

# 天井照明に設置したセンサデバイスを用いた 食事画像自動記録システムの提案

佐藤 琢磨<sup>1</sup> 前川 卓也<sup>1</sup>

**概要:** 生活習慣病は現代において解決すべき重要な問題であり、その原因の1つに日々の食生活の乱れが挙げられる。一方スマートフォンの普及により、スマートフォンのカメラで撮影された食事画像を記録し、日々の食生活を管理するオンラインサービスが登場したり、食事画像から食事のカロリーを推定する研究などが行われつつある。しかし、これらのアプリケーションに利用するための食事画像はユーザ自身で用意する必要があり、ユーザが食事画像の撮影を毎日欠かさず継続することは難しい。そこで本研究では、ウェブカメラとモーションセンサを備えた小型センサデバイスを食事が行われる部屋の天井照明に設置することで、食事を最も多く摂るであろう家庭での食事を撮影し、食事ログに適した画像を自動的に取得するシステムを提案する。提案システムではカメラとモーションセンサデータから特徴抽出を行い、機械学習を用いて食事準備が完了したときの画像、すなわち、すべての料理が食卓上に用意されたときの画像を自動的に特定する。これにより、上記のようなアプリケーションに用いられる画像を自動的に記録できる。

**キーワード:** 食事ログ, ライフログ, センサ, デバイス

## Proposing A System for Automatic Recording of Meal Photographs with a Ceiling Light-attached Small Sensor Device

TAKUMA SATO<sup>1</sup> TAKUYA MAEKAWA<sup>1</sup>

**Abstract:** Lifestyle disease is a serious problem that should be solved and the cause of the disease is reported to be irregular dietary life style. Due to the recent proliferation of smartphones, we can easily take the meal photograph with a smart phone and upload it to an on-line service for managing daily dietary life. Also, many researchers have recently tried to estimate calories of meals by using a meal photograph. However, users need to prepare meal photographs by themselves for these applications and it is difficult for them to continue taking meal photographs every day. So this research proposes a system for automatic recording of meal photograph with a ceiling-light attached sensor device. This system is designed to capture meals at home because we eat meal many times in our life. With the device, we try to automatically find a photograph taken immediately after all the food has been placed on the dining table, i.e., just after the meal preparation has been completed by employing machine learning approaches. This system enables us to automatically record meal photographs that are used for the above applications.

**Keywords:** meal log, life log, sensor device

### 1. はじめに

近年、肥満や高血圧といった生活習慣病を患う人は年々

増え続けており、厚生労働省が平成24年に行った国民健康栄養調査 [1] によると、生活習慣病の1つである肥満と診断された人は20代から60代の男女でそれぞれ30%と20%と報告されている。またいずれかの生活習慣病を患う人の割合はおよそ日本国民の半分とも言われており、生活習慣病は現代において解決すべき重要な問題となっている。こ

<sup>1</sup> 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

の原因の1つとして食生活の欧米化により肉類の消費が増加し、野菜類や魚類の消費が減少したことによる偏った食生活があげられる。食生活を改善する方法の1つとして食事の内容を記録し、自身が食べたものを把握する方法があるが、従来はノートなどに手書きで記録するという方法しかなかったため、その煩わしさから継続して記録をすることは困難であった。一方でスマートフォンやデジタルカメラが普及したことにより、手軽に写真を撮影することができるようになった。そのため食事をスマートフォンで撮影することで、食事を画像として記録して食生活の改善を試みようとする人が増えてきた。撮影された画像はTwitterやFacebookといったSNSにアップロードされ、SNSは食事ログの場としても活用されるようになった。このような健康意識の高まりを受け、Evernote Food<sup>\*1</sup>やmiil<sup>\*2</sup>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。また、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やご飯やおかず、野菜のバランスは適切かどうか判定する研究など、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、これらのアプリケーションに用いられる画像はユーザ自身が用意する必要があり、従来よりも手軽になったとはいえ、食事を毎日忘れずに撮影し続けることは難しい。そこで本研究では食事ログを自動的に取得するシステムを提案する。このとき食事の大半は家庭で摂取されることを考慮し、家庭における食事ログ自動取得を試みる。

