

## 自律移動ロボットのための直接例示による対話動作の習得方法

## An Acquisition Method of Human-Robot interactive Manners through Direct Exemplifications

長谷川 高輔†  
Kosuke Hasegawa大盛 善啓†  
Yoshihiro Ohmori石井 健太郎†  
Kentaro Ishii今井 倫太†  
Michita Imai

## 1 まえがき

近年、人型ロボットの開発が進み、実社会で人間と接する作業をすることが期待されている。その実現には様々な技術が必要だが、人間と自然に対話するための技術は極めて重要な研究課題のひとつである。

本研究では、人間がお辞儀するとロボットもお辞儀を返す、といった、人間とロボットの間での対話動作の習得方法を扱う。従来のロボットでは、プログラマが対話動作を発動させるための条件式を1つ1つ作り込む必要があった。相手が人間ならば、言葉や身振りで対話動作を教えられるが、ロボットに対話動作を教えるにはどうすればよいだろうか。

人間はロボットをあたかも自分であるかのように操作できる。ラジコンがその良い例である。本研究ではこの性質に着目して、自律移動ロボットに対話動作を直接例示して習得させる方法を提案する。

提案手法の有効性を確かめるために、我々が開発した自律移動ロボット Uni を用いて実験を行う。提案手法によると、5種類の対話動作を約10分で習得でき、限定された環境ならば対話者の動作を85%以上の確率で識別して対話動作を実行できる。

## 2 関連研究

従来からロボットと人間の対話に関して研究されている。神田らは、人間とロボットの対話に関する性質を調べて、ロボットが人間を対話に引き込むにはジェスチャが重要であることを示した [1]。この結果は本研究で対話動作を扱うことを動機付けている。

入力信号に反応した動作を扱う点で、機械の自動制御は提案手法と類似している。Katoらは、プロのカメラマンがカメラを制御する方法を分析して、サッカーの試合を自動撮影する方法を提案した [2]。この研究では、動作を人手で分析してロボットに組み込む必要があった。

動作をロボットや機械に習得させる点で、手本を見せて真似させることによって動作を習得する模倣学習は、提案手法と類似する。稲色らは、人型ロボットが、他の人型ロボットの動作を見て習得する手法を提案した [3]。これまでの模倣学習に関する研究では、手本と同じ動作を習得することに焦点が当てられており、人との対話動作を習得したものはなかった。

このように、従来はロボットに対話動作をさせるには、プログラマが大変な労力を払って、対話動作を発動させるための条件式を1つ1つ作り込む必要があった。

## 3 直接例示による対話動作の習得

図1に、提案手法による対話動作習得システムの構成図を示す。本システムは、動作習得部(図中実線で囲まれた部分)と動作実行部(破線で囲まれた部分)とから構成される。以下にそれぞれの詳細を説明する。

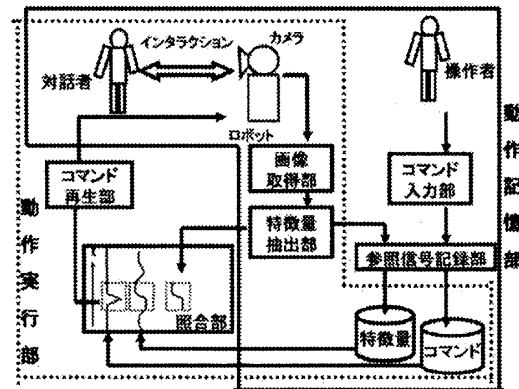


図1 対話動作習得システムの構成図

## 3.1 対話動作の習得

対話動作の習得は、ロボット、ロボットと対話する対話者、およびロボットを操作する操作者の3者で行う。

動作習得部では、例えば対話者がお辞儀をすると、操作者がロボットを操作して、ロボットにお辞儀をさせる。その際、ロボットに搭載されたカメラで画像を取得し(画像取得部)、特徴量を抽出する(特徴量抽出部)。ロボットには、お辞儀、ハイタッチ、などの一連の動作をあらかじめ組み込んで、操作者は対話者の動作に合わせて、動作に対応したキーボードのキーを押して操作する(コマンド入力部)。操作者はロボットの腕や首などの各部を詳細に操作することもできるが、本研究では、まず対話動作を発動させるための条件式の習得に関して実験する。

