7Q-03

低コスト手振り家電操作AIシステムの開発

伊藤　優太†　　久米　陽弾‡　　ラシキア　ジョージ¶

中京大学

1. はじめに

近年の新型コロナウイルスの影響により，非接触型のユーザーインタフェースが注目を集めている．特に，手話を代表とする複雑な意思疎通が可能なハンドジェスチャは，スマートウォッチの操作や医療現場でのモニター制御など，多岐にわたる応用が見込まれている．しかしながら，これらのシステムには電気信号を認識するための特定の装置や，Kinect や Leap Motion などのセンサが必要となる課題がある．ハンドジェスチャ認識（HGR）においては，Haar 特徴や HOG 特徴などの古典的な画像処理手法(1)（2）(3)が提案されているが，これらには手動で作成された特徴量と分類に依存する課題が存在する．その結果，複雑な環境下でのパフォーマンスの安定性に課題が生じる．特に，画像ソースのばらつきや照明の不均一性により，システムの性能が不安定になることがある．HGR システムにおいて最も重要なことは，優れたロバストな特徴表現の選択であり，手動による特徴設計は非常に難しい．加えて，認識可能なジェスチャ数の制約や，新しいジェスチャの追加に関する手間の問題も取り上げられている．

近年，深層学習（DL）に基づく HGR の数々の提案がされている．これらの提案の多くは，手の姿勢認識を対象とした静的ジェスチャ（3）(4)と手の動きを捉える動的ジェスチャ（5）(6)がある．静的ジェスチャはジェスチャ数に制約があり利便性に欠ける一方で，動的ジェスチャは RNN や 3DCNN といった処理が複雑で高コストな GPU などの機材が必要となる課題がある．このような課題に対応する手法として，静的ジェスチャの追跡を組み合わせたハイブリッド HGR（7）に注目する．

今回提案する手法は，低コストなハードウェアのみを用いて実用的なシステムを構築することを目的としている．このため，低コストな Raspberry Pi4 と Pi カメラをハードウェアとして選択した．Raspberry Pi 上での CNN 利用に伴い制限付きメモリと低速なプロセッサの問題が挙げられる．

小林氏が提案した MobilenetV3Small-Yolov3（MNSY）を用いた HGR 家電操作システム（8）は Yolov3 の backbone を MobilenetV3Small に置き換え，同モデルの Head に Depthwise Separable Convolution を適応したモデルでパラメータ削減による軽量化と高速化を可能にしたモデルであるが，実際に移動ジェスチャを利用する際は肩を軸にした回転運動によるジェスチャの動作が多く行われ，手の側面が多く映るが，ここでの精度が低下し動作に問題が生じるという課題が存在する．

本研究では手の側面の画像データを追加し MNSY の再学習を行う．加えて新たな最先端モデルに対し，ファインチューニングを行う．また，最先端モデルにモデルの軽量化技術を適用して独自モデルの作成と学習を行う．学習したモデルの容量，速度，精度の比較を行い最適なモデルでシステムを構築する．誤認識抑制のためにデータの後処理にも注目し，対策を提案する．

2. 提案手法

〈2･1〉 提案モデル　本システムの目的は処理が軽量で実用的なシステムの開発である．そのため，動的 HGR よりも処理の軽量なハイブリッド HGR に注目した．実装のために物体のクラスと位置を返す物体検出ニューラルネットワークを利用する．データセットは手の側面を含むデータセットを作成し，近年発表されたモデルについて Fine-tuning を行い，パフォーマンス比較を行うことで，システムに最適なモデルを選択しリアルタイムハンドジェスチャ家電操作システムの開発を行う．また，システムの面でも改良を行い，精度の安定性を図る．本研究では小林氏の提案した MNSY の再学習を行う．近年注目されている YOLO の最新モデルである YOLOv5，YOLOv6，YOLOv8 にも注目し学習を行った．また，YOLOv6，YOLOv8 のモデルの軽量化のために Ghost モジュール（9）を適応したモデルを新たに提案し，それぞれ YOLOv6‐Ghost，YOLOv8‐Ghost と呼ぶ．MNSY を参考にし，Backbone を MobileNetV3Small に置き換え Head に Depthwise Separable Convolution を適応したモデルを新たに提案しそれぞれ YOLOv6 - MobileNetV3，YOLOv8 - MobileNetV3 と呼ぶ．以上のモデルを COCO データセットで事前学習させたのち，Fine-tuning を行った．

〈2･2〉 学習　モデルを学習するために Creative Senz3d Dataset(10)（11)をもとに手の画像を集めた．Creative Senz3d Dataset に含まれていない，３の姿勢の手を自作で集めた． MNSY は手の側面の認識率が低下する問題があったため，手の側面の画像も集めた．クラ

「Development of a low-cost hand gesture appliance operation AI system using object tracking technology」
†「Yuta Ito・Chukyo University」
‡「Hibiki Kume・Chukyo University」
¶「George Lashikia・Chukyo University」

スは 12 種類にし，画像の枚数は合計約 36,000 枚を用意した．学習データに 8 割，検証データに 1 割，テストデータに 1 割を使用した．ニューラルネットワークの作成には PyTorch を使用した．

