

BoDiD: コミュニケーションロボットのための対話者判定システムの提案

BoDiD: Proposal of interlocutor judgment system for communications robot

桑山 裕基*¹
Yuki Kuwayama

Daniel Wickner*¹

大澤 博隆*¹
Hirotaka Osawa

今井 倫太*²
Michita Imai

*¹慶應義塾大学 理工学研究科

Graduate School of Science and Technology, Keio University

*²慶應義塾大学 理工学部 情報工学科

Department of Information and Computer Science, Keio University

In the present study, it proposes system BoDiD that judges the interlocutor detecting the direction of person's body for the communications robot. BoDiD has the adjustment module on measurement side of laser range finder by face detection and the presumption module in the direction of person's body by measurement of laser range finder. BoDiD enables whether the user of the communications robot is an other party of the conversation to be judged.

1. はじめに

近年、多数の知能ロボットが開発されており、コミュニケーションを目的としたロボットが注目されている。心理学の分野で、人と人のコミュニケーションにおいて、お互いの向きや位置関係は重要なものであるとされる。[Ruusuuvuori 01]では医者と患者のコミュニケーションにおいて、患者に医者が対話者となっているかどうかという判定に、体の向きが影響していることを示している。ロボットは、ユーザの体の向きや位置関係を理解し、適切なインタラクションを行える必要がある。

対話を円滑にするためには、人の注意から対話者をロボットが認識する必要がある。つまり、ロボットが人の振る舞いや身体情報をセンサによって検出し、誰に、もしくは何に注意が向いているか判定しなければならない。

ユーザの注意対象を特定する研究は多く行われている。脇山らは眼球運動から注意対象を検出することで、ユーザの興味モデルを作成している [脇山 他 07]。また、対話相手特定する研究では長尾らの研究が挙げられ、複数人での対話における頭と目の動きを表現できるエージェントのインタラクションについて述べられており、エージェントの顔の向きによる話者交代を行っている [Nagao *et al.* 94]。Langらは複数のユーザが存在する環境で、コミュニケーションロボットがユーザの注意に基づいたインタラクションについて述べている [Lang *et al.* 03]。Langらの研究では、カメラによる顔検出とレーザレンジファインダによる脚検出を行い、人の距離、方向、高さ情報を推定している。

人の体の向きを検出することで、ユーザの注意対象を推定することが可能であり、対話相手の推定に有用である。しかし、注意対象を推定する既存の研究の多くは画像処理によって顔の向きや視線方向を利用して行うものである。体の向きに着目している研究でも、定量的に示されていない。より円滑なインタラクションを行うために、体の向きを検出することのできるシステムが必要である。

コミュニケーションロボットが実際に活躍すると考えられる場合は、公共の場や家庭環境が考えられる。たとえば案内ロボ

連絡先: 桑山裕基, 慶應義塾大学大学院理工学研究科, 神奈川県横浜市港北区日吉 3-14-1, TEL/FAX 045-560-1070, kuwayan@ayu.ics.keio.ac.jp

ットが次々に現れるユーザに対して動作するような場合を考えると、ユーザにセンサを取り付ける必要がない方が望ましい。また、複数のカメラを利用して3次元モデルを作る場合、設置型のカメラでは移動型のロボットには適用することが難しく、計算量が膨大になってしまう。

本研究ではレーザレンジファインダを利用した体の向きの検出システム BoDiD を提案する。BoDiD はユーザの対話相手がロボット自身にあるかを知るために、相対するユーザの体の向きを調べるシステムである。本研究では、肩の高さの個人差や、ユーザとロボットの位置関係を考慮する。状況に合わせて測定する平面の方向を変えるために、コミュニケーションロボットの頭にレーザレンジファインダを取り付け、ユーザに合わせてロボットを動作させることで、肩の位置で計測ができるシステムを作成した。肩での計測後、得られたデータから体の向きを検出し、ユーザが対話相手であるかを推定する。本研究では、人の体の向きの検出システム BoDiD を用い、コミュニケーションロボットにおける対話相手の推定を実現する。

