

認知リハビリテーションのための 調理ナビゲーションに基づく調理行動理解

大井 翔^{1,a)} 児島 宏樹¹ 佐野 睦夫^{2,b)}

概要：高次脳機能障害者の認知リハビリテーションとして、生活行動を活用したリハビリテーションは、日常生活に復帰するためにも重要であり、またリハビリテーションとしての効果も高い。認知リハビリテーションで重要なことは対象者自身が気づきを得ることである。われわれはこれまでに、自身の行動を映像とコメントで振り返る認知リハビリテーションの方式を提案し、調理を題材とした振り返り認知リハビリテーションにおいて、対象者のリハビリテーションへの意欲付けや気づきを誘発することを示唆した。既報において、リハビリテーションの自動化を目指し、動作認識を行った結果として単一の動作認識として平均 84%の精度であった。そこで本研究では、振り返り認知リハビリテーションにおける対象者の行動を自動でシステムが判断できるように調理行動理解に焦点をあてる。具体的には、従来の腕の使い方と調理器具の情報に加えて、調理ナビゲーションに記されたレシピをトップダウン情報として扱い、調理動作を一連の流れとして食材の情報も考慮した方式の検討を行った。

Cooking Behavior Understanding based on Cooking Navigation for Cognitive Rehabilitation

SHO OOI^{1,a)} HIROKI KOJIMA¹ MUTSUO SANO^{2,b)}

1. はじめに

近年、日本における高次脳機能障害者数は 50 万人以上いるといわれており [1]、今後も増加すると考えられている。高次脳機能障害は、注意機能障害や遂行機能障害などの高次の精神活動が障害された状態である。高次脳機能障害の原因としては、高次脳機能障害全国実態調査報告によると [2]、いずれの障害でも脳血管障害（脳梗塞、脳出血）が多く、障害ごとに脳外傷や変性疾患などがある。高次脳機能障害は認知症とは違い、認知リハビリテーション（高次脳機能障害に対するリハビリテーションを認知リハビリテーションとする）を行うことで、軽度～中度であれば

回復することが可能である。認知リハビリテーションにおいて、日常生活の能力改善を図り、社会参加の問題にアプローチすることは重要である。認知リハビリテーションを実施するにあたり、立神 [3] は、Rusk 通院プログラムにて神経心理ピラミッドと呼ばれる認知機能を中心とした心理学的機能を階層的に捉え、認知リハビリテーションにおいては、“気づき”は重要であることを示唆している。

当研究室において、これまで認知リハビリテーションとして、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、自分自身の“気づき”を与えることを目的とした「振り返り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み、調理を題材とした「振り返り認知リハビリテーション」が障害者自身の“気づき”を誘発し、リハビリテーションへの意欲向上することを確認した [4]。振り返り認知リハビリテーションプログラムの実施時は、一部手動でモニタリングしている部分があり、現在、自動化を目指している。自動化の一環として、筆者はこれまで調理行動の認識として手の動かし方

¹ 大阪工業大学大学院
Graduate School of Osaka Institute of Technology, Kitayama
1-79-1, Hirakata-shi, Osaka 573-0196, Japan

² 大阪工業大学
Osaka Institute of Technology, Kitayama 1-79-1, Hirakata-
shi, Osaka 573-0196, Japan

a) SHO.OOI@outlook.jp

b) mutsuo.sano@oit.ac.jp

に着目した方式 [5] や手の動かし方と調理器具を統合した方式 [6] にて、5種類の調理動作（切る、剥く、加える（入れる）、（かき）混ぜる、炒める）に対して、平均精度 84 % の精度で認識することが可能であることを示した。

しかし、これまでの研究では単一の動作の認識に着目した認識を行っていた。そこで、本研究では、これまで我々が提案した方式に加えて、トップダウン情報として、レシピデータを用いて、食材ごとに動作を管理することで一連の調理動作を認識する手法を提案する。

2. 従来研究

これまでに、調理行動に着目した動作認識として、ユビキタスセンシング系とウェアラブルセンシング系があり、ユビキタスセンシング系では、宮脇ら [7], [8] は、動作同期型調理ナビゲーションの実現のため、AR マーカや振動センサ、リードスイッチを用いて調理動作有無の判断し、山肩ら [9], [10] は、調理器具にセンサを添付、調理動作ごとに「シンク」「調理台」「コンロ」と場所を区切り光学カメラやサーマルカメラを用いて腕などを分離し、調理動作認識を試みている。ユビキタスセンシング系は精度が高い反面、コストなどに問題がある。

一方、ウェアラブルセンシング系では、久原ら [11] は CHLAC (Cubic Higher-order Local Auto-Correlation; 立体高次局所自己相関) 特徴量を用いて調理映像中の各フレームにおける調理者の動きを表現し、「切る」「混ぜる」「加える」といった調理操作の認識を試みており、井上ら [12] は視線情報を用いて調理動作の認識を試みている。我々も、ウェアラブルカメラを用いて、腕の動かし方に着目した方式 [5] や腕の動かし方と調理器具に着目した方式 [6] を提案してきた。

