

機械学習を用いた摂食行動認識手法の実現と 食画像ラベリング環境の構築

佐藤 琢磨¹ 安田 陽介¹ 中井 大輔¹ 増田 彬¹ 前川 卓也¹

概要: 近年のスマートフォンやデジタルカメラの普及により、日々の食事を写真に撮影してクラウド上に記録・管理する人が増えてきている。しかし、写真のみを用いた食事ログでは何を食べたかの情報しか記録できない。何をどれだけ食べたか、どういった順で食べたかなどの食事の摂り方はユーザーの健康に大きな影響を与えるとされており、そのような情報を記録することができる食事ログ手法が求められている。そこで本研究では、RGB-D カメラにより食卓を上部から撮影し、機械学習を用いてより精度の高い皿のトラッキングおよび摂食行動認識を実現することを目指す。具体的には、追跡する皿の画像から特徴量を抽出し、特徴に応じて適切な追跡手法を自動的に切り替える手法を提案する。追跡手法としてテンプレート追跡、深度追跡、SIFT 追跡などが一般的だが、色に特徴が無い白米や餅などは SIFT 追跡やテンプレート追跡に向かず、深度に特徴がない平皿などの追跡には深度追跡は向いていない。これらを皿の特徴に応じて切り替えることで、高精度な追跡を行う。また、色情報と深度情報から手の追跡を行い、皿と手の相対位置などの特徴量を抽出し、隠れマルコフモデルを用いて摂食行動の認識も実現する。さらに、本研究では機械学習を用いた食行動認識・ライフログを容易にするため、クラウドソーシング型のラベリングシステムを開発する。システムでは、ユーザはウェブ上から任意の画像のラベリングを GUI を用いて容易に行うことができる。またユーザの功績がひと目で分かるようにラベリング数によるランキング機能も備える。

キーワード: ライフログ, 食事, 行動認識, 画像処理, RGB-D センサ

Implementing Recognition Method for Eating Activity based on Machine Learning and Constructing A Labeling Environment for Food Image

TAKUMA SATO¹ YOUSUKE YASUDA¹ DAISUKE NAKAI¹ AKIRA MASUDA¹ TAKUYA MAEKAWA¹

Abstract: Due to the recent proliferation of smartphones and digital cameras, the number of people who record and manage the daily dietary life by taking photos of dishes and upload them to an on-line service has been increasing. However, it is reported that the amount and order of food intake greatly influence human health and so a new meal log method that can record such information has been required. Therefore this research aims to achieve precise eating activity recognition by precisely tracking dishes and human hands based on machine learning from RGB-D images taken from ceiling. Specifically, we propose a method automatically switching an adequate tracking method according to feature value of a dish tracked. While template tracking, depth tracking and SIFT tracking are popular tracking method, SIFT and template tracking are not suitable for tracking such colorless food as rice and rice cake and depth tracking is not suitable for thin plates. Precise tracking is implemented by switching method according to dish's features. Eating activity recognition is implemented by using the Hidden Markov model after tracking hand using color and depth values and then such feature values as relative positions between a hand and a dish are extracted. Moreover we develop a labeling system based on crowdsourcing that permit us to easily implement eating activity recognition and life logging based on machine learning. The system enables us to label uploaded images on the web by GUI easily. The system also has a function that ranks contributors based on their contributions of labeling.

Keywords: meal log, life log, sensor device, image processing, RGB-D sensor

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、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つ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のみであ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おり、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、写真を用いた手法で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、摂食を始める直前の食事の写真だけでなく、食事の様子を常に記録しておく必要がある。そのため、食事の様子を撮り続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。

本研究では、この連続的に撮影された画像を用いて、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。まず、深度データを用いて画像内のどの領域が机(食卓)の面かを推定する。次に、皿の色などの特徴量を抽出し、特徴に応じた皿追跡手法の自動切り替えを行う。これは皿に盛られている料理によって、既存の追跡手法に得手不得手があるためである。また同時に、肌色抽出と深度データを用いて手領域を検出し、パーティクルフィルタを用いて逐次的に追跡する。それらの追跡位置を用いて各時刻における摂食皿(料理)の判定を行う。

2. 関連研究

2.1 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査した。野菜から先に摂取する場合、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り、摂食順序の変更が糖尿病の治療に役立つことが示された。

