

修士論文

調理における下処理分類を利用した食材認識

指導教員 美濃 導彦 教授

京都大学大学院情報学研究科  
修士課程知能情報学専攻

橋本 敦史

平成20年2月8日

## 調理における下処理分類を利用した食材認識

橋本 敦史

### 内容梗概

調理は毎日行われる作業であり，しかも日によって違うレシピに挑戦するため，伝統的にレシピ本のように調理を教示するメディアに対する需要は高い．本研究では，計算機により，ユーザの調理進行を自動的に把握して，調理台上で映像や音声を用いてレシピを教示するシステムの実現を目指して，食材の色と下処理分類を利用する食材認識手法を提案する．下処理では「根がある」「皮がある」といった食材の種類に依存した特徴によって行われる調理操作系列が決まる．そのため，(食材認識を行うために必要十分な詳細さで調理操作を分類し，その分類に基づいて記述される)調理操作系列を識別すれば(上述の特徴に応じた)食材の分類が出来ると考えられる．そこで，まず，食材認識に必要な十分な詳細さを持つ調理操作分類を Web から収集した 100 個のレシピを解析することで定めた．これらのレシピの中で，識別が困難な食材と考えられる，同じレシピに登場し，かつ同じ色に分類される食材の組合せ 170 事例のうち，最も多い 57 事例を占める茶色の食材組合せを調べた．その結果，下処理で行われる調理操作を「洗う」「切る」「剥く」「茹でる」の 4 つのクラスに分類して識別すれば，47 事例が認識可能となることがわかった．このうち「切る」と「剥く」の調理操作クラスは短時間の観測では同じ動作に見えることもあり，短時間の観測結果から抽出された特徴量では区別が付かない場合がある．そこで操作中に短時間観測で得られた特徴量を繰り返し抽出し，その特徴量から各調理操作が行われたことを支持する確率を計算し，各時点での確率を元に下処理中に行われた調理操作の系列を求めることを考える．この考えに基づいて，実際に行われた下処理の観測データから，上述の調理操作分類で記述される操作系列を識別し，その結果から食材を認識する手法を提案し，実験により評価を行った．実際に茶色の食材に対する下処理を調理台において行った観測データへ提案手法を適用した結果，平均 75.0%の精度で食材が認識出来ることを確認した．

## Recognizing Food Material by Grouping Food Preparation on Cooking

Atsushi HASHIMOTO

### Abstract

Cookbooks have long been in used for getting recipes for our daily cooking activities. Instead of conventional cookbooks, “*Smart Kitchen*” gives cooking instructions automatically to user step-by-step in user’s home kitchen using computer. This paper focuses on recognizing food material assuming that the result of the recognition is used for giving the instructions to the user by Smart Kitchen. Cooking activity can divided into two procedures. One is main procedure which is given by a recipe. Another is a procedure for arranging food materials to be ready for beginning the main procedure (e.g. removing the root or the skin of the food). This paper calls the latter procedure “food preparing steps”. To recognize food material, this paper uses food colors and the food preparing steps. The way of preparing a food material depends on the features of the food material (e.g. “it has the root” and “it has the skin”), which characterize the food material. It means that food material can be classified by observing the process of preparing the food material. This study investigated 100 recipes collected from the WWW and found that 4 classes of cooking actions including “wash”, “cut”, “peel”, and “boil” can classify different sort of food materials with the same color Brawn. To recognize the 4 action classes, the features observed in short period are not always available because “cut” and “peel” classes are sometimes not distinguishable in a short-time observation. The proposed method recognizes food materials by calculating the probability for the occurrence of each action classes based on the extracted features at each frame, and recognizing the sequence of the performed actions through the whole preparing step from the calculated probabilities, instead of determining performed action for each frame. The proposed method was applied to 36 food materials and achieved a recognition rate of 75.0%.

# 調理における下処理分類を利用した食材認識

## 目次

第1章	緒論	1
第2章	関連研究	6
2.1	レシピガイドシステムに関する研究	6
2.2	調理台上の食材認識に関する研究	7
第3章	食材認識の為の下処理分類	10
3.1	下処理を利用した食材認識	10
3.2	食材の個体差と調理者の習慣による下処理への影響	10
3.3	食材認識に必要な十分な詳細さを持った下処理分類の調査	11
第4章	下処理分類を利用した食材認識器の実装	16
4.1	食材認識器の概要	16
4.2	特徴量からの各調理操作が行われた確率の計算	18
4.3	調理操作を表す特徴量の抽出と評価	21
4.4	各食材に対する下処理が行われた確率の計算	26
第5章	実験	30
第6章	結論と今後の課題	33
	謝辞	35
	参考文献	36

## 第1章 緒論

家庭で料理を作る場合には，加熱具合や調味に関係する情報など，料理を完成させる上で必要不可欠な情報を調理中に確認するために，紙媒体のレシピを参照しながら調理することが多い．しかし紙媒体のレシピで用いられる文章と写真による教示では，調理操作をどのように実行するべきかが曖昧な場合がある．また，調理中に度々調理から視線を外す必要があり煩わしい．

文章による教示の曖昧性や視線を外す煩わしさを解決する媒体として映像や音声それぞれ有用である．例えば「カスタードを作るために，ナイフで切れる硬さになるまで卵白を泡立ててください」という教示文が，そのような硬さになるまで泡立てられた卵白の写真と共に示された場合，それらの情報から実際にどのくらいの硬さにする必要があるのかを想像するのは容易ではない．映像による教示であれば，映像中で泡立てられる卵白の動き方を，実際に自分が泡立てたときの動きと比べることができ，より具体的に硬さを伝えることが出来る．映像や静止画を使わなくとも十分理解可能な教示については，文章の代わりに音声を用いることができる．音声による教示では，ユーザが調理を行う手を休めることなく，また調理から視線をはずすことなく教示を受けることが出来るという利点がある．

映像と音声により教示を行うものとしては，料理番組が広く普及している．しかし，料理番組では調理中の閲覧を想定しておらず，調理者が行っている調理の進行とは関係なく教示が進行してしまう．そのため，通常は調理中に料理番組が閲覧されることはない．

調理の進行はレシピ本などでは一つの教示文が指示する調理操作（あるいは調理操作集合）毎に区切られる．この調理操作（集合）を“調理ステップ”と呼ぶことにする．最近では，ユーザの行っている調理の進行を調理ステップ毎に把握し，映像や音声による教示コンテンツを調理の進行にあわせて提示するレシピガイドシステムが研究されている [1, 2, 3]．これらのレシピガイドシステムではユーザがどの調理ステップを行ったかをユーザからの報告により把握することで，調理の進行に合わせて教示コンテンツを調理中に閲覧することを可能にしている．ユーザは調理ステップを終える度にシステムに次の教示コンテンツを提示するように要請する．システムは教示コンテンツを全順序で保持し，要求があった回数をインデックスとして，次に行う教示コンテンツを提示する．

レシピガイドシステムの日常的な利用を考えた場合には上述のレシピガイドシステムには問題がある。日常的に調理を行っているユーザがレシピ本を参照しながら料理を作る場合、ユーザがレシピ本に求める情報は主に加熱具合や調味、あるいは上のカスタードの例での卵白の硬さのようなレシピ特有の注意点であり、調理の順序は重要ではない。実際にレシピ本を見ながら調理を行う場合には、調理ステップの順序はユーザが並行で行っている作業や、コンロの数などの調理環境により異なるため、レシピ本で指示されている調理ステップの実行順序をユーザが入れ替えることは日常的に行われている。それに対して既存のレシピガイドシステムでは、提示される教示コンテンツは予め全順序で並べられている。そのため、ユーザが調理ステップの順序の入れ替えを行うと、以降で既にユーザが行った教示内容が提示されたり、事前にやらなければならない調理ステップが完了していないうちに先の調理ステップが教示されるなど、ユーザの調理進行状況に合わない教示コンテンツが提示されることになってしまうため、調理ステップの順序の入れ替えを行うことができない。

本研究はユーザに調理ステップの順序を自由に入れ替えることを許してレシピガイドを行うことを目標としたスマートキッチンシステムの実現を目指す研究のひとつである。調理ステップの順序の入れ替えは、同じ食材に対する調理ステップ同士では不可能であり、異なる食材に対する調理ステップ同士が入れ替えられる。そこで、教示コンテンツの順序関係を全順序ではなく、同じ食材を対象とした調理ステップ間だけで順序関係が定義された半順序関係で記述する。このようにした場合には、各調理ステップで対象となっている食材がわかれば、食材毎には全順序関係が定義されているため、食材毎にどの調理ステップまでが行われたかがわかる。従って、各食材の次に行うべき調理ステップのうち、そのステップで扱われる各々の食材に対して、既に終了したステップの次のステップを提示することで調理進行状況に合った教示コンテンツを提示することが可能になる。

本研究では、上のような考えに基づいて各調理ステップで対象となる食材をシステムに知らせることを考え、レシピに登場する食材と調理台上の各物体とを対応付ける方法を検討する。これによって調理台上の物体がレシピに登場する食材と対応付けられれば、調理台上の物体を追跡する森ら [4] の手法と組み合わせることで、調理ステップがどの食材に対して行われたのかがわかる。

レシピに登場する食材と調理台上の個体とを対応付ける戦略としては、「認識

を行う」、「ユーザに報告させる」、「RFID タグに代表される電子的なラベリングを利用する」などの方法が考えられ、それぞれ長所と短所がある。

まず、電子的なラベリングを利用する方法は、RFID タグなどが予め食材のパッケージ等に取り付けられていることを仮定し、冷蔵庫や調理台にタグリーダーを設置することで食材をセンシングしようというものである。この方法は食材とタグが同時に移動する場合には有効であるが、例えばパッケージ内の一部の食材だけを取り出し、それを調理台に移動した場合のように、パッケージと食材が分離してしまうと、もはやセンシングが出来ず、システムに食材を把握させることが不可能であることが予想される。

一方、ユーザに報告させる方法は、常に誤りなく対応付けることが出来るが、調理台上の食材一つ一つに、対応する食材名を入力する作業を行う必要がある。調理は毎日行われる作業であり、登場する食材の個体数も少なくないため、日常的な利用を前提とした調理支援システムでは、調理の開始時に毎回、全ての食材名を入力するのは煩わしい作業となることが予想される。作業内容の簡略化の為には出来るだけ、このような入力作業を削減することが望ましい。

そこで本研究では、上のような方法ではなく、認識による方法を検討する。調理台上の食材認識が可能であれば、パッケージと食材が分類している場合にも対応することが出来る。また、認識を行うためにユーザからの報告は必要なく、煩わしい入力作業を省くことが出来る。ただし、認識による方法では、認識誤りが問題となる。これを完全に無くすることは出来ないが、ユーザに報告させる作業と組み合わせることで、入力作業を誤認識が起こった場合のみに限定することができる。ある程度頑健な認識が実現されれば、少数の誤認識に対してユーザに誤り訂正を求めても問題ないと考えている。

