

生活支援ロボットのマルチモーダル対話のためのジェスチャ認識

日大生産工(院) ○木林 龍一
日大生産工 岡 哲資

1 はじめに

今日、さまざまなロボットが災害現場をはじめ、さまざまな場所で活躍している。しかし、それらの多くは特別な知識を持った専門家が操作している。そんな中、生活支援ロボットの一つであるルンバは、世界中で500万台販売された。将来、高齢化社会を背景に、多目的の生活支援ロボットが普及すると我々は考えている。そのようなロボットは、特別な知識なしで利用できる必要がある。

本研究では、生活支援ロボットを利用するために基本となる、ロボットをユーザに向かせる、近くに來させる、適切な距離を保つ、の3つの命令に必要なコミュニケーションに焦点を当てる。このようなコミュニケーションには、音声認識と手の三次元的な動きの認識を組み合わせ、ユーザの意図を理解する事が有効である。このような複数の伝達方式によるロボットの命令をマルチモーダル命令と呼ぶ。また、ロボットからも音や画面表示によって情報伝達を行い、マルチモーダル対話を可能にすることが重要である。例えば、ユーザは、現在命令が可能かどうかなど、ロボットの状態を把握する必要がある。

本研究では、前述の基本的なコミュニケーションを目的とした生活支援ロボットとのマルチモーダル対話を可能にするためのジェスチャ認識を実現し、ユーザ評価を行った。ユーザ評価の目的は、初心者が短時間で正しくジェスチャを行う事ができるか、実現したシステムがジェスチャを正しく認識できるか、の二点を検証することである。

これまで、いくつかのマルチモーダル命令の研究例がある。それらでは、音声にボタン入力[1]、ロボットのボディの接触[1]、二次元的なジェスチャを組み合わせさせた命令[1,2]によって、初心者がロボットにいくつかの命令ができることを示した。

しかし、これらの命令方法は、ロボットに三次元位置や距離の情報伝達が行えない。また、タッチスクリーンを用いて目標物や目標地点の三次元情報を伝達し、音声と組み合わせる研究も行われている[3]。しかし、この方法は、手元にデバイスが必要であること、ユーザ自身の位置を伝えづらいこと、距離の伝達が直感的でないこと、などが問題である。これに対して、本研究では、手を動かすことでユーザ自身の位置、目標位置、短い移動距離などを直感的に伝える方法を採用した。例えば、「こっちをむけ」という音声と手を上げるジェスチャを組み合わせることで意図を伝達する。音声のみやジェスチャのみでは、このような情報伝達は容易に行えない。例えば音声だけでは、「右に73° 旋回しろ」などの不自然な表現になってしまい、また角度や距離などの数値を特定することは難しい。また、ジェスチャのみでは、さまざまな種類の命令を行う事が困難である。

最近では従来よりも安価で性能の良い距離画像センサーKinectが販売されている。これを利用すれば、手の3次元位置を実時間で取得できる。そして、取得した手の3次元位置から手の動きを計測してジェスチャ認識ができる。また、音声認識エンジンの性能も年々高まってきている。

本研究で実現したジェスチャ認識システムは、4種類のジェスチャを検出する。これらは、以下のマルチモーダル命令に利用するためのものである。

- ・「こっちを向け」といいながら、手の動きで旋回目標を伝達する命令
- ・「こっちに來い」といいながら、手の動きで自身の位置を伝達する命令
- ・「下がれ」といいながら、移動量を伝達する命令

Gesture detection based on 3D tracking for multimodal communication with a life-supporting robot

Ryuuichi, Kibayashi and Tetsushi OKA

・「来い」といいながら、移動量を伝達する命令

移動量の伝達には、手の三次元での移動量を利用する方法と、移動の開始と停止の合図を送る方法の二種類を考えた。前者は、より直感的であるが、短い移動量のみには適用できない。後者は、長い移動量の指定に利用することを想定している。

具体的なジェスチャは、手を上げる、下げる、前に出す、手前に引く、の4種類である。ユーザ評価では、生活支援ロボットに全くの初心者11名がジェスチャを行い、10分以内に正しくジェスチャを行った。また、正しいジェスチャの認識率は100%であった。これらの結果と質問への答から、生活支援ロボットとのマルチモーダル対話に手の3次元位置によるジェスチャ認識が効果的であることが分かった。

2 ジェスチャ認識システム

以下では、ジェスチャ認識システムについて説明する。

2-1では生活支援ロボットに命令するための4種類のジェスチャについて、2-2ではジェスチャを行うユーザの発見の仕組み、2-3ではユーザの右手の発見の仕組みを説明する。2-4では各4種類のジェスチャの判定方法を説明する。

