

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

大井翔^{†1} 池ヶ谷剛^{†1} 佐野睦夫^{†1}

概要：近年、ウェアラブルカメラを用いた行動認識が盛んである。ウェアラブルカメラを用いることで簡単に人間の行動情報を取得することができ、ライフケアやライフログなどへの応用が期待されている。これまで我々は、生活に必要な調理に着目し、一人称視点から得られた動画から調理動作を単一的にとらえて識別を行ってきた。

本稿では、調理を一連の動作列として、トップダウン情報を用いた方法で調理動作の認識率向上を目指す。具体的には、トップダウンの情報としてレシピデータを学習することで、調理動作を「下準備」「本調理」「盛り付け」の3種類のステップに分ける。調理動作をステップごとに分けて認識することにより、調理動作の誤認識を防ぎ、認識を向上する方法を提案する。今回は対象として、「切る」「炒める」「加える」「かき混ぜる」「剥く」とした。「切る」動作は「下準備」のみのように、対象とした調理動作は、各ステップのみで使用する動作もあり、不要な動作を対象としないことで調理動作の識別の向上を目指す。

キーワード：ウェアラブルカメラ、調理動作認識、ライフケア、トップダウン情報

A Study of Cooking Behavior Recognition based on Top-Down using Egocentric Vision

SHO OI^{†1} TSUYOSHI IKEGAYA^{†1} MUTSUO SANO^{†1}

Abstract: The behavior recognition using a wearable camera is studied by several researchers. The wearable camera can get behavior data a person. So, the data can be used to life care and life log. We focused a cooking needed for daily living. And, we studied a single cooking behavior recognition based on egocentric vision. In this paper, we aim to improve recognition accuracy of cooking behavior using top-down data. Specifically, we recognize the cooking behavior using a method of top-down data. The top-down data is gotten from knowledge of recipe. The cooking behavior can separate to 3 steps. They are preparation behavior, main behavior and dishing behavior. We propose recognition method included 3 steps structure. For example, “cut” is used only for preparation behavior. So, we think to improve accuracy of cooking behavior. In this research, the cooking behavior is 4 behaviors. They are “cut”, “beat”, “stir” and “peel”.

Keywords: wearable camera, cooking behavior recognition

1. はじめに

近年、Google社のGoogle Glassをはじめ、頭部に装着することのできるウェアラブルカメラが普及しつつある。そのため、一人称カメラを用いた行動認識は盛んに行われている。一人称カメラを用いた行動認識は、個人の情報が簡単に取得でき、ライフケアやライフログの分野で注目されている。ウェアラブルカメラ以外にも、加速度センサなどを観測した身体の一部に装着することで、行動認識する方法が研究されている[1]。加速度センサは導入・管理のコストが小さくなるが、取得できるデータは装着した部位のみであり、複数の部位の動きを知りたい場合は、複数個所に装着する必要があること、オブジェクトの利用を伴う行動に関してはオブジェクトが何であるのかということが取得することが難しい。

また、ライフケアの一環として我々はこれまでに高次脳機能障害者に対して、調理を題材とした認知リハビリテー

ションを実施している。調理は、頭を使ったり身体を使う動作であり、日常において必要な行動の1つである。

これまでに、生活動作の中で調理に関する研究は多くされている。調理動作に関する研究として、宮脇らが、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をしている[2-3]。山肩らは、調理レシピを作成するために、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、調理動作認識をしている[4-5]。Murakamiらは、多数のセンサを埋め込んだキッチン空間を作成し、調理手順のサポートのためにマルコフ連鎖モデルを用いた方法を提案している[6]。久原らは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いる[7]。このように、複数のセンサを用いた調理動作の認識は盛んである。ただし、認識精度は高い結果であるが導入コストや設置の問題もある。

^{†1} 大阪工業大学大学院 情報科学研究科

Graduate of Information Science and Technology, Graduate School of Osaka Institute of Technology.

