

ハンドジェスチャによる 3 次元ポインティングを用いたロボット制御手法 A Robot Control Method Using 3D Pointing with Hand Gestures

篠塚 晃希[†] 田村 仁[‡]
Kouki Shinotsuka Hitoshi Tamura

1. はじめに

近年、人手不足やコロナの対策を理由にロボットが一部の作業を代替するといった形で共働を行う事例がある。しかし、現在実用化されているこのようなロボットの多くは、与える指示がタッチパネル等を用いての決まった形でしか行えない。

人が働く現場では、常に状況が変化する。繰り返して同じ作業を行うだけでなく様々な業務を柔軟にこなせるロボットであれば、現在は人にしかできない作業等に協力できるようになると期待できる。そのためにロボットは高い判断力を持つ他に常に人間からの指示を受けることのできるインターフェースが必要であり、指示を与える方法にタッチパネルやコントローラなど、指示者自身の作業を大きく阻害するものを用いることは望ましくないと考える。指示者の負担の低減や、誰もが比較的簡単にロボットに指示を与えることができる方法の 1 つとして、ジェスチャによりロボットへ指示を与える方法がある。その中に、指差しによるジェスチャがある。物や場所を指し示すために使われるジェスチャで、感覚的に使うことができる。

本研究では、小型ロボットにジェスチャのうち指差しにより指示を与えることで、指している先に移動するロボットの実現性について検証することを目的とし、特に指差し地点の認識の精度について確認する。

2. 関連研究

ジェスチャでロボットを操作する研究として、[1], [2] などが挙げられるが、人間、及びジェスチャの認識をステレオカメラや赤外線カメラ等を用いて行われている。

近年では[3]などの単画像から 2 次元の骨格座標を推定する方法も存在する。また、単眼カメラでの 3 次元の骨格座標の推定も様々な試みられているが、実用化に至るほどの正確性を持った方法は存在しない。単眼カメラでの指差し位置の推定を行う研究として[4]が挙げられるが、指差しの基準を目の間と指先の直線上として定めている。

3. 提案手法

小売店など、比較的小規模な店内において商品などを運搬する際に指差しによる指示で移動することのできるロボットを想定するため、対象の人物が特定地点へ指差しを行い、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、その情報をもとに簡易的に一部骨格の 3 次元座標を推定し、指をさしている地点を予想してその場

まで移動することを目的とする。

本研究ではこのうちカメラで撮影した画像をもとに指をさした先の位置を推定する方法について研究する。

移動可能なロボットである必要があるため、コストが低く、ロボットの処理の負担を抑えることができることから単眼カメラを使用する。

ロボットは人と同じ平面上に存在し、その面を地面として考えたとき、地点の指定は右肩と手首の関節座標を通る直線と地面の交点とする。関節座標は表 1 に示す通り ID に対応した座標を取得することができる。

ロボットには GPU を搭載したマイコンを使用し、ロボットに搭載した状態のカメラで対象人物を撮影するが、撮影画像は球状に歪んでいるため、画像にキャリブレーションを行い平面になるよう補正する。その画像をもとに人物の姿勢や関節の位置の検出を行い、図 1 に示すように、体の関節座標をそれぞれ得ることで、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(XI, YI)として推定し、その座標まで移動を行う。座標の計算等については 3.1 で述べる。

本研究では、対象の人物が特定地点へ指差しを行い、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、得られた関節座標をもとに人物の 3D 座標を推定、人物の立っている地面を原点とした

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 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。

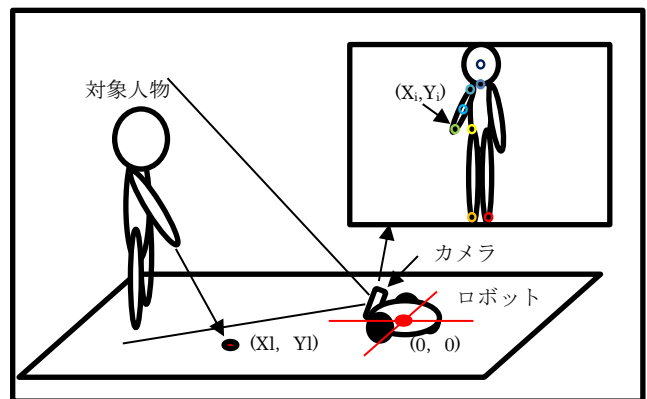


図 1 指示のイメージ

3.1 座標の推定方法

予想地点の座標はロボットの位置を基準として(XI, YI)として求める。XI, YI はロボットの置かれている平面上に限定し、カメラから得られる画像の下辺も同様に平面上に