提案システムに用いる小型デバイスは日々の食事を撮影し続けることを想定しているため、その電源を確保することが重要となる。そこで、日々の食事が行われる食卓の天井照明に注目し、照明から電源の供給を受け、照明が点灯している間に画像を連続的に撮影するようデバイスを設計・実装する。デバイスは天井に設置されるため、食卓を上から撮影できる。デバイスから連続的に撮影された食卓の画像系列は、その他のセンサデータとともにサーバに転送される。そして、サーバは受信した画像とその他のセンサデータから特徴量を抽出し、機械学習を行うことで画像系列の中からアプリケーションの利用に適切な画像、すなわち、その食事のすべての料理が食卓に並んだ瞬間の画像を推定する。推定に用いる特徴量として画像からは色のエントロピーや画像から検出されたエッジの数などを抽出し、その他のセンサデータからは食卓付近で人の動きが検出された回数などの特徴量を抽出する。この特徴量を用いて食事準備完了画像を推定するモデルを設計する。既存研究では食事ログを利用する研究は多く行われているが、食事ログ自体を自動的に取得する研究は少ない。本研究ではユーザの負担にならず、かつ実用的な食事ログ自動取得システムの開発を目指す。

\*1 <http://evernote.com/intl/jp/food/>

\*2 <http://miil.me/>

以降では、2章で関連研究について説明したあと、3章で提案デバイスと食事準備完了画像を推定する手法について説明する。4章で提案手法の評価実験を行い、5章で本稿のまとめを行う。

## 2. 関連研究

### 2.1 食事に含まれる栄養素推定に関する研究

健康管理に必要な栄養素をセンサにより推定する研究について紹介する。橋本ら [2] はまな板に負荷センサと音センサを取り付け、包丁で食材を切るときの力の負荷や切るときの音を計測することで料理に使われた材料の推定を行っている。しかし、料理に含まれる栄養素は調理方法に大きく依存するものであり、材料のみを推定する手法では、最終的に摂取される栄養素は推定できない。北村ら [3] は食事画像をいくつかのブロックに分類し、それぞれのブロックを主食、主菜、副菜などに分割することで食事バランスの推定を行っている。さらにこの研究を発展させたものが宮崎ら [4] によるカロリー推定に関する研究である。この研究では画像からRGBのヒストグラムやHSヒストグラムなど合計4つの特徴量を抽出する。そして、それぞれの特徴量を用いてカロリーを推定するモデルを構築し、単一特徴による推定値を求める。さらに、単一特徴量により求めた複数の推定値から重回帰分析を行い、各推定値に重み付けを行うことで最終的なカロリー値を推定する。しかし、これらの推定に用いる画像はユーザ自らが用意する必要がある上、写真を撮影する角度によって精度が異なるという問題点もある。

### 2.2 健康管理のための食事メニューの提案に関する研究

実際に食べた食事をセンシングする研究ではなく、調理する食事メニューを推薦することで健康管理を行う研究について紹介する。飯塚ら [5] はアレルギーなどにより食事に制限のある人に対し、安全な食事メニューを提案するシステムを提案している。Kuoら [6] はユーザが指定した使いたい材料を含む料理に加えて、その料理に合うその他の料理を提案することで、食事メニュー全体の提案を行っている。また長谷川ら [7] は、ユーザが携帯電話で撮影した食事画像を管理栄養士に送信し、管理栄養士がその栄養素や分量から次の献立を推奨するシステムの提案を行っている。長谷川らは実験のために学生に食事画像の提出を求めたが、3日間の食事データを集めるのに3週間かかった学生も存在した。そこで食事画像の取得部分を自動化することで、容易に健康管理を行うことができるようになる。

### 2.3 ウェアラブルデバイスを用いたライフログに関する研究

ウェアラブルデバイスを用いたライフログ収集に関する研究について紹介する。Arabら [8] はカメラを備えたウェ

アラブルデバイスを食事のときのみ首から下げて、食事ログの取得する方法を提案している。Blumら [9] はカメラだけではなく、マイクや加速度センサも備えたセンサデバイスから、取得したセンサデータを用いて、ユーザの行動や興味をもった瞬間を自動でクリッピングするシステムを提案している。Edisonら [10] は首から下げたスマートフォンを用いて一人称視点の画像を取得し、画像による顔検出やユーザの現在地を用いたフィルタリングを行うことで食事に関係のない画像を削除している。しかし、ウェアラブルデバイスを用いた手法の欠点としてバッテリーの容量に制限があることがあげられる。定期的なバッテリーの充電を必要とするような手法は、ユーザに大きな負担を強いるため、定期的な記録を継続することは難しい。またデバイスを身につけるといふことに対して、負担を感じる人も多い。