本来、食材は画像特徴のみから認識するのは難しい問題である。例えば、形状やテクスチャなどの画像特徴は個体差が大きかったり、特定の食材しか特徴付けられない。多くの食材を特徴付けられる画像特徴としては色がある。多くの食材が数種類の色素のいずれかを持つため、色による種類差が少ない。そのため、色特徴による同じ色素を持ついくつかの食材からなるグループへの分類は可能であるが、それだけで食材を認識することは難しい。この問題に対して、山肩らは食材を色を使って分類することに加え、食材に対して行われた操作の履歴とレシピに書かれた、各食材に対して行われる調理操作系列を照合することで、未編集の調理映像へのタグ付けを目的とした食材の認識を行った [5]。

本研究では、食材認識結果をレシピガイドシステムで手順の入れ替えを許すために利用しようとしている。レシピのガイドを行うシステムではユーザが自身の判断によって行うことが出来る行動については一々教示を行う必要はないが、レシピが決まって初めて決まるような調理ステップは教示をしなければユーザが調理を進めることが出来ない。例えば、ジャガイモの皮を剥くという作業は、ユーザが自身の判断によって行うため、一々教示をする必要はないが、皮を剥いたジャガイモを千切りにするのか、いちょう切りにするのか、あるいは柏木切りにするのかといったことは、レシピがわかって初めて決まる調理ステップであり、レシピのガイドを行うシステムが必ず教示を行わなければならないと考えられる。このレシピが決まって初めて決まるような調理ステップの集合をレシピ手順と呼ぶ。本研究では、レシピガイドシステムの為の食材認識を想定しているために、レシピ手順が開始されるよりも前に食材の認識が完了する必要がある。しかし、上述の従来手法で認識が完了するのはレシピ手順が観測された後であるため、この手法をそのまま本研究に適用することは出来ない。

レシピ手順が開始される前に食材を認識するために、本研究では食材への下処理で行われる調理操作系列から得られる情報を利用した食材認識手法を提案する。下処理は「洗う」や「皮を剥く」、あるいは「根や葉を切り落とす」といった調理操作からなる調理操作系列である。どのような調理操作を、どういう順番で行うかは、その個体に「皮がある」「根がある」といった特徴を認識したユーザ自身の判断により決定される。このような下処理はレシピには依存せず、個体のもつ上述の特徴をユーザが認識した結果に依存して決まるため、レシピ手順には含まれない。また、下処理は同じ食材クラスに属する個体をレシピ手順の中で一律に扱うために予め食材を整えることを目的とするため、必ずその食材に対するレシピ手順の開始前に観測される。そこで、その食材に対して施された下処理を観測し、その中で行われる調理操作系列によって食材を分類することによって、レシピ手順の開始前に上述の「皮がある」「根がある」といった特徴を反映した食材分類を行うことが出来ると考えられる。本研究では、従来手法でも用いられていた色による分類と、この下処理による分類とを組み合わせることで食材を認識する。

下処理によって食材をどの程度詳細に分類出来るかは、下処理で行われる調理操作をどの程度詳細に分類するかによって決まる。そこで、まず、実際のレシピに登場する同色の食材組合せが下処理により区別が出来るように、調理操



作の分類を定める．レシピ手順に比べて，下処理では行われる調理操作の種類が少ないために，この調理操作分類は従来手法で用いられていた調理操作クラスよりも詳細な分類となる．従来，調理操作分類を識別するためには，調理者の立ち位置が用いられていた．しかし，下処理で食材を認識するために必要な，従来より詳細に分類された調理操作クラスを識別するためには，立ち位置の特徴だけでは不十分であり，他に，何らかの特徴を用いる必要がある．本研究ではその特徴として調理者の手元の，動きがある領域の形状を用いる．

ただし，これらの特徴を用いる場合でも，上のような詳細な調理操作分類の中には，良く似た調理操作クラスが存在するため，短時間の観測結果では，観測の間に行われていた調理操作クラスを頑健に識別することが難しい場合もある．そこで，短時間の観測を繰り返すことで，その間に行われた調理操作の系列を正しく識別することを目指す．下処理から食材を分類するためには，下処理の中で行われた調理操作の系列を識別すればよく，必ずしも，各時点で行われていた調理操作を一意に決定する必要はない．そこで各時点での各調理操作クラスが行われた確率を基に，下処理の過程を通して，各食材に対応する下処理が行われた確率を計算することで，最も確率の高い食材を認識結果として出力する．

まず，2章では本研究の関連研究について述べる．3章では食材を認識するための下処理分類を定めるために実際のレシピを調査した結果について述べる．4章では，実際の映像から3章で定められた下処理分類を識別する手法と，それを利用した食材認識手法を述べる．5章では提案した食材認識手法を実際の下処理映像に適用し，その有効性を評価する．最後に6章で結論と今後の課題を述べる．

## 第2章 関連研究

### 2.1 レシピガイドシステムに関する研究

本研究で実現を目指すスマートキッチンと同様にレシピをガイドすることを目的としたコンピュータ強化キッチンには CounterActive[1] , eyeCOOK[2] や Kitchen of the Future[6] 上で動く Cooking navi[3] がある .

CounterActive[1] は , ビデオや音声を利用した調理ガイダンスを調理台で提示できるようにデザインされたシステムである . このシステムでは調理台上方に取り付けられたプロジェクタにより , 作業領域を想起させるスポットライトで照らされたような画像と教示用のビデオ , テキストを調理台上に投影し , 調理の初心者にも直感的にわかりやすい教示が行える調理台による調理支援を提案している .

このシステムのように , 調理台上で映像や音声により教示を行う場合には , 映像や音声をユーザの調理進行にあわせて提示する必要がある . 従って , システムは何らかの方法で調理の進行状況を把握しなければならない . CounterActive において , この調理進行状況の入力は , プロジェクタにより調理台にレシピや教示内容の映像を投影し , これに触れることで実現されている . このような調理者の手を用いた調理進行状況の入力による方法では , 調理中に調理用具を持ち替えたり手を洗ったりする場合に , 調理を妨げる可能性がある .

そこで eyeCOOK[2] では , 視線検出と音声認識を併用することで , 調理用具の持ち替えなどによるユーザの調理を妨げることなく , システムへの調理進行状況の入力を可能にすることを目指した . 調理の進行状況は調理者が , ディスプレイ上の該当する調理ステップを見ている状態で , その調理ステップを実行するための音声コマンドをユーザが発話することでシステムに与えられる .

しかし , キッチンでは調理音などのノイズ音が多く , この方式では , 調理者は音声コマンドの発話を要求されるにも関わらず , システムが発話内容を認識することに失敗してしまう可能性がある . Kitchen of the Future[6] では , 足によって操作するフットスイッチにより , ユーザの入力操作を確実に , 誤りなくシステムに伝えることを可能にしている . フットスイッチは調理台下部 , ちょうどユーザの足元あたりにある大きなスイッチで , 足のつま先で押し込むことで操作を行うことが出来る . フットスイッチを採用することで , eyeCOOK のように視線検出と音声認識を併用する方式とは異なり , ユーザによる調理進行

状況の入力がユーザの手を使わず、かつ誤りなく確実にシステムに伝えることが可能となっている。

CounterActive, eyeCOOK, Kitchen of the Future は全て、調理進行状況についての情報をユーザからの入力によって獲得している。これに対して、本研究が想定しているスマートキッチンでは、システムがユーザに入力操作を要求することなく調理進行状況についての情報を獲得するための認識技術の開発を目指している。本研究では、その中でも、調理者が行った調理ステップを特定し、調理の進行状況を把握するために重要な、調理台上の食材を認識することを目標としている。

また、前述のような調理進行状況の入力方法以外に従来システムとスマートキッチンにはいくつか相違がある。例えば、上述のシステムのうち、特に Kitchen of the Future にはいくつかの調理支援アプリケーションが実装されている。そのうち、本研究のようにレシピをガイドするシステムとして、Cooking navi[3] という調理支援アプリケーションがある。Cooking navi は調理経験の少ないユーザが特に難しいと感じる問題の一つである、調理ステップの順序の最適化をユーザの代わりに行うためのシステムである。しかしこの研究の中で行われたアンケートでは、調理に習熟した調理者にとっては提示された最適な手順に不満が残るという結果が出ている。システムが提示した調理ステップの順序に不満を感じたときに、ユーザの判断で調理ステップの順序を自由に入れ替えることが出来れば、調理に習熟した調理者も不満なくレシピガイドシステムを利用できる。そのために、スマートキッチンはユーザによる順序の入れ替えを許してレシピをガイドすることを目指している。

## 2.2 調理台上の食材認識に関する研究

調理台上の食材を認識する先行研究としては山肩らによるものがある。これは調理コンテンツの自動作成のためのレシピと調理映像との対応付けを目的としており、レシピ中に表れる食材や調理操作を未編集の調理映像から認識する手法 [5] を提案している。この手法は本研究と同様、食材の色と、食材に対して行われた調理操作から得られる情報を組み合わせることにより食材を認識している。

この手法では色による食材の大まかな分類と、その食材に施される調理操作の履歴や混ぜられる食材との関係を利用して食材の認識が行われる。食材の分

類には、ほぼ全ての食材に対して有効な分類を与えることが出来る色を利用した大まかな分類が用いられている。この色による大まかな分類は「赤・茶・黄・緑・黒・白」の6分類の色ラベルで与えられる。本研究でも、食材を分類するために、この色ラベルを用いる。

食材に施される調理操作の履歴や混ぜられる食材との関係は、レシピから得られる木構造を用いて表現される(図1)。これをレシピの木構造モデルと呼ぶ。この木構造モデルでは葉が食材、根が料理の完成品にそれぞれ対応し、それ以外の頂点は、下流にある葉の食材全てを対象として行われるなんらかの調理操作と対応する。根以外の各頂点は、対応する食材の色や調理操作クラスの情報をラベルとして保持する。このとき葉から一代ずつ親を辿り、辿った順にその頂点に対応する調理操作を書き下すことで、その葉に対応する食材に施される調理操作が、時系列順に与えられる。図1の例でいえば、食材  $X$  に対して行われる調理操作は  $A \rightarrow C \rightarrow F$  の順に行われる。

このレシピモデルを利用して食材は以下のように認識される。まず、映像から食材の色、調理操作クラスの識別結果と、その順序関係から、レシピモデルと同様の木構造を作成する。次に、その木構造をレシピモデルと一致するように編集することで映像から得られた識別結果とレシピモデルを対応付ける。最後に、レシピモデルの葉に一つ一つ対応する食材をレシピテキストに現れる食材から選ぶことにより食材が認識される。この対応付けは、レシピモデルに記述されている、各食材に施される調理操作の履歴や混ぜられる食材との関係を利用して行われる。葉と食材の一対一の対応付けが出来るためには、モデル上で区別がつかない二つの葉があってはならない。二つの葉が区別可能であるかどうかは、その二つの葉の共通の祖先の子を根とし、二つの葉がそれぞれ属する二つの部分木を比較することでわかる。例えば、図1のレシピモデル内の緑の二つの食材  $Y, Z$  の共通の祖先は  $F$  であり、それぞれ  $C, E$  を根とした部分木に属する。この二つの部分木が、その各頂点が保持するラベルも含めて、全く等しい場合には、それら二つの葉と、その二つの葉に対応すると思われるレシピテキストに現れる二つの食材との間の対応関係を特定することが出来ない。したがって同じ色ラベルを持つ二つの葉に対応する食材を認識するためには、少なくとも二つの葉の属する部分木で異なる頂点に対応する調理操作が観測されるまで待つ必要がある。そのため、この手法では多くの場合、レシピ手順がある程度進行するまで食材が認識できない。

