

画像を用いた人間の形態分類に関する一考察

Human Motion Classification with Vision

○平塚誠良*, 大場光太郎*, 梶川伸哉**, 猪岡光*

○Shigeyoshi Hiratsuka*, Kohtaro Ohba*, Shinya Kajikawa**, Hikaru Inooka*

*東北大学, **宮城高専

*Tohoku University, **Miyagi National College of Technology

キーワード: マン・マシン・インターフェース (Man-Machine Interface),
人間とロボットの協調 (Human-Robot Collaboration), 人間の行動認識 (Recognition of Human Motion),
K-L 展開 (K-L Expansion), 固有空間 (Eigen Space),

連絡先: 〒 980-77 仙台市青葉区荒巻字青葉 東北大学 情報科学研究科 システム情報科学専攻
平塚誠良, Tel.: (022)217-7021, Fax.: (022)217-7019, E-mail: hiratsuka@control.is.tohoku.ac.jp

1. はじめに

1.1 人間とロボットの共存

近年, 人間と日常生活空間を共有し, 人間の能力の介助や共同で作業を行うロボットが求められてきている. 例えば, 病院内での介護ロボットや, 老人福祉向けのお手伝いロボットがある.

人間とロボットが生活空間を共有するうえで, 人間と協調して動作を行う機能が必要不可欠である. こうしたロボットには, 人間の行動に柔軟に適應することが求められる. このような協調性は, ロボットが人間の行動を理解, 予測, そして阻害しないことにより実現される. ^{1,2,3,4)}

現在, 接触状態では, 人間とロボットとの協調動作に関しては, 互いに力を及ぼしあうことから, この力による情報交換を用いた意思伝達が有効であることがわかっている (Fig1(a)). ⁵⁾ しかし, 非接触状態では力による情報交換ができないため, 何かの意思伝達手段が必要となる (Fig1(b)). この

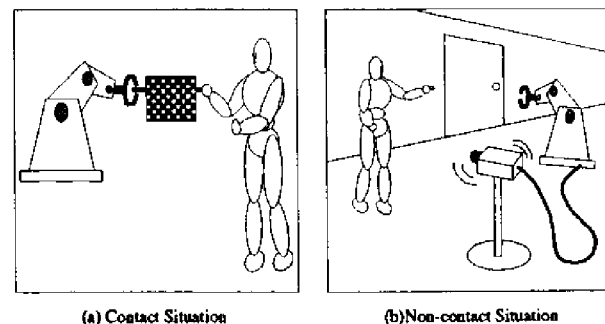


Fig. 1 人間とロボットの協調動作

ような情報交換のためには, 人間の視覚のように画像情報を用いた手段が有効であると考えられる.

こうしたマン・マシン・インタフェースが必要なので, 画像を用いた人間の手足の動きを認識する研究 ^{6,7)} や人間の行動の検出や認識をする研究 ^{8,9)} が盛んに行われている.

1.2 画像による人間の行動認識

画像から人間の行動を認識することを目的とした従来の方法として、トップダウン・アプローチや3次元復元のように人間の幾何学的なモデルを2次元の画像に当てはめるモデルフィッティングに基づく方法^{10,11,12,13)}、また、ボトムアップ・アプローチのように画像から得られる特徴量をヒューリスティックに利用する方法^{14,15)}などがあつた。

トップダウン的アプローチである幾何学的なモデルに基づく方法のメリットは、得られる情報が関節角パラメータなどの有用性の高いものであり、様々な応用が可能である。また、3次元の幾何学的な情報が復元できれば、それをもとにして姿勢や行動の解析や推定などが比較的容易に実現可能となる。しかし、問題点として、実画像の曖昧性やノイズの影響から、モデルと実画像とのフィッティングに時間を要し、また、安定して処理を行うことは困難であることが挙げられる。

