

把持の順序と外見の変化モデルを利用した調理作業における食材追跡

橋本 敦史[†] 森 直幸[†] 船富 卓哉^{††} 山肩 洋子^{††}
棕木 雅之^{††} 角所 考^{†††} 美濃 導彦^{††}

Tracking Ingredients on Food Preparation Process Using Grasped Order and Appearance Change Model

Atsushi HASHIMOTO[†], Naoyuki MORI[†], Takuya FUNATOMI^{††}, Yoko YAMAKATA^{††},
Masayuki MUKUNOKI^{††}, Koh KAKUSHO^{†††}, and Michihiko MINOH^{††}

あらまし 本研究では調理を観測した映像において把持中に切削を受ける食材を追跡することを目的とする。食材のように掌に収まる大きさの物体は、把持中に完全に遮へいされ、カメラから観測できない。また、把持の間に切削などの加工を受けて外見も変化する。このような場合でも正しく食材を追跡するため、本研究では把持の前後で消失・出現する物体を検出し、それらに対応づけるアプローチをとる。この際、切削された物体を正しく対応づけるために画像特徴の類似性に加えて、物体が把持された順序や、あらかじめモデル化した外見の変化と観測された画像特徴の変化との類似度を利用する。実験では切削を含む調理作業の映像で観測された把持において提案手法により 83.6%の精度で正しく食材に対応づけられることを確認した。

キーワード 把持物体追跡, 物体検出, 環境知能化, 調理作業理解, 料理メディア

1. ま え が き

映像中の動物体を追跡する処理はコンピュータビジョンにおいて重要な処理の一つである。物体追跡は異なるフレームに異なる位置・状態で現れる同一物体を特定することで達成される。これまでに様々な条件における物体追跡手法が提案されているが、本研究では特に調理中の切削を伴う加工作業における食材追跡を想定する。

調理は毎日行われる日常的な作業である。しかしながら、作業の内容はレシピによって決まり、そのレシピは毎日変わる。そのため、近年、調理中にレシピに従ってガイダンスを行う様々な支援システムが研究されている [1]~[9]。この中でも特に [6]~[9] はカメラや

その他のセンサを利用して調理の進行状況を自動的に理解しようと試みている。しかしながら、これらのシステムでは食材を追跡するために同時に把持する物体を一つに制限する [6]~[8]、各調理ステップで利用する容器を事前に決めておく [9] 等の制約がある。一般家庭では日常的にこのような制約を守りながら調理を行うことは負担が大きい。本研究ではこのような制約のない食材追跡の実現を目指す。

人や車を対象とした一般的な物体追跡では物体の位置や状態が各フレーム間で連続的に変化する。多くの従来手法がこの位置や状態の連続性を前提としている。しかし、食材は掌に隠れる程度の大きさである場合が多く、把持されて移動する際には完全にカメラから隠されてしまうことがある。また、把持中に切削などの加工を受けて外見が変化する場合がある。そのため、位置や状態の連続性を前提とした多くの従来手法は食材追跡に用いることができない。

本研究では位置や状態の連続性の代わりに、把持された順序を物体の同一性を判定する指標として利用する。また、画像特徴が変化する場合に対処するため、従来から用いられている単純な画像特徴の類似性に加えて、レシピから得られる知識をもとにあらかじめ画

[†] 京都大学大学院情報学研究科, 京都市
Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto-shi, 606-8501 Japan

^{††} 京都大学学術情報メディアセンター, 京都市
Academic Center for Computing and Media Studies, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto-shi, 606-8501 Japan

^{†††} 関西学院大学理工学部, 三田市
School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda-shi, 669-1337 Japan

像特徴の変化をモデル化しておき、そのモデルと観測された変化の類似度を新たに導入する。

本論文では、まず 2. で、調理作業における食材追跡の問題点を述べ、本研究が対処する問題を明らかにする。また 3. において提案手法を詳説する。続いて 4. では実際に調理作業を観測した映像に対して提案手法を適用し、その有効性を評価する。5. において関連研究を、6. でまとめと今後の課題を述べる。

2. 食材追跡問題

本章では、調理作業において食材を追跡する際に問題となる点を整理し、提案手法の適用範囲を明らかにする。また、手法が前提とする観測環境について述べる。

2.1 追跡対象としての食材

食材は調理者に把持されたときにのみ移動する。本論文では把持とは調理者が食材に接触してから離すまでを指す。食材の多くは掌に収まる程度の大きさである。また、そうでない場合も口に入る大きさに切断された結果、掌に収まる大きさに切りそろえられる。そのため、把持中の食材はカメラでは観測できないことが多い。このような場合、食材の追跡問題は把持の開始時に画像から消失した物体と、把持の終了時に画像に再び現れた食材を正しく対応づける問題に帰着される。

一般的に同一物体同士を対応づけるためには画像特徴の比較が行われる。しかし、切削加工された物体は形状が変化する。更に一部の食材は表皮と内部が異なる色をしており、切削により色も変化してしまう。このような場合には単純に画像特徴の類似性を評価しても同一物体同士を対応づけることは困難である。このような表皮と内部の色が異なる食材が調理作業に登場する頻度を調べるため、Web から収集した 1312 個のレシピデータを解析した。その結果、材料表に現れた延べ 5339 個 (466 種) の食材のうち、約 35.3% にあたる 1887 個 (136 種) が表皮と内部で色が異なる食材であった。これらの食材の登場頻度を種類ごとに集計したヒストグラムを図 1^(注1) に示す。このヒストグラムで黒で示された食材は切削などの加工により色が変化する食材を指す。また、レシピに登場した食材の具体例を、その食材の登場頻度、色変化の有無とともに表 1 に示す。集計にあたり各食材の色が切削などの加工により変化するかどうかは、一般的にスーパーマーケットで販売されている状態 (例えば魚ならば切

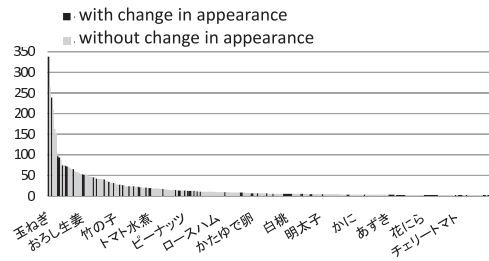


図 1 食材がレシピに登場した頻度、及び、各食材の色変化の有無を示すヒストグラム。一度しか登場しなかった食材は省略した。

Fig.1 Histogram of ingredients appeared in the recipes. The color of bars directs whether the corresponding ingredient changes its color. Ingredients appeared only once are skipped.

