

人物追跡と腕指しによる複数人物下での家電操作システム

○横田 雅恵 (中央大学), Pathak Sarthak (中央大学), 梅田 和昇 (中央大学)

Home Appliance Operation System by Person Tracking and Arm Pointing in Multi-person Environment

○ Masae YOKOTA (Chuo University), Sarthak PATHAK (Chuo University), and Kazunori UMEDA (Chuo University)

Abstract: In recent years, there has been a lot of research into methods of operating home appliances based on gesture recognition. In this method, it is important that the gestures used for operation are intuitive and simple enough to be repeated on a daily basis. Conventional systems that allow users to operate power supplies by pointing at home appliances with their arms have the problem of operating poorly in an environment where multiple people are on camera. In this study, we solved this problem by identifying and tracking the operator.

1. 緒言

日常生活において、我々は多くの家電製品を利用している。家電製品一つに対して一つのリモコンで操作を行うのが一般的であるが、家電製品の多機能化により、リモコン上のボタン数が増加し、ボタンのサイズや文字が縮小化している。近年では音声認識による操作手法が普及しているが、声を出せない状況や周囲の環境音などに認識率が左右されてしまう課題がある。これにより、ハンドジェスチャなどを用いることで、指先を上手く動かせない人や声を出しにくい状況下の人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。

ハンドジェスチャでの家電操作システムの中で、本研究では腕指し指示による家電選択操作を用いる。腕や指での指示による家電操作システムの先行研究において、家電の位置を事前にデータとして用意していることが前提になっている^{[1][2]}。しかし、その前提では頻繁に場所の変わる家電には対応することが出来ない。そこで、以前より本研究では物体検出アルゴリズムを用いることでその課題に対応している^[3]。

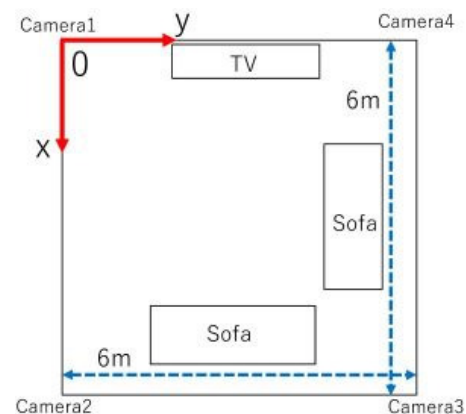
一方で、人の姿勢を算出するために骨格点検出アルゴリズムを用いているが、複数人が撮像内に映っている際に操作者を特定することが困難であるという課題があった。そのため、複数カメラ画像間で操作者を特定し追跡する必要がある。そこで、本研究では物体検出アルゴリズムを利用することで、操作者を特定・追跡し、よりロバストな家電操作システムを構築した。

2. 提案手法

2.1 本システム概要

本システムの環境は Fig.1 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。提案手法の大まかな流れを Fig.2(a) に示す。天井四隅のカメラから得られた 4 枚の画像を結合した画像に対して、YOLO^[4] を用いて家電と人の検出を行う。検出されたオブジェクトはバウンディングボックス (Bbox) で囲まれ示される。操作者を特定した後は 15 フレームに一度 OpenPose^[5] を用いて人物の骨格点推定を行い、操作者の姿勢計算を行う。これにより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、腕指しによる指示を終えると家電操作を終了したとみなし、再び挙手するまでは家電操作を行うことが出来ないようにすることで誤認識を防ぐ。実際の複数人状況下で本システムを用いた画像は 3 に示す。



