

# マルチモーダルセンシングに基づく料理中のマイクロ行動認識の提案

## A Proposal for Micro Activity Recognition during Cooking Based on Multimodal Sensing

石山 時宗<sup>†</sup>      松井 智一<sup>†</sup>      藤本 まなと<sup>†</sup>      諏訪 博彦<sup>‡</sup>      安本 慶一<sup>†</sup>  
Tokimune Ishiyama   Tomokazu Matsui   Manato Fujimoto   Hirohiko Suwa   Keiichi Yasumoto

### 1. はじめに

居住者のヘルスケアや宅内サービスの高度化を目的として、人感センサや環境センサなどの設置型センサや電力情報などのデータを基に、居住者の宅内行動認識を行う研究が多数提案されている。特に、入浴や料理、食事、睡眠などの日常生活行動の認識は、居住者の生活リズムを自動的にデータ化でき、健康寿命延伸のための行動変容やヘルスケアに有用である。

しかしながら、従来提案されている日常生活行動の認識手法では、入浴や睡眠行動などの行動は場所と行動が対応づけられており、1日の中で行われる時間が大まかに決まっていることから、人感センサやドアセンサなどの設置型センサで比較的容易に認識が可能である。一方で、料理や食事、PC 仕事、読書などの行動は、1日のうちに行われる時間が定まっておらず、さらに行われる場所も他の行動と重なる場合が多いため、設置型センサでは認識が難しいという課題がある。

ウェアラブルセンサやカメラを利用した行動認識手法も存在する。ウェアラブルセンサの利用に関して、近年 Apple Watch をはじめとした高機能スマートウォッチや、Fitbit などのフィットネストラッカーが販売されており、各種センシングに利用されている。また、カメラの行動認識への利用に関して、一般的なプライバシー問題への対策として、画像内の RGB 情報を削除し骨格情報を抽出する手法が存在する。しかしながら、ウェアラブルセンサを常に装着する手間や、カメラを設置すること自体に対する忌避感などの問題から、宅内行動認識のためのセンサとして利用することは難しい。

本研究では、ピエゾセンサを利用した振動センシングにより得られたデータと、いくつかの設置型センサから得られたデータを分析し、従来認識が難しかった日常生活行動の認識を目指す。その中でも、特に料理行動のマイクロ認識を行うことを目的とする。料理行動を対象とする理由は、料理行動が、食材を「切る」、「焼く」、「煮る」、「炒める」行動や、「煮っぱなし」、「下ごしらえ」行動などの時間的に連続して行われない行動をはじめとしたマイクロ行動から構成されており、従来の日常生活行

動認識の枠組みでは認識が難しいためである。

なお、振動情報のみでは食材を「焼く」、「煮る」行動のセンシングが難しいことが考えられる。そのため、人感センサや環境センサなどのセンサを補助的に利用したマルチモーダルなセンシングによって、振動センサのみでは取得が難しい行動の認識を行う。

本論文では、様々な行動認識手法と、料理に関する行動認識の関連研究を紹介し、その課題について述べる。次に、マルチモーダルセンシングに基づく料理のマイクロ行動認識手法と実験計画について説明する。最後に、本論文の結論をまとめる。

### 2. 関連研究

本章では、行動認識に利用される代表的なセンサであるカメラを用いた行動認識に関連する既存研究と、本研究で取り扱う振動センサに基づく行動認識に関連する既存研究、および本研究の対象行動である料理行動に関する既存研究について述べる。

#### 2.1 カメラを利用した行動認識

撮影された画像を利用した行動認識に関する研究は多数行われている [1, 2]。画像に基づく行動認識の代表的なアプローチとして、CNN をはじめとした深層学習モデルを利用した行動認識や、被撮影者のプライバシーに配慮するために映像内の人物の骨格情報のみを算出し、関節の位置・加速度情報などから行動認識を行う研究が存在する。[3, 4] しかしながら、カメラで撮影できる情報は限られており、細かい動作を網羅的にセンシングするためには多数のカメラが必要となる。また、撮影した情報から骨格情報を抽出するなどして、プライバシーを考慮した行動認識を行う場合においても、宅内にカメラを設置し撮影し続けることは、カメラ設置に対する居住者の抵抗感が課題となる。