2.2 物体追跡

物体追跡とは、与えられた動画像から、指定した対象が画像上のどこに移動するかを推定する問題である。追跡手法には領域ベースの追跡、特徴点ベースの追跡、深度ベースの追跡などがある。

¹ 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University, Suita, Osaka, Japan

2.2.1 領域ベースの追跡

動画像上の物体の追跡をするには追跡対象の画像上の領域を指定する必要がある。領域ベースに物体を追跡する場合、特徴量としてユーザの指定した矩形領域から得られるカラーヒストグラムを用いる。領域ベースの追跡手法としてテンプレートマッチングやパーティクルフィルタなどがある。

テンプレートマッチングは探索領域内をラスタスキャンする手法であるが、回転やスケールを考慮した場合、膨大な計算コストがかかるという問題がある。パーティクルフィルタは Brasnett ら [5] により提案され、過去の情報と現在の観測情報から現在の対象の状態を推定する手法である。しかし、色情報を用いた手法は、色情報が乏しい場合に十分な追跡精度を得ることができない。

2.2.2 特徴点ベースの追跡

カラーヒストグラムを特徴量とした追跡では、照明変化が起こる環境下では、追跡が困難である場合がある。そこで、追跡対象領域内の照明変化を受けにくい追跡に有効な特徴点を検出、マッチングし、その特徴点の移動量を求める追跡手法が提案されている。主な特徴点追跡手法として SIFT 特徴量を用いた手法がある。

SIFT 特徴とは画像の回転・拡大・縮小や照明変化に頑健な特徴 [6] であり、特徴点周辺領域の持つ勾配情報を用い、SIFT descriptor と呼ばれる 128 次元の特徴量で記述される。また、128 次元の各特徴ベクトルの長さは、ベクトルの総和で正規化するため、照明変化の影響を受けにくい特徴量となる。都築ら [7] は、回転やスケールの変化に対して頑健な、SIFT 特徴量を用いた Mean Shift 探索による特徴点記述法を提案している。この手法では、SIFT により記述された特徴量を用いて Mean Shift による画像空間とスケール空間を交互に局地探索することにより、特徴点の移動量とスケールを同時に探索し追跡を行う。

2.2.3 深度ベースの追跡

深度情報を用いて、色特徴によらない物体追跡を実現した研究は多く行われており、Xia ら [8] は深度画像から人物の頭部を検出し、その結果をもとに人物輪郭を取り出し、その輪郭に含まれる 2 次元画像平面上での領域の重心によって人物を追跡する手法を提案している。Gokturk ら [9] は様々な位置・姿勢の上半身の深度情報から形状に関する特徴量を抽出し、k-means 法を用いたクラスタリングにより分類モデルを学習することで頭部位置の推定を行っている。

2.3 手領域検出

従来から 2D カメラから得た情報をもとに手領域の検出する研究が数多く行われてきた。Starner ら [10] は、ユーザの肌の色を検出し、手の位置と形を推定する手法を提案している。Diamanti ら [11] は、色だけでなく輪郭抽出を用いて手の位置と形を推定する手法を提案している。しか

しこれらの手法は、肌の色には個人差があること、照明の影響を大いに受けること、背景に似たような色があると誤認識してしまうことなどからロバスト性に問題があるとされている。

3. 提案手法

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、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図 1 のように設置する。撮影デバイスとしては RGB 画像と深度データを取得できる Microsoft Kinect v2 を用いる。



図 1 天井にデバイスを設置するイメージ図

図 2 にデバイスを用いて取得したデータによる摂食行動認識手法を示す。センサデータを用いて手と皿を追跡し、それらの位置関係からどの皿を摂食したかを認識する。提案手法は、まず、画像内のどの領域が机(食卓)の面かを推定する。次に、皿の色などの特徴に応じた皿追跡手法の自動切り替えを行う。これは皿に盛られている料理によって、既存の追跡手法に得手不得手があるためである。また同時に、肌色抽出と深度データを用いて手領域を検出し、パーティクルフィルタを用いて逐次的に追跡する。それらの追跡位置から摂食皿(料理)の判定を行う。以降ではこれらの手法の詳細を説明する。

