

腕差しでの物体指示による家電操作システムの構築

Construction of a Camera-Based System for Controlling Home Application via Arm Pointing

○学 横田雅恵 (中央大) 眞嶋宗一郎 (中央大)
Sarthak Pathak (中央大) 正 梅田和昇 (中央大)

Masae YOKOTA, Chuo University, yokota@sensor.mech.chuo-u.ac.jp
Soichiro MAJIMA, Chuo University
Sarthak PATHAK, Chuo University
Kazunori UMEDA, Chuo University

Recently, in addition to the use of remote controllers and voice recognition, gesture-based methods have been studied as a way to operate home appliances. In this research, we construct an intuitive system in which home appliances are selected and operated via arm pointing. The conventional method uses 3D information of home appliances prepared in advance, but our proposed system can handle home appliances from only image without requiring their positions in advance. We conducted an evaluation experiment of the proposed system and obtained a high gesture recognition rate. In addition, a usability study also indicated good results for real world use.

Key Words: Intelligent room, Human machine Interface, Gesture recognition, Control system of appliances

1 序論

我々は日常的に多くの家電製品を利用して生活している。一般的に、一つの家電に対して一つのリモコンで操作を行う。一方で、日本では総人口に対して65歳以上が28.9%を占めており、高齢化が深刻である[1]。しかし、近年では家電製品の高機能化・高性能化に伴い、リモコンのボタンの縮小化や操作の複雑化という課題がある。これにより、指先を細かに動かしづらい高齢者にとってリモコン操作が難化している。そのため、ハンドジェスチャなどを用いることで、手先をうまく動かせない人でも簡単に家電を遠隔操作できるシステムの研究が進められている。

先行研究においてハンドジェスチャを用いた家電操作システムには、操作する家電を指差して選択するもの[2]-[7]、操作ごとにジェスチャを紐付けするもの[8]-[13]、操作する空間ごとにコマンドを分けるもの[14]-[16]などがある。Muranakaら[13]のように操作ごとに対応したジェスチャを用意すると、同じ電源操作でも家電ごとにジェスチャを区別する必要がある。入江ら[14]は複数種類のハンドジェスチャと家電操作のコマンドを紐づけ、室内に複数設置したカメラでジェスチャを認識するインテリジェントルームを構築した。しかし、この手法では、コマンドと操作の対応関係やコマンドの位置をユーザが記憶する必要がある。また、斎藤ら[7]は家電の位置座標や実寸サイズなどのデータセットを事前に用意し、家電の位置関係を利用した家電選択システムを構築した。しかし、腕差しで操作したい家電を選択できる一方で、家電の位置や種類が変わるたびにデータを更新する必要があるため、ユーザの負担が大きいと考えられる。

そのため、本研究では家電の位置情報を事前に用意せず、ユーザの指示方向と家電の中心座標情報を組み合わせることで任意の位置の家電を選択し電源の操作を行うシステムを構築した。

2 提案手法

2.1 提案手法概要

提案手法の大まかな流れを図1に示す。本システムの環境は図2のようになっており、天井四隅に設置したそれぞれのカメラから室内の撮影を行う。得られた4枚の画像に対して、10フレーム目にYOLO[17]を用いて家電の検出を行う。最初の家電検出以降は3フレームに一度OpenPose[18]を用いて人物の骨格点推定を行い、10フレームに一度YOLOを用いて家電検出を行う。これにより図3aのように家電の中心点と骨格点それぞれの座標から腕差しによる家電選択と電源操作を行う。また、誤操作を防ぐために図3bのような家電操作を開始・終了するジェスチャも設定した。

