

認知機能自動評価システムのための 一人称視点による調理動作認識

池ヶ谷剛^{†1} 大井翔^{†1} 佐野睦夫^{†1}

概要: 我々は認知症や高次脳機能障がい者に代表される認知障がいの日常生活の自立を促すために、調理行動を始めとする生活行動をリハビリ対象として、在宅で認知リハビリテーションを行う方式の確立に取り組んでいる。本稿では、特にリハビリ効果の高い調理における認知機能を評価するため、ウェアラブルカメラから取得した一人称視点映像を視覚情報とし、深層学習を用いた調理器具認識から得られる使用調理器具情報と腕の軌跡情報に着目した、調理動作認識手法について検討する。

Cooking Behavior Recognition based on First Person Vision for Cognitive Function Automatic Evaluation System

TSUYOSHI IKEGAYA^{†1} SHO OOI^{†1}
MUTSUO SANNO^{†1}

1. はじめに

近年、日本は高齢化社会を迎えており、認知症患者は年々増加傾向にある。厚生労働省のデータでは65歳以上の高齢者のうち、認知症患者は推計15%で、2012年時点で約462万人以上に上ると報告されている[1]。予備軍とされる「軽度認知障害」(MCI)の約400万人を加えれば、65歳以上の4人に1人が該当することになる。また、脳血管障害や脳症、脳炎などの病気や事故による脳外傷などによって脳が損傷されたために、認記憶障がい、注意障がい、遂行機能障がい、社会的行動障がいなどの症状を患った高次脳機能障がいの数は全国で約7万人と推定されており、将来的にはさらに増加すると考えられている[2]。一方で、これらの患者数に対して介護福祉士やリハビリテーションの専門家の人数が不足している状態が続いており、患者が十分な介護を受けることができないだけでなく、介護福祉士やリハビリテーションの専門家への負担が大きいのが近年のリハビリテーション業界の現状である。

我々はこれまでに、認知症や高次脳機能障がい者に代表される認知障がいの日常生活の自立を促すために、調理行動を始めとする生活行動をリハビリ対象として、「遠隔認知リハビリテーション」という新しい方式の確立に取り組んできた。また、振り返り支援として図1の様に自身の行動に対しての認知状態を点数化し、振り返りを行うことでリハビリへの意欲や自分自身の気づきに効果があることを検証してきた[3]。

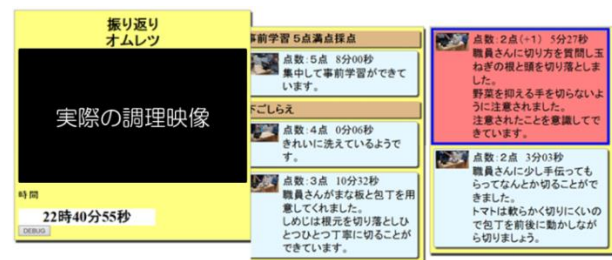


図1 振り返り支援システム

Fig.1 Reflection Support System

そこで我々は、認知リハビリにおける患者の行動をセンシングし、認知機能(注意力・記憶力・遂行力等)の評価を自動で算出する自動評価システムを目指している。従来の認知機能評価に以下のような問題点が存在する。

- ・リハビリ担当者が目視で行っているため負担が大きい
- ・担当者により評価のばらつきがある

そのため、継続的に在宅での認知リハビリテーションの実施が困難とされてきた。しかし、ウェアラブルデバイス等のセンシングによって得られた情報から認知機能評価の自動化を実現することで、介護福祉士の負担を軽減するだけでなく、患者の負担も軽減することにつながるため在宅でのリハビリテーションが可能となる。

本研究では、実践的で並行タスクが多く含まれ、注意力や記憶力、遂行能力など非常に高度な能力を要求されるため認知機能のリハビリ効果が高いとされている調理における認知機能を評価するために必須の調理動作認識手法について記述する。

^{†1} 大阪工業大学
Osaka Institute of Technology.

主な調理中における認知機能の例を以下に示す。

- ・遂行機能：患者自らレシピ(手順)を作成し、自身の計画が完遂できたかを評価
- ・注意機能：調理中の行動に着目し、注意ができていてかを評価（切っている最中によそ見をしていないか等）

こうした調理時における認知機能を評価するために本研究では、ウェアラブルカメラから取得した一人称視点映像を視覚情報とし、深層学習を用いた調理器具・腕領域認識から得られる使用調理器具情報と腕の軌跡情報の情報統合による調理動作認識手法の検討を行う。

2. 従来・関連研究

2.1 料理リハビリテーション

従来から取り組んでいる料理リハビリテーションの流れを図2に示す。認知障がいに応じた適切な料理行動をナビゲーションするシステムを用いて、調理を必要に応じてナビゲーションし、完成した料理を一緒に味わい、料理行動を振り返るサイクルを繰り返すという認知リハビリテーションを行い日常生活の自立を促進する研究を行っている[3]。