(a) The overview of the room

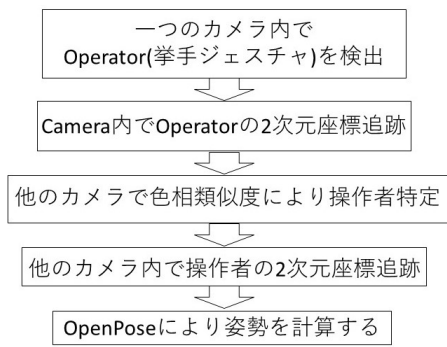


(b) The captured Image

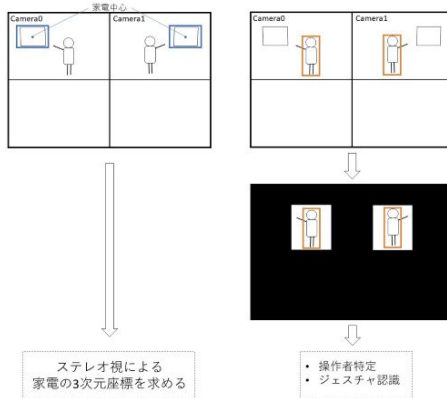
Fig.1 Environment of this study

2.2 操作者特定・追跡

挙手ジェスチャを操作者特定のジェスチャとする。camera0 からの画像内で挙手ジェスチャが認識される



(a) Flowchart of this tracking system



(b) Crop process from combined image

Fig.2 Process of this system



Fig.3 Image of system execution in a multi-person situation

と operator Bbox が描画される。数フレーム連続で中心座標のフレーム間ユークリッド距離がしきい値以下である operator Bbox 内の人物を操作者とみなし、操作者の operator Bbox の中心の 2 次元座標と、中心周辺の矩形領域の色相をそれぞれ保存する。camera0 内の操作者が挙手ジェスチャをやめると person Bbox が描画され、直前のフレームでの operator Bbox の中心座標と person Bbox の中心座標との距離がしきい値以下の時、その person Bbox を操作者のものと特定する。続いて、camera1 のすべての person Bbox の中心周辺の矩形領域の色相と、はじめに保存した操作者の色相の類似度を計算し、しきい値以上の Bbox の人物を操作者と特定する。以降は camera0, camera1 それぞれで操作者と特定した Bbox の中心座標のフレーム間距離から操作者の追跡を行う。

追跡している Bbox をそのまま切り出してしまうと、身体の一部が写っていない場合があるため、可能な限り、切り出す領域を Bbox の縦のサイズを基準とした正

方形領域とする。この領域に OpenPose を用いて、操作者の腕差し方向を計算し、家電選択状態を判定する。本論文では、肘から手首へのベクトルを腕差し方向とする。

3. 実験

実験環境は Fig. 1 のとおりである。30 秒程度の動画を用いて実験を行った。本実験を通してパラメータ決定と、それぞれのカメラでの追跡精度を評価した。以下の値を明らかにした。

- 1) カメラ間で操作者を特定するのに用いる色相類似度のしきい値
- 2) camera0, camera1 それぞれで操作者を追跡するための Bbox 中心のフレーム間移動距離のしきい値
- 3) それぞれのカメラでの 2 次元追跡精度

3.1 カメラ間で操作者を特定するのに用いる色相類似度のしきい値

camera0 で特定した操作者を camera1 内で見つけるために、camera0 内操作者の person Bbox の中心周辺の矩形領域の色相を保存し、基準とする。camera1 内のすべての person Bbox の中心周辺の矩形領域の色相と保存した色相とで類似度を計算し、しきい値以下の person Bbox を操作者のものと特定している。この際の色相類似度のしきい値を実験によって決定する。本実験では 20×40 の領域を切り取った。

まず、保存した領域の色相と、camera1 内のすべての person Bbox の中心周辺矩形領域の色相の類似度を算出し、0.1 ごとの幅で各値の出現頻度を比較したところ、Fig. 4 のようになった。これにより、しきい値となる値の範囲を絞ることが出来た。

この結果より、色相類似度が 0.8 以下の際の HSV 画像の目視による色相類似度の比較を行ったところ、Fig. 5 のように色相類似度の変化と HSV 画像の違いを確認できた。この結果より、操作者特定に用いる色相類似度のしきい値は 0.45 と決定した。