また、筆者らはトップダウン情報として、レシピ情報から調理を主として火を扱わない「下準備」、火を主として扱う「本調理」、調理の最後の動作である「盛り付け」と場面ごとに分割し、ステートチャートを用いて調理動作の認識を試みた [13]。しかし、実際の調理では下準備中であっても先に火を通しておくなどといった事も想定される。そこで、本研究では、これまで我々が提案した方式に加えて、トップダウン情報として、レシピデータを用いて、食材ごとに動作を管理することで一連の調理動作を認識する手法を提案する。

3. 調理動作認識手法

3.1 対象とする調理動作

調理動作は数百種類あり、分類の方法も複数存在する。そこで、我々は実際の調理レシピサイトである味の素のサイトからレシピデータ 9,515 件（2016 年 10 月時点）でのデータを対象として、MeCab を用いて形態素解析を行い、手動にて動作部分のみを抽出し、分類を試み、上位 75 %

を占め、かつ能動的に動作を行う動作である、「切る」、「剥く」、「加える（入れる）」、「（かき）混ぜる」、「炒める」を対象とする。

3.2 調理ナビゲーション

当研究室では、これまで高次脳機能障害に向けた調理ナビゲーションシステムを開発してきた [14]。このナビゲーションシステムは、図 1 に示すように、静的メディアと映像などの動的メディアを効果的に組み合わせ、レシピの内容も短文で暗黙知を排除する構成にすることで高次脳機能障害者の負担を軽減することが可能である。

しかし、このナビゲーションでは、決められた手順でしか作業をすることができず、融通が利かないシステムであった。つまり、作業を入れ替えたりすることができないというデメリットもあった。例えば、下ごしらえの場面にて、複数の食材があった場合、これまでのナビゲーションシステムでは、1つの食材ごとに作業を行っていた。具体的には、カレーを作るレシピにおいて、ニンジンとジャガイモがあった場合、調理ナビゲーションシステムでは、「ニンジン washing」→「ニンジンの皮を剥く」→「ニンジン cutting」→「ジャガイモ washing」→「ジャガイモの皮を剥く」→「ジャガイモ cutting」のように、食材ごとに作業を進めていた。しかし、調理作業は個人によって異なることがあり、動作ごとにまとめる作業を行う人もいる。

そこで、本研究では調理を「下準備」「本調理」「盛り付け」と分割し、調理フローを作成することとした。作成した調理フローの例を図 2 に示す。

3.3 作業管理テーブル

図 2 に示したように、食材ごとに着目すると動作の手順は決まっている。例えば、ジャガイモに着目した場合、「切る」動作は「剥く」動作が終わってからでないと作業できない。そこで、食材ごとに行う作業を管理し、これを作業管理テーブルと定義する。図 2 のレシピフローより作成し



図 1: 調理ナビゲーション
Fig. 1 Cooking navigation.

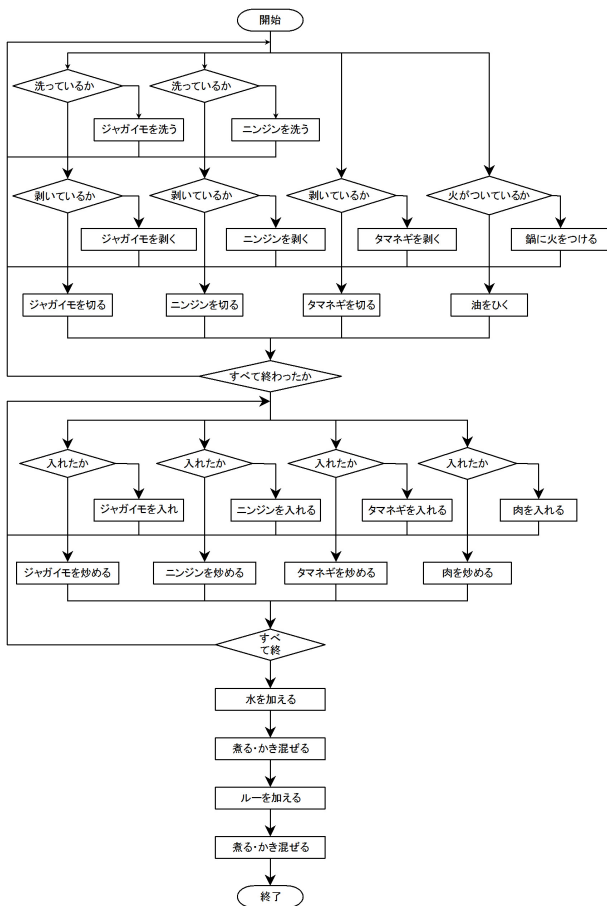


図 2: 調理レシピフローの例
Fig. 2 Example of recipe flow.

