動き情報を用いたジェスチャ認識*

1AB−2

岩井 儀雄 畠 直志 谷内田 正彦†
大阪大学大学院基礎工学研究科‡
e-mail: iwai@sys.es.osaka-u.ac.jp

1 まえがき

日常社会において，人間は知らず知らずのうちにジェスチャをし，言語以上に意志を伝達することが良くある．そのジェスチャをコンピュータに認識，理解させることは，意志伝達をすみやかに行うという点でも重要な課題である．近年，コンピュータが人間の動作，ジェスチャ，表情などを理解する研究が盛んに行われ，インタラクティブな意志の疎通の実現を目指した研究が進んでいる．従来のジェスチャ認識の研究として，特徴点を確実に得るために体にセンサを取り付けたり，マーカーを付けた画像を用いたりすることが多い．しかしながら，接触型のセンサは体に取り付ける煩わしさや身体の動きを拘束するといった問題がある．ることによりジェスチャ認識を行った研究である．また，画像列を直接利用する方法は，背景が変化したり，照明が変化すると利用することができなくなってしまう．

本研究では，動き情報を用いるため背景などの環境条件に左右されない点，同様の理由で画像の濃度補正などの正規化を必要としない点，ジェスチャ認識にHMMを利用することで雑音や時間的な伸縮に対してロバストである点，学習に基づいて認識系を構築するので高い認識率が達成できるなどの点で有効な手法を提案する．

2 動領域の抽出とKL展開によるジェスチャ空間の構築

KL展開による認識は，対象の平行移動に弱く，正しく認識するためには，人物とカメラとの相対位置が固定されてしまう．本研究では，人物を検出することで，動領域をセグメンテーションし，人物の平行移動にロバストな方法を利用する．

2.1 人物の動領域の発見

ジェスチャを認識をするためには，まず最初に動画像中の情報から人物を発見し切り出す必要があるが，容易なことではない．そこで，まず画像中の人物の顔を発見することで，顔の位置を抽出し[1]，それを基準に人物の動領域の切り出しを行う．この抽出法によって，人物が画像中のどこにいても安定して動領域だけを抽出することができる．図1は，顔の位置を基準とした動き領域の抽出図である．図中，顔が小さな枠で囲まれて抽出されており，さらに動領域が大きな枠で囲まれた枠で抽出されている．

2.2 KL展開による情報圧縮と特徴抽出

KL展開を画像に適用する場合に，画素の輝度値を直接多変量の要素として画像の特徴抽出することが多いが，それでは環境の変化に対応できないため，本論文では，オプティカルフローを要素とする．さらに，KL展開によって多次元の多数のモデルをパターン空間に記述する．具体的には，各ジェスチャのモデル動作から動き情報(オプティカルフロー) X_i を計算して，計N個を採用し，

$$X = (X_1, X_2, \cdots, X_N) \tag{1}$$

とすると、KL展開して固有ベクトル

$$\phi = (\phi_1, \phi_2, \cdots, \phi_K) \tag{2}$$

を得る．ここで，固有値λ_iに対応する固有ベクトルϕ_iの要素は次式で表される．

$$\phi_i = (a_{i1}, a_{i2}, \cdots, a_{i,2N})^T \tag{3}$$

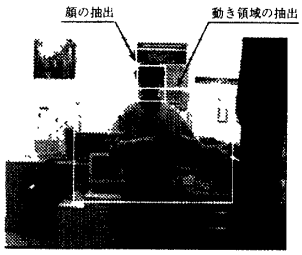

図 1: 顔の抽出による人物の動領域の抽出

3 HMMによるジェスチャ認識

HMMによる認識は，環境の違いなど雑音に対する影響や時間的な伸縮による影響を吸収することができる点で，有効な認識手法である．ジェスチャ認識は，各ジェスチャiに対応するシンボル系列をHMMλ_iにあらかじめ学習させておき，観測されたシンボル系列$Y = y_1 y_2 \cdots y_T$がジェスチャiである確率$P(Y|\lambda_i)$が最大になるHMMを選択することによって行う．

3.1 主成分得点のシンボル化

パターン空間に射影して得られる各主成分得点をシンボル化するためには，モデルデータをクラスタリングすることにより[2]，その代表点である標準パターンとの距離の比較によりシンボルに変換するのが，比較的簡単で学習に基づいてモデル化ができる点で有効な方法である．

第i番目のモデルとなる特徴量Z_iは第K主成分までの

$$Z_i = (z_1, z_2, \cdots, z_K)^T \tag{4}$$

で表され，すべてのモデル$(Z_1, Z_2, \cdots, Z_i, \cdots)$に対してクラスタリングを行うことになる．

そして，主成分をシンボルに変換する際には，あらかじめクラスタにラベル付けをしておき，入力主成分のベクトル$X = (x_1, x_2, \cdots, x_K)$と第$i$クラスタの代表点

*Gesture recognition by using image motion
†Yosio IWAI, Tadashi HATA, and Masahiko YACHIDA
‡Graduate School of Engineering Science, Osaka University