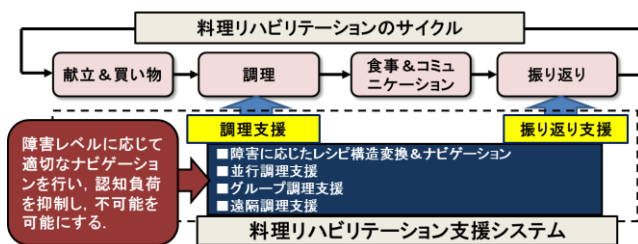


図2 料理リハビリテーションの流れ
Fig.2 Flow of Cooking Rehabilitation

2.2 調理センシングに関する研究

調理のセンシングに関する研究として、当研究室では、手首に装着した加速度センサから得られた加速度・角速度を用いた調理動作認識を従来から研究してきた。また、他には山肩らの研究[4]が挙げられる。山肩らはカメラやセンサから得られた観測データをもとに食材と調理動作を認識し、また各食材をトラッキングすることによりフロアグラフの構築を行い、同じくフロアグラフで表現されたレシピとマッチングすることでユーザが作っている料理のレシピの動的認識を行っている。しかし、上記の研究では、調理行動認識にはキッチン上に取り付けるサーマルカメラ・光学カメラ、食材認識にはRF-IDタグ、温度センサ、調理器具認識には3次元位置センサ等が必要となるが、これらの装置を一般の家庭で用意することは難しい。

そこで本研究では、なるべく調理者の負担を減らすため、

ヘッドマウントカメラのみを使用し、得られた情報によって調理動作認識を行う手法について提案する。

2.3 First Person Vision に関する研究

また、近年 Google Glass などのメガネ型ウェアラブルデバイスが注目を浴びており、一人称視点での映像認識についての研究が進んでいる。一人称視点による行動認識についての研究には、Sanath Narayan らの研究[5]や松尾らの研究[6]が挙げられる。松尾らの研究では物体検出と手動作検出により「Active object」か「passive object」に分類を行い、さらに顕著性マップ(saliency map)を用いた Visual attention により「salient object」または「non-salient objects」に分類し、そこからユーザが何を手に持ち、何処に注意を向いているかを判断し、ユーザの振る舞いの認識を行っている。他には、久賀らの研究[7][8]が挙げられる。久賀らは事前学習したディープニューラルネットワークを用いることにより、ユーザが利用したオブジェクトを一般物体認識により認識し、認識したオブジェクトの名前とあらかじめ定義した日常行動の名前との類似度を Web 上の知識を用いて計算することで、学習を必要としない行動認識を実現している。

これら一人称視点映像(First Person Vision)による行動認識の多くは「料理をする」や「テレビを見る」と言った大きな行動を認識対象としている。しかし料理とは、「切る」「剥く」「炒める」いった細かな動作により形成されている、そのため調理における認知機能を評価するためには、そのような調理動作を認識する必要がある。本研究では、ウェアラブルカメラを利用して、調理における基本動作である「切る」「剥く」「混ぜる」「炒める」「加える(入れる)」「(計量スプーンを用いて)加える」の6種類の一人称視点による調理動作認識を行う。

3. 提案手法

3.1 調理動作認識の枠組み

本研究では、調理動作において重要な情報である使用している調理器具と腕の動きに着目し、取得した映像(一人称視点)を視覚情報として腕の動き/調理器具の認識を行い、腕の動きからは調理動作の軌跡情報、調理器具からは使用している調理器具情報を取得する。そしてこれらの2つの特徴量を組み合わせ調理動作の認識を行う。また、本研究では上記で述べた調理器具特徴量と腕の動き特徴量を2つの統合手法を用いて調理動作認識を行い、どちらが調理動作認識において有効であるかを検証する。

まず、1つ目の統合手法は図3で示すような初期統合手法である。初期統合手法とは、入力特徴量の段階で統合を行う手法である。本研究での調理動作認識では、調理器具特徴量と腕の動き特徴量を結合し、調理器具・腕の動き特徴量

として認識を行い、特徴量レベルで統合し調理動作の認識を行う。

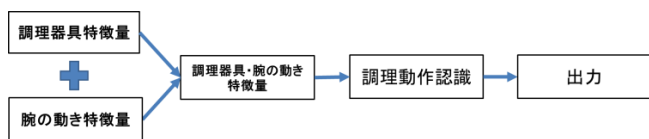


図 3 初期統合法

Fig.3 Initial Integration Method

2つ目の統合方法は、図 4 で示すような結果統合法である。別々の過程で処理を行い、結果を統合する手法である。本研究での調理動作認識では、調理器具認識による調理動作推定結果と腕の動きによる調理動作推定結果を結合することで調理動作の認識を行う。