一方、画像から得られる特徴量に基づく、ボトムアップ的アプローチでは、実画像での実現も多く比較的ロバストであると言える。現在のところ、監視システムでの、人数のカウントの用途に対しては実用レベルに達しつつある。しかし、問題点として、対象や状況に応じて、画像内で使用する特徴(例えば、エッジやコーナー)を人間がヒューリスティックに決定していること、比較的単純で限られたシーンでないと成功しにくいことが挙げられる。

また、実画像全体のパターンマッチングを行う場合は画像のデータベースが巨大になるなどの問題点がある。

そこで本報告では、モデルフィッティングによらない、K-L展開に基づく固有空間法^{16,17,18,19)}を用いた、人間の行動認識の手法について検討する。そして、その基礎実験として人間の形態分類の例を取り上げる。

K-L展開に基づく固有空間法では、画像データの次元を小さくすることが可能であり、処理を高速化することができる。

2. K-L展開に基づく固有空間法

2.1 固有値と固有ベクトル

K-L展開を用いた固有空間の作成方法について述べる。まず、学習する M 個の実画像より、次のマトリックスを作成する。

$$[z_1 \ z_2 \ \cdots \ z_M] \quad (1)$$

ここで、画像データは、それぞれ $N \times N$ の行列であり、長さが N^2 の列ベクトルに変換されている。このとき、行列(1)は、 $N^2 \times M$ の行列となる。

次に、式(1)で表される行列から平均画像 c を差し引き、行列 Z を作る。

$$Z = [z_1 - c \ z_2 - c \ \cdots \ z_M - c] \quad (2)$$

これより、 $N^2 \times N^2$ の共分散行列 Q が計算される。

$$Q = ZZ^T \quad (3)$$

そこで、K-L展開により固有方程式、

$$Qe_i = \lambda_i e_i \quad (4)$$

を解くと、固有値 λ_i と固有ベクトル $e_i (i = 1, \dots, N^2)$ が求まる。そして、固有値の大きい方から k 個とり、それに対応する固有ベクトル e_1, \dots, e_k を基底ベクトルとすることにより固有空間が得られる。一般的に、画像データの共分散のように次元数の大きな行列の固有ベクトルの計算は困難であるが、画像数が少ない場合には特異値分解やSTA法などを利用することにより計算することが可能である。²⁰⁾

また、式(4)より Q を固有値が λ_i である N^2 個の固有ベクトルに分解できることがわかる。このように、それぞれの画像データを、固有値と固有ベ

クトルによって表すことができる。このとき、固有値 λ_i は固有ベクトル e_i の重みとなっている。

2.2 実画像の固有空間表現

学習用サンプル画像データを用いて、一度、固有ベクトルを計算すれば、この固有ベクトルの行列、

$$E = [e_1 \ e_2 \ \dots \ e_k] \quad (5)$$

を用いて、以下の式(6)で、次元が N^2 である画像を固有空間上の点 g_i に投影することができる。

$$g_i = E^T(z_i - c) \quad (6)$$

この固有空間表現を用いることで画像(次元 N^2)を固有空間(次元 k)にまで小さくすることができる。

2.3 固有値と累積寄与率

固有空間において、次元を小さくしようとする場合、どの程度の次元が必要であることを示す尺度として、

$$W_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^{N^2} \lambda_i} \geq T_s \quad (7)$$

を用いる。ここで、 T_s は、しきい値、 N^2 はもとの実画像の次元、 k は固有空間の次元であり、 k はもとの N^2 に比べてかなり小さくなる。

式(7)より、累積寄与率は画像を再構成したときに、どの程度の2乗誤差で近似的に再構成できるかを示した指標である。

また、固有値 λ_i は固有ベクトル e_i の重みとなっているため、式(7)より求めた次元まで、実画像の次元を小さくできる。このため、ある物体に対する固有空間は、その物体の集合を表現するのに適した空間となる。

3. 人間の形態の固有空間への投影

人間とロボットが生活空間を共有するうえで、人間と協調して動作を行う機能が必要不可欠である。こうしたロボットには、人間の行動に柔軟に適應することが求められる。このような協調性は、ロボットが人間の行動を理解、予測、そして阻害しないことにより実現される。

