

RGB-D センサを用いた皿および手の追跡による 摂食行動記録システムの提案

安田 陽介^{1,a)} 前川 卓也^{1,2,b)}

概要: 日々の食生活を記録する食事ログにおいて、RGB-D センサを用いることでユーザが食事を摂る際の詳細な食行動を自動的にセンシングできる食事ログシステムを提案する。デジタルカメラなどを用いた写真による食事の記録では、画像処理により食事全体のカロリーは取得できるが、ユーザがどの料理をどのような順序で摂食したかを記録することはできない。本研究では、RGB-D センサを用いて取得した時系列データから、皿とユーザの手の位置を追跡することにより詳細な摂食情報を取得する。これにより、ユーザはより詳細な食事記録を食事ログアプリケーションに利用することができる。

1. はじめに

近年、自身の生活を記録して共有することが一般的になり、個人のあらゆる行動や体験のデジタルアーカイブ化を目指すライフログ研究への注目が高まってきている。そのようなライフログを行う対象として、睡眠時間、食事内容、運動量といったものや twitter でのつぶやきなどが挙げられる。それらの中でも食事の記録は、近年特に注目されているライフログ対象の一つである。年々患者数が増加している肥満や高血圧、糖尿病といった生活習慣病は、偏った食生活が原因であるとされ、日々の食事内容を記録し自身の食べたものを把握することで、食生活を見直すことを可能とする食事ログは、そのような偏った食生活を改善する方法の一つであるとされている。そのため病院では、従来から患者の食事内容の記録が行われており、近年ではその記録は患者ごとにデジタル管理されている場合もある。しかし、これは提供される食事の予定が決まっている病院などの環境でのみ低コストで行えるものであり、そのような環境でも食べ残しなどがあった場合はその量などを患者や看護師が手動で入力する必要がある。このような記録を健康な人が普段から継続して行うのは非常に煩わしい。

一方、近年のスマートフォンの急激な普及により、手軽に高品質な画像を用いて食事内容の記録が可能になったこ

とで、写真を用いた食事ログが注目されている。さらに、記録した後は自身で確認するだけだった従来の食事ログと異なり、撮影した画像は Twitter や Facebook といった SNS にアップロード・共有され、SNS そのものが食事ログの場としても活用されるようになった。また健康意識の高まりを受け、Evernote Food や miil といった食事ログ専用の SNS サービスも開始されるようになった。さらに、食事画像からその食事に含まれるカロリーを推定する研究 [1] やご飯やおかず、野菜のバランスが適切かどうかを判定する研究 [2] など、食事画像を用いた研究が盛んに行われるようになった。しかし、このような食事写真のみを用いた食事ログでは、実際に摂取した量や摂取した順序、すなわち食事の過程の情報を記録することはできない。この食事の過程の情報は、健康管理・病気の治療において非常に重要であり、近年の糖尿病研究において、食事の内容とともに食事の摂取順序が食後の血糖値に大きく影響することも分かっている [3][4]。しかし、現在提供されている食事ログサービスでは、1 枚ないし数枚の食事画像から食事内容の記録を行っているだけで、食事時のログ、すなわち摂食順序などのログは記録していない。皿に添付された RFID タグと重量センサを用いる手法や、皿に添付した AR マーカーとガラステーブルの下部に設置されたカメラを用いる手法で摂食順序の記録を行う研究もなされているが、どちらも装置が非常に大掛かりになり、それぞれの皿や箸などにタグを添付する必要があるため、一般的に利用できるとは考えにくく、写真を用いた食事ログシステムの実現が現実的である。しかし、摂食順序のログを取得するためには、完成した食事の写真だけでなく、食事時の様子を常に記録しておく必要がある。そのためには、食事の様子を撮り続

¹ 大阪大学
Osaka University

² 国際電気通信基礎技術研究所
Advanced Telecommunications Research Institute International (ATR)

a) yasuda.yosuke@ist.osaka-u.ac.jp

b) maekawa@ist.osaka-u.ac.jp

けた映像を用いて食事ログシステムを実現する必要があるだろう。

本研究では、この連続的に撮影された映像を用いて、机の上の皿とユーザの手を追跡し、どの時刻にどの皿の食事を摂ったかを記録することで摂食情報の取得を試みる。提案システムでは、机の上の皿や手を追跡するため、家庭の食卓天井照明に着目する。食卓上の照明は食事の際に点灯されるため、照明から電力の供給を受けるセンサデバイスを用いて、食卓をユーザの頭上から真下に向けて撮影する。また、電源を照明から取得できるため導入も比較的容易である。その際、撮影用カメラには通常のRGBカメラに加えて深度センサを搭載したMicrosoft社のKinectを用いる。深度センサのデータを用いることで、机の領域を推定したあと、机上の皿および手領域の認識を行う。そして認識された皿および手領域を、SIFT特徴[5]およびパーティクルフィルタを用いて追跡する。最後に、追跡した皿と手領域の接触時間の情報を用いてその皿の食事を摂ったかどうかを判定する。以降では、2節で関連研究について説明した後、3節で提案手法について説明する。4節で提案手法の評価実験を行い、5節で本稿のまとめを行う。