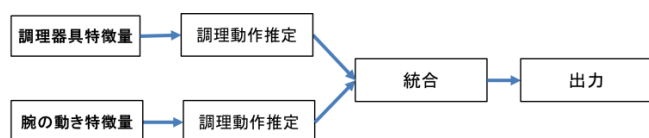


図 4 結果統合法

Fig.4 Result Integration Method

本研究では、調理における基本動作である「切る(Cut)」「剥く(Peel)」「(かき)混ぜ(Stir)」「炒める(Beat)」の 4 種類と、調味料等をそのまま「加える(Add)」動作と、調味料等を「(計量スプーンを用いて)加える(Add with Tool)」動作の 2 種類、計 6 種類の調理動作を本研究での調理動作認識における対象動作とした。

3.2 腕の軌跡情報

3.2.1 腕領域認識

本研究では、調理動作を認識する手法として腕の動き方に着目した。まず、腕の軌跡情報を取得するためには、腕領域を認識し、腕先を推定する必要がある。そこで、肌色抽出に加えて深層学習用のフレームワークである Chainer を用いた腕領域認識を行う。具体的には、画像データを RGB 色空間から HSV 色空間に変換し、肌色抽出後、収縮・膨張処理でノイズ除去を行い、ノイズ除去を行った画像に対してラベリング処理を行うことで腕領域候補を抽出する。この時点では、腕領域だけでなく肌色に近い物体等も腕領域候補として選出されている。そこで、選出された腕領域候補に対して深層学習を用いた画像認識によって「右手」「左手」「両手」「その他」に分類する。図 5 に用いた腕の学習画像の一部を示す。

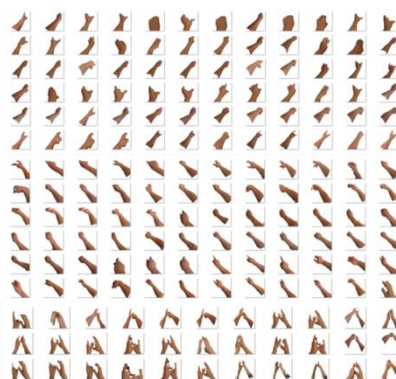


図 5 腕領域の学習画像の一部

Fig.5 A part of the Learning Images of Arm Area

3.2.2 腕先推定

本研究では、腕の軌跡情報を取得するために、認識した腕領域から右手・左手の腕先を推定し、推定した腕先の動きをそれぞれの腕の軌跡情報として調理動作認識に用いている。具体的には、認識した腕領域から回転を考慮した外接矩形を算出し、算出した外接矩形の頂点を腕先の推定結果とする。実際の腕先推定の結果を図 6 に示す。図に表示されている赤い点が推定した左腕の腕先であり、同様に青い点が推定した右腕の腕先である。

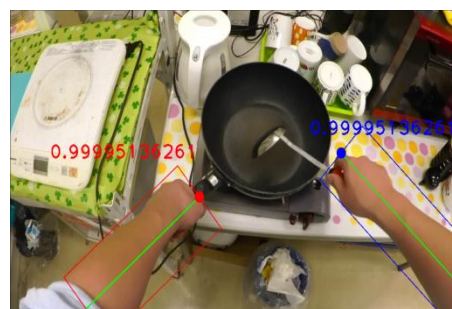


図 6 腕先の推定

Fig.6 Estimation of Arm Tip

3.3 調理器具情報

調理動作認識において、腕の軌跡情報だけではあいまいな腕の動きが起きた場合には、誤認識の確率が高くなってしまふ。そこで、本研究では使用調理器具が調理動作の認識において重要な情報であると考え、深層学習を用いた調理器具認識を行うことで使用調理器具情報を取得する。

3.3.1 調理器具認識

調理器具認識では、調理で主に使用する調理器具（まな板・包丁・フライパン・鍋・ピーラー・計量スプーン・お玉・調味料・その他）に対して、腕領域認識でも使用した深層学習用のフレームワークである Chainer を用いた調理器具認識を行う。

3.3.2 物体候補検出

物体認識を行うにあたりウェアラブルカメラから取得した画像では調理器具の位置や大きさは不明であり、また複数の物体が写っている可能性があるために物体候補領域検出を行う必要がある。こうした、物体候補領域検出を行うには一般的に、DeepLearningでの候補領域を選出するアルゴリズムの一つであるSelective Search[9]を用いてオブジェクト候補の抽出を行うことが多い。Selective Searchとはピクセルレベルで類似する領域をグルーピングしていくことで候補領域を選出する手法であり、処理対象となる領域が多くなりすぎる上に、対応できる形状やサイズに制限が発生するExhaustive Searchとは違い、似たような特徴を持つ領域を結合していき、1つのオブジェクトとして抽出することにより高速な物体候補領域検出が可能である。

しかし、Selective Searchでは候補オブジェクトが多いためまだまだリアルタイムでの検出は難しい。また、パラメータの調整が難しく動画像に対しては、安定した精度での物体候補の検出が行えない。

