論文

指差しによるロボット指示のための一点較正法

A Calibration Method using One Point for Instructing Robot by Pointing Gestures

正会員 上野 智史[†], 正会員 内藤 整[†], Tsuhan Chen^{††}
Satoshi Ueno[†], Sei Naito[†] and Tsuhan Chen^{††}

あらまし 人が指差しを利用して対象を示す動作には個人差があり, また指差し指示の際には人は必ずしも指先を凝視しないことがある. これは目標物への視線と, 利き目と指先を結ぶ直線が一致しないことを意味し, この場合は指差しを観測したロボットが指差しから目標物を特定することは難しい. そこで筆者らは, 人が指差しにより指示した目標物をロボットが高精度に特定するために, 指差しから視線を簡単に学習する較正法を提案する. 提案法では, 人がカメラに向かって指差しをする映像を利用して, 頭を中心の位置と指先を検出した後にカメラへの視線と指先の差を算出し個人ごとに補正值として保存する. 評価実験の結果, 提案法は従来検討されていた特定部位の検出に基づく推定法より精度よく目標物を特定できることを確認した.

キーワード: 指差し, 較正, ロボット, Pointing Gesture, Calibration, Robot Interaction

1. ま え が き

方向を指示する直感的な身振りとして指差しがある. 本論文では, 人が指差しにより指示した目標物をロボットが特定できるように人の視線を推定することが目的である. 指差しによる視線をロボットが推定するためには, 一般的には人の利き目(以下単に目とする)と人差し指の指先(以下単に指先とする)から生成される直線方向を利用する. ロボットに目標物が存在する平面(例えば机, 床, 壁など)に関する事前知識がある場合は, 視線を検出した後にロボットは指示方向とその平面の交点の位置を探索することで, 目標物を特定できる. 目標物を特定できれば, ロボットはそれを人に持って行く, 捨てる, 等の作業を実現できる. しかしながら, 人の指差しを映像で観測してその指示目標物を特定することは依然として難しい問題である. 問題の一つは人の指差し動作に個人差があることであるが, より本質的な問題は指差し指示の際には人は必ずしも指先を凝視しない場合があることである. これは目標物への視線と, 上記の目と指先から生成される直線が一致しないことを意味する. この場合はロボットが仮に人間の目と指先を両方とも正確に抽出できたとしても, 目標物を特定するためには別の補正が必要となる.

そこで筆者らは, 指差しから目標物を特定するために視線の較正法を提案してきた¹⁾²⁾. 提案法では, ロボットは位置が既知の目標物に対して, 人の指差しを観測して補正

値を学習することにより視線の較正を行う. 本手法による補正の目標は人物が机上の小物(コップやリモコンなど)を指差しをした時に, ロボットが目標物を特定できる精度を簡単な較正で実現することである. 提案法は, 人の負担を軽減させるために一点の目標物への指差し指示から学習することが特徴である.

本論文は以下のように構成される. 2章では従来法とその課題について述べる. 3章では提案法について述べる. 4章で評価実験に関して述べ, 5章で本論文のまとめを述べる.

2. 従来法とその課題

ロボット操作を目的とした方向推定に関してこれまで多くの研究がある. 研究の初期では, 人物の検出や指差しの身振り認識に注力されていた^{3)~5)}. 一方で, 屋内環境に限定すればKinectTMなどのRGB-Dセンサを利用することにより, 人物の検出や身振りの認識は十分な精度で実現されつつある. そこで筆者らは人物検出および身振り認識はKinectによる出力結果を利用し, 指差し方向推定に注力することにする.

指差し方向は, 大まかには目と指先の2点から生成される直線で表現できる. しかしながら, 上記二つの部位は身体の中で比較的小さい部位のため検出が難しく, 画像処理に基づくこれまでの解析では上記2部位の検出の代わりに, 頭-手部位^{6)~8)}, 頭-指先部位⁷⁾⁹⁾¹⁰⁾, 前腕部位⁷⁾⁸⁾¹¹⁾, 顔の向き⁸⁾¹²⁾を検出して指差し方向を推定している. 頭の向きによる手法は指先の位置を利用せず, 音声認識による目標物の別の特徴との併用が必要であった¹²⁾. また前腕の向きによる手法は, 部屋の中でロボットに行くべき方角を指示する場合には効果的である¹¹⁾が, 複数の手法の性能評価を

2014年9月30日受付, 2014年11月17日採録

[†] 株式会社KDDI研究所

(〒356-8502 ふじみ野市大原2-1-15, TEL 049-278-7881)

^{††} Cornell University

(Ithaca, NY 14850 USA)

