gestures. Motion prediction is to predict the subsequent posture of the person making a gesture by using the early recognition result. This paper provides and evaluate a prediction model with eigen-gestures which express frequent deformations of each category as a solution for high-accuracy prediction.

1. はじめに

本論文では、早期認識に基づく動作（ジェスチャ）予測の一手法について述べる。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「万歳」の冒頭部であると認識するような処理である。一方、動作予測とはその認識結果に基づきユーザ（動作主）の数フレーム後の状態を推定するものである。

動作予測は、より自然なシステムとのインタラクションを実現する上で必要とされている技術である。動作者がジェスチャを終える前にシステムが次にとる行動を前もって予測（先回り）することでシステムが持つ遅れを補償することが可能となり、自然な対話が期待出来る。

本論文では、文献¹⁾の従来手法をより高精度なものとするために、新たに固有ジェスチャを用いた動作予測モデルの導入を検討する。

2. 従来の早期認識ならびに動作予測¹⁾

前述のように早期認識とはジェスチャの終了を待たずに認識結果を出力する方式である。現フレームを τ とすれば、早期認識は、入力パターン I の区間 $[0, \tau]$ がどの標準パターン

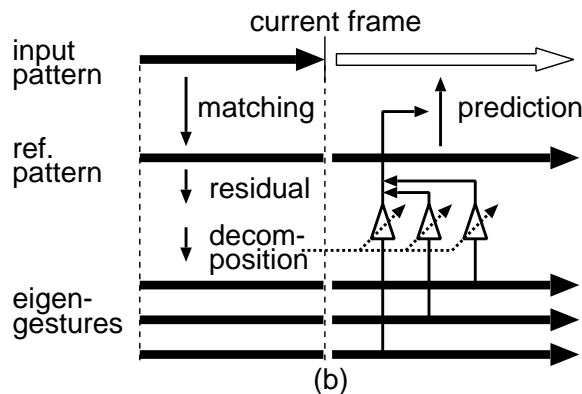
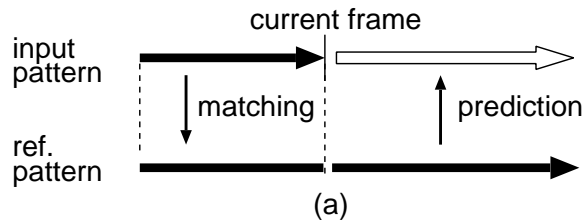


Fig.1 動作予測の原理. (a) 従来法¹⁾. (b) 本手法.

R_c (c はジェスチャのクラス) の、どの区間 $[0, t]$ に対応するかを見出す問題に帰着する (Fig.1(a)). すなわち I と R_c の部分マッチングを行ないながら最適な c と t を探索することになる。文献¹⁾では、連続 DP マッチング²⁾を拡張した手法でこの問題を効率的に解いている。

今、早期認識により、現フレーム τ がジェスチャ c^* の第 t^* フレームに最適対応することがわかったとすると、 δ 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、

$$\hat{I}_{\tau + \delta} = R_{c^*, t^* + \delta} \quad (1)$$

により、動作予測が実現する。

3. 固有ジェスチャの線形結合モデルの導入

以上の従来法では、予測値として標準パターンそのものが出力される。従って、予測精度は標準パターンに大きく依存する。動作主が変わるなどして、実際入力されたパターンと用意した標準パターンとの間に乖離が生じれば (例えば手の振り幅の違いなど)、予測精度は落ちることになる。一方、大量の標準パターンを準備するという改善策も考えられるが、事前にそれら多数の標準パターンを準備するのは容易ではないことに加え、処理能率が著しく落ちることが予想され、動作予測という実時間処理を前提としたタスクにおいては現実的ではない。

そこで、本論文では標準パターンそのものに加え、それからの変動分を効率的に表現する固有ジェスチャの線形結合モデルの導入を検討する。本手法の概要を Fig.1(b) に示す。ジェスチャ c の固有ジェスチャ $u_{c,k}$ とは、学習パターンと標準パターンの差分 (すなわち変動ベクトル) の第 k 主成分であり、主成分分析法により得られた累積寄与率が高い上位幾つかが保存される。これら固有ジェスチャは標準パターン R_c からの固有変形であるため、従って、ジェスチャ c において起こり易い変動を表現している。

この固有ジェスチャを用いた場合の動作予測式は

$$\hat{I}_{\tau + \delta} = R_{c^*, t^* + \delta} + \sum_k \alpha_k u_{c^*, k, t^* + \delta} \quad (2)$$

となる。ここで α_k は加重係数であり、差分 $I - R_c$ とモデル $\sum_k \alpha_k u_{c,k}$ との間の二乗誤差が最小になるように、各フレーム τ において、区間 $[0, \tau]$ の部分パターンを用いて決定される。

以上では簡単のため、時間軸方向の非線形伸縮 (すなわち、フレーム数の違い) への対処の詳細を省略した。実際には DP により伸縮を補償しながら、現フレーム τ に対応する t や、 α_k を決定している。

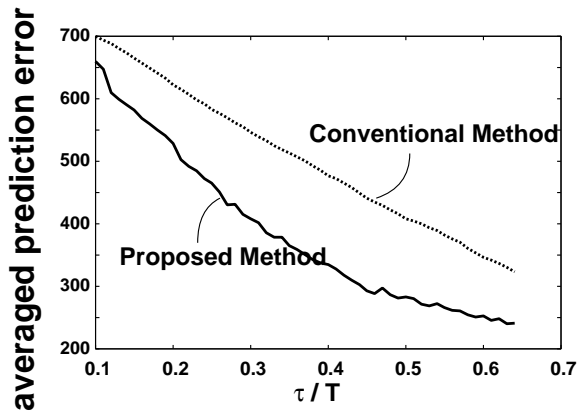


Fig.2 本手法による動作予測の精度.