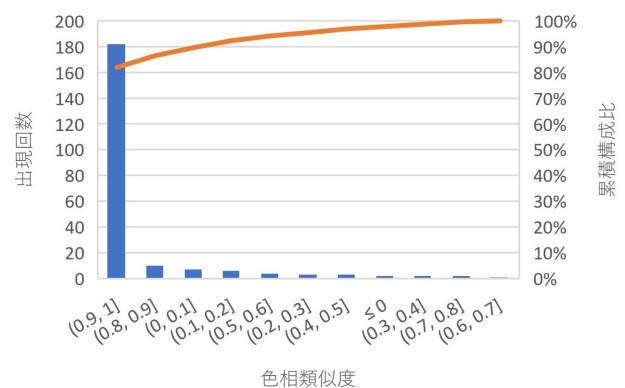


Fig.4 Pareto plot of hue similarity

3.2 Bbox 中心のフレーム間移動距離のしきい値

前後のフレーム間のみで Bbox 中心の移動距離をプロットしたものが Fig. 6 である。この結果から、フレーム間での Bbox 中心の移動距離が 50 以下であるとき、同一人物の Bbox であるとする。この時、身体の向きがカメラに対して正面を向いている状態かた横を向

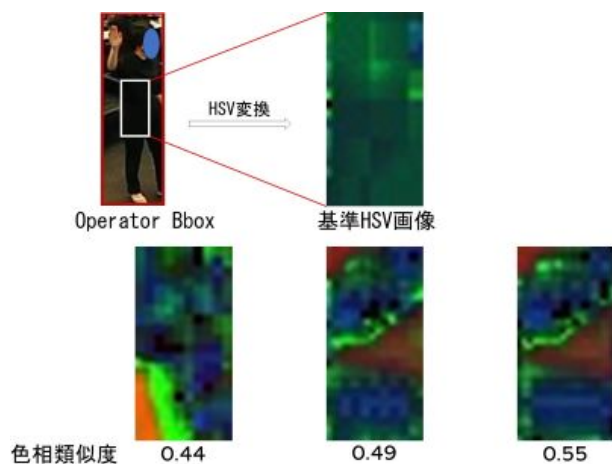


Fig.5 Comparison of HSV images and hue similarity

いている状態に切り替わることで Bbox 中心の移動量が大きくなっていることを確認できた。また、急な歩行速度の変化も、Bbox 中心のフレーム間移動量に影響を及ぼしていると考えられる。