上の研究では、食材の認識結果を、後で調理映像を調理の教材として利用する際にシーンを検索するためのタグとして利用することを前提としていることから、食材が認識できるのであれば、それがレシピ手順が全て完了した後であっても支障はない。これに対して本研究では、食材の認識結果をレシピ手順が行われている最中に提示する教示内容の決定に用いることを前提としていることから、レシピ手順の開始よりも前に食材が認識される必要があり、上の研究で用いられた手法をそのまま本研究に用いることが出来ない。

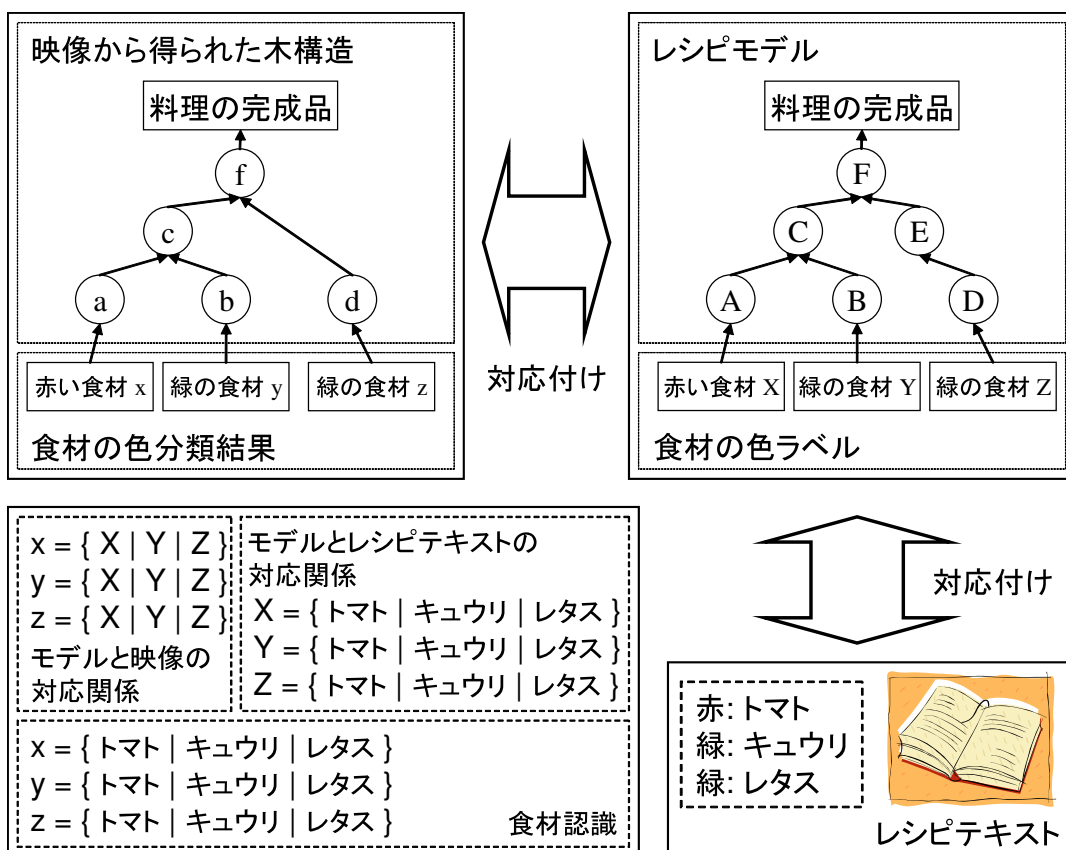


図 1: 映像から得られたモデルとレシピテキストから得られたモデルとのマッチングによる認識結果の修正と、認識結果を利用した食材の同定。

## 第3章 食材認識の為の下処理分類

### 3.1 下処理を利用した食材認識

前章で述べたように先行研究の手法では、レシピ手順の開始時点で食材を認識することが出来ない。この問題に対して、本研究ではレシピ手順の開始よりも前に行われる下処理を利用することで食材認識を行う。

下処理前の食材の特徴は「根がある」、「皮がある」、「種がある」、「土が付いている」、「あくが多い」などである。これらの特徴は、食材の種類によって決定されるため、食材を認識するのに有効な特徴量である。しかし、食材には様々な形状のものがあリ、また調理台上で食材の正確な形状や細かいテクスチャを観測することが難しいため「根がある」や「土が付いている」といった外見の特徴を画像処理で頑健に抽出することは難しい。また、「皮がある」、「種がある」、「あくが多い」といった特徴は食材の内部構造や成分に関わる特徴であり、非接触、非破壊型のセンサーでこれらを抽出することも難しい。

下処理では、上述の特徴をユーザが認識した結果に応じて、行われうる調理操作系列がある程度絞られる。例えば、根がある場合には、「根を切り落とす」という調理操作が行われる。また、種がある場合には、種を取り除ける状態にするために、まず「切る」操作が行われ、次に「種を取り除く」という操作が行われる。従って、下処理で行われた調理系列ごとに食材を分類することで、ユーザによる認識・行動結果を通して、間接的に上述の食材特徴を反映した食材分類を得ることが出来ると考えられる。

ただし「油抜き」など、一部の下処理は、複数の実行方法があるため、上述の特徴に対して、行われうる調理操作系列が一意には定まらない。次節では、この例のように複数の実行方法がある場合や食材の個体差により、同じ食材クラスに対して行われうる下処理が一对多の関係となるような場合に対して、どのように対処するかを述べる。

### 3.2 食材の個体差と調理者の習慣による下処理への影響

同じ種類の食材であっても、その個体に対して行われうる下処理にはいくつかの調理操作系列のパターンがある。この原因としては、そもそも下処理を決定付ける食材の特徴自体にも個体差があることや、その特徴に応じてユーザが下処理を実行する際に有効な実行方法が複数あることが挙げられる。

まず、下処理を決定付ける食材の特徴の個体差について議論する。どのような下処理を行うべきかを定める「根がある」、「皮がある」といった食材の特徴は、本来、同じ種類の食材であれば同じとなるはずである。しかし、実際には調理台上に現れるまでの流通過程で既に加工作られるなどして変化する。その結果、例えば予め端を切り落とす加工を施された状態で販売されていた個体に対する下処理では、未加工状態の場合に行われる調理操作系列から、「端を切り落とす」操作が省略された調理操作系列が行われることになる。あらゆる流通過程を想定して、どのような調理操作系列が行われるかを予めシステムに与えておくことは難しいため、本研究では、食料品店などで販売されている一般的な加工状態を想定する。

次に、ある食材の特徴に対する有効な下処理の方法が複数あるために起こる下処理の変化について原因を考える。有効な下処理の方法が複数ある例としては、例えば「油抜き」と呼ばれる調理操作がある。「油抜き」を行う場合にはシンクで食材に熱湯をかける方法とコンロで鍋に張った熱湯に湯通しする方法がある。そのため、「油抜き」に対応する食材特徴を持つ食材に対して行われる下処理方法には2つのパターンが存在する。このような場合に、どちらの方法を優先するかは調理者の習慣などにより決定される。

食材の個体差や調理者の習慣が異なることから、予め、ある食材クラスに対して下処理として行われる調理操作系列を頑健に予測すること難しい。本研究では、下処理のパターンが複数存在する問題に対して、どの調理操作系列が行われるかを予測することは避け、どのような調理操作系列が行われても、十分に食材を認識でき、かつ頑健な識別が可能な粒度で下処理を分類することで対処する。

### 3.3 食材認識に必要な十分な詳細さを持った下処理分類の調査

二つの食材クラスが同じ下処理のパターンを持たなければ、調理台上の物体に対して行われた下処理のパターンに基づいてその物体の食材クラスを一意に特定することが出来る。下処理のパターンは調理操作の系列であるから、ある二つのパターンが同一であると見なされるかどうかは、調理操作の分類粒度に依存する。調理操作の分類が詳細であるほど、下処理のパターンも詳細に分類することが出来る。しかし、行われた調理操作系列を識別する際には、調理操作の分類は粗い方が頑健な認識が可能になる。そこで、食材を認識するために

必要十分な粒度の調理操作分類を求めることを考える．

以降の議論を明確にするため，レシピ，食材クラス，調理操作，下処理，画像特徴の関係を以下のように定義する．まず，ユーザによって指定された，これからガイドを行おうとしているレシピを  $R$  とし， $R$  に登場する食材クラスの集合を  $F_R$  とする． $f \in F_R$  はレシピ  $R$  に登場するある食材クラスである．また，食材クラスの色ラベルを  $c$  で表す． $F_R^c$  は色ラベル  $c$  を持つものからなる  $F_R$  の部分集合である．一方，下処理に現れる全ての調理操作分類からなる集合を  $A$  で表す．このとき下処理は  $A$  の要素を記号とした記号列  $seq^A = a_0a_1\dots a_{m-1}$ ， $m > 0$ ， $a_i \in A$  ( $0 \leq i < m$ ) で表すことが出来る．3.2 で述べたように，同じ食材クラスであっても，個体差により，実際に施される下処理は多少の変更がなされる．そこで食材  $f$  に対して行われうる可能性のある下処理の集合を  $SEQ_f^A = \{seq_0^A, seq_1^A, \dots, seq_{n-1}^A\}$ ， $n > 0$  と表す．

同じ色ラベルを持つ2つの食材クラス  $f_0, f_1$  がいかなる場合でも下処理によって認識可能であるためには，調理操作分類  $A$  が以下の条件を満たす必要がある．

条件1  $SEQ_{f_0}^A \cap SEQ_{f_1}^A = \phi$ ， $\forall f_0, f_1 \in F^c$

これが成り立たない場合には， $seq \in SEQ_{f_0}^A \cap SEQ_{f_1}^A$  となるような下処理  $seq$  が行われたときには，例えば映像から正しい下処理分類である  $seq$  が抽出できたとしても，その下処理を施された食材を一意に特定できないため，食材を認識することが出来ない．

そこで実際のレシピを調査して，条件1を満たすような下処理分類を求めることを試みる．システムが複数のレシピを一度にガイドすることを想定した場合には，どのようなレシピの組合せが起こりうるかを予測することは難しいため，システムが一度にガイドするレシピは一つだけであると仮定し，同じレシピ内で共起し，同じ色ラベルを持つ食材組み合わせ  $\{f_0, f_1\} \in F_R^c$  に対して， $A$  が条件1を満たす食材組み合わせ事例を調べた．

調査対象として，2章で紹介した山肩らの関連研究で使用されたものと同じ，Web から収集した100種類のレシピを用いた．ここでは，食材の6種類の色による画像特徴分類が正確に行えるものと仮定し，各食材に対する色ラベルは予め人手で与えた．これらのレシピ集合には，同じレシピに登場する2種類の食材の組み合わせが960事例が含まれている．この960事例は，実際にその食材組み合わせが共起する頻度を反映するために，同じ食材組み合わせが何度も登場した場合にも，それらを重複して数え上げたものである．以降，事例という