た作業管理テーブルを表 1 に示す。ただし、今回は対象とする調理動作は 5 種類のため、「洗う」動作は本研究では対象外とする。

作業管理テーブル、つまりレシピ情報を用いることで、動作の制約をつけることができる。例えば、表 1 に示した状態において動作に着目した場合、推定される動作は「作業可能な工程」の項目に記載されている動作の「洗う」「剥く」の 2 種類のみであり、「切る」「入れる（加える）」「（かき）混ぜる」「炒める」は推定されない。

例えば、次の手順のように作業を行った場合、作業管理テーブルは図 3 に示すような遷移となる。

●作業手順例

- (1) ジャガイモを洗う
- (2) ニンジン洗う
- (3) ニンジン剥く
- (4) ジャガイモを剥く
- ...
- (17) かき混ぜる

ここで、重要になるのが、何の食材を扱うのかによって推定される動作が変わるということである。つまり、使用する食材の認識が必要となる。

表 1: 作業管理テーブル
Table 1 The table of work management.

食 材	作業可能 な工程	次 の 工 程				混ぜる
		洗う	剥く	切る	入れる 炒める	
ジャガイモ	洗う	剥く	切る	入れる	炒める	混ぜる
タマネギ	剥く	切る	—	入れる	炒める	
ニンジン	洗う	剥く	切る	入れる	炒める	
お肉	—	—	—	入れる	炒める	

※網掛け部分がユーザが現在行うことのできる動作
※太字斜体の文字は前の動作がすべて終わった場合に行える動作

(1)ジャガイモを洗う

食 材	作業可能 な工程	次 の 工 程				混ぜる
		洗う	剥く	切る	入れる 炒める	
ジャガイモ	洗う	剥く	切る	入れる	炒める	混ぜる
タマネギ	剥く	切る	—	入れる	炒める	
ニンジン	洗う	剥く	切る	入れる	炒める	
お肉	—	—	—	入れる	炒める	

※網掛け部分がユーザが現在行うことのできる動作
※太字斜体の文字は前の動作がすべて終わった場合に行える動作

(2)ニンジン洗う

食 材	作業可能 な工程	次 の 工 程				混ぜる
		洗う	剥く	切る	入れる 炒める	
ジャガイモ	剥く	切る	—	入れる	炒める	混ぜる
タマネギ	剥く	切る	—	入れる	炒める	
ニンジン	洗う	剥く	切る	入れる	炒める	
お肉	—	—	—	入れる	炒める	

※網掛け部分がユーザが現在行うことのできる動作
※太字斜体の文字は前の動作がすべて終わった場合に行える動作

(3)ニンジン剥く

食 材	作業可能 な工程	次 の 工 程				混ぜる
		剥く	切る	入れる	炒める	
ジャガイモ	剥く	切る	入れる	炒める	混ぜる	
タマネギ	剥く	切る	入れる	炒める		
ニンジン	剥く	切る	入れる	炒める		
お肉	—	—	入れる	炒める		

※網掛け部分がユーザが現在行うことのできる動作
※太字斜体の文字は前の動作がすべて終わった場合に行える動作

(4)ジャガイモを剥く

食 材	作業可能 な工程	次 の 工 程				混ぜる
		剥く	切る	入れる	炒める	
ジャガイモ	剥く	切る	入れる	炒める	混ぜる	
タマネギ	剥く	切る	入れる	炒める		
ニンジン	切る	—	入れる	炒める		
お肉	—	—	入れる	炒める		

※網掛け部分がユーザが現在行うことのできる動作
※太字斜体の文字は前の動作がすべて終わった場合に行える動作

(17)かき混ぜる

食 材	作業可能 な工程	次 の 工 程				混ぜる
		洗う	剥く	切る	入れる 炒める	
ジャガイモ	混ぜる	—	—	—	—	—
タマネギ						
ニンジン						
お肉						

※網掛け部分がユーザが現在行うことのできる動作
※太字斜体の文字は前の動作がすべて終わった場合に行える動作

図 3: 作業管理テーブルの遷移例

Fig. 3 Example of transition of work management table.

3.4 食材認識

使用する食材を正しく識別できれば、トップダウン情報の制約条件の下でどのような調理動作を行うのか判断できる。しかし、調理における食材は時間とともに色・形状など変化していく。例えば、ジャガイモは皮をむくことで色変化が起き、切り方によっては形状が異なる。そこで、本研究では、事前に食材認識を行い、食材がどの程度識別できるのかを求め、求めた結果を用いることで食材から調理動作の推定する。対象とする食材の数 n 、調理数 m とし、ある食材 x の時に動作 y である確率 $p(y|x)$ とした場合、食材からの調理動作の推定確率 $p(F)$ は式 (1) で求める。

$$p_m(F) = \sum_n p(y_m|x_n) \quad (1)$$

事前の食材認識の結果例えば、ジャガイモの事前の食材認識の結果が(ジャガイモ:ニンジン:タマネギ) = (0.75:0.15:0.10) であり、作業手順が図 3(4) に示した状態の場合(ジャガイモ = peel, ニンジン = cut, タマネギ = peel) であれば、食材からの調理動作確率 $p(F)$ は、(add:stir:cut:peel:beat) = (0.00:0.00:0.15:0.85:0.00) となる。作業管理テーブルを使用しない場合は、食材ごとの動作は均等であるとする。