2. インタラクションにおける体の向き情報の重要性

Ruusuuvuori は医者と患者のインタラクションにおいて、医師の状態(正面を向いている、カルテを書いている)によって、医者が対話相手となっているかどうか、という患者の判断に影響を与えていることを示している [Ruusuuvuori 01]。

ロボットとユーザのインタラクションにおいても、ユーザが受ける印象はロボットとユーザの位置関係や方向などの条件によって変化するものだと考えられる。エージェントの体の位置や方向について述べられている研究として、鈴木らの研究 [鈴木 他 08] が挙げられる。鈴木らは画像内に表現される人型のエージェントの体の位置や方向を変化させて、ユーザが受ける印象を比較する実験を行っている。しかしながら、体の向きは正面、背面方向といった大まかなものである。

コミュニケーションロボットがユーザとインタラクションを行う際、ユーザが対話相手となっているかを判断する必要がある。対話相手を判断するために、ユーザの体の向きが重要であると考えられる。本研究では、ユーザの体の向きを定量的に検知できるシステムの作成を行う。

3. BoDiD のシステム構成

全体のシステムを図1に示す．図中の LRF はレーザレンジファインダを指す．BoDiD はレーザレンジファインダによって，ロボットに相対するユーザの体の向きを検出を行う．BoDiD は LRF 計測面の調整モジュールと体の向きの推定モジュールの2つで構成される．

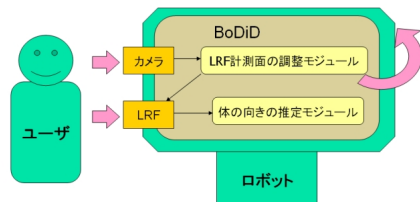


図1: BoDiD のシステム図

1. LRF 計測面の調整モジュール

カメラから得られた画像を得て，ロボットを動作させ，レーザレンジファインダの計測面が肩の位置になるようにする．計測面が決定されると，レーザレンジファインダに計測開始命令を送る．

2. 体の向きの推定モジュール

体の向きの推定モジュールはレーザレンジファインダから計測データを取得し，人間の体の範囲の決定，体の向きの推定，対話者であるかの判定を行う．

3.1 LRF 計測面の調整モジュール

LRF 計測面の調整モジュールでは，ユーザの顔の位置の検出し，レーザレンジファインダの照射方向の設定，レーザレンジファインダへ計測開始命令の送信を行う．レーザレンジファインダは照射平面上の距離を計測する．レーザレンジファインダはロボットの頭部に固定されているため，ロボットの頭部を動作させてユーザの肩の位置にレーザが照射されるように調整を行う．調整が終了すると，レーザレンジファインダに計測開始命令を送信する．

ロボットに送信するモータ値はカメラで得られた画像上での平均化された顔の座標値と画像上のある目標点と比較して生成される．カメラで得られた画像上の目標点にユーザの顔があるとき，レーザレンジファインダが人の肩の部分にあたるように，カメラとレーザレンジファインダのピッチ方向を設定しておく．

モータ値は，現在のモータ値に差分量を足し合わせることで新しいモータ値を得る．一度に動く最大のモータ値の変化量 (*MotorStep*) を設定する．目標点と顔の座標値の横と縦の差分 (dx, dy) を算出し，画像の横と縦の大きさ (ix, iy) との比を求め，最大のモータ値の変化量との積を求める．横方向の比はロボットのヨー方向のモータ値 (*NewMotorValueX*) に反映され，縦方向の比はロボットのピッチ方向のモータ値 (*NewMotorValueY*) に反映される．ロール方向のロボットのモータ値は変化させない．新しいモータ値の決定と，ロボットの動作を繰り返すことにより，最終的に顔の座標値は目標点に達する．つまり，ロボットは検出される顔の座標値が目標点と重なるように動作する．以下にモータ値の算出に使用される式を示す．

$$\begin{aligned} \text{NewMotorValueX} &= \text{MotorValueX} + \text{MotorStep} * \frac{dx}{ix} \\ \text{NewMotorValueY} &= \text{MotorValueY} + \text{MotorStep} * \frac{dy}{iy} \end{aligned}$$

