

指差しジェスチャと音声指示によるロボットの制御手法

篠塚 晃希[†] 田村 仁[‡]日本工業大学 工学部[†] 日本工業大学 先進工学部[‡]

1. はじめに

近年、自立稼働するロボットの開発が進んでおり、人との共働をするロボットへの関心も高まっている。人と共働する際に起こりうる、変化する状況に対応する上で、その場で与えられた指示をもとに判断を行う機能は重要である。

指示を与える方法として、ジェスチャなど非言語でのコミュニケーションを用いることができれば、コントローラを用いるより多くの場面での活躍が期待できると考える。

ジェスチャでロボットを操作する研究として[1], [2]などがある。ロボットが人間、及びジェスチャの認識を行う際、ステレオカメラや赤外線カメラが使用されている。

近年では、単画像から2次元の骨格を推定する方法[3]が存在する。また、単眼カメラでの3次元座標の推定も様々に試みられているが、どれも安定した性能を示すには至っていない。本研究では、手首、肩などの2次元座標から指差し方向の3次元座標を簡易的に推定、加えて音声での補助によってロボットへの指示とする手法を提案する。

2. 提案手法

本研究では、対象の人物が特定地点へ指さしを行い、対面したロボットが撮影した画像からその人物の関節の座標を取得し、その情報をもとに指をさしている地点を推定し、その地点まで移動することを目的とする。

ロボットはJetson nanoを搭載したJetbot[4]を使用し、カメラには単眼カメラを用いる。人物の姿勢や関節の位置の認識にはJetson nano上で使用できるtf-pose-estimation[5]を用いる。

図1に示す通り、頭、首、右肩、右肘、右手首、右腰、両足首の関節座標を得ることで、現在の腕の端から端までの長さ、本来の腕の長さ、人物の縮尺などを求める。それらの情報をもとに、指している場所をロボットの走行面上の座標として求

め、その座標まで移動を行う。その際発生した誤差を、「右」、「前」のような特定の発話音声によって修正する。

推定座標を $(X1, Y1)$ とし、ロボットと人の間の距離を L_p 、画像の全高を h 、図1のそれぞれの関節の座標を (X_i, Y_i) としたとき、

$$X1 = (-(-((h - Y_2) - (h - Y_4))/(X_2 - X_4)) + (h - Y_2))/((h - Y_2) - (h - Y_4)/(X_2 - X_4)) * L/2$$

$$Y1 = L_p + \left(\frac{-L_p \times \left(\frac{A_z}{\sqrt{A_e^2 - A_z^2}} \right) + (h - Y_2)}{\left(\frac{A_z}{\sqrt{A_e^2 - A_z^2}} \right)} \right) \times L/5$$

として計算を行った。

画像内の頭から足までの距離 h_1

$$h_1 = \sqrt{\left(X_0 - \frac{X_{10} + X_{13}}{2} \right)^2 + \left((h - Y_0) - \frac{(h - Y_{10}) + (h - Y_{13})}{2} \right)^2}$$

本来の身長を今回1.65[m]と固定して h_1 で割った数を L として、

$$L = 1.65/h_1$$

肩から腰までの長さを A_e とする。

$$A_e = \sqrt{(X_1 - X_8)^2 + ((h - Y_1) - (h - Y_8))^2}$$

腕から手首までの距離 A_z を求める。

$$A_z = \sqrt{(X_2 - X_4)^2 + ((h - Y_2) - (h - Y_4))^2}$$

腕を地面から垂直にしたときの2, 4及び1, 8間の長さが等しいものとする。

Robot control methods with pointing gestures and voice instructions

[†]Kouki Shinotsuka,

[‡]Tamura Hitoshi

[‡]Nippon Institute of Technology



図 1 各関節の位置と番号

3. 実験方法

現段階では、音声認識とロボットの移動制御が未実装であるため、単眼カメラでの指差し位置の推定精度の確認を行った。ロボットと人物の距離を 1.5[m]とし、ロボットの前 1.0[m]の空間に図 2 に示す通り、0.5[m]ごとにポイントを設置した。対象人物が肘関節を伸ばした状態でそのポイントにまっすぐ腕を伸ばしレーザーポインタを当てた時のロボットの推定地点との差を評価する。

