

ヘッドマウントディスプレイを使用した ユーザのための調理動画の自動同期システム

森 亘佑† 河野恭之†

関西学院大学大学院理工学研究科†

1.はじめに

本研究は、ヘッドマウントディスプレイ(以下HMD)のみを使用して、調理動画を参考に調理を行うユーザのためにユーザと同じタイミングで進むように調理動画の一時停止と再生を行うことで調理の進行度を同期させるシステムの開発を目指す。現代ではインターネット上に多くの調理動画がアップロードされている。新型コロナウイルスの流行によって家庭で自炊を行う頻度が増加し[1]、アップロードされている動画を参考に調理する機会が増加してきていると考えられる。しかし、ユーザは調理という分野において初心者であるが故にその調理動画を視聴していることが想像でき、動画内で調理を行っているのは料理研究家など調理の専門的な知識を持っている人物であるため、一般のユーザに比べて手際が良い。このことから動画に対してユーザの進行度が遅れることが考えられる。アップロードされている動画は視聴者が見やすいように編集されている場合が多く、無駄な箇所はカットされているため、これらの条件下でユーザが動画と同じタイミングで調理を行うことは難しい。本研究ではこれらの問題を解決するため、調理動画を参考にしながら調理を行う初心者のユーザが動画から遅れたタイミングで動画を一時停止し、ユーザが動画に追いつくと再生することで動画とユーザの調理工程の同期を図るシステムの開発を目指す。火や刃物を使用するのでユーザが動画視聴のために視界を大きくうごかしてしまうと危険なため視聴する動画はHMD内で再生する。その例を図1に示す。左の画像がユーザの手元の画像で右が調理動画の画像である。



図1. 本研究の想定している実行例

2.先行研究

ユーザの調理をサポートする様々な研究が存在する。森岡ら[2]はカメラ2台とプロジェクタ3台、対話ロボットを使用してユーザにプロジェクタによるガイドと対話ロボットの指示で調理の支援を行った。しかし、カメラ2台とプロジェクタ3台、対話ロボットを使用するとデバイスの数が多くなってしまいコストがかかってしまうため一般家庭での運用は実用的ではない。是松ら[3]は、「カット」「グリル」「その他」の調理行動を行った際に発生する音を用いて推定した。この手法では環境音のノイズが存在する場合、その影響を大きく受けてしまい精度が安定しないことが考えられる。加藤ら[4]は包丁に取り付けた加速度センサと調理動画のメタデータを使用することで動画内とユーザの調理の同期を図った。しかし、包丁が関連した動作以外の認識ができないため全ての調理工程に対応することができない。本研究は、HMDから得られた映像と動画内の映像を比較して動画の停止・再生を行うため、低コストで周りのノイズに影響されることなく精度を保ったままユーザのサポートを行うことができる。

3.研究概要

本研究では、ユーザにHMDを使用してもらい、動画内の手元の映像とユーザの手元の映像を比較することで、ユーザが動画に対して遅れている場合に動画を停止させ、ユーザが動画に追いついた場合再生する。大井ら[5]は調理の際に出現する動作は「加える」「切る」「混ぜる」「焼く・炒める」の動作が数ある調理動作の中で特に多く出現するとし、調理動作の認識を行った。そこで本研究も同じ調理動作を用いて作られる料理を対象にユーザスタディを行う。本システムの流れを図2に示す。まず、ユーザにHMDを装着してもらい、ユーザの視界を映像として取得する。その得られた映像から調理中の食材の認識、調理に使用されている器具の認識を行う。動画内の映像にも同じ処理を実行し、比較することで停止するか決定し、停止した場合はユーザが停止した箇所まで追いついた時に動画を再生させる。なお、比較する際にはユーザは動画を見ながら調理を行うため動画に対して少しの遅れが発生することが予想される。そのため同じタイミングの映像を比較するのではなく調理中のユーザの手元の映像と、数秒前の動画の映像で比較を行う。