そこで本稿は、頭部 1 か所のみ装着するだけで使用できるウェアラブルカメラのみを用いて、トップダウン情報に基づいた調理動作の認識手法について検討する。具体的には、調理とは一般的にレシピに沿って行われることが多いということを前提として、トップダウンの情報としてレシピデータを用いることで状態数を把握し、また、調理煮には流れという概念があり、調理の流れの状態遷移を管理するためにステートチャート図を用いた方式を提案する。

2. 調理動作の認識手法

これまでに我々は、調理動作の認識において動作単体に対して調理動作の認識を行ってきた。しかし、調理動作は前後の流れを考慮して考えられる。そこで、本章では調理の流れを考慮した調理動作認識の手法について記述する。

2.1 調理動作の分類

調理動作は数百種類の動作があり、分類の仕方も複数存在する。本稿では、調理動作を実際のレシピサイトである味の素のサイトからレシピデータ 9515 件（2016 年 10 月時点）に対して、MeCab を用いて形態素解析を行った。形態素解析を行った結果に対して、人手にて動作部分のみを抽出し、調理に関係する動詞のみを仕分けした。仕分けした調理動作の総数は 117,295 件あり、表 1 に示すように調理操作として 10 種類に整理した。

表 1 調理行動の分類

Table 1. The classification of cooking behavior

調理操作	出現頻度	調理動作例		
追加	28.00	加える	詰める	-
加熱	24.07	焼く/炒める	煮る/炊く	-
切碎	14.69	切る	剥く	-
混合	9.23	混ぜる	合わせる	-
装飾	4.77	盛る	添える	-
浸漬	2.24	浸す	漬ける	-
分離	1.84	絞る	振う	-
冷却	1.29	冷やす	冷ます	-
洗浄	0.95	洗う	-	-
固定	0.41	抑える	固める	-
その他	12.5	溶かす	取る	抜く

本稿では、調理の大まかなカテゴリとして「調理操作」、詳細なカテゴリとして「調理動作」と定義した。また、表 1 においてボールド調で記した調理動作は動作中において人の手が必要となる動作であり、ボールド調でない動作は人の手が加えられた後、放置をすることが多い動作である。

ここで上位 4 カテゴリの調理操作「追加」「加熱」「切碎」「混合」を認識することで調理操作の 75%はカバーすることができる。そこで、本稿では上位 4 カテゴリを中心として、その中の「加える」「焼く/炒める」「切る」「剥く」「混

ぜる」の 5 種類の調理動作を認識対象とした。

2.2 トップダウンを用いた調理手順

調理の細かい流れは個人差があるが、大きな流れは大きな差が見られない。それは、調理はレシピというアウトラインがあるからである。レシピデータにより大まかな調理の流れを考えることができる。調理の大きな流れは「下ごしらえ」→「本調理」→「盛り付け」の順で行われる。これらの流れは一方的な流れであり、「本調理」の後に「下ごしらえ」がくることはない。ここで、表 2 に表 1 で示した調理操作を 3 種類の調理手順に整理した。

表 2 調理手順の分類

Table 2. The classification of cooking procedure

手順	調理操作								
下ごしらえ	追加	切碎	混合	浸漬	分離	冷却	洗浄	固定	その他
本調理	追加	加熱	混合	冷却	その他	-	-	-	-
盛り付け	装飾	-	-	-	-	-	-	-	-

調理操作ごとに分類すると、「下ごしらえ」と「本調理」に分類される調理操作があるが、調理動作単位でみると分離できる調理操作もある。たとえば、「追加」という調理操作はどちらの手順にも含まれるが、一般的には「詰める」という動作は「下ごしらえ」の手順に出現する。一方「加える」の動作は下準備の味付けや調理中の味付けで使用されるため、どちらにも含まれる動作となる。ここで、調理手順のステートチャートを図 1 に示す。



図 1 調理手順のステートチャート

Figure 1. The state chart of cooking procedure

また、各手順工程の中でも順序の流れはある。ここで、「下ごしらえ」の手順に着目する。例えば、ニンジンやジャガイモの下ごしらえであれば、通常の調理動作手順として、「洗う」→「(皮を) 剥く」→「切る」の流れであり、「剥く」と「切る」の調理動作が入れ替わることはない。このように、調理動作を動作間の関係により制約することで調理動作の認識を行う方式を提案する。

2.3 調理レシピモデル化

調理のレシピとして、これまでに調理ナビゲーションというインタフェースを作成していた[8]。調理ナビゲーションは通常のレシピとは違い、文字だけの表現ではなく、実際の動画を同時に提示することができるというメリットが