実際には目標点を画像上のある程度の領域でとり，範囲内に

入ったらレーザレンジファインダでの計測開始命令を出す．これはユーザが少し動くことによってレーザレンジファインダの計測が停止してしまうのを避けるためである．

3.2 体の向きの推定モジュール

LRF 計測面の調整モジュールでレーザレンジファインダの照射面が決定すると，レーザレンジファインダの計測が行われる．レーザレンジファインダで得られた計測データからユーザの肩部分の計測データを特定する．肩部分の計測データからユーザの体の向きを推定する．最後に，推定された体の向きから，ユーザが対話相手であるかどうかを判定する．

3.2.1 ユーザの肩の範囲の推定

得られるデータは，測定範囲内のレーザ照射面上の距離データであり，得られた距離データの中からユーザの肩部分のデータを特定する必要がある．レーザレンジファインダから得られるデータは計測範囲内のステップ角ごとの連続した距離データのため，近くの距離データ間の値の差をとり，差分の大きい部分をエッジとし，距離が小さくなるエッジと，距離が大きくなるエッジのペアを検出する．エッジのペアの角度とエッジ部分での対象との距離がわかるので，大体の大きさを推定することができる．エッジのペアから，平均的な距離が比較的近く，人がとりえそうな範囲にないデータを除くことで，ユーザの肩部分の範囲を推定した．ユーザの肩部分の範囲内の距離データを体の向きの推定に使用する．本研究では，カメラ上での体の位置とレーザレンジファインダ上での体の位置の対応をとっていないので，同時に対象とできるユーザは1人となる．

3.2.2 体の向きの推定

推定されたユーザの肩部分のデータはレーザレンジファインダからユーザまでの距離データであるので，直交座標系 $(x, y) = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ への変換を行う．本研究では，ユーザの体を平面として扱う．そのため，ユーザの体から得られた座標を最小二乗法で直線に近似し，直線の法線方向を体の向きとした．最小二乗法による近似された直線 $y = ax + b$ の a, b は以下のように求められる．

$$a = \frac{n \sum_{k=1}^n x_k y_k - \sum_{k=1}^n x_k \sum_{k=1}^n y_k}{n \sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k \right)^2} \quad (1)$$

$$b = \frac{\sum_{k=1}^n x_k^2 \sum_{k=1}^n y_k - \sum_{k=1}^n x_k y_k \sum_{k=1}^n x_k}{n \sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k \right)^2} \quad (2)$$

推定される体の向きの値は，ロボットの頭部の水平方向の向きとユーザの体の向きが向かい合っているときを0度とし，ユーザが真横を向いている状態を90度または-90度とした．Robovie に対して左側がプラス，右側がマイナス方向となる．

レーザレンジファインダでの計測は必ず水平方向に行われるわけではない．ユーザの体を近似する平面のモデルはレーザレンジファインダの計測面に対して垂直な面となる．このため，レーザレンジファインダのピッチ方向の傾きによって推定されるべき体の傾きの値は変化することになる．しかし，本研究ではピッチ方向の変化は少ないことを仮定し，計測面のピッチ方向への変化に対して考慮していない．また，実際には人間の体は平面でないので，体の向きが左右に大きくなるほど，体の厚みにより誤差が大きくなると考えられる．これに対応するた

め、補正なしの状態、つまり完全に人間を平面と仮定して算出された体の向きから、補正を行えるようにした。

3.2.3 対話相手の判定

推定された体の向きから、ユーザの対話相手がロボット自身にあるかを判定する。対話相手の判定においては、ユーザの体のロボットに対する角度の閾値として自由に設定できるようにした。