2. 関連研究

2.1 食事ログに関する研究

食事ログに対するニーズの高まりを受け、食事ログに関する研究が広くなされるようになってきた。そのような食事ログに関する研究の中から、本研究と関連性の高いものをいくつか紹介する。北村ら[2]は、食事画像解析において特徴量の定量的な評価を行い、検出された特徴点を色情報と勾配情報によって記述することで、食事画像の栄養バランスの推定する食事ログシステムを提案した。このシステムでは、SIFT特徴などの局所画像特徴量を、Bag of Featureモデルに適用し、SVMを用いることで画像の分類を行う。また、食事画像解析のみでログの取得を行うため、食器やテーブルが制限されることなく、外食の記録にも対応している。川嶋ら[6]は、RFIDタグを用いた摂食モニタリングシステムを提案した。食器の近接情報をRFIDタグと感圧センサにより検出し、摂食量を重量センサで推定するため、食事の摂取量・摂取順序の記録が可能である。瀬戸ら[7]は、ARマーカとカメラを用いた追加品目推薦システムを提案した。手と皿の裏面に添付したARマーカを、机の下方から透明なガラステーブルの裏を撮影するように設置したカメラで認識し、摂食順序の記録を行う。これらの手法は、摂食順序の記録を想定していないか、もしくは摂食順序の記録は行うが、使用する装置が非常に大掛かりになるという問題点がある。

2.2 食事の摂取順序に関する研究

食事の順序の情報は、健康管理・病気の治療において非

常に重要であるとされている。糖尿病の研究において、食事の摂取順序に関するものを紹介する。今井ら[3]は、2型糖尿病患者を対象に、野菜を米飯の後に摂取した場合と米飯の前に摂取した場合の、摂取後の血糖値および血清インスリン値を無作為クロスオーバー法により調べた。野菜から先に摂取すると、米飯から先に摂取した場合と比較して、30分後、60分後の血糖値は低い値を示した。また、インスリン値も同様に30分後、60分後共に有意に抑制された。これらの結果より、「食べる順番」を重視した容易な教育方法が食事指導に重要であると述べている。金本ら[4]は、10人の健常成人を対象に、野菜サラダ(キャベツ、オリーブ油、酢)と米飯の摂取順序を変えた時の食後の血糖値とインスリン値のプロファイルの変化を調べた。その結果、米飯摂取後に野菜サラダを摂取した場合と比較して、米飯摂取前に野菜サラダを摂取した場合には、食後20分、30分、45分での血糖上昇値は有意に低下し($p < 0.01$)、最高血糖値に到達する時間は約40分遅延した($p < 0.01$)。これらの結果より、野菜サラダは米飯よりも先に摂取するほうが食後の血糖上昇を抑制するために有効であると結論付けている。これらの研究では、摂食順序の変更が糖尿病の治療に役立つことを示している。また、血糖値の乱高下を抑えられ腹持ちが良くなることから、近年では新たなダイエット手法として注目されてきている。このことから、摂食順序の記録が可能な食事ログシステムを実現することで、これまでになかった効果的な摂食管理が可能になると考える。

3. 提案手法

3.1 提案手法の概要

本研究で提案する手法の概要を図1に示す。手法は以下の4つの手順から構成される。

- (1) Kinectを用いたデータの取得・深度データの補正
- (2) 皿の追跡
- (3) 手の追跡
- (4) 摂食の判定

提案する手法では、Kinectをユーザの頭上から真下に向けて設置することで、連続的に画像と深度データを取得する。まず取得した深度データに対してずれや測定不能値などの補正を行ったあと、補正されたデータを用いて皿の追跡・手の追跡を行う。皿の追跡では、まず深度データか

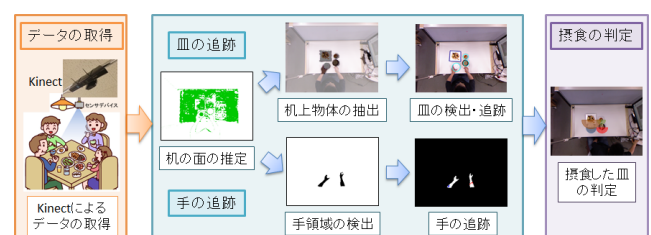


図1 提案手法の概要

ら机の面の推定および机の領域の推定を行う。この推定された机の面・領域は以降の手の追跡でも利用する。次に、推定された机の面より高い位置にある物体（机上物体）を抽出する。抽出された机上物体に対しノイズ除去を行った後、円検出・矩形検出を用いて皿の検出を行う。検出された皿に対しては SIFT 特徴を用いて以降の時刻において追跡を行う。すなわち、検出された皿と以前の時刻から追跡されている皿から重複を除いたものをその時刻に机上にある皿とする。手の追跡では、まず肌色抽出と深度データを用いて手領域の抽出を行う。抽出された手領域は、皿（料理）などに比べて濃淡の変化が少ないため、SIFT 特徴などを用いて追跡するのは難しい。そのため、本研究ではパーティクルフィルタを用いて手の追跡を行う。最後に、これらを追跡することで得られた皿・手の位置情報から摂食の判定を行う。以降ではこれらの手法の詳細を説明する。