ここでは、その基礎実験として、基本的な人間の形態として立位・座位を取り上げ、その分類を行う。処理は、形態を学習する段階と、学習した情報を利用して実画像データから形態を分類する段階に分けられる。まず初めに、立位と座位の形態の学習について述べ、次にその情報を用いた形態分類について述べる。

3.1 形態の学習

3.1.1 学習に用いた画像データ

形態の学習用サンプル画像データはCCDカメラを用いて収集する。形態の条件としては、被験者にカメラの画面中央で左方向を向いて、立つ・座るの動作を繰り返してもらう。(3)

CCDカメラからコンピュータへ取り込まれる画像の解像度は512(pixel)×512(pixel)である。まず初めに、実画像からバックグラウンドを取り除き、人間の領域を抽出する。次に、人間の領域

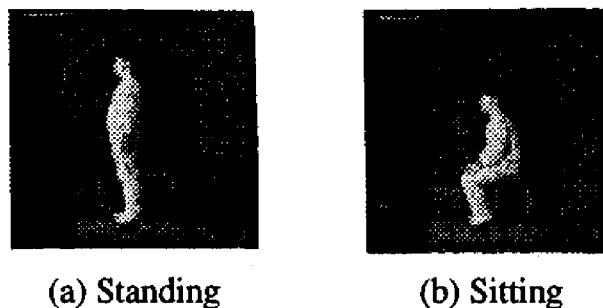


Fig. 2 学習用サンプル画像データ

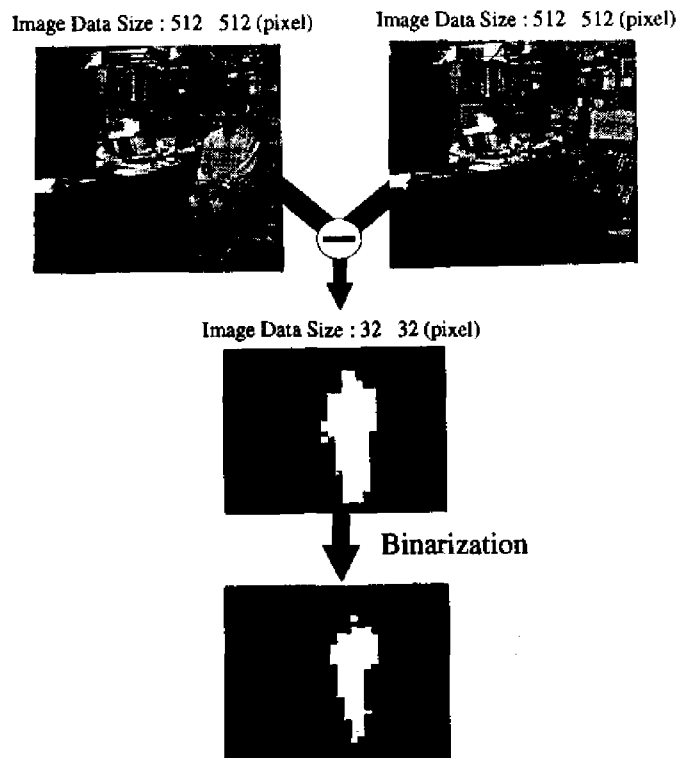


Fig. 3 画像データの処理過程

を抽出した画像の解像度を 32 (pixel) × 32 (pixel) に変換し、さらに、人間のシルエットを抽出する (Fig(3))。この画像データより式 (1) で表されるトレーニングセットを作成する。今回は、立位と座位、あわせて 100 フレーム分のデータを用いた。

3.1.2 固有空間の生成

学習用サンプル画像データを用いて、式 (1) で表される行列を作成し、式 (4) より固有値と固有ベクトルを計算する。ここでは、式 (5) における固有空間の次元は、 $k = 10$ とした。

