

K-15 HMMを用いたジェスチャー認識における混合数の検討

The number of Gaussians for HMM-based gesture recognition

井上 涼子[†] 柳生 雄午[†] 酒向 慎司[†] 宮島 千代美[†] 徳田 恵一[†] 北村 正[†]
 Ryoko INOUE Yugo YAGYU Shinji SAKO Chiyomi MIYAJIMA Keiichi TOKUDA Tadashi KITAMURA

1. まえがき

近年、コンピュータとの様々な対話手段に関する研究が行われているが、ジェスチャーも非接触型の対話手段として注目されている [1]~[3]. 我々はこれまでに、主成分分析 (PCA) による特徴抽出を用いたジェスチャー認識手法について検討し、その有効性を示した [1]. 我々は、ジェスチャーのモデル化に隠れマルコフモデル (HMM) を用いているが、これまでは、各状態の確率分布を一つのガウス分布で表現していた。しかし、実際には、個人によって動作の大きさなどに差があるため、複数のガウス分布を混合して混合分布とすることによって、より柔軟なモデル化を行う必要があると考えられる。そこで、本研究では HMM の混合数を変化させ、その認識率との関係について検討する。

2. 主成分分析

1枚の画像の全画素値をならべた L 次元のベクトルを $\mathbf{x} = [x_1, x_2, \dots, x_L]^t$ とする。学習用の N 枚の画像集合に対して PCA を行うことにより平均 2 乗誤差の観点において最適に近似する正規直交基底 $S = [w_1, w_2, \dots, w_N]$ が得られる。また、ある画像 \mathbf{x} に対する主成分スコア \mathbf{y} の各成分 y_i は、画像 \mathbf{x} を表現するための各固有ベクトル w_i の貢献度を表していると解釈できる。再構成画像は、 S と主成分スコア \mathbf{y} を用いて、再構成画像 $\hat{\mathbf{x}} = S\mathbf{y}$ のように表すことができる。このとき S と \mathbf{y} の次元 n を N より小さくすることで、低次元で近似することが可能となる。

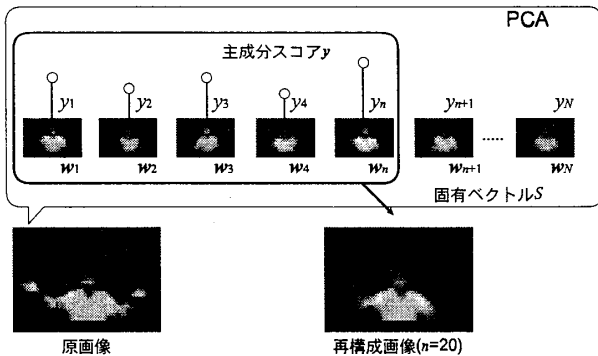


図 1: PCA と固有値展開

3. 認識実験

本実験では、表 1 に示す条件でジェスチャー認識実験を行った。

3.1 データベース

本研究では、RWC マルチモーダルデータベース [4] 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 (a~y) の動作が 4 回ずつ、発話とともに収録されている。画像サイズは 320×240 pixel, 30 枚/秒のフレームレートである。ここでは、このデータベース

表 1: 実験条件

データベース	RWC マルチモーダルデータベース	
使用データ	動作内容	ジェスチャー (h~y)
	時期	4 回分
	階調数	grayscale, 256 階調
	画像サイズ	横 280 × 縦 200 pixel
HMM	モデル	h~y の 18 モデル
	状態数	9, 18, 27, 36, 45
	混合数	1, 2, 4, 8
	使用ベクトル	主成分スコア, Δ, Δ^2
学習データ	24 名 × 18 動作 × 4 回分 = 1728 データ	
テストデータ	24 名 × 18 動作 × 4 回分 = 1728 データ	

のうち h~y の 18 ジェスチャーの 4 回分を用いた。24 名分を学習データ、残りの 24 名分をテストデータとした。図 2 に女性と男性の同一動作の一部を示すが、大きさがかなり異なり、大きさの正規化などの前処理を行わない場合には複数のガウス分布が必要なことが予想される。