した研究では、頭-手、頭-指先の検出に基づく手法が前腕の向きによる手法より方向推定精度が高いと結論づけている⁷⁾⁸⁾。筆者らの手法も基本的には頭-指先の検出に基づく手法であるが、従来法⁷⁾⁹⁾¹⁰⁾では考慮されていなかった指差しの個人差を考慮することが特徴である。

指差しの個人差を複数回の指差し学習データを利用して回帰モデルで表現する研究も報告されている¹³⁾¹⁴⁾。これらの手法は方向推定精度は高くなるものの、学習のために位置が既知の目標物に対して人間に指差し指示を複数回行わせる必要があり作業の負荷があった。さらに、学習するために部屋の中に位置が既知の目標物を複数設置する必要があるが、この環境を構築する負荷も軽視できない問題であった。

3. 提案法

提案法は、前記の従来法の課題を解決するためにカメラを目標物として一度指差しをすることにより個人差に対応する較正法である。提案法は、まずKinectが抽出した人物骨格位置を元にして、頭の中心位置と指先を算出する。続いて人がKinectに向かって指差しをした動作から個人に対応する補正値を学習する。課題と提案法で定義した補正値の位置関係を図1に示す。図1のように、目標物を見る視線の直線と目と指先を結ぶ直線との乖離を、人物の側面(胴体)の平面と並行で目標物を含む平面上の座標上で算出し、水平方向の乖離値を σ_h 、鉛直方向の乖離値を σ_v とする。筆者らは鉛直方向の乖離値が水平方向の乖離値より大きいと仮定し、主に鉛直方向の乖離値の補正を行う。また提案法は、従来法¹³⁾¹⁴⁾と比較して指差しの大量の学習データを必要としないことと、位置が既知の目標物を複数設置することなく実現できることが特徴である。

本論文では、評価のために静止したKinectをロボットのカメラと見立てる。また座標系は、Kinectで利用されるカメラ座標系を利用する。すなわち、 x 軸はカメラから左向きであり、 y 軸はカメラから上向きであり、 z 軸はカメラから人物に向かう向きである。カメラは水平に配置するため、鉛直方向は y 軸と平行になる。Kinectが出力する人物骨格情報(頭の中心、手など20箇所)を、以下単に骨格情報と表記する。

提案法により補正値を算出する前に、Kinectによる骨格情報の補正(前処理)を行い、その後カメラ中心に向かった指差しから個人の補正値を取得する。

3.1 前処理

(1) 頭の位置の補正

Kinectによる頭の中心の検出は、手により頭の一部が隠される場合と、側面からの撮影の場合に精度が低い場合がある。位置検出精度の誤差は x 軸および z 軸の方向に大きく、 y 軸方向には小さいことが実験的に確認できるため、骨格情報の胴体の中心と両肩の中心を用いて、二つの点の延長線上に頭の位置を補正する。補正した頭の中心を P_h

図1 指差しと視線の関係図

図2 頭の中心補正の模式図

とし、模式図を図2に示す。胴体の中心と両肩の中心が腕により隠されている場合は、Kinectが出力した頭中心の骨格情報を、補正した頭の位置 P_h として利用する。

(2) 指先の位置の補正

Kinectは骨格情報と同時に人物領域を点群で出力する。そこで骨格情報と点群の人物領域を併用して指先を抽出する。本論文では、指先 P_f は肘と手の骨格位置を基準に肘から手に向かった半直線上に存在する人物領域の端点とする。深度センサは境界検出の精度が高くないため、手の骨格位置と指先 P_f が閾値以上離れる場合は手の骨格位置を指先とする。

3.2 指の長さおよび仰角の補正

(1) 指の長さの補正

前述のように、画像処理で指先を正確に抽出することは困難であるため指の長さの補正を行う。図3に補正を行った指先の定義を示す。補正した後の新しい指先の位置 P_{ef} は、頭の中心の位置 P_h 、胴体の中心の骨格情報、カメラから生成される平面と、肘と手の骨格情報を結ぶ直線と平行で指先 P_f を始点とした肘から手の向きへの半直線との交点である。カメラに向かって正面から指差しをする場合、指示をする上腕の向きはカメラの画像平面にほぼ垂直になる。この環境下ではKinectから見て肩または肘が手に隠れることがあるため、骨格情報の抽出に失敗することがある。

図3 指の長さ補正の定義

図4 仰角補正の定義

そのため、指差し時に肩、肘、手の位置に関して整合性の確認をする。直立時と指差し時でそれぞれの関節の長さを比較し、一定閾値以上の変化が生じている場合は整合性がとれていないと判定する。3点の整合性が確認できない場合は、腕から生成される直線の信頼性が低いと判定し、指の位置の補正值 o_p を0にする。