2-1 4種類のジェスチャ

実現したジェスチャ認識システムは、生活支援ロボットに命令するための4種類のジェスチャを認識する。これらは、生活支援で想定される以下に示すような意図伝達に用いる。

- ・ユーザの方を向かせる命令
- ・ユーザの近くに來させる命令
- ・ロボットとユーザの距離を調整する命令
- ・アクションの開始と終了の合図

以上の意図伝達のために片手を使った以下のジェスチャを認識する。

- ・手を上げる
- ・手を下げる
- ・手を前に出す
- ・手を手前に引く

なお、全てのジェスチャは、一定以上の速度で一定以上の距離直線的に手を動かして行う。

2-2 ユーザの発見

ユーザの発見は、以下の順に行う。

1. 視野に入っている人物のラベル付け
2. 各人物の頭の3D座標の取得
3. 一番近くに座っている人物の特定

まず、Kinect用のSDKであるOpenNIを用い、人物を複数人検出し、それぞれにラベル付けを行う。次に、各ラベルの一番高い位置を発見し、それぞれの3D座標を取得する。最後に、視線より低い位置にあり、一番近くにある頭のラベルをユーザとして特定する。

2-3 ユーザの右手の発見と追跡

発見したユーザの領域で、頭から右端まで輪郭をたどる。その際、最も低い場所を肘とし、肘と領域の右端の間で一番高い画素を決定する。その画素の座標を右手の初期位置とし、OpenNIの手の追跡関数を利用して、右手の実時間追跡を行う。これによって、右手の三次元位置が毎秒30回獲得される。OpenNIの関数が追跡失敗の判定をした時は、再度右手発見のための処理を行う。

2-4 4種類のジェスチャの判定

以下で、4種類のジェスチャの判定について説明する。上下方向のジェスチャ検出では、平均0.3[m/s]以上、前後方向では0.15[m/s]以上の手の動きのみを判定対象とする。

2-4-1 手を上げるジェスチャ

手を上げるジェスチャの判定条件は、以下の通りである。

- ・各フレーム0.003[m]以上かつ1秒以内の連続した移動
- ・手の停止位置が頭より高い
- ・手の動きだしの位置が頭より低い
- ・手の移動が0.13[m]以上
- ・左右の移動が0.08[m]未満
- ・1フレームあたり0.028[m]以上の移動を含むこと

2-4-2 手を下げるジェスチャ

手を下げるジェスチャの判定条件は、以下の通りである。

- ・各フレーム0.003[m]以上かつ1秒以内の連続した移動
- ・手の停止位置が頭より低い
- ・手の動きだしの位置が頭より高い
- ・手の移動が0.09[m]以上
- ・左右の移動が0.08[m]未満
- ・1フレームあたり0.03[m]以上の移動を含むこと

2-4-3 手を前に出すジェスチャ

手を前に出すジェスチャの判定条件は、以下の通りである。

- ・各フレーム0.0015[m]以上かつ1秒以内の連続した移動
- ・手の停止位置が頭からよりも0.4[m]以上前方
- ・手の動きだしの位置が頭から0.4[m]前方までの範囲内
- ・手の移動が0.019[m]以上
- ・左右の移動が0.11[m]未満
- ・上下の移動が0.15[m]未満

2-4-4 手を引くジェスチャ

手を前に出すジェスチャの判定条件は、以下の通りである。

- ・各フレーム0.0015[m]以上かつ1秒以内の連続した移動
- ・手の停止位置が頭から前方0.4[m]までの範囲内
- ・手の動きだしの位置が頭よりも0.4[m]以上前方
- ・手の移動が0.019[m]以上
- ・左右の移動が0.20[m]未満
- ・上下の移動が0.15[m]未満

3 ユーザ評価実験

生活支援ロボット（図1）を用いてジェスチャ認識システムのユーザ評価実験を行った。ロボットにはKinect、ノート型PC、小型液晶ディスプレイが搭載されている。



図1 生活支援ロボット

ジェスチャ認識システムを、ノートPCで動作させ、小型ディスプレイにシステムの状態や認識結果を表示した。ロボットについて何も知らない初心者11人が、ロボットから2m離れた位置に座り、4種類のジェスチャを行った。

3-1 実験手順

11人のうち6人の実験は、以下のように行われた。まず、4分以内のジェスチャ方法の説明ビデオをみせた。次に、ジェスチャの練習を

2分間行わせた。練習終了後、ユーザへ質疑応答と失敗したジェスチャへのアドバイスをを行った（後述）。

その後、4種類のジェスチャをまとめて3回ずつ行わせ、質問紙に記入させた。（後述）残りの5人の実験手順も同様であるが、以下の点が異なる。

- ・練習時間を限定せず、各ジェスチャが正しくできることが確認されるまで練習させた点
- ・練習中、適宜アドバイスをを行った点
- ・テスト中、各ジェスチャを1回ずつ区切って行わせた点

実験手順を変更した理由は以下である。

- ・1つのジェスチャが成功するまで何度も練習し、練習時間が無くなる事があった
- ・ジェスチャが失敗した理由を知らせ、正しいジェスチャを練習する機会が必要と考えた
- ・ジェスチャが検出されない場合、手を戻す動きをジェスチャと判定し、ユーザが混乱した

