

インタラクションに基づく親和性の向上に関する研究

～人とロボットが同居するためのロボット要素技術に関する研究～

小川 行宏 三浦 達也* 山本 和彦*
王 剛* 稲葉 昭夫 杉山 正晴

Improving Human Friendly Interface based on Interaction with Robot - Development of Basic Technologies for Coexistence of Human and Robots -

Yukihiro OGAWA Tatsuya MIURA* Kazuhiko YAMAMOTO*
Gang WANG* Akio INABA Masaharu SUGIYAMA

あらまし 本研究は人とロボットのインタラクションによりロボットに対する親近感の創出を目的とし、人とロボットが共存するための要素技術として新たな手法を提案する。自然に人とロボットがインタラクションを行うためには、対応する相手が誰であるかを知り、どのような意図をもっているのかを知る必要がある。本研究ではカメラ画像を用いて人の意図を認識し、その情報を用いてインタラクティブに行動するためのロボットシステムについて提案する。人物認識および顔ポーズの認識実験を行い実環境で使用できる精度であることを確認した。さらに状況に応じた行動をするためにロボットの行動学習に顔ポーズを利用することによる強化学習の枠組みの適用を提案した。

キーワード インタラクション, 画像処理, 人物認識, 顔認識, 行動学習, 顔ポーズ

1. はじめに

現在成長が期待されている産業の一つとしてロボット産業がある。これまでのロボットは産業用ロボットに見られるように人の代わりに重労働や過酷な環境での作業に用いられ、生産の高度化、効率化、省力化のために活用されてきた。しかし今後、ロボットがさまざまな形で人社会に関わり、人の日常生活を支援し、人と共存することが予想される。これら人と共存して生活するロボットは信頼感や親近感を感じられるロボットであることが重要になる。

しかしながら、人がロボットに対して信頼感や親近感を持つためには様々な問題が存在する。鉄の塊であるロボットが動作することに対して恐怖を感じるといったハードウェア的な問題や、人の生活環境をいかにセンシングするか、どのように動作したらよいかというソフトウェア的な問題などが存在する。

これまでに人と共存するロボットの研究開発において様々な手法が検討されてきている^[1-5]。個人を識別し個人ごとにアクションする手法や、人とロボットのインタラクションにおけるインターフェースのモデルの提案などが挙げられる。しかしインタラクションの中で相手の意図を検出し、それに基づいて行動するものはない。

本研究では新たに人がロボットに親しみを感じるため

の手法を提案する。基本戦略として人とロボットのインタラクションにおける人の意図に着目する。ロボットは人の意図を検出することにより、意図に応じて行動し、行動を学習する。ロボットが各個人に対して人同士のよう自然にインタラクションを行うことができ、各個人がどのような意図をもっているのかを知ることができれば信頼感や親近感を創出できると考える。

本報ではカメラ画像から得られた情報を用いることにより、インタラクティブに行動するためのシステム構成について提案する。主にカメラから得た画像を用いた人物の認識、および人の意図の検出について述べる。また、その画像から得た情報を利用することによりロボットが各個人に対応した行動の学習を行う手法について提案する。図1に人とロボットのインタラクションのイメージ図を示す。



図1 人とロボットのインタラクションイメージ図

* 岐阜大学 工学部 応用情報学科

2. 親近感創出のための戦略

本章では人のロボットに対する親近感を創出するための基本戦略の概要および基本戦略の中心となる意図の認識について述べ、それを実現するためのシステム構成を示す。

2.1 基本戦略

人とロボットが人同士のようにインタラクティブに接することができれば、人のロボットに対する信頼感や親近感創出が期待できる。本研究では人とロボットの親和性を向上させる手段として、人とロボットのインタラクションにおける人の意図に着目した手法を提案する。手法の概要を図2に示す。インタラクションにおいて相手が誰であるかを特定し、どのような意図を持っているのかを知ることが重要である。本手法では、ロボットは環境の認識において人物認識と人物の意図認識を行う。この認識結果により人の意図に応じて行動を切り替える。さらに、ロボットは行動した後人の意図を認識することによって、ロボットの行動に対する評価・学習をする。提案した手法により、環境の認識と行動(決定・学習)の操作を繰り返し行うことにより人とロボットの間徐々に親近感が創出されることを期待する。

2.2 意図の認識

人の意図という非常に広い意味を持っている。例えば手を上げるということは動き自体が意図である、また行動していなくても人がこちらを向いているだけでも意図であると言うことができる。このような中で本研究では意図の認識に関して顔の表情に着目する。顔の表情は意図の中で目に見える形で認識可能なものの一つであり、意図の表現の中で重要な役割を果たしていると考えられる。この例として「人の顔色を伺う」という言葉があるように、その人物が何を考えているのかを知るために顔を見るということが挙げられる。しかしながら表情は各個人によって異なる。また、同じ表情でも意図が一意に定まるものではない。

