Vol.2017-CVIM-205 No.15
2017/1/19

食生活の振り返り支援のための摂食場面の抽出

松浦 亮太[1,a)]　築地原 里樹[1]　GARCIA RICARDEZ Gustavo Alfonso[1]
丁 明[1]　高松 淳[1]　小笠原 司[1]

概要：健康管理や生活の質の向上のために，生活を記録，評価し，改善することは重要である．一方，日常生活を記録した長時間の映像から，ユーザにとって有益な情報を効率的に得るためには，要約作業が必要である．本研究では，日常生活の中でも健康との関係が深い「食事」に着目し，食事中の映像から摂食場面のみを抽出するシステムを提案する．提案手法では，不特定多数の摂食場面を認識するのではなく，初回の摂食場面の映像のみから教師データを作成し，その回の食事に特化した認識器を構築する方法を選択する．これにより，教師データを準備する手間を軽減できる．本システムを用いることで，特定の被験者に対して摂食場面の抽出ができることを示す．

Extraction of Eating Scenes to Support the Review of Eating Habits

Matsuura Ryota[1,a)]　Tsuichihara Satoki[1]　Garcia Ricardez Gustavo Alfonso[1]
Ding Ming[1]　Takamatsu Jun[1]　Ogasawara Tsukasa[1]

1. はじめに

健康管理や生活の質の向上のために，生活を記録，評価し，改善することは重要である．これまで，人手により調査票を記入し生活を記録する方法 [1] や，スマートフォンを用いて食べ物の写真を撮影して食事内容を記録する方法 [2]，箸をセンサ化して食事動作を検出する方法 [3] などが提案されている．これらの手法では，生活時間や食事の内容，タイミングなどの特定の情報を抽出し記録することに成功している．

一方，生活を詳細に振り返るためには，人の行動や周囲の状況などを映像で記録し，解析することが有効である．映像からユーザにとって有益な情報を得るためには，特定の行動が含まれる場面を抽出する要約作業が必要である．特に，発生する頻度が少ない行動を要約なしに振り返るためには，映像の中から対象とする行動をユーザ自身で抽出する必要があり，負担が大きい．

映像要約を目的として特定の場面を抽出する手法として，音声や脈拍，位置情報などを利用する方法 [4] が提案されている．また，料理番組の要約に限定されるものの，あらかじめ与えられた料理レシピと画像から推定した人の調理動作を照合する方法 [5] なども提案されている．

映像要約のための手掛かりとなる人の行動や姿勢を認識する手法として，機械学習を用いた汎用的な手法 [6], [7], [8], [9] が提案されている．これらの手法は多くの教師データを用いて機械学習を行う．しかし，本研究で注目する発生頻度の低い行動を扱う場合，教師データそのものを集めることが困難である場合もある．

本研究では，日常生活の中でも健康との関係が深い「食事」に着目し，食事中の映像から摂食場面のみを抽出するシステムを提案する．本稿では，食べ物を箸などの道具を使って口に入れる動作を摂食動作，それ以外の動作を非摂食動作と定義し，観測された動作が摂食動作であるか判定することを摂食判定と定義する．また，入力する画像はカメラの正面に向かって着座している人が，中央付近に写るように撮影したものとする．提案手法では，教師データを準備する手間を軽減するために，初回の摂食場面の映像のみから教師データを作成し，今回撮影された食事映像に特化した認識器を構築する．認識器の適用範囲を限定するこ

[1] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Ikoma, Nara 630–0192, Japan
[a)] matsura.ryota.mm2@is.naist.jp

(a) Pick up (b) Carry (c) Eat

図 1 摂食動作の段階
Fig. 1 Eating steps

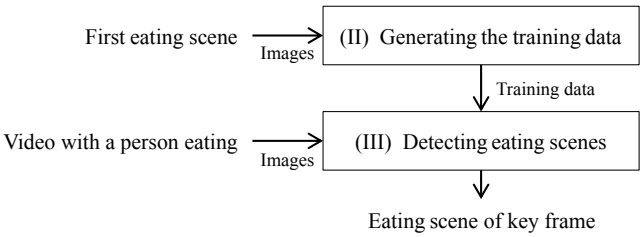


図 2 摂食判定の流れ
Fig. 2 Process of eating scenes detection

とで，少ない教師データからでも解析を可能にすることを目指す．実験では，特定の被験者の食事場面に対して摂食判定を行い，55 回の摂食動作のうち 41 回の摂食動作において摂食と判定されたことを確認し，全ての摂食場面の約 75%の摂食場面を抽出できることを確認した．

