認識に基づいた動作予測に関する一検討 ー固有ジェスチャの線形結合モデルによる予測ー

中島, 正登
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

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

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

他

出版情報：電気関係学会九州支部連合大会, 2005-09. 電気関係学会九州支部連合会
バージョン：
権利関係：

認識に基いた動作予測に関する一検討
– 固有ジェスチャの線形結合モデルによる予測 –

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 *
(*九州大学)

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．本論文では，文献 [1] の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の線形結合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]

文献 [1] の早期認識ならびに動作予測の原理を図 1(a) に示す．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 $\boldsymbol{I}$ の区間 $[0,\tau]$ がどの標準パターン $\boldsymbol{R}_c$ (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．すなわち $\boldsymbol{I}$ と $\boldsymbol{R}_c$ 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 $\boldsymbol{R}_{c^*,t^*+\delta}$ に似た動作が現れると予測できる．すなわち，$\boldsymbol{I}_{\tau+\delta}$ の予測値を $\widehat{\boldsymbol{I}}_{\tau+\delta}$ とすると，

$$\widehat{\boldsymbol{I}}_{\tau+\delta} = \boldsymbol{R}_{c^*,t^*+\delta} \tag{1}$$

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，入力パターンと標準パターンに乖離が生じれば (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．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，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．本手法の概要を図 1(b) に示す．ジェスチャ c の固有ジェスチャ $\boldsymbol{u}_{c,k}$ 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，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

$$\widehat{\boldsymbol{I}}_{\tau+\delta} = \boldsymbol{R}_{c^*,t^*+\delta} + \sum_k \alpha_k \boldsymbol{u}_{c^*,k,t^*+\delta} \tag{2}$$

となる．ここで α_k は加重係数であり，差分 $\boldsymbol{I}-\boldsymbol{R}_c$ とモデル $\sum_k \alpha_k \boldsymbol{u}_{c,k}$ 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．

4 動作予測実験

3 クラスのジェスチャを用いて動作予測実験を行なった．各クラスについて 100 サンプルを準備し，固有ジェスチャ $\boldsymbol{u}_{c,k}$ を求めた．その際の標準パターン $\boldsymbol{R}_c$ は全 100 サンプルの平均パターンとした．

予測実験の結果を図 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 $\sum_{\delta=0}^{T-\tau} \|\widehat{\boldsymbol{I}}_{\tau+\delta} - \boldsymbol{I}_{\tau+\delta}\|$ の平均値である (T は入力パターン長)．

この結果から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が分かる．また従来法と比べても平均して高精度な予測が可能である．

5 まとめ

本論文では，文献 [1] 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検討し，その精度を評価した．

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.

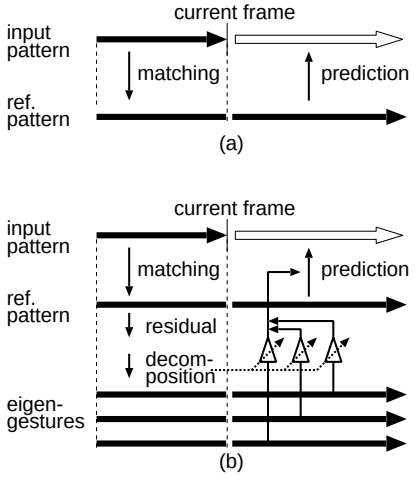

図 1: 動作予測の原理．(a) 従来法 [1]．(b) 本手法.

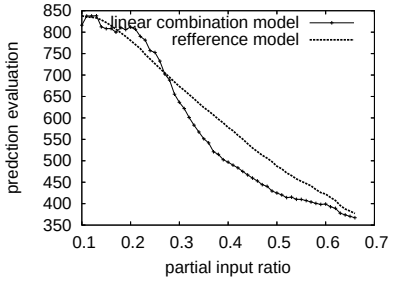

図 2: 本手法による動作予測の精度.