4. BoDiDの実装

本研究では、コミュニケーションロボットに Robovie を使用する。Robovie に搭載したレーザレンジファインダとカメラの様子を図 2 に示す。ロボット自身の目となる 2 つのカメラが存在するが、レーザレンジファインダとの校正を行いやすくするため、レーザレンジファインダの真上のカメラを追加した。



図 2: Robovie とレーザレンジファインダとカメラ

4.1 LRF 計測面の調整モジュール

4.1.1 カメラによる顔検出

640 × 480 ピクセルの解像度を持つカメラを使用した。顔の検出には OpenCV で haarcascade を利用した。顔の座標を安定して取得するため、本研究においては、5 回のスキャンから顔の座標の平均を出すように実装した。

4.1.2 モータ値の決定

Robovie のモータ値は -5.0 から 5.0 の間で設定されている。そのため、最大のモータ値の変化量は 10 以下で設定されなければならない。本研究では、インタラクション中にユーザがあまり大きく動かないことを考慮し、最大のモータ値の変化量を 2 に設定した。-5.0 より小さい、または 5.0 より大きい場合はそれぞれ -5.0, 5.0 に設定される。画像上の目標点は、画像の中心である $(x, y) = (320, 240)$ に設定した。また、目標範囲は横方向に ± 32 ピクセル、縦方向に ± 24 ピクセルとした。

4.2 体の向きの推定モジュール

4.2.1 ユーザの肩の範囲の推定

レーザレンジファインダ URG-04LX では、前方 ± 120 度の範囲で、本体周囲の 20mm から 4000mm までの計測が行われる。距離データは 1mm 単位で出力され、角度分解能は約 0.36 度 (360 度/1024 分割) である。

取得するデータはレーザレンジファインダ前方の約 90 度 (± 45 度) の範囲に設定した。顔検出により、ユーザがレーザレンジファインダの正面にすることが前提となるためである。エッジ部分の推定として、隣り合うデータ間の差が 1000mm 以上あるデータ部分をエッジとした。つまり、検出できる範囲は 3000mm までの範囲となる。エッジは距離が 1000mm 以上小さくなるエッジと 1000mm 以上大きくなるエッジを 1 組とし

てとらえ、この間を対象のデータとする。また、常にユーザが正面にいと仮定できるので、1 組のエッジ間で正面方向のデータを含むことを条件とした。

4.2.2 体の向きの推定

人の肩の部分の距離データの始まりと終わりがわかると、始まりと終わりの間のデータから、式 1 と式 2 にしたがって最小二乗法による近似直線 $y = ax + b$ を求める。しかし、レーザレンジファインダからわかるデータは角度と距離のデータであるので、まず、X-Y 座標に変換し、計算を行った。

直線の傾きがわかるので、体の向き θ は a の正接の逆関数をとればよい。以上で体の向きが求まる。

5. 評価

5.1 距離による比較

図 3 に 0.5m から 2.5m まで、0 度から 80 度まで 20 度ごとに体の向きを変化させたデータを示す。それぞれ 100 回の計測から平均値を求めた。被験者には身長約 170cm の平均的な体格の 20 代男性を選んだ。距離によって、Robovie の頭部のピッチ方向の変化が変化する。また、レーザレンジファインダの角度分解により、計測されるひとつの角度 (ステップ角) あたりの計測面上の長さが変化するため、一般に距離が遠くなる方が、レーザレンジファインダで計測される距離データが少なくなる。実際の体の向きが 0 度のとき、誤差はほぼ 0 に近い値が推定された。20 度、40 度になるにつれて推定された体の向きの各距離間のばらつきが大きくなっている。体の向きが正面方向から左右に離れるにつれて、計測を行うときの体のゆらぎやレーザレンジファインダの計測面上にあった、体の凹凸により、影になったりならなかったりする部分の影響が大きくなると考えられる。実際の体の向きが 60 度を超えると、計測ではかなりの誤差が出ていることがわかる。計測すべき肩の部分で、レーザレンジファインダから影になってしまう部分が大きく増加するためだと考えられる。

