

修士論文

調理操作に起因する  
画像特徴変化のモデルを用いた  
食材領域の追跡

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

森 直幸

平成 21 年 2 月 6 日

## 調理操作に起因する 画像特徴変化のモデルを用いた食材領域の追跡

森 直幸

### 内容梗概

調理の様子を観測した映像アーカイブの各シーンに対し，そのシーンで使用されている食材名を注釈として付与するためには，調理者が扱っている食材を常時特定している必要がある．調理中の食材は，切断・混合・加熱などの操作によって外見が変化するので，食材を継続的に追跡することによって個体を特定するアプローチをとる．

操作時に容器を用いる混合・加熱の操作では，容器に取り付けたセンサなどを用いて食材を追跡することが可能である．一方，手で直接行う操作である切断では，センサ情報による追跡が不可能である．そのため本研究では切断などの直接手で操作を施している最中の食材を追跡することを目的とする．

食材は切断されると，多数の断片になったりひとまとめにされたりするため，映像中での食材領域は，多数の分裂と統合を繰り返す．追跡を実現するためには，各フレームで分裂・統合した領域をフレーム間で対応付ける必要がある．しかし，分裂・統合した領域を対応付けるとき，その候補は膨大になりうるし，各領域の画像特徴は切断とともに大きく変化しうる．そのため，考えうるすべての候補から画像特徴のみを用いて正しい対応付けを求めるのは困難である．

従来手法では，各時刻で分裂・統合しうる食材領域が1種類のみであるという制約のもと，追跡を実現していた．対応付け候補が多数あっても，画像特徴が変化する食材領域が1種類であることがわかれば，対応付けは容易である．しかしながら，この制約は，実際の調理では実現性が低い．

本研究では実際の調理での有効性を考慮して上述の制約を緩和し，同色の食材については分裂・統合しうる食材は1種類である，という制約のもと，追跡を行う手法を提案する．制約の緩和により，複数種類の食材についての対応付けが一意に定まらなくなる．そこで，食材領域の画像特徴変化モデルを導入することで，食材領域の対応付け候補について尤度を求め，尤度が高い候補を採用することで，対応付けを定める．

以上で述べた手法の有効性の確認のため，画像特徴変化モデルの妥当性を評価する実験を行った．実際の調理を観測した映像から，食材領域の画像特徴変

化モデルにより，異色の食材からなる領域間の対応付け候補を棄却できることを示した．また，提案制約が実際の調理でどの程度満たされるかについて，調査を行った．多数のレシピをもとに，従来手法と提案手法が導入している制約が成立する割合を検証し，実際の調理における有効性が向上することを確認した．

# Food Tracking under Image Feature Transition Caused by Cooking Process

Naoyuki MORI

## Abstract

Annotation of food stuff to a cooking video is convenient for searching scenes in which the food stuff is used. In order to annotate food stuff in the video, identification of the food stuff is required in the whole video. Food stuff undergoes various changes through cooking processes: cut, mix, boil, etc. Since it is difficult to recognize food stuff in a single captured image, we track the food stuff from the beginning of the cooking.

In a scene of mixing or boiling food stuff, the cook puts the food stuff in some container: bowl, pan, etc. We can track easily the food stuff by a sensor attached to the container. In a scene of cutting food stuff, the cook manipulates the food stuff directly by the hands. We can not attach sensors directly to the food stuff. Thus we aim to track food stuff handled directly by cook's hands in a video.

The cooking processes mentioned above will split the food stuff into pieces, or gather up pieces of food stuff. This is observed in the video as split and merge of regions of the food stuff. Tracking the food stuff in the video requires finding correspondence of split and merge of regions between frames. The number of candidates of the correspondence in the whole video can be vast and the appearance of the regions of each correspondence can be different between beginning and end. Therefore, it is difficult to determine the correct correspondences from all the candidates by their appearance between beginning and end.

Previous work introduces the constraint that the cook processes only one kind of food stuff. Under the constraint, the number of species of food stuff is one that causes split and merge of the regions simultaneously between frames. Since, under the constraint, a number of candidates in a frame originate from only one species of food stuff, finding correspondence is easily obtained. However, the constraint is often unsatisfied.

In this work, we relax the constraint to one that is easily satisfied. We introduce the constraint that the cook processes separately the food stuff that

has different appearance. For reducing ambiguity in finding correspondences caused by the relaxation, we introduce an appearance transition model of food stuff. We adopt the candidate with the highest likelihood among the candidates, and thus tracking food stuff is obtained.

We implemented the method mentioned above, and evaluated the capability of calculating the likelihood and determining the correct correspondence. In addition, we assessed the possibility of the proposed constraint satisfied. The results showed that the method can increase the feasibility in real activities.

# 調理操作に起因する 画像特徴変化のモデルを用いた食材領域の追跡

## 目次

第1章	緒論	1
第2章	本研究が対象とする調理	4
2.1	調理を観測する環境の設定	4
2.2	想定する調理	5
第3章	調理中の食材と従来手法が用いた制約	7
3.1	調理映像中の食材追跡問題の分析	7
3.1.1	レシピに記載される調理操作と食材追跡する上で扱う調理操作	8
3.1.2	消滅ノードと湧出ノード	8
3.1.3	調理操作による食材領域の分裂	9
3.1.4	調理操作による食材領域の統合	11
3.2	従来手法の検討	12
3.2.1	従来手法における食材領域の分裂に対する処理	13
3.2.2	従来手法における食材領域の統合に対する処理	14
3.2.3	従来手法のまとめ	14
3.2.4	従来手法の問題点	15
第4章	食材領域の追跡	17
4.1	本研究での調理に対する制約	17
4.2	調理者の手による食材の遮蔽の検出	18
4.3	画像特徴変化モデルの導入による食材領域の対応付け	19
4.3.1	食材領域を表現する画像特徴量の選定	19
4.3.2	食材の見えの変化と画像特徴の分布	20
4.3.3	尤度関数の設計	20
4.4	時間コンテキストを利用した物体領域抽出	22
4.5	追跡アルゴリズム全体の流れ	23
第5章	実験・評価	24

5.1	実験環境	24
5.2	実験 1:画像特徴の変化モデルに対する尤度関数の評価	24
5.3	実験 2:提案制約の適用可能性の検討	27
5.4	実験 3:実際の調理映像中での食材追跡	28
<b>第 6 章</b>	<b>結論</b>	<b>31</b>
	謝辞	33
	参考文献	34
	付録	A-1
A.1	実験で使用したレシピ	A-1
	A.1.1 カレー	A-1
	A.1.2 筑前煮	A-1

## 第1章 緒論

近年，膨大な量の映像アーカイブが蓄積されるようになってきた．映像アーカイブに対し，映像中に登場する物体の種類や物体に対する注釈を付与することで，目的のシーンを検索することができれば有用である．調理の様子を観測した映像アーカイブでは，映像中の物体に対する注釈として，その調理で使用される食材の注釈を付与しておくことが考えられる．閲覧者がこの注釈を用いてアーカイブを検索することで，特定の食材の扱い方を知ることが可能となる．たとえば「たまねぎ」というクエリに対しては，映像の中から食材「たまねぎ」の注釈が付与されたたまねぎを扱っているシーンを抽出して閲覧者に提示することが考えられる．

映像に対して食材の注釈を自動的に付与するには，映像中で食材種類を特定することが必要である．これを目指した研究としては山肩ら [1] の研究がある．[1] では，レシピに記載されている調理操作の順序を制約として，調理アーカイブ中の食材を認識している．調理操作の順序を制約として利用するためには，各食材に対する調理操作の系列が必要である．このとき，各調理操作でどの食材が対象になっているかを得るため，調理操作間で食材を対応付ける必要がある．このため，映像中での食材追跡を必要要件としている．

映像中での物体追跡は，ある時刻の画像中の物体と，それ以降の時点での画像中の物体について，フレーム間の対応付けをとる問題である．対象物体として，人物や自動車といった物体を追跡する場合，フレーム間で，個々の物体を1対1で正確に対応付ける必要がある．対応付けの際に問題となるのは，対象物体が移動することにより，画像中での見えの変化が生じることである．フレーム間で「物体の見えの変化が微小である」「物体の移動距離は短い」といった制約のもと，物体を対応付ける．

一方，調理を観測した映像中で食材領域を追跡することを考えると，物体の見えの変化の仕方およびフレーム間での対応付け方に違いがある．食材に調理操作が施されるとき，食材は調理者の手に把持された状態で断片化し，それ以降，断片ごとに異なる時点で解放されうる．この把持・解放の間，食材は位置，姿勢，形状が変化する．これを観測した画像中では，食材領域が消滅し，それ以降，多数に分裂した食材領域がそれぞれ異なる時刻で湧出する，のように観測される．この消滅・湧出間で食材領域は，位置，見え，領域数が変化する．領



領域数の変化には、食材領域が分裂することにより領域数が増加する場合と、統合されることにより領域数が減少する場合がある。見えの変化には、皮を剥かれることにより物体の内部色が露出する場合、切断によって形状が変化することによる場合、などがある。食材の領域数の変化に対応するため、フレーム間で多対多の対応付けを行う必要がある。ただし、映像に対して食材種類の注釈を付与する目的であれば、食材の断片に該当する個々の領域を正確に対応付ける必要はない。食材の個々の領域の対応付けに誤りがあっても、食材の種類を正しく把握できればよい。

山肩らの手法 [1] では、以下の制約を課すことで食材追跡を実現している。

1. 同時に施される調理操作は 1 種類である
2. ひとまとまりの食材領域は 1 種類の食材で構成される。

以降の文章では「山肩らの制約 (1)」という呼称で参照する。この制約のもとでは、ある時刻で消滅した領域とそれ以降のもっとも近い時刻で湧出した食材は、同一種類の食材に該当する領域である。このような時間的に最も近接した消滅・湧出領域を最近傍の消滅・湧出領域と呼ぶことにすると、最近傍の消滅・湧出間に対応付けることによって、対応付けの候補の多さが問題とならず、追跡を実現することができていた。ただし、この制約 (1) の下では、作業の効率化のために左右の手で異なる食材を持って同時に移動させる、といった操作ができなくなる。また、全長が長い食材や、葉の面積が広い食材などでは、調理者が意図せず複数種類の食材を隣接して調理台上に置くことがありうるが、この操作ができなくなる。このような操作は多くの調理者にとって必要とされるものであり、これらの制約が実際の調理で成立するとは考えにくい。