そこで、本研究において使用調理器具を認識するにあたり、一人称視点調理映像の「動作中は手が常に観測できる」「使用する物体（調理器具）は手元付近に存在する」といった特徴を利用し、手元付近を抽出することで、できるだけオブジェクト候補を少なくすることで処理時間の短縮を行う。つまり、3.2.2で述べた腕領域認識から求めた腕先座標を中心として、そこからある一定の領域内を調理器具認識における調理器具候補領域と設定する。

実際に取得した一人称視点画像に対して、Selective Searchによる候補領域認識結果と本研究で設定した調理器具の候補領域認識結果を図7に示す。

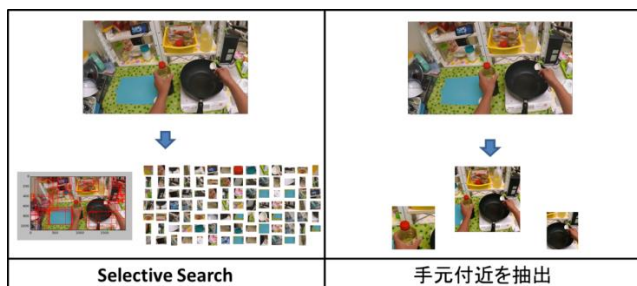


図7 調理器具候補領域の検出

Fig.7 Result of Candidate Area Recognition

そして、検出した調理器具の候補領域に対して深層学習を用いた調理器具の認識を行う。本研究では一人称視点での画像認識を行うため、データセットにはウェアラブルカメラから取得した一人称視点画像を利用することにより、一人称視点映像に強いデータセットの作成を行う。認識カテゴリは、(まな板・包丁・フライパン・鍋・ピーラー・菜箸・計量スプーン・お玉・コショウ・醤油・みりん・その他)の12種類とし、学習画像はそれぞれの調理器具に対し

て約3000枚とした。本研究で使用した、調理器具認識の学習画像の一部を図8に示す。



図8 調理器具認識の学習画像の一部

Fig.8 A part of the Learning Images of Cooking Utensils

3.4 初期統合法

初期統合法では、腕の軌跡情報と調理器具情報を入力特徴量の段階で統合し、調理器具・腕の動き特徴量として認識を行う。具体的には、図9で示すように腕領域認識から得られた腕先座標のシーケンスをフレーム間の角度に変換後、角度方向によって分類し、分類結果を腕の軌跡情報の特徴量とする。また、調理器具認識から得られたフレームごとの使用調理器具の結果を調理器具情報の特徴量とし、それぞれの特徴量を統合し、調理器具・腕の動き特徴量を求める。そして、統合した特徴量に対して隠れマルコフモデル(HMM)を用いて調理動作の認識を行う。

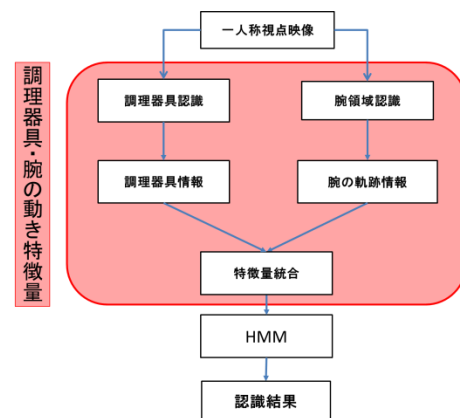


図9 初期統合法の流れ

Fig.9 Flow of Initial Integration Method

3.4.1 調理器具・腕の動き特徴量 -初期統合法-

本研究では、初期統合法においては腕の軌跡情報と調理器具情報を統合した調理器具・腕の動き特徴量として認識を行う。以下に「調理器具情報」と「腕の動き情報」それぞれの取得方法について詳しく述べ、最後に調理器具認識に用いる調理器具・腕の動き特徴量について記述する。

・調理器具情報

調理器具情報では、まず以下のようなそれぞれの調理器具に対応する要素をもつ配列を作成する。

[Other, Chopsticks, Kitchen knife, Ladle, Measuring spoon, Peeler, seasoning, Cutting Board, Frying pan]

そして、1 フレームごとの調理器具候補領域に対する調理器具認識結果をもとに認識した調理器具に対応する要素に 1 を代入、認識されていない調理器具は 0 と代入する。例えば、調理器具候補領域 3 か所の認識結果が「まな板」「包丁」「その他」であった場合、調理器具情報は[1, 0, 1, 0, 0, 0, 0, 1, 0]となる。なお本研究では、基本的に今回の認識対象である調理動作において複数同じ調理器具を使用することはないと仮定し、同じ調理器具が複数認識された場合でも、認識された調理器具の要素は 1 とした。つまり、調理器具候補領域 3 か所の認識結果が「まな板」「まな板」「包丁」の場合でも調理器具情報は[0, 0, 1, 0, 0, 0, 0, 1, 0]となる。