距離にかかわらず、40 度までの体の向きで考えると、ほぼ同様に推定が行われていることがわかる。実際の体の向きが 40 度までの推定値から直線を近似して算出し、補正を行えば実際の値に近い値を推定できると考えられる。

5.2 個人差による比較

体の向きの推定は様々なユーザに対して対応しなければならない。そこで、体格、性別の異なるユーザで評価を行った。身長 180cm の男性、身長 170cm の男性、身長 150cm の女性の 3 人を被験者とした。それぞれの被験者で 0 度、20 度、40 度、60 度、80 度で各 100 回の計測から平均値をとった。それぞれの被験者の推定された体の向きと実際の体の向きを比較した図を図 4 に示す。

ユーザ同士での誤差は、実際の体の向きで 0 度、20 度、40 度では ± 2 度以下となっている。60 度では多少のばらつきが生じ、80 度ではほぼ計測できていない状態であることがわかる。データ間でのばらつきは、距離が違ったときと同様な理由が挙げられる。また、ユーザが違うので、体型や身長、服装などもユーザ間のデータのばらつきに影響しているものと思われる。しかし、40 度までのデータではほぼ同じように推定されている。よって、40 度までの推定値から直線を近似して算出し、補正を行えば実際の値に近い値を推定できると考えられる。

5.3 対話者の判定

本研究では、ユーザが対話者となっているかどうか判定するための角度の閾値の設定のための予備実験として、Robovie

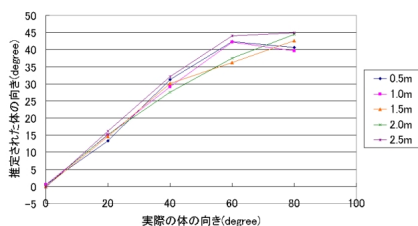


図 3: 距離による変化

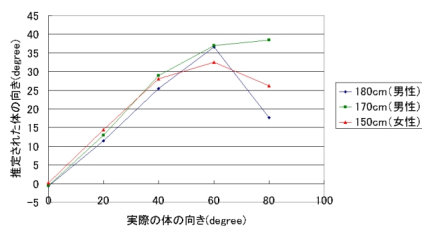


図 4: ユーザごとの比較

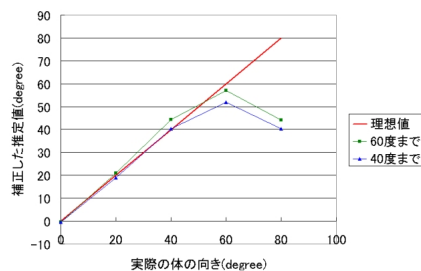


図 5: 補正を適用した推定値の平均

と Robovie の横に 2 人並んだ状況を設定した．それぞれ立つ位置は 0.6m ほど離れている．Robovie の隣の人を A，さらに A の隣の人を B とする．被験者は Robovie の正面 1m のところに立ち，Robovie と Robovie の横に並んだ人と一人ずつ会話するイメージで立つ．それぞれの状態について，Robovie から被験者数人の体の向きを計測した．その状況を図 6 に，得られた被験者の体の向きのデータから平均と分散を求め，それぞれの正規分布を表したものを図 7 に示す．

