

## 主成分分析を用いた基本動作 HMM によるジェスチャー認識\*

3R-5

柳生 雄午 酒向 慎司 宮島 千代美 徳田 恵一 北村 正†

名古屋工業大学 知能情報システム学科‡

## 1 まえがき

人間同士の対話では、音声言語に加えてジェスチャーを用いることにより、相互理解を深めることができる。しかし、ジェスチャーにはさまざまな種類があり、すべてのジェスチャーを網羅して個別のモデルを作成するには限界がある。そこで、ジェスチャーを基本動作単位に分割する方法が考えられる [1]。また、画像ベースの認識法では次元数が增大するため、効率的な特徴抽出法が求められる。我々はこれまでに主成分分析 (PCA) を特徴抽出に用いたジェスチャー認識手法を提案し、その有効性を示した [2]。本研究ではこれを用いて特徴抽出を行い、ジェスチャーの基本動作を隠れマルコフモデル (HMM) に基づいてモデル化し認識を行う。

## 2 ジェスチャーの基本動作モデル

異なる意味を表すジェスチャーでも局所的に見た場合、互いに同じ基本的な動きを含んでいることがある。こうした基本的な動作を 1 つの単位として扱うことによって、少ないモデル数で多くのジェスチャーを表現することができる。また、既存の基本動作モデルを用いて十分な学習データが得られないジェスチャーをモデル化して認識することも可能となる。図 1 はあるジェスチャーを基本動作 HMM で連結してモデル化した様子を示している。

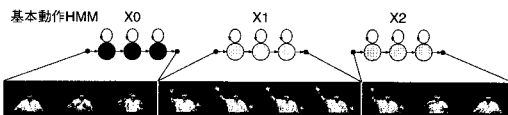


図 1: 基本動作 HMM の連結によるモデル化

一連の動作の中には動きが停止する点 (停留点) がある。本研究では、視察により得られた停留点をもとに周辺の動作を 1 つの基本動作とする。このようにジェスチャーを分割して得られた 67 種類の基本動作を HMM によりモデル化する。

## 3 主成分分析 (PCA) による次元圧縮

1 枚の画像の全画素値を並べた  $M$  次元ベクトルを  $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$  とすると、 $N$  枚の画像  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  の平均画像  $\bar{\mathbf{x}}$  は次のように求められる。

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (1)$$

また、画像  $\mathbf{x}$  から平均ベクトルを引いたものを  $\hat{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$  と表す。このとき、平均除去した  $N$  枚の画像から得られる行列  $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_N]$  に対して PCA を行うことによって正規直交基底  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N]$  が得られる。これは固有ベクトルと呼ばれる。ある画像  $\mathbf{x}$  から求めた  $\hat{\mathbf{x}}$  は  $\mathbf{U}$  の各固有ベクトルの線形結合によって次のように表すことができる。

$$\hat{\mathbf{x}} = \mathbf{U}\mathbf{y} \quad (2)$$

ここで、主成分スコア  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$  は元の画像  $\mathbf{x}$  の持つ特徴量と考えることができる。また、 $\mathbf{y}$  の次元を小さく取ることによって特徴空間の次元を圧縮することが可能となる。

本研究では、学習用データの中からランダムに選んだ 1000 枚の画像に対して PCA を行い、図 2 のような固有ベクトルを得る。次に、固有値の大きいものから 20 次元の固有ベクトルを使用する。この 20 次元に対して式 (2) を適用し、平均ベクトルを加えたものが図中の再構成画像となる。実際に HMM でモデル化する特徴ベクトルのは主成分スコアおよびその  $\Delta, \Delta^2$  であり、少ない次元数 (本実験では 60 次元) でジェスチャーをモデル化することができる。

## 4 ジェスチャー認識実験

## 4.1 実験条件

本研究では、RWC マルチモーダルデータベースを用いる。このデータベースには男性 23 名、女性 25 名の 25 種類のジェスチャーが 4 回ずつ収録されている。画像サイズは  $320 \times 240$  画素、フレームレートは 1/30 秒である。今回はこのうち  $h \sim y$  の 18 種類のジェスチャーを 1 時期使用し、学習データとして 23 名分、テストデータとして学習と異なる 23 名分を用いて認識を行った。テスト総数は 414 となる。

## 4.2 認識実験 1 (基本動作モデルによる認識)

本実験では、基本動作モデルによるジェスチャー認識を行い、ジェスチャー単位のモデルによる従来法と比較する。HMM の状態数は従来法では 9, 18, 27, 36,

\* Gesture Recognition Based on Sub-Gesture HMMs Using Principal Component Analysis

