

# FMCW レーダーに基づく距離アウェアなハンドジェスチャ認識

飯田 耀介<sup>†</sup> 範 明陽<sup>†</sup> 馬 建華<sup>†</sup>

法政大学大学院 情報科学研究科<sup>†</sup>

## 1. はじめに

ハンドジェスチャは、人がコミュニケーションを取るための手段として、話すことと同様のように頻繁に使われている。近年では、機械が手話やデバイス操作を認識するために、ハンドジェスチャ認識の研究が盛んに行われている。課題点として、ウェアラブルデバイスでは装着の手間が掛かる上に拘束感があり、音声やカメラ認識では発話が困難な人やプライバシーを気にする人への配慮が必要である。

一方で、周波数変調連続波 (FMCW) レーダーは上記に配慮した上で、単一の非接触デバイスとして細かい動きを検出できる。しかし、既存研究の実験方法は、レーダーと人の位置を固定している場合[1]、位置や角度を変更しても人の手のみをレーダーに写るように調整している場合[2]がほとんどである。そこで本研究では、レーダーから人までの距離を変更した場合でも、距離の違いやジェスチャを分類できるような処理や識別方法を提案する。

## 2. レーダーの概要とデータ処理

FMCW レーダーとは図1のように、送信アンテナから連続して発信される変調的な周波数の波と、それが何らかの物体に反射して戻り受信アンテナに着信した時の波の差から中間周波数 (IF) 信号を取得する。この IF 信号を処理することにより、レーダー波に反射した物体に関する距離や速度などの情報を得ることができる。また、一連の変調的な周波数の振幅をチャープ信号と言い、この振幅の大きさを時間関数として扱う。

次にデータ処理について説明する。取得した IF 信号に高速フーリエ変換 (Range-FFT) を施し、時間と距離データを取得する。このデータのチャープ信号をさらに FFT 処理 (Doppler-FFT) を施すことで、速度と距離データが取得できる。この時、特定の距離における速度データを取得するために、短時間フーリエ変換 (STFT) を施すことで、時間と速度データが取得できる。最終的に、約3秒間のジェスチャをスペクトログラム画像として深層学習の構築のために利用する。

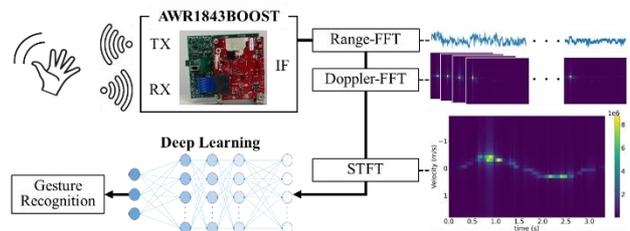


図1 データ処理から深層学習までの流れ

## 3. データ取得実験

スマートフォン、テレビ、そしてカーナビといった IT 機器に向けた非接触入力インターフェースを想定して、図2に示すように8種類のジェスチャを用意した。これらのジェスチャはそれぞれ25回ずつ行い、1回の実験で合計200データを取得できる。レーダーと人の手の位置を比較する実験では30cm, 40cm, 50cm, 60cm, 90cmの距離で実施し、合計で1,000データ収集した。実験は、図3に示すように Texas Instruments 社の AWR1843BOOST を使用し、人と対面するように設置した。このレーダーの周波数帯域は76-81GHzと高く、細かい動きの特徴検出が期待できる。また FMCW レーダーのデータ転送を補助する役割として、DCA1000EVM を補填している。

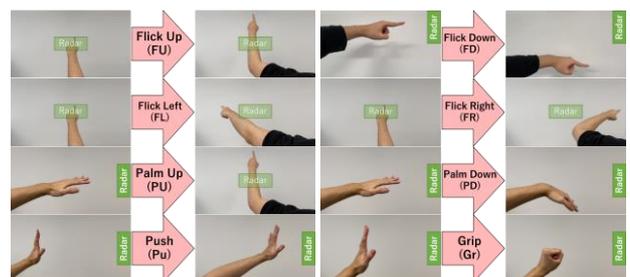


図2 実験で用いる8つのジェスチャ

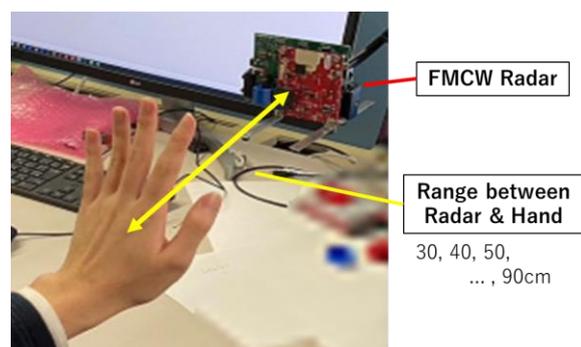


図3 実験環境の例

FMCW Radar Based Range-aware Hand Gesture Recognition  
Yosuke Iida<sup>†</sup>, Mingyang Fan<sup>†</sup>, Jianhua Ma<sup>†</sup>  
Graduate School of Computer and Information Sciences,  
Hosei University<sup>†</sup>