・腕の動き情報

腕の動き情報では、腕領域認識から得られた右腕・左腕それぞれの腕先座標のシーケンスをフレーム間の角度に変換後、角度方向によって 8 方向, 16 方向, 32 方向に分類し、その分類結果を腕の動き情報とした。

・調理器具・腕の動き特徴量

そして、上記で述べた調理器具情報と腕の動きを合わせたものを初期統合法による調理動作認識に用いる調理器具・腕の動き特徴量とした。つまり、調理器具・腕の動き特徴量は以下の特徴量として形成される。

- 調理器具・腕の動き特徴量
 [分類結果:右腕, 分類結果:左腕, Other, Chopsticks, Kitchen knife, Ladle, Measuring spoon, Peeler, seasoning, Cutting Board, Frying pan]

例えば、あるフレームにおいて、調理器具認識により調理器具候補領域 3 か所の認識結果が「菜箸」「フライパン」「フライパン」であり、腕領域認識により 16 方向による腕の動き情報が、16 方向:[12 (分類結果:右腕), 4 (分類結果:左腕)]と取得された場合、

[12, 4, 0, 1, 0, 0, 0, 0, 0, 0, 1]

がそのフレームにおける調理器具・腕の動き特徴量として取得する。つまり、本研究で使用する調理器具認識での調理器具・腕の動き特徴量は、先ほどの 1 フレーム間の調理器具・腕の動き特徴量を調理動作の間だけ取得した時系列データ集合となる。

3.4.2 HMM を用いた調理動作認識手法 -初期統合法-

本研究では、腕の軌跡情報と調理器具情報を統合した特徴量に対して隠れマルコフモデル(HMM)を用いて調理動作の認識を行う。まず、最初に HMM を用いた調理動作認識における学習工程について記述する。学習工程では、まず各調理動作に対して複数の学習データ(調理器具・腕の動き特徴量)を用意し、それぞれの調理動作ごとに HMM を作

成後、Baum-Welch アルゴリズムによってパラメータを学習する。そして、得られた複数の HMM のパラメータを平均化し、各調理動作の HMM モデルを作成する。ここまです、学習工程とし、次に HMM を用いた調理動作認識における認識手順について記述する。まず、学習工程で作成した各調理動作の HMM モデルを読み込む。そして、テスト用データ集合(調理器具・腕の動き特徴量)を読み込む。各調理動作の HMM でシーケンス毎に Viterbi アルゴリズムを適用し、それぞれの調理動作の HMM に対しての尤度を算出する。そして、最も尤度が高い調理動作の HMM を本研究での HMM を用いた調理動作認識手法、つまり初期統合法による認識結果とした。

3.5 結果統合法

結果統合法では、調理器具認識による調理動作推定結果と腕の動きによる調理動作推定結果を結合して、調理動作の認識を行う。具体的には、図 10 で示すように、腕の動き情報から調理動作推定を行い、調理器具情報からレシピ分析に基づく調理器具に含まれる動作情報をもとに調理動作推定を行う。そして、これらの情報の共起性をもとにして調理動作の認識を行う。調理動作と調理器具の共起性を利用するため、腕の動きによる調理動作推定に加え、調理器具に含まれる動作情報による調理動作推定により、調理動作認識を行う。調理器具に含まれる動作情報とは、包丁を認識した場合は切る動作である可能性が高いといった調理器具の動作割合を実際の調理映像から、調理器具がどのような調理動作で用いられているかを調理動作率として算出したものである。

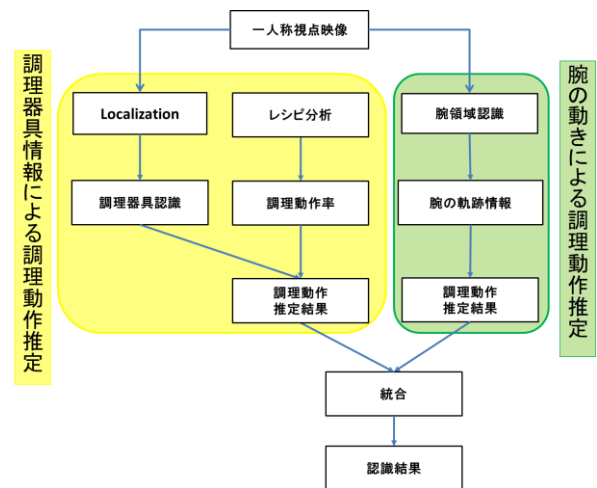


図 10 結果統合法の流れ

Fig.10 Flow of Result Integration Method

3.5.1 調理器具認識による調理動作推定 -結果統合法-

結果統合法では、調理器具認識による調理動作推定結果と腕の動きによる調理動作推定結果を結合して、調理動作の認識を行う。そのために、調理器具認識から調理動作を

推定しなければならない。そこで、本研究ではレシピ分析に基づくことで調理器具に含まれる動作情報を取得し、調理器具認識による調理動作推定を求める。以下に、調理器具に含まれる動作情報を取得するためのレシピ分析について、そしてレシピ分析から得られた調理器具に含まれる動作情報から調理動作を推定する方法について記述する。