調理操作が食材に施されるとき、その食材の見えの変化は、食材の種類および調理操作の種類に依存する。このため、画像中の食材領域の消滅・湧出間での見えの変化は、どのような変化でも起こりうるものではなく、変化の仕方として妥当ではないと判断できるものがある。また、複数種類の食材を調理台上に置くとき、それ以降の工程で混合する食材でなければ、必ずしも食材を同時に置くわけではなく、時間をずらして台上に置くことがある。これは観測画像中では、ある時刻で湧出した食材領域とそれ以降の時刻で湧出した食材領域が連結する、のように観測される。このため、連結領域であっても時間軸上でみたときに相違点があるため、別々の食材領域であると判断できるものがある。

本研究では、調理映像に対する注釈付与のため、調理映像中の食材を追跡す

ることを目標とする．そこで，現実の調理に対する実現性を高めるため，山肩らの制約(1)(2)を

- 1'. 同色の異種食材に対して同時に施される調理操作は1種類である
- 2'. 異種の食材が同時にひとまとめにして置かれることはない

という制約に緩和する．それぞれの制約を緩和することによって，次の2つの問題が起こる．(1) 調理操作による見えの変化をする食材領域が複数存在しうることになり，追跡が困難になる．この問題に対して，食材領域の見えの変化モデルを導入し，食材領域の見えの変化の妥当性を評価する．対応付けの際，フレーム間での食材領域の見えの変化が妥当でない対応付けを棄却する．これにより，妥当な見えの変化をしているフレーム間の対応付けのみを残すことができる．(2) 異種の食材が隣接して調理台上に置かれると，その時点での画像でこれらを分離することが困難になる．そこで，時間コンテキストを用いた領域分割を行うことで，これらの食材領域を分離する．

まず，2章では調理状況観測の設定と，本研究が対象とする調理について述べる．3章では，調理映像中の食材を追跡する上で生じる問題点についての分析と，問題に対処するため従来手法で課していた制約について述べる．4章では，本研究が従来手法の制約を緩和し，調理操作に起因する食材の見えの変化モデルを導入することによって，対応付けを行うことを述べる．5章では，実際の調理を観測した映像に対して，本研究の追跡手法の有効性を検討する．最後に6章で，結論と今後の課題を述べる．

## 第2章 本研究が対象とする調理

### 2.1 調理を観測する環境の設定



図1: カメラの設置位置による見え方の違い



図2: 調理台上方に設置したカメラで観測される画像例

本節では、調理中の食材を観測するためのカメラの設置位置について述べる。通常、調理台は室内に固定されているので、固定カメラによって調理の作業が行われる領域を観測することができる。すると映像中には、食材、調理者、調理器具、調理台が映る。映像中で調理台は固定されているので適当な背景モデルを用いて調理台を除去することができる。背景を除去した映像中には、食材、

調理者，調理器具が前景領域として残る．

画像中から抽出した対象物体を見失うことなく食材を追跡するためには，食材が調理者の手や調理器具による遮蔽ができるだけ起きないようにする必要がある．観測画像中で食材が遮蔽されるのは，(a) 食材同士の遮蔽，(b) 調理者による食材の遮蔽，(c) 調理器具による食材の遮蔽，の3パターンである．カメラを調理台の上方(図1)に設置することで，(a) 食材同士の遮蔽，を防ぐことができる．(b) および (c) の遮蔽については，4.2 節で対処方法を述べる．実際に調理台を観測した画像を図2に示す．白色の広い領域が調理台天板であり，調理台の上端よりも上側の領域は床面である．

ここで，第1章で述べたような調理操作の技術を知ることが目的とした映像コンテンツでは，調理台上方以外から撮影した映像のほうが目的にかなっている場合がある．例えば，包丁で食材を切断するときの様子を閲覧するときには，真上からの映像だと包丁の形状上，包丁の動きを把握することが難しい．横方向から撮影した映像であれば，包丁の動きがよくわかる．そこで，調理台上方から撮影した映像は食材の注釈を付与することを目的とした食材追跡に使い，その注釈を用いて検索などを行う目的には別の方向から撮影した映像を別途用意することとする．上方向および別方向から撮影した映像にはタイムコードを付与することで時間同期をとる．

## 2.2 想定する調理

本研究が対象とする調理に対する想定を述べる．調理で扱うメニューとしては，山肩らの研究[1]で使用されたものと同じの，Web から収集した100種類のレシピを用いる．このレシピに記載してある事項は既知のものであるとし，調理中で使用される食材および行われる調理操作の順序をシステムが保持していることを仮定する．

調理ではさまざまな用途で調理器具(包丁，まな板，皮むき器，ボール，バット，皿，鍋など)が用いられる．これらの物体は映像から食材を抽出する際，食材と共に抽出されてしまう．そこで，これらの器具には接触型の位置センサを装着することで位置を検出し，食材と峻別する．調理器具は形状や色が変わらないので，他の調理器具と並べたり重ねたりして調理台上に置かれていても，検出は可能である．調理者が把持している調理器具によって食材が遮蔽されてしまう場合には，追跡対象の食材を観測することができなくなる．この場合には，

遮蔽を検出し，その前後で食材を対応付ける．詳細は 4.2 節で述べる．接触型の位置センサとしては，超音波位置センサやマーカ―など [2][3] を使用することが考えられる．以降，映像中から調理器具が除去されているものとして議論する．

これらの調理器具のうちボールやバット，鍋などでは，容器に食材が入った状態で調理操作が行われる．容器に食材が入っていることがわかれば，上述の位置センサによって調理が行われている位置および食材の位置を検出できる．そこで，混ぜる，加熱する（煮る，焼く，など），容器に入れて移動させる，の調理操作が施されている間の食材は，調理器具の位置センサからの情報によって追跡する．

調理操作について，次にあげるものを想定する．

- 切る
  - 皮をむく
  - 手で直接移動させる（以降，誤解を生じない限り「移動させる」と記す）
- 「切る」および「皮を剥く」では調理操作の対象となる食材は 1 種類である！「移動させる」では，複数個の対象食材を同時に移動させうる．

調理者が食材に調理操作を施すとき，調理者の手が食材を隠し，その後食材が再び現れると領域数や形状，見えが変化する．この点では「調理者の手が食材を遮蔽する」ときにも同様の変化があるため，これらの調理操作と混同してしまう．そこで「調理者の手が遮蔽する」については，操作の前後で食材に見えの変化がほぼないことを用いて別途検出処理を行い，混同を防ぐ．詳細は，4.2 節で述べる．

これらの調理操作では，容器に入れられた状態で施される調理操作と異なり，食材の位置を接触型センサで取得することができない．そのため，非接触型センサであるカメラで得た情報で追跡する必要がある．

## 第3章 調理中の食材と従来手法が用いた制約

本章では，調理における食材追跡の問題について述べた後，従来提案されている手法の有効性と問題点について述べる．

### 3.1 調理映像中の食材追跡問題の分析

食材追跡とは，調理映像の各フレームで食材領域を抽出し，それをフレーム間で対応付ける問題である．フレーム間での対応付けを難しくするのは，画像中での食材の領域数の変化および見えの変化である．

食材の領域数の変化は，食材が切断されて断片に分裂すること，および，断片がひとまとめに統合されること，によって起こる．分裂の際，画像中では，1個の食材領域が多数の領域に分裂するが，分裂後に何個の領域になるかは未知である．統合の際，画像中では，多数の食材領域が1個の領域に統合されるが，統合によって何個の領域がまとめられるかは未知である．以上から，分裂・統合では，フレーム間で領域数未知の多対多の対応付けが必要である．ただし，調理映像に対する食材種類の注釈付与という目的の下では，食材の種類さえ正確であればよい．すなわち，切断により生じた食材の断片に該当する個々の領域の対応付けは正確でなくても問題とならない．

調理操作が食材に施され，位置および姿勢に加えて形状が変化することによって，食材領域の見えの変化が起こる．特に，かぼちゃのように外部と内部で色が異なる食材は，切断によって内部色が露出するため，見えが大きく変化することがある．

一方で，調理操作が施されていない食材は，画像中で該当する領域に位置・見え・領域数のいずれも変化もない．調理者の手が食材領域付近にあるとき，キャストシャドウによって見えが変化しうるのみである．追跡対象物体のこのような変化は，従来の物体追跡問題で扱われていたため，食材追跡をする上では本質的な問題ではない．

以上のことから，調理操作を受けた食材の領域数および見えの変化をいかに対応付けていくかが，食材追跡の問題である．調理では「切る」などによって食材が多数の断片になるため，映像中に現れる分裂した食材領域の個数は膨大である．また，食材の断片を「移動させる」ことなどでひとまとめにすることにより，統合される食材領域の個数の変化も膨大である．これらすべての領域

について，見えの変化によって対応付けを求めようとする，その計算の量は膨大なものになってしまい現実的でない．そこでまず，食材領域の分裂・統合による食材領域数の変化に対処し，対応付けられる可能性のある領域の数を減らす．その後，見えの変化に対して処理を行うことにする．

食材が調理操作を受けている最中，画像中の食材領域は調理者の手によって完全にあるいは部分的に隠蔽される．このため，この間の食材は追跡を行わない．調理操作の前後で手に隠蔽されていない食材の見えの変化から，食材の対応関係を求める．

### 3.1.1 レシピに記載される調理操作と食材追跡する上で扱う調理操作

食材追跡問題における，調理操作の粒度について述べておく．食材に調理操作「切る」がほどこされて，画像中で食材領域数が変化する例を図3に示す．時刻  $[0, T]$  間の調理で，時刻  $[t_0, t_5]$  の部分に注目している．図中の黒い丸の1個が，ある時刻の画像中で観測される食材領域1個を表す．

レシピに記載してある調理操作「切る」は，食材個体を細かい断片にするまでの，開始・終了間で行われる操作を指す．図中では，時刻  $t_0$  から  $t_5$  の区間に該当する．他方，フレーム間で食材の領域数や見えが変化する場合には「真っ二つに割る」「細かく刻む」「断片をまとめる」といった，より細かい粒度の動作が行われている．図中では，食材  $U, W$  間，および  $B, C$  間でそれぞれ細かい粒度の操作「真っ二つに割る」「断片をまとめる」が行われている．これらの動作を「動作プリミティブ」などと読んで詳細な検討を加えている研究がある [4] が，本研究で行う食材追跡では，このような細かな動作を分類する必要はない．そのため，食材  $U, W$  間，および  $B, C$  間でより細かい粒度の動作が行われている箇所についても，調理操作「切る」が施されることによって，食材に変化が生じているとする．

### 3.1.2 消滅ノードと湧出ノード

以降の議論のため，術語の定義を行う．調理操作によって調理者の手に隠蔽されてしまう直前の食材を消滅ノードとする．調理操作による消滅ノードが現れたら，その時刻が調理操作の開始時刻である．