3.5 動作特徴量

動作特徴量として、移動方向ヒストグラムを用いる [5], [6]。移動方向ヒストグラムは、筆者らがこれまでに調理行動認識で扱った特徴量である。一人称視点映像から腕の領域を抽出し、腕の移動量をヒストグラムとした特徴量であり、以下のように定式化している。左右の手のオプティカルフローをそれぞれ求める。全時刻 T において、ある時刻 t における x 成分のオプティカルフローの画素値を $I_t(x)$ 、 y 成分のオプティカルフローの画素値を $I_t(y)$ 、強度 m 、角度 θ 、ヒストグラム h とすると以下の式 (2)~(4) ように定式化できる。

$$m_t(x, y) = \sqrt{I_t(x)^2 + I_t(y)^2} \quad (2)$$

$$\theta_t(x, y) = \arctan \frac{I_t(y)}{I_t(x)} \quad (3)$$

$$h(\theta') = \sum_t \sum_x \sum_y m_t(x, y) \cdot \delta[\theta', \theta_t(x, y)] \quad (4)$$

ここで δ は Kroencker のデルタ関数であり、勾配方向が量子化した勾配方向と同じヒストグラム要素である場合は 1 を返す。今回、量子化レベル θ' として 16 方向、32 方向とした。次に、すべてのデータ N 、方向ヒストグラムの分解能 c において移動方向ヒストグラムを算出後に正規化を行う。今回は、得られた全データの中で最も大きな値を 1 となるよう式 (5), (6) に示すように正規化を行った。

$$h_{norm}(\theta'_{i,n}) = \frac{h(\theta'_{i,n})}{h_{max}} \quad (5)$$

ただし、 $i = 1, \dots, c, n = 1, \dots, N$ である。

$$h_{max} = \underbrace{\text{maximize}}_{i=1, \dots, c, n=1, \dots, N} h(\theta'_{i,n}) \quad (6)$$

図 4 に「切る」動作の移動方向ヒストグラムを示す。切る動作では利き手である右手の動作が上下していることがよくわかる。食材を抑える手は上下左右に均等的に動いており、全体としてあまり特徴がない様子が分かる。

本研究では、移動方向ヒストグラムから推定される調理動作の確率を $p(H)$ とする。

3.6 調理器具による調理動作率

調理動作において、調理器具は重要な要素である。例えば、包丁を使っていれば「切る」や「剥く」などどのような調理動作を行うのか判断することができる。そこで、本研究では実際の調理動作から調理器具がどのような調理動作に使用されているのかを分布にした調理動作率を用いて判断を行う。本研究では調理器具としてまな板 (cutting board), ボウル (bowl), 包丁 (kitchen knife), お玉 (ladle), フライパン (frying pan), ピーラー (peeler), お鍋 (pot), 計量スプーン (measuring spoon), 菜箸 (chopsticks), フライ返し (food turner), 泡だて器 (egg whisk) の 11 種類とし、表 2 に調理ナビゲーションにある調理動画 40 種類から作成した調理動作率を示す。

ここで、調理器具数を n 、調理動作数 m 、調理器具の認識確率を $p_n(V)$ 、表 2 に示した調理動作率を $p_{n,i}(B)$ 、調理器具から推定される調理動作を $p(U)$ とすると式 (7) のように調理動作を求める。

$$p_i(U) = \sum_n p_n(V) p_{n,i}(B) \quad (7)$$

本研究では、調理器具数は表 2 に示すように 11 種類である。具体的には、事前に調理器具認識を行い、その結果として包丁の識別精度が(まな板:ボウル:包丁:お玉:フライパン:ピーラー:お鍋:計量スプーン:菜箸:フライ返し:泡だて器) = (0.00:0.00:0.75:0.05:0.00:0.05:0.00:0.05:0.00:0.00:0.10) とする。このとき、使用している調理器具が「包丁」であった場合に推定される調理動

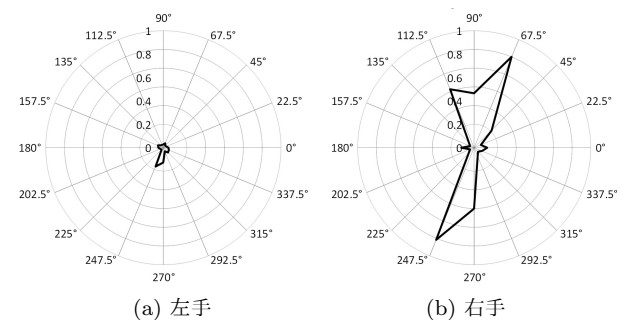


図 4: 「切る」動作の移動方向ヒストグラム
Fig. 4 Movement orientation histogram of cutting.