#### 2.2 振動センサを利用した行動認識

振動センサを利用した行動認識に関する研究も多数行われている。VibroSense[5] は、レーザードップラー振動センサによりセンシングした振動データをディープラーニングにより解析し、ドライヤーやシャワーなどの、18 種類の行動を認識するシステムを提案している。この際、天井のある一点にレーザードップラー振動センサを設置することで、微少な振動を測定する。OAC[6] では、机

<sup>†</sup> 奈良先端科学技術大学院大学, Nara Institute of Science and Technology

<sup>‡</sup> 理化学研究所, RIKEN

の上にセンサーを設置し、周囲の構造物の振動を計測することで歩く、タイピング、話すなどの活動を識別する。振動センサを利用した手法は、振動はテーブルや床などの構造物を伝わるため、カメラなどを利用する手法と比較して死角が存在しないという利点を持つ。一方、欠点として、構造物と接触を必要としない行動は原理的に区別することが不可能である。

### 2.3 料理行動に関する研究

数ある日常生活行動の中でも、料理行動を扱う研究は多数提案されている。例えば、調理時のマイクロ行動認識による最適調理手順作成法が提案されている [7]。この提案では、調理に不慣れな初心者が調理時間を短縮するために、複数レシピの並行調理をタスクスケジューリング問題として捉え、タスクスケジューリング問題の最適化アルゴリズムを用いて、最適調理手順を作成している。画像により調理行動を識別する研究 [8] では、CNNにより調理作業のマイクロ行動認識手法を提案している。提案では、対象とする調理作業として「調理器具を取り出す」「食材を取り出す」「洗う」などを想定している。また、深度カメラ KinectV2 を用いて、調理作業の認識を行っている研究 [9, 10] がある。RGBD カメラの深度データから計算した骨格情報をもとに、調理者のキッチンにおける相対位置および胴体と上肢の位置関係・角度を算出する。そして、これらを特徴量として機械学習を行うことで、5種類の調理作業を平均 72.3% の精度で分類可能である。これらの提案では、設置型センサを利用していることから、ユーザーの行動を阻害せずに調理を行うことが可能である。しかし、「洗う」のように正解率の高い動作がある一方で、「食材を取り出す」や「粉をまぶす (49.5%)」のように正解率の低い行動がある。

## 3. 提案システム

本章では、提案するマルチモーダルセンシングシステムについて述べる。まず初めに、提案システムの概要と使用するセンサについて説明した後、本システムの動作フローについて説明する。

### 3.1 提案システムの目的と概要

ピエゾセンサなどいくつかのセンサから得られマルチモーダルなデータを分析し、従来では難しかった料理のマイクロ行動認識を行うことを目的とする。図 1 に、本研究で目標とするマルチモーダルセンシングシステムを示す。振動や画像などの多数のモーダルの計測を行い、機械学習モデルにより入力され、処理される。出力として、行動が識別される。それぞれの結果により、アンサンブル学習を行った出力を、識別結果とする。

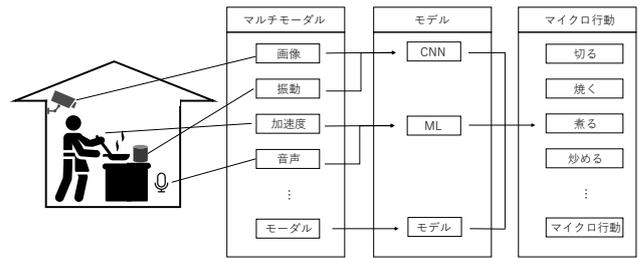


図 1: マルチモーダルセンシングシステム

### 3.2 使用するセンサ

本節では、提案システムで使用するセンサについて説明する。料理行動におけるマイクロ行動を認識するため、提案システムでは、主に 3 つのセンサ（振動センサ、人感センサ、環境センサ）を使用する。振動センサは、主に料理行動時に発生する振動をセンシングするために使用する。また、人感センサと環境センサは、振動センサでは取得できない位置情報や周辺の環境情報をセンシングするために利用する。以下に、使用するセンサについて具体的に述べる。