#### 4. 深層学習モデルの構築

入力データは RGB 画像であり、ピクセル数が 496×369 であるが、処理コストを低減するために約 10 分の 1 の 50×37 のサイズに縮小した。データセットは学習データ、検証データ、テストデータを 8 対 1 対 1 の割合で分割し、深層学習モデルの認識精度を全てのデータが一度はテストデータとなるように 10 分割交差検証から評価した。モデル生成のために、学習データと検証データを用いて、画像認識に強みを持つ畳み込みニューラルネットワーク (CNN) から学習させた。より具体的なモデルの構造定義は、図 4 が示す通りである。畳み込み層の第 1 引数はフィルタ、第 2 引数はカーネルである。Dense 層の第 1 引数もフィルタを示し、最終的なフィルタはその時に学習させる分類数と等しくさせた。また、過学習を抑えるために Dropout を適宜使用している。この影響により学習が鈍化することを踏まえ、エポック数は 300 回に設定した。

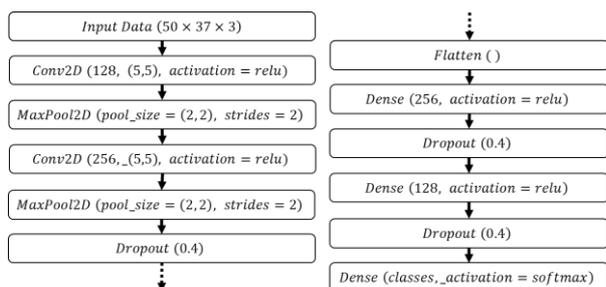


図 4 モデルの構造定義

#### 5. 推論結果

学習モデルの推論について、その学習に関与していないテストデータを用いて推論を行う。その結果について、表 1 の左側に各ジェスチャのデータのみと全てのジェスチャを含んだ ALL データを利用した距離分類、右側に各距離のデータのみと全ての距離を含んだ ALL データを利用したジェスチャ分類の平均精度、表 2 の混同行列には ALL データの混同行列をそれぞれ示す。

これらの表から考察を行う。まず、どちらの ALL データも各分類の平均を取るような精度である。5 種類の距離分類の精度について、FLICK ジェスチャ群は 90%以上であり、その他は約 80-90%であることがわかる。混同行列から特徴は、距離が中心であるほど精度が下がることがわかる。8 種類のジェスチャ分類の精度について、40cm の距離が最も高く、それを中心に距離が離れるほど精度も下がることがわかる。混同行列からは、FLICK ジェスチャ群が約 90-95%、それ以外は約 95-100%の精度であることがわかり、距離分類の特徴とは逆になった。

表 1 平均精度のまとめ  
距離分類 (左) とジェスチャ分類 (右)

Range Classification	ACC (%)	Gesture Classification	ACC (%)
ALL	88.8	ALL	94.1
FLICK UP	94.3	30cm	92.5
FLICK DOWN	92.3	40cm	95.0
FLICK LEFT	90.0	50cm	94.2
FLICK RIGHT	96.7	60cm	94.0
GRIP	82.7	90cm	92.1
PALM UP	80.3	Ave.	93.7
PALM DOWN	89.7		
PUSH	85.7		
Ave.	88.9		

表 2 ALL データによる混同行列の精度結果  
距離分類 (左) とジェスチャ分類 (右)

	30cm	40cm	50cm	60cm	90cm	Acc.
30cm	93.5	1.0	2.5	1.5	1.5	93.5
40cm	2.0	85.5	10.0	2.0	0.5	85.5
50cm	5.0	8.0	78.5	8.0	0.5	78.5
60cm	2.0	1.5	6.0	90.5	0.0	90.5
90cm	2.5	1.5	0.0	0.0	96.0	96.0
ALL						Ave. 88.8

	FD	FL	FR	FU	Gr	PD	PU	Pu	Acc.
FD	94.4	1.5	2.5	0.0	1.5	0.0	0.0	0.0	94.4
FL	1.6	87.9	6.4	3.3	0.0	0.0	0.8	0.0	87.9
FR	0.0	7.2	86.4	5.6	0.0	0.0	0.0	0.8	86.4
FU	0.0	4.0	4.9	90.3	0.0	0.0	0.8	0.0	90.3
Gr	0.8	0.0	0.0	0.0	99.2	0.0	0.0	0.0	99.2
PD	0.8	0.0	0.0	0.0	2.3	95.3	1.6	0.0	95.3
PU	0.0	0.0	0.0	0.0	0.0	0.8	99.2	0.0	99.2
Pu	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
ALL									Ave. 94.1

#### 6. まとめと今後の展望

本研究は、レーダーと人との距離が異なる場合での距離の違いやジェスチャを分類し、その精度結果を比較した。また、分類時に用いるデータを分散することで、高精度を維持すると同時に、計算コストを削減できることを示した。本研究の課題として、被験者数の不足、10cm 単位ではなく数 cm 単位での変化への対処法、精度向上のためにさらなる調整、もしくは他の認識技術の利用が考えられる。

今後、本研究のような FMCW レーダーを用いたハンドジェスチャ認識の研究を発展させるために、レーダーと人との距離だけでなく、縦横の角度や操作速度の比較、被験者非依存の検証、そしてこれらを組み合わせた実験が必要である。それによって、堅牢性のある認識技術とそれをを用いたアプリケーションが実現できる。

#### 参考文献

[1] P. Zhao, C. Lu, B. Wang, N. Trigoni and A. Markham, "CubeLearn: End-to-end learning for human motion recognition from raw mmWave radar signals", Computer Science, Machine Learning, Signal Processing (eess. SP), arXiv:2111.03976, 7 November 2021.

[2] W. Jiang, Y. Ren, Y. Liu, Z. Wang and X. Wang, "Recognition of dynamic hand gesture based on mm-wave FMCW radar micro-doppler signatures", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4905-4909, 13 May 2021.