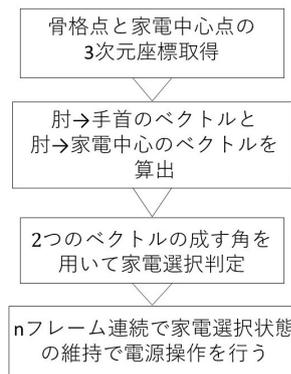


Fig.1: Flowchart of this system

2.2 家電の検出

本システムではYOLOv4[19]を用い、家電の検出を行った。検出したい物体はTV、加湿器、ソファ、ロボット掃除機であったが、あらかじめ用意されたCOCOデータ[20]では加湿器とロボット掃除機を検出できなかったため、これらを検出できる検出器を作成した。本システムの空間を撮影した122枚の写真でアノテーションを行い、データを拡張し、2928組のデータセットを得た。このうち2050枚を学習に用い、878枚を検証に用いた。これにより、TV、加湿器、ロボット掃除機、ソファの平均適合率はそれぞれ100%、99.12%、99.14%、100%となり、高精度の検出器が作成できた。この検出器を用いることで検出した家電の周りに図4のようなバウンディングボックスが描画される。

2.3 骨格点の検出

OpenPoseを用い、骨格点を検出する。本システムでは右肘と右手首の骨格点をジェスチャ認識に用いる。この際、信頼度の低い骨格点をジェスチャ認識に採用すると誤った姿勢を認識してしまうため、しきい値を設ける必要がある。ここでは、経験的に0.60と設定した。

2.4 骨格点と家電中心点の計算

同一家電のバウンディングボックスの中心の二次元座標が複数得られるため、ステレオ視の原理[21]から家電の三次元中心点が求められる。同様に、同一人物の右肘、右手首の骨格点の

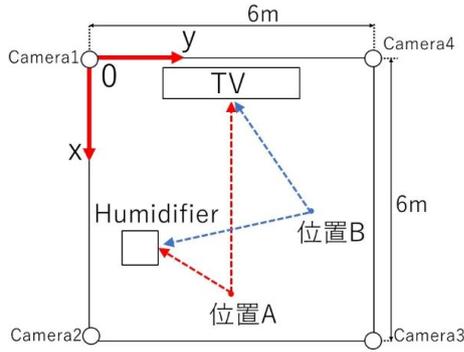
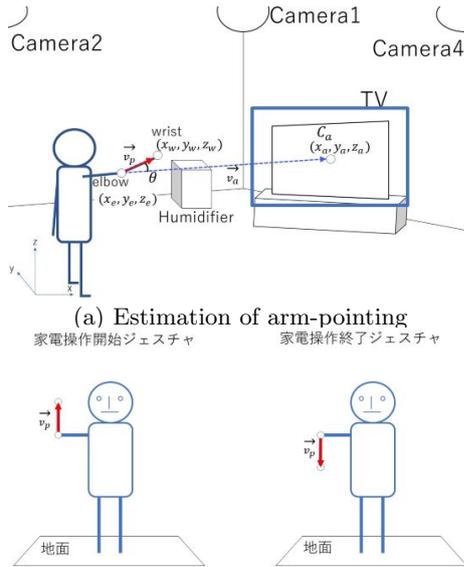


Fig.2: The overview of the room



(a) Estimation of arm-pointing
家電操作開始ジェスチャ 家電操作終了ジェスチャ

(b) Gestures for switching start and end appliance operation

Fig.3: Gestures used in this method

二次元座標が複数得られるため、ステレオ視の原理から三次元座標を求められる。家電の三次元中心点を $C_a = (x_a, y_a, z_a)$ とし、右肘、右手首の骨格点の三次元座標をそれぞれ (x_e, y_e, z_e) , (x_w, y_w, z_w) とする。

2.5 ジェスチャ認識

右肘から右手首への腕差しのベクトル \vec{v}_p と右肘から家電中心へのベクトル \vec{v}_a はそれぞれ式 (1), 式 (2) に示すように求められる。

$$\vec{v}_p = \begin{bmatrix} x_e - x_w \\ y_e - y_w \\ z_e - z_w \end{bmatrix} \quad (1)$$

$$\vec{v}_a = \begin{bmatrix} x_a - x_e \\ y_a - y_e \\ z_a - z_e \end{bmatrix} \quad (2)$$