表 2: 調理動作率
Table 2 Cooking behavior distribution.

	cooking behavior distribution(%)				
	add	stir	cut	peel	beat
cutting board	0.00	0.00	0.75	0.25	0.00
bowl	0.57	0.40	0.03	0.00	0.00
kitchen knife	0.00	0.00	0.91	0.09	0.00
ladle	0.09	0.91	0.00	0.00	0.00
flying pan	0.25	0.00	0.00	0.00	0.75
peeler	0.00	0.00	0.00	1.00	0.00
pot	0.46	0.18	0.00	0.00	0.36
measuring spoon	0.93	0.07	0.00	0.00	0.00
chopsticks	0.36	0.21	0.00	0.00	0.43
food turner	0.00	0.00	0.00	0.00	1.00
edd whisk	0.00	1.00	0.00	0.00	0.00

作は (add : stir : cut : peel : beat) = (0.05 : 0.15 : 0.68 : 0.07 : 0.00) となる。

つまり、事前における調理器具認識精度が 100%であれば、表 2 の値である。しかし、実際には 100%の精度となることは難しい。また本研究では、1 枚の画像から 1 つの調理器具のみを結果として出力している。

3.7 調理動作推定

これまでに、調理動作として、1) 食材からの調理動作確率 $p(F)$, 2) 移動方向ヒストグラムによる調理動作確率 $p(H)$, 3) 調理器具による調理動作確率 $p(U)$ とし、推定する調理動作確率 $p(C)$ を式 (8) で求め、式 (9) で示すように、動作数 n の中で最も確率の高い動作を認識結果とする。

$$p(C) = p(F)p(H)p(U) \quad (8)$$

$$Cooking\ behavior = \arg \max_n p_n(C) \quad (9)$$

4. 実験

本研究では、研究室で行った調理 3 種類 (カレー 2 種類, 肉じゃが 1 種類) を対象とする。そこで、3.4~3.6 に示した認識率について実験を行い、それぞれの動作確率を求める。対象とした食材はカレーで使用する {ニンジン, ジャガイモ, タマネギ, ホウレンソウ, カボチャ} の 5 種類 (うち今回調理で使用したのは 3 種類), 調理器具は前述しているように 11 種類, 調理動作は 5 種類である。使用したデータ数を表 3, それぞれの画像を図 5,6 に示す。

本研究で扱った画像は、Web 上から画像取得を行い、イラストなどを手動にて排除した画像を利用した。

次に、実際の調理動画 2 種類に対してそれぞれの作業管理テーブルを図 7 に示す。かっこで示した数字が本研究の動作認識対象数である。

今回、食材・調理器具認識は Deep Learning のツールである Caffe を用いて認識を行い、モデルとして、認識率が

表 3: 調理動作率
Table 3 Cooking behavior distribution.

	学習数 (個々の数)	評価数 (個々の数)	画像・ 映像サイズ
食材	4,500(900)	1,800(360)	100px × 100px
調理器具	1,100(100)	550(50)	100px × 100px
調理動作数	50(10)	30(6)	480px × 270px
調理動画	—	2 (カレーを作成)	480px × 270px

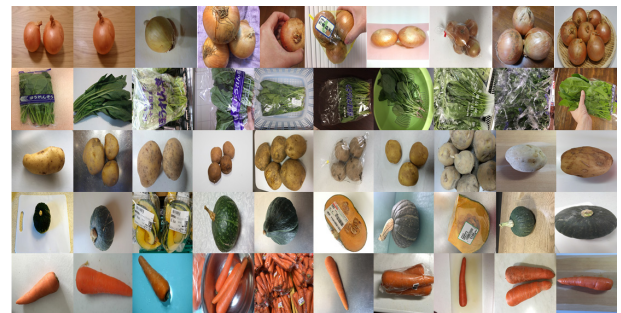


図 5: 食材の画像例

Fig. 5 Example of foodstuff.



図 6: 調理器具の画像例

Fig. 6 Example of cooking utensils.

食 材	作業可能 な工程	次 の 工 程				
		切る(2)	入れる(1)			
ジャガイモ	剥く(2)	切る(2)	入れる(1)		炒める(6)	混ぜる(7)
タマネギ	切る(5)	—	入れる(1)			
ニンジン	剥く(1)	切る(1)	入れる(1)			

(a) 実験参加者 A

食 材	作業可能 な工程	次 の 工 程				
		切る(1)	入れる(1)	炒める(1)		
ジャガイモ	剥く(2)	切る(1)	入れる(1)	炒める(1)	混ぜる(1)	
タマネギ	切る(1)	—				
ニンジン	剥く(1)	切る(1)				

(b) 実験参加者 B

図 7: 実際の調理における作業管理テーブル

Fig. 7 The table of work management by cooking.