本研究では問題を簡単にするため予め顔のポーズを表情として定義する。例えば、「顔をしかめている」場合は「怒っている表情」とし、「目を大きく開けて口を大きく開いている」場合は「驚いている表情」とする。この定義が全ての場合で成立するわけではないが一般的な表情においては多くの場合に成立すると考えられる。顔のポーズを人の意図であると仮定し、それを認識することによりシステムを構成する。



図2 システム概要

2.3 システム構成

図3に提案した手法を実現するシステムの構成を示す。ロボットにはカメラとマイクを装着する。カメラからの画像を用いて人物認識および顔ポーズ認識を行う。マイクからの音声を用いて音声認識を行う。画像認識、音声認識ともに常に入力を受け付けること可能で、双方の認識結果に基づいて行動を決定し、実際に行動に移る。ロボットの行動にはロボット動作および音声を使用する。ロボットの動作に対する人のリアクションがロボットに対する次の入力となる。さらに、ロボットは顔ポーズの認識を行うことにより自分のとった行動を評価し学習する。人物認識・顔ポーズ認識について3章で詳細に述べ、行動学習については4章で提案する。

3. 環境の認識

3.1 人物認識

人とロボットがインタラクションを成立させる上で重要なこととして、相手が誰であるかを理解する必要があると考えられる。そこでカメラから得られる画像を用いて人物認識を行う。入力画像から人物認識の結果を出力するまでの流れを図4に示す。カメラ画像から色情報を用いて人物の肌色らしき領域を抽出し、それを顔領域とする^[6]。肌色領域の抽出には、人の色に対する感覚に近い均等知覚色空間のひとつであるLUV表色系を用い、ダイナミックに閾値を変化させ、環境の変化や個人差に対応した顔領域抽出を行う。またエッジ特徴である4方向面特徴を抽出し、線形判別分析により特徴量を算出しあらかじめ用意されている辞書とユークリッド距離を比較することにより人物認識を行う^[7]。線形判別分析はより少ない次元でクラス間の分散を最大限強調する線形写像を構成する多変量解析手法である。4方向面特徴については次節で詳しく述べる。

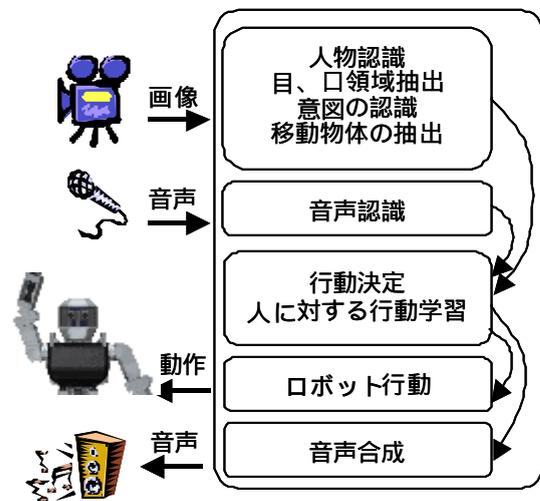


図3 システムの内部構成

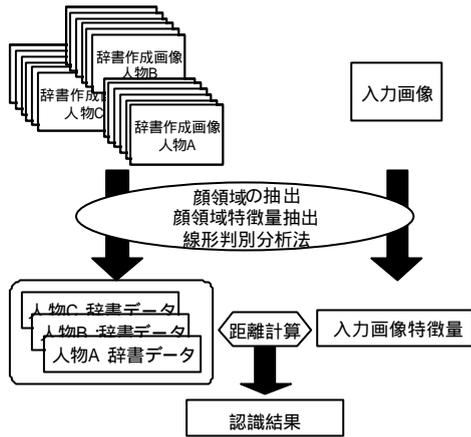


図4 人物認識の流れ

3.2 4方向面特徴

本研究では人物の認識を行うための特徴として4方向面特徴を用いる。一般的にエッジ特徴は文字認識や人物認識の分野で良く用いられている方法である。その中でも4方向面特徴は濃淡の入力画像に対し方向検出フィルタを用いて、方向毎にエッジ強度を濃淡特徴とする手法である。ここで4方向面とは水平方向面、右上がり方向面、垂直方向面、右下がり方向面のことである。

4方向面特徴は高次特徴面に分けた後、低解像度化しているため、1つの画像を4方向面に分ける前にエッジを抽出し解像度を落とす場合と比べて、高解像度でのエッジ情報が残る。また、4方向面特徴では非常に低い解像度まで落とすため、顔自体や顔部品の位置のわずかなずれが特徴量に及ぼす影響が少なくなるという性質を持っている。図5に方向検出フィルタを用いて4方向面特徴を抽出した例を示す。