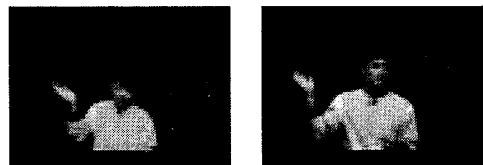


図 2: 動作の個人差

3.2 前処理

本実験では画像データに対し、以下の処理を行う。画像の切り出し 人間の体の位置の中心を求め、そこから左右にそれぞれ 140pixel ずつ切り出す。また、テーブルと体の境界から上に 180pixel, 下に 20pixel だけ切り出す。切り出し後の大きさは 280×200 pixel になる。

サブサンプリング 画像を $m \times m$ のブロックに分割し、各ブロック毎に画素の持つ輝度の平均値を取り、その平均値をブロックの代表する値とする。本実験では、 5×5 のサブサンプリングを行ってから PCA を行った場合と、原画像をそのまま PCA した場合について、それらと比較する。

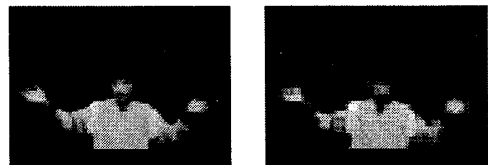


図 3: サブサンプリング (5 × 5)

[†]名古屋工業大学 知能情報システム学科,
 Dept. of Computer Science, Nagoya Institute of Technology

PCA 学習用の全画像から無作為に選んだ 1000 枚の画像に対して PCA を行うことにより、固有ベクトルが得られる。次に、 n 次元の固有ベクトルとその主成分スコアを使用し、全ての原画像を固有値展開する。その後、HMM によって主成分スコアおよび、その $\Delta, \Delta\Delta$ をモデル化し、認識を行う。

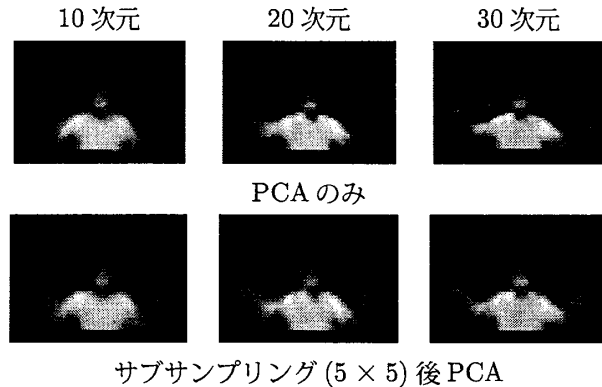


図 4: 再構成画像

3.3 ジェスチャー認識実験

本実験では HMM の状態数と混合数を変化させて、認識率を比較する。PCA 20 次元の場合の認識率を図 5 に示す。混合数を 2, 4 に増やすことによって、どの状態においても単混合に比べて認識率が向上した。しかし、混合数 8 では全般的に認識率は低下している。最高の認識率は、状態数 45, 混合数 4 の場合で 92.6% であった。次に、サブサンプリング後に PCA を行う方法についても同様の実験を行った。その結果を図 6 に示す。この場合も混合数 2, 4 において単混合に比べて向上が見られ、状態数 45, 混合数 2 の場合に最高で 92.8% の認識率となった。最後に、PCA の次元数を 30 次元として同様の実験を行った。その実験を図 7, 8 に示す。30 次元を用いた場合は、2, 4, 8 のどの混合数の場合においても単混合より高い認識率が得られた。また、20 次元の場合より全体的に認識率が高くなり、図 7 において状態数 36, 混合数 2 の場合、および状態数 45, 混合数 8 の場合に最高で 94.0% の認識率が得られた。

4. むすび

本研究では HMM に基づくジェスチャー認識における混合数の検討を行った。HMM の各状態を混合分布することにより、単混合の場合より認識率が向上し、最高で 94.0% の認識率が得られた。