3.2 Kinect を用いたデータの取得

机の上の皿や手を追跡するため、連続的に画像を取得するデバイスは、机全体を俯瞰できる位置にあることが好ましい。そこで本研究では天井照明に着目する。机上の照明は食事の際に点灯されるため、照明が点灯した際にそこから電力の供給を受けるようにデバイスを設置する。照明に設置することで図2のように机全体をユーザの頭上から真下に向けて撮影できる。また、撮影デバイスとして Microsoft 社の Kinect を用いる。Kinect を用いることで、同期した RGB 画像と深度データを容易に取得できる。

図3は取得した深度データをグレースケールの濃淡画像として表示した例であり、各ピクセルの濃淡値は深度の値に対応している。また、深度が浅すぎるピクセル、深すぎるピクセルは緑色で表示している。図3の右部分と上部分の空白は、RGB カメラでは観測できているが、深度センサでは観測できていない領域である。これは、RGB カメラと深度センサの観測範囲が異なるためである。また深度データには、計測不能点があくつか存在する。これら計測不能点を、その点の周囲の値の平均で補完することで以降の過程にて発生するノイズを減少させる。

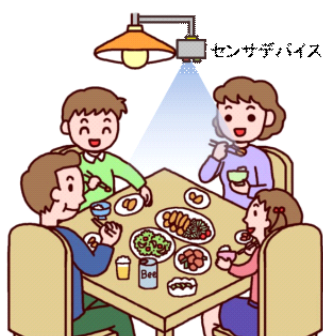


図2 天井にデバイスを設置するイメージ図

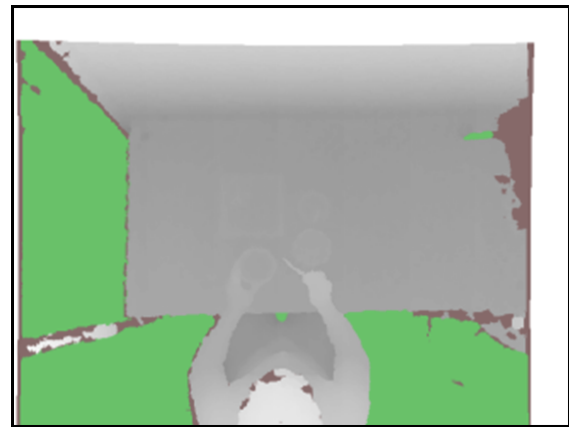


図3 深度データの濃淡画像の例

3.3 皿の追跡

皿および手の追跡の前準備として、まず深度データから机の面推定および机の領域の推定を行う。本手法において撮影デバイスは机を上部から撮影するため、画面内の最も広い平面が机であると考えられる。本研究では、まず深度画像から無作為に選んだ3点に対して、式1の p_x に x 座標、 p_y に y 座標、 p_z に深度を代入することで得た3元連立方程式を解くことで、その3点が構成する平面の式を求める。

$$ap_x + bp_y + cp_z + d = 0 \quad (1)$$

同様の方法で200個の平面の式を導出し、それらの面の中で、深度データの値が面の式で導出される深度値の許容誤差(5mm)以内に含まれるピクセルの数が最も多い面を机の面(の式)とする。この方法により、深度画像に含まれる最も大きい平面、すなわち机を発見する。次に、深度画像の行・列それぞれに対し、推定された机の面に含まれるピクセルの割合をカウントし、図5に示すようにその割合が閾値を超える範囲を机の領域と設定する。この推定された机の面・領域は以降3.4節で手の追跡にも利用する。また以降は、この机の領域内に含まれるデータのみを取り扱う。