表 1 食材と色変化の有無の例

Table 1 Example of ingredients and their change in appearance.

頻度上位 10 個			その他の代表例		
食材名	頻度	色変化	食材名	頻度	色変化
玉ねぎ	338	有	きゅうり	71	有
にんにく	264	無	レモン	65	有
卵	239	有	牛肉	58	無
パセリ	210	無	生しいたけ	51	有
人参	162	無	いか	23	無
トマト	153	無	鮭	18	無
じゃがいも	96	有	キウイ	14	有
生姜	93	有	油揚げ	12	有
ピーマン	88	無	チョコレート	9	無
長ねぎ	75	無	かまぼこ	2	有

り身の状態)を初期状態と想定して手動で判定した。調理に登場する食材のうち、約 1/3 は色が変化的ことから、切削などの加工による外見の変化は日常的に発生していると考えられる。

更に、食材は把持中に切削加工を受けると複数に分裂する。そのため、把持開始時に画像から消失した物体は把持終了後に再び現れる食材と 1 対 1 ではなく、1 対多で対応づけられる。1 対多の対応付けは自由度が高く、問題を難しくする要因となる。

2.2 観測環境

食材を追跡するのに適した観測環境を決めるため、天井から撮影した場合と側面の壁から撮影した場合の 2 種類を比較した。食材を頑健に追跡するためには、観測期間中ずっと食材が遮へいされずに観測画像中にとらえられていることが望ましい。しかし (i) 把持中に手で食材が完全に握り込まれた場合には撮影方向に

(注1)：詳細なデータは <http://hdi.handle.net/2433/139378> からダウンロードできる

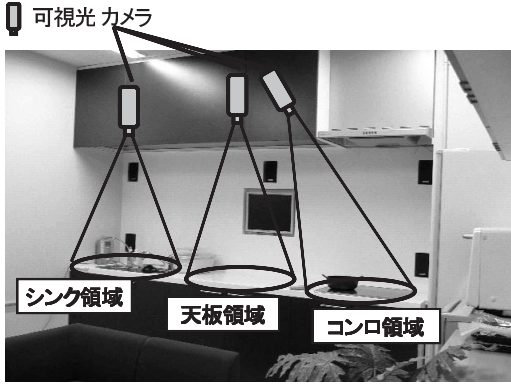


図 2 調理作業の観測環境

Fig. 2 Our observational environment for cooking activity.

ならず食材が観測できなくなる。次に (ii) 食材が把持されていない場合、調理中に伸ばした腕などの体の一部が食材とカメラの間に入った場合には、食材が観測できなくなる。このタイプの遮へいも撮影方向によらずに起こるが、調理者がそこから移動すれば再び遮へいされていた物体が観測できる。最後に (iii) ボウルなどの底が深い容器による遮へいは側面の壁から撮影する場合のみで起こる。この最後のタイプの遮へいを避けるために、本研究では天井から撮影を行う観測環境を構築した (図 2)。

本研究では (i) の遮へいの問題に対して、把持されることで調理者に遮へいされ観測されなくなった物体と、把持後に台の上に置かれ、新たに観測された物体を対応づけるというアプローチをとる。ただし、観測画像中の食材の遮へいには (i) と (ii) の二つのタイプがあり、これらを区別する必要がある。以降、観測画像中から遮へいにより物体が観測できなくなることを物体の「消失」と呼ぶ。物体が消失した場合、(i) か (ii) の遮へいのどちらかが生じたことが分かる。特に消失した物体が遮へい終了後にその場からなくなっていた場合には (i) のタイプの遮へいであり、物体が把持されたことが分かる。また、把持された物体が観測画像中に再び現れることを「出現」と呼ぶ。

3. 消失・出現物体の検出及び対応付け

本章では提案手法を消失・出現物体を検出する物体検出部と、検出されたそれらの物体を対応づける対応付け部の二つに分けて説明する。図 3 に提案手法の大きな流れを示す。物体検出部 (図 3 上) は観測映像

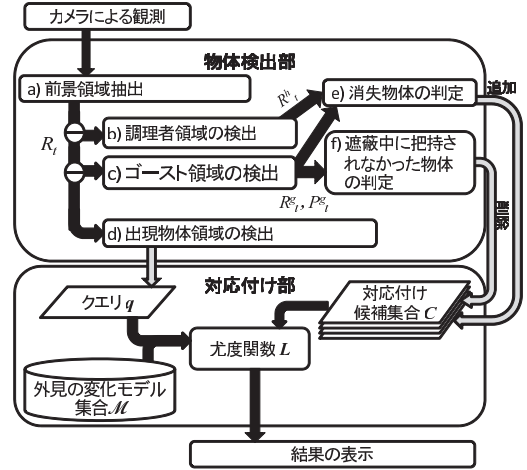


図 3 提案する把持物体追跡システムの概要

Fig. 3 The overview of the proposed system for tracking objects in hand.