・レシピ分析

調理において、使用する調理器具によってある程度の調理動作は推定することができる。例えば、まな板や包丁を使用しているならば切る動作・ピーラーを使用していれば剥く動作・フライパンを使用していれば炒める動作である確率が高くなるはずである。そこで本研究では、宮脇らが作成した調理ナビゲーション[10]にあるレシピ30種類(回鍋肉、肉じゃが等)のプロによる調理映像から調理器具がどの調理動作で使用されているかを目視で確認し、表1に示すような調理器具がどのように使われているかを分布にした調理動作率を作成した。

表1 調理動作率

Table 1 Cooking Behavior Distribution

	Add	Beat	Cut	Peel	Stir
Chopsticks	0.43	0.39	0.00	0.00	0.18
Kitchen Knife	0.00	0.00	0.76	0.24	0.00
Ladle	0.00	0.00	0.00	0.00	1.00
Measuring spoon	1.00	0.00	0.00	0.00	0.00
Peeler	0.00	0.00	0.00	1.00	0.00
seasoning	1.00	0.00	0.00	0.00	0.00
Cutting Board	0.00	0.00	0.62	0.38	0.00
Frying pan	0.32	0.68	0.00	0.00	0.00
Pot	0.38	0.00	0.00	0.20	0.42

・推定方法

調理器具認識により認識した使用調理器具に対して、上記で述べたレシピ分析から得られた調理器具による調理動作率をもとに調理器具認識による調理動作の推定結果を求める。例えば、調理器具認識の結果が「包丁」と「まな板」であった場合には表1より、
 「包丁」の調理動作率 (0.76, 0.24, 0.00, 0.00, 0.00)
 「まな板」の調理動作率 (0.62, 0.38, 0.00, 0.00, 0.00)
 を調理動作ごとにすべて足し合わせ、調理器具認識した枚数で割ることで求める。つまり、「包丁」と「まな板」が認識されたときの調理動作率は

(Add, Beat, Cut, Peel, Stir) = (0.69, 0.31, 0.00, 0.00, 0.00)となり、これを調理器具認識による調理動作推定の結果とした。なお、調理器具認識による調理動作推定に使用する調理器具認識の結果は、3.5で述べた初期統合法で使った調理器具情報の合計を利用する。つまり、調理動作間の時系列の調理器具情報を足し合わせて、認識した枚数の合計を割ることで算出する。ただし、調理器具認識結果が「そ

他(Other)」であった場合は、合計数には含まない。つまり、「包丁」「その他」「まな板」の場合でも調理動作率は、(0.69, 0.31, 0.00, 0.00, 0.00)となり、これを「包丁」「その他」「まな板」が認識されたときの調理器具認識による調理動作推定の結果とした。

3.5.2 腕の動きによる調理動作推定 –結果統合法–

結果統合法では、3.5.1で述べた調理器具認識による調理動作推定結果と腕の動きによる調理動作推定結果を結合して、調理動作の認識を行う。そのために腕の動きから調理動作を推定しなければならない。そこで、3.5でも述べたHMMを用いて、腕の動きから調理動作を推定する。なお、腕の動き情報としては3.4.1で述べた腕領域認識から得られた右腕・左腕それぞれの腕先座標のシーケンスをフレーム間の角度に変換後、角度方向によって分類した結果を使用する。つまり、本研究で使用する腕の動き特徴量は、1フレーム間の腕の動き特徴量を調理動作の間だけ取得した時系列データ集合となる。そして、3.4.2で述べたHMMを用いた調理動作認識手法を用いて腕の動きから調理動作を推定する。

しかし、調理器具認識による調理動作推定結果は、(Add, Beat, Cut, Peel, Stir) = (0.69, 0.31, 0.00, 0.00, 0.00)の様な調理動作率、つまり確率で表されている。そのため、結果統合するにあたり、腕の動きによる調理動作推定の結果に対しても調理動作率のように確率に変換しなければならない。そこで、算出したHMMに対しての尤度を使いソフトマックス関数により、確率に変換を行う。

ソフトマックス関数は、 i 番目の調理動作の尤度を a_i 、調理動作のカテゴリ数を D とすると、式(1)(2)で表せる。

$$y_i = \frac{\exp(a_i)}{\sum_j^D \exp(a_j)} \quad (1)$$

$$\sum_i^D y_i = \frac{\sum_i^D \exp(a_i)}{\sum_j^D \exp(a_j)} = 1 \quad (2)$$

例えば、それぞれの調理動作のHMMに対しての尤度が (Add, Beat, Cut, Peel, Stir) = (10, 15, 10, 8, 14) であった場合、式(1)より

$$(Add, Beat, Cut, Peel, Stir) = (0.0049, 0.7234, 0.0049, 0.0007, 0.2661)$$

となり、このように尤度を確率に変換したものを腕の動きによる調理動作推定の結果とした。

3.5.3 結果統合法による調理動作認識手法–結果統合法–

最後に、3.5.1で求めた調理器具認識による調理動作推定の結果と3.5.2で求めた腕の動きによる調理動作推定の結果を掛け合わせることで結果統合法による調理動作認識手法も結果とした。つまり、調理器具認識による調理動作推定結果P(H)、腕の動きによる調理動作推定結果P(U)、とすると、結果統合法による調理動作認識手法は式(3)と定義し