\vec{v}_p と \vec{v}_a の成す角 θ がしきい値 θ_{th} 以下である時、家電選択が行われたとみなす。 θ は式 (3) のように求められる。

$$\theta = \arccos \left(\frac{\vec{v}_p \cdot \vec{v}_a}{|\vec{v}_p| \cdot |\vec{v}_a|} \right) \quad (3)$$

家電操作の開始と終了を切り替えるジェスチャの認識するための式を設定した。 \vec{v}_p と地面とのなす角度が 1.25rad 以上 1.58rad 以下となるように腕を上げ、手首が肘より上に来るジェスチャを行うと家電操作が始まる。 \vec{v}_p と地面とのなす角度が 1.25rad 以

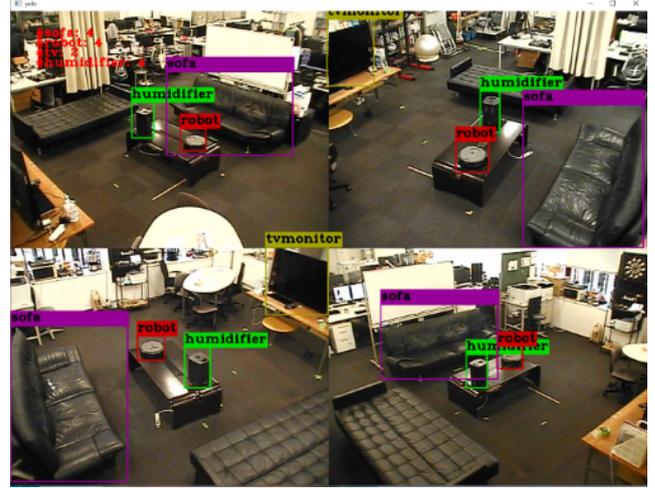


Fig.4: Object detection results: bounding boxes

上 1.58rad 未満で、手首が肘より下にあるとき、家電操作を終了するジェスチャである。

また、腕差しでの家電選択の判定を行う式を設定した。ユーザーと家電の距離が遠くなるほど、ユーザから見て家電は点に近似できるため家電の中心を指しやすくなる。そのため、家電と右肘の距離 $|\vec{v}_a|$ の値に応じて θ_{th} の値を変化させた。この変化を式 (4), 式 (5) のように家電ごとに経験的に設定した。ここで、本システムで操作する家電は TV と加湿器である。

$$(TV)\theta_{th} = 0.50 - 0.1 \cdot |\vec{v}_a| \quad (4)$$

$$(Humidifier)\theta_{th} = 0.46 - 0.05 \cdot |\vec{v}_a| \quad (5)$$

家電選択状態が任意の n フレーム連続で維持されると、家電の電源操作が行われたとみなす。

3 実験

本システムの認識精度とユーザビリティを評価する実験を行った。実験は、前腕を挙上できる成人 10 名の被験者を対象に行った。本手法の簡単な説明の後、実験を開始した。

実験 1 では、家電操作開始と終了を切り替えるジェスチャの認識率を評価した。また、実験 2 では、腕で家電を指し示すことで家電の電源を ON/OFF する操作の認識精度を評価した。さらに、家電の電源動作のために、家電の選択状態を連続して維持するフレーム数を 6 フレームとした。いずれの実験でも、被験者には図 5 のようなフィードバック画面を提示した。左上に赤い文字で controlling と現れると、操作開始状態に切り替えられたことを表している。また、青いバウンディングボックスが家電の周りに現れることで電源操作を行ったとみなしたことを示す。最後に、実験 3 では、SUS (System Usability Scale[22]) を用いて本システムのユーザビリティを評価した。