調理名： カレー

使用材料（具体的に） お肉、ジャガイモ、ニンジン、玉ねぎ、カレー	
下準備（番号を付けておく） 1. ジャガイモを洗う 2. ジャガイモの皮をむく 3. ニンジン洗う 4. ニンジンの皮をむく 5. 玉ねぎの皮をむく 6. ジャガイモを切る 7. ニンジン切る 8. タマネギを切る	
調理工程 下準備が終わって初めに行う工程	鍋に火をつけ、油をひく 1. お肉を入れる 2. お肉を炒める 3. タマネギを入れる 4. ニンジンを入れる 5. ジャガイモを入れる 6. 全体を炒める 7. 水も少し入れる 8. 全体を煮る 9. カレーを入れる 10. 全体を煮る 11. 混ぜる（煮込む）

図2 調理レシピの記入例

Figure 2. Example of recipe form

ある。また、通常のレシピでは複数の作業を1つの文章に書いているケースが多いが、調理ナビゲーションでは単文に構成になっているため、調理の初心者でも容易に調理をすることができる。しかし、調理ナビゲーションは決められた手順でしか作業をすることができない。つまり、作業を前後することができないというデメリットもあった。たとえば、下ごしらえにおいて、ニンジンとジャガイモを切る順番は決まっていたが、基本的には個人差があり、どちらから取り掛かってもよいはずである。また、食材も決められた食材でしか手順が書かれていないため、個人の調理として新たな食材を追加したりすることが困難であった。そこで、我々は柔軟に対応できるように、調理のベースとなるレシピは個人ごとに適用できるようにするために、あらかじめ調理者に必要な項目を記入してもらう方式をとる。実際に記入したレシピの例を図2に示す。

しかし、調理に慣れている人であれば簡単にレシピを作成することができるが、調理に不慣れな人はそもそも調理手順が分からない場合がある。そのような場合は他のレシピを参考にして作成してもらうこととする。図2の記入用紙における下準備が調理手順の下ごしらえ、調理工程が