調理操作中の調理者の手による隠蔽から解放された直後の食材を湧出ノード，とする．ここで，湧出ノードによって調理操作の終了を検出できない点で，消滅ノードとの対称性がないことに注意を要する．調理者がある食材に調理操作を施すとき，同時に他の食材を調理者の手で隠蔽することが起こりうる．ある

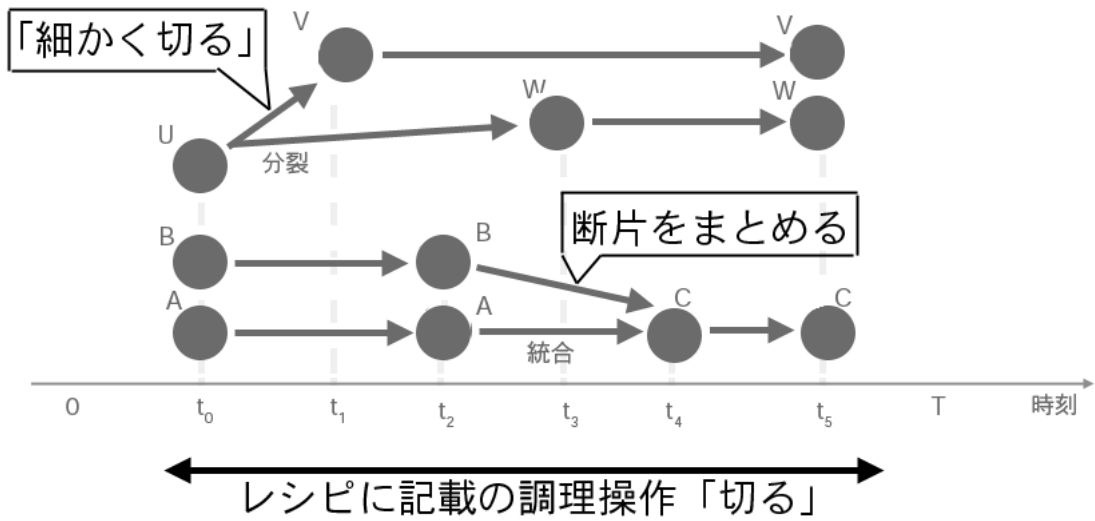


図 3: 調理操作による食材領域数の変化の模式図

時刻で湧出ノードが現れたとき，それ以降の時刻で調理者の手による隠蔽から解放されて湧出ノードが現れる可能性がある．すなわち調理操作の終了時点を求めるのは，open endの問題になっている．

### 3.1.3 調理操作による食材領域の分裂

調理操作によって食材が断片に分割され，映像中で食材領域が分裂する際の模式図を図 4 に示す．なお，図中で消滅ノード，湧出ノード，通常ノード（消滅・湧出のいずれでもないノード）を色分けしてある．後の図でも同様のノード表記を用いる．

図 4 では，次のような調理中の 1 シーンを想定している．時刻 0 で調理が開始し，時刻 T で終了する．調理台上に 2 個の食材 U, A があり，これらの食材が調理操作「切る」によって切断され，5 個の食材 V, W, Z, B, C に分裂するとする．

まず，時刻  $t_0$  で食材 U が切断され，3 個の断片 V, W, Z に分裂する．時刻  $t_0$  から  $t_1$  で調理操作「切る」が施され，この間，調理者の手によって食材が隠蔽され，映像中では観測されない．食材 U の断片 V は時刻  $t_1$  で調理者の手による隠蔽から解放され，映像中に現れる．この時点で食材 W, Z は分裂した状態で調理者の手に遮蔽されていて，観測映像中に現れない．食材 W, Z を遮蔽した状態で，調理者は時刻  $t_2$  で食材 A の切断を開始する．断片 W は時刻  $t_3$  で遮蔽から解放されて映像中に現れる．断片 Z は食材 A の切断およびその断片の分裂が終わってから，時刻  $t_6$  で遮蔽から解放されて映像中に現れる．



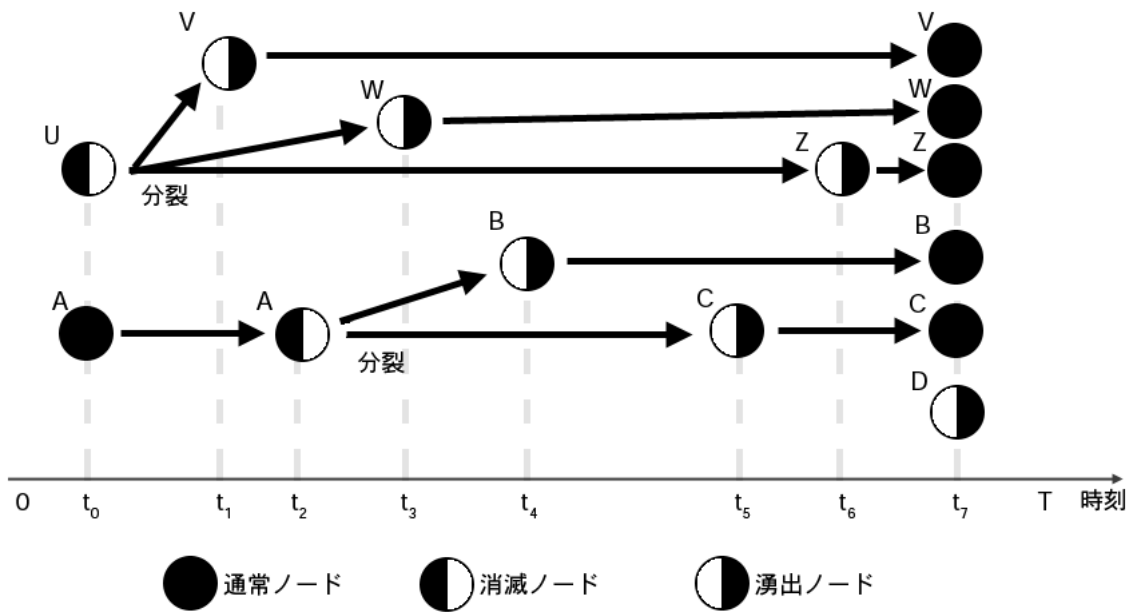


図 4: 分裂による食材領域数の変化の模式図

時刻  $t_2$  で調理台上にあった食材 A は「切る」によって切断され，2 個の断片 B, C に分裂する．食材 A の片方の断片 B は時刻  $t_4$  で調理者の手による隠蔽から解放され，画像中に現れる．他方の断片 C は時刻  $t_5$  で画像中に現れる．

調理操作によって食材は形状や位置が変化する．このため，U と V，U と W，U と Z，A と B，A と C は同一の食材であるが，見えが全く異なっている可能性がある．したがって，見えの類似度によって分裂前後の食材を対応付けるのは困難である．

すでに述べたように，分裂によって現れた個々の断片の対応付けは正確でなくても問題とならない．仮に U と A が同種の食材であるとすると，U と B，A と Z の対応付けによって，B および Z には正しく食材種類が割り当てられる．したがって U と B，A と Z の対応付けは正解としてよい．

一方，食材 U と A が異種の食材である場合，U と V，U と W，U と Z，A と B，A と C を正確に対応付けなければならない．

ここで，調理操作によって分裂したことが既知であるとして，時刻  $t_7$  で画像中に現れた新たな湧出ノード D について，対応付けを考える．図中の食材で湧出ノード D に対応付けられる消滅ノードとしては，U もしくは A がありえる．さらに，U よりも前の時刻で調理操作を受けた食材が，隠蔽から解放されて，時

刻  $t_7$  で湧出ノードとして出現することもありえる．したがって，一般的には，食材 D は調理が開始された時刻 0 以降での消滅ノードすべてと対応する可能性を考慮する必要がある．

ただし実際の調理では，食材が調理者の手に遮蔽される時間が極端に長いことは稀であるし，調理者が調理台を一時的に離れるためにすべての食材が遮蔽なく観測されることがしばしば起こる．上述の「すべての消滅ノードと対応する可能性」を何らかの制約で軽減することは可能であると考えられるが，本稿では触れない．

### 3.1.4 調理操作による食材領域の統合

調理操作によって食材がひとまとめにされ，映像中で食材領域が統合する際の模式図を図 5 に示す．

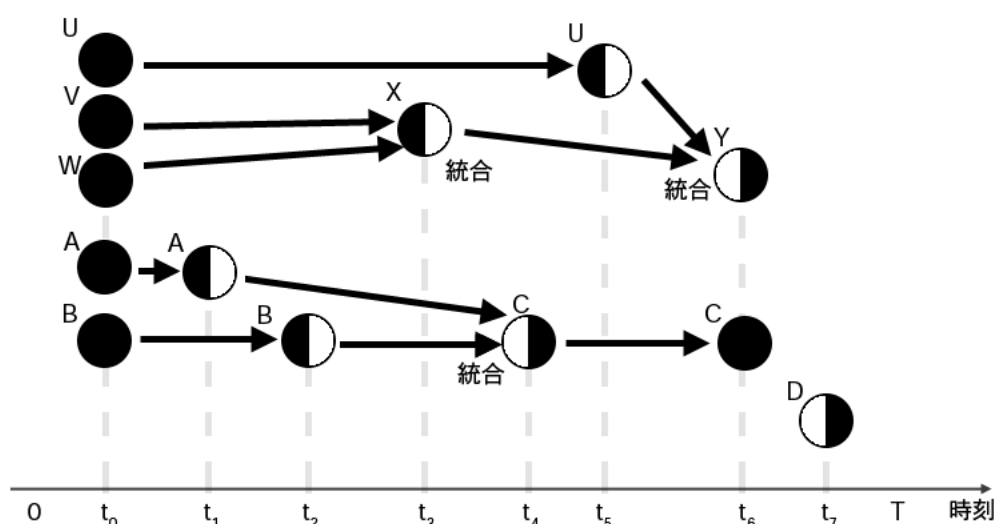


図 5: 統合による食材領域数の変化の模式図

図 5 では，次のような調理中の 1 シーンを想定している．時刻 0 で調理が開始し，時刻 T で終了する．調理台上に 5 個の食材 U, V, W, A, B がある．これらの食材が調理操作「移動させる」によってひとまとめにされ，U, V, W が食材 Y に，A, B が食材 C に統合するとする．

まず，時刻  $t_0$  で食材 V, W が「移動させ」られ，1 個の食材 X に統合する！「移動させる」の最中，食材 V, W は調理者の手によって隠蔽され，映像中では観測されない．食材 V, W を遮蔽した状態のまま，調理者は時刻  $t_1, t_2$  で食材 A, B の