特徴量とコマンドは1秒間に30回同期して入力し、参照信号としてファイルに記録する(参照信号記録部)。特徴量は、カメラから得られる画像を $3 \times 3$ の9等分して、それぞれの領域の輝度の平均値を等間隔の4段階(2bit)に量子化する。9個の量子化された値を連結して18bitの符号にする。

## 3.2 対話動作の実行

対話動作の実行は、ロボット、およびロボットと対話する対話者の2者だけで行う。

動作実行部では、ロボットに搭載されたカメラで対話者の動きを撮影し(画像取得部)、特徴量を算出して(特徴量抽出部)、入力信号を得る。10秒間の入力信号を取得すると、これをテンプレートとみなし、参照信号上でテンプレートと同じ長さの10秒の窓をスライドさせながら距離を算出する。その距離を用いて照合し、類似する部分が見つかった場合、その部分に対応付けられたコマンドを再生してロボットに対話動作をさせる。

式(1)に参照信号と入力信号の照合で用いる距離の算出方法を示す。 $d_j$ は窓の先頭から $j$ 番目の参照信号と入力信号の距離を表す。 $u_i$ ,  $v_i$ は参照信号, 入力信号のフレーム画像中での $i$ 番目の領域の平均輝度値を表し, $d_j$ は $u_i$ と $v_i$ の

†慶應義塾大学大学院 理工学研究科

†慶應義塾大学 理工学部

SAD (Sum of Absolute Distance) として算出する。参照信号の位置  $t$  における入力信号との距離  $D(t)$  は、窓に含まれる 300 個 ( $30[\text{fs}] \times 10[\text{s}]$ ) の  $d_j$  の総和として算出する。

$$D(t) = \sum_{j=1}^{300} d_j, \quad d_j = \sum_{i=1}^9 |u_i - v_i| \quad (1)$$

照合結果は距離が最も小さい位置  $t_{\text{match}}$  とする。ただし、距離がしきい値  $T$  を下回らない場合は照合に失敗したとする。式 (2) に算出方法を示す。また、図 2 に、照合の様子を示す。

$$t_{\text{match}} = \begin{cases} \arg \min(D(t)) & D(t_{\text{match}}) < T \text{ の時} \\ -1 & \text{それ以外} \end{cases} \quad (2)$$

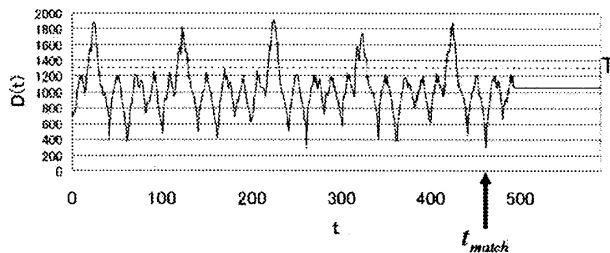


図 2 照合の様子

## 4 実験

### 4.1 実験方法

5 種類の対話動作を何回記憶すれば、対話者の動作を実用的に識別できるかを調べた。表 1 にロボットに記憶させる対話者の動作の種類を示す。記憶回数が 1, 3, 5 回の場合について、動作時の識別率を調べた。

図 3 に対話動作を記憶させる様子を示す。対話動作の記憶は、背景が白い部屋にロボットを置いて、1 から 5 の動作を順番に記憶させた。対話者は初めロボットの視界の外にいて、記憶時にロボットの正面に移動して対話動作した後、ロボットの視界の外に移動した。対話者の動作が連続すると動作を区切ることが必要となるため、今回の実験ではこの影響を排除することにした。1 回の対話動作の記憶は約 20 秒とした。1 回、3 回の参照信号は、5 回の参照信号からランダムに回を切り出して生成した。