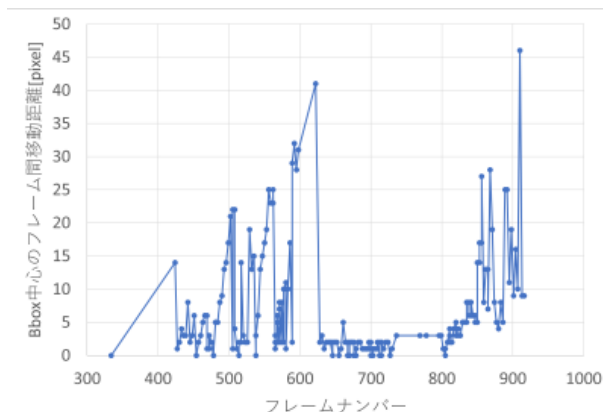


Fig.6 Bbox center shift between frames

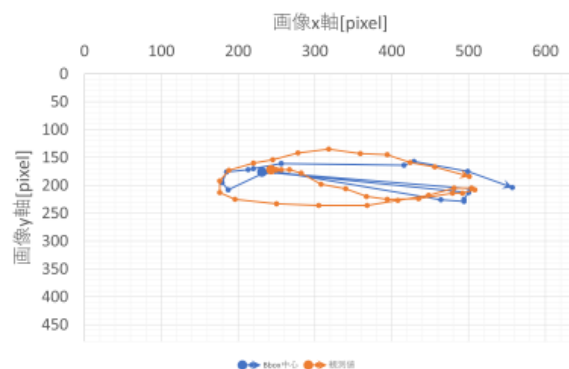
3.3 それぞれのカメラでの追跡精度

Camera0, Camera1 内での操作者の追跡が出来ていることを確かめるため、操作者の Bbox 中心と、15 フレームに一度操作者の身体中心手動でマーキングし、それぞれ撮像内の軌跡を描画したものを Fig. 7 に示す。青色の線が Bbox 中心の軌跡をプロットした線、オレンジ色の線が手動で身体中心の軌跡をプロットした線である。横軸、縦軸は共にカメラ画像の x 軸、y 軸と対応している。

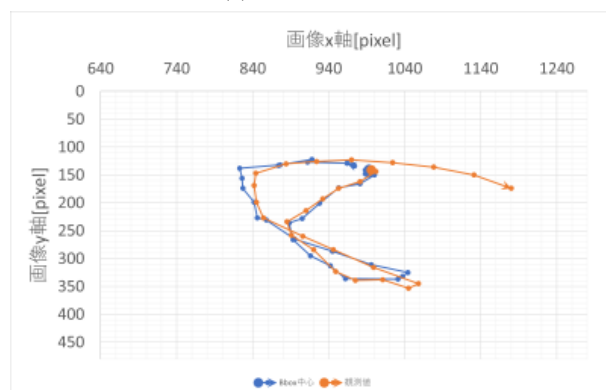
Bbox 中心と観測した身体中心のユークリッド距離を誤差とすると、Camera0 での平均誤差は 56.80 かつ標準偏差は 70.02, Camera1 での平均誤差は 55.79 かつ標準偏差は 50.61 であった。この誤差の原因としては、足元などが隠れている時の Bbox の縦方向の幅が小さくなり、Bbox の中心が実際の身体中心よりもずれてしまうことが考えられる。また、標準偏差が大きい原因としては、移動速度の急な変化によるものと考えられる。

4. 結言

本論文では、複数人状況下でもロバストなジェスチャ認識を行うことが出来るシステムを実現するため



(a) In Camera 0



(b) In Camera 1

Fig.7 The trajectory of the center of the Bbox and the observed trajectory of the body center for each camera

の、物体検出アルゴリズムによる人物追跡手法の実験を行った。実験により、別カメラ画像間での操作者特定に用いる色相類似度のしきい値、追跡に用いる Bbox 中心のフレーム間移動量のしきい値と、それぞれのカメラでの Bbox 中心による追跡精度を明らかにした。これらの結果から、今後の課題として、以下が挙げられる。

- 1) 操作者特定における色相類似度がしきい値よりも下回った際の冗長化の手段。
- 2) Bbox 中心のフレーム間移動量の急な歩行速度の変化への対応手段。

今後は以上の二点を明らかにし、混雑状況下でも対応できる腕指しによる家電操作システムを実現する。

参考文献

- [1] 片山喜章 齋藤雄太 金銘煥: “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 = *IEICE technical report*: 信学技報 117.451, pp. 89–96, (Feb. 2018). ISSN: 09135685. URL: <https://cir.nii.ac.jp/crid/1520009410126176896>.
- [2] Ana Fernández et al.: “A Kinect-based system to enable interaction by pointing in smart spaces”. *2015 IEEE Sensors Applications Symposium (SAS)*. 2015, pp. 1–6. DOI: 10.1109/SAS.2015.7133613.

- [3] Masae Yokota et al.: “Intuitive Arm-Pointing Based Home-Appliance Control from Multiple Camera Views”. *The 32nd IEEE International Conference on Robot and Human Interactive Communication*, (2023).
- [4] Joseph Redmon et al.: “You Only Look Once: Unified, Real-Time Object Detection”. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [5] Zhe Cao et al.: “OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.1, pp. 172–186, (2021). doi: 10.1109/TPAMI.2019.2929257.