統合を開始する．統合された食材  $X$  は時刻  $t_3$  で調理者の手による隠蔽から解放され，画像中に現れる．時刻  $t_3, t_5$  で食材  $X, U$  は「移動させ」られ，1 個の食材  $Y$  に統合される．食材  $Y$  が画像中に現れるのは，食材  $A, B$  の統合が終わって，時刻  $t_6$  で遮蔽から解放されたときである．時刻  $t_1, t_2$  で調理台上にあった食材  $A, B$  は「移動させ」られ，1 個の食材  $C$  に統合する．食材  $C$  は時刻  $t_4$  で調理者の手による隠蔽から解放され，画像中に現れる．

食材の統合の際，食材の個々の断片の対応付けは正確でなくても問題とならない．仮に，食材  $A, B$  が同種である場合について考える． $A$  と  $C$  もしくは  $B$  と  $C$  のいずれか一方の対応付けさえできていれば，もう片方の対応付けを行わなかったとしても， $C$  には正しく食材種類が割り当てられる．このように複数個の同種食材の統合の際には，そのうち 1 個の食材さえ正確に対応付けていて，かつ，それ以外の食材の対応付けを行わなかったとしても，正解としてよい．

一方，食材  $A$  と  $B$  が異種の食材である場合， $A$  と  $C$ ， $B$  と  $C$ ，を正確に対応付けなければならない．

ここで，調理操作によって統合したことが既知であるとして，時刻  $t_7$  で画像中に現れた湧出ノード  $D$  について，対応付けを考える．図中の食材では  $U, V, W, X, A, B$  のいずれか 2 つ以上が，食材  $D$  に統合したものと考えられる．しかし， $V$  よりも前の時刻で調理操作を受けた食材が，遮蔽から解放されて，時刻  $t_7$  で湧出ノードとして出現することもありえる．つまり，食材  $D$  は調理が開始された時刻 0 以降での消滅ノードすべてと対応関係を評価する必要がある．

ただし，前小節で述べたのと同様に，「すべての消滅ノードと対応する可能性」を何らかの制約で軽減することは可能であると考えられるが，本稿では触れない．

## 3.2 従来手法の検討

これまでに述べてきた状況設定と問題設定のもと，従来手法がどのように食材追跡にアプローチしていたかを検討する．

山肩らの手法 [1] では，膨大な数の消滅・湧出ノードに対処するため，調理に以下の 2 つの制約を課していた．これらの制約を課すことで，対応付けの可能性のある消滅・湧出ノードの数を削減していた．

1. 同時に施される調理操作は 1 種類である
2. ひとまとまりの食材領域は 1 種類の食材で構成される．

### 3.2.1 従来手法における食材領域の分裂に対する処理

まず，制約 (1) によって，食材の分裂に対処していたことを述べる．図 6 では，山肩らの制約 (1) が想定している，分裂による食材領域数の変化の例を示している．時刻  $0, T$  間で，食材  $A, U$  がそれぞれ切断され，食材  $V, W, Z, B, C$  に分裂している．山肩らの制約 (1) により，同時に施される調理操作が 1 つであることから，これらの調理操作は時間が重複しない．したがって，消滅ノード  $U$  で開始する調理操作は，遅くとも，消滅ノード  $A$  が出現する時刻  $t_4$  までには終了している．湧出ノード  $C$  が画像中に現れる時刻  $t_6$  において，その時刻に行われている可能性がある調理操作は，消滅ノード  $A$  から開始する調理操作のみである．したがって，湧出ノード  $C$  は消滅ノード  $A$  に調理操作が施されたことによって出現した食材であると判断できる．同様に，湧出ノード  $V$  は消滅ノード  $U$  に調理操作が施されたことによって出現した食材であると判断できる．このように，制約下 (1) では，湧出ノードに対してそれ以前の時刻でもっとも近い消滅ノードを対応付ければ，正しい対応付け得られることになる．この湧出ノード以前の湧出ノードとは対応付けを行う必要はない．

なお，食材の分裂の際には，制約 (2) は寄与しない．

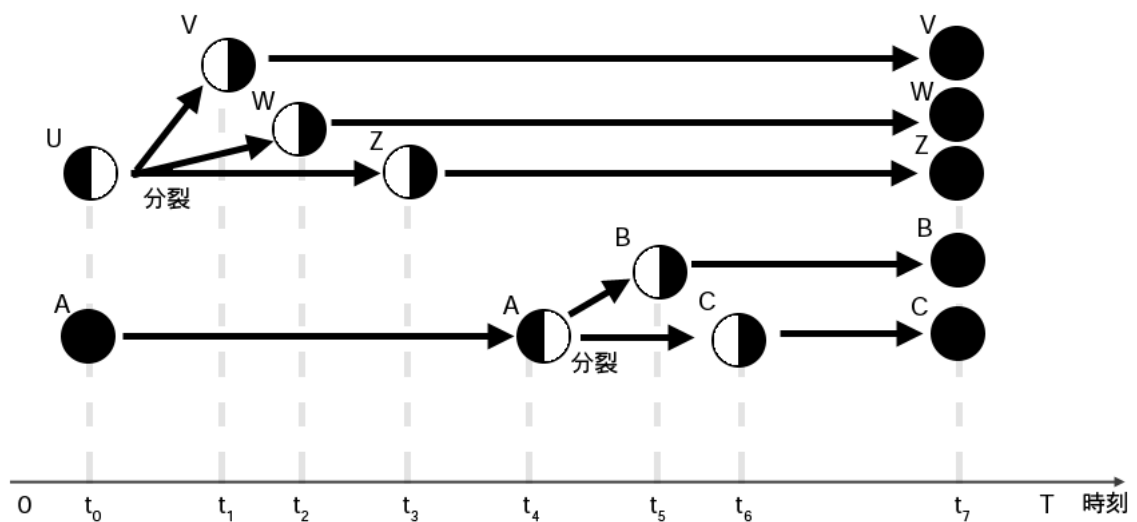


図 6: 山肩らの制約 (1) のもとでの，分裂による食材領域数の変化の模式図

### 3.2.2 従来手法における食材領域の統合に対する処理

次に、山肩らの制約 (1) および (2) によって、食材の統合に対処していたことを述べる。山肩らの制約 (1) が課されていない場合には、複数個の食材に同時に見た目の変化が起こりうるので、図 5 のように食材の統合が複数同時に起こる。山肩らの制約 (1) が課されている場合、同時に見た目の変化が起こる食材は 1 種類である。このとき、図 7 のように食材の統合は 1 個の食材でしか起こりえない。したがって、これらの調理操作は時間が重複しない。

制約 (2) が課されている場合、統合によって生じる湧出ノードは 1 種類の食材から構成される。図 7 の湧出ノード C では、C に統合する消滅ノード A, B は同種の食材である。既に述べたように、複数個の同種食材の対応付けでは、個々の食材断片の対応付けが間違えていても問題とはならない。湧出ノード C の対応付けでは、消滅ノード A, B のうちいずれか 1 個の食材さえ対応付ければ、食材 C の食材種類を正しく把握できる。それ以外の対応付けなかった消滅ノードは対応付けを行う必要はない。

この制約 (2) の実現性を高めるため、温度画像を用いたクロマキー処理によって、画像中から食材領域を抽出している。可視光画像から背景モデルを構築し、それを用いて背景差分を行う方法では、背景画像と食材の画素値が類似している場合、背景と食材を分離できないことがある。すると、近接して調理台上に置かれた異種食材領域が連結してしまうことが起こりうるため、制約が満たされない。そこで、食材が冷やされているとの仮定のもと温度画像を用いることで、常温の背景と低温の食材を分離している。

### 3.2.3 従来手法のまとめ

山肩らの手法が食材の分裂・統合に対処するため、2 つの制約を課していることを述べた。これらの制約下での消滅・湧出ノードの対応付けは、次のようなルールにまとめることができる。

- 分裂・統合のいずれの場合でも、湧出ノードをそれよりも前の時刻の最近傍にある消滅ノードに対応付ける

このルールのもとでは、ある湧出ノードに対応する可能性のある消滅ノードのうち、最近傍にあるもののみに対応付ければよい。したがって、膨大な数の消滅・湧出ノードに対処することができていた。

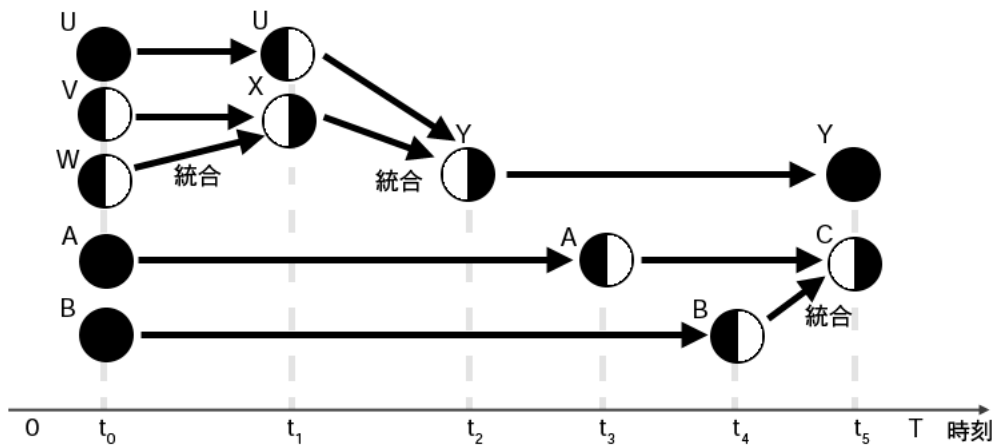


図 7: 山肩制約 (1)(2) のもとでの、統合による食材領域数の変化の模式図

### 3.2.4 従来手法の問題点

山肩らの制約について、現実性に疑問が残る。制約 (1) が成り立たない例を図 8 に示す。図 8 では、相異なる食材 A,U に対して調理操作が行われている。食材 U が時刻  $t_0$  で「切られる」ことによって、時刻  $t_1$  で分裂した食材 W が現れている。食材 A,B が時刻  $t_2$  で「移動させ」られることによって、時刻  $t_4$  でひとまとめに統合されて、食材 C となって現れている。つまり、食材 W が「移動させ」られるのが終了する前に、食材 A,B を「移動させる」が開始されている。これは同時に行われる調理操作が 1 種類であるという、山肩らの制約 (1) に反している。山肩らの手法の対応付けルールを適用すると、湧出ノード W に対応付けられる消滅ノードは、W が現れる時刻  $t_3$  よりも前にあって、かつ、W に最も近接した消滅ノードである。このため、湧出ノード W に対して、消滅ノード A もしくは B が対応付けられてしまい、追跡を誤る。