上述した研究はユーザによる画像の準備を必要としたり、ライフログを取るためにユーザに負担を強いるなど、アプリケーションの継続的な利用を妨げる要因が存在した。そこで本研究ではアプリケーションの継続した利用を実現するために、食卓の照明に着目した食事ログ取得専用のセンサデバイスを作成することで、ユーザの負担を抑えた上で半永久的に利用できる食事ログ自動取得システムを目指す。

### 3. 提案手法

#### 3.1 システム概要

食事ログに用いられる画像は1回の食事で食べられるすべての料理が写っている必要がある。一人称視点の場合、奥にある料理が手前にある料理に隠れて見えないことが考えられる。そこで画像を取得するデバイスは、食卓全体を俯瞰できる位置にあることが好ましい。また、デバイスが毎日の食事画像を取得し続けることを想定しているため、安定して取得可能な電源が必要である。そこで図1のようにデバイスを天井に設置し電源を照明から確保する。デバイスは食卓を上から撮影できるため、食卓上の全ての料理を取得することができる。



図1 提案デバイスの設置イメージ

本システムは小型デバイスによるデータ取得部とサーバによる画像選択部から構成される。その概要を図2に示す。まず小型デバイスにより時系列の画像とモーション

センサデータを取得する。食事中は人が食卓にいるため、モーションセンサにより人の動きが必ず検出される。そこでモーションデータを用いて人が検知されていない、すなわち、無人のときに撮影された画像を削除する。削除されなかった画像とそれに対応するモーションセンサデータはサーバに送信される。サーバでは受信したデータから、1枚の画像とその画像に対応するモーションセンサデータごとに特徴抽出を行い、特徴ベクトルを作成する。その特徴ベクトルを基に機械学習を行い、準備完了画像を推定する。

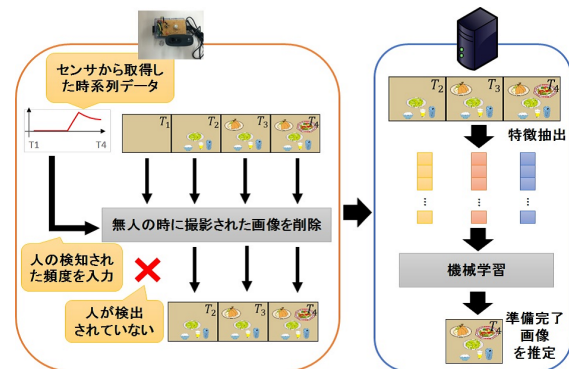


図2 システム概要

#### 3.2 提案デバイス

本研究では小型コンピュータ Raspberry Pi を基に小型センサデバイスを作成した。Raspberry Pi はGPIOポートを備えているため、各種センサを取り付けることが可能であり、このGPIOポートに人の動きを検知するためのモーションセンサを、Raspberry Pi のUSB端子にはウェブカメラを取り付けた。モーションセンサは1秒間に1回サンプリングを行い、動きが検知されたか否かの2値を出力する。ウェブカメラは約10秒から20秒に1回の頻度で写真を撮影する。実際に開発したデバイスを図3に示す。



図3 開発したデバイス



図4 天井に設置されたデバイス

本研究では食事の画像を永続的に撮影する必要があるため、デバイスの電源を外部から確保する必要がある。また、食事ログに適切な画像を取得するためには食卓全体を撮影する必要があること、および、ほぼすべての部屋に照明が備えられていることから、デバイスを天井照明に設置し、その電源を天井照明の電気ソケットから確保することとする。部屋の照明が点灯するとデバイスの電源がオンに

なり、画像とモーションセンサーデータの取得を自動的に開始するように設計した。

### 3.3 特徴抽出

図2に示すように、サーバ上では受信した画像とセンサーデータから特徴抽出を行ったあと、特徴ベクトルを作成し、機械学習により食事準備完了画像の推定を行う。各特徴量について以降で詳細に説明する。