[†] 日本工業大学 機械システム工学専攻,
Nippon Institute of Technology Department of Mechanical
Systems Engineering

[‡] 日本工業大学 先進工学部 ロボティクス学科,
Nippon Institute of Technology Department of Advanced
Engineering Faculty of Robotics

あると想定する．関節座標 tf-pose-estimation により，表 1 に示す通り ID に対応した各関節の(x, y)座標を取得する事ができる．tf-pose-estimation の説明は 3.1 で行う．

X 座標は，カメラが撮影した画像の中心の x 座標を零としたときの，肩と手首の関節座標上を通る直線と画像下辺との交点の x の値とする．

Y 座標は，ID2 の z 座標を零としたとき，画像から腕の長さ ID2，4 間の距離を求め，ID1，8 の長さと比較し，ID2，4 の画像内の z 座標を求め，座標(z, y)の ID2，4 を通る直線と画像底面との交点の z の値とする．

この時，腕を真下に自然体で下ろしたときの，ID2，4 の長さ，直立時の ID1，8 の長さは，画像からは等しく見えているものとする．また，ID2 を z 座標の零点とする．

指示の方向は地面の特定地点を指さすことから，腕の向きを水平方向から下方 180° の範囲に限定するが，カメラに対して人が腕を前後に 180 度させる場合，求まった Y 座標に対して正負 2 通りの状態が存在してしまうため，本研究では体の後方 90 度を対象外とする．

ロボットと人との距離を L_p ，画像の全高を h ，図 1 のそれぞれの関節の座標を表 1 の ID を i として， (X_i, Y_i) とする．画像内の頭から足までの距離 h_1 を求める．

$$h_1 = \sqrt{(X_0 - \frac{X_{10} + X_{13}}{2})^2 + ((h - Y_0) - \frac{(h - Y_{10}) + (h - Y_{13})}{2})^2} \quad (式 1)$$

画像内の距離を現実の距離に変換するため，身長を元にした足首から頭までの高さを定数 h_h とし， h_1 で割った数として L を求める．

$$L = h_h / h_1 \quad (式 2)$$

首から腰までの長さを本来の腕の長さ A_e として求める．

$$A_e = \sqrt{(X_1 - X_8)^2 + ((h - Y_1) - (h - Y_8))^2} \quad (式 3)$$

肩から手首までの距離を，現在の腕の長さ A_z として求める．

$$A_z = \sqrt{(X_2 - X_4)^2 + ((h - Y_2) - (h - Y_4))^2} \quad (式 4)$$

これらの式より X_l 及び Y_l を，

$$X_l = \left(-\frac{-(\frac{(h - Y_2) - (h - Y_4)}{X_2 - X_4}) + (h - Y_2)}{(h - Y_2) - \frac{h - Y_4}{X_2 - X_4}} \right) * L / 2 \quad (式 5)$$

$$Y_l = L_p + \left(L_p - \frac{h - Y_2}{\frac{A_z}{\sqrt{A_e^2 - A_z^2}}} \right) * L / 5 \quad (式 6)$$

として求める．

ここまでを先行研究[6]により行ってきたが， L_p を一定の距離として固定して実験を行っていたため，自由な位置から指示を行うことができなかった．そのため，本研究では，ロボットと人との距離である L_p を固定せず，指示者の距離を変えることができるような機能の追加を行うための検