から消失・出現した物体を検出し、かつ、2.2 で述べた (i), (ii) の 2 種類の遮へいを区別する。対応付け部 (図 3 下) では消失物体が検出されるたびに、その物体を対応付けの候補として蓄えておく。こうして作成される対応付け候補の集合を C とする。出現物体が検出された場合は候補集合 C の中から同一と考えられる物体を検索することで 1 対多の対応付けを行う。 C との関係から出現物体をクエリ q と呼ぶ。以下では物体検出部と対応付け部の処理を順に説明する。

3.1 消失・出現物体の検出

物体の把持の開始と終了は映像中では以下のように観測される。まず、把持の開始時は物体が調理者の体により遮へいされる。若しくは、フレームレートが低い場合には単純に物体がその場所から消えてなくなることもある。把持の終了時には単純にそれまでなかった物体が調理台上で観測される。ただし、調理者による遮へいが起こった場合であっても、その遮へいが 2.2 における (ii) のタイプの遮へいであった場合には把持は起こっていない。遮へいが (i) のタイプか (ii) のタイプかの区別は遮へいされた物体があった領域から調理者が退いたときに判明する。

以上を踏まえ、我々は消失・出現物体を以下のように設計した。大きな流れは図 3 にも示されている。以降の本文中に現れる (a-f) のインデックスは図 3 と対応している。 I_t を観測映像中の t 番目のフレームとする。最初の処理では背景差分により調理台上の変化を検出する (a)。我々の実装では図 4 に記したパッチ

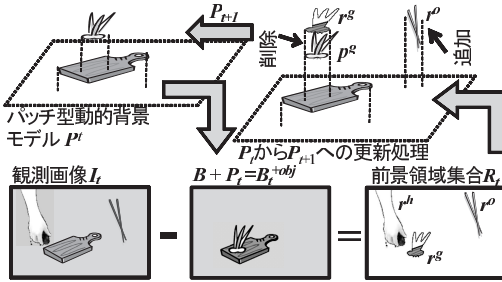


図4 パッチ型の台上物体管理モデルから得られる前景領域の例。背景画像に既に存在する台上物体を重畳した台上画像 B_t^{+obj} から得られる前景領域は人物領域 r^h 、ゴースト領域 r^g 、及び、出現物体領域 r^o の3種類に分類される。

Fig. 4 An example of the patch-based model for objects on a table, and foreground regions obtained from a generated background image B_t^{+obj} . Each foreground region should be a human (r^h), a ghost (r^g), or an appearing object (r^o).

型動的背景モデル P_t に基づいて、背景画像 B_t^{+obj} を毎フレーム生成する。 B_t^{+obj} には時刻 $t-1$ において台上に存在していたすべての物体が含まれている。これは t より前の各フレームで本節で述べる処理結果に基づいて P_t を更新することで実現されている。

このような背景画像 B_t^{+obj} と観測画像 I_t との背景差分の結果得られる前景領域の集合を R_t とする。 R_t は調理者領域集合 R_t^h (b), ゴースト領域集合 R_t^g (c), 及び (3) 出現物体領域 R_t^o (d) (ただし $R_t^h \cup R_t^g \cup R_t^o = R_t$ かつ $R_t^h \cap R_t^g \cap R_t^o = \phi$) の3種類の部分集合に分けられる。これらの集合は以下のような領域が含まれたものと定義する。

$r^h \in R_t^h$: 調理者の体と把持中の物体領域

$r^g \in R_t^g$: 物体がその場からなくなった領域

$r^o \in R_t^o$: 把持終了後に出現した物体領域

ここで r^o は正に検出対象となる出現物体そのものであり、対応付け部にクエリ q として送られる。残りの2種類の領域 r^h, r^g は消失した物体を検出するのに利用される(図3における(e-f))。なお、人物領域 r^h に把持中の物体も含めることで完全に遮へいされた物体以外も遮へいされた場合と同じように扱っている。

人物領域 r^h は t 番目のフレームにおいて新たに遮へいされた物体を検出するために用いられる(e)。その遮へいが把持されたのか(i)、隠されただけで把持されていないのか(ii)が判明するのは調理者が退いた後となる。提案手法ではガイダンスなどのリアルタイ

ム性が要求される応用を想定し、遮へいが起きた場合には、いったん、その物体は把持されたと考えて対応付け部の候補集合 C に加える。後の処理(f)で把持されていないことが判明した場合には、その物体を C から取り除く。また、それまでにその物体と対応づけられた出現物体があった場合には、その対応付けは誤りであると考えられるので、更新された C を用いて、もう一度対応付けを計算し直す。

領域 r^g は物体が移動した後に、物体の後ろにあった領域が観測されることで背景画像との差分を生じ、検出される領域である。このような領域はゴーストと呼ばれ、通常の背景差分では誤検出として扱われる[10]。ゴースト領域を用いれば物体が把持されてその場からなくなったのか、まだそこに置かれたままなのか分かる。提案手法ではこれを用いて遮へいが(i)のタイプか(ii)のタイプかを判別する。

パッチ型背景モデル P_t は台上に置かれている物体をパッチとして保持する(図4)。ここでパッチとは台上の一つの物体を観測した領域のみからなる部分画像であり、各画素の値と観測画像内における自身の位置の絶対座標の両方を保持したものである。また、同じ場所に複数の物体が置かれている場合には、その重なり関係も保持する。出現物体 r^o が観測されるたびに r^o をパッチ p^o として P_t に加える。また、ゴースト領域 r^g が検出された場合には対応するパッチ $p^g \in P_t$ を取り除く。これらの更新処理により P_t を何も置かれていない調理台を観測した画像 B に重畳して生成される背景画像 B_t^{+obj} は常に時刻 $t-1$ における台上に置かれた物体を含む画像となる。

以上の設計に基づき、(a-f)の個々の処理は以下のように実装した。

(a) 前景領域抽出

I_t と B_t^{+obj} との背景差分には調理者による影などの影響を受けにくい正規化RGBによる手法を用いた。また、差分の結果得られた前景画素を、互いに連結しているかどうかに基づいて画素ごとにグループ化し、前景領域とした。

(b) 調理者領域の検出

我々の観測環境では調理者は常に調理台の縁をまたいで観測される。これを利用して、ある領域 $r \in R_t$ が事前に設定した調理台の境界線と交差した場合には r が R_t^h に含まれると判断した(図5)。

(c) ゴースト領域の検出

もし領域 $r (\notin R_t^h)$ と同じ位置に同じ形状のパッチ

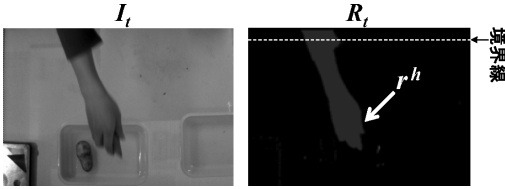


図 5 境界線をまたぐ領域は作業者領域、それ以外は非作業者領域と判定する

Fig. 5 Any regions intersecting the border line are regarded as human's body.