単位が使われた場合には，全て，このように同じ食材組合せの重複を許して数え上げたものである．

これらの事例のうち，それぞれの食材が共に茶色の色ラベルを持つ組み合わせの事例について，実際の調理映像や料理の教則本 [7, 8, 9]，インターネット上の情報を参考に下処理手順の集合  $SEQ_{f_0}^A, SEQ_{f_1}^A$  を定め，条件 1 を満たす事例の割合を調べた．なお，組合せの両方の食材が同色 ( $f_0, f_1 \in F_R^c$ ) となる 170 事例のうち，調査の対象とした茶色の食材組合せは最も多い 57 事例が存在し，33.5% を占める．

これらの茶色の食材組合せについて，どの程度的事例が条件 1 を満たすかを識別可能性を考慮して定義した調理操作分類に対して調査する．定義する調理操作分類  $A$  を粗い分類から順に詳細にしていくことで，必要十分な詳細さを持った調理操作分類を求める．ただし，57 事例のうち，7 事例は調理台上での下処理が行われないハムなどの加工食品同士の組合せであった．これらは，調理操作分類の詳細さに関わらず，下処理では認識が不可能であるため，これら 7 事例を除外した茶色の食材組合せ 50 事例を用いて調理操作分類を評価した．

ユーザの作業位置による調理操作分類は頑健に識別可能であることが山肩らの研究 [5] で示されている．そこで，下処理に現われる調理操作を，その操作が行われる場所毎に以下のように定義し，これを頑健に識別可能な調理操作分類の初期値  $A_0$  とした．

- *wash*: シンクで水を使って行われる操作
- *cut/peel*: 天板上で食材の部位を分離する操作
- *boil*: コンロで水を使い，加熱する操作

表 1 は調査対象となった茶色の食材クラス  $f$  と， $A_0$  により調理操作を分類した場合のその食材クラスに対して行われうる下処理の集合  $SEQ_f^{A_0}$  をまとめた表である．この表では，同じ下処理集合  $SEQ_f^{A_0}$  を持つ食材は一つにまとめ，その食材集合を表すカテゴリ名を与えた．例えば  $SEQ_f^{A_0} = \{\{wash, cut\}, \{cut, wash\}, \{cut\}\}$  となる食材  $f$  は，一つにまとめられて「きのこ類」というカテゴリ名が与えられている．

$A_0$  による調理操作分類では条件 1 を満たす茶色の食材組合せ事例は 27 事例 (54.0%) となり，まだ 23 事例 (46.0%) が条件 1 を満たさなかった．これは食材を認識するためには不十分であると考えられる．そのため， $A_0$  の 3 つの調理操作分類を分割することでより細かい調理操作分類を作った．

表 1: 茶色の食材クラスと  $A_0$  で記述した下処理の対応関係

食材クラス ( $f$ )	下処理 ( $SEQ_f^{A_0}$ )
乾物	{wash}
きのこ類	{wash, cut}, {cut, wash}, {cut}
根菜類, 筍	{wash, cut}, {cut}
玉葱類	{cut}
油揚げ類	{boil}, {wash}
貝類	{boil} . {wash, boil}

この際、シンク、コンロで下処理の際に行われる調理操作のほとんどが、それぞれ「洗う」、「茹でる」であるため、シンクで行われる調理操作  $wash$  やコンロで行われる調理操作  $boil$  は、それ以上細かく分類しても、下処理による食材認識率への寄与が期待できないと考えられる。一方、天板上で行われる  $cut/peel$  は「一端を切り落とす」、「皮を剥く」、「種を取り除く」など、頻繁に行われる調理操作を複数含んだ調理操作クラスである。特に「切り落とす」作業と「皮を剥く」、「種を取る」といった、それ以外の作業を分けることで、分離されたものが「根」や「葉」などの包丁を使って分離される別のものであるか、あるいは「皮」や「種」であるかなど、食材の特徴をさらに区別することができると考えられる。実際には「皮を剥く」という調理操作はピーラーや包丁を使用する場合から、指で剥く場合も含み、「種を取る」や「魚のうろこをとる」など、他の多くの分離操作と類似している。一方、「切り落とす」場合には食材をまな板の上で固定し、包丁を利用する作業であり、他の分離操作との区別が比較的容易であると考えられる。

そこで、 $A_0$  中の  $cut/peel$  を「切り落とす」分離操作と「切り落とす以外」の分離操作に分割することで以下のような調理操作分類  $A_1$  を作成した。また、表 2 は、調査対象となった茶色の食材クラス  $f$  と  $A_1$  でそれらに対して行われうる下処理の集合  $SEQ_f$  を表したものである。

- $wash$ : シンクで水を使って行われる操作
- $cut$ : 天板上で行われ、まな板の上で、かつ包丁を用いて食材の部位を分離する操作
- $peel$ : 天板上で行われる、 $cut$  以外の分離操作
- $boil$ : コンロで水を使い、加熱する操作

表 2: 茶色の食材クラスと  $A_1$  で記述した下処理の対応関係

食材クラス ( $f$ )	下処理 ( $SEQ_f^{A_1}$ )
乾物	{wash}
きのこ類	{wash, cut}, {cut, wash}, {cut}
根菜類	{wash, peel}, {peel}
筍	{wash, peel, cut}, {peel, cut}
玉葱類	{cut, peel}
油揚げ類	{boil}, {wash}
貝類	{boil}, {wash, boil}

$A_1$  では 47 事例 (94.0%) が条件 1 を満たし, どの調理手順を実行されても認識可能と判断された.  $A_1$  で条件 1 を満たさないものは 3 事例 (6.0%) であった. これらの 3 事例は下処理の方法を決定する食材特徴がそもそも同じために, 同じ下処理が行われる. そのため, これらの 3 事例についても, 下処理が行われないハムや加工食品同士が組み合わせられた 7 事例と同様に, 下処理分類の詳細さに関わらず, 下処理を利用した区別が不可能である. 従って  $A_1$  よりも細かい調理操作分類を利用しても, 認識が可能な食材組合せ事例の増加は見込まれない. 本研究では, この程度の事例数ならば, ユーザに対応付けの操作を求めても, それほど負担にならないと考え, 食材を認識するための調理操作分類として  $A_1$  を用いた下処理分類を採用する.

## 第4章 下処理分類を利用した食材認識器の実装

### 4.1 食材認識器の概要

前章までで、食材認識を行うために必要な調理操作分類  $A_1$  が与えられたため、次にこの  $A_1$  を用いて、観測データから食材を認識する手法を述べる。本節では、 $A_1$  の系列として記述される下処理を利用して観測データから食材認識を行う手法の基本的なアイデアを述べ、アイデアを実現するための具体的な計算方法などは4.2以降で説明する。

ある食材に施される下処理は、第3章で定義した  $A_1$  の各調理操作クラスを、実行される時系列順に並べて得られる調理操作系列毎に分類される。第3章では、同じレシピに登場する同色の食材の組合せの多くに対して、 $A_1$  で記述された調理操作系列と食材が多対一の対応関係となることから、調理台上の個体に対して下処理で行われた調理操作系列がわかれば、その個体がどの食材であるかを特定することが出来ることを示した。従って、下処理の観測データから食材を認識する為の基本的な流れは以下ようになる。まず、調理台上の個体の色を抽出し、色ラベルを持った食材領域の追跡をする。追跡をしながら、各領域に施された調理操作クラスを各時点で識別する。それらの識別結果から下処理全体を通してどのような調理操作系列が行われたかを識別する。最後に、色ラベルと識別された調理操作系列から対応する食材を出力する。以上の手順で食材認識は完了する。

$A_1$  で定義された *cut* と *peel* の認識には調理者の手元の動きに関する特徴が有用であると考えられるが、良く似た操作であるため、短時間の観測から得られる情報だけでは区別がつかない場合がある。そこで、まず、二つの調理操作を明確に区別できる時点での観測結果を確実に含むように同じ操作が行われている間に何度も短時間の観測を繰り返す。次に各時点での観測結果から得られた特徴量を元に、その時点で各調理操作クラスが行われていた確率を求める。

この確率の系列から、調理操作を明確に識別できる場合だけを考慮して、下処理全体を通して行われた調理操作系列を求める。以下ではこのための手法と、その手法によって各時点で観測される特徴量だけでは区別がつかない場合にも調理操作系列を識別できる理由を述べる。

時刻  $t$  に調理台上から観測された調理操作の特徴量を  $x(t)$  とし、同時刻に調理台上において調理操作クラス  $a \in A_1$  が行われている確率を  $P_t(a)$  とする。以

降では，このように時刻  $t$  における確率を全て  $P_t$  と表す．このとき，ある時刻  $t$  での観測に基づいて2つの調理操作クラス  $\alpha, \beta (\alpha, \beta \in A_1)$  が存在する可能性がある場合には， $x(t)$  が観測されたときに  $\alpha, \beta$  が行われている条件付確率が近い値をとり，

$$P(\alpha|x(t)) \simeq P(\beta|x(t)) \quad (1)$$

となる．一方，区別がつく場合にはある調理操作が行われた確率だけが高くなる．このときにユーザによって行われた調理操作を  $\alpha$  とすると，

$$P(\alpha|x(t)) \gg P(\beta|x(t)) \quad (2)$$

が  $A_1$  に含まれる  $\alpha$  以外の任意の調理操作  $\beta$  に対して成り立つ．

このような性質を持った確率による出力を利用して，時刻  $t-1$  までにある調理操作系列が行われた確率から，時刻  $t$  までにある調理操作系列が行われた確率を伝播させる確率伝播計算を各時刻で特徴量が抽出されるたびに繰り返し，下処理の終了時点までに行われた調理操作系列の仮説を確率付きで求める．時刻  $t$  までに  $seq$  という調理操作系列が行われた確率を  $P_t(seq)$  とする．また， $seq$  の後に調理操作  $a$  が行われるときの調理操作系列を  $\{seq, a\}$  と表す．このとき，時刻  $t$  で  $\{seq, a\}$  が行われたという仮説に対する確率は次式で与えられる．

$$P_t(\{seq, a\}) = P_{t-1}(seq)P(a|x(t)) \quad (3)$$

観測された特徴量  $x$  だけでは区別できないような2つの調理操作クラス  $\alpha, \beta$  が存在する場合，仮説  $\{seq, \alpha\}$  と  $\{seq, \beta\}$  に対する確率は，式 (1) と式 (3) より，同等となり，どちらの仮説が行われた確率が高いかは保留される．

一方，観測された特徴量  $x(t)$  で調理操作クラス  $\alpha$  が行われた確率が他の調理操作クラスが行われた確率に対して十分に大きい場合には，式 (2) と式 (3) より仮説  $\{seq, \alpha\}$  に対する確率だけが高くなる．

従って，各時点における出力を  $P(a|x(t))$  とすることで， $x(t)$  で区別がつかない調理操作クラスが存在する場合には，調理操作系列の識別結果を一意に決定することを保留し，明確に区別がつく時点で得られた特徴量を反映した調理操作系列の識別結果を得ることができる．各時点で，各仮説に対して式 (3) を計算することで，下処理終了時点で，認識対象の食材  $f \in F_R^c$  に対応する下処理  $seq \in SEQ_f$  が行われた確率  $P(seq)$  を求めることができる．