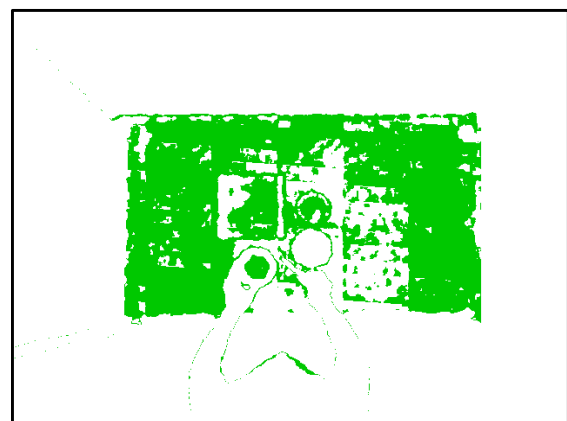


図4 机の面に含まれるピクセルを緑色で強調表示した画像

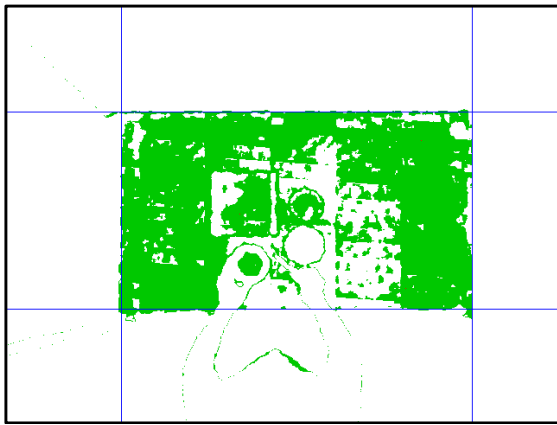


図 5 図5に机の範囲を描画した画像
中心の領域が検出された机の範囲に対応する

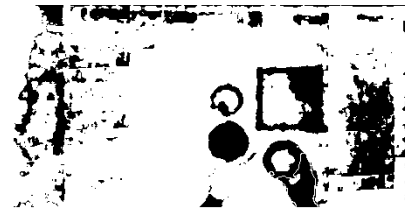


図 6 机上物体抽出

3.3.1 机上物体抽出

皿の検出のために、まず机の上に存在している物体（机上物体）を抽出する。本研究における深度データは、机を上から Kinect より撮影したデータであるため、より深度の浅い物体は、より高い位置にある。そこで、深度の値が机の面の式から導出される深度よりある閾値 (5mm) 以上浅いピクセルの集合を机上物体とする。すなわち、机上にある物体を検出する。しかし、図 6 に示すように検出された机上物体にはノイズが多く含まれている。そこで、以下の 3 つの方法を用いてノイズの除去・修正を行う。

- (1) 局所平面除去
- (2) SLSF アルゴリズムによる穴埋め
- (3) 輪郭抽出によるノイズ除去

まず、深度データから局所的に平面である範囲は机の面に対応していると考え、その範囲のデータは机上物体から除去する。その結果を図 7 に示す。次に SLSF (スキャンライン・シード・フィル) アルゴリズムを用いた穴埋めを行う。この SLSF アルゴリズムとは、初期シードが含まれる閉平面を塗りつぶすためのアルゴリズムである。まず初期シードを含む列を走査し、両端まで走査が終わると、走査済みの空白ピクセルを塗りつぶす。その後、その上下に次シードを配置し、次シードに対して同じ処理を繰り返し行う。図 7 の画素のない領域群に対して SLSF アルゴリズムを適用し、机上外と繋がっていない画素の無い集合を塗りつぶす。皿の中に食べ物が入っている場合、赤外線を用いた深度センサでは、食べ物が赤外線を吸収してしまうため、皿の内部は正確な深度が得られず空白ピクセルとなってしまう。そのため、縁だけ抽出された皿の内部を塗りつぶすことでこの問題に対応する。最後に輪郭抽出を行い、閾値以下の大きさを持つ領域をノイズと判断して取り除く。その結果を図 8 に示す。

3.3.2 皿の検出

前節で抽出した机上物体から皿の検出を実現する。検出手法を決定するために、まず一般的に使われている皿の形

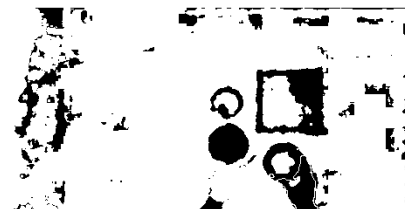


図 7 局所平面除去適用後

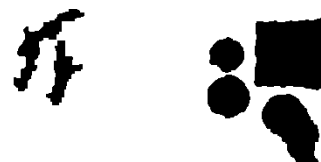


図 8 SLSF アルゴリズムによる穴埋めおよび輪郭抽出によるノイズ除去適用後

状の調査を行った。英語圏を中心に広く利用されている Allrecipes.com というレシピ投稿サイトから、料理写真をランダムに 180 枚選び、写真に写っている皿の形状を調査した。形状としては大まかに、円形、四角形、それ以外に分類された。円形に関しては真円および楕円に詳細に分類し、四角形に関しては正方形・長方形および角や辺に丸みのある四角形として詳細に分類した。さらに、コップ・グラタン皿を想定した持ち手の有無に関する特徴でも分類した。その結果を表 1 に示す。9 割弱が円形のもの、1 割強が四角形のものに分類された。また想定とは異なり、円や四角以外の例外の形の皿は見られなかった。

次に、日本の一般家庭 2 世帯から、日常的に使われている計 247 枚の皿の形状を調査した。形状の分類は先ほどと同様に行った。結果を表 2 に示す。三角形の皿や、キャラクターの形をした皿が数枚見られたものの、9 割弱が円形のもの、1 割強が四角形のものといった結果となった。