Automatic synchronization method of cooking videos using HMD for users

†Kosuke Mori, Yasuyuki Kono

Graduate School of Kwansai Gakuin University

動画に対する処理 ユーザの手元映像の処理

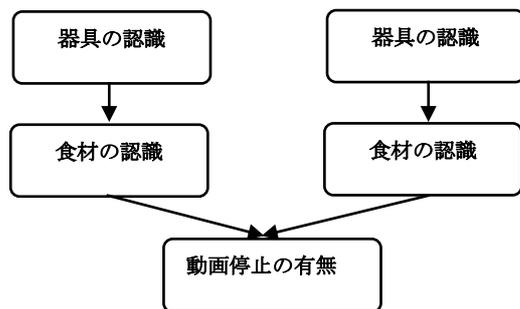


図 2. システムの流れ

3.1.調理器具の認識

調理器具の認識は DeepSORT[6]を用いて認識・追跡する。なお、DeepSORT のデフォルトの認識モデルに今回対象となる調理器具で含まれていないものがあるため、その調理器具は自前で用意した画像で学習を行い約 88%の精度で認識を行うことを可能とした。

3.2.調理食材の認識

調理中の食材がユーザのものと同動画内のものが同一あるか判別する必要があるため、セグメンテーションモデルを作成した。このモデルにより調理中の食材のマスク画像を生成しマッチングを行う。周ら[7]は、ユーザのジェスチャを利用して関心のあるオブジェクトのマスク画像を生成し、アノテーションを簡易化させた。これを参考に手元の画像と人の腕のセグメンテーション画像を一組として学習データ、正解データには注目している物体のセグメンテーション画像を入力しモデルを作成した。モデルの作成には約 6000 枚の画像を使用した。マッチングのアルゴリズムは AKAZE を用いる。AKAZE は回転拡大縮小に強いアルゴリズムで、本研究の環境ではユーザの手元と動画は撮影する画角が異なるためである。教師データのイメージを図 3、正解データのイメージを図 4 に示す。

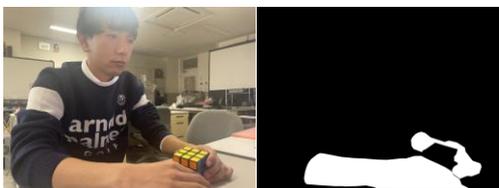


図 3. 教師データのイメージ

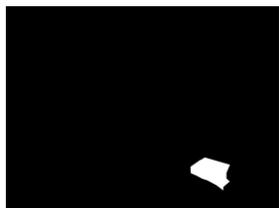


図 4. 正解データのイメージ

4. ユーザスタディ

実際にユーザにこのシステムを使用してもらい、評価を行う。今回は炒め物、透明もしくは半透明の汁物、色がついて具材の見えない汁物の 3 種類の料理を対象に行う。この理由は、料理の種類によって精度がどれほど変化しユーザにどのような感想を抱かせるか検証するためである。またユーザスタディには、普段から定期的に調理を行う人物とそうでない人物に協力してもらい感想の比較を行う。

5. おわりに

本稿では HMD を使用したユーザのための調理動画の自動同期システムを提案した。今後は発表に向け、前述したユーザスタディを実施し、評価を行うと共に課題点の考察、システムの質の向上を行っていく。

参考文献

- [1]Eczine 編集部: コロナ下で約 6 割が自炊増加、若年層ほど顕著 理由 8 割が「感染予防で外出控えた」/ポケットマルシェ調査|, Eczine 編集部(オンライン),入手先 <<https://eczine.jp/news/detail/8364>> (参照 2023-03-22)
- [2]森岡俊介, 上田博唯: カメラとプロジェクタと対話ロボットが連携する調理支援システム, Proceedings Human-Agent Interaction Symposium2011, (2011).
- [3]Korematsu, Y., Saito, D., and Minematsu, N.: Cooking state recognition based on acoustic event detection, CEA '19: Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities, pp41-44, (2019).
- [4]加藤岳大, 横窪安奈, ロペズギヨーム: SynCook: 動画メタデータと加速度センサを用いたレシピ動画進行度自動同期システム, Proceedings WISS 2020, (2020).
- [5]大井翔, 池ヶ谷剛, 佐野睦夫: Dp マッチングと移動方向ヒストグラムに基づく一人称視点調理行動認識. 画像電子学会誌, Vol. 46, No. 4, pp. 570-578, (2017).
- [6]Wojke, N., Bewley, A., and Paulus, D.: Simple Online and Realtime Tracking with a Deep Association Metric, Cornell University(online), available from <<https://arxiv.org/abs/1703.07402>> (accessed 2023-10-3).
- [7]Zhou, Z., Yatani, K.: Gesture-aware Interactive Machine Teaching with In-situ Object Annotations, Proceedings UIST2022, pp1-14, (2022).