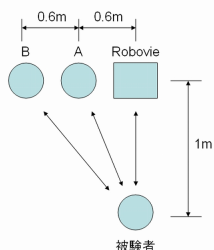


図 6: 予備実験

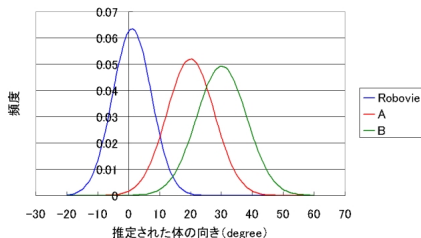


図 7: 閾値の設定

Robovie と A を判別したい場合と Robovie と B を判別したい場合は，インタラクションや周囲の環境によって変化すると考えられる．本研究では，コミュニケーションロボットとユーザが一對一でインタラクションをすることを考えるので，Robovie と A を判断する境界を定めることを考える．Robovie と A の正規分布は推定された体の向きの 10 度のところで交差しているので，閾値を ± 10 度に設定した．この値は，補正なしの状態であるので，実際の体の向きとは異なる可能性がある．本研究においては補正なしの状態で得られたユーザの推定された体の向きが ± 10 度までである場合，ユーザが対話相手であると判定する．

5.4 補正值の検討

本研究では様々なユーザがコミュニケーションロボットを使うことを考慮し，ユーザを変えて実験したものから補正值の検討を行う．実際の体の向きが 40 度までの部分で補正を行うことを考える．比較のために，実際の体の向きが 60 度のときまで考慮した補正值も算出した．

補正值の算出を行う．補正值を求めるための関数を求めるため，推定された体の向きのデータの平均値で最小二乗法による直線への近似をする．式 1, 2 を用いることにより，直線を求めた．実際の体の向きで 40 度まで考慮した場合と，60 度まで考慮した場合はそれぞれ， $y = 0.68x$ ， $y = 0.62x$ が求められた．求められた直線の傾きと，理想値である直線の傾きを比較し，直線の傾きの比率を求める．40 度までの場合 $1.0/0.68$ ，60 度までの場合 $1.0/0.62$ となる．求められた比率を推定され

た体の向きに適用する．図 5 に補正された値を示した．ここで，ユーザが対話者であるかどうか判定をする閾値の値も変化する．閾値は体の向きを推定することによって得られた値なので，推定値を補正するのと同じく，閾値も補正する必要がある．閾値は ± 10 度としたので，補正に使用する比率をかければよい．つまり，40 度までを考慮した比率を使う場合の閾値となる推定値は ± 14.7 度，同様に，60 度までを考慮した比率を使う場合は ± 16.1 度となるこれにより補正を行った場合でも閾値を決定することができ，対話者を判定することができた．

6. 結論

本研究では，体の向きを検知することでユーザの対話相手がロボット自身にあるかどうかを判定するために，レーザレンジファインダを用いて人の体の向きを検出するシステム BoDiD を提案し，実装を行った．

本研究では，レーザレンジファインダを用いて人の体の向きを検知するシステム BoDiD を用いることで，2.5m 以内にある人の体の向きを検知し，ユーザの対話相手の特定に利用できるシステムを実現した．

参考文献

[Ruusuuvuori 01] J. Ruusuuvuori: Looking means listening: coordinating displays of engagement in doctor-patient interaction, *Social Science & Medicine* 52, pp.1093-1108, 2001.

[脇山 他 07] 脇山 孝貴, 吉高 淳夫, 平嶋 宗: 注目の検出に基づいた興味モデルの作成と絵画推薦, *情報処理学会誌* Vol.48 No.3, pp.1048-1057, 2007.

[Nagao et al. 94] K. Nagao and A. Takeuchi: Social Interaction: Multimodal Conversation with Social Agents, *AAAI-94* Vol.1, pp.22-28, 1994.

[Lang et al. 03] S. Lang, M. Kleinehagenbrock, S. Hohenner, J. Fritsch, G. A. Fink, and G. Sagerer: Providing the Basis for Human-Robot-Interaction: A Multimodal Attention System for a Mobile Robot, *ICMI'03* November5-7, pp.28-35, 2003.

[鈴木 他 08] 鈴木 聡, 森島 泰則, 中村 美代子, 槻館 尚武, 武田 英明: 身体化エージェントの身体方向・登場位置がユーザに与える影響, *知能と情報 (日本知能情報ファジィ学会誌)* Vol.20 No.4, pp.513-525, 2008.