Fig(4) は固有値 λ_i と累積寄与率 W_k を示している。これより、10 次元の固有空間でもとの実画像の 90% を再構成することができることがわかる。

3.2 形態の分類

形態の学習段階で得られた固有ベクトル (式 (5), 次元は $k = 10$) を用いて、式 (6) により、立

Table1: 立位の平均値と標準偏差 ($\times 10^3$)

	e_1	e_2	e_3
平均値	0.4042	-1.2549	0.4058
標準偏差	0.1287	0.0447	0.0363

Table2: 座位の平均値と標準偏差 ($\times 10^3$)

	e_1	e_2	e_3
平均値	1.2055	-0.7643	0.3647
標準偏差	0.1281	0.1319	0.0248

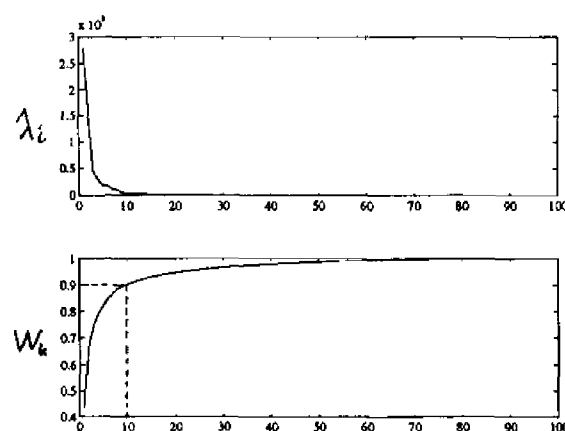


Fig. 4 固有値と累積寄与率

位または座位状態の画像データを固有空間に投影する。ここで用いられる立位・座位の画像データは学習段階で用いた画像データと同一の被験者から収集したものである。

そのときの結果を Fig(5) に示す。ここでは、実画像から固有空間への投影の計算で用いられている固有空間の次元は $k = 10$ であるが、プロットの関係上、固有ベクトルの上位、3 次元をプロットしている。

Fig(5) より、立位と座位の、形態の違いによる、固有空間での分布の特徴を観察することができる。ここで、立位、座位、それぞれ 3 次元の固有ベクトルの平均値 ($\bar{\omega}_j$) と標準偏差 (σ_{ω_j}) を Table1 および Table2 に示す。 ω_j は、同じ形態の集合 ($\omega_j = \{g_j | g_j \in (\text{同じ形態})\}$) を表す。

3.3 クラスタリング

実画像 (Z_{in}) を固有空間に投影したとき、人間の形態の違いにより、その投影点 (g_{in}) の分布に特徴がある。そこで、投影点のクラスタリングを行うことによって、人間の形態の分類を行う方法について考える。

形態が同じ場合、固有空間内の、ある限られた領域に投影されると考えられることから、その投影点の集合を考え、平均値と標準偏差を用いてクラスタリングを行う方法について検討する。

平均値と標準偏差を用いたクラスタリングの例として、簡単なアルゴリズムを以下に示す。

```

if
   $\|g_{in} - w_j\| \leq \sigma_{w_j}$ 
then
   $g_{in} \leftarrow w_j$ 
  