安定して高くなる AlexNet を用いた。ただし、最終出力数は食材数 5 または調理器具数 11 としている。AlexNet のパラメータの詳細について、図 8 に示す。

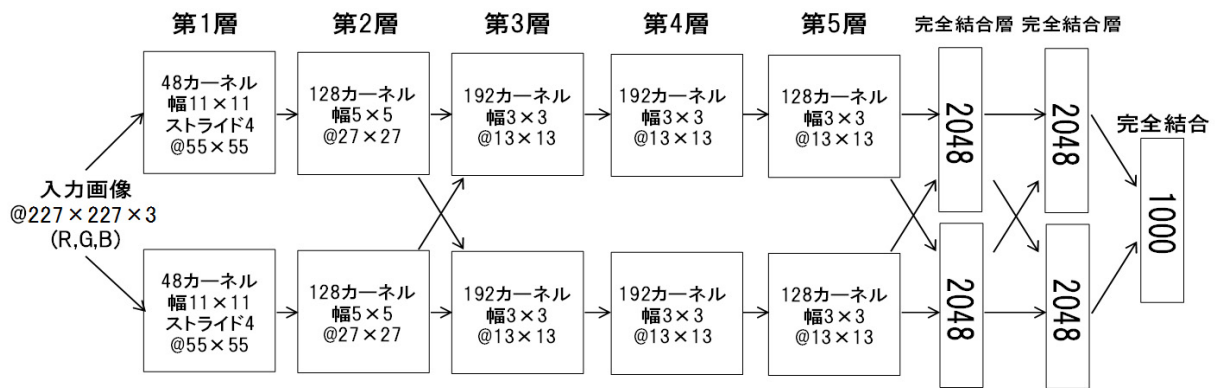


図 8: AlexNet パラメータ
Fig. 8 AlexNet parameter.

5. 結果

食材・調理器具の認識結果と移動方向ヒストグラムおよび3要素から推定した調理動作認識の結果を示す。

5.1 食材認識の結果

食材認識の結果を図9に示す。食材認識の平均精度は55.7%であった。各食材に着目するとホウレンソウ (spinach) が最も精度がよく85.0%であった。反対に、タマネギ (onion) の精度が最も悪く25.0%であった。

本研究では、この表を基に食材から調理動作を推定する。ただし、実際の動画では3種類の食材 (タマネギ、ニンジン、ジャガイモ) しか扱っていないが、今後他の食材も使用することを考慮し、この値で推定することとした。

5.2 移動方向ヒストグラムの結果

移動方向ヒストグラムの結果として、以下のパターンで検討を行った。

- (a) k-NN (k=1)
- (b) Support Vector Machine (RBF カーネル)
- (c) ニューラルネットワーク (隠れ層: 10)

図10に移動方向ヒストグラムの各パターンの認識結果を示す。今回の、移動方向ヒストグラムで平均精度が最も良いパターンであったSVMの結果を移動方向ヒストグラ

		Prediction Label				
		carrot	onion	potato	pumpkin	spinach
True Label	carrot	0.489	0.219	0.222	0.075	0.000
	onion	0.244	0.250	0.469	0.036	0.003
	potato	0.117	0.106	0.658	0.089	0.031
	pumpkin	0.103	0.069	0.056	0.542	0.231
	spinach	0.003	0.000	0.022	0.125	0.850
accuracy		0.557				

図 9: 食材認識の結果
Fig. 9 Recognition of foodstuff.

ムから推定される動作確率の結果として扱うこととした。

5.3 調理器具認識の結果

調理器具認識の結果を図11に示す。11種類の平均認識精度は66.3%であり、各調理器具に着目するとフライパン (flying pan) が最も良く92%であった。泡立て器 (edd whisk) を除き、全体的に50%以上の精度で識別できている。また、お玉や包丁は精度は50~64%であったが、他の

		Prediction Label				
		add	beat	cut	peel	stir
True Label	add	0.667	0.167	0.133	0.033	0.000
	beat	0.000	1.000	0.000	0.000	0.000
	cut	0.000	0.200	0.100	0.633	0.067
	peel	0.000	0.033	0.000	0.900	0.067
	stir	0.000	0.900	0.000	0.000	0.100
accuracy		0.553				

(a) k-NN

		Prediction Label				
		add	beat	cut	peel	stir
True Label	add	0.933	0.000	0.067	0.000	0.000
	beat	0.200	0.467	0.200	0.000	0.133
	cut	0.333	0.067	0.167	0.333	0.100
	peel	0.133	0.000	0.067	0.733	0.067
	stir	0.000	0.100	0.233	0.000	0.667
accuracy		0.593				

(b) SVM

		Prediction Label				
		add	beat	cut	peel	stir
True Label	add	0.367	0.333	0.167	0.033	0.100
	beat	0.000	0.867	0.000	0.000	0.133
	cut	0.000	0.133	0.367	0.200	0.300
	peel	0.000	0.033	0.000	0.933	0.033
	stir	0.067	0.700	0.033	0.000	0.200
accuracy		0.547				

(c) ニューラルネットワーク

図 10: 移動方向ヒストグラムの認識結果
Fig. 10 Recognition of movement orientation histogram.