#### 3.3.1 画像に含まれる色の分布

北村ら研究によると、食事の画像には食事独特の色味が含まれると言われている [3]。このことから食事の準備が整うにつれて、画像中に食事独特の色が含まれていくと考えられる。そこで、画像にどのような色がどの程度含まれているかを特徴とする。そのため、まずトレーニングデータから1枚の画像ごとに5000ピクセルをランダムサンプリングし、K-means法を用いて12個のクラスタを作成する。このとき、各クラスタは食事の際によく観測される色に対応する。そして、テストデータの各ピクセルを各クラスタに分類し、各クラスタに分類されるピクセル数を特徴として用いる。このとき、ピクセルは12個のクラスタのうち、もっとも近いものに分類される。このとき、色ピクセル間の距離計算には、通常のRGB空間ではなく、HSV空間を用いる。HSV空間は色相 (Hue)、彩度 (Saturation)、明度 (Value)、の3つの値で色が指定される色空間である。本研究では照明の変化による影響を可能な限り抑えるため、明度に対応するVの値を無視し、色相と彩度の値のみを用いて距離計算を行う。

#### 3.3.2 エッジ検出により求めた画像の複雑さ

食事準備中は、食卓上に乗せられている皿 (料理) の数が少ないため、画像は単純でありその複雑さは小さい。すなわち、画像から検出されるエッジの数は少ない。例えば、図5はテーブルの上に何も置かれておらず、隣り合うピクセル同士の値の変化が小さいため、エッジは検出されない。図6は料理がテーブルの上であり、テーブルと皿の間でピクセル値の変化が大きくなり多くのエッジが検出される。準備が進むにつれて食卓上の皿の数は増えていくため、皿のエッジが数多く検出されるようになる。したがって、エッジを用いて画像の複雑さ [11] を計算することで、食事準備中か食事中かを識別することができると考えられる。エッジを用いた画像の複雑さは式1で計算される。

$$K = \sum_{\substack{0 \leq i \leq X \\ 0 \leq j \leq Y}} |a_{ij}|^2 + 2 \sum_{\substack{0 \leq i \leq X \\ 0 \leq j \leq Y}} |b_{ij}|^2 + \sum_{\substack{0 \leq i \leq X \\ 0 \leq j \leq Y}} |c_{ij}|^2 \quad (1)$$

Xは画像の行の数、Yは画像の列の数である。 $a_{ij}$ は画像をx方向に2回微分した画像のi行目、j列の画素値、 $b_{ij}$ は画像をx方向に1回微分、y方向に1回微分した時の画像のi行目、j列の画素値、 $c_{ij}$ は画像をy方向に2回微分した画像のi行目、j列の画素値である。

#### 3.3.3 画像に含まれる色の分布より求めた画像の複雑さ

食事準備が整うにつれ、数多くの皿が食卓に乗せられる。このとき、画像内には料理の多彩な色が含まれるようになると考えられる。図5は画像全体がテーブルの色で染まっており、色分布に大きく偏りが出るため、色分布の複雑さは小さくなる。それに対し、図6はテーブルの色だけでなく、ご飯の色や主菜の色など画像の中に様々な色が存在しているため、色の分布が分散し複雑さは大きくなる。そこで、画像に含まれる色の偏り (色の分布のエントロピー) から食事準備中か食事中かを識別できると考えられる。エントロピーは上述のK-means法により求めた12個のクラスタに属するピクセル数から計算する。その計算式はクラスタiに属するピクセルの全面素に対する割合を $P_i$ とすると式2で表される。

$$I = \sum_{i=1}^{12} \log \frac{1}{P_i} \quad (2)$$

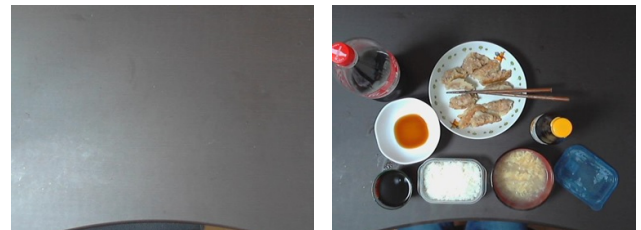


図5 複雑さが小さい画像

図6 複雑さが大きい画像

#### 3.3.4 肌色が検出されたピクセル数

食事の準備が整うにつれてエッジ検出による画像の複雑さ、色の分布から求められる複雑さは増加していく。しかし、これらの特徴は手や顔などの体の一部が画像に含まれた場合に影響を受け、画像の複雑さが増加すると考えられる。そこで画像に含まれる肌色のピクセル数を特徴量に加える事で、体の一部が写っている画像を判別できると考えられる。