被験者は図 2 に示す、位置 A、位置 B からそれぞれ操作を行う。



Fig.5: Feedback screen

3.1 実験 1

家電操作開始・終了を切り替える操作を位置 A, 位置 B それぞれで起立状態で 10 回ずつ行い, 5 秒以上経過して認識されなければ操作失敗とみなした。

各操作の位置ごとの平均認識率を表 1 に示す。

Table 1: 操作開始・終了の切り替え操作の平均認識率 [%]

	位置 A	位置 B	平均認識率
操作開始	89	73	81
操作終了	87	70	78.5

操作開始ジェスチャは, 位置 A での認識率が 89%, 位置 B での認識率が 73%, 平均で 81% となった。操作終了ジェスチャは, 位置 A での認識率が 87%, 位置 B での認識率が 70%, 平均で 79% であった。

以上のことから, 家電操作の開始・終了を切り替える操作は高確率で正しく認識されており, 誤操作を防ぐことが出来ると考えられる。一方で, 位置 B での認識率が位置 A に比べやや低い理由としては, 図 2 の Camera3 からではユーザの身長によっては右前腕がユーザの頭部によって隠れてしまい, うまく認識されない場合があったためであると考えられる。

3.2 実験 2

次に, 実際に TV と家電を指示し, 電源の操作を行った。起立状態, 着席状態それぞれで家電を腕指しし, 5 秒以上経過して認識されなければ, 操作失敗とみなした。これを位置 A, 位置 B で行った。

位置ごとの各操作の平均認識率を表 2 に示す。

Table 2: 電源の操作の平均認識率 [%]

	位置 A	位置 B	平均認識率
TV 操作 (起立)	57	75	66
TV 操作 (着席)	34	61	47.5
加湿器操作 (起立)	95	72	83.5
加湿器操作 (着席)	92	88	90

位置 A で起立したまま TV を操作した際の平均認識率は 57%, 位置 B でだと 75% となり, 平均認識率は 66% であった。位置 A で着席したまま TV を操作した際の平均認識率は 34%, 位置 B でだと 61% となり, 平均認識率は 48% であった。また, 位置 A で起立したまま加湿器を操作した際の認識率は 95%, 位置 B でだと 72% となり, 平均認識率は 84% であった。次に, 位置 A で着席したまま加湿器を操作した際の認識率は 92%, 位置 B でだと 88% となり, 平均認識率は 90% であった。着席時の TV 操作を除いた 3 つの場面で平均認識率は 50% を超える結果となった。

表 2 から, 加湿器の電源操作に比べ, TV の電源操作がかなり低い認識率であると分かる。これは, TV のほうが加湿器よりもユーザにとって指示しにくかったためであると考えられる。実験中, 被験者が TV に腕差しを行う際に, 指示方向をたびたび変えている場面があった。そのため, 家電が大きくなると中心を捉えにくくなるため, 認識率が下がったと考えられる。

また, 両位置において TV 指示時に, 起立時より着席時のほうが認識率が下がっていることが分かる。これは, 視線が上下することで, 同じところを指示しているつもりでも, 手首から肘へのベクトル \vec{v}_0 と, 肘から家電へのベクトル \vec{v}_1 の角度 θ が変動し, 指示判定の範囲から出てしまうためと考えられる。

加えて, TV 操作において, 起立時着席時ともに, 位置 A より位置 B でのほうが認識率が高いことがわかる。これは, ユーザが目線の変化を踏まえた操作に慣れて上達したためと考えられる。