4. 動作予測実験

実際に3クラスのジェスチャを用いて本手法による動作予測実験を行なった. 用意したジェスチャは「万歳」「パイパイ」「指をさす」のそれぞれの顔, 両手の3次元座標位置の時系列情報で表現される. 各クラスについて約100サンプル程度を準備し, 固有ジェスチャ $u_{c,k}$ を求めた. その際の標準パターン R_c は各クラス c における平均パターン (分布の中心) とした.

予測実験の結果を Fig. 2 に示す. 横軸は予測に用いる区間 $[0, \tau]$ の入力パターン全体に占める割合であり, 従って値が小さいほどより早期の段階での認識結果を用いて, 長期の予測を行なったことになる. 縦軸は予測精度評価量であり, 予測誤差の総和 $\sum_{\delta=0}^{T-\tau} \|\hat{I}_{\tau+\delta} - I_{\tau+\delta}\|$ の平均値である (T は入力パターン長). また, 合成に必要とした固有ジェスチャの数は累積寄与率 0.85 において各クラス平均 8 個程度を用いた.

この結果から従来手法に比べ高い精度において予測が行われていることが分かる. 特にジェスチャの分散が最も高い中間部において, 高精度な予測モデルを作成出来ていることを表しており, これは固有ジェスチャの導入によりクラス内における変形の起こりやすさが考慮されたことによるものであると推測される. 改善した一例として Fig.3 に「指をさす (右手)」というジェスチャを行った際の右手の三次元座標値の遷移軌道の様子と予測したモデルとの比較図を示す. (a) が従来手法によるものに対して, (b) が本論文で提案した手法によるものである. 従来手法においては標準パターンそのものを予測モデルとして用いた為, 大きく標準パターンから逸脱した入力が行われた際には, 適した対応がなされなかった. それに対し, 本手法ではジェスチャ初期における分散状況を考慮し予測区間を合成している為, より実際の入力軌道に類似した軌道を予測している. なお, 予測に用いた区間 (τ/T) は 0.3 である.

5. まとめ

本論文では, 文献¹⁾ による動作予測の手法をより高精度なものとするために, 標準パターンからの変動を表現する固有ジェスチャの線形結合モデルの導入を提案した. それにより従来手法における問題点であった標準パターンへの依存を脱却し, クラス内における変形傾向を考慮してやることで柔軟な予測を可能とした. また実験により本研究における予測精度を評価し, 従来手法と比べ高精度な予測が可能であることを確認した.

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の支援を受けた.

参考文献

- 1) 森, 内田, 倉爪, 谷口, 長谷川, 迫江, “ジェスチャの早期認

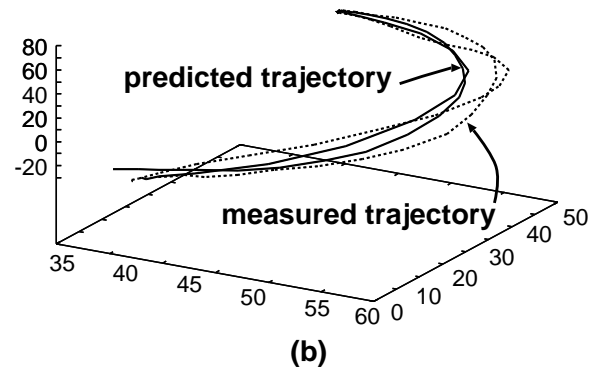
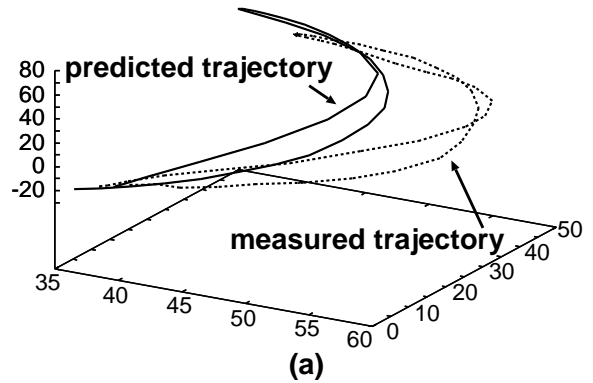


Fig.3 予測ジェスチャ軌道の比較:(a) 従来手法 (b) 本手法.

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, 2005.

- 2) 岡, “連続 DP を用いた連続単語認識,” 音響研資, S78-20, 1978.