謝辞

本研究の一部は、立松財団研究助成により行われた。

参考文献

- [1] 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, p.283, Mar. 2001.
- [2] Takio Kurita and Satoru Hayamizu, “Gesture Recognition Using HLAC Features of PARCOR Images and HMM Based Recognizer,” In Proc. of FGR98, pp.422-427, 1991.
- [3] 柳生雄午, 酒向 慎司, 宮島千代美, 徳田 恵一, 北村 正, “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,” 情報処理学会第 63 回全国大会講演論文集, 3R-5, pp.213-214, Sep. 2001.
- [4] マルチモーダルデータベースサブ委員会, “RWC マルチモーダルデータベース,” <http://www.rwcp.or.jp>

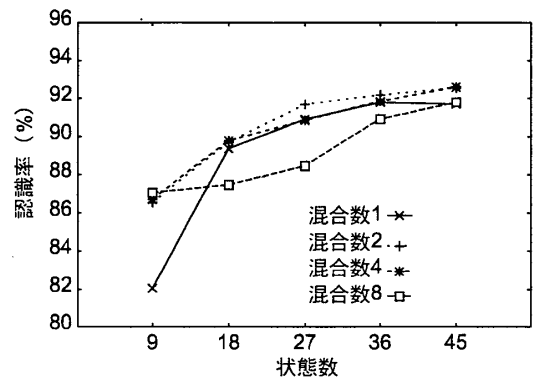


図 5: HMM の状態数と認識結果 (PCA20 次元)

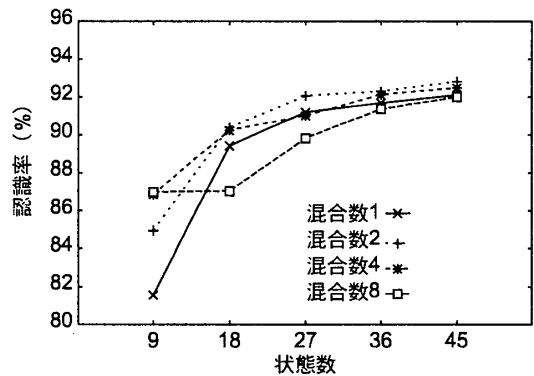


図 6: HMM の状態数と認識結果 (サブサンプリング+ PCA20 次元)

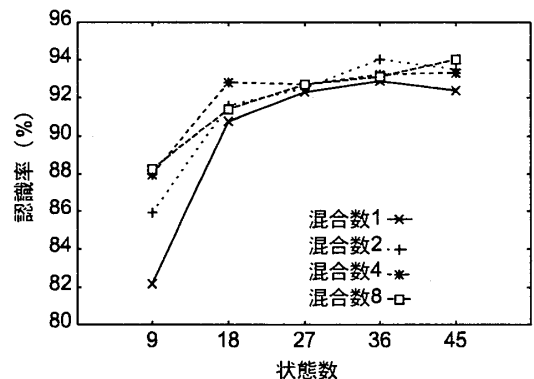


図 7: HMM の状態数と認識結果 (PCA30 次元)

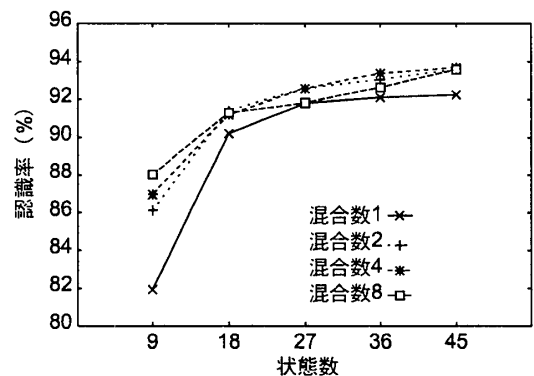


図 8: HMM の状態数と認識結果 (サブサンプリング+ PCA30 次元)