た。

$$P(H, U) = P(H)P(U) \quad (3)$$

例えば、調理器具認識による調理動作推定結果 $P(H)$ が、
 (Add, Beat, Cut, Peel, Stir) = (0.69, 0.31, 0.00, 0.00, 0.00)
 腕の動きによる調理動作推定結果 $P(U)$ が、
 (Add, Beat, Cut, Peel, Stir)
 = (0.0049, 0.7234, 0.0049, 0.0007, 0.2661)

の場合には、結果統合法による調理認識は

$$(Add, Beat, Cut, Peel, Stir) = (0.0034, 0.2242, 0, 0, 0)$$

より、認識結果は「Beat」となる。

4. 実験

実験として、ウェアラブルカメラより撮影した調理動作映像から手動で切り抜いた対象調理動作計 397 に対して、提案した初期統合法、結果統合法によって認識できているかを判定する。本研究では提案手法の有効性を示すために以下の認識率の比較を行った。

- (1) 腕の動作のみ (8 方向, 16 方向, 32 方向) (従来)
- (2) 腕の動作 (8 方向, 16 方向, 32 方向)
 + 調理器具認識 (初期統合法)
- (3) 腕の動作 (8 方向, 16 方向, 32 方向)
 + 調理器具認識 (結果統合法)

なお、調理における基本動作である「切る(Cut)」「剥く(Peel)」「かき混ぜ(Stir)」「炒める(Beat)」の 4 種類と、調味料等をそのまま「加える(Add)」動作と、調味料等を「(計量スプーンを用いて)加える(Add with Tool)」動作の 2 種類、計 6 種類の調理動作を本研究での調理動作認識における対象動作とした。HMM の状態数を 50, 100, 250 に設定し、Baum-Welch アルゴリズムのイテレーション回数を 1 回, 5 回, 10 回とし検証を行った。

5. 結果と考察

「腕の動きのみ」「初期統合法」「結果統合法」による認識結果の中で、それぞれの手法において最も平均認識率の高かった結果の比較を表 2 に示す。また、それぞれの手法の平均認識率を表 3 に示す。

表 2 認識手法の比較

Table 2 Comparison of Recognition Methods

		Add	Add with Tool	Beat	Cut	Peel	Stir	平均認識率
腕の動きのみ	8方向	58.33	45.45	69.23	66.67	58.14	89.47	64.55
	16方向	46.67	45.45	85.90	81.33	56.98	90.79	67.85
	32方向	43.33	54.55	85.90	76.00	56.98	90.79	67.93
初期統合法	8方向	95.00	90.91	100.00	100.00	100.00	97.37	97.21
	16方向	81.67	95.45	98.72	100.00	100.00	96.05	95.32
	32方向	55.00	95.45	100.00	100.00	100.00	98.68	91.52
結果統合法	8方向	50.00	63.64	87.18	81.33	58.14	90.79	71.85
	16方向	48.33	63.64	85.90	78.67	59.30	90.79	71.11
	32方向	63.33	50.00	67.95	76.00	59.30	92.11	68.12

表 3 平均認識率

Table 3 Average Recognition Rate

	Add	Add with Tool	Beat	Cut	Peel	Stir	平均認識率
腕動きのみ	49.44	48.48	80.34	74.67	57.37	90.35	66.78
初期統合法	77.22	93.94	99.57	100.00	100.00	97.37	94.68
結果統合法	53.89	59.09	80.34	78.67	58.91	91.23	70.36

腕の軌跡情報と調理器具情報を調理器具・腕の動き特徴量として、認識を行った初期統合法では、8 方向・16 方向・32 方向、そしてどの状態数を設定しても、その状態数における最も高い平均認識率は 90% 以上であった。特に認識率が高かったのは 8 方向・状態数 250 であり平均認識率 97.21% となった。また、調理動作ごとに認識率を確認してみても「炒める(Beat)」、「切る(Cut)」、「剥く(Peel)」、「かき混ぜ(Stir)」においては、かなり高い認識精度が確認できた。これは、「炒める(Beat)」、「切る(Cut)」、「剥く(Peel)」、「かき混ぜ(Stir)」といった調理動作において用いる調理器具を認識することができたからだと考えられる。一方で、「加える(Add)」、「(計量スプーンを用いて)加える(Add with Tool)」においては、他の動作に比べて認識率が下がっている、これは、加える動作は切るや炒めるといった動作と比べて、調味料をさっと加えるだけなど短い動作のため特徴が小さいといったことや、切る動作なら「包丁」、剥く動作なら「ピーラー」といった動作に使用する調理器具が決まっているが、加える動作には決まった調理器具が存在しないため、調理器具情報を加えても認識精度が向上しなかったと考えられる。