3.1 机の面の推定

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

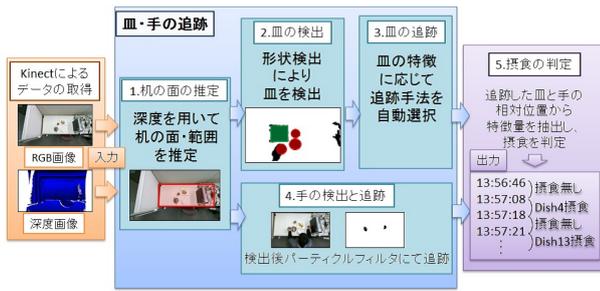


図 2 提案手法の概要

める。

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

同様の方法で無作為に生成した多数の平面 (5000 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。その後、推定した机に含まれる深度ピクセルから机の領域を設定する。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。

3.2 皿の検出

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 に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。



図 3 机上物体抽出の結果

しかし、図 3 に示すように検出された机上物体にはノイズが非常に多く含まれている。そのため、ノイズの除去・修正を行う。まず、画像を小さなウィンドウ (一辺 15 ピクセル四方の正方形) に区切り、ウィンドウ内のピクセルの深度の分散が閾値以下の範囲を局所的に平面な範囲であるとする。ウィンドウ内が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。また、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い皿の内部の正確な深度が得られず空白ピクセルになってしまう。そのため、SLSF (スキャンライン・シード・フィル) アルゴリズム [12] を用いた穴埋めを行うことでこの問題に対応する。

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。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。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 4 に示す。その後、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、皿の検出を行う。これは事前調査により、一般的な皿の形状は約 98 % が円形か四角形であることが明らかになったためである。



図 4 ノイズ除去適用後の机上物体

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・追跡された皿に対して追跡を行う。本研究では、追跡手法としてテンプレート追跡、深度追跡、SIFT 追跡の 3 手法を用い、これらの中で追跡対象の皿にとってどの手法が最も有効な追跡手法なのかを機械学習手法を用いて自動的に判定することによって追跡精度を向上させる。また、学習手法は C4.5 決定木を用いる。手法の概要を図 5 に示す。まず追跡対象の皿から、追跡手法の決定に有効な特徴量を抽出する。抽出する特徴量は RGB 値や深度値の平均・分散を用いる。抽出した特徴を入力し、手法選択器により追跡に使用する手法を選択、追跡を行う。手法選択器には C4.5 決定木を用いている。すなわち、特徴量から 3 つの追跡手法のいずれかを出力する分類器であるテンプレート追跡は追跡対象の皿の RGB 画像をテンプレートとし、現在時刻の取得画像全体に対してテンプレートマッチングを適用し、最も尤度の高い位置を推定位置とする手法である。テンプレートとして追跡対象の皿の RGB

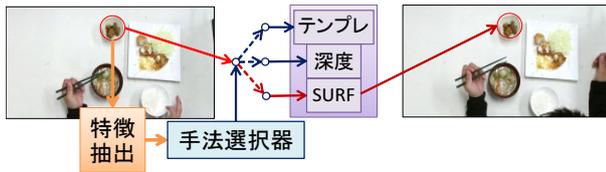


図 5 追跡手法概要

画像全体を用いるため、皿の上を何かに遮られたり、大きく色が変わることがあると追跡に失敗する。本研究では固定した Kinect が撮影した連続画像を用いるため、全体としてフレーム間の皿の色特徴の大きな変化が少なく、結果として 3 手法の中で最も高精度な追跡手法である。一方で色特徴の乏しい料理の載った皿 (例:白米, 餅) 等の追跡には不向きである。



図 6 テンプレート追跡の例

深度追跡では、追跡対象の皿の領域の深度情報から後述する相対深度ベクトルを作成する、そして現フレームの各ウインドウにおいて同様に作成した相対深度ベクトルとの類似度を計算することで追跡を行う。詳細には以下の手順で追跡する。