■ **振動センサ** 料理行動を行う際に発生する多種多様な振動を取得するための振動センサとして、ピエゾセンサを利用する。本研究において使用するピエゾセンサは、直径 5cm 程度の大きさで、非常に薄型であるため、環境に溶け込むように配置でき、また、多少の障害物が周辺に存在しても、回折効果によって振動を比較的捉えやすいという利点を有する [11]。本センサは、FET（電界効果トランジスタ）とオペアンプから構成される増幅回路に接続して使用するものとし、料理行動によって引き起こされる振動を効果的に取得できるようチューニングされたものを利用する。

■ **人感センサ** 人感センサは、人体から発せられる赤外線を検知することで人間の存在を検知する。本研究では、検出範囲が扇形に広がる焦電型赤外線センサモジュールを複数使用し、被験者の位置（例えば、調理場にいるのかいないのか、など）を検出する。検出したデータは、EnOcean を用いて、サーバにデータをアップロードする。

■ **環境センサ** 環境センサは、調理場に 1 つ設置し、様々な環境データを収集するために用いる。本研究で使用する環境センサは、温度、湿度、照度、気圧、騒音などの情報を取得できる。加熱調理器具による温度変化や、水道の使用による湿度の変化、作業時に発生する音などの検知を行う。環境センサには、通信規格として、BLE（Bluetooth Low Energy）モジュールが搭載されている。

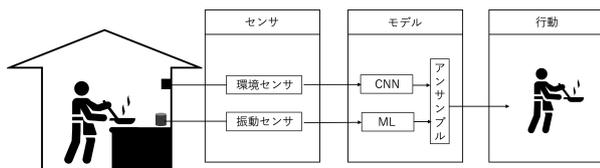


図 2: 実験システム

ため、BLE 規格のワイヤレス通信を用いて、センシングデータをサーバにアップロードする。

■ その他 上記で説明したセンサ以外にも、例えば、行動時の動作速度を検知可能なドップラーセンサ [12] や、対象の表面温度の 2 次元画像を取得するためのサーモカメラの使用を想定する。

### 3.3 動作フロー

図 2 に、使用する実験システムの概要図を示す。システムは、ユーザの行動を認識するため、様々なモダルのセンシングを行う。センシングされたデータは、1 つのサーバに集約される。集約されたデータは、前処理されたあと、機械学習モデルの入力とする。例えば、行動により引き起こされる振動は、ピエゾセンサにより計測され、短時間フーリエ変換により 2 次元画像として CNN に入力する。機械学習モデルは、それぞれ独立して学習され、出力は多数決を取るというアンサンブル学習を行うことを想定している。学習されたモデルにより、ユーザのマイクロ行動を推定する。

## 4. 実験方法

### 4.1 概要

本研究では、マルチモーダルセンシングによる料理中のマイクロ行動の認識を目指す。被験者の行動は、複数のセンサを使用して、対象となる活動をセンシングする。センシングされたデータにより学習された機械学習モデルを、クロスバリデーションを用いて評価を行う。

### 4.2 実験条件

実験は、奈良先端大内に設置されたスマートホームで行うことを想定している。スマートホームは、365 日いつでも In-the-wild な環境を想定した実験ができるため、本研究の実験場所として適している。具体的な実験としては、3 つのセンサ（振動センサ、人感センサ、環境センサ）を使用して、対象のマイクロ行動である、「切る」、「焼く」、「煮る」、「炒める」行動や、「煮っぱなし」、「下ごしらえ」を識別する。被験者が行う 1 時間程度の料理行動のセンサデータを取得する。5 人の被験者を対象として、1 つのレシピによる料理行動を使用し、行動時の

センサデータを収集する。特徴量としては、振動センサ、人感センサ、環境センサによるセンサデータを使用する。

### 4.3 評価方法

本システムで推定した行動の精度を検証するため、適合率、再現率、F 値を計算する。また、個人の行動のばらつきに対するロバスト性を評価するため、5 人の被験者で実験を行う。Leave One Person Out 交差検証を採用し、4 人分のトレーニングデータにより学習したモデルを 1 人のテストデータにより検証することを計画している。