(2) 仰角の補正

指の長さを補正後，仰角の補正を行う．仰角の補正值 θ_0 は頭から指先の直線と，頭からカメラ中心への直線のなす角である．図4に仰角補正を行った指先の位置 $\mathbf{P}_{cf}$ の定義を示す．人が腕を大きく動かさない動作（小さな身振り）で指差しをする時¹⁵⁾，仰角は相対的に大きくなる．提案方式では仰角を補正するのでこのような小さな身振りの場合でも視線の校正が可能である．

3.3 視線の算出

個人ごとに補正値を取得した後、指差しに対して補正値を元に視線を校正する。指先の位置 $\mathbf{P}_f$ から肘と手の骨格情報から生成される直線方向に $\mathbf{o}_p$ を加算し、頭中心を軸に仰角の補正 $\mathbf{o}_\theta$ を鉛直方向に加算する。これにより校正した視線は、頭中心 $\mathbf{P}_h$ と校正した指先 $\mathbf{P}'_{ef}$ を結ぶ直線になる。二つの補正値はカメラに向けた一回の指差しからフレームごとに算出し、その平均値を個人ごとに保存する。また、それぞれの補正値は負の値もとりうる。さらに、Kinectの分解能の制約と大きな外れ値を取り除くために、補正値の絶対値には以下のように下限と上限を設定した。上限値以上の補正値は上限値を採用する。下限値以下の補正値の場合は、値が定まらず正負の値を交互にとることがあるため一

律0で処理する.

$$0.05 \leq |o_p| \leq 0.25 \text{ (m)}$$

$$1.0 \leq |o_\theta| \leq 10.0 \text{ (deg)}$$

4. 評価実験

4.1 実験環境

屋内環境で指差しによる目標物推定精度の評価を行った。筆者らは、部屋の中で日常の生活用品をロボットに持ってこさせたり捨てさせたりする際に対象物を指差して指定することを想定している。そこでテーブルの上に目標物を複数設置し、指差して小物を特定する識別精度を評価した。図5に実験環境の模式図を示す。16個の目標物が4×4に並べられており、それぞれの間隔は0.25 mである。また、テーブルの高さは0.7 mである。

さらに，人物とカメラの向きと指差しの推定精度の評価を行うために，(a) 正面から，(b) 斜めから，(c) 側面からの3視点から撮影した映像を利用した．すべてのカメラの高さは1mである．それぞれのカメラ視点からのカメラ座標系に対する目標物の位置（正解位置）は，Kinect深度センサの画像から目標物を手で特定してKinect深度センサの座標値を利用した．

評価は被験者10人(男性9名, 女性1名, 20~30代)で実施した。被験者の身長は1.6~1.9 mとさまざまで, 利き腕は9人が右腕, 1人が左腕で, それぞれ利き腕で指差しを行った。較正には正面からの指差し映像を利用し, すべての身振りは指定位置で起立した状態で行った。一人の被験者ごとに各目標物を3回指示し, 合計480回(10人×16箇所×3回)の指差しの映像データを取得した。取得した映像例を図6に示す。

観測による目標物の特定方法は、手法ごとに生成した直線に最も近い目標物を検出目標物とした。評価値は0から1の値を取り、正解目標物を特定できた場合は1、不正解時は0として、評価値が大きいほど精度が良いことを示す。本論文では、以下の手法の評価を行った。

図5 実験環境

図6 3方向から撮影した映像例

手法(1)：骨格情報の頭と手の直線

手法(2)：骨格情報の頭と提案法(前処理)の指先の直線

手法(3)：Kinect顔検出による目の位置と提案法の指先の直線

手法(4)：提案法1:前処理後の頭と指先の直線

手法(5)：提案法2:提案法1に加え、指の長さを補正した直線

手法(6)：提案法3:提案法2に加え、仰角を補正した直線

手法(7)：ガウスモデルを利用した回帰による手法¹³⁾

手法(3)に関して、(b)斜めと(c)側面からの視点からでは目を検出できない場合があります、その場合は目の位置の代わりに骨格情報の頭の位置を代用する。手法(7)に関して、回帰モデル生成のために取得した映像の半分を用いて、16点の目標物を対象にした指差しの学習データを生成した。すべての手法の評価は、残り半分の映像データを対象として実施した。