最後に、各食材に対応する各下処理  $P(seq)$  から、下処理が施された個体が属すると考えられる食材クラス  $\hat{f}$  を求める。この際、区別のつきにくい3つ以上の調理操作系列が存在する場合には、最も確率の高い下処理に対応する食材だけを選ぶ方法が考えられるが、これには問題がある。例えば、ユーザによって行われた確率が高いと判断された上位3位の下処理  $seq_0, seq_1, seq_2$  ( $P(seq_0) > P(seq_1) > P(seq_2)$ ) があるとする。このうち最も確率が高い下処理  $seq_0$  に対応する食材を  $f_a \in F_R^c$  とおく。最も確率の高い下処理に対応する食材を選ぶ単純な方法では  $f_a$  が認識結果として選ばれる。しかし、 $seq_1$  と  $seq_2$  の両方に対応する食材  $f_b \in F_R^c$  が存在し、かつ、 $P(seq_1) + P(seq_2) > P(seq_0)$  であった場合には  $seq_0$  に対応する食材  $f_a$  を選択するこの単純な方法では問題がある。この場合、 $f_a$  と  $f_b$  を区別するだけならば  $seq_1$  と  $seq_2$  を同じ下処理分類  $seq_b$  と見なすことができる。このとき、下処理  $seq_b$  が行われた確率  $P(seq_b)$  は  $P(seq_1) + P(seq_2)$  と等しい。従って、本来は  $P(seq_b) > P(seq_0)$  であるにも関わらず、 $seq_0$  と  $seq_1$  を区別したために最も確率の高いものが入れ替わる現象が起きる。

そこで、本手法では、次式により、いずれかの下処理が行われた確率が最も高い食材  $\hat{f}$  を出力する。

$$\hat{f} = \operatorname{argmax}_{f \in F_R^c} \sum_{seq \in SEQ_f^{A_1}} P(seq) \quad (4)$$

このように計算することで、認識対象となる食材の組合せによっては区別する必要がない調理操作系列(上述の例では  $seq_0$  と  $seq_1$ ) を区別せずに、一つの調理操作系列として扱った場合と同じ認識結果を得ることが出来る。

以降の議論では、次の順番で上述のアイデアの実装についての説明を行う。まず、4.2 で特徴量  $x(t)$  から  $P(a|x(t))$  を計算する手法について述べ、次に4.3 で具体的にどのような特徴量を  $x(t)$  として用いるのかを説明する。これは、特徴量  $x(t)$  として有用である特徴量が不明であり、 $P(a|x(t))$  を用いて特徴量の評価を行うためである。次に、各時刻で各調理操作クラスについて出力される  $P(a|x(t))$  から式(4)を計算する手法について説明する。

## 4.2 特徴量からの各調理操作が行われた確率の計算

各時点で抽出される特徴量を用いて、 $A_1$  の調理操作分類に含まれる各調理操作クラスが行われた確率の計算を行う。そのために、教師つきデータを用いて、

ある時刻に観測された  $d$  次元特徴量  $x$  に対して，その特徴量が観測されたときに行われている調理操作が  $a$  である条件付確率  $P(a|x)$  を統計的に求める．限られたサンプル数で信頼できる統計を得るには，特徴量  $x$  を離散化し， $d$  次元の特徴量に対して各離散値ごとに十分なサンプル数を確保する必要がある．

離散化には，サンプル集合  $X$  の分布に応じた離散化が可能な方法として，クラスタリングによってサンプルをいくつかのグループに分ける方法をとった．クラスタリングの方法としては，画像処理ライブラリである OpenCV を用いて簡単に実装が出来るという理由から，サンプル集合  $X$  の分布が混合正規分布 (GMM: GaussianMixture Model) に従うと仮定し，GMM 内の各正規分布をクラスタとする方法を用いた．このとき， $x \in X$  は GMM を用いて以下のように表現される．

$$P(x; a_k, S_k, \pi_k) = \sum_{k=1}^m \pi_k P_k(x) \quad (5)$$

$$P_k(x) = \varphi(x; a_k, S_k) \quad (6)$$

$$= \frac{1}{(2\pi)^{d/2} |S_k|^{1/2}} \exp\left(-\frac{1}{2}(x - a_k)^T S_k^{-1} (x - a_k)\right) \quad (7)$$

ここで， $m$  は GMM の混合数， $P_k$  は平均が  $a_k$ ，共分散行列  $S_k$  で  $k$  番目の分布としての重み  $\pi_k$  を持つ正規分布密度である．また， $\pi_k \geq 0, \sum_{k=1}^m \pi_k = 1$  となる．

混合数  $m$  を指定すれば，EM アルゴリズムを用いてサンプル集合  $X$  の分布を最もよく近似する GMM のパラメータ  $a_k, S_k, \pi_k$  を推定することが出来る．この EM アルゴリズムの E-step では，その時点で得られているパラメータ推定値を用いて， $i$  番目のサンプル  $x_i$  が  $k$  番目の混合分布に属する確率  $P_{i,k}$  (以下の式では  $\kappa_{i,k}$  と記述) を以下のように求める．

$$\kappa_{i,k} = \frac{\pi_k \varphi(x_i; a_k, S_k)}{\sum_{j=1}^m \pi_j \varphi(x_i; a_j, S_j)} \quad (8)$$

M-step では，E-step で計算された  $\kappa_{i,k}$  を用いて以下のように推定値を高精度化する．

$$\pi_k = \frac{1}{N} \sum_{i=1}^N \kappa_{i,k} \quad (9)$$

$$a_k = \frac{\sum_{i=1}^N \kappa_{i,k} x_i}{\sum_{i=1}^N \kappa_{i,k}} \quad (10)$$

$$S_k = \frac{\sum_{i=1}^N \kappa_{i,k} (x_i - a_k)(x_i - a_k)^T}{\sum_{i=1}^N \kappa_{i,k}} \quad (11)$$

ここで  $N$  はサンプル集合  $X$  の要素数 (サンプル数) である。この際、GMM が  $X$  に過度にフィッティングすることを避けるため、 $S_k$  は対角成分以外が 0 であると仮定して計算を行った。

このクラスタリングにおいて、各クラスタは単に特徴量の類似性によりまとめられているだけである。このように特徴量の類似性のみでまとめられたクラスタを、調理の状況を信号的に類似したものにまとめたものという意味からシーンクラスタと名付け、シーンクラスタの集合を  $SC$  で表す。GMM のパラメータを推定した後は、サンプル  $x$  が属するシーンクラスタを  $sc(x) (\in SC)$  とすると、 $sc(x)$  は以下の式により求められる。

$$sc(x) = sc_{\hat{k}} \quad (12)$$

$$\hat{k} = \underset{k}{\operatorname{argmax}} \pi_k P_k(x) \quad (13)$$

サンプル  $x$  が属するシーンクラスタ  $sc(x)$  は、特徴量を離散化した際の離散値であると考えることができる。 $sc$  が観測された場合に各調理操作が行われている条件付確率  $P(a|sc)$  は、全ての調理操作  $a \in A_1$ 、全てのシーンクラスタ  $sc \in SC$  毎に数え上げることで統計的に得ることが出来る。この  $P(a|sc)$  を、調理操作とシーンクラスタをそれぞれ縦軸、横軸としてまとめた表は条件付確率表 (CPT: Conditional Probability Table) と呼ばれる。本研究では、調理を観測したデータから各サンプルがどのシーンクラスタに属するかを計算し、そのシーンクラスタから条件付確率表に従って各調理操作が行われた確率を求める。ただし、サンプル内での  $a$  の生起確率の偏りについては、特徴量  $x$  以外の要因による統計データの偏りを防ぐため、学習サンプル内で全ての調理操作の生起確率が等しくなるように均等化をしてから集計を行った。サンプル集合  $X$  のうち、教師ベクトルが調理操作クラス  $a$  を表すものの集合を  $X(a)$ 、クラスタ  $sc$  に属するサンプルの集合を  $X(sc)$  とすると、 $P(a|sc)$  は次式で求められる。

$$P(a|sc) = \frac{\rho(a|sc)}{\sum_{b \in A_1} \rho(b|sc)} \quad (14)$$

$$\rho(a|sc) = |X(sc) \cap X(a)| / |X(a)| \quad (15)$$

式 (12) と式 (14) から  $P(a|sc(x))$  を計算することができる。連続的な値である特徴量  $x$  から、統計的に  $P(a|x)$  を求めることは難しいが、近似値として  $P(a|sc(x))$  を用いることで、特徴量  $x$  から調理操作  $a$  が行われた確率を求めることが出来る。



📷 可視光カメラ: 平均7.5 frame/sec. 680\*480 pixels

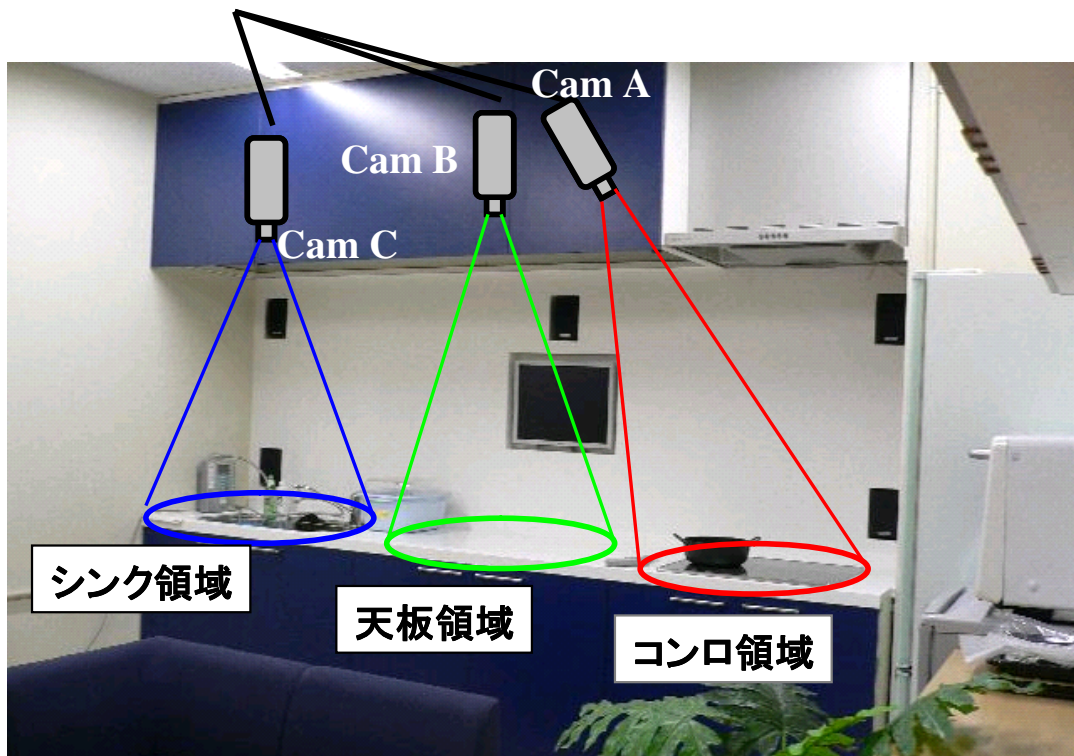


図 2: 可視光カメラによる調理観測環境を備えたキッチン

#### 4.3 調理操作を表す特徴量の抽出と評価

次に、 $A_1$  を区別するための、観測データからの特徴量抽出について述べる。調理者の行動を観測するために、スマートキッチンでは図2のような観測環境を用意した。まず、一般的な調理台としてシンク、調理作業を行うための天板、コンロを備えたものを用意した。この調理台上には作業を観測するためのセンサとして、調理台上方の天袋内にシンク領域、天板領域、コンロ領域を撮影する3台の可視光カメラが設置されている。それぞれのカメラにより撮影された映像は図3のようになる。