このように、ある食材に対する調理操作が終了する前に、他の食材に対する調理操作が開始されると、その切り替わりの際に、追跡を誤る。調理操作の切り替わりは、2 種類以上の食材を扱うレシピに含まれる。調理では、ほとんどのレシピが 2 種類以上の食材を扱うため、ほとんどのレシピで制約 (1) が成り立たなくなる可能性がある。この点において、山肩らの制約 (1) は現実的でない。

また、制約 (2) についても、異種の食材が隣接するように調理台上に置くことは、頻繁に起こりうる事象である。特に、ごぼうのように全長の長い食材や、ほうれん草など調理台を占有する面積が大きい食材を用いる場合、これらの食

材が意図せず隣接してしまうことが起こりうる．そのため，制約 (2) も現実的ではないといえる．

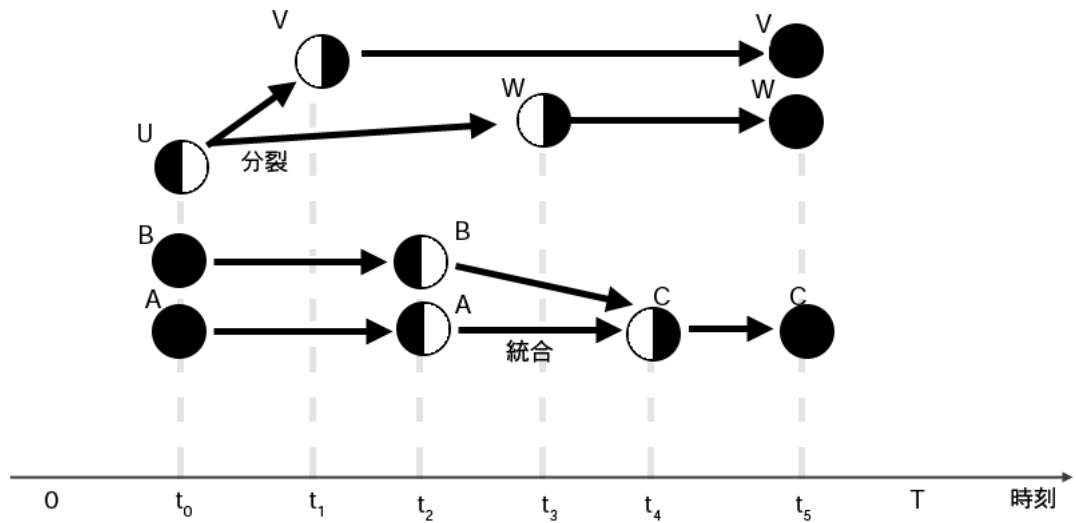


図 8: 山肩制約 (1)(2)のもと，分裂・統合が同時に起こって対応付けを誤る例の模式図

## 第4章 食材領域の追跡

本章では、まず、山肩らの制約を、現実性を考慮した制約に緩和することについて4.1節で述べる。この際、山肩らの制約(1)を緩和することで、対応付けのあいまい性が発生する。これに対処するため、調理者による食材の遮蔽を検出することを4.2節で述べ、食材の見えの変化モデルを導入することを4.3節で述べる。さらに、山肩らの制約(2)を緩和することで、食材領域の分割の問題が発生する。これに対処するため、時間コンテキストを用いた領域抽出をすることを、4.4節で述べる。最後に、食材追跡アルゴリズムの全容について4.5節で述べる。

### 4.1 本研究での調理に対する制約

山肩らの制約に対して、以下のように変更を加える。

- 1'. 同色異種の食材に対する調理操作が同時に施されることはない。
- 2'. 異種の食材が同時にひとまとめにして置かれることはない。

以降、これらの制約を提案制約(1)、などの呼称で参照する。

山肩らの制約(1)ではすべての食材に対して調理操作が同時に起こらないとされていた。これに対して、提案制約(1')では、同色異種の食材に対しては調理操作が同時に起こらない、のように制約を緩和している。つまり異色異種の食材については、同時に調理操作が行われてもよいとしている。このとき、山肩らの制約下では起きなかった、異色異種食材が同時に分裂・統合することが起きうる。同時に分裂・統合が起きることが原因で、3.2節で述べたように、最近傍の消滅・湧出ノードを対応付ければよいとしていたルールが適用できなくなる。湧出ノードに対して、最近傍を含めた複数の消滅ノードの中から、妥当な対応付けノードを探す必要が生じる。

そこで、この妥当な対応付けを求めるため、調理者による食材の遮蔽の検出と、食材の見えの変化のモデルを用いる。

食材の遮蔽の検出では、調理者の手によって食材が遮蔽されることで生じる、消滅・湧出ノードを対応付けることができる。このとき、食材の見えの変化は、キャストシャドウによるもののみであることを用いる。

食材の見えの変化のモデルでは、妥当性未知の食材の見えの変化が与えられたときに、その変化とモデルとの類似度(尤度)を求めることができる。この



尤度を対応付けの妥当性とする。フレーム間の対応付けの際、モデルと照らし合わせることで、見えの変化が妥当でない対応付けがあれば、それを棄却する。見えの変化が妥当な対応付けのみを残すことで追跡を実現する。

図8の例で山肩らの対応付けルールを適用した場合、湧出ノード W とそれに最近傍の消滅ノード A (および B) を誤って対応付けてしまうことが問題である。ここで、事前に異色異種食材の見えの変化モデルが獲得されているとする。W と A は異種類の食材であるので、W から A へ見えの変化は、モデル化されていない。そのため、食材 W から食材 A へ見えの変化を仮定すると、その見えの変化はモデルとの類似度が低い。結果、食材 W, A の対応付けを棄却することができる。食材 W, B 間の対応付けも同様に棄却する。次に、A よりも前の時刻で W に最近傍の消滅ノードである食材 U との対応付けをおこなう。食材 W と U は同種類の食材であるので、W から U へ見えの変化は、モデル化されている。そのため、食材 W から食材 U へ見えの変化を仮定すると、その見えの変化はモデルとの類似度が高い。結果、W と U の対応付けは妥当であると判断する。湧出ノード W に対応付けられる可能性のある消滅ノードは、時刻  $t_0$  以前にも多数ある。しかし、湧出ノード W は消滅ノード U と同一の食材であると判断していて、かつ、提案制約 (2) により、湧出ノード W は 1 種類の食材のみによる食材領域である。時刻  $t_0$  以前に対応付け先があったとしても、湧出ノード W を構成する食材種類は変わらない。したがって、湧出ノード W に対応付けられる可能性のある消滅ノードをこれ以上探す必要はない。

山肩らの制約 (2) では、異種の食材がひとまとめにされることはない、としていた。これに対して、提案制約 (2') では、異種の食材が同時にひとまとめにされることはない、と制約を緩和している。つまり、同時でなければ異種の食材をひとまとめにして調理台上に置いてよいとしている。このとき、調理台上に置かれている食材に隣接して、異種の食材がひとまとめにして置かれるので、これらの領域を分離する必要がある。そこで、異なる時刻で調理台上に置かれるこれらの食材を分離するため、観測画像中の食材領域を抽出する際、時間コンテキストを用いる。

## 4.2 調理者の手による食材の遮蔽の検出

本節では、調理者の手によって食材が遮蔽されていることを検出する方法について述べる。

遮蔽の前後では，画像中で調理台上にあった食材が，一旦調理者の手に隠され，その後再び画像中で観測される．このため，調理操作が施された食材の対応付けと同様に，遮蔽直前の消滅ノードと直後の湧出ノード間の対応付けを行う必要がある．このとき，食材に調理操作が施されることはないので，食材が内部露出，形状変化，姿勢変化などによって見えが変化することはない．見えの変化は，調理者の手のキャストシャドウによるもののみである．そこで，食材領域内の正規化距離 [5] の和を閾値処理することで検出する．具体的には，消滅ノードの領域内の位置  $(x, y)$  の画素値  $(r, g, b)$  を  $\mathbf{i}(x, y) = (r, g, b)$  で表し，湧出ノードでも同様に画素値を  $\mathbf{j}(x, y)$  とし，この食材が画像中で占める領域を  $R$ ，閾値を  $T$  として，

$$\sum_{(x,y) \in R} \left| \frac{\mathbf{i}(x, y)}{|\mathbf{i}(x, y)|} - \frac{\mathbf{j}(x, y)}{|\mathbf{j}(x, y)|} \right| < T \quad (1)$$

ならば，両領域間で見えの変化が小さいと判断し，調理者が食材を遮蔽したものとみなし，消滅・湧出ノードを対応付ける．ただし  $|\cdot|$  は画素値の大きさを表す．

### 4.3 画像特徴変化モデルの導入による食材領域の対応付け

山肩らの制約 (1) を提案制約 (1') に緩和することで発生する対応付けのあいまい性を，食材の見えの変化モデルの導入によって対処することを述べる．

#### 4.3.1 食材領域を表現する画像特徴量の選定

まず，観測画像中の食材領域の見えを記述する画像特徴の選定について述べる．特徴量としては，色，形状 [6]，テクスチャ [7] などが考えられる．調理では，彩りの豊かさを考慮して食材が選ばれるため，色特徴を用いると同一メニューに用いられる食材については特徴の分散が大きくなることが期待できる．一方，形状特徴は異なる種類の食材で似通ったものが多く存在する．また，テクスチャ特徴は未調理状態で表面が一様な食材が多く存在する．従って，形状やテクスチャの特徴は，色特徴には劣ると考えられる．

山肩らの研究では，食材領域の各画素値のヒストグラムの最頻値を用いている．しかし，これは最頻値以外の特徴を破棄してしまう，局所的な情報のみからなる特徴であるため，葉菜類，根菜類のように主要な色が類似した食材の区別が困難になる．そこで本稿では，食材領域の各画素値のヒストグラムの各ピンの値を並べたベクトルを特徴量として扱う [8]．こうすることで，局所的な情報

である主要な色を食材領域の全体的な色特徴をおさえつつ、主要な色が類似した食材でも主要色以外の大域的な特徴の差異を記述することができるため、食材を表現する特徴として適している。