これらの 3 種類の参照信号に対して、それぞれ 20 回の対話動作を行い、識別率を算出した。入力信号はあらかじめ記録しておいて、各試行で同じ入力信号を用いた。

表 1 対話動作の種類

1	お辞儀をする
2	ロボットに近づいて顔をのぞき込む
3	ハイタッチをする
4	手を振る
5	ロボットの前を歩き来して正面で立ち止まる

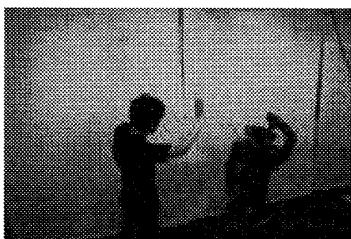


図 3 対話動作を記憶する様子

## 4.2 実験結果

図 4 に記憶回数と識別結果の関係を示す。対話動作 2(近づく)と 4(手を振る)は、記憶回数が 1 回でも識別率が高かった。その他の動作は記憶回数が増えると識別率が上がって、5 回ではいずれの動作も 0.85 以上だった。

記憶回数 1 回では、対話動作 3(ハイタッチ)の識別率が低い。これは対話動作 4(手を振る)と多く間違えたためだった。

記憶にかかった時間は、1, 3, 5 回がそれぞれ 2, 4, 6 分だった。探索にかかった時間は、最大で約 6 秒だった。

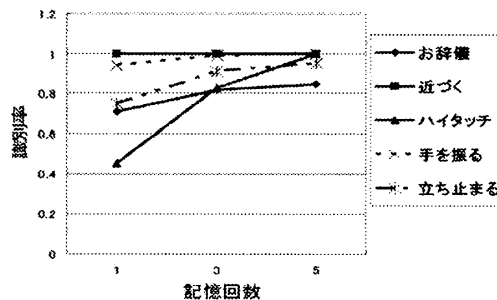


図 4 記憶回数と識別率の関係

## 5 考察

対話者の立ち位置や動きのタイミングが記憶時と動作時で異なるために、誤識別が生じると考えられる。記憶回数が増えると、類似する動きが参照信号に存在する可能性が高くなるために識別率が高くなったと考えられる。

対話者の動作が似ていても誤識別が生じると考えられる。例えば、対話動作 3 (ハイタッチ) と対話動作 4 (手を振る) は参照信号が似ているために対話動作 3 (ハイタッチ) の識別率が低くなったと考えられる。このように識別が難しい似た動作でも、記憶回数を増やすと識別できることがわかった。

5 種類の対話動作をプログラミングにより実装するには、それぞれの動作に対して個別の条件式を設計しなければならず、莫大な時間がかかり負担も大きい。これに対して提案手法では、5 種類の対話動作を同じ手法で実装でき、時間の点でも、5 回習得したとしても約 10 分しかかかっておらず、作業者の負担を大幅に軽減できる。

## 6 まとめと今後の課題

人間がロボットを操作して、お辞儀などの対話動作をロボットに直接例示することにより、ロボットに対話動作を習得させる手法を提案した。自律移動ロボットを用いた実験により、5 種類の動作をわずか 10 分程度で記憶して、約 85% の識別率で実行できた。

今後は、照合時間を短くして対話動作の反応を改善する。また、身体を持たないエージェントなどでの検証実験も行う。

## 参考文献

- [1] Takayuki Kanda, et al : "Development and Evaluation of Interactive Humanoid Robots", Proc. of the IEEE, Vol.92, No.11, Nov. 2004.
- [2] Daiichiro Kato, et al : "Automatic Control of Robot Camera for Broadcasting Based on Cameramen's Techniques and Subjective Evaluation and Analysis of Reproduced Images", Jour. of Physiological Anthropology and Applied Human Science, 19 (2), pp.61-71, 2000.
- [3] 稲邑 哲也, ほか: "ミメシス理論に基づく見まね学習とシンボル創発の統合モデル", 日本ロボット学会誌, Vol.22 No.2, pp256-263, 2004.