3.3 実験 3

本システムのユーザビリティ評価のために, 被験者に対してアンケート調査を行い, SUS を用いて評価した。

この調査により, システムや製品に対するユーザの主観的満足度を評価することができ, 全体的な満足度評価を行うことができる。

アンケートでは, 「全くそう思わない」を 1 とし, 「全くそう思う」を 5 とした, 5 段階の尺度で表したリッカート尺度 [23] を用

い, 各設問に対して 5 段階のどのレベルに該当するかを被験者に回答してもらった。

アンケートの内容は以下のとおりである。

Q1. このシステムをしばしば利用したいと思う。

Q2. システムが不必要に複雑だと思った。

Q3. 使いやすいシステムだと思った。

Q4. このシステムを使うために, 技術者のサポートが必要だと思う。

Q5. いろいろな機能がうまく統合されていると思った。

Q6. このシステムは矛盾が多いと思った。

Q7. ほとんどの人がすぐに使いこなすことができると思う。

Q8. 非常に使いにくいシステムだと思った。

Q9. 私はこのシステムを使うことに非常に自信を持った。

Q10. このシステムを使いこなすには, いろいろなことを学ぶ必要があった。

SUS においての, 被験者ごとのスコアの算出方法は以下の通りである。

- 奇数番号の設問の回答スコアから 1 を引く。
- 偶数番号の設問の回答スコアを 5 から引く。
- 変換後のスコアを全て足し合わせて 2.5 倍する。

以上の方法を用いて得られた計算結果を表 3 に示す。

Table 3: 被験者ごとのスコアと全体のスコア [点]

a	b	c	d	e	f	g	h	i	j	平均
52.5	85	77.5	75	75	90	65	82.5	70	50	72.75

Sauro は 5000 件以上の SUS スコアの測定値に基づいて, SUS とパーセンタイルの関係を明らかにしたデータから, SUS スコアの平均値は 68.1 点であることを導いた [24]。本手法の SUS スコアは 72.75 点であるとわかり, 68.1 点を上回る値であることから, 優れたユーザビリティのシステムであると言える。

4 結論

本研究では, 覚えられる範囲の複数のジェスチャを用いて, 少ないジェスチャ数で操作が可能かつ, 家電の位置や大きさの事前情報が不要なシステムを提案した。本システムは, ユーザが前腕で家電を指示し続けることで家電の電源操作を行うことができるシステムである。また, 誤操作を防ぐために, 家電操作の開始と終了を切り替えるシステムも設けている。実験により, 提案手法の評価を行った。

今後の課題と展望を以下にまとめる。

- 高機能化とユーザビリティ

本研究では, 家電の操作は電源操作のみである。今後, より多様な操作を行うために, ジェスチャの種類や手順が増えていくとユーザが煩わしく感じることは容易に想像できる。そこで, 多くの機能を利用しながら高いユーザビリティを維持するために, 今よりも自然で継続して行いやすいジェスチャや, 初めてシステムを使うユーザにも分かりやすい操作方法を導入する必要がある。

- 認識精度向上とフィードバック

本研究で行った実験で, 着席時の TV 操作の平均認識率は 47.5% であり, これは実用的とは言えない結果であると考えた。認識精度を向上させるためには, 前述の操作方法の改良だけでなく, 操作時にユーザに提示するフィードバックを改善し, よりユーザに伝わりやすいものにするのが重要であると考えられる。

- 移動する物体への適用

近年では, 自動で掃除を行うロボット掃除機や, ユーザの身体負担を減らすために身の回りの世話をを行うことを目的とされたロボットが開発され, 徐々に普及している。そこで, ロボットとのインタラクションとして本手法を利用することができる。本研究での操作対象は静止している家電であったが, 今後は移動し続ける物体に対して物理的な指示による遠隔操作の研究も進めていきたい。