以上の調査より、机上物体から円形および四角形の物体を検出することにより、ほとんどの皿の検出が可能であると考えられる。したがって、本研究では机上物体を抽出した画像に対して、Hough 変換を用いた円検出・矩形検出を用いて、皿の検出を行う。Hough 変換を用いた円検出・矩形検出手法の詳細については [8] を参照されたい。その結果を図 9、図 10 に示す。図 9 中において、皿同士が重なっていたため検出できていない皿が存在しているが、一度検出された皿に関しては、検出後に画像特徴を用いた追跡を行うため、連続して撮影される画像の皿が遮蔽されていないタイミングにおいて検出できれば問題ないと考えられる。

3.3.3 皿の追跡

机上物体から、その時刻での皿の検出を行うと同時に、以前の時刻にて検出された皿の SIFT 特徴を用いた追跡を行う。SIFT(Scale-Invariant Feature Transform) は、特徴点の検出と特徴量の記述を行うアルゴリズムである。検出した特徴点に対して、画像の回転・スケール変化・照明変化等に頑健な特徴量を記述することができる。現時刻の画像に対して SIFT 特徴点の検出を行い、以前の時刻にて検出された皿の SIFT 特徴点とのマッチングを行う。マッチングされた特徴点対から皿の移動・回転・スケール変化を推定し、現時刻での皿の位置を推定する。この処理を毎時



図 9 皿が一部検出できていない画像

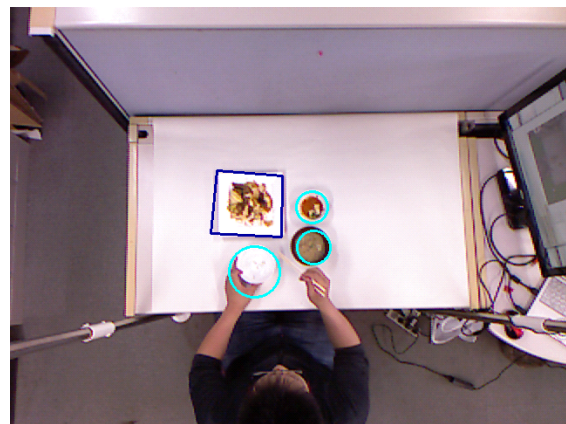


図 10 皿がすべて検出されている画像

刻ごとに行うことで皿を検出後以降から追跡できる。ただし、皿の中の料理は徐々に消費され、皿内の SIFT 特徴も徐々に変化するため、追跡に用いる SIFT 特徴は毎時刻ごとにアップデートしている。これら検出された皿と追跡した皿から重複を除いたものをその時刻に机上にある皿とする。

3.4 手の追跡

3.4.1 手領域の検出

手の追跡では、まず肌色抽出と深度データを用いて手領域の抽出を行う。RGB 画像に対して肌色抽出を適用し手領域の候補を抽出する。結果を図 11 に示す。手領域だけでなく、皿の中の肌色に近い食材の領域も誤って抽出されているため、深度データを用いて食材の領域を取り除く。肌色抽出された領域に含まれるピクセルのうち、深度の値が机の面の式から導出される机の高さより一定の範囲内の高さ(+5cm~+30cm)に含まれるピクセルのみを手領域であるとす。その結果の例を図 12 に示す。

3.4.2 手の追跡

人は利き手に箸やフォークを持って摂食行動を行うため、右手と左手をそれぞれ追跡する必要がある。すなわち、検出された手領域に対して、右手と左手のラベルを付与す

表 1 Allrecipes から取得した皿の分類結果

概形	形状	特徴	該当数	割合	
円	真円	持ち手無し	151	83.9%	88.3%
		持ち手付き	4	2.2%	
	楕円	持ち手無し	3	1.7%	
		持ち手付き	1	0.6%	
四角形	正方形	丸みなし	4	2.2%	11.7%
		丸みあり	10	5.6%	
	長方形	丸みなし	3	1.7%	
		丸みあり	4	2.2%	
その他			0	0.0%	0.0%

表 2 日本の 2 家庭から取得した皿の分類結果

概形	形状	特徴	該当数	割合	
円	真円	持ち手無し	176	71.3%	87.4%
		持ち手付き	29	11.7%	
	楕円	持ち手無し	6	2.4%	
		持ち手付き	5	2.0%	
四角形	正方形	丸みなし	6	2.4%	10.1%
		丸みあり	8	3.2%	
	長方形	丸みなし	6	2.4%	
		丸みあり	5	2.0%	
その他			6	2.4%	2.4%