上方から撮影することでユーザの体以外によるオクルージョンが生じることなく、調理台上の食材を観測できる。また、背景が調理台で固定となるため、簡単な処理で調理台上で動きのある領域を区別することができる。

この観測環境のもとで、 $A_1$  を区別できるような特徴量を次のように選んだ。まず、「切る」「剥く」などの操作は単純な周期運動の繰り返しであるが、調理者の動作の周期や、より複雑な時系列変化に関する特徴は、食材の形状、大きさ、



Cam A: コンロ領域      Cam B: 天板領域      Cam C: シンク領域

図 3: 3 台の可視光カメラにより撮影したコンロ，調理台，シンクの画像

硬さなどの影響を受けて大きく変化するため，操作の特徴量としては期待できない．そのため，利用可能な特徴量は調理中に動きがある領域の形状や位置程度であることが予想される．そこで，データマイニングで用いられる特徴量評価手法により，後で特徴量の選抜を行うことを前提として，動きのある領域を表す候補特徴量を抽出した．

調理作業が行われている領域では，ユーザの動きにより，フレーム間差分が大きい画素が集中する．また，それ以外の領域では移動する物体はないため，調理作業領域外の画素のフレーム間差分は生じない．そこで， $t-1$  番目のフレーム画像  $I(t-1)$  と  $t$  番目のフレーム画像  $I(t)$  から得られるフレーム間の差分画像を  $I_{diff}(t)(=I(t)-I(t-1))$  として， $I_{diff}(t)$  から得られる 2 種類の二値画像  $I_1(t)$ ， $I_2(t)$  を用いて候補特徴量の抽出を行った．

画像  $Img$  の中で輝度が最大の画素値を  $\max(Img)$  と表すと  $I_1(t)$  は  $I_{diff}(t)$  を閾値  $Th_1/\max(I_{diff}(t))$  によって二値化したものである．この閾値は調理者の服の色が未知であるために，調理者の動きがあった領域でも，どの程度の差分が生じるかがわからないという問題に対応するために，各差分画像の最大の画素値に比例した値とした． $I_1(t)$  は時間応答性に優れるが，調理操作中にはユーザの動きは大きくないため， $S/N$  比が低いという欠点がある．

$I_2(t)$  は過去のフレーム間差分画像の加重平均画像  $\tilde{I}_2(t)$  を閾値  $Th_2/\max(\tilde{I}_2(t))$  によって二値化したものである (図 4 左)．観測開始フレームを  $t_0$  としたとき  $I_2$  は次式で示される  $\tilde{I}_2(t)$  を二値化して得られる．

$$\tilde{I}_2(t) = \sum_{s=t_0}^t \gamma^{t-s} I_{diff}(s) \quad (0 < \gamma < 1) \quad (16)$$

ここで  $\gamma$  は平均を取る際にかけられる重みであり， $0 < \gamma < 1$  とすることで， $t_1$  に近い  $I_{diff}(t)$  の重みがより大きくなるようにすることが出来る． $I_2(t)$  は過去

のフレームの影響を受けるため  $I_1(t)$  とは逆に時間応答性には劣るが、繰り返し動作が起こる領域の信号が強化されるため c 調理操作中の S/N 比が高いという特性を持つ。なお、閾値  $Th_1, Th_2$  は手動で設定を行った。

この  $I_1, I_2$  から計算される候補特徴量として、まずユーザの位置や動きの大きさを反映すると思われる、二値画像で値が 1 となる画素の重心 (2 次元) と画素数 (1 次元) を用いた。また、動きがある領域を、その領域に外接する矩形により表現することで、ユーザの手元の動きを反映した特徴量を得る (図 4 右)。

この矩形からは矩形の幅 (1 次元)、高さ (1 次元)、図 3 の縦方向を  $x$  軸、横方向を  $y$  軸とした場合の左右の辺の  $x$  座標 (2 次元)、上下の辺の  $y$  座標 (2 次元)、面積 (1 次元)、濃度 (=画素数/面積, 1 次元) を用いた。時間軸方向の細かい位置変化に影響を受けにくい値と、時間軸方向の位置変化のばらつきを示す値として、これら全 11 次元の特徴量に、さらにそれぞれの特徴の近傍 10 フレームにおける平均、分散を加えた 33 次元を  $I_1, I_2$  についてそれぞれ計算した。最終的に候補特徴量は 66 次元となった。

なお矩形はノイズの影響をさけるため、値が 1 となる画素のうち、画像の上下端から  $Th_3$  番目の画素を通る水平線と左右端から  $Th_4$  番目の画素を通る垂直線によって得ることが出来る。ここで  $Th_3, Th_4$  はそれぞれ定数の閾値である。作業領域では値が 1 となる画素がある程度の幅を持って密集しているため、閾値が多少増減しても、水平線や垂直線の位置は変わらず、これらの閾値の設定は難しくない。なお、調理台上のノイズは、一箇所に固まって発生する傾向があったため、領域を縮退させてノイズを除去する方法は避けた。



図 4: 作業領域特徴抽出に用いるフレーム間差分の加重平均画像 (左) と矩形領域を重畳した、左の画像の二値化画像 (右)

上述の計算によって得られる 66 次元の特徴量は，下処理が行われている調理台上を短時間の間に観測した結果から得られる．これは調理者の手元の動きを反映しているものと考えられるが，66 の特徴の全てが  $A_1$  を特徴付ける値であるかどうかは定かではない．通常，認識に用いる特徴が多くなると，学習にかかるコストが大きくなるだけでなく，サンプル集合  $|X|$  の数は固定であるために次元数  $d$  が大きい方が過学習が起りやすく，認識率が低下することがある．そこで，この 66 次元の特徴量から， $A_1$  の各調理操作クラスを区別するのに最も効果的と思われる特徴の組合せを選ぶことで，次元削減を行う．

特徴の選抜の際には，計算量の問題から，全ての組合せについての評価を行うことは難しいため，どのような組合せに対して評価を行うかなど，いくつかの戦略がある．それぞれの戦略の特徴は Jain ら [10] により評価がなされているが，ここでは，単純な方法として，最も認識に寄与しない特徴を順に取り除く方法を採用した．まず，上述の計算により得られる 66 次元の候補特徴を集合  $T$  により表し，これを初期値として  $\tilde{T}$  へ代入する． $\tilde{T}$  から特徴  $s \in \tilde{T}$  を抜いた部分集合  $\tilde{T} - \{s\}$  を  $\tilde{T}$  の全ての要素を  $s$  とした場合に対して作成し，評価関数  $H$  を用いて  $\arg \min_{s \in \tilde{T}} H(\tilde{T} - \{s\})$  となるような特徴  $s_{worst}$  を選び， $\tilde{T} - \{s_{worst}\}$  を新たな  $\tilde{T}$  として繰り返す．これを  $|\tilde{T}|$  が十分小さくなるまで繰り返し， $H(\tilde{T})$  が最大となる特徴の集合  $\tilde{T}$  を選抜された特徴量とする．

評価関数  $H$  としてどのようなものを用いるかについても様々な議論があるが，ここでは特徴量  $x$  から計算されるのは調理操作の認識結果ではなく各調理操作クラスが行われた確率  $P(a|sc(x))$  であるため，評価関数には特徴量が属するシーンクラスが判明した場合に得られる平均情報利得を用いた．この平均情報利得を利用した評価関数  $H$  は，混合数を  $m$  として得られた GMM の  $k$  番目のシーンクラスを  $sc_k^m$  とすると，以下の式で定義される．

$$H(\tilde{T}) = \min_m \left( - \sum_{k=0}^m P(sc_k^m) \sum_{a \in A_1} P(a|sc_k^m) \log P(a|sc_k^m) \right) \quad (17)$$

$m$  が大きすぎると各クラスを構成する学習サンプル数が足りなくなってサンプル集合  $X$  に対して GMM が過度にフィッティングし，また， $|A_1|$  未満になると，何れかの調理操作が，それが行われた確率が有意に高くなるような  $sc$  を持たなくなってしまう．GMM の過度のフィッティングや，クラス数不足が起きない範囲で探索するために， $m$  の探索範囲は 5 から 20 とした．

特徴量の評価に用いたデータは第5章で観測したサンプルから無作為に選ばれたものである。このデータにより選抜された特徴量は表3にまとめた8種類となった。また、その際、評価関数を最大とするシーンクラスタの数  $n$  は14であった。これらの8次元の特徴量で構成されたCPTを表4にまとめる。表4で  $P(sc)$  の項目は全サンプルの中でシーンクラスタ  $sc$  が占める割合を示しており、それ以外の値は横軸の  $sc$  と縦軸の  $a$  に対して  $P(a|sc)$  を表している。例えば表4から、 $sc_3$  が観測された場合に  $wash$  が行われている条件付確率  $P(wash|sc_3)$  は0.985であることがわかる。

表3: 選抜された特徴量

	raw data	average	variance
$I_1$	重心の $x$ 座標, 矩形の高さ	矩形の上下の辺の $y$ 座標	-
$I_2$	重心の $x$ 座標	重心の $x, y$ 座標, 面積	-

特徴量評価により選抜された8種類は全て、調理操作が行われている領域の位置に関するものであり、動き方の時間方向のばらつきを示す分散からは選出されなかった。また表4を見るとどのシーンクラスタが観測された場合でも、そのシーンクラスタが観測されたときに従来頑健な認識が可能であった  $A_0$  の調理操作分類に相当する  $wash, cut/peel, boil$  の3つのうち、いずれが起きたか

表4: 評価関数が最大となった条件付確率表

$SC$	$sc_0$	$sc_1$	$sc_2$	$sc_3$	$sc_4$	$sc_5$	$sc_6$
$P(sc)$	0.102	0.079	0.055	0.112	0.047	0.058	0.018
$wash$	0.123	0.000	0.000	0.985	0.619	0.113	0.000
$cut$	0.754	0.298	0.428	0.000	0.288	0.000	0.000
$peel$	0.123	0.702	0.572	0.015	0.093	0.133	1.000
$boil$	0.000	0.000	0.000	0.000	0.000	0.754	0.000
$SC$	$sc_7$	$sc_8$	$sc_9$	$sc_{10}$	$sc_{11}$	$sc_{12}$	$sc_{13}$
$P(sc)$	0.032	0.070	0.107	0.068	0.009	0.035	0.206
$wash$	0.000	0.898	0.000	0.000	0.000	0.806	0.000
$cut$	0.392	0.056	0.454	0.661	0.104	0.028	0.000
$peel$	0.608	0.046	0.546	0.339	0.896	0.166	0.000
$boil$	0.000	0.000	0.000	0.000	0.000	0.000	1.000

は明確に優位な調理操作を決めることが出来る．例えば， $sc_2$  では  $cut$  と  $peel$  は区別が難しいが，これらを同じ調理操作とみなせば， $P(cut/peel|sc_2) = 1.00$  となっている．他のシーンクラスタについても， $P(cut/peel|sc)$  を同様に計算すれば， $wash$ ， $cut/peel$ ， $boil$  の3つについてはどのシーンクラスタでも明確に区別可能であることが確認できる．一方で  $cut$  と  $peel$  の区別は難しく， $P(sc)$  が比較的大きい値をとるシーンクラスタの中でこれらが良く区別が来ている  $sc_1$  では  $peel$  の確率が 70.2% であるものの， $sc_9$  ではほとんど区別がついておらず，やはり  $cut$  と  $peel$  の認識は難しいことがわかる．