$p \in P_t$ が存在する場合、 r が R_t^g に含まれると判定する。また、このときのパッチ p を台上から取り除かれたパッチの集合 P_t^g とする。

p と r が同じ位置で同じ形状をしているかの尺度として類似度 S_{shape} を次式により定義する。

$$S_{shape}(p, r) = \frac{\#(\partial r \cap \partial p)}{\#(\partial r)} \quad (1)$$

ここで ∂r は領域 r の輪郭領域を表しており、以下のように求めた。まず領域 r と、 r をモフォロジー処理により 1 回膨張させた画像 r^+ との XOR により r の輪郭線を得る。次に、影等の影響による多少のずれを吸収するために得られた輪郭線に対して更にモフォロジー処理により 2 回膨張させたものを輪郭領域 ∂r とした。また $\#(\cdot)$ は領域の面積を計算する関数である。 S_{shape} がしきい値以上となるような前景領域 $r \in R_t$ とパッチ $p \in P_t$ の組合せがあれば、 p を P_t^g に、 r を R_t^g に加える。 P_t^g に対応する物体は既に台上に存在しないので P_{t+1} には含めない。経験的にこのためのしきい値には 0.2 を用いた。

(d) 出現物体領域の検出

(b), (c) の結果から、 $R_t^o = R_t \cap \overline{R_t^h} \cap \overline{R_t^g}$ とした。ここで $\overline{X}$ は集合 X の補集合を表す

(e) 消失物体の判定

時刻 t に調理者により遮へいされている物体を P_t から探すため、重なり具合を表す関数 S_{locate} を次式で定義する。

$$S_{locate}(p, r^h) = \frac{\#(r^h \cap p)}{\#(p)} \quad (2)$$

S_{locate} が経験的に設定したしきい値 0.9 を超えた場合、 $p \in P_t$ を現在遮へいされている物体の集合 P_t^h に加える。この処理の後、 $P_t^h \cap \overline{P_{t-1}^h}$ であるパッチは時刻 t で消失したとみなす。この他、フレームレートが低い場合には r^h と重なることなく物体が消失する。

このような物体は $P_t^g \cap \overline{P_{t-1}^h}$ として検出する。

(f) 遮へい中に把持されなかった物体の判定

時刻 $t-1$ において遮へいされていた物体のうち、遮へいが終了した後に同じ位置で観測され、 C から除去する必要がある物体は $P_{t-1}^h \cap \overline{P_t^h} \cap \overline{P_t^g}$ により得られる。

3.2 消失・出現物体の対応付け

前節では調理者によって把持され、台上から消失した物体と、把持後に台上に出現した物体を検出した。提案手法では以下の二つの仮定のもとで、これらに対応づける。まず、一つ目の仮定は出現した物体は、最近消失した物体と対応づく傾向にあるという仮定である。これは従来研究 [6]~[8] の「調理者は同時に一つしか食材を把持しない」という制約を緩和したものに相当する。通常、調理者は二つ程度の食材は同時に把持できるものの、数多くの食材は同時に把持できないことから、この緩和された制約は直感的に正しいと考えられる。二つ目の仮定は追跡器にとって調理で使用する食材が既知であるという仮定である。通常、レシピのガイダンスを行うシステムでは事前に調理者によりレシピが指定されていることが多い。このような場合には二つ目の仮定も十分に成り立つと考えられる。

これらの二つの仮定に基づき、消失物体と出現物体の対応付けゆう度を以下のように計算する。まず、対応づけ候補 C の各要素をその消失した時刻に応じて降順に並べ換え、最後から i 番目 ($0 \leq i < |C|$) に消失した物体を c_i とする。ここで、前節の処理で検出された出現物体 q と消失物体 c_i の画像特徴をそれぞれ $\mathbf{x}_q, \mathbf{x}_{c_i}$ と表す。なお、対応付け候補 c_i とクエリ q は 1 対多で対応づくため、 c_i は q と対応づいても C から取り除かれず、引き続き後から検出された出現物体の対応付け候補となる。

一つ目の仮定に基づき、候補集合 C の中で先頭に近い要素から優先的に対応づけるようなゆう度関数 L を以下のように定義する。

$$L(c_i, q) = \begin{cases} 0 & \text{if } \max_{j < i} S(\mathbf{x}_{c_j}, \mathbf{x}_q) \geq \Theta_{match} \\ S(\mathbf{x}_{c_i}, \mathbf{x}_q) & \text{otherwise} \end{cases} \quad (3)$$

ここで、 S は q と c_i の画像特徴に基づいた類似度を計算する関数である。また Θ_{match} は C 内の探索終了のためのしきい値である。 $i = 0$ から画像特徴の類似度の評価を行い、 $L(c_i, q) \geq \Theta_{match}$ となれば c_i と

q を同一物体として対応づける．また，その場合には $j > i$ となる要素 c_j の類似度は 0 として探索を終了する．もし $L(c_i, q) \geq \Theta_{match}$ となる要素 c_i が見つからなかった場合には，その物体は冷蔵庫や棚など，観測範囲外から持ち込まれたものだとして判断する．