証を行う．これにより，決められた範囲内であれば任意の位置から指示を行うことのできる機能の実装を目的とする．

L_p の推定には，カメラ画像に映る人物の画像上の身長と，距離 a [mm]の位置にて取得した画像上の身長 l_i との割合から求める．

$$L_p = l_i / l_h * a \quad (式 7)$$

今回の a と l_i は，自身の身長をもとに測定を行い，決定した．

表 1 関節と ID の関係

ID	関節
0	頭
1	首
2	右肩
3	右肘
4	右手首
8	右腰
10	右足首
13	左足首

3.1 tf-pose-estimation

OpenPose の tf-pose-estimation[5]という機械学習モデルが存在する．OpenPose の機能が TensorFlow で実装されており，CPU や組み込みデバイスなど性能が高い GPU を使用せずに動作させることができるため JetsonNano に実装して使用する．

本研究では，それぞれの特徴点の X, Y 座標をポイント ID として取得できる機能を利用して，右肩，右手首，首，右腰，頭，右足首，左足首の関節座標を図 2 のように取得する．それぞれの座標との対応は表 1 に示した通りである．



図 2 tf-pose-estimation の取得画像[6]より引用

4. 実験

カメラに全身が写る範囲内で任意の位置から指示を行うためには 500[mm]単位の大まかな位置の推定である場合でも指差し位置の推定の時点でも誤差が発生するため、 L_p はできるだけ誤差が少なければならない。そのため、指差し地点推定の機能と統合しての評価の前に、提案手法の誤差の程度を確認するため、Jetson Nano と単眼カメラを用いての測定を行い誤差の評価を行った。

測定は、ロボットの搭載するカメラから、ポイント A の 1500[mm]、ポイント B の 2000[mm]、ポイント C の 2500[mm]のそれぞれ 3 地点の距離に人物を直立させた状態で撮影を行い、得られた 100 フレーム分の画像上の身長から平均値をとり、それを式 7 の l_h に代入する形で計算を行い、正しい距離との比較を行った。また、本実験で使用するデータは 4 人分のものである。

l_i は表 2 の 4 番の人物の 1500[mm]にて取得した画像上の身長のデータを用いている。

表 3 距離ごとの画像上の身長の値

人物	ポイントA	ポイントB	ポイントC
1	205.1	164.0	139.4
2	186.9	145.0	121.6
3	206.6	165.7	139.8
4	205.6	165.6	136.7

5. 結果

結果は表 3 の通りとなった。表 3 は、人物ごとにそれぞれの距離に対して推定を行った結果を表している。1500[mm]、2000[mm]、2500[mm]、はそれぞれ測定した際のロボットと人との距離である。

それぞれの距離は 500[mm]離れているが、少なくとも 300[mm]は離れていることが検知できていた。

人物ごとによる距離のばらつきは、人によりほとんど存在しない場合もあれば、200[mm]程存在する場合もあることが確認できた。

4 番の人物の 1500[mm]の値をもとに計算をしたにもかかわらず、4 番の 2000[mm]で 138.9[mm]、2500[mm]では 244.5[mm]の誤差が発生してしまっていた。

全体的には距離の変化に対して致命的な誤差は発生しておらず、大体ではあるが位置の推定ができていたことが分かった。

表 4 距離の推定結果

人物	1500[mm]	2000[mm]	2500[mm]
1	1503.2	1880.1	2211.7
2	1649.8	2127.3	2535.7
3	1492.8	1861.0	2205.9
4	1500.0	1861.9	2255.5

6. 考察

6.1 実験の考察

1500[mm]こそ、ほとんどの人物の位置の推定が数ミリメートル単位でできていたが、距離が離れると誤差が非常に大きくなった。また、1500[mm]地点の値を l_i に使用した 4 番の人物の他の地点の推定結果に大きな誤差が見られた。これは、画像上の身長に対しての距離との関係が式 2 で正しく表せていなかったからであると考えられる。

人物ごとによる推定距離のばらつきに、個人差があることが挙げられるが、本実験では 2020 年の成人男性の平均値である 172[cm]に最も近い 171[cm]の人物である 4 番の値を使用して推定を行っていたため、身長が大きく異なる人物にも対応できるような方法に変更する必要があると考える。