#### 4.4 各食材に対する下処理が行われた確率の計算

本節では，時刻  $t$  における特徴量  $x(t)$  により計算された調理操作クラス  $a(\in A_1)$  が行われた確率  $P(a|sc(x(t)))$  を利用して各食材に対する下処理が行われた確率の計算を行う方法を述べる．レシピ  $R$  のガイドをする場合には，同じ色ラベル  $c$  をもつ食材集合  $F_R^c$  に含まれる各食材に対して行われうる下処理についてのみ，それらの下処理がユーザによって行われた確率を求めれば，式 (4) の右辺に含まれる，各食材に対する下処理が行われた確率  $\sum_{seq \in SEQ_f^{A_1}} P(seq)$  を計算することが出来る．

この計算を行うために  $f$  に対して行われうる全ての下処理分類の集合  $SEQ_f$  を変換して得られる有向グラフ  $G^f(V, E)$  上での確率伝播計算を用いる．変換されたグラフ  $G^f$  はソース  $u_{src} \in V$  とシンク  $u_{sink} \in V$  を一つずつ持つ連結グラフである． $V$  に含まれる  $u_{src}, u_{sink}$  以外の頂点は  $A_1$  の調理操作クラスの何れかに対応している．頂点  $u$  に対応する調理操作クラスを  $a(u) (\in A_1)$  で表す． $u_{src}$  から  $u_{sink}$  までのパス  $P$  上で  $u_{src}$  に近い頂点に対応する調理操作から順に並べることで得られる系列は調理操作分類を記号とした記号列となる．ここで， $u_{src}$  から  $u_{sink}$  までの全てのパスからなる集合を  $\mathcal{P}^{G^f}$  とするとき， $\mathcal{P}^{G^f}$  から  $SEQ_f$  への写像は全単射となる．

図 5 は表 2 のうち，きのこ類，根菜類，玉葱類，油揚げ類に対する下処理集合  $SEQ_f^{A_1}$  のグラフによる表現である．このグラフは，確率伝播計算時に無駄な計算を省くため，上述の性質を満たすようなグラフのうち， $|V|$  が最小となるものを用意した．

このグラフ上で確率伝播計算により，各時刻で，仮説の確率を更新し，最終

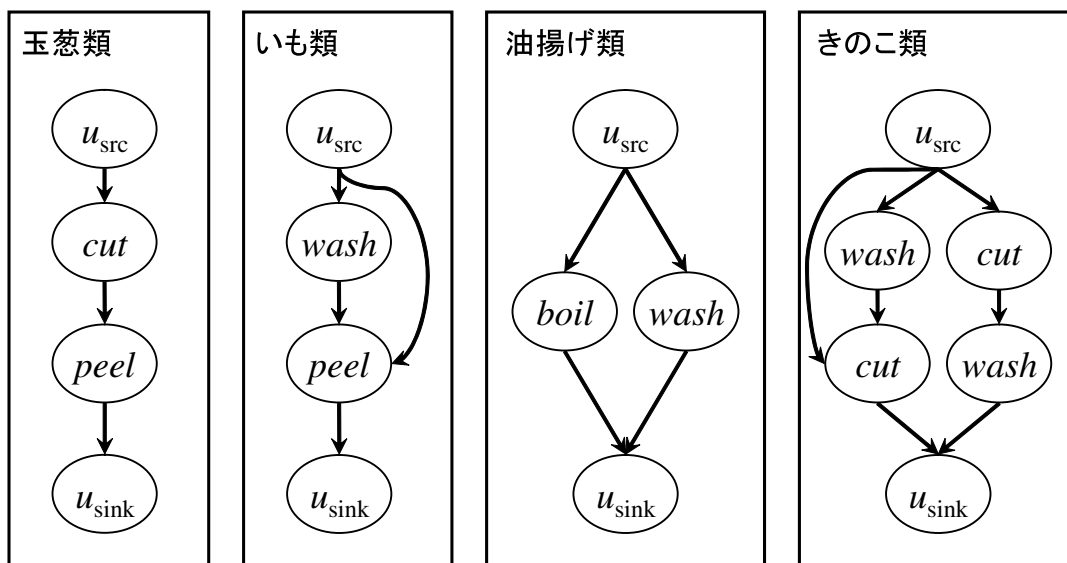


図 5: 4 つの茶色の食品群がもつ下処理のグラフ表現

的に下処理終了時点で  $\sum_{seq \in SEQ_f^{A_1}} P(seq)$  を得る方法を述べる．グラフ  $G^f$  上で， $u_{sink}$  以外の各頂点は，現在ユーザが行っている調理操作がその頂点に対応する作業である確率  $P(u)$  を保持する．ここで，頂点  $u$  に対応する作業とは， $u_{src}$  と  $u$  を結ぶ何れかのパスに対応する調理操作系列が行われたあとで行われる  $a(u)$  に対応する調理操作のことを指す．頂点  $u$  への枝を持つ全ての頂点の集合を  $Parent(u)$  とすると，各頂点に対応する作業が行われる確率  $P(u)$  は，直前に  $Parent(u)$  の何れかが行われており，かつ，現在行われている調理操作が  $a(u)$  である確率と，直前まで  $u$  に対応する作業が行われており，かつ現在も引き続き  $a(u)$  が行われている確率の和となる． $P_t(u)$  は次式で表される．

$$P_t(u) = P_t(a(u))\{P_{t-1}(Parent(u)) + P_{t-1}(u)\} \quad (18)$$

$$P(Parent(u)) = \sum_{v \in Parent(u)} P(v) \quad (19)$$

式 (18) は式 (3) を，全ての仮説に対して計算することに対応している．なお，この確率伝播計算の初期値は，まだ何も下処理をしていない状態を表すように， $P(u_{src}) = 1.00$  とし，それ以外の頂点への確率を 0.00 して行う．図 6 は，式 (18) に従った確率伝播の過程を示したものである．

$P_t(u)$  は  $u$  を通る何れかのパスに対応する下処理分類が  $u$  までが行われたという仮説に対する確率に他ならない．そのため，下処理が終了した時点をも  $t_{fin}$ ，食材クラス  $f$  へ下処理をモデル化したグラフ  $G^f$  上のシンクを  $u_{sink}^{G^f}$  とすると，



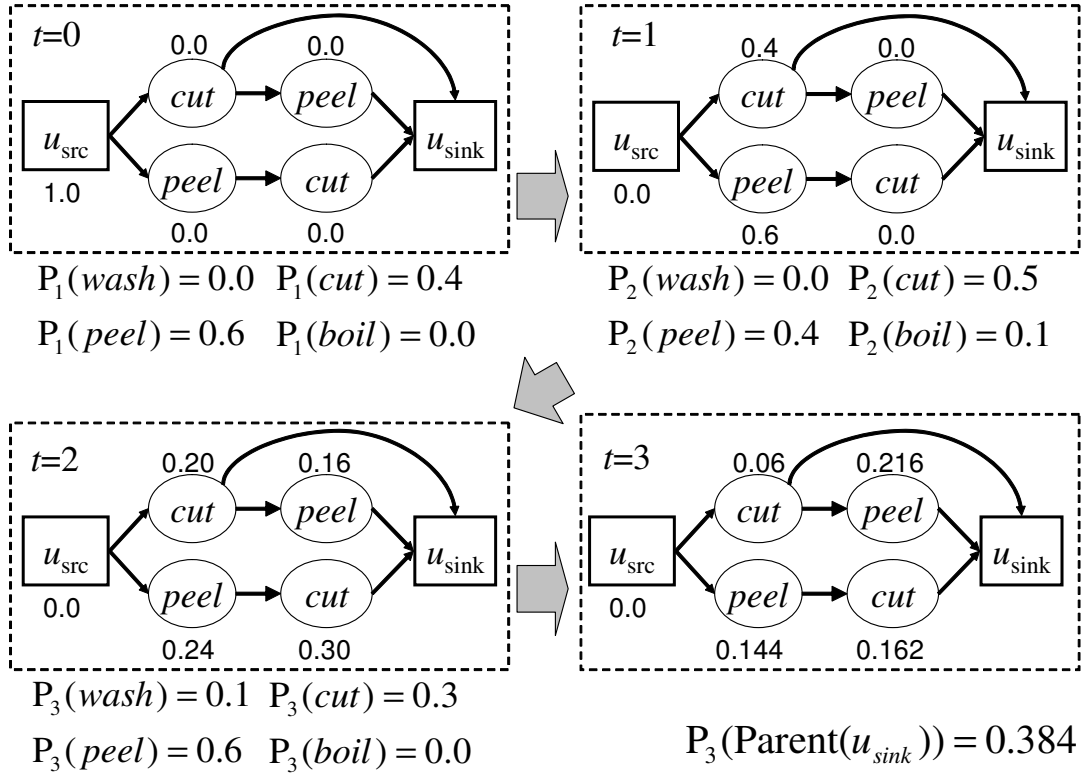


図 6:  $G^f$  上での確率伝播の過程

$\text{Parent}(u_{sink}^{G^f})$  を通るパスの集合は  $\mathcal{P}^{G^f}$  に等しく,  $P_{t_{fin}}(\text{Parent}(u_{sink}^{G^f}))$  は  $SEQ_f^{A_0}$  に含まれる下処理の何れかが行われた確率に等しくなる. 従って, 式 (4) は以下のように変形される.

$$\hat{f} = \operatorname{argmax}_{f \in F_R^c} P_{t_{fin}}(\text{Parent}(u_{sink}^{G^f})) \quad (20)$$

調理台上の食材に対する下処理が終了した時点で, 式 (20) により得られる  $\hat{f}$  を認識結果として出力することで, 下処理を利用した食材認識が完了する.