画像特徴に基づく類似度 S は，使用される食材が既知であるという二つ目の仮定に基づき，次のように設計した．まず，食材の画像特徴 $\mathbf{x}$ としては山肩らの先行研究 [6] で定義された赤，黄，茶，緑，白，黒の 6 色をピンとして利用した色ヒストグラムを用いた．把持中に切削加工を受けた場合，物体の形状は食材の種類などに関係なく切りそろえられてしまう．そのため，切削時には形状は物体の同一性を表す特徴として適切ではない．そのため，提案手法では色ヒストグラムのみを特徴として利用する．ほとんどの場合，食材の色は限られた種類の色素でその色が決まる．同一の色素をもつ食材では色の区別がつかないことが多く，ヒストグラムのピンはこれ以上細かく分割しても対応付けの精度には寄与しない．また，特に小さな食材に対しては得られる画素数が少ないために細かいピンを採用した場合にはヒストグラムの形状を不安定にしまうため，色素ごとの区別を付けられる必要最小限の分割数として 6 色のピンを採用した．

この色ヒストグラム特徴を用いて類似度を計算する関数 S を定義する．切削による特徴の変化に対応するため，我々は画像特徴の変化モデル M を導入した．変化モデル M は二つの外見モデル $\mathbf{m}_{bef}$ と $\mathbf{m}_{aft}$ からなる． $\mathbf{m}_{bef}$ と $\mathbf{m}_{aft}$ はそれぞれ切削による変化前，変化後の外見を表したモデルであり， $\mathbf{x}$ と同様に 6 色の色ヒストグラムで与えられる．もし， $\mathbf{x}_{c_i}$ が $\mathbf{m}_{bef}$ と類似しており，かつ $\mathbf{x}_q$ が $\mathbf{m}_{aft}$ と類似しているならば， c_i は q と同一であると考えることができる．このような考えに基づいて，類似度関数 S を次式で定義した．

$$S(\mathbf{x}_c, \mathbf{x}_q) = \max\{S_{ins}(\mathbf{x}_c, \mathbf{x}_q), S_{mod}(\mathbf{x}_c, \mathbf{x}_q)\} \quad (4)$$

$$S_{ins}(\mathbf{x}_c, \mathbf{x}_q) = \frac{1}{\exp(2D(\mathbf{x}_c, \mathbf{x}_q))} \quad (5)$$

$$S_{mod}(\mathbf{x}_c, \mathbf{x}_q) = \max_{M \in \mathcal{M}} \frac{1}{\exp(D(\mathbf{x}_c, \mathbf{m}_{bef}) + D(\mathbf{x}_q, \mathbf{m}_{aft}))} \quad (6)$$

ここで S_{ins} は把持中に変化がなかったと仮定して，単純に $\mathbf{x}_c$ と $\mathbf{x}_q$ の近さに基づいた類似度を計算する関数であり， S_{mod} と直接比較できる形になっている． $\mathcal{M}$

は二つ目の仮定に基づいてレシピから得られる，使用されるすべての食材の変化モデルの集合である．6 色のヒストグラムは単純に食材一般を区別するには能力が不足するが，区別する対象をレシピに登場する食材のみに絞った場合には十分に有意な特徴となり得る． D は二つの特徴量の距離を計算する関数であり，予備実験の結果から (1-ヒストグラムの交差 [11]) を用いた．また， Θ_{match} は経験的に $D(\mathbf{x}, \mathbf{m}) = 0.3$ の場合に式 (6) がとる値として 0.548 を用いた．

4. 実験

提案手法の有効性を示すため提案手法と 2 種類の比較用追跡アルゴリズム手法 A，及び手法 B を用意し，実際に切削作業を観測した映像に対して食材追跡処理を行った．

手法 A はクエリを受け取るたびに $\mathbf{x}_q$ にかかわらず，その時刻 t_q における c_0 を盲目的に対応づける．手法 B は式 (3) において類似度 S の代わりに変化モデルを利用しない類似度 S_{ins} を用いる．手法 A，手法 B と提案手法を比較することで，把持の順序及び S_{mod} それぞれが追跡にどのように寄与するのかを確認する．

この実験に用いた映像は食材を切削する作業を 3 回行ったものである．調理においては切削以外の作業も行われるが，切削後の食材の移動はボウルやトレイ等の容器と一緒に扱われることが多く，従来手法 [12]～[14] でも頑健に追跡が可能と考えられる．そのため，今回は特に追跡が難しい切削作業が行われる図 2 の Table Top Area のみを評価対象とし，それ以外の場所での作業における追跡は評価しなかった．また R_t にノイズのみからなる領域が含まれても除去できるように，作業領域の背景色，用いる容器及びまな板の色には食材のとり得る 6 色の色分類に現れない青を採用した．

3 回の切削作業のうち，1 回はカボチャとナスの 2 種類，2 回は，それらにジャガイモとニンジンを加えた計 4 種類の食材を切った調理である．ただし切削された後は複数に分割される．また同じ食材が複数把持されることもある．そのため発生した対応付けの回数は個体数とは一致しない．これらの食材に対して与えた変化モデルを表 2 左に示す．ただしニンジンは切削されても色が赤のまま変化しないため変化モデルをもたない．この 3 回の切削作業を用いて手法 A，手法 B と提案手法との比較を行った．