- (1) 追跡皿の領域内から全体の 1% のピクセルを無作為に選びそのうちの 1 つを基準点 (図 7 中央の黄色の点)、残りを比較点 (図 7 中央の緑色の点) とするサブセットを設定、その後各比較点における基準点との相対深度を計算し、それらを連結することで相対深度ベクトルを作成
- (2) 現フレームにて皿領域と同じ大きさのウインドウを作成、そのウインドウ内にて (1) のサブセットと同様の基準点および比較点を用いてウインドウの相対深度ベクトルを作成
- (3) 全てのウインドウに対して追跡皿の相対深度ベクトルとのユークリッド距離を計算し、最も近いウインドウに皿が移動したとする

図 7 に示す追跡例では、ウインドウ B が最も相対深度ベクトル間の距離が近く、皿の移動先とされた。深度追跡は他の 2 手法と違い色特徴を用いず、深度情報のみを用いて皿の追跡を行う。そのため、色特徴の乏しい皿でも追跡できる。一方で、深度に特徴の見られない大きな平皿や、深度値取得エラーの起きやすい汁物に弱く、また深度情報自体、RGB 画像に比べて取得精度が不安定なため 3 手法の中で最も精度が低い手法である。

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

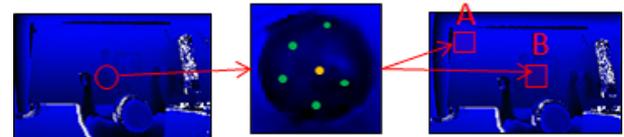


図 7 相対深度ベクトルを用いた追跡の例

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。SIFT 追跡では、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決定する。SIFT 追跡ではテンプレート追跡と同様に、RGB 画像から得た画像特徴量を用いて追跡を行っているが、テンプレート追跡が皿画像全体をテンプレートとし、その尤度によって追跡を行っているのに対し、SIFT 追跡では特徴点同士のマッチングによって追跡を行う為、急に手や頭で皿の一部が遮蔽されたとしても、観測できている一部の特征点同士でマッチングができれば皿の追跡が可能である。そのため、テンプレート追跡や深度追跡では追跡できず見失ってしまう皿も一部追跡可能となる。一方で色勾配を特徴量として記述するため、全体的にのっぺりした色合いの皿の追跡には不向きな手法である。



図 8 SIFT 追跡における特徴点マッチングの例 (赤い円は特徴点、矢印はマッチング先の特徴点を示す)

3.4 手の追跡

3.4.1 手領域の検出・追跡

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。初めに 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。肌色抽出では、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しまう。そのため、深度データを用いて食材の領域を取り除く。検出した手領域は以降、パーティクルフィルタ [13] を用いて追跡する。

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。まず、得られた手・皿の位置から HMM の学習のための特徴量を抽出する。抽出する特徴量は表 1 の 4 つである。左手と皿の距離は、皿を持つ

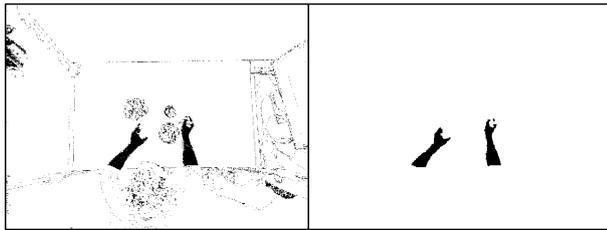


図 9 手領域抽出結果の例 (左:肌色抽出のみ 右:深度データ使用後)

て摂食する場合に有効である。右手と皿の距離は、皿を持たず箸で食品を摂食する場合に有効である。皿と頭の距離は、皿から直接摂食する汁物などの認識に有効である。「手の速度と皿との距離から算出した評価関数」は、手の速度が小さくかつ手と皿との距離が短いときに値が大きくなるように設計した評価関数である。この関数の値は、食品をほぐすなどの動作をする食事に有効である。これら抽出した特徴量から、「摂食」と「非摂食」の2つのHMMを学習する。そして、学習したHMMを用いて摂食の認識を行う。認識を行う際は、皿ごとに2つのHMMセットを用意し、皿ごとに「摂食」と「非摂食」の判定を行う。すなわち用意したHMMセットにより、各時刻ごとにその皿の料理を摂食しているか否かが判定できる。