		prediction labels										
		cutting board	bowl	kitchen	ladle	flying pan	peeler	pot	measuring spoon	chopsticks	food turner	egg whisk
true labels	cutting board	0.88	0.04	0.00	0.00	0.02	0.02	0.00	0.02	0.00	0.02	0.00
	bowl	0.04	0.80	0.00	0.00	0.04	0.02	0.04	0.00	0.04	0.02	0.00
	kitchen knife	0.08	0.02	0.64	0.00	0.04	0.04	0.00	0.06	0.06	0.06	0.00
	ladle	0.02	0.00	0.00	0.50	0.00	0.16	0.00	0.06	0.02	0.24	0.00
	flying pan	0.00	0.00	0.00	0.00	0.92	0.06	0.02	0.00	0.00	0.00	0.00
	peeler	0.06	0.00	0.00	0.02	0.02	0.74	0.02	0.08	0.00	0.06	0.00
	pot	0.00	0.02	0.00	0.00	0.16	0.00	0.78	0.04	0.00	0.00	0.00
	measuring spoon	0.06	0.06	0.00	0.00	0.00	0.08	0.06	0.54	0.04	0.12	0.02
	chopsticks	0.02	0.02	0.02	0.02	0.02	0.14	0.00	0.00	0.68	0.06	0.02
	food turner	0.00	0.00	0.02	0.00	0.02	0.14	0.00	0.06	0.06	0.70	0.00
	egg whisk	0.04	0.00	0.00	0.04	0.04	0.22	0.00	0.04	0.08	0.44	0.10
	accuracy	0.663										

図 11: 調理器具認識の結果

Fig. 11 Result of cooking utensils recognition.

器具からの誤認識は少ない結果であった。

5.4 調理動作認識の結果

今回、調理器具・食材は手の付近にあると仮定し、手の先端・付近で取得した画像に対して認識を行った。認識は動作中ランダムに選択し、対象となった画像から手の先端・付近を切り取った。実際には、YOLO などの検出アルゴリズムもあるが、今回は手動にて行った。図 12 に調理動作認識を行った結果を示す。

結果として、作業管理テーブルがない場合は平均精度が 62.9% であり、作業管理テーブルを用いることで約 23% 程度向上し、平均精度が 85.7% になった。また、動作ごとにみるとすべての動作のパターンで精度が向上している。

6. 考察

各実験の考察について述べる。

6.1 食材認識の考察

タマネギの精度が悪かった理由として、タマネギの状態が剥いている場合と剥いていない場合で異なり、剥いている場合がジャガイモの色や形状に類似していたため、誤認識してしまったのではないかと考える。また、ニンジンとの間にも誤認識があるが、色がそれぞれ干渉し合っていたため、精度が低下したと考えられる。

また、ハウレンソウの精度が高かったのは、他の食材と異なる色、形状であったため精度が最も高い結果となったと考えられる。

6.2 移動方向ヒストグラムの考察

移動方向ヒストグラムの結果として、3 種類の手法を行ったが、全体の平均精度は大きな差はなかった。動作ごとに確認してみると、どの方式も切る動作の識別精度が悪く、剥く動作へ誤認識しているケースが多い。これは、切る動作と剥く動作の移動方向ヒストグラムが縦方向で類似していたことが原因であると考えられる。次に炒める動作の精度について k-NN 法では精度が良いが SVM とニューラル

A				B		全体の精度	0.629
peel	0	add	1	peel	0	Aのみの精度	0.654
peel	0	beat	0	cut	1	Bのみの精度	0.556
peel	0	add	1	cut	0	add	1.000
cut	1	beat	1	peel	0	beat	0.667
cut	1	beat	1	peel	0	cut	0.636
cut	1	beat	1	cut	1	peel	0.000
cut	1	stir	1	beat	1	stir	0.875
cut	1	stir	1	add	1		
cut	0	stir	1	stir	1		
cut	0	stir	0				
cut	0	stir	1				
add	1	stir	1				
beat	0	stir	1				

(a) 作業管理テーブルなし

A				B		全体の精度	0.857
peel	1	add	1	peel	1	Aのみの精度	0.846
peel	1	beat	0	cut	1	Bのみの精度	0.889
peel	1	add	1	cut	0	add	1.000
cut	1	beat	1	peel	1	beat	0.667
cut	1	beat	1	peel	1	cut	0.727
cut	1	beat	1	cut	1	peel	1.000
cut	1	stir	1	beat	1	stir	1.000
cut	1	stir	1	add	1		
cut	0	stir	1	stir	1		
cut	1	stir	1				
cut	0	stir	1				
add	1	stir	1				
beat	0	stir	1				

(b) 作業管理テーブルあり

図 12: 調理動作認識の結果

Fig. 12 Recognition of cooking behavior.

