

ロボット制御のための指差しジェスチャの解析

Analysis of pointing gestures for robot control

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、今まで行ってきた作業の一部を代替するといった、人と共働を行うロボットに対しての注目が集まっている。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場面もあると考えられる。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には、従来から用いられるコントローラや、音声指示、ダイレクトティーチングの他に、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。

ジェスチャの認識には、ロボット側のカメラ等を用いる方式や、操作者がグローブなど体に装着する形でセンサを取り付けて行う方法などがあるが、本研究では、ある程度利用環境が制限されるが、個人に合わせて調整を細かく行う必要のない、ロボットに搭載する方法での研究を行う。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の広角カメラ(図 1)を使用する。



図 1 広角カメラ

[†] 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

2. 関連研究

360 度画像から人物のジェスチャを認識する手法として、[1], [2], [3]などが挙げられる。このような広角レンズを用いたカメラから認識を行う際の問題として、[4]で提案されるような、ミラーを用いたカメラと比較して推定に必要な部分の画素数が少なくなってしまうことや、360 度画像のキャリブレーション及び展開の困難さが挙げられる。そこで本手法では、取得した関節座標をもとに機械学習を行った判別器を用いることで指差し地点推定を行う手法を提案する。機械学習を用いることで最低限のキャリブレーションで多様な撮影条件への対応を可能とする。

3. 提案手法

小型の掃除用ロボットへ掃除の範囲の指定ができる機能を想定した実現を目指し、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が指をさしている地点を推定する方法として機械学習を用いる。

本研究では機械学習により指差し地点を推定することを目指す。そのために考える方法として、大きく 2 つを考える。1 つ目は画像認識によって画像から指示を行っている人物の関節の 2 次元座標を取得し、それを入力として扱う。2 つ目は、360 度画像を入力としたものである。出力はロボットの中心が原点となるような平面の座標を、正解ラベルとして出力されるよう予測器を作成する。関節座標の取得は、OpenPose^[5]を用いて行う。

本実験では、OpenPose を用いた方法と、用いない方法の 2 種類のモデルを作成し、計算時間がどの程度に収まるのか、及び推定地点の精度を比較、評価する。一連の流れは図 2 の通りである。

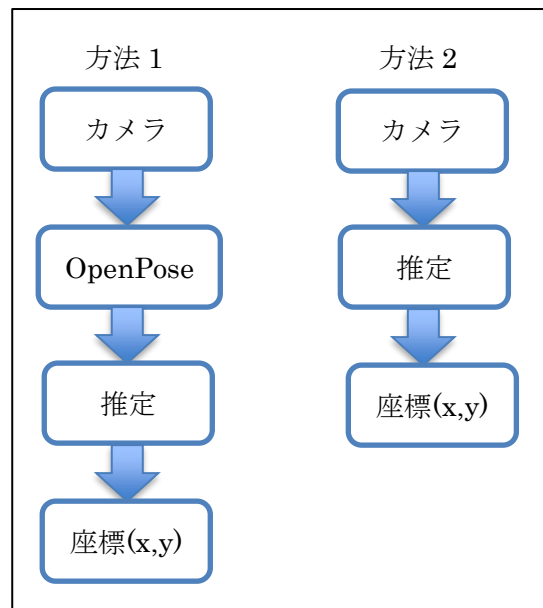


図 2 推定までの流れ

3.1 関節座標を用いる方法

1 つ目は、OpenPose によって取得した関節座標を入力としたモデルである。OpenPose を用いることで、事前学習に必要なデータセットの負担を減らす狙いがある。図 3 のように対象人物が任意の地点に指差ししているとき、図 4 のように関節ごとにカメラから見た 2 次元の座標を取得する事ができる。このように取得した関節座標を入力座標として判別を行うことで、指差しの判別自体には服装や背景などの要素を取り除いて判別を行うことができると考える。本研究では 360 度画像を用いるため、図 4 のような単眼カメラの画像と異なり、写る人物の大きさや歪み方、人物の関節の位置関係が同じ姿勢でもロボットの向き等によって変わってしまうため、それによって判別に与える影響についても考慮する必要がある。