参考文献

- [1] 内閣府. 令和 4 年版高齢社会白書(全体版), 2022. https://www8.cao.go.jp/kourei/whitepaper/w-2022/html/zenbun/s1_1_1.html, (2023/1/26).
- [2] J. R. B. Bodollo, J. Daniel V. Cortez, E. R. P. Maraya, E. V. Navarro, R. Q. L. Saquing and R. E. Tolentino, "Selection of Appliance Using Skeletal Tracking and 3D Face Tracking for Gesture Control Home Automation," 2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE), Bangalore, India, pp. 1-7, 2019.
- [3] H. Asano, T. Nagayasu, T. Orimo, K. Terabayashi, M. Ohta and K. Umeda, "Recognition of finger-pointing direction using color clustering and image segmentation," The SICE Annual Conference 2013, Nagoya, Japan, pp. 2029-2034, 2013.
- [4] A. I. D. Viaje, P. S. Bernardo, K. N. Manuel, G. M. Pacheco, K. -R. C. Barroma and R. E. Tolentino, "Selection of Appliance Using Skeletal Tracking of Hand to Hand-tip for a Gesture Controlled Home Automation," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, pp. 575-580, 2020.
- [5] A. Fernández, L. Bergesio, A. M. Bernardos, J. A. Besada and J. R. Casar, "A Kinect-based system to enable interaction by pointing in smart spaces," 2015 IEEE Sensors Applications Symposium (SAS), Zadar, Croatia, pp. 1-6, 2015.
- [6] M. A. Iqbal, S. K. Asrafuzzaman, M. M. Arifin and S. K. A. Hossain, "Smart home appliance control system for physically disabled people using kinect and X10," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, pp. 891-896, 2016.
- [7] 齋藤雄太, 金鎔煥, 片山喜章, "ジェスチャを用いた正確な家電制御を実現するためのインタラクティブな家電操作インタフェースについて", 電子情報通信学会技術研究報告, 2018-UBI-57(13), pp. 1-8, 2018.
- [8] M. S. Verdadero, C. O. Martinez-Ojeda and J. C. D. Cruz, "Hand Gesture Recognition System as an Alternative Interface for Remote Controlled Home Appliances," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, pp. 1-5, 2018.
- [9] S. Kaur, Anuranjana and N. Nair, "Electronic Device Control Using Hand Gesture Recognition System for Differently Aabled," 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, pp. 371-375, 2018.
- [10] Y. Muranaka, M. Al-Sada and T. Nakajima, "A Home Appliance Control System with Hand Gesture based on Pose Estimation," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), Kobe, Japan, pp. 752-755, 2020.
- [11] T. -H. Tsai, Y. -J. Luo and W. -C. Wan, "Live Demonstration: Home Appliance Control System with Dynamic Hand Gesture Recognition base on 3D Hand Skeletons," 2022 IEEE 4th International Conference on Artificial Intelligence Circuits and Systems (AICAS), Incheon, Korea, Republic of, pp. 503-503, 2022.
- [12] R. A. Urme, N. S. Prome and T. Ahmed, "Hand Gesture-Based Home Automation System," TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON), Hong Kong, Hong Kong, pp. 1-5, 2022.
- [13] Y. Muranaka, M. Al-Sada and T. Nakajima, "A Home Appliance Control System with Hand Gesture based on Pose Estimation," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), Kobe, Japan, pp. 752-755, 2020.
- [14] 入江耕太, 若林直弘, 梅田和昇, "ジェスチャ認識に基づくインテリジェントルームの構築," 日本機械学会論文集 C 編, Vol.73-725, pp.258-265, 2007.
- [15] A. Dongre, R. Pinto, A. Patkar and M. Lopes, "Computer Cursor Control Using Eye and Face Gestures," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, pp. 1-6, 2020.
- [16] B. Singh, N. Kandru and M. Chandra, "Application control using eye motion," 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Greater Noida, India, pp. 206-210, 2014.
- [17] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 779-788, 2016.
- [18] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021.
- [19] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, "Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao," arXiv preprint arXiv:2004.10934, 2004.
- [20] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár, "Microsoft COCO: Common Objects in Context," 2014.
- [21] Richard Hartley and Andrew Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [22] Brooke J. "SUS—a quick and dirty usability scale," Usability Evaluation in Industry, pp. 189-194. 1996.
- [23] Likert, R. "A Technique for the Measurement of Attitudes," Archives of Psychology 140, 55, 1932.
- [24] J. Sauro; "A Przsctical Guide to the System Usability Scale," Measuring Usability LLC, 2011.