#### 3.3.5 画像に含まれる円の数

皿は丸い場合が多いこと、食事準備完了時には料理がすべてテーブルの上に乗っていることから、準備完了画像から検出される円の数最大になると考えられる。すなわち、円の数最大になった画像が準備完了画像に対応する可能性が高いと考えられる。そこで、画像にHough変換 [12] を行うことで画像に含まれる円を検出し、特徴として用いる。図7は、食事画像からHough変換を用いて円を抽出した際の結果である。

#### 3.3.6 動作が検知された頻度

図8はモーションセンサーから得られる時系列データの例である。横軸は時間を表しており、青色のバーはモーションセンサーが動きを検出したタイミングを表している。また、その上部に対応する画像を示している。51番目の画像



図 7 画像から Hough 変換により検出された円

が食事準備完了直後の画像である。図のように、食事の準備中は食卓の上に料理(皿)を運ぶときのみモーションセンサに人の動作が検知されるため、その動作の検知頻度は小さくなり、準備が完了して食事を食べ始めるとテーブル上の料理を摂取するための動きが頻繁に検知されると考えられる。すなわち、検出頻度が高くなり始めた時点で食事準備が完了したと考えられる。そこで、注目する画像とその直前に撮影された画像の間に取得されたモーションセンサデータから、人の動きが検出された頻度を計算し、それを特徴として用いる。

### 3.4 準備完了画像識別手法

準備完了画像の推定を行うためには、準備中から食事中へと状態が遷移する瞬間を識別する必要がある。しかし、センサデータは時系列性を有するため準備中の最終状態と、準備完了時と食事中の最初の状態は近い値を取ると予想される。すなわち、得られるセンサデータは時間とともに徐々にその値が変化すると考えられる。そのためデータの時系列性を考慮したモデリングと準備完了画像の識別性能を高めるモデリングを行う必要がある。そこで本研究では、準備中と食事中を識別する2クラス分類器と系列データを扱うことができる隠れマルコフモデル [13] を組み合わせて準備完了画像を推定する。図9にその手法の流れを示す。提案手法は、まず2クラス分類器により特徴ベクトル系列に含まれる「食事準備中」と「食事中」の2クラスに分類する。そして、2クラス分類器の出力(分類確率の系列データ)を隠れマルコフモデルを用いて認識する。以下に詳細を説明する。

まず、トレーニングデータを準備中と食事中のクラスに分類する2値分類器を学習する。このとき、準備完了画像は食事中のクラスに属するとする。トレーニングデータを用いて分類器の学習を行ったあと、テストデータを分類器入力して食事中のクラス分類される確率を計算する。今回は分類器としてSVMを用いた。SVMではまず準備中と食事中の2クラス分類を行う分類面を学習し、各インスタンスの分類面までの距離を計算し、これを確率の代わりに用いる。この結果を時系列順にソートして確率の時系列データを取得する。

隠れマルコフモデルの学習には、2値分類器が出力した分類確率の系列データを用いる。準備中と食事中の2つのleft-to-right HMMを用意し、準備中HMMの最終状態から食事中の最初の状態に遷移するよう2つのHMMを連結する。HMMの出力は2値分類器が出力した分類確率に対応する。HMMの学習にはEMアルゴリズム [14] を用いる。このHMMでテストデータ(分類確率の系列データ)の認識を行い、準備中から食事中に遷移したときの画像を準備完了画像として選択する。