なお，実際のアプリケーションでは同種の食材であ

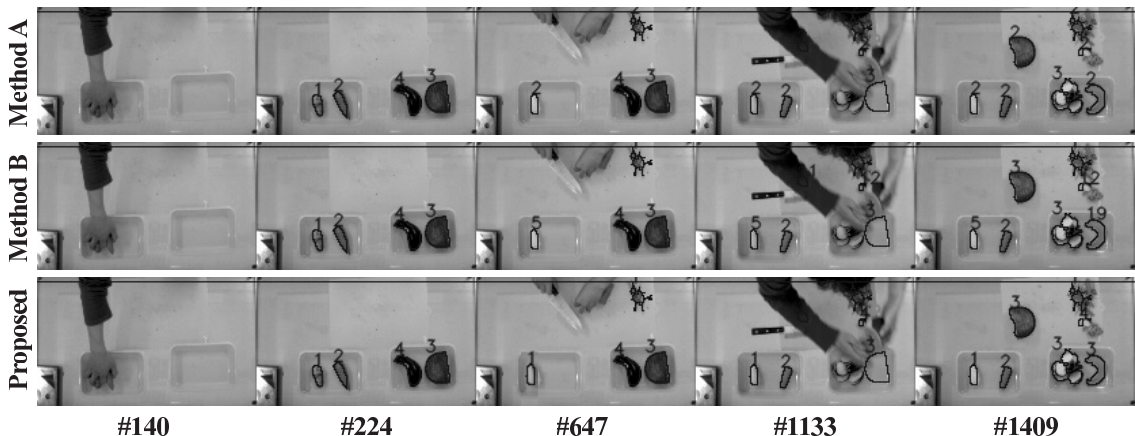


図 7 手法 A、手法 B と提案手法の比較。それぞれの列で同一物体と判定された領域には同じ ID 番号が割り振られている。
Fig. 7 Comparative view of our method with Method A and Method B. In subsequent image frames, object regions are assigned the same numeric ID when they are judged to be the same object.

表 2 実験に用いた変化モデル
Table 2 Change models used in the experiment.

M	m_0	m_1	タマネギ	茶	白
カボチャ	緑	黄	シイタケ	茶	茶, 白
ナス	黒	白, 黒	ショウガ	茶	黄
ジャガイモ	茶	白			

れば個体の区別まで必要となることは考えにくい。そこで、クエリ q が同種の食材からなる候補 c と対応づけられた場合は正解とした。また、用いた映像は 2.5 フレーム/秒、解像度が 256×192 の映像を用いた。本手法では位置の連続性を用いないため、低いフレームレートでも致命的な情報の見落としが起きない。また、画像特徴に 6 個のピンからなる粗いヒストグラムを用いているため、解像度が低くても食材の色の傾向が十分反映されたヒストグラムを得ることができる。

3 回の切削作業に対して手法 A、B 及び提案手法を適用した際の正答率を図 6 に、それぞれの手法の結果を観測映像に重畳して比較した様子を図 7 に示す。なお、上記の三つの観測を通じて対応づけは 61 回行われた。そのうち、手法 A、手法 B はそれぞれ 39 回 (63.9%)、37 回 (60.7%) を正しく対応づけた。これに対し、提案手法は 51 回 (83.6%) を正しく対応づけた。手法 A でも半分以上の場合に正しく対応付けができたものの、複数の物体が同時に把持された直後に誤対応が連続して発生した。また、変化モデルを利用しない手法 B では切削が起きた場合にすべて誤対応となった。一方で提案手法においては上記のような誤りはほ

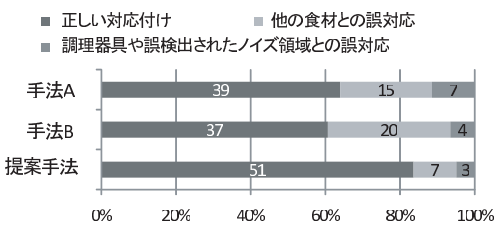


図 6 手法 A、手法 B 及び提案手法による対応付けの結果。他の種類の食材との誤対応と、調理器具や誤検出されたノイズ領域との誤対応は別々に数え上げた。
Fig. 6 Matching results for Method A, Method B and our proposed method. Two types of false matching (with other food materials, and devices and false-detected noise regions) are counted separately.

とんど起きなかった。
ただし、作業者の影になるなどして暗くなった場合などに別の食材と似たようなヒストグラムとなってしまう場合があり、その場合には手法 B とともに対応付けを誤った。また、トレーにまとめて置かれたナスは黒と白の両方が同程度ずつ含まれる領域として観測されるのに対し、切削作業中は内部の白い面のみがカメラに向けて、黒い皮の部分が観測されず、一時的にジャガイモと誤対応を起こすことが確認された。この誤対応はまな板からトレーに移動した際に、切削前のナスと対応づけられることで復旧することができた。
以上の実験では画像特徴に明確に違いが出る場合を想定していた。実際のレシピでは m_{bef} が同一となる

食材が複数登場する場合もあり、そのような食材の組合せに対しては類似度関数 S_{mod} は有効に働かない。このような提案手法が苦手とする状況の極端な場合を想定して、上記の実験とは別に m_{bef} が共通となる 4 種類の食材（ジャガイモ、タマネギ、生姜、シイタケ）を切削する作業を観測した映像に対し、提案手法を適用した。これらの食材の変化モデルを表 2 右に示す。全体で 59 回の対応付けが行われ、提案手法により 44 回 (74.6%) が正しく対応づけられた。この結果は m_{bef} が同一の食材に対しても把持の順序による制約が有効に働いたことを示唆している。提案手法でも対応付けができなかった 15 回のうち、典型的な例は複数の食材が同時に把持された場合であった。 m_{bef} が同じ物体が複数同時に把持された場合は提案手法でも原理的に対処できない。

調理に登場する食材の種類や複数同時に把持する回数などはレシピの好みや調理の習慣などに応じた偏りがあるため、統計的にどの程度の割合で原理的に対処できない状況が生じるのかを見積もるのは難しい。ただし、それぞれを独立に見積もった場合、まず m_{bef} とほぼ同等のものとして、食材の代表色が同一の食材が登場するレシピの割合は 24.8% 程度であることが山肩らにより示されている [6]。また、日本における一般的な家庭料理の一つとして筑前煮を対象に、異なる人物が調理した映像 2 本に対して、同時把持が何回起こるかを数え上げたところ、それぞれの調理において同時把持回数/全把持回数は 8/30、と 19/36 となった。これらの結果から、 m_{bef} が同一の物体が複数、同時に把持されるという状況の発生回数は全体の把持のうちわずかであると考えている。