表 1 HMM の学習に利用する特徴量

左手と皿の距離	右手と皿の距離
皿と頭の距離	手の速度と皿との距離から算出した評価関数の値

4. クラウドソーシング型ラベリング環境の構築

本研究の摂食判定では皿と手の位置情報から特徴抽出を行い、機械学習により摂食判定を行う。また、皿の追跡手法選択器でも、それぞれの追跡手法の追跡結果を用いて、学習を行う。そのため、皿の正解位置をラベリングする必要がある。このように画像を用いた機械学習ではトレーニングデータを作成するために画像に対して人手でラベリングを行う必要があるが、多数の画像に対してラベリング作業を研究者が単独で行う場合、その負担は大きい。そこで今回、Webベースで複数のユーザがラベリング作業を共同で行うことができるクラウドソーシング型ラベリングシステムを構築した。このラベリング環境は、今後実現を予定している皿内部の食材の判定にも利用できる。

4.1 システム概要

本システムの概要を図10に示す。本システムは画像アップロード機能、画像ラベリング機能を備えており、研究者はラベリングを行いたい画像をシステムへとアップロードし、ユーザはウェブ上から任意の画像のラベリングをGUIを用いて容易に行うことができる。またラベリング数によ

るランキングを表示する機能も備えているため、ユーザの功績を一目で確認することができ、研究者はユーザの功績に応じて報酬を与えることができる。以降では本システムの主要な機能である画像アップロー機能、画像ラベリング機能とランキング機能について詳細を述べる。

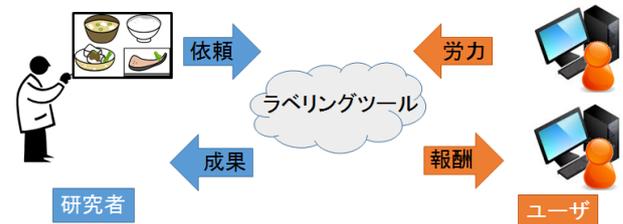


図 10 ラベリングシステム概要

4.2 画像アップロード機能

図11は研究者がラベリングを行いたい画像をアップロードするための画面である。複数の研究者がシステムを利用することを想定しているため、データベース内では画像はプロジェクトに紐づけして管理されている。そこで画像をアップロードする時にはどのプロジェクトにアップロードするかを選択する必要がある。まず既存のプロジェクトにアップロードする場合には画面中央にあるドロップダウンリストから、新規にプロジェクトを作成したい場合にはその右にあるボタンからプロジェクトを作成する。その後ファイルを選択と書いてあるボタンをクリックしファイルを選択する。

またラベリング対象はプロジェクトごとに異なるため、プロジェクトごとにメインラベルとサブラベルを指定することが可能である。たとえば、画像の中から食事かそうでないかを判別したい場合はメインラベルには食べ物という大きな分類を指定し、サブラベルにはハンバーグ、エビフライなど細かい分類を指定する。サブラベルを指定することで、さらに研究を進めたい場合にラベルの貼り直しをせずに研究を行うことができる。



図 11 画像アップロード画面

4.3 ラベリング機能

図12はプロジェクトに参加したユーザが実際に画像に対してラベリングを行う画面である。どの画像にラベリングを行うかを選択する必要があるため、画面上部のドロップダウンリストからプロジェクトを選択し、データベース

からプロジェクトに対応した画像を読み込む。次に画面右のドロップダウンリストから画像を選択し、画面中央に画像を表示させる。その後、画面左の Labeling Menu からラベルの形状を選択する。ラベルの形状は円、矩形、多角形の3種類が選択可能となっている。そしてラベルに添付するメインタグ及びサブタグを選択し、画像上でラベルの形状を調節し、画像に対してラベリングを行う。例えば、図12の食事画像内には複数の食べ物が存在するため、タグ付けを行う場合、それぞれの食べ物に適した形状を選択してラベリングを行う。また、画面右の Label List では、自身もしくは同じプロジェクトに参加している他者がラベリングしたラベルの閲覧と削除を行うことができる。ダウンロードボタンを押すことで、プロジェクトに含まれるすべてのラベルデータをダウンロードすることができる。