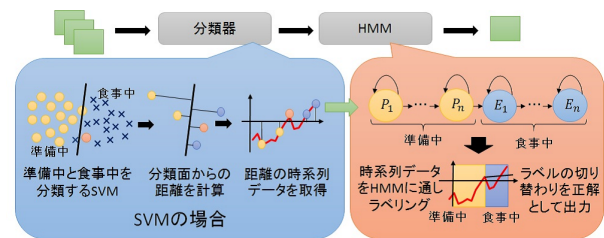


図 9 識別型分類器と HMM を用いた手法の流れ

## 4. 評価実験・考察

### 4.1 データセット

今回の実験では各家庭のダイニングルームの天井照明に提案デバイスを設置してデータの取得を行った。一人暮らしが2世帯、2人で住んでいる世帯が1世帯、3人で住んでいる世帯が1世帯の合計4世帯から朝食、昼食、夕食を含む合計45回の食事を取得した。以下では1回の食事をセッションと呼ぶこととする。1セッションあたりの画像の平均枚数は84枚で、平均正解画像数は約2.3枚であった。表1に各世帯のセッション数、1セッションあたりの平均画像数、1セッションあたりの平均正解数を示す。またデバイスが取得したデータの分析はオフラインで行った。

表 1 各世帯から取得したデータの詳細

	取得したセッション数	1セッションの平均画像数	1セッションの平均正解画像数
家庭 A	10	40	1.6
家庭 B	12	84	2.1
家庭 C	12	88	2.3
家庭 D	11	111	3.2

### 4.2 評価方法

食事画像の内、1回の食事で消費される料理がすべて含まれている画像をすべて正解画像とした。ただし、料理がすべて含まれていても体の一部で料理が隠れている画像や、食事を食べ始めている画像は不正解とした。評価手法として Leave-One-Session-Out 交差検定 (以下 LOS) と Leave-One-Home-Out 交差検定 (以下 LOH) を用いた。

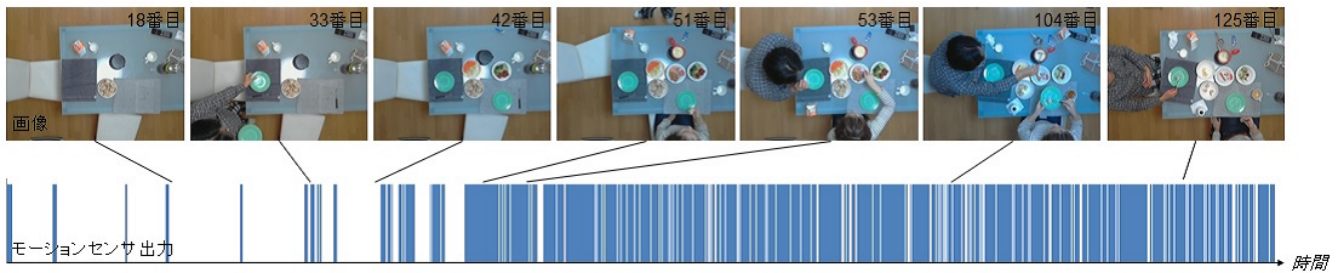


図 8 モーションセンサデータと画像の系列データの例

LOS とは同じ家庭内で 1 つのセッションをテストデータとし、それ以外をトレーニングデータとして交差検定を行う手法であり、LOH とは 1 つの家庭をテストデータ、それ以外の家庭をトレーニングデータとする手法である。評価項目として、

- 正解画像と推定画像との時間のずれ（時間誤差）
- 推定画像の正解率
- 推定画像の前後も含めた計 3 枚に正解画像が含まれる割合

を用いた。ただし、1 つのセッションに正解画像が複数枚存在する場合、そのセッションの時間誤差は誤差の中で一番小さい値のものを採用した。例えば正解画像 A と B が存在し、正解画像 A と推定画像の時間の誤差が 30 秒、正解画像 B との時間の誤差が 40 秒であった場合、そのセッションの時間の誤差は 30 秒となる。また比較手法として以下の 2 つの手法を用いた。

- SVM 手法：正解画像と不正解画像を分類する SVM を学習し、正解画像である確率が一番高い画像を推定画像とする手法。
- HMM 手法：準備中と食事中をの HMM をそれぞれ学習し、準備中の最終状態と食事中の最初の状態を結合した HMM を作成する。そして準備中から食事中へと遷移した時の画像を推定画像として出力する手法。

### 4.3 結果と考察

LOS 交差検定を行ったときの提案手法と比較手法の結果を図 10 に示す。SVM のみを用いた場合は平均時間誤差は 189.45 秒、正解率は 20.00%、HMM のみを用いた場合は平均時間誤差は 112.16 秒、正解率は 17.78% であった。データが時系列性を有するため、SVM では正解画像と不正解画像を分類する面を学習できなかったことで、SVM の精度が HMM に比べて悪くなったと考えられる。また、提案手法の平均時間誤差は 58.16 秒、正解率は 26.67% であった。提案手法はすべての項目において比較手法を上回っており、データの時系列性を考慮しつつ準備完了画像の識別性能を高めたモデリングが行えたと考えられる。また提案手法を用いて推定された画像とそのときの実際の正解画像を図 11 と図 12 に示す。推定画像には米飯が写っていなかった