3-2 ジェスチャのアドバイス

前述の6名に対しては、ジェスチャの練習後に、5種類のアドバイスをを行った。それぞれ、「画面にReadyの文字が出てからジェスチャを行う事」、「Readyが表示されない時は手を右上に動かし止める事」、「ジェスチャ終了後、すぐに手を下す事」、「手を速く動かす事」、「手を真っ直ぐ動かす事」である。別の5名に対しては、練習後ではなく、練習中に上述のアドバイスを適宜行った。特に、Ready表示についてのアドバイスは全員に徹底して行った。

3-3 質問紙の内容

質問項目は以下の通りである。

- 1) 4種類のジェスチャの使い方は分かりましたか？ 使い方が分からなかったジェスチャはどれですか？（はい いいえ どちらともいえない）
- 2) 4種類のジェスチャのできるタイミングは分かりましたか？ タイミングが分からなかったジェスチャはどれですか？（はい いいえ）
- 3) 練習後のテストで、ジェスチャに反応しないことはありましたか？ 反応しなかったジェスチャと回数を教えてください（はい いいえ）
- 4) 練習後のテストで、行ったジェスチャが別のジェスチャと認識されることはありましたか？ その種類、認識結果と回数を記入してください（はい いいえ）

- 5) 練習後のテストで、何もしていないのにジェスチャを認識する事がありましたか? 回数を記入してください (はい いいえ)
- 6) 4種類のジェスチャが出来るようになりましたか? できるようになっていないジェスチャを教えてください (はい いいえ)
- 7) 画面の表示は分かりやすかったですか? どのようなところが分かりにくかったですか? (わかりやすい どちらかといえばわかりやすい どちらともいえない あまりわかりやすくない わかりにくい)
- 8) やりにくかった点、お気づきの点があればご記入ください

4 結果

ジェスチャ認識システムは、初心者11名の正しいジェスチャを100%認識した。また、「手を速く動かす事」や「手を真っ直ぐに動かす事」のアドバイスを受けたユーザは全員正しくジェスチャを行った。しかし、「画面にReadyの文字が出てからジェスチャを行う事」、「ジェスチャ終了後、すぐに手を下す事」はその通りに動かさない事もあった。前述の6名のうち一人は、これらのアドバイスなしで正しくジェスチャを行った。残りの5名にはアドバイスをし、3名がその通りに行った。テスト後、アドバイスに通りに動かさなかったユーザに理由を尋ねたところ、「反射的に動かしてしまった」と答えた。前述の後者5名に対しては、全員にアドバイスし、4名がその通りに動かした。また、ユーザが手を顔の横に出した時にロボットがジェスチャと判定し、ユーザが混乱して手を前に出す事があった。

質問紙の1つ目と6つ目の質問には、11名全員が「はい」と答えた。8つ目の項目で、「ジェスチャを行うタイミングが分かりにくい」と答えた初心者が全体で2名いた。そのうちの1名は最も長く練習し、4分以上行った。

5 考察

実験結果から、初心者ユーザと生活支援ロボットのマルチモーダル対話に手の3次元位置によるジェスチャは効果的であることが分かる。初心者全員が正しくジェスチャを行い、簡単なアドバイスを与えられた後、短時間でジェスチャを覚えた。また、質問に対して、全員がジェスチャの使い方を理解し、出来たと答えた。初心者が正しいジェスチャを覚えた後、意図的に手を動かした時、その意図をロボットは100%認識した。ジェスチャを行う前や検出されなかった時など、ユーザが意図していない手の動きを、システムがジェスチ

ャと判断する事はあったが、音声と組み合わせる命令を行う際には問題にならない。例えば、「下がれ」という音声と手を前に出すジェスチャが近い時間に検出されたときのみ、命令と判断すればよい。一度ジェスチャを検出してから2秒間検出しなかったのは、ジェスチャ認識のみで意図伝達を行う実験のためであり、音声認識を組み合わせる場合は、常時ジェスチャの候補としてイベントを検出すればよい。Ready表示がないときにジェスチャを行ったユーザは、ジェスチャ検出画面などをReadyの画面と混同した可能性がある。より分かりやすい画面表示や効果音の利用によって、このような混同は避けられるはずである。

6 おわりに

本研究では、マルチモーダル対話のためのジェスチャ認識システムを作成し、ユーザ評価を行った。その結果、ユーザ全員が短時間で正しくジェスチャを行い、システムはジェスチャを正しく認識した。以上から、生活支援ロボットとのマルチモーダル対話に手の3次元位置によるジェスチャ認識が効果的であることが分かった。今後、音声認識とジェスチャ認識を組み合わせ、マルチモーダル対話を実現する。

参考文献

- [1] Tetsushi Oka, "Multimodal Command Language to Direct Home-use Robots", *Advances in Human-Robot Interaction, INTECH*, pp.221-232, December 2009
- [2] Tetsushi Oka, Kaoru Sugita and Masao Yokota, "Commanding a humanoid to move objects in a multimodal language", *Artificial Life and Robotics, Vol.15, No.1*, pp.17-20, 2010
- [3] 岡哲資 松本大賢 木林龍一: 音声とタッチディスプレイによるロボットの物体移動の命令方法, 日本機械学会ロボティクスメカトロニクス講演会2011講演論文集, 2011