調理器具認識による調理動作推定の結果と腕の動きによる調理動作推定の結果を掛け合わせることによる、結果統合法においては、8 方向・16 方向・32 方向、そしてどの状態数を設定しても、その状態数における最も高い平均認識率は 70% 程度であった。結果統合法では、腕の動きのみと同様に Baum-Welch アルゴリズムの学習回数 (イテレーション回数) を増やすことで認識率の向上が見られた。ただ、結果統合法では「加える(Add)」、「(計量スプーンを用いて)加える(Add with Tool)」動作での認識率が大きく下がっている。これは、「加える(Add)」と「(計量スプーンを用いて)加える(Add with Tool)」において認識される調理器具がほぼ同じであるからだと考えられる。

本研究で提案した腕の動きのみ、初期統合法、結果統合法の認識率を比較してみると、初期統合法が最も精度の高い調理動作認識手法であることが分かった。初期統合法では、腕の動きのみよりも約 28% 以上、結果統合法よりも約 24% 以上の向上が確認された。また、結果統合法では腕の動きのみよりも約 4% 向上した。

このことから、行動理解において、使用する物体 (本研究では調理器具) が重要な情報であることが分かる。つま

り、使用する物体を正確に認識することにより行動認識が良い精度で行えると考えられる。

今回、結果統合法では約4%の向上に止まったが調理器具情報と腕の軌跡上を適切な重みづけにより結合することでさらなる認識率の向上が考えられる。

6. おわりに

今回、調理リハビリテーションにおける認知機能自動評価のための調理動作認識において重要な情報である使用している調理器具と腕の動きに着目し、取得した映像(一人称視点)を視覚情報として腕の動き/調理器具の認識から腕の軌跡情報と使用調理器具情報を取得し、それらの情報を初期統合法・結果統合法2つの情報統合手法による調理動作認識手法を提案し、検証を行った。

検証の結果、提案した動作認識手法において使用する物体情報を加えることで認識率の向上が可能であることが分かった。そのため、物体認識の認識精度を向上させることで行動認識の精度の向上が見込まれる。そして、認識物体を増やすことで日常生活における行動認識に活用することが期待できるため、今後は調理動作だけでなく、他の日常生活行動に対して活用することで、生活中における認知機能評価の自動化を目指していく予定である。

謝 辞

本研究の一部は文部科学省研究費補助金(基盤C15K00368)の支援を受けた。

文 献

- [1] 厚生労働省, 平成24年厚生労働省研究班調査報告, 東京, 2012
- [2] 中島八十一, 高次脳機能障害の現状と診断基準. 中島八十一, 寺島彰編: 高次脳機能障害ハンドブック—診断・評価から自立支援まで. 医学書院, 東京, pp.1-20, 2006.
- [3] 佐野睦夫, 宮脇健三郎, 米村俊一, 大出道子: 高次脳機能障害者の自立に向けた料理リハビリテーション支援, 信学技報, vol. 111, no. 424, WIT2011-54, pp. 19-24, 2012年1月.
- [4] 山肩洋子, 角所考, 美濃導彦, “食材および調理操作の流れからなる調理行動認識に基づく動的なレシピ認識手法の提案”, シンポジウム報告集 / 京都大学学術情報メディアセンター研究開発部 編, pp66-74, 2004.
- [5] Sanath Narayan., Mohan S. Kankanhalli., Kalpathi R. Ramakrishnan., “Action and Interaction Recognition in First-Person Videos”, CVPR2014 Workshop, 2014
- [6] Matsuo K., Yamada K., Ueno S., Naito S., “An Attention-based Activity Recognition for Egocentric Video”, CVPR2014 Workshop, 2014
- [7] 久賀 稜平, 前川 卓也, 松下 康之, “Web上の知識を用いた1人称視点映像におけるオブジェクトの利用を伴う行動の認識に関する検討”, 2016
- [8] 久賀 稜平, 前川 卓也, 松下 康之, “一人称視点映像を用いたWeb上の知識に基づく環境非依存な行動認識手法の提案”, 2016

- [9] Koen E. A. van de Sande, Jasoer R. R. Uijlings, Theo Gevers, Arnold W. M. Smeulders, “Segmentation as Selective Search for Object Recognition,” Proc. International Conference on Computer Vision, 2011.
- [10] 宮脇健三郎, 佐野睦夫, 米村俊一, 大出道子, 松岡美保子, “高次脳機能障がい者向け調理ナビゲーションのためのレシピおよび掲示メディアの構造化”, 映像メディア学会 Vol.64, No.12, pp.1863-1872(2010.10)