ネットワークの精度が悪い。これは、SVM やニューラルネットワークの境界面が k-NN より複雑になっており、詳細な部分も考慮していたのが原因ではないかと考える。

6.3 調理器具認識の考察

泡だて器の精度が非常に悪かった原因として考えられるのが、泡だて器の Confusion Matrix を確認すると、フライ返しへ誤認識しているケースが多かった。これは、泡だて器とフライ返し、取っ手部分と先端部分が太くなっているという形状が類似していたため、誤認識したのではないかと考えられる。また、ほとんどのケースで泡だて器と判断しているケースが少なかった。これは、泡だて器のデータが他のデータと区別することがうまくできていないのではないかと考えられる。

6.4 調理動作認識の考察

作業管理テーブルを用いることで、初期段階の動作（炒めるや入れる動作は後から）など拘束することができたことにより誤認識を減らすことができ、精度が向上したと考えられる。しかし、われわれは 90% 以上の精度がでると考え

ていたが、実際は 85.7%という結果になった。これは、それぞれの特徴量から求められた動作の確率分布が影響していたことが分かった。例えば、移動方向ヒストグラムの結果は「切る」動作であるが、本研究においては、移動方向ヒストグラムの結果から推定される調理動作確率は「切る」動作よりも「剥く」動作の確率の方が高くなっており、誤認識の原因となっている。また、今回は検出する器具は 1 種類のみとしたが、実際には包丁とまな板など複数を用いることがあり、それぞれの共起性も考慮することで誤認識を防ぐことができるのではないかと考える。他の原因として、学習データは不要なオブジェクトなどない環境であったが、実験の動画では、実際の調理では食器をそのまま置いていたり、食材においては角度や調味料などにより色の変化があったりしたことが精度が悪くなった原因であると考えられる。

7. おわりに

本研究では、トップダウン情報としてレシピ情報を活用し、食材ごとに調理作業手順があるとした方式を用いて、調理動作の認識を検討した。

その結果、食材認識単体での精度は 55.7%、移動方向ヒストグラムでの動作推定精度は 59.3%、調理器具認識の精度は 66.3%であり、これらの推定確率を用いた調理動作の認識精度は 62.9%、さらにトップダウン情報としてレシピデータから作成した調理管理テーブルを使用した場合は 85.7%の結果であった。今後、複数の調理器具の認識と、推定するための動作確率の向上、複数のレシピに対しての実験を行っていく予定である。

謝辞 本研究の一部は、JSPS KAKENHI Grant Number JP 15K00368 の支援を受けた。

参考文献

- [1] 東京都高次脳機能障害者実態調査検討委員会：東京都高次脳機能障害者実態調査報告書，2008.
- [2] 日本高次脳機能障害学会高次脳機能障害全国実態調査委員会：“高次脳機能障害全国実態調査報告”，高次脳機能研究，Vol.31，pp.19-31，2011.
- [3] 立神粧子：前頭葉機能不全その先の戦略 Rusk 通院プログラムと神経心理ピラミッド（大橋正洋，Ben-Yishay，Y.，監修），医学書院，東京，2010.
- [4] 大井翔，佐野睦夫，渋谷咲月，水野翔太，大出道子，中山佳代：“高次脳機能障害者の自立に向けた調理行動振り返り支援システムに基づく認知リハビリテーション”，認知リハビリテーション，Vol. 20，No. 1，pp. 51-61，2015.
- [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] S. Ooi, T. Ikegaya, M. Sano：“Cooking Behavior Recognition using Egocentric Vision for Cooking Navigation”，Journal of Robotics and Mechatronics (JRM)，Vol.29，No.4，pp.728-736，2017.
- [7] 宮脇健三郎，佐野睦夫，西口敏司，池田克夫：“動作同期型調理ナビゲーションのためのユーザ適応型調理タスクモデル”，情報処理学会論文誌，Vol. 50，No. 4，pp.1299～1310，2009.
- [8] 宮脇健三郎，佐野睦夫，近間正樹，上田博唯：“動作同期型調理ナビゲーションのための実世界対応型調理タスクモデル”，画像電子学会誌，Vol. 36，No. 3，pp.252～263，2007.
- [9] 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”，京都大学学術情報メディアセンター開発部シンポジウム報告集，No. 2，pp.64-74，2004.
- [10] 山肩洋子，角所考，美濃導彦：“調理コンテンツの自動作成のためのレシピテキストと調理観測映像の対応付け”，電子情報通信学会論文誌，Vol. J90-D，No. 10，pp.2817-2829，2007.
- [11] 久原 卓，出口大輔，高橋友和，井手一郎，村瀬 洋：“CHLAC 特徴の周期性解析による料理映像中の繰り返し調理動作区間の抽出と識別”，電子情報通信学会技術研究報告，Vol. 110，No. 457，pp.61?66，2011.
- [12] 井上裕哉，平山高嗣，道満恵介，川西康友，井手一郎，出口大輔，村瀬洋：“調理者の視線運動パターンに基づく調理動作識別手法”，Proc. of Human Communication Group Symposium 2015(HCG2015)，HCG2015-D-9-3，pp.547-552，2015.
- [13] 大井翔，池ヶ谷剛，佐野睦夫：“一人称視点映像を用いたトップダウン情報に基づく調理動作認識の検討”，研究報告ユビキタスコンピューティングシステム（UBI），2017-UBI-53(11)，pp.1-6，2017.
- [14] 宮脇健三郎，佐野睦夫，米村俊一，大出道子，松岡美保子：“高次脳機能障害者向け調理ナビゲーションのためのレシピおよび提示メディアの構造化”，映像情報メディア学会誌，Vol. 64，No. 12，pp.1863-1872，2010.