5. 関連研究

作業対象を追跡することによる行動支援システムとして、組立作業の支援が Zauner ら [13] などにより、AR の分野で研究されている。これらのシステムでは、物体の追跡は ARToolKit のマーカや電子的なタグを事前に追跡対象に取り付ける必要がある。しかし、事前に食材がどのように切断されるのかは分からないため、マーカや電子的なタグなどをあらかじめすべての断片に取り付けておくことは難しい。

画像処理に基づくアプローチとしては SIFT を用いて、机上に置かれた書籍の増減を監視し、新たに置かれたものを Web 上のコンテンツと結び付ける手法が Kim ら [15] により提案されている。この手法では机

上で起こる動きが書類の移動のみであると仮定し、フレーム間差分により変化が検出された画素数がしきい値以下となったときに画像内の物体の増減を比較する。しかし、常に作業者が台上で作業を行い続ける今回の問題では物体が移動するとき以外でも動きが発生するため、この仮定は成り立たない。そのため、我々は各フレームにおいて、作業者の占める領域を検出するとともに、台上に置かれている物体を管理するパッチ型動的背景モデルを利用することにより、物体の消失・出現を毎フレーム確認している。

また、ビデオ会議システムなどの映像を対象に顔を頑健に追跡する手法として三上らは顔の位置や向き、あるいは外見特徴を過去の追跡結果からオンラインでサンプリングし、ノンパラメトリックにモデル化する手法を提案している [16], [17]。この手法では Particle Filter による追跡を行う際に、過去の観測データをもとに顔の位置や向き、外見の状態空間における事前分布を学習することで、遮へいなどにより追跡が不可能なフレームが連続し、かつ遮へい中に顔の向き及び外見が変化した場合でも、頑健に復旧可能な追跡を実現している。ただし、本論文の提案手法のように物体が複数に分割され、別の時刻に別の場所に置かれるような状況は考慮されておらず、また、追跡対象が複数同時に遮へいされた場合に関しても特に議論されていない。これに対し、提案手法では把持物体の追跡という状況に特化した方法として把持順序に基づく対応付けを外見特徴と組み合わせることにより、物体が分割されたり、複数の物体が同時に遮へいされる場合にも対処している点が異なる。

これらとは別に、把持により移動する物体を追跡するために、津吹らは手領域検出を利用した手法を提案した [12]。この手法では、まず、可視光カメラと赤外線カメラを併用することで頑健に手領域を検出する。手領域の近くにあり、かつ、手領域と同様に移動する領域は同一物体であると考えすることで、物体の画像特徴などによらない追跡を実現している。ただし、位置の連続性を利用する手法であるために、把持中の物体領域がカメラで観測されている必要があり、掌に収まるような小さな物体に対しては頑健な動作が望めないことが報告されている。

また、津吹らのように追跡対象と同様に移動する物体を追跡の手掛りとして利用する手法として Grabner らの手法も提案されている [14]。この手法ではまず画像中から多数の特徴点を抽出し、遮へいのない状態で

追跡対象と同様に動くサポーターと呼ばれる特徴点を学習する。遮へいが起きた場合には追跡対象ではなく、サポーターとの相対的な位置関係により追跡対象の位置を頑健に推定することができる。このサポーターは追跡対象に依らずオンラインで学習できるため、津吹らの手法をより一般的なシーンへ拡張したものと考えることができる。ただし、この手法でもサポーターが遮へいのない状態で追跡対象と同様の動きをするシーンを1フレーム以上観測する必要がある。このため、食材追跡のように対象が掌に納まる程度の大きさである場合には、把持の瞬間から対象が完全に隠されてしまつてサポーターが有効に働かない場合も多いと考えられる。

これに対し、提案手法は把持の瞬間から観測不可能な小さな物体に対しても利用可能な手法である。また、把持中も一部が観測できるような大きな物体に対しては、津吹らや Grabner らの手法と適切に組み合わせることでより頑健な追跡が可能になることが期待される。

6. む す び

本研究では調理作業において食材を追跡するための手法を提案した。食材は掌に収まるため把持されている間はカメラで観測できない。更に、把持の間に切削作業を受けて外見が変化するとともに、複数に分割される。これらの問題に対処するため、提案手法では把持の前後で消失・出現した物体を検出し、それらを消失した順序と外見の変化モデルに基づく類似度という二つの基準を新たに導入した。

実験では、調理の中でも特に追跡が難しいと考えられる切削作業を観測したシーンを対象に提案手法を適用し、その有効性を確かめた。3回の切削作業を通じて、4種類の食材の切削を行った結果、発生した61回の対応付けのうち、83.6%にあたる51回の対応付けに成功した。また、提案手法が苦手とする変化前の外見が似た食材4種類に対しても手法を適用し、74.6%の精度を得た。

なお、変化前の外見が似た食材が複数、同時に把持された場合には原理的に提案手法でも正しく追跡を行うことができない。このような状況が発生する頻度は限られているものの、より頑健な追跡の実現には、把持中に行われた作業の内容や別の食材との対応付け結果など、同一性を示す新たな指標を導入することが必要である。また、初めて観測画像に登場した物体を検出するためには冷蔵庫の開閉ドアなどの外部機器から