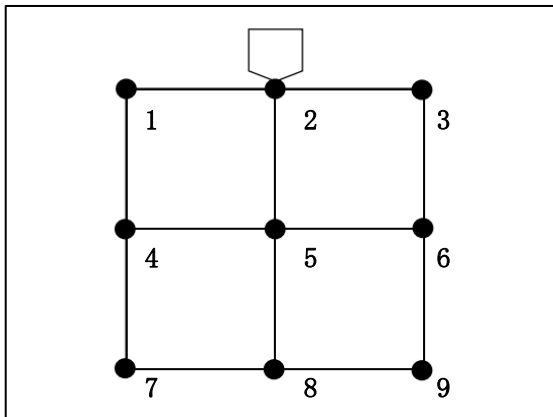


図 2 ポイントの配置

4. 実験結果

この実験にて得られた座標と、実際に指をさしていた座標を比較した表を表 1 に示す。表 1 の推定座標の値は、一つのポイントに指をさしている 50 フレーム分の画像の平均値を取っている。

X 軸, Y 軸ともに、位置の変化に合わせて座標が変化していることが分かった。

誤差は 1.0[m]以内に収まることが確認できた。

X 軸は、対象人物に近い位置になるほど誤差が小さくなることが分かった。

Y 軸は、ロボットと対象人物の軸線上のポイント 2, 5, 8 では誤差に大きな差はなく、それ以外のポイントでは、対象人物に近い位置になるほど誤差が小さくなった。

表 1 実験結果

ポイント	座標		推定値	
	X[m]	Y[m]	X[m]	Y[m]
1	0.500	0.00	0.937	0.560
2	0.00	0.00	-0.354	-0.210
3	-0.500	0.00	-0.991	0.907
4	0.500	0.500	0.565	0.829
5	0.00	0.500	0.1160	0.711
6	-0.500	0.500	-0.833	0.959
7	0.500	1.000	0.479	1.169
8	0.00	1.000	0.106	1.191
9	-0.500	1.000	-0.443	1.293

5. 考察

機体と人の軸線上から外れたポイントは、人から離れるほど誤差がひどくなっていった。

ポイント 1, 4, 7 を指差しした際に腕が体の前方に来るため腕関節の座標が認識できなくなることがあった。認識率を上げることができれば、より精度を上げることができるのではないかと考える。

6. おわりに

本研究では、ロボットへ指差しのジェスチャによって指をさした先へ移動する指示を行う事を目的とし、単眼カメラによる 3 次元座標の簡易的な推定の精度の確認を行った結果、現段階では実用的ではないものの単眼カメラで指差し指示の大きな 3 次元座標を推定する事ができた。今後は、精度の向上や走行機能、音声の補助の実装も含めた総合的な検証を行いたい。

参考文献

- [1] 大久保 康基, 岡田 慧, 稲邑 哲也, 稲葉 雅幸, 2A1-S-049 日常生活におけるロボットへの対象指示のための指差し認識に関する研究(ヒューマノイド3, 生活を支援するロボメカ技術のメカインテグレーション), ロボティクス・メカトロニクス講演会講演概要集, 2005, 2005 巻, pp. 174-178,
- [2] 山口 亨, 櫻井 翔一郎, 佐藤 英理, 指差し動作による人間-ロボット間のナチュラルインターフェース, 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集, 2006, 22 巻, 第 22 回ファジィ システム シンポジウム, pp. 639 - 642
- [3] CMU-Perceptual-Computing-Lab
/openpose, <<https://github.com/CMU-Perceptual-Computing-Lab/openpose>>2021 年 1 月 8 日アクセス
- [4] NANO JETBOT KIT - AN EDUCATIONAL AI ROBOT BASED ON NVIDIA JETSON NANO., <<https://www.siliconhighwaydirect.com/product-p/jetbot-kit.htm>>2021 年 1 月 8 日アクセス
- [5] ildoonet/tf-pose-estimation, <<https://github.com/ildoonet/tf-pose-estimation>>2021 年 1 月 8 日アクセス