† Yugo YAGYU, Shinji SAKO, Chiyomi MIYAJIMA, Keiichi TOKUDA, Tadashi KITAMURA

‡ Department of Computer Science, Nagoya Institute of Technology

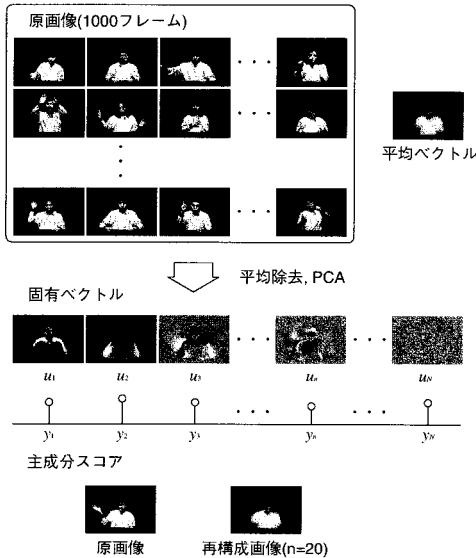


図 2: 主成分分析と固有ベクトル

基本動作単位では 2~6, 9 とし, 混合数は 1 とした. 特徴抽出にはサブサンプリングを使用している.

認識結果を図 3 に破線で示す. 図中の ( ) は各 HMM の状態数を表す. ジェスチャー単位の HMM を用いた場合, 最高で 79.0%, 基本動作モデルを用いた場合でも, これに近い最高で 78.0% の認識率を得た. また, システムの総状態数が少ない場合は基本動作 HMM がジェスチャー単位 HMM の認識率を上回ることがわかった.

4.3 認識実験 2 (PCA を用いた認識)

次に, 実験 1 と同様の実験を PCA を用いた特徴抽出法で行い, 基本動作単位での認識法とジェスチャー単位の認識法を比較する.

認識結果を図 3 に実線で示す. 図中の ( ) は各 HMM の状態数を表す. 特徴抽出法に PCA を用いることにより, ジェスチャー単位, 基本動作単位どちらの場合も認識率がサブサンプリングの結果と比べて 10% 前後向上することが確認できた.

4.4 認識実験 3 (学習外ジェスチャーの認識)

次に, 十分な学習データが得られないジェスチャー (学習外ジェスチャー) を正しく認識できるかについて検討する. 18 種類のジェスチャーのうち 17 種類のジェスチャーは 23 名で学習し, 残りの 1 種類のジェスチャー (学習外ジェスチャー) については 2 名で学習する. これを各ジェスチャーについて繰り返す leave-one-out 法に基づいて実験を行う. HMM の状態数は総状態数を同程度とするため, ジェスチャー単位では 18,

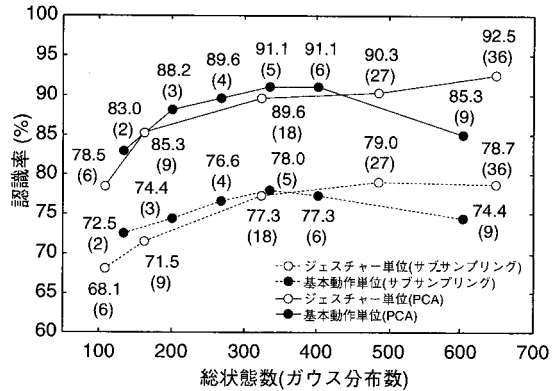


図 3: システムの総状態数と認識率の関係

基本動作単位では 5 のものを用いた. 認識結果を表 1 に示す. モデルの作成は文献 [1] に基づいて行った.

表 1: ジェスチャー単位と基本動作単位での平均認識率

HMM	学習外	学習内	全体
ジェスチャー単位モデル	0%	90.7%	85.6%
基本動作単位モデル	45.7%	91.1%	88.6%

実験の結果, ジェスチャー単位での学習外ジェスチャーの認識は 0% であったのに対し, 基本動作 HMM による認識率は 45.7% となり, 半分程度の認識が可能であることがわかった. 全体の認識率も平均すると基本動作 HMM では 88.6% となり, ジェスチャー単位 HMM での 85.6% を上回る結果となった.

5 むすび

本研究では, 基本動作 HMM に基づくジェスチャー認識手法を提案し, その有効性の検討をした. 実験では, ジェスチャー単位のモデルによる従来法に比べて少ない状態数で同程度の認識率を得ることができた. また, 特徴抽出法としてサブサンプリングと PCA を比較し, PCA を用いることによってより高い認識率が得られることが確認できた. 更に, 学習外ジェスチャーの認識についてもある程度適応できることがわかった.

参考文献

[1] 桑原昌也, 徳田恵一, 北村正, “基本動作の HMM に基づくジェスチャー認識”, 信学総大, D-12-115, Mar. 2001.  
 [2] 中谷博美, 酒向慎司, 徳田恵一, 北村正, “固有ジェスチャーを用いた HMM に基づくジェスチャー認識”, 信学総大, D-12-116, Mar. 2001.