```

4. シーンを用いた実験

前節では、立位と座位の実画像データを固有空間に投影し、形態の違いによる固有ベクトルの分布を観察した。また、分布の特徴から、クラスタ

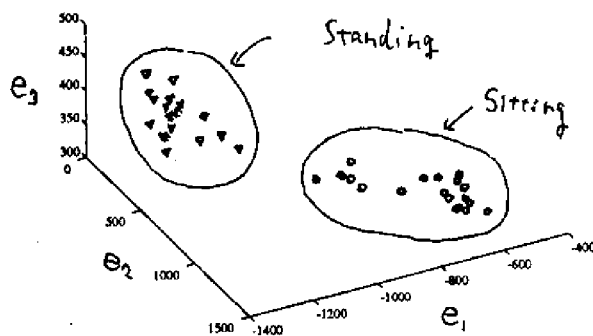


Fig. 5 立位・座位の固有空間への投影

リングによる形態分類の方法について検討した。

ここでは、前節で学習した固有ベクトルを用いて、一連の行動の中から立位・座位の形態分類を試みる。

4.1 実験条件

被験者に画面の右側から歩いて入って来てもらい、中央の椅子のある場所で立つ・座るの動作を数回、繰り返してもらい、その後、画面左側へ歩いて出ていってもらう (Fig(6)) という動作を行ってもらう。これらの実画像を前節で学習した固有ベクトルを用いて、固有空間に投影する。また、前節同様、計算に用いる固有空間の次元は $k = 10$ としている。

4.2 結果

このときの結果を Fig(7) に示す。固有空間内で、歩行・立位・座位・立位・歩行の状態の移り変わりを確認することができる。

また、被験者が学習時と実画像の投影時で違っていても、形態の分類行われていることがわかる。

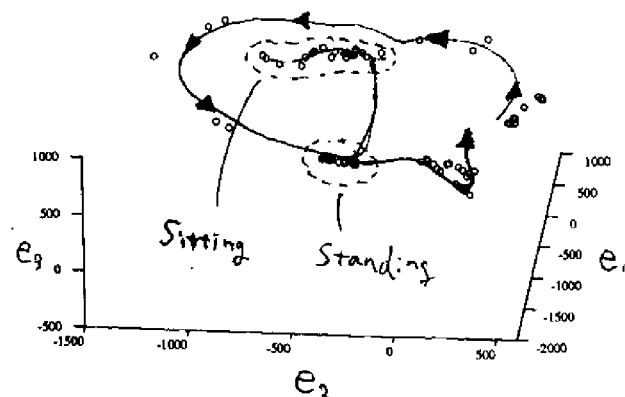


Fig. 7 一連の動作の固有空間への投影

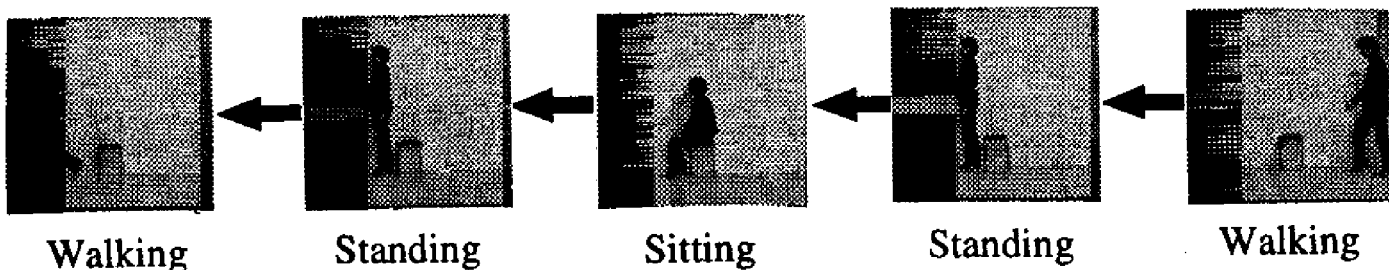


Fig. 6 一連の動作

5. おわりに

本報告では、モデルフィッティングによらない、K-L展開に基づく固有空間法を用いた、人間の行動認識の手法について検討した。そして、その基礎実験として人間の形態分類を行い、考察を行った。

結果として、画像データの次元を小さくし、処理を高速化することができた。また、実画像を固有空間へ投影することで、人間の形態の違いにより、分布が異なることが観察できた。それにより、固有空間上でのクラスタリングによる形態分類の方法について検討した。

今後の方針として、1) 固有空間上で形態を分類する際のクラスタリングの手法について、さらに検討を行うこと。また、2) 固有空間上の投影点の時系列データを用いた行動推定法について検討すること。3) 固有空間と状態遷移モデルの融合に関して検討すること、などが挙げられる。