## 5. まとめ

料理行動は、食材を「切る」、「焼く」、「煮る」、「炒める」行動や、「煮っぱなし」、「下ごしらえ」行動などの時間的に連続して行われない行動をはじめとしたマイクロ行動から構成されており、従来の日常生活行動認識の枠組みでは認識が難しい。本論文では、マルチモーダルセンシングを用いて、料理のマイクロ行動をセンシングするための新たな手法を提案し、実験計画について述べた。今後、実際に調理時のマイクロ行動認識システムを構築し、実証実験を行うことで、本研究を進めていく。

## 謝辞

本研究の一部は、科研費基盤研究 (B)(No.20H04177) の助成によって行った。

## 参考文献

- [1] Zhiwei Deng, Arash Vahdat, Hexiang Hu, and Greg Mori. Structure inference machines: Recurrent neural networks for analyzing relations in group activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4772–4781, 2016.
- [2] Muhammad Ehatisham-Ul-Haq, Ali Javed, Muhammad Awais Azam, Hafiz MA Malik, Aun Irtaza, Ik Hyun Lee, and Muhammad Tariq Mahmood. Robust human activity recognition using multimodal feature-level fusion. *IEEE Access*, Vol. 7, pp. 60736–60751, 2019.
- [3] Tsukasa Okumura, Shuichi Urabe, Katsufumi Inoue, and Michifumi Yoshioka. Cooking activities recognition in egocentric videos using hand shape feature with openpose. In *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*, pp. 42–45, 2018.

- [4] Sumaira Ghazal, Umar S Khan, Muhammad Mubasher Saleem, Nasir Rashid, and Javaid Iqbal. Human activity recognition using 2d skeleton data and supervised machine learning. *IET image processing*, Vol. 13, No. 13, pp. 2572–2578, 2019.
- [5] Wei Sun, Tuochao Chen, Jiayi Zheng, Zhenyu Lei, Lucy Wang, Benjamin Steeper, Peng He, Matthew Dressa, Feng Tian, and Cheng Zhang. Vibrosense: Recognizing home activities by deep learning subtle vibrations on an interior surface of a house from a single point using laser doppler vibrometry. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 3, September 2020.
- [6] Amelie Bonde, Shijia Pan, Mostafa Mirshekari, Carlos Ruiz, Hae Young Noh, and Pei Zhang. Oac: Overlapping office activity classification through iot-sensed structural vibration. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pp. 216–222, 2020.
- [7] Jin Nakabe, Teruhiro Mizumoto, Hirohiko Suwa, and Keiichi Yasumoto. Optimal cooking procedure presentation system for multiple recipes and investigating its effect. *Algorithms*, Vol. 14, No. 2, 2021.
- [8] Shinya Michibata, Katsufumi Inoue, Michifumi Yoshioka, and Atsushi Hashimoto. Cooking activity recognition in egocentric videos with a hand mask image branch in the multi-stream cnn. In *Proceedings of the 12th Workshop on Multimedia for Cooking and Eating Activities, CEA '20*, pp. 1–6. Association for Computing Machinery, 2020.
- [9] Teruhiro Mizumoto, Alberto Fornaser, Hirohiko Suwa, Keiichi Yasumoto, and Mariolino De Cecco. Kinect-based micro-behavior sensing system for learning the smart assistance with human subjects inside their homes. In *2018 Workshop on Metrology for Industry 4.0 and IoT*, pp. 1–6, 2018.
- [10] Alberto Fornaser, Teruhiro Mizumoto, Hiroiko Suwa, Keiichi Yasumoto, and Mariolino De Cecco. The influence of measurements and feature types in automatic micro-behavior recognition in meal preparation. *IEEE Instrumentation & Measurement Magazine*, Vol. 21, No. 6, pp. 10–14, 2018.
- [11] Keisuke Umakoshi, Tomokazu Matsui, Makoto Yoshida, Hyuckjin Choi, Manato Fujimoto, Hirohiko Suwa, and Keiichi Yasumoto. Non-contact person identification by piezoelectric-based gait vibration sensing. In *AINA (1)*, pp. 745–757, 2021.
- [12] Shinya Misaki, Keisuke Umakoshi, Tomokazu Matsui, Hyuckjin Choi, Manato Fujimoto, and K. Yasumoto. Non-contact in-home activity recognition system utilizing doppler sensors. *Adjunct Proceedings of the 2021 International Conference on Distributed Computing and Networking*, 2021.