実装に際しては, 次に述べる, 調理操作の内容とは無関係な動作による影響を考慮した. 調理を観測する場合, 調理操作中でも, 手を伸ばして, まな板から転がり落ちた食材を取るなど, ユーザが  $A_1$  の調理操作クラスとは無関係の動きする場合がある. この無関係な動作から抽出された特徴から計算された  $P(a)$  が入力されると, 実際に行われている調理操作とは関係のない調理操作と誤認識されてしまう場合があり, 式 (18) の計算の際に, 実際にユーザが行っている作業と対応している頂点  $u$  の保持する確率  $P(u)$  が小さな値になってしまう. その場合には, Left-to-Right の今回のグラフモデルでは回復手段がなく, 式 (18)



の計算を行う前に，無関係な動作による影響を受けにくくする工夫が必要となる．この無関係な動作は多くの場合，1，2 フレームだけ連続して観測される．そのため，特徴量からの  $P(a)$  の計算結果を， $Th_5$  フレーム毎に平均した  $\bar{P}(a)$  を用いて， $Th_5$  フレーム毎に式 (18) の確率伝播計算を行うことで，各フレームから出力される  $P(a)$  の平滑化を行い，無関係な動作による影響を抑えた．なお， $Th_5$  の値は手動で与えた．

## 第5章 実験

第4章で実装した食材認識器を用いて、実際にスマートキッチンで行われた下処理の観測を通じた食材の認識を試みた。下処理分類の食材認識への有効性を確認するため、画像特徴である色については誤りなく抽出できるものと仮定して、茶色の食材のみを対象として実験を行った。対象とした食材は表2のうち、季節の問題から入手が困難な「筍」(茶の場合は皮付きのものであり、春にしか市場に出回らない)、「長時間水につける」など、特殊な下処理をもつ「貝類」と「乾物」、そもそも下処理がなされない「加工食品」を除いた4種類の食品群「玉葱類」「いも類」「きのこ類」「油揚げ類」を用いて実験を行った。それぞれの食品群に対応する下処理は図5で示されている。

実験に際して、サンプルとしては、それぞれの食品群から9個ずつ、合計36個の個体を用意した。また、下処理を行うユーザとして被験者を一名用意した。これらのサンプルに対して、使用する調理器具を替えながら、被験者の経験から自由に下処理を行ってもらった実験を行った。その際、交差確認法 (cross-variation method) として、各食材1個体ずつ、計4個体をテストサンプルとし、それらを除いた32サンプルから獲得されたGMMとCPTを用いて、テストサンプルを認識する操作を繰り返し、全36サンプルについて提案手法を適用し、食材認識を行った。なお、4.3で行われた特徴量選抜では、各食材から無作為に5個ずつ選んだ、計20個のサンプルを用いたものであり、実験でも表3の特徴量と表4のCPTを用いた。

また、この手法を適用するには、ユーザが調理操作中であるかどうかの判定を別途行う必要がある。この判定については、操作対象の食材が固定されているために、調理操作中にはユーザの動く範囲が限定的になること、また調理操作の直前、直後には調理操作対象の食材をそれぞれ作業領域内、作業領域外へ移動するために、ユーザの動く範囲が広くなることから、計算機による判定も将来的には可能だと考えている。しかし、今回の認識器とは別に、調理操作の開始前後、終了前後のユーザの動きの時系列変化を特徴として用いる必要があり、下処理の内容を対象とした今回の問題とは異なる認識問題となることから、本研究では人手で正解を与えている。なお、観測は平均3.75 frame/secで行い、特徴量抽出の際のフレーム間差分から得られる二種類の画像  $I_1, I_2$  の二値化のための閾値を  $Th_1 = Th_2 = 10$ 、矩形のノイズ除去のための閾値を  $Th_3 = Th_4 = 60$

とし、各調理操作が行われた確率の平滑化と、確率伝播計算を行う周期である  $Th_5$  を 20 フレームとした。

この実験結果を表 5 の正誤表にまとめた「油揚げ類」については全て正しく認識することができたことが表から確認される。この「油揚げ類」は *cut* や *peel* が行なわれない。一方、他の食材クラスでは必ず *cut* か *peel* のいずれかが行われる (図 5)。従って、*cut/peel* とそれ以外の調理操作クラス (*wash*, *boil*) はある程度頑健に区別が出来ているものと考えられる。それ以外の食材クラスでは、例えば「きのこ類」のサンプルのうち、「いも類」と誤認識されたサンプルでは、*cut* が行われていた時間が短かったため、短時間の観測で *peel* と誤認識されたものが、そのまま操作系列の識別結果の誤認識を引き起こしていたことが、実際のデータから確認できた。同様に「たまねぎ類」では最初に施された *cut* が *peel* と誤識別されたことに起因すると思われる「いも類」への誤認識が起こっている。また「いも類」では 3 つのサンプルで「きのこ類」への誤認識が見られた。これは、1 つのサンプルについては、調理者が皮を剥き終わった後で食材を入念に洗うという、下処理のモデルにはない行動を取ったことに起因するものであり、残り 2 つについては、包丁を使って *peel* が行われている間中、常に調理者の頭が手元を隠していたために、調理操作識別結果に *peel* が行われている確率が低い特徴量しか観測されず、繰り返し観測しても正しく認識が出来なかった。

まとめると、今回の実験で誤認識の原因となったものは、主に *cut* が行われている時間が短いために、十分な回数、特徴量を抽出できなかった場合と、手元が操作中常に調理者の頭により隠されてしまい、手元の動きを観測できなかった場合であった。これらの原因は研究当初は予想していなかったものであるが、これらの原因に対しては、下処理で行われる *cut* と *peel* は、実行するのにかかる時間が *cut* の方が短い場合が多く、また手元が隠されていても実行時間は観測可能であることから、実行時間の長さを考慮して下処理グラフ上での確率伝播計算を拡張することで対処が出来ると考えている。正しく認識された事例の中には、常に正しく行われた調理操作を識別することが出来る事例も存在したが、提案手法が改善を狙った、調理操作が行われている間中、常に正しく調理操作識別が出来ていないような例でも、提案手法により、食材は正しく認識されている事例が確認された。

なお、今回の実験では、同じ被験者が行った下処理を実験データとして用い

表 5: 食材の認識結果を示す正誤表．縦軸が正解食材，横軸が認識結果を表す

正解 \ 認識結果	玉葱類	いも類	きのこ類	油揚げ類	正答率	サンプル数
玉葱類	66.7%	33.3%	0.0%	0.0%	66.7%	9
いも類	0.0 %	66.7%	33.3%	0.0%	66.7%	9
きのこ類	0.0 %	11.1%	66.7%	22.2%	66.7%	9
油揚げ類	0.0 %	0.0%	0.0%	100.0%	100.0%	9
平均正答率					75.0%	36

た．これは一般家庭では調理を行う人物が限定されていることから，認識器がクラスタリングや手入力による誤り訂正などを利用してその人物に適応していった場合に見込まれる性能を評価したものである．システムがその家庭に導入され，運用が開始された時点では個人に適応していないため，そのような場合の性能評価についても考える必要がある．また，特徴量選抜に用いた 20 サンプルが，それ以外のサンプルよりも有意に高い認識率となるようなことは起こらなかった．

## 第6章 結論と今後の課題

本研究ではレシピガイドシステムを想定した食材認識手法について提案を行った。食材を画像特徴のみから認識するのは難しいことから、既存の食材認識手法ではレシピに書かれた手順と実際にユーザが行った手順とを対応付けることで食材を認識していた。しかし、レシピガイドシステムでの利用を前提として食材認識を行う場合には、レシピ手順の開始よりも前に食材が認識できている必要があり、レシピ手順を観測しなければならない既存の手法を適用することが出来なかった。そこで、本研究ではレシピに関わらず行われる下処理の手順を利用することでこの問題を解決した。

下処理を利用した食材認識の実現可能性を調べるため、まず、Web から収集した 100 個のレシピにおいて、同じレシピに登場する同色の食材組合せが下処理により区別可能であるかどうかを調べた。また、その調査の中で、区別可能であるかどうかの判定に用いる調理操作分類を徐々に詳細にしていくことで、食材認識に必要な十分な詳細さを持つ調理操作分類を定めた。その結果、茶色の食材では、82.5%の事例が下処理で区別可能であると判断された。また、下処理の中で行われる調理操作を *wash, cut, peel, boil* の 4 つに分類することで、食材認識に必要な十分な詳細さを持つ下処理の分類を行った。

次に、実際にそれらの下処理分類を利用して、下処理の観測映像から食材を認識する手法を提案した。上述の下処理分類をする際には、短時間の観測では区別が難しい場合がある調理操作クラスを含んだ調理操作系列を識別する必要がある。そこで、各時点での観測特徴量から、その時点で行われていた調理操作を一意に特定することは避け、各調理操作が行われていた確率をもとに下処理全体を通して行われた調理操作系列を識別し、食材を認識した。実際に茶色の食材に対して下処理を行い、平均で 75.0%の認識率を達成した。

実際のレシピガイドシステムでこの手法を利用する場合には、ユーザが食材に対して、何らかの調理操作を行っているかどうかを認識する必要があり、また、なんらかの方法で下処理の終了時点を把握する必要がある。前者については、調理中には操作対象の食材が固定されているために、ユーザの動く範囲が限定的になること、また調理操作の直前、直後には調理操作対象の食材を移動するためにユーザの動く範囲が広くなることから、認識が可能だと考えている。後者についても、システムにとってレシピは既知であるため、レシピ手順の最

初のステップが行われたかどうかを認識することで、あるいは下処理終了時点でのユーザからの問い合わせを受け付けることで、ユーザが下処理を終了してそのままレシピ手順を行った場合と、ユーザが下処理終了時点でレシピ手順をレシピガイドシステムに問い合わせた場合の両方に対応可能であると考えている。また、本稿では6色の色分類は頑健に識別可能であると仮定したが、実際には各色の境界付近で色の誤識別が起こる可能性がある。そのため、色が未知の食材に対して本稿で提案した下処理を使った認識手法を適用する場合には色の確信度も考慮する必要がある。この、調理操作中かどうかの判定、下処理終了時点の検出、色の誤識別を考慮した手法の提案については今後の課題である。

## 謝辞

本研究を行うにあたり多くの御教示，熱心な御指導を賜りました美濃導彦教授に深く感謝致します。また，本論文をご査読いただき有用な御意見を賜りました角康之准教授に深くお礼を申し上げます。さらに，研究について数多く助言を頂き、本報告書の作成においても御指導を頂きました美濃研究室の角所考助教授，船富卓哉助教に感謝の意を表します。最後に，本研究を進めるにあたり貴重な議論の場を提供していただきました認識グループの方々，並びに，美濃研究室の皆様にお礼を申し上げます。

## 参考文献

- [1] Ju, W., Hurwitz, R., Judd, T. and Lee, B.: CounterActive: an interactive cookbook for the kitchen counter, *Conference on Human Factors in Computing Systems*, pp. 269–270 (2001).
- [2] Bradbury, J. S., Shell, J. S. and Knowles, C. B.: Hands on cooking: towards an attentive kitchen, *Conference on Human Factors in Computing Systems*, pp. 996–997 (2003).
- [3] Hamada, R., Okabe, J., Ide, I., Sakai, S. and Tanaka, H.: Cooking navi: assistant for daily cooking in kitchen, *Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 371–374 (2005).
- [4] 森 直幸, 船富 卓哉, 山肩 洋子, 角所 考, 美濃 導彦: 調理者の手の動きを時空間制約とした調理中の食材追跡, *信学技法*, Vol. 107, No. 454, pp. 45–50 (2008).
- [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 D*, Vol. Vol. j90-D, No. 10, pp. 2817–2829 (2007).
- [6] 椎尾 一郎, 浜田 玲子, 美馬 のゆり: Kitchen of the Future : コンピュータ強化キッチンとその応用 (〈特集〉インタラクティブソフトウェア), *コンピュータソフトウェア*, Vol. 23, No. 4, pp. 36–46 (2006).
- [7] 成瀬宇平: 下ごしらえ便利事典 食材の知識と仕込みの基本, 柴田書店 (2005).
- [8] 小林公成: キッチンの必需品! 食材の下ごしらえと保存, 世界文化社 (2007).
- [9] 八代義剛: 材料の下ごしらえ百科, 主婦と生活社 (2000).
- [10] Jain, A. and Zongker, D.: Feature Selection: Evaluation, Application, and Small Sample Performance, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 2, pp. 153–158 (1997).