ため不正解画像となっている。この 2 つの画像の時間のずれは 70 秒であり、提案手法の平均時間誤差である 58 秒と近い値である。また提案手法の正解率は低かったが、正解画像がすべての画像の枚数の 3% しかないことを考慮すると、高い精度で推定が行えたといえる。また、正解画像として推定された画像のうち 78% は図 11 のように、皿が 1 枚足りないか料理を少しだけ食べ始めているもので、限りなく正解画像に近かった。すなわち、正解画像として推定された全ての画像のうち約 78% が正解画像とほぼ同じ画像であった。

表 2 は提案手法を用いた時の世帯ごとの結果を示している。家庭 A, B, D では良い結果が得られた。一方で世帯 C ではその結果が悪かった。世帯 C では食卓のある部屋をリビングとしても利用しており、食事準備の間でもテーブルの周りに人が座っており、常にモーションセンサが検出されてしまったことや、家の中が全体的に赤みを帯びており、人の肌だけでなくカーペットやテーブルも肌色検出で検出されてしまったことが、精度が悪かったことの原因として挙げられる。

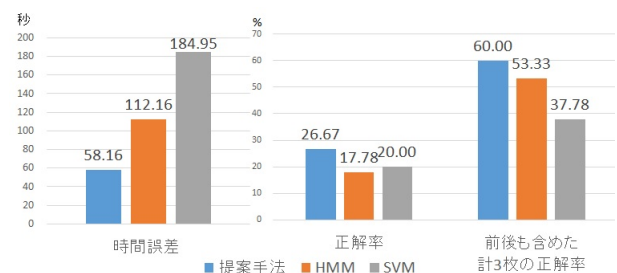


図 10 LOS 交差検定の結果

表 2 LOS 交差検定の結果

	平均時間誤差 [秒]	正解率	前後計 3 枚の正解率
家庭 A	56.10	30% (3/10)	70% (7/10)
家庭 B	33.33	42% (5/12)	67% (8/12)
家庭 C	98.58	17% (2/12)	42% (5/12)
家庭 D	45.09	18% (2/11)	64% (7/11)

次に LOH 交差検定を行ったときの提案手法と比較手法の



図 11 実際に推定された画像

図 12 正解画像

結果を図 13 に示す。このとき、各家庭により食卓の色などが異なるため、色の分布に関する特徴量は取り除いて機械学習を行った。SVM のみを用いた場合は平均時間誤差は 305.98 秒、正解率は 11.11%，HMM のみを用いた場合は平均時間誤差は 144.33 秒、正解率は 15.56%，提案手法の平均時間誤差は 98.35 秒、正解率は 28.89%であった。LOH 交差検定でも提案手法は比較手法を上回った。しかし、LOS 交差検定を行ったときに比べて時間誤差と推定画像の前後 1 枚ずつ計 3 枚に正解画像が含まれる確率は悪化した。これは色の分布に関する特徴量を取り除いたことで、画像に現れた食事の色みを学習できなかったことが原因であると考えられる。また提案手法における家庭ごとの結果を表 3 に示す。家庭 C のみ他の家庭に比べ、精度が大きく悪化している。LOS 交差検定と同じ理由に加えて、家庭 C は 3 人の世帯であり、学習した世帯の中に複数人で住んでいる世帯が家庭 B しかなかったことも原因の 1 つとして考えられる。

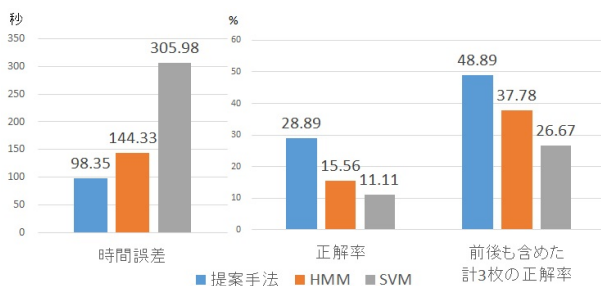


図 13 LOH 交差検定の結果

表 3 LOH 交差検定の結果

	平均時間誤差 [秒]	正解率	前後計 3 枚の正解率
家庭 A	26.30	60% (6/10)	70% (7/10)
家庭 B	90.17	25% (3/12)	58% (7/12)
家庭 C	214.00	8% (1/12)	8% (1/12)
家庭 D	35.00	27% (3/11)	64% (7/11)