4.2 実験結果

(1) 乖離値の傾向検証

3章で述べた、水平方向と鉛直方向の乖離値の検証を行う。乖離値を直接観測することは困難であるため、実験で利用した目標物に対して、人に向かって前後方向と左右方向にグループ分けを行いその識別性能で乖離値の傾向を検証する。例えば水平方向の性能評価は、4×4に並んだ16個の目標物を人に向かって前後に1列ずつの4グループに分け、左右に4つの大きなオブジェクトがあると仮定した時のオブジェクトの識別性能である。同様に鉛直方向の性能評価は、4×4に並んだ16個の目標物を人に向かって左右に1列ずつの4グループに分け、前後に4つの大きなオブジェクトがあると仮定した時のオブジェクトの識別性能である。カメラ視点別、手法別の精度を表1に示す。評価値は0から1の値をとり、正解オブジェクトを特定できた場合は1、不正解時は0として、評価値が大きいほど精度が良いことを示す。

表1の結果より、すべての手法において水平方向の目標物の特定精度は鉛直方向の特定精度と比較して高いことが確認できる。これにより水平方向の乖離値が鉛直方向の乖離値と比較して小さいことが予想される。これは、目が左右についているため指先の水平方向の識別感度が鉛直方向の識別感度より高いことが原因の一つと考えられる。また、提案法は鉛直方向の識別精度向上に寄与していることが確認できる。

(2) 目標物識別精度

図7に手法(1)から(4)のカメラ視点別の目標物特定精度

表1 カメラ視点別の目標物特定精度の評価値

カメラ視点	手法	水平方向	鉛直方向
(a) 正面	(1)	0.49	0.24
	(2)	0.90	0.69
	(3)	0.91	0.64
	(4)	0.91	0.75
	(5)	0.96	0.81
	(6)	0.98	0.82
	(7)	0.98	0.85
(b) 斜め	(1)	0.80	0.24
	(2)	0.90	0.71
	(3)	0.83	0.67
	(4)	0.95	0.83
	(5)	0.99	0.88
	(6)	0.98	0.90
	(7)	0.98	0.93
(c) 側面	(1)	0.80	0.24
	(2)	0.86	0.40
	(3)	0.86	0.40
	(4)	0.93	0.83
	(5)	0.93	0.87
	(6)	0.93	0.89
	(7)	0.98	0.86

図7 手法(1)から手法(4)のカメラ視点別の目標物特定精度

図8 手法(4)から手法(7)のカメラ視点別の目標物特定精度

の平均値を示し、図8に手法(4)から(7)のカメラ視点別の目標物特定精度の平均値を示す。誤差範囲は標準誤差値を示す。図7から提案法の前処理により特に側面からの視点に対して精度が大幅に向上していることが確認できる。また、図8から提案法がそれぞれ手法(4)から(6)と提案法の段階を経るごとに精度が向上していることが確認できる。さらに提案法(6)は、全点を対象とした学習法(7)の精度には若干劣るものの近い精度を実現していることが確

認できる。手法 (7) の学習の作業量と学習環境の構築の手間を考慮すると、提案法はテーブル上の小物を識別するための精度に到達するためにはより簡単な校正法であるといえる。なお、提案法による校正した視線と、頭を中心と目標物の正解位置とを結ぶ直線の平均乖離度は2.1度であった。これは2m離れた目標物の平均識別誤差は7.3cmであることを意味し、屋内でのテーブル上の小物の識別するシーンには有効であると言える。

さらに、人物の向きに対するカメラ視点に関して考察する。最も精度が良いのは、(b) 斜めからの視点であり、(a) 正面および (c) 側面からの視点はほぼ同じ精度であることが確認できた。これは、正面からの撮影では指差し時に手と肘がオクルージョンになりやすく、また側面からの撮影では人物から見て左右方向の指先の特定精度が低くなることが原因と考えられる。従来法の多く^{4) 7) 8) 10) 12) 15)}は (a) 正面からの映像を解析しており、その理由は正面からの映像が人物領域を最も簡単に抽出できるからと予想される。しかしながら、Kinectを利用すると人物検出に関しては (a) 正面および (b) 斜めからでも精度は同程度である。このことから、もしロボットが人の指差しをするタイミングを予測できる (例えば、人によるロボットへの事前の声掛けのタイミングなど) のであれば、ロボットは人を斜めから観測できる位置に移動してから指差しによる視線推定をするべきであることがわかる。

5. むすび

本論文では、人の指差し方向を精度よく推定するために簡単に実施できる校正法を提案した。提案法は、観測するカメラに向かって人が一度指差しをすることにより実現することができるため容易に実現でき、かつ目標物の特定精度を高めることができることが特徴である。提案法は、従来の頭と手、頭と指先、目と指先の人物部位の検出を基本とした方向推定法より精度よく指示方向を推定できることを確認した。