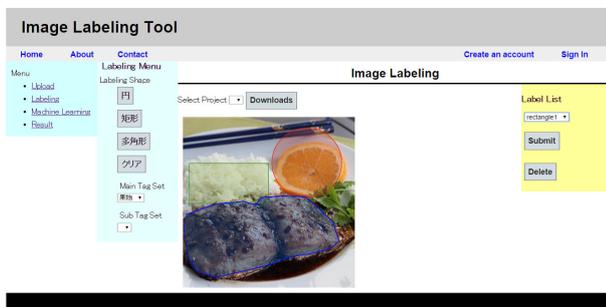


図 12 ラベリング画面

4.4 ランキング機能

図13はラベリング作業量のランキングを表示する画面である。各ユーザに対してラベリング数などのランキングを提示することにより、報酬に対する競争意欲を喚起し、各ユーザのラベリング作業に対するモチベーションの維持及び向上を図る。作業を依頼した研究者はこのランキングを基に各ユーザに報酬を与える。

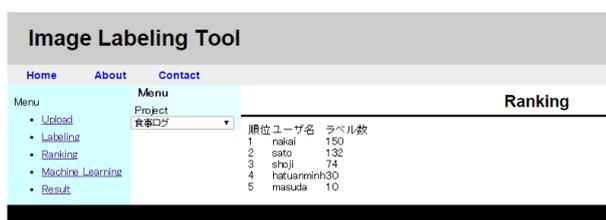


図 13 ランキング画面

5. 評価実験

提案手法の有用性を示すため、皿の追跡および摂食の判定の評価を行った。皿の追跡では、基本となるテンプレート追跡、深度追跡、SIFT 追跡に加え、それら3手法の中か

ら最も追跡対象皿に適した手法を選択する提案手法と比較手法として基本3手法の推定位置の平均位置を手法の推定位置とする平均手法を用い、計5手法での皿の追跡精度の比較を行った。また、提案手法によって基本3手法の中から有効な手法が選択されているかどうかの確認を行った。摂食の判定では、提案手法を用いた摂食の判定と皿と手の位置関係のみを用いた単純な手法との判定制度を比較し、提案手法の評価を行った。以降ではその結果を詳細に示し、結果に対する考察を行う。また今回の実験では、撮影デバイスとして Kinect v2 for Windows を用いた。Kinect は地上約 2.4m、机 上 1.4m の位置に頭上から真下に向けて設置した。

またデータの取得には、大阪大学工学部食堂にて購入した食事を使用し、摂食に使用した皿および箸も、大阪大学工学部食堂にて使用されているものである。データのラベリングには、本研究で構築したラベリングツールを用いた。

5.1 皿の追跡の評価

まず皿の特徴によって有効な追跡手法を学習し、自動選択することによって皿の追跡精度が向上するかの確認を行った。評価実験では Kinect v2 を用いて取得した5セッションのデータ・約 25 分間 (約 2800 フレーム) を全て用いた。

5.1.1 評価方法

評価には、Leave-one session クロスバリデーションを用いた。追跡精度の検証には、前フレームの皿の正解位置を入力とし、正解位置にある皿に対して各手法を用いた追跡を行い、その追跡先が一定閾値以内の誤差なら追跡成功と判定、成功率を追跡精度とし、評価を行った。また比較手法として、テンプレート追跡、深度追跡、SIFT 追跡の3手法の推定位置の平均位置を手法の推定位置とする、平均手法も評価手法に加えた。

5.1.2 実験結果

図14に追跡成功率を示す。提案手法により平均を出力する手法より 4.8%、テンプレート追跡より 1.7% 高い追跡成功率を達成した。これは3つの追跡手法を最適に選択した場合の追跡成功率ある、92.8% とほぼ同等の性能である。

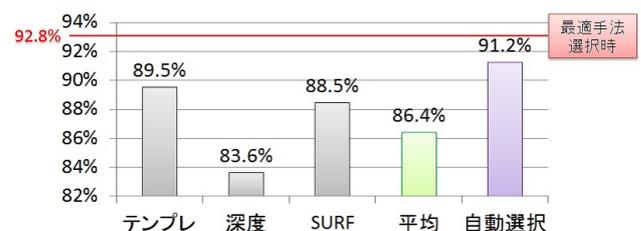


図 14 追跡手法の評価結果

5.2 摂食の判定の評価

次に、提案手法にて追跡した手と皿の位置から摂食の識別に有用な特徴量を抽出し、HMMの学習・認識を行った。評価実験では Kinect v2 を用いて取得した 5 セッションのデータ・約 25 分間 (約 2800 フレーム) を全て用いた。