〈2･3〉 モデルの評価 実験を Raspberry Pi4 上で行った． OS は Raspberry Pi OS を使用した．今回は認識速度，認識精度，容量，についてモデルの比較を行った．また，低スペックデバイスでの高速な実行を実現するため，tensorflow-lite を利用しモデルの最適化を行ってから比較を行った．YOLOv5 はモデルの標準バージョンは精度が高いが，パラメータ数が膨大であるため高速に実行することのできる YOLOv5-n を比較対象とした．同じ理由で YOLOv6 は軽量モデルである YOLOv6-n を選択し，YOLOv8 は YOLOv8-n を選択した．

精度評価に mean Average Precision (mAP)の 0.5-0.95 を使用した．実行速度を評価する指標として一般的に使われる frames per second (fps) を使用した．得られた実験結果を Table 1 に示す．結果，YOLOv8-n MobileNet は mAP0.5-0.95 が 74.358，処理速度が 5.025fps と精度と速度のバランスが良いため，システムに YOLOv8-n MobileNet を利用した．

Model	params	mAP0.5-0.95	FPS
MNSY	19014434	75.045	0.428
YOLOv5-n	2505284	73.83	3.922
YOLOv6-n	4234932	75.597	2.882
YOLOv6-n ghost	2733748	73.312	4.149
YOLOv6-n MobileNet	1252450	72.4	5.917
YOLOv8-n	3013172	74.997	3.650
YOLOv8-n ghost	2728372	74.875	3.745
YOLOv8-n MobileNet	2345954	74.358	5.025

表 1 ：Table 1. Performance Comparison Among Different Models

〈2･4〉 提案システム 本システムに用いる主な部品は，Raspberry Pi4，Pi カメラ，irMagician である．全体構成を図 1 に示す．

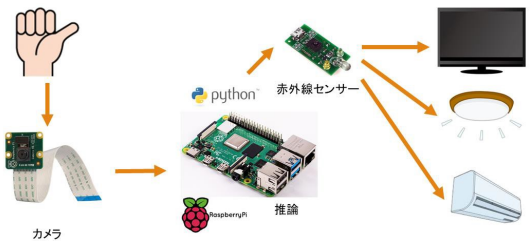

図 1 ：全体構成

なお，本システムはすべての処理を Raspberry Pi 上で完結するため，ネットワーク環境は必要ない．本システムを動作させると，カメラが起動し，撮影された映像のリアルタイム物体認識が行われる．ジェスチャを認識すれば一定時間同行を追跡し，対応する赤外線信号を出力する．赤外線との対応付けはシステム内の GUI により設定できる．ジェスチャは複数フレームを用いて決定している．また，誤認識抑制とよりスムーズなジェスチャ認識の為しきい値の調整を行った．最初のしきい値を高い値にして誤認識を抑制し，動的ジェスチャ中は cosine カーブに従い移動距離に応じて徐々にしきい値を下げることでジェスチャ中の認識はずれの問題を解消した．

3 まとめ

本研究で家電操作システムの開発を行った．提案手法として DL に注目し，安価なデバイスで実行できるモデルを比較し，本システムに最適なモデルを採用した．誰でも気軽に利用できるようにエンドツーエンドリアルタイムシステムを開発した．開発したシステムは GitHub で公開する予定である．

参考文献

[1] 牛丸太希, 佐藤一誠, 中川裕志, “3 次元 Haar 特徴量を用いたハンドジェスチャ認識”, 研究報告数理モデル化と問題解決 (MPS), 2014.

[2] 山下大輔，間博人，山本泰士，本田雄亮，三木光範，“モバイル端末のアプリケーション利用時における内蔵照度センサを用いたハンドジェスチャ認識手法の提案”，情報処理学会論文誌 vol.59, no.2, pp. 715-722, 2018

[3] H. Lahiani and M. Nejib, “Hand Gesture Recognition Method Based on HOG-LBP Features for Mobile Devices”, Procedia Computer Science, vol. 126, pp. 254–263, 2018.

[4] S. Ameen and S. Vadera, “A Convolutional Neural Network to Classify American Sign Language Fingerspelling from Depth and Colour images”, Wiley Expert Systems, 2016.

[5] V. Adithya and R. Rajesh, “A Deep Convolutional Neural Network Approach for Static Hand Gesture Recognition”, Procedia Computer Science, Elsevier, vol. 171, pp. 2353-2361, 2020.

[6] P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree and J. Kautz, “Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3d Convolutional Neural Network”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[7] O. Köpüklü, N. Köse and G. Rigoll, “Motion Fused Frames: Data Level Fusion Strategy for Hand Gesture Recognition”, IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.

[8] 小林 優太, ラシキア 城治, “低コストのディープニューラルネットワークベース家電操作用リアルタイム手振り認識システム”, 情報処理学会論文誌 vol.141, no.7, pp. 822-831, 2021.

[9] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, Chang Xu “GhostNet: More Features from Cheap Operations” 2019

[10] A. Memo, L. Minto and P. Zanuttigh, “Exploiting Silhouette Descriptors and Synthetic Data for Hand Gesture Recognition”, STAG: Smart Tools & Apps for Graphics, 2015.

[11] A. Memo and P. Zanuttigh, “Head-mounted Gesture Controlled Interface for Human-computer Interaction”, Multimedia Tools and Applications, 2017.