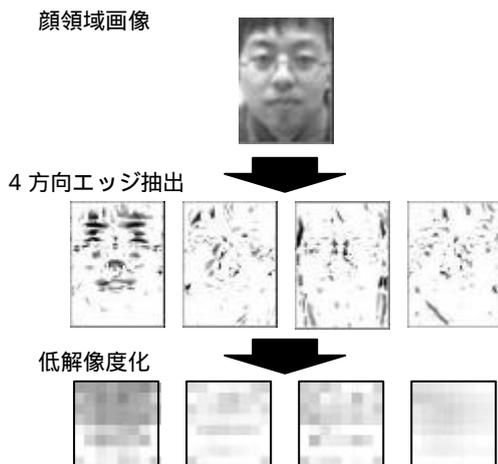


図5 4方向面特徴の抽出



図6 顔ポーズ認識の実験画像

3.3 顔ポーズ認識

人物認識と同様に得られた顔画像領域から4方向面特徴を抽出し、あらかじめ用意されている辞書と比較することにより意図の認識を行う。本研究では図6に示す3パターンの顔ポーズに無表情を加えた4つの顔を表情として定義し顔ポーズ認識を行う。

3.4 認識実験

本節では4方向面特徴を用いた認識実験を行い、認識システムが使用できるものであるか検討する。

各認識実験は5人を対象とする。実験に用いる画像は連続して7.5フレーム/秒で撮影した個人、ポーズごとに400枚、一人につき1600枚の画像とする。400枚のうち100枚を辞書作成用の学習データとし、使用していない残りの画像のうち100枚を未学習データとして使用する。

人物認識の辞書作成用の画像は各個人の無表情のデータを使用する。また、人物認識実験の未学習データには各顔ポーズ画像(無表情を含む)を100枚1人につき400枚ずつの合計2000枚の画像を使用する。人物認識実験の結果を図7に示す。縦軸は認識率である。未学習データに顔ポーズが変化したものを使用したが、非常に高い認識率を得ることができた。

顔ポーズ認識は各個人の1つのポーズごとに100枚、合計2000枚を学習データとして使用した。同様に1つのポーズごとに100枚ずつの未学習データを使用して認識実験を行った。顔ポーズ認識実験の結果を図8に示す。縦軸は認識率である。各ポーズともに学習データに比べ未学習データでは認識率が同じもしくは低下するものの、高い認識率となり、今回使用した辞書は特徴空間をうまく分割しており信頼性の高いものといえる。

3.5 実環境での認識実験

次に前節で使用した辞書データを使用して実環境での認識実験を行う。実験は顔ポーズ認識で辞書に登録して

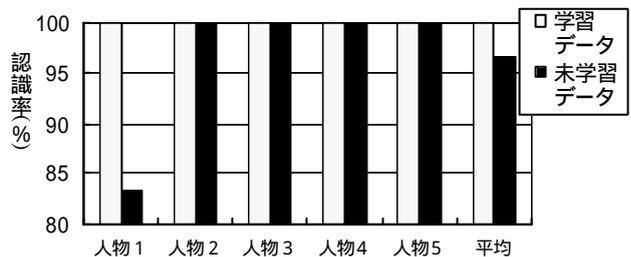


図7 人物認識の認識率

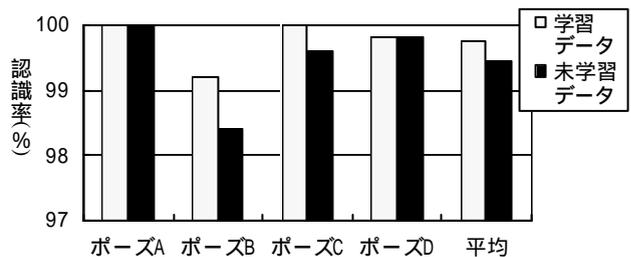


図8 顔ポーズ認識の認識率

ある1名(人物3)を対象に行う。カメラの前に誰もいない状態から辞書に登録された人物が現れ、顔ポーズを変化させカメラから出て行く様子をフレームごとに認識する実験を行った。実験データは辞書データの作成時とは別の時間に実際に認識動作しながら約2フレーム/秒で撮影したものである。

各フレームの認識結果を図9(a)に示す。横軸はフレーム数、縦軸は各フレームの認識結果である。縦軸は認識結果が被験者の各顔ポーズ(4種類)、他人、無人状態の6つのどの状態かを示したものである。また、図の上に被験者が実際にどの顔ポーズ行っているかを示す。各フレームで認識した場合は、1フレームや2フレーム程度の認識失敗が散らばっており認識が安定しないことがわかる。