図 11 肌色抽出のみを用いた手領域抽出結果の例



図 12 深度を利用した手領域抽出結果の例

る必要がある。2つの手領域が検出されているとき、多くの場合においてそれらの位置で左右を判別することはできるだろう。しかし皿が置かれている位置によっては、必ずしも右手が左手より右にあるとは限らない。また、食事を口元に運んだ際は手が頭に隠れてしまうため、手領域が検出されなくなってしまう。このような状況においても、ロバストに左右の判別ができるように、パーティクルフィルタを用いて左右の手の追跡を行う。パーティクルフィルタ [9] は、非線形な状態遷移を行うシステムの状態を予測するために用いられる。そのアルゴリズムは、サンプリング、重み計算、リサンプリングの3ステップからなる。本手法では、重み計算の代わりに、パーティクルのスコアリングを行う。まず、初めて一定以上の大きさの手領域が2つ抽出された時、手領域の左右の初期化を行う。検出された手領域の位置から単純に手の左右を決定し、それぞれの領域ごとにその中心からガウス分布に従って20個の初期パーティクルを生成する。その後新たな時刻の手領域が検出された時、それぞれの領域に対して左右両方のフィルタを適用し、左右の手の追跡を行う。以下、追跡の流れを説明する。

まずサンプリングにおいては、時刻 t のパーティクル p_t から、新たなパーティクルを作成し、移動と静止に対応す

る2つの移動モデルに基づいてそのパーティクルを移動させる。このパーティクルがトラッキング対象の状態(位置)を表す。1つ目の移動モデルは、 $p_{t+1}^i = Ap_t + w_v$ である。ここで、 A は遷移行列であり、今回は等速直線運動を仮定したものを用いている。すなわち、 p_t と p_{t-1} から求めた速度をそのまま利用している。 w_v は速度に比例した分散を持つ平均0の正規分布であり、これによりガウシアンノイズを含んだ移動を表現している。2つ目の移動モデルは、 $p_{t+1}^i = p_t + w$ である。 w は定数値の分散を持つ平均0の正規分布であり、これによりガウシアンノイズを含んだ静止を表現している。パーティクルフィルタでは、これらの移動モデルを用いて1つの p_t からガウシアンノイズに従って複数の p_{t+1}^i をサンプリングする。 $(i$ はパーティクルの識別子。) 今回の実装では1つのパーティクルをそれぞれの移動モデルにつき5つずつ、計10個に分裂させている。

スコアリングにおいては、観測(検出された手領域)を用いてパーティクルへのスコアリングを行う。スコアリングは左右の手ごとに行う。本手法では、検出された手領域に近いパーティクルが多いほど、実際にその場に手が移動した可能性が高いと考えられるため、近いパーティクルに高いスコアを与える。 p_{t+1}^i のスコアは、式2を用いて計算する。このとき、 D は検出された手領域の中心と p_{t+1}^i とのユークリッド距離とし、 p_{t+1}^i が検出された手領域内にあるならば、 D を0として計算する。

$$score = \frac{1}{D+1} \quad (2)$$

全てのパーティクルのスコアを計算し、その平均を手領域に対するフィルタのスコアとする。検出された手領域に対して左右のフィルタのスコアを比較することで、その手領域が左右どちらの手の移動したものなのかを判別する。リサンプリングにおいては、スコアの高いパーティクルから順に一定数を保持する。今回の実装では、高い順に20個のパーティクルを保持する。図13にリサンプリングされたパーティクルを示す。この3ステップを繰り返すことで手の追跡を行う。

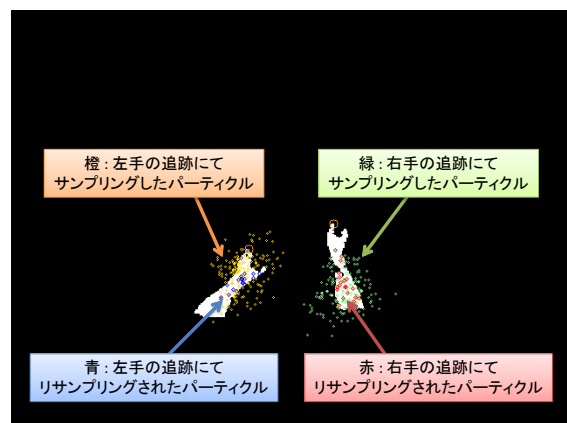


図 13 リサンプリング後のパーティクル+手領域の画像

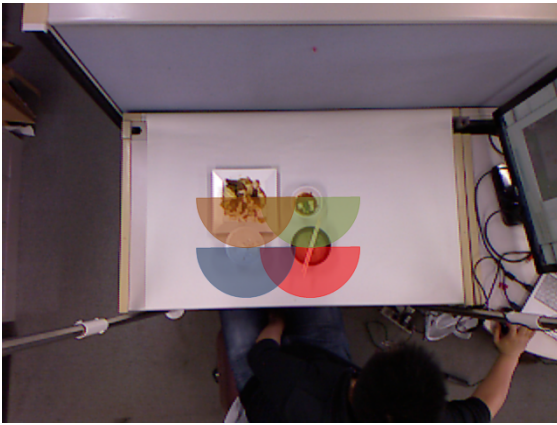


図 14 設定した摂食判定領域の例

3.5 摂食判定

皿と手を追跡することにより得られたそれぞれの位置情報から、摂食の判定を行う。本研究では、手の先端が皿の接触判定領域内に一定時間以上続けて存在していたとき、ユーザはその皿にアクセスし、皿の料理を摂食したと判定する。図 14 に接触判定領域の例を示す。接触判定領域は皿と同じ中心と、一定の大きさの半径をもつ上半分が欠けた半円とする。箸やスプーンといった道具を使って食事を摂る場合、皿にアクセスした手の先端が皿の範囲内に入るとは限らない。そのため上記のような範囲を接触判定領域と設定した。