得られる情報を導入するなどの工夫が必要である。これらは今後の課題である。

文 献

- [1] W. Ju, R. Hurwitz, T. Judd, and B. Lee, "CounterActive: an interactive cookbook for the kitchen counter," Conf. on Human Factors in Computing Systems, pp.269–270, 2001.
- [2] J.S. Bradbury, J.S. Shell, and C.B. Knowles, "Hands on cooking: towards an attentive kitchen," Conf. on Human Factors in Computing Systems, pp.996–997, 2003.
- [3] Q. Tran, G. Calcaterra, and E. Mynatt, "Cook's collage: Déjà vu display for a home kitchen," Proc. Home-Oriented Informatics and Telematics, pp.15–32, 2005.
- [4] C.H.J. Lee, L. Bonanni, J.H. Espinosa, H. Lieberman, and T. Selker, "Augmenting kitchen appliances with a shared context using knowledge about daily events," Proc. 11th International Conference on Intelligent User Interfaces, pp.348–350, 2006.
- [5] 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," コンピュータソフトウェア, vol.23, no.4, pp.36–46, 2006.
- [6] 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," 信学論 (D), vol.J90-D, no.10, pp.2817–2829, Oct. 2007.
- [7] P.Y. Chi, J.H. Chen, H.H. Chu, and J.L. Lo, "Enabling calorie-aware cooking in a smart kitchen," Proc. 3rd Int'l Conf. Persuasive Technology, pp.116–127, 2008.
- [8] A. Hashimoto, N. Mori, T. Funatomi, Y. Yamakata, K. Kakusho, and M. Minoh, "Smart kitchen: A user centric cooking support system," Proc. Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp.848–854, 2008.
- [9] 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," 情処学論, vol.50, no.4, pp.1299–1310, 2009.
- [10] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," IEEE Trans. Pattern Anal. Mach. Intell., vol.25, no.16, pp.1337–1342, 2003.
- [11] M.J. Swain and D.H. Ballard, "Color indexing," Int. J. Comput. Vis., vol.7, no.1, pp.11–32, 1991.
- [12] 津吹陽介, 小阪拓也, 亀田能成, 中村裕一, 大田友一, "さりげなく作業支援を行うメディア 物体変化の認識と作業過程の同定," 信学技報, vol.104, no.369, pp.13–18, 2004.
- [13] J. Zauner, M. Haller, A. Brandl, and W. Hartmann, "Authoring of a mixed reality assembly instructor for hierarchical structures," ACM Int'l Symp. on Mixed and Augmented Reality, vol.1, pp.237–246, 2003.
- [14] H. Grabner, J. Matas, L.V. Gool, and P. Cattin, "Tracking the invisible: Learning where the object might be," CVPR2010, 2010.

- [15] J. Kim, S.M. Seitz, and M. Agrawala, "Video-based document tracking: unifying your physical and electronic desktops," Proc. ACM Symp. on User Interface Software and Technology, pp.99-107, 2004.
- [16] 三上 弾, 大塚和弘, 大和淳司, "メモリベースパーティクルフィルタ: 状態履歴に基づく事前分布予測を用いた頑健な対象追跡," 信学論 (D), vol.J93-D, no.8, pp.1313-1328, Aug. 2010.
- [17] 三上 弾, 大塚和弘, 大和淳司, "姿勢とアピランスの変化に頑健な対象追跡を実現するアピランス統合メモリベースパーティクルフィルタ," MIRU, 2010.
- (平成 22 年 9 月 24 日受付, 23 年 1 月 7 日再受付)



橋本 敦史 (学生員)

平 17 京大・工・情報卒. 平 20 同大学院修士課程了, 現在, 同大学院博士課程に在籍中. 人間行動観測, 画像認識, 物体追跡に関する研究に従事.



森 直幸

平 19 京大・工・情報卒. 平 21 同大学院修士課程了. 人間行動観測, 物体追跡に関する研究に従事.



船富 卓哉 (正員)

平 14 京大・工・情報卒. 平 16 同大学院情報学研究科修士課程了. 平 18 日本学術振興会特別研究員 DC2. 平 19 京都大学大学院情報学研究科博士後期課程了. 平 19 京都大学学術情報メディアセンター助教, 現在に至る. 三次元モデル, 人間行動観測に関する研究に従事. 博士 (情報学).



山肩 洋子 (正員)

平 12 京大・工・情報卒. 平 17 同大学院博士後期課程単位認定退学, 同年京都大学研究員. 平 18 (独) 情報通信研究機構専攻研究員, 平 22 から京大学術情報メディアセンター特定講師, 現在に至る. 画像・映像認識, 音声対話システム, 立体音響システムなどの研究に従事. 博士 (情報学). H14 年度人工知能学会研究会優秀賞受賞.



松木 雅之 (正員)

平 3 京大・工・情報卒. 平 8 同大学院工学研究科博士後期研究指導認定退学. 同年同大学工学部助手. 平 12 同大学総合情報メディアセンター助手. 平 14 広島市立大学情報科学部助教授, 平 21 京都大学学術情報メディアセンター准教授, 現在に至る. 博士 (工学). 画像認識, コンピュータビジョン, 映像メディア処理の研究に従事. 情報処理学会会員.



角所 考 (正員)

昭 63 名大・工・電気卒. 平 5 阪大大学院工学研究科通信工学専攻博士課程了. 平 4~6 日本学術振興会特別研究員. 平 5~6 スタンフォード大学ロボティクス研究所客員研究員. 平 6 大阪大学産業科学研究所助手. 平 9 京都大学総合情報メディアセンター助教授. 平 19 同准教授. 平 21 関西学院大学理工学部教授. 視覚メディア処理, コミュニケーション, インタラクションに関する研究に従事. 博士 (工学). IEEE, ACM, 情報処理学会, 人工知能学会各会員.



美濃 導彦 (正員: フェロー)

昭 53 京大・工・情報卒. 昭 58 同大学院博士課程了. 同年工学部助手, 昭 62~63 マサチューセッツ州立大学客員研究員, 平元京都大学工学部附属高度情報開発実験施設助教授, 平 7 同教授, 平 9 京都大学総合情報メディアセンター教授, 平 14 京大学術情報メディアセンター教授. 画像処理, 人工知能, 知的コミュニケーション関係の研究に従事. 工博. IEEE, ACM, 情報処理学会, 画像電子学会, 日本ロボット学会各会員.

正 誤

平 23-07 (Vol. J94-A, No.7, pp.509-518) 特集論文

橋本敦史, 森 直幸, 船富卓哉, 山肩洋子, 椋木雅之, 角所 考, 美濃導彦:

把持の順序と外見の変化モデルを利用した調理作業における食材追跡

ページ	誤	正
p.510 脚注(注1)	http://hdi.handle.net/2433/139378	http://hdl.handle.net/2433/139378