調理の際、調理者の影が食材に映ることで食材の見ための明るさが変化してしまうことが多い。そのため、明度の変化による影響を避けるために、画素値を表す色空間の計算に HLS 表色系を用いて明度成分を分離する。ただし、HLS 表色系では明度成分が上界・下界に近づくときには彩度・色相の各値が不安定になる性質があるので、これらの画素用にヒストグラムにピンを別途用意する。明度成分に上限・下限を設定し、その範囲外にある画素については、別途用意したピンにて度数を計算する。

#### 4.3.2 食材の見えの変化と画像特徴の分布

食材の画像特徴として色特徴を用いた場合の、特徴の分布について述べる。例として、未調理のニンジン、切られたニンジン、未調理のナス、切られたナスの4つについて考える。4.1節で述べたように、食材の見えは、食材種類と調理操作種類に依存した分布をする。未調理のニンジンおよび切られたニンジンでは、色が類似しているため、それぞれの特徴の分布は近接していると考えられる。一方、未調理のナスおよび切られたナスでは、内部色の露出によって色が類似しないため、それぞれの特徴の分布は離れていると考えられる。これを図9に示す。このように、調理操作によって特徴の分布が変わってしまうようなケースで、食材の対応付けを考える必要がある。以降、調理操作によって食材の画像特徴が属する分布が変わることを、画像特徴が遷移する、と記す。

#### 4.3.3 尤度関数の設計

調理操作によって食材の画像特徴が遷移することに対処するため、この遷移を考慮した尤度関数を設計する。観測画像中の消滅・湧出ノードから抽出した画像特徴を特徴空間でプロットすると、消滅ノードを始点、湧出ノードを終点とするベクトルとみなすことができる。図9の黒矢印  $V_1$  は未調理のナスを把持して移動させた際に、画像から観測される特徴をプロットしたベクトルである。黒矢印  $V_2$  は未調理のナスに対して皮をむく調理操作を施した際に、画像から観測される特徴をプロットしたベクトルである。逆にベクトル  $V_3$  が観測された場合、これは画像中のニンジンあるいはナスに該当する領域から抽出した特徴をプロットしたベクトルではないであろうと判断できる。

消滅・湧出ノード間の対応付けは、湧出ノードに対応する可能性がある複数

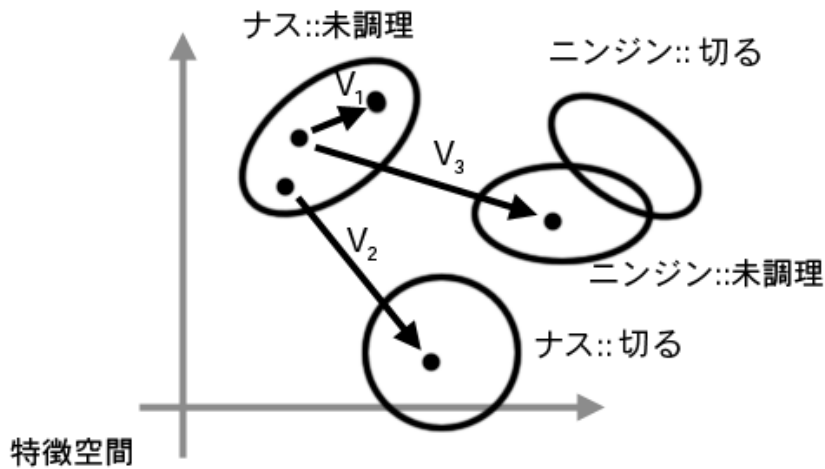


図 9: 食材領域の画像特徴の分布．各楕円が食材種類と調理操作種類ごとの画像特徴クラスタを示す

の消滅ノードのうち，尤度が低いものを棄却することによって行う．これは，画像特徴空間での複数のベクトルのうち，始点・終点の遷移が妥当でないベクトルを棄却することに帰着する．ベクトルの妥当性を評価するため，事前に多数獲得しておいた妥当なベクトルの集合をモデルとして用いる．妥当性が未知のベクトルと，モデルとの類似度を尤度として，採択・棄却を判定する．

妥当性未知のベクトルとモデルの尤度は，モデルに含まれるベクトルのうち最も類似度が高いベクトルとの類似度で定める．事前に獲得しておいた  $M$  個の妥当なベクトルの集合を  $\{V_k^{pre}\}_{k=1}^M$  として，妥当性が未知のベクトル  $V_i$  の尤度を

$$Likelihood(V_i | \{V_k^{pre}\}_{k=1}^M) = \max_k \{1 - Similarity(V_i, V_k^{pre})\} \quad (2)$$

とする．ベクトルの大きさを 1 に正規化しておくこと，式 2 の値域は  $[0,1]$  となる．ただし，ベクトル  $V_i, V_j$  間の類似度  $Similarity(V_i, V_j)$  は，次式のように始点・終点それぞれの距離の積の逆数で定義する．ベクトルの始点・終点の特徴量をそれぞれ  $S(V_i), S(V_j), E(V_i), E(V_j)$  として，

$$Similarity(V_i, V_j) = \frac{1}{|(S(V_i) - S(V_j))| * |(E(V_i) - E(V_j))|} \quad (3)$$

ここで  $|\cdot|$  は特徴量の絶対値を表す．

そして最終的な食材追跡の定式化を行う。

ある時点での観測画像中で注目している湧出ノードの特徴量を  $E_{target}$  とする。この湧出ノードに対応すると思われる消滅ノードの候補が次時点の観測画像中で  $N$  個検出され、これらの特徴量がそれぞれ  $S_1^{candidate}, S_2^{candidate}, \dots, S_N^{candidate}$  であるとする。注目している湧出ノードと対応する消滅ノード候補間の  $N$  個のベクトルについて尤度を求める。事前に獲得しておいた  $M$  個のベクトルの集合を  $\{V_k^{pre}\}_{k=1}^M$  として、

$$\arg_i \max_i \{Likelihood(V_i | \{V_k^{pre}\}_{k=1}^M)\} \quad (4)$$

によって与えられる  $i$  番目のベクトル  $V_i$  (始点が消滅ノード  $S^i$  , 終点が湧出ノード  $E_{target}$ ) を正解として選択する。

#### 4.4 時間コンテキストを利用した物体領域抽出

本節では、山肩らの制約 (2) を提案制約 (2') に緩和することで、食材領域の分割の問題が発生するため、時間コンテキストを用いた領域抽出を行う。



図 10: 時間コンテキストを用いた食材領域分割

異種の食材 A,B がそれぞれ時刻  $t_1, t_2$  で調理台上に置かれるとする。模式図を図 10 に示す。時刻  $t_2$  で食材 B は食材 A に隣接して置かれる、このとき山肩らの制約 (2) が満たされない。

そこで、時間コンテキストを用いる。食材が調理台上に置かれた時刻ではその食材領域を前景領域とみなし、それ以降の時刻では背景領域とみなす。こうすることで、異なる時間に調理台上に置かれた食材は、食材領域が隣接していたとしても、これらの食材領域を分離して扱うことができる。

## 4.5 追跡アルゴリズム全体の流れ

これまでの議論をまとめて、本稿で提案する調理映像中の食材領域を追跡するためのアルゴリズムを示す。

1. 各時点 ( $0, \dots, t-1, t, t+1, \dots$ ) で a から h を適用
  - a. 観測画像を取得
  - b. 画像から前景領域 (食材, 調理者, 調理器具) を抽出
  - c. センサ情報によって調理器具を抽出
  - d. 調理者を抽出
  - e. 画像中の消滅・湧出ノードを抽出
  - f. 以前の時刻の消滅ノードと現時刻の湧出ノードのうち, 遮蔽によるものを対応付け
  - g. 以前の時刻の消滅ノードと現時刻の湧出ノードに対して, 事前に獲得した食材の見えの変化モデルを用いて, 妥当でない対応付けを棄却
  - h. 現時刻の消滅ノードは次時刻以降で対応付ける
2. 終了

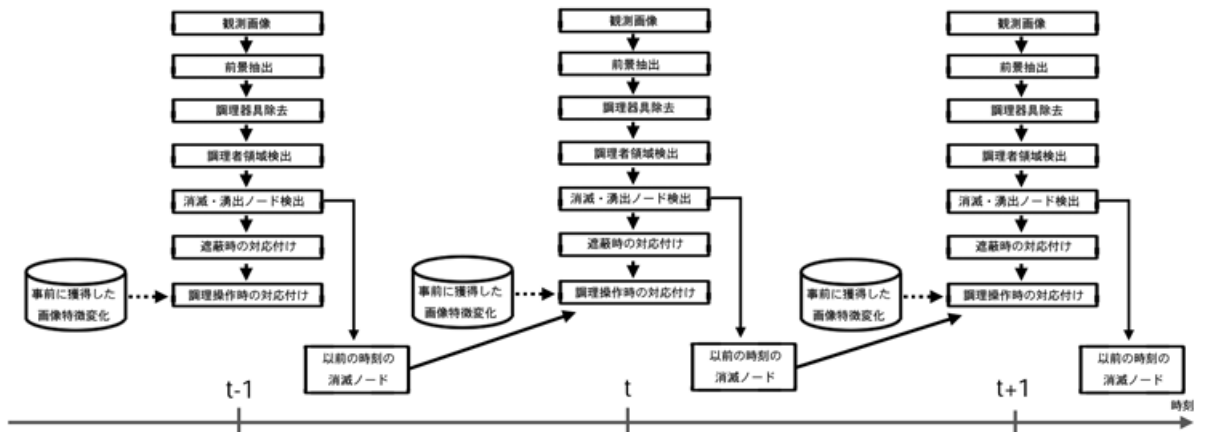


図 11: 追跡アルゴリズム全体の流れ

## 第5章 実験・評価

### 5.1 実験環境

まず実験環境について説明する．調理は図 12 に示すような環境で行った．カメラの設置は，2.1 節で述べたように，調理台ほぼ真上の天袋の内部に取り付けている．

このカメラ設定で実際に観測した画像を図 13 に示す．背景差分の簡単化のため，調理台天板に液晶ディスプレイを埋め込み，これに常時 RGB(0,0,255) の一様な画面を表示させた．また，食材追跡に直接関与しない容器，バット，まな板は青色のものを使用した．このような青い色の外見・内面を持つ食材は存在しないので，クロマキー処理によって背景領域と調理器具の領域を除き，食材領域を抽出することが容易に行えた．



図 12: 調理台とカメラの設置位置

### 5.2 実験 1:画像特徴の変化モデルに対する尤度関数の評価

4.3.3 小節で述べた尤度関数の性能を調べるため実験を行った．追跡では，ある時刻で 1 箇所でも対応付け間違いが生じると，それ以降のすべての時刻での対応付けにその影響が波及してしまう．したがって，尤度関数は，正解の取りこぼしがないよう，再現率が高くなるよう設計することが望ましい．