## 5. おわりに

本研究では小型センサデバイスを天井の照明に設置することで、食事ログの自動記録を行うシステムを提案した。提案システムでは、食事の準備開始から食事の終わりまで

をカメラで撮影し、得られた複数の画像から食事ログに適切な画像を推定する。食事ログに適した画像を推定する手法を設計するためには、データの時系列性を考慮したモデリングと準備完了画像の識別性能を高めるモデリングが必要とされる。そこで、時系列データの扱いに強い隠れマルコフモデルと識別問題に強い分類器を組み合わせることで、準備完了画像の推定を行った。

評価実験では提案した手法により、約 30% の精度で準備完了画像を推定でき、準備完了画像を推定できることが確認できた。また推定画像の前後も含めた正解率は約 50% であった。しかし、画像に含まれる色の分布や肌色検出などが環境によっては機能しないこと、皿検出において丸い皿以外は検出できないうえに丸い皿でも検出されない場合があると言った問題点が見つかった。

今後は背景差分や塊認識を適用し、これらの特徴量を改良することで更なる精度の向上を目指す。

## 参考文献

- [1] 厚生労働省. 平成 24 年国民健康・栄養調査報告. 平成 26 年, Vol. 5, ; 2014.
- [2] Atsushi Hashimoto, Jin Inoue, Kazuaki Nakamura, Takuya Funatomi, Mayumi Ueda, Yoko Yamakata, and Michihiko Minoh. Recognizing ingredients at cutting process by integrating multimodal features. In *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities*, pp. 13–18. ACM, 2012.
- [3] 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: 映像情報メディア, Vol. 63, No. 3, pp. 376–379, 2009.
- [4] 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.
- [5] Kayo Iizuka, Takuya Okawada, Kouki Matsuyama, Sui Kurihashi, and Yasuki Iizuka. Food menu selection support system: considering constraint conditions for safe dietary life. In *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities*, pp. 53–58. ACM, 2012.
- [6] Fang-Fei Kuo, Cheng-Te Li, Man-Kwan Shan, and Suh-Yin Lee. Intelligent menu planning: recommending set of recipes by ingredients. In *Proceedings of the ACM multimedia 2012 workshop on Multimedia for cooking and eating activities*, pp. 1–6. ACM, 2012.
- [7] 長谷川聡, 吉田友敬, 江上いすず, 横田正, 村上洋子. ケータイ栄養管理システムによる食育と栄養教育. コンピュータ & エデュケーション, Vol. 21, pp. 107–113, 2006.
- [8] Lenore Arab, Deborah Estrin, Donnie H Kim, Jeff Burke, and Jeff Goldman. Feasibility testing of an automated image-capture method to aid dietary recall. *European journal of clinical nutrition*, Vol. 65, No. 10, pp. 1156–1162, 2011.
- [9] Mark Blum, Alex Pentland, and Gerhard Troster. In-sense: Interest-based life logging. *IEEE MultiMedia*, Vol. 13, No. 4, pp. 40–48, 2006.
- [10] Edison Thomaz, Aman Parnami, Jonathan Bidwell, Irfan Essa, and Gregory D Abowd. Technological approaches for addressing privacy concerns when recogniz-

- ing eating behaviors with wearable cameras. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pp. 739–748. ACM, 2013.
- [11] Robert Rich. Image contrast, complexity, and stability. *Computer vision, graphics, and image processing*, Vol. 26, No. 3, pp. 394–399, 1984.
- [12] Dana H Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, Vol. 13, No. 2, pp. 111–122, 1981.
- [13] Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in hidden Markov models*, Vol. 6. Springer, 2005.
- [14] Arthur P Dempster, Nan M Laird, Donald B Rubin, et al. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal statistical Society*, Vol. 39, No. 1, pp. 1–38, 1977.