4. 評価実験・考察

本節では手の追跡および摂食の判定の評価を行い、その結果に対する考察を行う。今回の実験では、撮影デバイスとして Kinect for Windows を用い、ユーザの頭上から真下に向けて設置して一回分の食事データの取得を行った。

4.1 手の追跡の評価

4.1.1 データセット

評価には取得したデータの内、左右の手が初めて同時に現れた後の 256 秒間、1939 フレームのデータを用いた。

4.1.2 評価方法

手の追跡の評価は、検出された手領域の適合率(式 3)と再現率(式 4)にて行う。

$$\text{適合率} = \frac{\text{左右が正しく認識された手領域の数}}{\text{検出された手領域の数}} \quad (3)$$

$$\text{再現率} = \frac{\text{左右が正しく認識された手領域の数}}{\text{画面内に手が出現した数}} \quad (4)$$

4.1.3 結果と考察

手の追跡の結果および評価を表 3 に示す。

表 3 に見られるように、今回の実験では全体の約 28% にあたる 603 個の手領域に対して、左右判定誤りが起きた。その原因を考察する。Kinect for Windows では RGB センサの撮影準備と深度センサの撮影準備が別々に行われてい

るが、本手法では同期されたデータを取得するために、両方の準備が完了し次第撮影を行う。そのため、何らかの理由でどちらかの処理が滞り、数秒から十数秒にかけてデータが取得できない場合がある。今回の実験では、1289 フレーム撮影後、1290 フレームを撮影するまで 12.74 秒かかった。表 4 に 1289 フレーム付近の撮影時刻を記載する。

本手法では、一つ前のフレーム間の速度から現在のフレームでのパーティクルの位置を推定する。そのため 1289 フレーム目前後では、フレーム間 0.04 秒のパーティクルの移動で作られた速度から 12.73 秒後のパーティクルの移動先が推定され、また 12 秒間の間に大きく手が移動してしまったため、1290 フレーム目の追跡において、右手用および左手用のパーティクル群の左右の位置が反転してしまった。そして本手法では、入れ替わってしまった左右の位置の修正を行っていないため、それ以降実験終了まで位置が反転した状態であった。表 5 に 1289 フレームまでの結果を示す。表 5 の結果においては左右判定の誤りは無く、画面内に手が一本しか写っていない場合でもその手がどちらか正確に判定できていた。今後は、Kinect for Windows のデータ取得エラーが起こった場合に左右の手の初期化処理などを行うことで、認識誤りを修正する必要がある。

表 3 手の追跡の結果および評価

認識結果	左	右	両
正しく左右を認識	667	785	1452
手が検出されず	12	42	54
誤った位置に手が検出	8	6	14
左右判定誤り	263	340	603
適合率	0.711	0.694	0.702
再現率	0.702	0.669	0.684

表 4 1289 フレーム付近の撮影時刻

フレーム	撮影時刻
1287	14 時 38 分 11 秒 48
1288	14 時 38 分 12 秒 17
1289	14 時 38 分 12 秒 21
1290	14 時 38 分 24 秒 94
1291	14 時 38 分 25 秒 77

表 5 1289 フレームまでの手の追跡の結果および評価

認識結果	左	右	両
正しく左右を認識	665	784	1449
手が検出されず	12	42	54
誤った位置に手が検出	6	6	12
左右判定誤り	0	0	0
適合率	0.991	0.992	0.992
再現率	0.974	0.942	0.956

4.2 摂食の判定の評価

4.2.1 データセット

評価には取得したデータの内の、左右の手が初めて同時に現れた後、Kinect for Windows のデータ取得エラーが起こるまでの 147 秒間、1289 フレームのデータを用いた。

4.2.2 評価方法

摂食判定の評価は、摂食判定の適合率 (式 5) と再現率 (式 6) にて行う。

$$\text{適合率} = \frac{\text{正しく摂食が判定された回数}}{\text{摂食と判定された回数}} \quad (5)$$

$$\text{再現率} = \frac{\text{正しく摂食が判定された回数}}{\text{実際に摂食した回数}} \quad (6)$$

実際に摂食した回数は画像から目視で計数し、11 回であった。また摂食の判定に用いるパラメータである $T1$, $T2$ を変化させて性能評価を行う。 $T1$ は皿と手の接触が検出されなくても接触状態が維持される最大時間 (秒), $T2$ は摂食と判定する最短の接触時間 (秒) を表す。