実験には，ナスとカボチャを切るシーンを用いた．このシーンでは，ナスを



図 13: 調理台の様子

切った後，調理者の手にナスの断片が隠蔽されたまま，カボチャを切り始める．複数の調理操作が同時に行われるため，山肩らの制約 (1) が満たされない．他方，ナスは外部が黒色，内部が白色，カボチャは外部が緑色，内部が黄色で，これらは異種異色の食材の組み合わせである．したがって，提案制約 (1') は満たしている．このシーンで，正しい対応付けによる見えの変化を正解，誤った対応付けによる見えの変化を不正解として，画像特徴の変化を収集した．正解・不正解とも 8 個ずつ，収集した．

食材の見えの変化のモデルを準備するため，事前に同様の調理を複数回行った．画像中でナスおよびカボチャの消滅・湧出ノードに該当する領域を手作業で切り出し，その画像特徴を抽出した．食材に調理操作が施される前後での画像特徴の遷移を多数獲得した．モデル構築のための画像から抽出する特徴ベクトルのサンプル数は 200 とした．

4.3.1 で述べた食材の画像特徴として HLS 表色系のうち，HS のみの値によるヒストグラムを用いた．HS の各値を  $16 \times 16 = 256$  次元のビンに切り分けた．併せて，L の値が上限下限を超える画素については別途ビンを用意し，合計 258 次元のベクトルを用いた．L の値域は  $[0, 255]$  であるが，上限・下限をそれぞれ 210, 50 とした．

4.3.3 小節で述べた尤度関数の式 (3) 中の特徴量の距離関数  $|\cdot|$  としては，2 乗



距離，交差距離 [8]，カイ 2 乗距離 [7]，Bhattacharyya 距離 [9] などが有効な手法として提案されている．これら 4 つの距離関数で予備実験を行い，2 乗距離

$$|\mathbf{x}| = \sum_{i=1}^N x_i^2 \quad (5)$$

を用いることとした．ただし上の式で， $N$  次元ベクトルを  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  で表した．

正解・不正解の各対応付けに対して，食材の見えの変化のモデルを用いて尤度を求めた．このときの，正解・不正解ごとの尤度の分布を集計したものを図 14 に示す．不正解の対応付けに対して低い尤度，正解の対応付けに対して高い尤度を付与できていることがわかる．

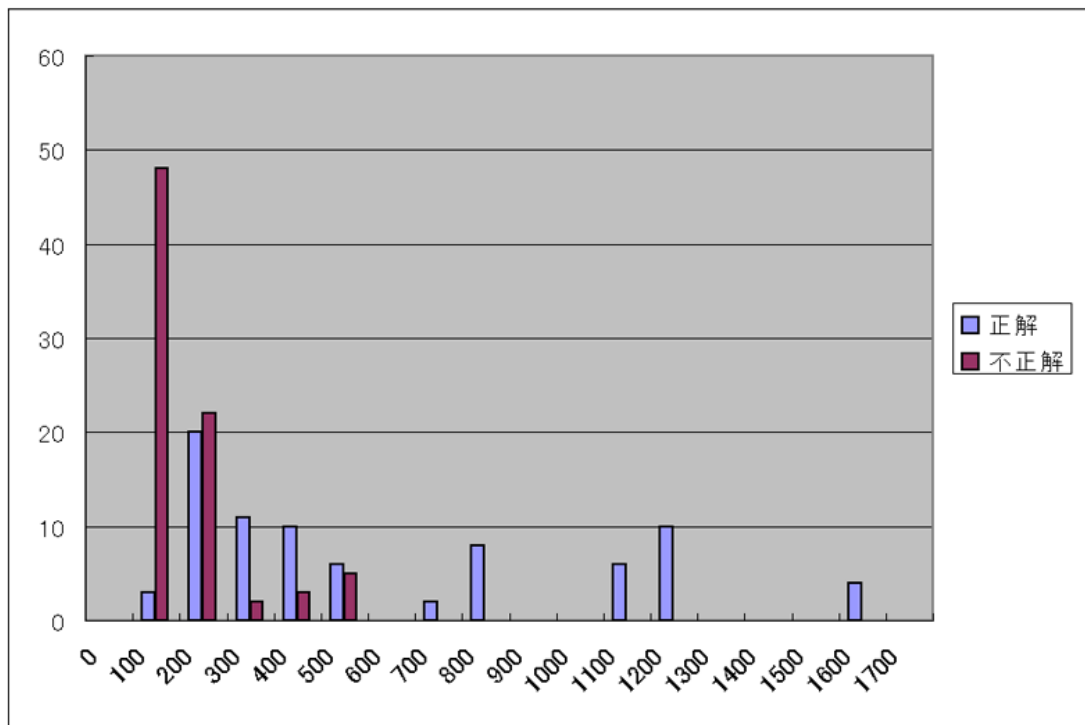


図 14: 正解・不正解ごとの尤度の度数分布．縦軸が度数．横軸が尤度

画像特徴変化に対して計算した尤度に対して閾値処理を施すことで，その画像特徴変化が妥当なものであるか否かを判定する．既に述べた通り，再現率ができるだけ高くなるようこの閾値を定める．図 15 に，図 14 の尤度 [96,272] の範囲の詳細な分布を示す．閾値を 188 とすることで，再現率:80/80=1.00，適合率:80/112=0.71 の結果を得た．

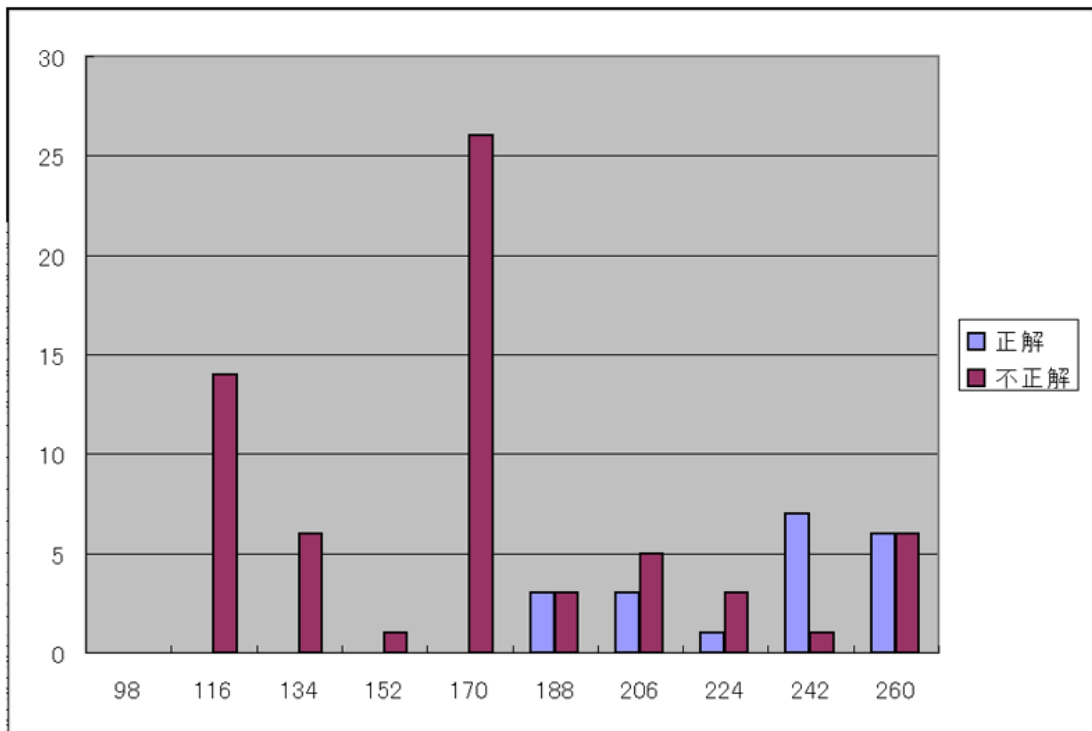


図 15: 正解・不正解ごとの尤度の度数分布．尤度 [96,272] の範囲の詳細な分布

### 5.3 実験 2: 提案制約の適用可能性の検討

前節で述べた尤度関数を用いて食材追跡を行うには，その調理において提案制約 (1') が満たされている必要がある．そこで制約の適用可能性について評価するため，提案制約 (1') と山肩らの制約 (1) が満たされないシーンが実際の調理でどの程度存在するかを検証した．ここで，4.1 節で述べたように，提案制約 (1') が満たされないのは「同色異種食材に対して同時に調理操作が行われるシーン」であり，山肩らの制約 (1) が満たされないのは「異種食材に対して同時に調理操作が行われるシーン」である．

以下に実験の詳細を記す．調理メニューとしてカレーと筑前煮を用いた．レシピは詳細な記述は付録 A.1 に記載してある．

このレシピを用いて，2 人の調理者 A, B がそれぞれ，A: カレー 3 回，筑前煮 1 回，B: カレー 1 回，筑前煮 1 回，の合計 6 回の調理をする様子を観測した．

調理で用いられる食材が同色か異色かの判定は見た目による評価とした．表 2 にその結果を記す．各食材の同色・異色の判定は，その食材に対してどのような調理操作が施されても，類似した色が現れることはないと考えられる食材の

組み合わせを異色とした．未調理状態で類似した色である食材，および，ある調理操作によって類似した色が現れうると考えられる食材の組み合わせを同色とした．

実験の結果を表3に記す．

カレーの場合，本研究が対象とする調理操作はレシピ中に7個記載されているので，4回の調理で合計28回の調理操作が対象となった．このうち，5回で異種食材に対して同時に調理操作が施され，1回で同色同種食材に対して同時に調理操作が施された．

筑前煮の場合，本研究が対象とする調理操作はレシピ中に5個に記載されているので，2回の調理で合計10回の調理操作が対象となった．このうち，3回で異色異種食材に対して同時に調理操作が施され，1回で同色同種食材に対して同時に調理操作が施された．

以上から，提案制約(1')が満たされるのは，38回の調理操作で，同色同種食材に対して同時に調理操作が施された2回を除いて，36回であった．一方，山肩らの制約(1)が満たされるのは，38回の調理操作で，異種食材に対して同時に調理操作がほどこされた8回を除いて，30回であった．したがって，山肩らの制約(1)と比較し提案制約(1')を導入することで，実際の調理において $6/38 = 16\%$ の追跡結果の向上があることが示される．