5.2.1 評価方法

評価には、Leave-one session クロスバリデーションを用いた。比較手法には、手が皿の付近に一定時間以上存在したとき、その皿を摂食したと判定する、単純な手法を用いた。また摂食の正誤判定は、皿と手の追跡により推定した摂食時間と正解の接触時間が一定割合以上オーバーラップしている場合、その摂食は正しく推定されたと判断し、正解とした。この時、推定摂食同士の間隔が短く、同一の正解摂食に対し、2 度正解判定を行わないようにした。摂食判定の評価は、摂食判定の適合率式 (2) と再現率式 (3) に行う。

$$\text{適合率} = \frac{\text{正しく摂食が判定された回数}}{\text{摂食と判定された回数}} \quad (2)$$

$$\text{再現率} = \frac{\text{正しく摂食が判定された回数}}{\text{実際に摂食した回数}} \quad (3)$$

5.2.2 実験結果

図 15 に摂食認識の適合率、再現率および F 値を示す。提案手法は比較手法より高い F 値を実現したが、十分な精度ではなく、今後の改善により向上させる必要があると考える。

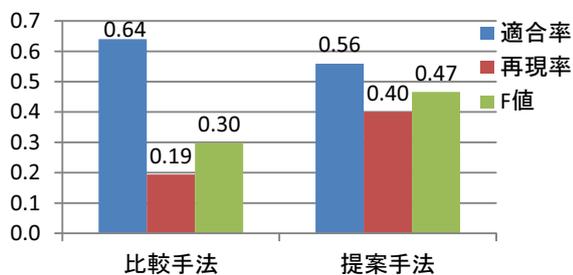


図 15 摂食判定の結果

6. 結論

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。まず、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皿を検出し、検出した皿の画像特徴量を用いて、有効な追跡手法を自動選択することで、高精度な皿追跡を実現した。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。評価実験では、摂食の認識

において、追跡した手と皿の相対位置から特徴量を抽出し、HMM により認識を行うことで、認識精度が向上することを確認した。しかし、まだ十分な認識精度を得られたとは言えず今後の手法改善により認識精度の向上を試みる必要があると考える。

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 *Multimedia (ISM), 2011 IEEE International Symposium on*, pp. 363–368. IEEE, 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—. *The Journal of The Institute of Image Information and Television Engineers*, 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. *糖尿病*, Vol. 55, No. 1, pp. 1–5, 2012.
- [4] 都男金本, 裕井上, 匡守内, 佳枝山田, 久子居村, 眞治佐藤. 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. *糖尿病*, Vol. 53, No. 2, pp. 96–101, feb 2010.
- [5] Paul Brasnett, Lyudmila Mihaylova, David Bull, and Nishan Canagarajah. Sequential monte carlo tracking by fusing multiple cues in video sequences. *Image and Vision Computing*, Vol. 25, No. 8, pp. 1217–1227, 2007.
- [6] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, Vol. 2, pp. 1150–1157. Ieee, 1999.
- [7] 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 CVIM*, Vol. 20, pp. 35–45, 2007.
- [8] Lu Xia, Chia-Chih Chen, and JK Aggarwal. Human detection using depth information by kinect. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pp. 15–22. IEEE, 2011.
- [9] Salih Burak Gokturk and Carlo Tomasi. 3d head tracking based on recognition and interpolation using a time-of-flight depth sensor. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 2, pp. II–211. IEEE, 2004.
- [10] Thad Starner, Joshua Weaver, and Alex Pentland. Real-time american sign language recognition using desk and wearable computer based video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 20, No. 12, pp. 1371–1375, 1998.
- [11] Olga Diamanti and Petros Maragos. Geodesic active regions for segmentation and tracking of human gestures in sign language videos. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pp. 1096–1099. IEEE, 2008.
- [12] Wen-ping Guo and Bang-qiang Long. Improvement of scan line seed fill algorithm [j]. *Journal of Tianjin Polytechnic University*, Vol. 2, p. 013, 2008.
- [13] Arnaud Doucet, Nando De Freitas, and Neil Gordon. *An introduction to sequential Monte Carlo methods*. Springer, 2001.