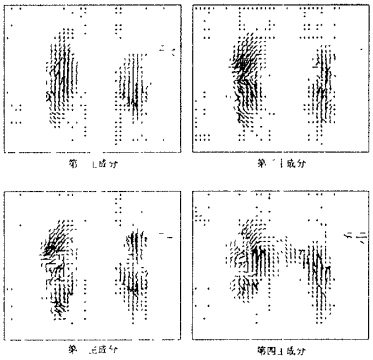

図 2: 各主成分に対する固有ベクトル

である標準ベクトル $Y_i = (y_{i1}, y_{i2}, \cdots, y_{iK})$ とのユークリッド距離

$$d_i = \left(\sum_{j=1}^{K} (x_j - y_{ij})^2 \right)^{1/2} \quad (5)$$

を計算し，

$$d_{min} = \min_{1 \le i \le N} d_i \quad (6)$$

となる i をシンボルにすることによりシンボル化をする．これを各フレームに対して求めることによりシンボル系列を作成する．

4 実験および考察

入力画像のサイズは 240x320 とし，ウィンドウの大きさは 165x195 とする．また，テンプレートは 9x9，各探索エリアは 27x27 とし，5pixel ごとにモーションベクトルを求めた．認識に用いたジェスチャは，楽器（ドラム，ギター，ピアノ，バイオリン，カスタネット）を演奏する５種類のジェスチャとする．ジェスチャ空間作成のためのモデル画像数は，一つのジェスチャにつき６０枚，合計３００枚用いた．そのサンプル画像を図3に示す．また，HMM の状態数は５とし，シンボル数を決定するクラスタ数は３５とした．

更に，第４主成分までの固有ベクトルを図 2に示す．この図から第１主成分は上下方向のフロー，第４主成分は左右方向のフローに影響があるのがわかる．また，これらを見ても分かるようにモデルから得られたフローによって固有ベクトルが決まるため，必要な所以外にフローが発生していても全く影響を及ぼすことがなく，背景が動いていたとしても認識が可能であることが分かる．

複数者の学習モデルを使用した場合，いずれのジェスチャにおいても 100%であったことからも分かるように，この２つのジェスチャにおいては個人差がかなり出てしまうようである．HMM を用いた認識の問題点としては，このような学習にないシンボル系列のデータを認識出来ない点があげられるが，複数の人物によるジェスチャ空間の作成と HMM の学習を行うことでこの点は解決された．また，この実験結果からも分かるように，背景や個人差などに依存することなく安定した認識系が構築できる．更に，この実験は人物を正面から撮影した画像を用いたが，左右に±２０度位までの回転した画像を用いても認識に成功している [3]．

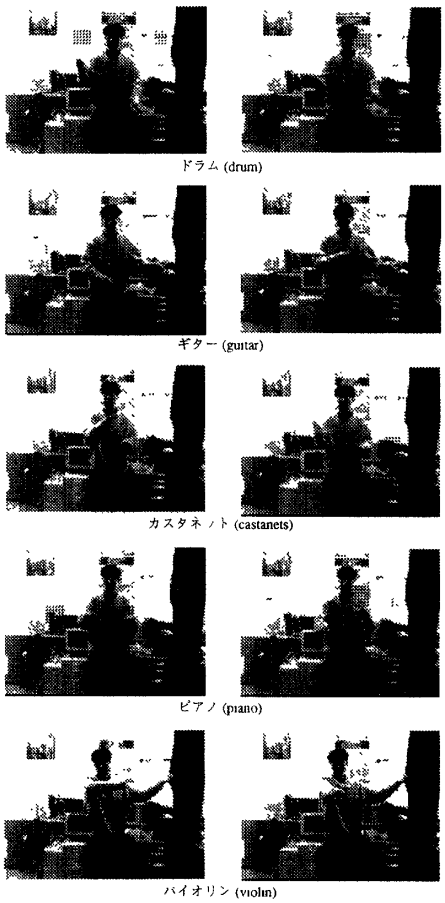

図 3: ジェスチャのサンプル画像

5 結論

本論文では，背景など環境の変化にロバストな人物のジェスチャ認識手法を述べた．カメラ画像からの動き情報を抽出し，それを固有空間上に表現することによってジェスチャ空間を作成，認識系に HMM を用いることによって，ロバストな認識が可能であることを実験で示した．今後は複数のカメラによる多眼視からの認識系を構築することによって，本手法の拡張性を検討していくつもりである．

参考文献

[1] Wu, H., Chen, Q. and Yachida, M.: A Fuzzy Theory Based Face Detector, *International Conference on Pattern Recognition*, Vol. 3, No. 13, pp. 406–410 (1996).

[2] Linda, Y., Buzo, A. and Gray, R. M.: An Algorithm for Vector Quantizer Design, *IEEE Trans. on Comm.*, Vol. 28, No. 1, pp. 84–94 (1980).

[3] IWAI, Y., HATA, T. and YACHIDA, M.: Gesture Recognition based on Subspace Method and Hidden Markov Model, *International Conference on Intelligent Robots and Systems*, (to appear).