提案法は、人の指差しの方法は人ごとに異なるものの同じ人物では変化が少ないことを仮定している。しかし人と目標物との距離や経年により較正值は変化する可能性がある。今後は個人傾向の変化に関して評価を実施する。

〔文 献〕

- 1) 上野, 内藤, Chen: “指差し方向推定のための校正法の検討”, 2014年信学総全大 (2014)
- 2) S. Ueno, S. Naito and T. Chen: "An Efficient Method for Human Pointing Estimation for Robot Interaction", IEEE International Conference on Image Processing (2014)
- 3) C. Martin, F-F. Steege and H-M. Gross: "Estimation of Pointing Poses for Visual Instructing Mobile Robots under Real World Conditions", Robotics and Autonomous Systems, 58, 2, pp.174-185 (2010)

- 4) M. Sigalas, H. Baltzakis and P. Trahanias: "Gesture recognition based on arm tracking for human-robot interaction", Intelligent Robots and Systems (IROS), pp.5424-5429 (2010)
- 5) S. Waldherr, R. Romero and S. Thrun: "A Gesture Based Interface for Human-Robot Interaction", Autonomous Robots, 9, 2, pp.151-173 (2000)
- 6) B. Burger, I. Ferrane, F. Lerasle and G. Infantes: "Two-handed gesture recognition and fusion with speech to command a robot", Autonomous Robots, 32, 2, pp.129-147 (2012)
- 7) Z. Li and R. Jarvis: "Visual Interpretation of Natural Pointing Gestures in 3D Space for Human-Robot Interaction", Control Automation Robotics & Vision (ICARCV), pp.2513-2518 (2010)
- 8) K. Nickel and R. Stiefelbogen: "Visual recognition of pointing gestures for human-robot interaction", Image and Vision Computing, 25, 12, pp.1875-1884 (2007)
- 9) P. Jing and G.Y. Peng: "Human-computer Interaction using Pointing Gesture based on an Adaptive Virtual Touch Screen", International Journal of Signal Processing, Image Processing and Pattern Recognition, 6, 4, p.81-92 (2013)
- 10) Y. Yamamoto, I. Yoda and K. Sakaue: "Arm-pointing gesture interface using surrounded stereo cameras system", 17th International Conference on Pattern Recognition (ICPR), 4, pp.965-970 (2004)
- 11) M.V. Bergh, D. Carton, R.D. Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlitz, D. Wollherr, L.V. Gool and M. Buss: "Real-time 3D Hand Gesture Interaction with a Robot for Understanding Directions from Humans", 20th IEEE International Symposium on Robot and Human Interactive Communication, pp.357-362 (2011)
- 12) R. Stiefelbogen, H.K. Ekenal, C. Fugen, P. Gieselmann, H. Holzapfel, F. Kraft, K. Nickel, M. Voit and A. Waibel: "Enabling Multimodal Human-Robot Interaction for the Karlsruhe Humanoid Robot", IEEE Transactions on Robotics, 23, 5, pp.840-851 (2007)
- 13) D. Droschel, J. Stuckler and S. Behnke: "Learning to Interpret Pointing Gestures with a Time-of-Flight Camera", sixth ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp.481-488 (2011)
- 14) N. Jojic, B. Brumitt, B. Meyers, S. Harris and T. Huang: "Detection and estimation of pointing gestures in dense disparity maps", Fourth IEEE international Conference on Automatic Face and Gesture Recognition, pp.468-475 (2000)
- 15) C-B. Park and S-W. Lee: "Real-time 3D pointing gesture recognition for mobile robots with cascade HMM and particle filter", Image and Vision Computing, 29, 1, pp.51-63 (2011)

上野 智史 2004年、東京大学工学部電子情報工学科卒業。2006年、同大学大学院修士課程修了。同年、KDDI (株) に入社。2012年～2013年、米コーネル大学客員研究員。現在、(株) KDDI 研究所超臨場感通信グループ研究主査。正会員。

内藤 整 1994年、早稲田大学理工学部電子通信工学科卒業。1996年、同大学大学院修士課程修了。同年、国際電信電話 (株) に入社。以来、同社研究所にて、主にHDTV、4K/8K 超高精細映像の動画像符号化および伝送方式、自由視点映像技術を始めとする超臨場感メディアの研究・開発に従事。現在、(株) KDDI 研究所超臨場感通信グループリーダー。工学博士。正会員。

Tsuan Chen 1987年、国立台湾大学電機工程学科卒業。1993年、米カルフォルニア工科大学大学院博士課程修了。同年、米ベル研究所に入社。1997年～2008年、米カーネギーメロン大学教授。2009年より、米コーネル大学教授。工学博士。IEEE フェロー。