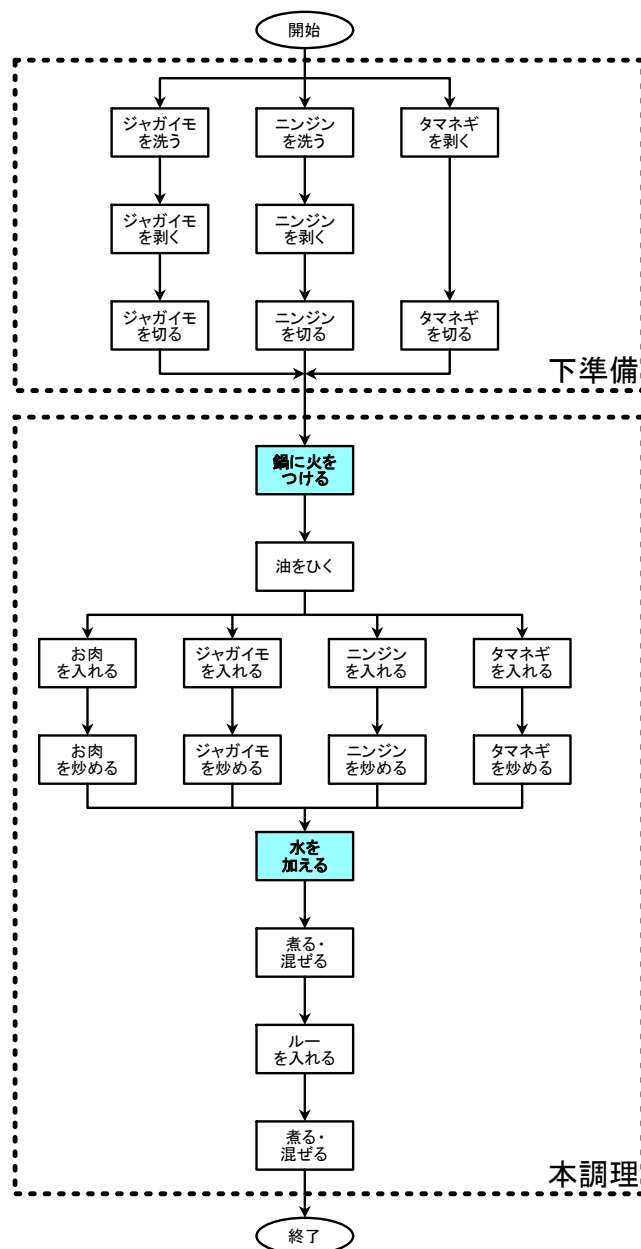


図3 調理フローの例

Figure 3. Example of recipe flow

本調理である。また、盛り付けは特にレシピ上に記入することはない。図2に示した調理レシピから生成したフローグラフを図3に示す。調理手順としては食材ごとに分割することが可能である。どの食材から作業を行ってもよいが、食材ごとの手順は一方方向のみとなる。つまり、図3に示した下準備工程において、ジャガイモ・ニンジン・タマネギのどの食材から手を付けても良いが、必ず上から順に行う必要がある。たとえば、「ジャガイモを洗う」→「ジャガイモを剥く」→「ニンジン洗う」→「タマネギを剥く」→・・・「タマネギを切る」という手順は受理できるが、「ジャガイモ洗う」→「ニンジン切る」→「タマネギ洗う」→・・・→「ジャガイモ切る」という工程は受理できない。これは、「ニンジン切る」という工程は「ニンジン洗う」と

表 5 作業管理テーブルの遷移

Table 5. The classification of cooking procedure

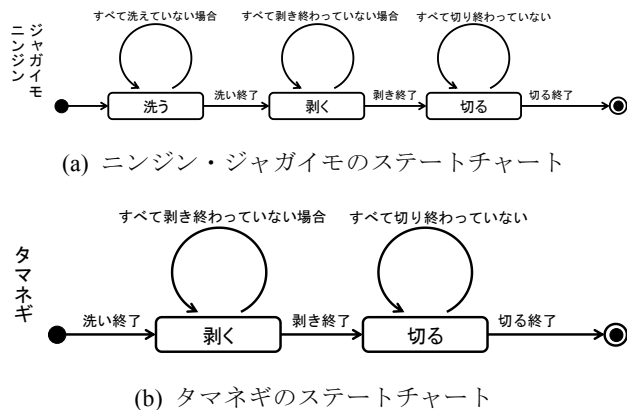


図 4 下ごしらえのステートチャート

Figure 4. The state chart of preparation behavior

表 3 調理手順の分類

Table 3. The table of work management

下 準 備			
食材	工程		
ジャガイモ	洗う	剥く	切る
タマネギ	剥く	切る	-
ニンジン	洗う	剥く	切る

※斜体文字が作業できる工程

いう工程がなければ作業できない工程であるからである。

図 3 における「下ごしらえ」における 3 種類の食材のステートチャートを図 4 に示す。

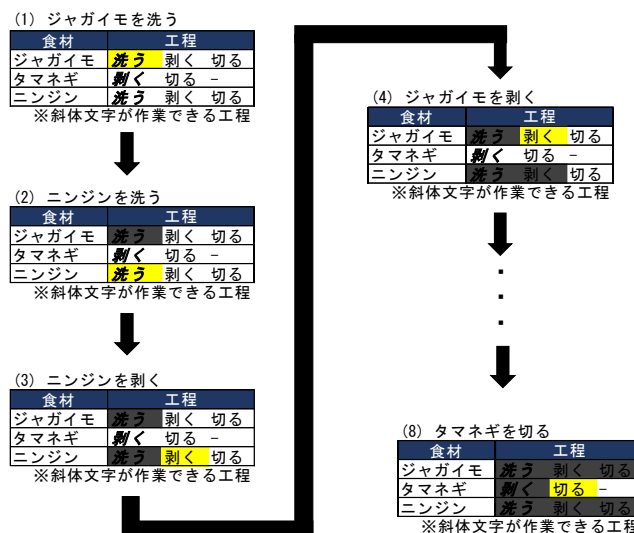
2.4 作業管理テーブル

図 3, 4 に示したように食材ごとに行う作業の手順は決まっている。そこで、どの食材のどの作業までを行ったのかを管理する必要がある。作業管理を行うために、表 3 に示すように作業管理テーブルを作成した。作業管理テーブルは食材ごとに必要な動作を列挙しており、必ず左から作業を行っていくこととなる。表 3 に示した作業管理テーブルは初期状態であり、行うことのできる作業は「ジャガイモを洗う」「ニンジンを洗う」「タマネギを剥く」という 3 種類の動作のみである。ここで、作業を以下に行った場合の作業管理テーブルを表 4 に示す。