4.2.3 結果と考察

摂食判定の結果および評価を表 6 に示す。 $T1$ を 1.0 と小さく、 $T2$ を 2.0 と大きく設定することで摂食判定の誤検出を削減した場合は、適合率が 0.667, 再現率が 0.363 となった。この場合において誤って摂食と判定された 2 つの皿は、手を休めている時に右手付近にあった皿、および皿の検出によって誤検出された皿であり、パラメータの設定でこれ以上誤判定を削減することは困難であると考えられる。そのため今後は、皿の内部の画像特徴を用いて、検出された皿の料理らしさにより皿の誤検出を削減する必要がある。 $T1$ を 1.5 と大きく、 $T2$ を 1.0 と小さく設定することで検出数を増加させた場合は、適合率が 0.380, 再現率が 0.727 となった。この手法において摂食と判定できなかった 3 つの皿は、追跡に失敗した皿であった。そのため今後は、皿の内部の画像特徴だけでなく、深度データに対しても SIFT 特徴量を抽出し、追跡を行うことで皿の追跡精度の向上を試みる。

5. おわりに

本研究では Kinect を天井の照明に設置することで、摂食順序を自動取得する食事ログシステムを提案した。提案システムでは、取得した RGB 画像、深度データから机の上の皿、ユーザの手を追跡し摂食の判定を行う。皿の追跡は、深度データから机の上に物体の存在するピクセル (机上物体) を抽出し、机上物体に対して円検出・矩形検出を用いることで行った。また手の追跡では、パーティクルフィルタ

を用いることで、片方の手しか検出されなかった時でも正確に左右の判定が可能な手法を実現した。しかし、Kinect for Windows のデータ取得にエラーがあった場合に識別精度が低下した。今後はこの問題に対処する必要がある。摂食の判定では、追跡した皿と手の位置から摂食した皿を判定する。しかし、皿の誤検出や追跡失敗により摂食の判定を失敗した例が多くみられたため、今後はこの問題にも対処する必要がある。

謝辞 本研究の一部は、財団法人大川情報通信基金 2014 年度研究助成の支援により行われた。

参考文献

- [1] Tatsuya Miyazaki, Gamhewage C de Silva, and Kiyoharu Aizawa: *Image-based Calorie Content Estimation for Dietary Assessment*, In IEEE International Symposium on Multimedia (ISM)2011, pp. 363-368. 2011.
- [2] 北村圭吾, 山崎俊彦, 相澤清晴: 食事ログの取得と処理—画像処理による食事記録—, 映像情報メディア学会誌, Vol.63, No.3, pp. 376-379, 2009.
- [3] 今井佐恵子, 梶山静夫: 食品の摂取順序を重視した糖尿病栄養指導の血糖コントロール改善効果, 日本糖尿病学会, Vol.55, No.1, pp. 1-5, 2012.
- [4] 金本, 郁男, 井上裕, 守内匡, 山田佳枝, 居村久子, 佐藤眞治: 低 *Glycemic Index* 食の摂取順序の違いが食後血糖プロフィールに及ぼす影響, 日本糖尿病学会, Vol.53, No.2, pp. 96-101, 2010.
- [5] 都築勇司, 藤吉弘巨, 金出 武雄: *SIFT* 特徴量に基づく *Mean-Shift* 探索による特徴点追跡, 情報処理学会論文誌, コンピュータビジョンとイメージメディア, Vol.49, No.SIG 6, pp. 35-45, 2007.
- [6] 川嶋稔夫, 谷杉泰苗, 光藤雄一: センシングトレイと *ID* ウェアを用いた摂食モニタリングシステム, 電子情報通信学会技術研究報告. WIT, 福祉情報工学, Vol.106, No.285, pp. 61-66, 2006.
- [7] 瀬戸優貴, 野口康人, 登坂蘭, 井上智雄: 実物体履歴による食事状況の認識に基づく追加品目推薦システムの開発, 電子情報通信学会技術研究報告. MVE, Vol.107, No.554, pp. 55-60, 2008.
- [8] 渡辺孝志, 石戸橋真: 線分近似による一般化ハフ変換の高速化と任意図形検出, 電子情報通信学会論文誌 D. The Institute of Electronics, Information and Communication Engineers, Vol.74, No.8, pp. 995-1003, 1991.
- [9] A. Doucet, N. De Freitas and N. Gordon: *Sequential Monte Carlo methods in practice*, Springer Verlag (2001), 2001.
- [10] D.G.Lone: *Object recognition from Local Scale-Invariant features*, Proc. of ICCV, pp. 1150-1157, 1999.

表 6 接触判定の結果および評価

$T1$	$T2$	正解数	不正解数	総数	適合率	再現率
1.0	2.0	4	2	6	0.667	0.364
1.5	1.5	6	6	12	0.500	0.545
1.5	1.0	8	13	21	0.380	0.727