2 番の人物の推定結果のみ、他の人物から大きく異なる値を示している。また、表 4 は、実験に使用した 100 フレーム分の画像上の身長の標準偏差である。偏差の少ない箇所は 1 程度なのに対し大きいところでは 30 以上もある。これらは、関節座標検出に用いている tf-pose-estimation が服装や体格などの要因によって値のばらつきが起きてしまっていると考えられる。

表 2 標準偏差

人物	ポイントA	ポイントB	ポイントC
1	10.22	2.799	1.504
2	34.29	13.20	30.16
3	11.21	4.693	3.205
4	1.860	3.327	1.359

6.2 先行研究の結果との考察

先行研究[6]では、ロボットと人物の距離が 1500mm のみでしか測定できなかったことに対して、本研究では 1500mm から 2500mm の間の任意の地点で測定することが可能になった。

しかし、距離の推定の誤差が表 3 の通りであるため、表 5 に表す先行研究の結果より本研究の手法で指差し地点の推定を行った場合、距離のずれはそのまま位置のずれにつながる。

例として表 3 の場合、3 番の人物が 2500[mm]で指差し指示を行ったとき、先行研究の手法より本研究の手法では 295[mm]の誤差が生じることとなる。

全体で見ても数百ミリメートル単位での精度の悪化が考えられるため、本手法のまま利用することができない。しかし、任意の地点から指差し指示を送る事ができる利点は大きいため、精度の改善が必須である。

人物までの距離推定の精度の改善の他、別の方法での精度向上も必要であることや、指差し位置の精度の問題を解決するため、2 次元の関節座標を 3 次元の関節座標に変換する方法の変更を検討する。

現在検討している方法として, A simple yet effective baseline for 3d human pose estimation [7]の手法がある.

表 5 指差し地点と推定結果[6]より引用

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

- [6] 篠塚 晃希, 田村 仁, “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”, 情報処理学会第 83 回全国大会 講演論文集, 2021, 分冊 4 pp143 – 144
- [7] Julieta Martinez, Rayat Hossain, Javier Romero, James J. Little” A simple yet effective baseline for 3d human pose estimation”, 2017, 1705.03098, arXiv,

6.3 軽量の 3 次元姿勢推定手法

A simple yet effective baseline for 3d human pose estimation[7]は, 2 次元の姿勢のデータから 3 次元の姿勢を推定する手法の一つで, 推定に画像を使用せず 2 次元の関節座標をもとに行うため, 軽量で高速に計算する事ができる. この手法による代替で誤差を少なくすることが可能だと考えたため, 今後 Jetson Nano 上で動作するかの実験, 及び先行研究との誤差の比較を行い, 指差し地点の推定精度の改善を行う予定である.

7. おわりに

本研究では, 非言語による人とロボットの意思疎通を目的としてジェスチャのうち特に指差しによる指示で目的地まで移動するロボットの研究を行うことを目的とし, 先行研究から人物の位置を問わず指示を受けられるよう改良を行った. 実際の距離と推定結果の比較を行い, 誤差の有無及び程度の大きさについての検証を行った. 結果は表 3 の通り数百ミリメートル単位の誤差が出てしまっており, 現時点でのこの誤差と指差し位置の推定の際に発生する誤差が合わさってしまうと機能の実現は難しいため, さらなる改善が必要である.

また, 先行研究からの問題であった指差し位置の推定精度の改善を手法の改善, 音声認識での補助機能の追加, 及び移動機能の実装を行うことで, 今後一連の動作を行う実験を行うことを目指す.

参考文献

- [1] 大久保 康基, 岡田 慧, 稲邑 哲也, 稲葉 雅幸, “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”, 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>”, (参照 2021 年 1 月 8 日)
- [4] 木村 理恵子, 阪本 邦夫, “単眼カメラ画像からの指先位置計測の一手法”, 島根大学総合理工学部紀要. シリーズ A 38 , 2004-12-15, pp. 33 – 38,
- [5] ildoonet/tf-pose-estimation,”<https://github.com/ildoonet/tf-pose-estimation>” (参照 2021 年 1 月 8 日)