●作業手順例

- (1) ジャガイモを洗う
- (2) ニンジンを洗う
- (3) ニンジンを剥く
- (4) ジャガイモを剥く
- ...
- (8) タマネギを切る

表 5 において、黄色で示した部分が現在の作業箇所、黒くなった部分が作業終了した部分を示す。



このように、調理動作において順序制約を行うことで調理動作の誤認識を防ぐことができる。つまり、表 2 で示したレシピにおいて下ごしらえに必要な動作は「洗う」「剥く」「切る」の 3 種類のみであり、「炒める」や「混ぜる」などの動作は認識対象外とする。下線部で示した動作は本稿で対象とした調理動作である。

2.5 動作特徴量

本稿において動作特徴量として、左右の手の移動方向ヒストグラムとした。移動方向ヒストグラムとは、左右の手の動きをオブティカルフローでとらえ、分解能 N の角度に対したヒストグラムのことである。

全時刻 T において、ある時刻 t における画像 $I(x, y)$ 、強度 m 、角度 θ 、ヒストグラム h とすると以下のように定式化できる。

$$m_t(x, y) = \sqrt{I_t^2} = \sqrt{I_t(x)^2 + I_t(y)^2} \quad (1)$$

$$\theta_t(x, y) = \arctan \frac{I_t(y)}{I_t(x)} \quad (2)$$

$$h(\theta') = \sum_t \sum_x \sum_y m_t(x, y) \cdot \delta[\theta', \theta_t(x, y)] \quad (3)$$

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。本稿では量子化レベル 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(4), (5)に示すように正規化を行った。

$$h_{norm}(\theta'_{i,n}) = \frac{h(\theta'_{i,n})}{h_{max}} \quad (i=1, \dots, c, n=1, \dots, N) \quad (4)$$

$$h_{\max} = \underset{i=1, \dots, C, n=1, \dots, N}{\operatorname{maximize}} h(\theta'_{i,n}) \quad (5)$$

図 5 に「切る」動作の移動方向ヒストグラムを示す。

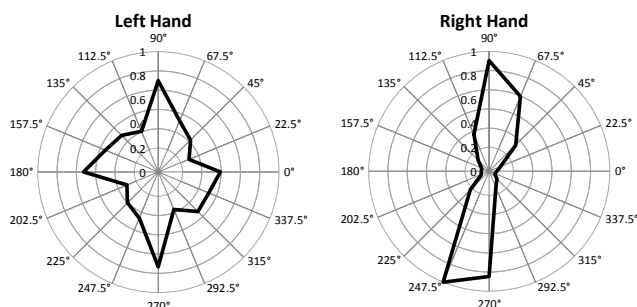


図 5 「切る」動作の移動方向ヒストグラム（16 方向）

左右の手領域の抽出方法は以下の処理で行う。

(1) 平滑化処理

画像に対してメディアンフィルタを用いて平滑化を行う。

(2) 色空間の変換と肌色領域の抽出

入力画像は RGB 色空間であるため、比較的照明環境に頑強な HSV 色空間へ変換し、肌色領域（H: S: V=0~15: 80~200: 60~200）の抽出を行う。事前実験の結果から最適なパラメータを設定した。