#### 5.4 実験3:実際の調理映像中での食材追跡

次に，従来手法が追跡を失敗するシーンで，提案手法が追跡に成功することを確認するため，実際に調理を行った様子を撮影した映像映像に対し提案手法を適用した．従来手法が追跡を失敗するシーンとして，シチューを作っている映像中の1シーンを手動で切り出した．このシーンでは，なすを切り，かぼちゃを切る．2つの調理操作の切り替わりのとき，2つの食材を同時に手で持って移動させるため，複数の調理操作が同時に起こらないことを制約とする従来手法は適用できない．映像の最初のフレーム内で抽出した食材領域に対して，

- なす：A
- かぼちゃ：B

というラベルを割り当てた．このラベルがシーンの終わりまで映像中で該当する食材領域に割り当てられていれば，追跡がうまくいっているといえる．

表 1: カレー使用食材の同色異色判定．各行・列の食材の組み合わせの同色・異色を判定している．

	ほうれん草	豚肉	なす	トマト	たまねぎ	にんじん	じゃがいも
じゃがいも	異色	異色	異色	異色	<u>同色</u>	異色	-
にんじん	異色	<u>同色</u>	異色	<u>同色</u>	異色	-	
たまねぎ	異色	異色	<u>同色</u>	異色	-		
トマト	異色	<u>同色</u>	異色	-			
なす	異色	異色	-				
豚肉	異色	-					
ほうれん草	-						

表 2: 筑前煮使用食材の同色異色判定

	ごぼう	こんにゃく	さといも	にんじん	鶏肉
鶏肉	異色	異色	異色	<u>同色</u>	-
にんじん	異色	異色	異色	-	
さといも	<u>同色</u>	<u>同色</u>	-		
こんにゃく	<u>同色</u>	-			
ごぼう	-				

表 3: 制約が満たされなくなる調理操作の発生回数

メニュー	カレー	筑前煮
レシピに記載されている調理操作	28	10
異種食材に対して同時に調理操作	5	3
同色同種食材に対して同時に調理操作	1	1

観測画像に追跡結果を重畳したものを図 16 に示す．各小画像の下部に示した数字（#825 など）がフレーム番号を示す．このシーンでは，#1 でなすを切り始め，#825 で切ったなすの断片を#1125 でバットに戻すとき，同時にかぼちゃの移動を開始している．#1275 でなすの移動が終了し，画像中に現れる．#4750 でこのかぼちゃが調理者の手による隠蔽から解放され画像中に現れる．正解の対応付けは，

- なす (A)： #825 #1275
- かぼちゃ(B)： #825 #4750

である．

図 16 に示したように，複数の調理操作が同時に行われる前後で，なすおよびかぼちゃの領域を正しく対応付けることができている．

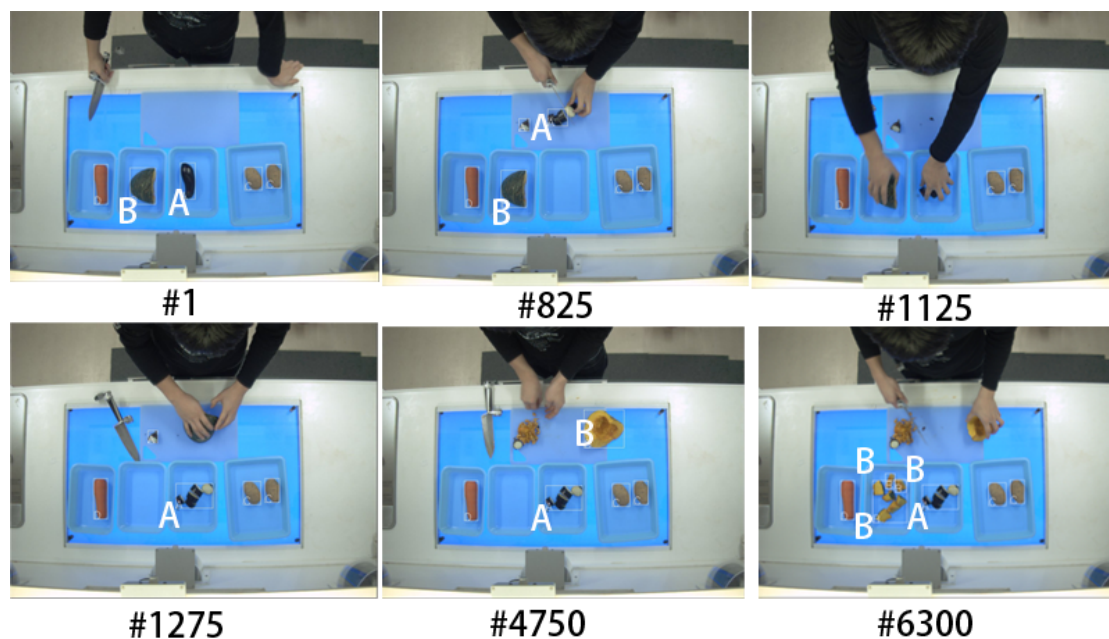


図 16: 実際の調理映像に対する提案手法の適用結果．画像下部の数字がフレーム番号を示す．

## 第6章 結論

本研究では，調理映像アーカイブに対する食材の注釈の付与を行うため，映像中で常時食材を特定するために，画像中の食材領域を追跡する手法を述べた．食材領域を追跡する上で問題となるのは，調理操作に起因する食材の領域数の変化および見えの変化であった．従来手法では，調理操作対象となる食材の種類数に強い制約を課すことで追跡を実現していたが，この制約は実現性に乏しかった．そこで，この食材の種類に対する制約を緩和した．制約の緩和によって追跡のあいまい性が生じるが，調理操作に起因する食材の見えの変化をモデル化し，妥当でない追跡候補を棄却することで，この問題を解決した．

食材の見えの変化のモデルの妥当性を確認するため，調理映像中で集めた画像特徴からモデルを構築し，見えの変化に対する尤度関数を設計した．実際の調理映像中，従来手法では対処できない「異色異種食材が同時に見えの変化をする」シーンを対象として尤度関数を設計したところ，再現率 100%，適合率 71% の結果を得て，有効性が確認できた．

また，提案手法の制約が実際の調理での有効性を検証するため，2人の調理者による6回の調理において，制約が満たされなくなるシーンの発生回数を検証した．提案手法の制約が満たされるのは，調理映像中の38回の調理操作のうち36回であった．従来手法の制約が満たされるのは，調理映像中の38回の調理操作のうち30回であった．したがって，16%の適用可能性の向上を確認できた．

本研究では，色特徴による食材の見えの変化の記述を試みたため，同色異種の食材を識別することができなかった．実験で検討したように，実際の調理では同色異種食材に同時に調理操作することは起こりうるので，この際にも追跡を実現することが望ましい．同色異種の食材の識別を扱うためには，食材の画像特徴量を拡張することが考えられる．このとき，物体認識の研究で用いられている，物体形状やテクスチャ，輝度勾配に関する記述について検討の余地がある．また，調理では食材以外に調理者や調理器具など多くの物体が登場するため，これらの物体間の時間・空間での関係をモデル化した，時空間コンテキストを導入することも検討の余地がある．

調理操作に起因する食材の見えの変化モデルによる追跡候補の尤度計算が，食材種類の識別に寄与する可能性について述べる．見えの変化モデルの構築の際に，各食材種類ごとにモデルを構築することで，食材の画像特徴の遷移の尤

度を計算する際に，食材種類および調理操作ごとに尤度を計算することができる．この尤度を用いることで食材種類を特定することが考えられる．食材種類を唯一に特定できないとしても，尤度が著しく低い食材種類を棄却することもできると考えられる．ただし，本研究では食材の種類の特定は目的としていないため，ここで軽く触れておくにとどめた．

## 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中村裕一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准教授，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] 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. 90-D, No. 10, pp. 2817–2829 (2007).
- [2] Nishida, Y., Aizawa, H., Hori, T., Hoffman, N., Kanade, T. and Kakikura, M.: 3D ultrasonic tagging system for observing human activity, *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, Vol. 1, pp. 785–791 vol.1 (2003).
- [3] AR-toolkit: <http://www.hitl.washington.edu/>.
- [4] 白井清昭, 大川寛志: アニメーション生成のための料理動作辞書の構築, *情報処理学会研究報告*, Vol. 2004, No. 108, pp. 123–128 (2004).
- [5] 松山隆司, 和田俊和, 波部斉, 棚橋和也: 照明変化に頑健な背景差分, *信学論*, pp. 2201–2211 (2001).
- [6] CT Zahn, R.: Fourier descriptors for plane closed curves, *IEEE Transactions on Computers*, Vol. c-21, No. 3, pp. 269–281 (1972).
- [7] Schiele, B. and Crowley, J.: Probabilistic Object Recognition Using Multidimensional Receptive Field Histograms, *Pattern Recognition, International Conference on*, Vol. 2, p. 50 (1996).
- [8] Swain, M. J. and Ballard, D. H.: Color indexing, *Int. J. Comput. Vision*, Vol. 7, No. 1, pp. 11–32 (1991).
- [9] N. A. Thacker, F. J. A. and Rockett, P. I.: The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data., *Kybernetika*, Vol. 32, No. 4, pp. 363–368 (1997).

## 付録

### A.1 実験で使用したレシピ

#### A.1.1 カレー

- 使用食材
  - － ジャがいも（5-6個）
  - － にんじん（2本）
  - － たまねぎ（2個）
  - － トマト（2個）
  - － ナス（2個）
  - － 豚肉 肩ロースの塊（500-600g）
  - － ほうれん草（1把）
- 調理手順
  1. ほうれん草を茹でる
  2. ほうれん草を切る．
  3. なすを切る
  4. ジャがいもの皮を剥く
  5. にんじんの皮を剥く
  6. たまねぎの皮を剥く
  7. 豚肉切る
  8. トマトを切る
  9. 豚肉，ジャがいも，にんじん，たまねぎ，なすを炒める
  10. 水，トマトを入れ，煮込む
  11. ルウを割り入れ，煮込む
  12. ホウレン草を加え，軽く煮込む

#### A.1.2 筑前煮

- 使用食材
  - － 鶏もも肉 1枚
  - － にんじん 大1本
  - － 小芋 中5~6個
  - － こんにゃく 1丁
  - － ごぼう 1本

- 調理手順

1. 鶏肉に切る
2. にんじんの皮を剥く
3. にんじんを切る
4. ごぼうの皮を剥く
5. ごぼうを切る
6. さといもの皮を剥く
7. さといものを切る
8. こんにゃくを切る
9. 鶏肉を炒める
10. にんじん, ごぼう, さといものを炒める
11. こんにゃく炒める
12. 煮る
13. 砂糖を入れる
14. 醤油を入れる