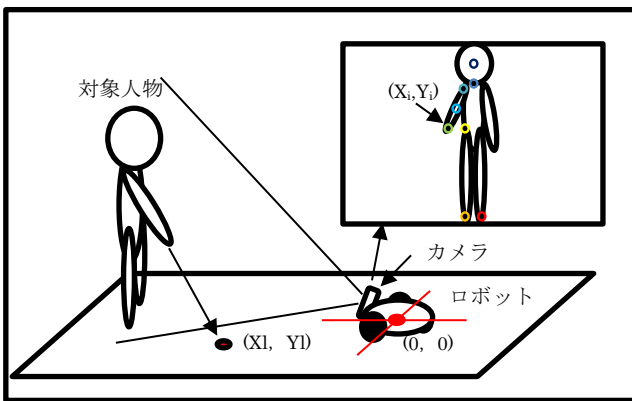


図 3 指示のイメージ

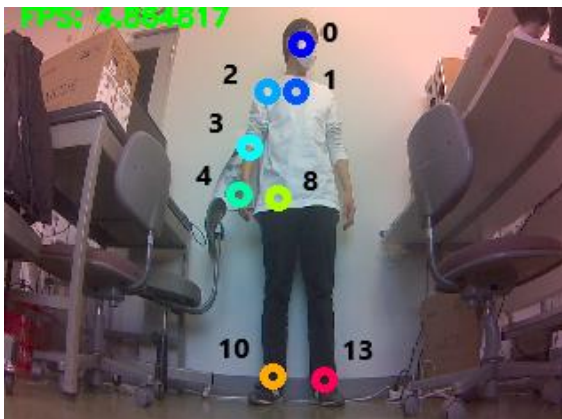


図 4 関節座標取得のイメージ

3.2 360 度画像をそのまま用いる方法

2 つ目は、図 5 のようなカメラから取得した 360 度画像を入力として、推定座標を出力とする判別を行うものである。ネットワークモデルには ResNet¹⁰⁾を用いる。

ResNet とは、Residual block を用いることで大幅なレイヤー数の増加が見込めるうえに、エラー率の低減を実現することができることから、近年画像の分類に多用されている手法の 1 つである。



図 5 広角カメラの画像

4.2 種類の方法の検討

データセットは、主に画像、もしくは関節座標のデータと指差し地点の正解座標を合わせたものを用いる。方法 1 は、OpenPose によって取得した図 3 のような関節の座標と方向のデータを組み合わせる。方法 2 は、広角カメラによって撮影した図 4 に示すような、指示をしている人間が写った画像を用いて学習、判別を行う。

関節座標を用いる方法と、360 度画像をそのまま用いる方法の 2 種類の方法を検討した時、関節座標を用いる方法の利点としては、人物の検出及び関節座標の推定の部分を既存の OpenPose で行うことによって、学習に必要な枚数を減らすことができると考えられることである。また、関節の座標データを入力として用いるため、画像そのものを入力とする 360 度画像をそのまま用いる方法と比較しネットワークモデルの構築が容易であると考えられる。欠点としては、本研究で用いる OpenPose は通常の RGB 画像での人物検出に用いるためのものであるため、広角カメラから得る 360 度画像に映る人物の検出に不安があると考えられる。また、一連の流れとして、2 つのモデルを稼働させることになり、リアルタイム性への影響が大きくなる可能性がある。

対して 360 度画像をそのまま用いる方法では、利点として、360 度画像からの人物検出を OpenPose に頼る方法よりも、精度の向上を期待する事ができる。また、関節座標を用いる方法よりも、計算時間を少なく行うことができるはずである。欠点としては、目的に完全に沿う形 of データセットを見つけることができなかつたことから、1,000 枚から 10,000 枚程度のデータセットを自身での用意が必要になると想定されることである。

5. データセット

関節座標を用いる方法のデータセットの準備として、関節座標と指差ししている場所の組み合わせができるようなデータセットが必要になる。そのため、カメラを搭載するロボットの周囲から様々な地点に指を指している画像から得た関節座標のデータが必要になる。これには関節座標を用いるため、服装や背景などの環境はある程度無視できると考えられる。このことから、データセットの総数を 100 枚から 1,000 枚程度を想定して準備する。

360 度画像をそのまま用いる方法のデータセットの準備としては、広角カメラから取得する 360 度画像と指を指している場所の組み合わせができるようなデータセットが必要になる。このことから、データセットの総数を 1,000 枚から 10,000 枚程度を想定して準備を行う。

ロボットにはメインボードに Jetson Xavier NX を搭載することを想定し、人物の姿勢や関節の位置の認識には Jetson Xavier NX 上でも使用できる trt-pose^[7]を用いる。

5.1 trt-pose

OpenPose を小型コンピュータボード上で動かすことを想定し、動作が軽量な特徴がある trt-pose という機械学習モデルを実験にて使用する。OpenPose の機能が TensorFlow で実装されており、trt を使用することで、CPU や組み込みデバイスなど性能が高い GPU を使用しない機器でも比較的高速で動作させることができるため Jetson Xavier NX に実装して使用する。

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 4 のように取得する。それぞれの座標との対応は表 1 に示す。

表 1 関節座標と 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

6. 実験予定

現在データセットの準備等が完了していないため、実験が行えていない。そのため、今後行う予定の実験について述べる。

まず、関節座標を用いる方法と 360 度画像をそのまま用いる方法の利点欠点の比較を行い、どちらの方法を用いるべきかの検討を行う。検討は、必要な計算時間及び精度の 2 点について行う。