参考文献

- 1) 橋野 賢, 他: 特集「医療福祉とロボティクス」, 日本ロボット学会誌, 14-5, 613/631 (1996)
- 2) 佐藤 知正, 溝口 博: 人にサービスするための意図, 行動の理解, 計測と制御, 34-4, 255/260 (1995)
- 3) 原文雄: 人間と協調するロボットの技術課題, 計測と制御, 34-4, 243/247 (1995)
- 4) 佐藤 知正, 西田 佳史, 市川 純理, 畑村 洋太郎, 溝口 博: ロボットによる人間の意図の能動的理解機能, 日本ロボット学会誌, 13-4, 545/552 (1995)
- 5) 池浦 良淳, 門田 猪岡 光: 人間とロボットの協調動作, 日本機械学会ロボティクス・メカトロニクス講演会講演論文集, 89/91 (1994)
- 6) 石淵 耕一, 岩崎 圭介, 竹村 浩雄, 岸野 文郎: 画像処理を用いた実時間手振り推定とヒューマンインタフェースへの応用, 電子情報通信学会論文誌, J79-D-II-7, 1218/1229 (1996)
- 7) 久野 義徳: コンピュータビジョンによる人間の指示行動の理解, 日本機械学会第1回 JSME ロボメカ・シンポジウム講演論文集, 96-8, 138/142 (1996)
- 8) Kosuke Sato, Tomoharu Maeda, Hirokazu Kato, Seiji Inokuchi: CAD-based Object Tracking with Distributed Monocular Camera for Security Monitoring, IEEE, 291/297 (1994)

- 9) 川角 浩亮, 関井 弘志, 榎本 暢芳, 大波多 元, 岡崎 彰夫: シルエットの投影パターンに基づく時系列データを用いた侵入者検出, 電気学会論文誌, 116-9, 1007/1014 (1996)
- 10) 山本 正信, 川田 聡, 近藤 拓也, 越川 和忠: ロボットモデルに基づく人間動作の3次元動画追跡, 電子情報通信学会論文誌, J79-D-II-1, 71/83 (1996)
- 11) 大和 淳司, 大谷 淳, 石井 健一郎: 隠れマルコフモデルを用いた動画からの人間の行動認識, 電子情報通信学会論文誌, J76-D-II-12, 2556/2563 (1993)
- 12) O'Rourke J, Badler N.: Model-Based Image Analysis of Human Motion Using Constraint Propagation, IEEE Trans. PAMI, PAMI-2-6, 522/536 (1994)
- 13) 亀田 能成, 美濃 導彦, 池田 克夫: シルエット画像からの関節物体の姿勢推定法, 電子情報通信学会論文誌, J79-D-II-1, 26/35 (1996)
- 14) 築山 俊史, 白井 良明: 連続する TV 画像からの人の動きの検出, 計測自動制御学会論文集, 19-11, 909/915 (1983)
- 15) 安富 敏, 森 英雄, 清弘 智昭: 歩行のリズムに基づく歩行者検出の一手法, 電子情報通信学会論文誌, J78-D-II-4, 608/617 (1995)
- 16) 村瀬 洋, シュリー K. ナイヤー: 多重解像度と固有空間表現による3次元物体のイメージスポッティング, 情報処理学会論文誌, 36-10, 2234/2243 (1995)
- 17) Hiroshi Murase, Shree K. Nayar: Visual Learning and Recognition of 3-D Objects from Appearance, International Journal of Computer Vision, 14-1, 5/24 (1995)
- 18) Takahiro Watanabe, Chil-Woo Lee, Masahiko Yachida: Recognition of Complicated Gesture in Real-Time Interactive System, IEEE International Workshop on Robot and Human Communication, 268/273 (1996)
- 19) Kohtaro Ohba, Katsushi Ikeuchi: Recognition of the Multi Specularity Objects using the Eigen-Windows, CMU Technical Report, School of Computer Science, CMU (1996)
- 20) Press, W., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T.: Numerical Recipes in C, Cambridge University Press, Cambridge (1988)