そこで前16フレームの結果から最も数の多いクラスを認識結果とすることにより安定化を図った。前16フレームによる安定化の結果を図9(b)に示す。前フレームの結果を使用することにより、多少の時間遅れが生じるものの認識結果が安定し、実環境において使用可能な認識率となった。

4. 行動学習

これまでに顔ポーズを用い、ロボットが環境を認識することについて述べた。本章では強化学習の枠組みを取り入れることによりロボットの行動を決定することについて述べる。

強化学習は環境とエージェントの相互作用を通じて目的の行動を獲得する手法として有効であると言われてい^[8]。本研究では強化学習の一手法であるQ学習をロボットの行動系として適用し、相手の状態に応じた行動を学習することを提案する。Q学習は強化学習の一手法でありロボットを含めた環境に関する先見の知識をほとんど必要としないという利点がある。

ロボットが観測する状態を人物および人物の顔ポーズとし、報酬はロボット行動後の相手の状態から受け取る。例えば相手を喜ばせるという政策であれば、顔ポーズが喜んだという状態になった場合に報酬を受け取る。このように状態、報酬を設定し行動を学習することで、相手の意図に添った行動を選択できるようになる。システムの実装については今後の課題である。

5. まとめ

本研究では人とロボットの親和性の創出を目的としたシステムを提案した。その中で重要な要素となる環境の認識として、人物認識、顔ポーズの認識を行い実環境でも使用できる精度であることを確認した。また、状況に応じた行動をとるために、ロボットシステムの行動学習への強化学習の枠組みの適用を提案した。

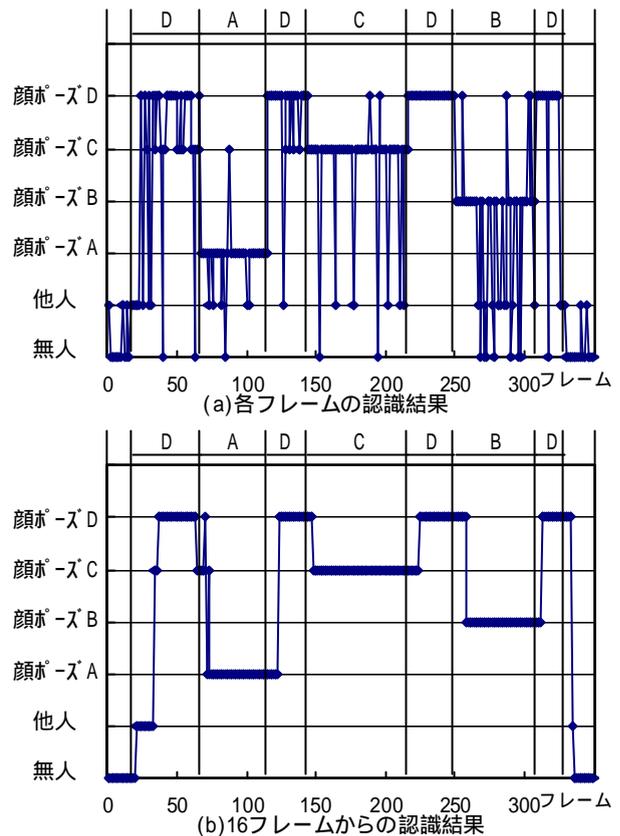


図9 実環境での実験結果

文献

- [1] 日本電気株式会社, “パーソナルロボット「Papero」”, <http://www.incx.nec.co.jp/robot/>
- [2] 藤田雅博, “ロボットエンタテインメント:小型4脚自律ロボット”, 日本ロボット学会誌 Vol.16 No.3, pp.313-314, 1998.
- [3] 株式会社国際電気通信基礎技術研究所, “日常活動型ロボットプロジェクト”, <http://www.mic.atr.co.jp/~michita/everyday/>
- [4] 小野哲雄, 今井倫太, 江谷為之, 中津良平, “ヒューマンロボットインタラクションにおける関係性の創出”, 情報処理学会論文誌 Vol.41 No.1, pp.158-166, 2000.
- [5] 佐部浩太郎, 藤田雅博, “エンターテインメントロボットの商品化”, 日本ロボット学会誌 Vol.18 No.2, pp.185-187, 2000.
- [6] 羽飼直記, 本郷仁志, 加藤邦人, 山本和彦, “顔画像からの眼部抽出方法の提案”, SII'98 ,D-4 pp.185-190, 1998.
- [7] 栗山聖, 山本和彦, 本郷仁志, 加藤邦彦, “顔認識システムにおける高次特徴の効果と性能”, MIRU2000, pp.475-480, 2000.
- [8] Richard S.Sutton, Andrew G.Barto著 / 三上 貞芳, 皆川 雅章訳, “強化学習”, 森北出版社, 2000.