行う実験内容としては、任意の地点に立った人物が指差しを行い、計測を行うカメラが設置されたロボットから見てどの方向に指を指しているかを推定できるかを第一段階とし、次に、どの範囲を指しているかの推定が可能かについての検討を、関節座標を用いる方法と 360 度画像をそのまま用いる方法の 2 通りの方法で行う。これを数人で行ったときの関節座標を用いる方法、360 度画像をそのまま用いる方法それぞれの結果から、どちらの方が妥当であるか

の検討を行う。図 6 は上から見たものである。具体的には、カメラの周囲の任意の地点に立った人が指差しを行う。判別の方法はまずロボットから見た方向のうち、360 度を 12 等分した簡易的な方向推定とする。このため、1 方向に対して様々な位置から指を指したときのデータセットが必要になる。

次に、誤認識や推定の誤差が許容範囲内に収まるかの確認を行う。本実験における許容範囲は、方向の推定では、正解の方向に対して隣接している方向までに収まっているか。範囲の推定では、指している方向と真逆ではないことを許容範囲と定める。この許容範囲の確認に加えて、他の手法との比較も行うべきであると考えられる。

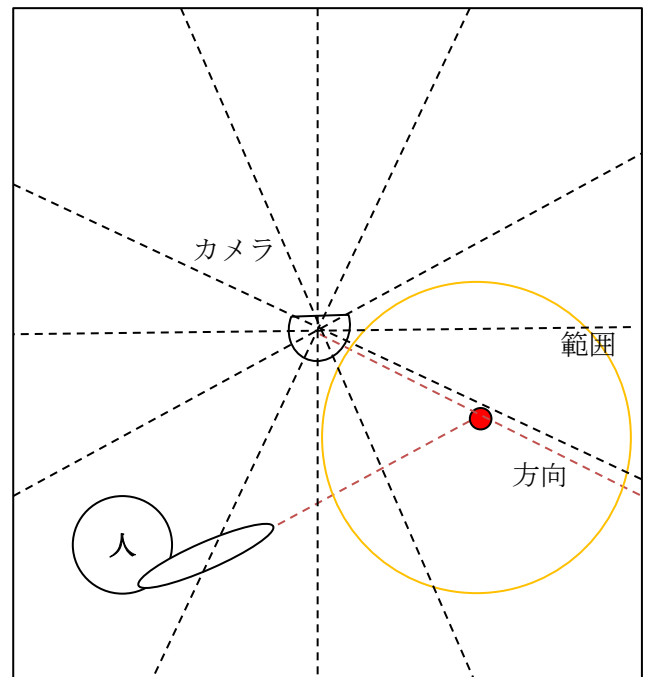


図 6 実験概要

7. まとめ

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、機械学習によって全方位から指差し指示が可能な方法についての研究を行っている。本稿執筆時点でデータセットの収集が完了しておらず、実験ができていないため発表時までに実験を完了する予定である。

ジェスチャを単眼の広角カメラからの画像で認識する方法のみでの指示では、誤差などの要因によって指示できる部分に制約や限界が存在する。そのため、今後は指示のジェスチャの認識に加えて、音声による補助も加えるなど複数の方法を用いての指示が可能な機能の実装が必要であると考えられるため、音声による補助についての実装を目指したいと考える。

また、先行研究からの課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、及び移動機能を図 7 に示す全方位に移動可能な移動体の実装

することで、今後一連の動作を行う実験を行うことを目指す。



図 7 全方位移動可能な移動体

参考文献

- [1] 吉田 享平, 日比野 文則, 高橋 泰岳, 前田 陽一郎, 魚眼カメラを用いた自己位置同定と人間指示認識によるナビゲーションシステム, 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, 2011, 27 巻, 第 27 回ファジィシステムシンポジウム, pp. 236 - 227
- [2] 吉本 寛, 久保 守, 村本 健一郎, “上向き魚眼カメラの水平回転移動によるステレオ視”, 情報処理学会, 2009-CVIM-169, 研究報告コンピュータビジョンとイメージメディア (CVIM), pp.1 - 7
- [3] 渡辺 博己, 本郷 仁志, 安本 護, 山本 和彦, マルチカメラを用いた全方位ポインティングジェスチャの方向推定, 電気学会論文誌 C (電子・情報・システム部門誌), 2001, 121 巻, 9 号, pp. 1388-1394
- [4] 山澤 一誠, 八木 康史, 谷内田 正彦, “移動ロボットのナビゲーションのための全方位視覚センサ HyperOmni Vision の提案”, 電子情報通信学会, 電子情報通信学会論文誌 D Vol.J79-D2 No.5 pp.698-707
- [5] Cao, Zhe and Hidalgo, Gines and Simon, Tomas and Wei, Shih-En and Sheikh, Yaser, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, 2018, arXiv,
- [6] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, “Deep Residual Learning for Image Recognition”, 2015, arXiv,
- [7] NVIDIA-AI-IOT/trt_pose,”https://github.com/NVIDIA-AI-IOT/trt_pose”,(最終閲覧日 2022/6/23)