(3) ノイズ除去（膨張・収縮処理）

膨張・収縮処理 2 回を行い、ノイズ除去を行う。

(4) ラベリング処理

ノイズ除去を行った画像に対して 8 近傍のラベリングを行う。ここで、一人称視点映像において、作業中には手領域が大きく映ると仮定し、小さい領域は削除した。

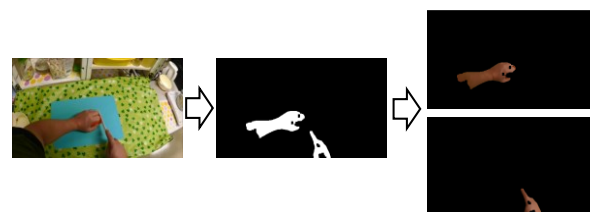
(5) 左右の手のセグメンテーション

一人称視点映像において、作業中であれば手が交差することがないという仮定から、画像において左側に見える部分を左手、右側に見える部分を右手とした。以上の処理を行い、手領域を抽出した結果を図 2(a)に 2 領域の場合、図 1(b)に 2 領域以外の場合を示す。

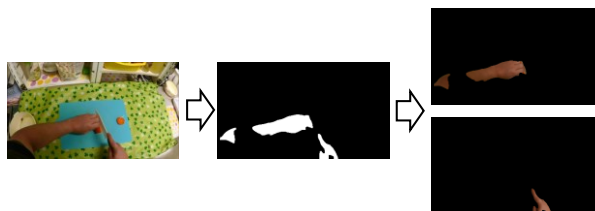
3. 実験

本稿では、ウェアラブルカメラとして Panasonic 製の HX-A500 を使用した。取得できる映像の大きさは 1920×1080 であるが、このままのデータを処理すると処理時間がかかってしまう。そこで、処理時間を考慮して 480×270 のスケールで処理を行った。今回、ダウンスケールとして 480×270 としたのは、比較的早く安定して処理ができたのでこのスケールを採用した。

調理動作のトレーニングデータとして、数名に対して 5 種類の動作映像データを 80 本の計 400 本用意した。実験としては、カレー、肉じゃが、焼きそばの調理に対して、調



(a)抽出した結果が 2 領域



(b)抽出した結果が 2 領域以外

図 2 手領域の抽出結果

理動作単体による認識と今回提案した方式による認識を示す予定である。

識別関数としては、(1)k-NN 法、(2)SVM(Support Vector Machine)、(3)3 層のニューラルネットワークで行う。

4. 結果・考察

現状、16 方向の移動方向ヒストグラム、32 方向の移動方向ヒストグラムにおける、調理工程を考慮しないで認識を行った認識率について表 6 に示す。

表 6 単一動作の認識率

Table 6. The accuracy of simple behavior

	16方向の移動方向 ヒストグラム	32方向の移動方向 ヒストグラム
k-NN	62.67	60.00
SVM	41.30	48.70
3層NN	68.67	70.67

表 6 の結果は、5 種類の動作全体における認識率である。k-NN 法においては 32 方向の移動方向ヒストグラムの精度が下がっているが、SVM と 3 層 NN においては精度が向上している。つまり、調理工程を考慮しない場合は 32 方向の移動方向ヒストグラムを採用するとよいということが分かった。

5. おわりに

本稿では、調理に着目したトップダウン情報に基づく調理動作の認識手法についての提案を行った。今後として、調理動作を認識するための特徴量として、画像データから何を使っているのかという調理器具を含めた調理動作の認

識を考えている。なぜなら、調理器具が分かればある程度調理動作が推定できるからである。たとえば、包丁とまな板が認識されればおそらく「切る」という情報が高くなるため。このように、調理器具の認識と手の動きの結果を組み合わせることで精度の高い結果が得られるだろうと考えられる。また、本稿で示した方式で高い精度を出すことができれば、他の生活行動に応用できると考えられる。今後として、調理以外に掃除や片づけといった行動を含めて、生活行動対象を広げていきたい。

謝辞 本研究の一部は、文部科学省研究費補助金（JSPS KAKENHI Grant Number JP 15K00368）の支援を受けた。

参考文献

- [1] 村尾和哉, 寺田 努: “加速度センサの定常性判定による動作認識手法”, 情報処理学会論文誌, Vol. 52, No. 6, pp.1968-1979 (2001).
- [2] 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).
- [3] 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).
- [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”, 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).
- [5] 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.
- [6] S. Murakami, T. Suzuki, A. Tokumasu, and Y. Nakauchi: "Cooking procedure recognition and support by ubiquitous sensors", Journal of Robotics and Mechatronics, Vol. 21, No. 4, pp.498-506(2009).
- [7] 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).
- [8] 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).