2. 提案手法

2.1 摂食判定の概要

摂食時に人がとる行動は以下の動作段階に分解できる．

(1) 箸などを使って食べ物を掴む（図 **1**(a)）
(2) 食べ物を口元に移動する（図 **1**(b)）
(3) 食べ物を口の中に入れる（図 **1**(c)）

本研究ではある特定個人のある食事内容のみを対象とするため，動作パターンおよび動作特徴のバリエーションは多くないと考えられる．

図 **2** に示すように，食事の開始から終了まで撮影された映像に対しての摂食判定の流れは以下の通りである．

手順 I ：基準摂食場面として，記録した映像の中から 1 回の摂食動作を選択
手順 II ：基準摂食場面から教師データを取得（2.3 節）
手順 III：得られた教師データをもとに，それ以外の映像に対して摂食判定（2.4 節）

ここで，手順 I は人手で教師データを取得するための映像を切り出す処理とする．

2.2 フレームごとの動作段階の判定

摂食判定を行うために，各フレームにおいて図 1 に示すどの動作にあるかを判定することから始める．動作段階を判定するために以下の処理を行う．

(1) 単一フレームにおける手と頭の位置を検出
(2) 与えられた手と頭の位置から，そのフレームが摂食動作のどの段階にあるかを判定

まず，(1) の単一フレームにおける手と頭の位置の検出器を確率関数で記述する．図 1 に示すように手と頭の見えは摂食動作の前後で刻一刻と変わっていくため，それらをいくつかの代表的なテンプレートで表現する．画像 I が与えられたとき，あるテンプレート r に対して手と頭の位置を $\boldsymbol{a}$，$\boldsymbol{b}$ と推定する確率は $p(\boldsymbol{a},\boldsymbol{b}|I,r)$ で表されるとすると，手と頭の検出位置は式 (1) で表される．

$$(\boldsymbol{a},\boldsymbol{b}) = \mathrm{argmax}_{\boldsymbol{a},\boldsymbol{b}} \mathrm{max}_r \left(p(\boldsymbol{a},\boldsymbol{b}|I,r)\right) \qquad (1)$$

確率 $p(\boldsymbol{a},\boldsymbol{b}|I,r)$ は式 (2) のように定義する．

$$p(\boldsymbol{a},\boldsymbol{b}|I,r) = p(I|\boldsymbol{a},r)p(I|\boldsymbol{b},r)p(\boldsymbol{a},\boldsymbol{b}|r) \qquad (2)$$

$p(I|\boldsymbol{a},r)$，$p(I|\boldsymbol{b},r)$ は手と頭の位置がそれぞれ $\boldsymbol{a}$，$\boldsymbol{b}$，テンプレートが r であるときの画像 I のもっともらしさである．また，$p(\boldsymbol{a},\boldsymbol{b}|r)$ は，テンプレート r が与えられたときの，手と頭の位置関係 $\boldsymbol{a}$，$\boldsymbol{b}$ のもっともらしさである．

次に，(2) の手と頭の位置が与えられた時に，動作段階を判定する判定器を確率関数で記述する．式 (3) で示すように，検出された手と頭の位置 $\boldsymbol{a}$，$\boldsymbol{b}$ が与えられた時に，動作段階 d である確率 $p(d|\boldsymbol{a},\boldsymbol{b})$ が最も高くなる d を選択することとする．

$$d = \mathrm{argmax}_d \left(p(d|\boldsymbol{a},\boldsymbol{b})\right) \qquad (3)$$

2.3 教師データの取得

摂食判定のための教師データとして以下の二種類のデータを仮定する．

- 動作中の代表的な手と頭の見えおよび位置関係
- 各動作段階の手と頭の位置関係

教師データを取得する作業では，ユーザに対する負荷が少ないことが望ましい．画像中のある特定領域を選択するといった作業に比べて，提示された情報に対してその情報が正しいか否かを判定することは，簡単な作業であると考える．そこで，本研究では領域選択の作業を極力させずに教師データを取得することを考える．

2.3.1 動作中の代表的な手と頭の見え

初回の摂食場面に対応する画像セット $\boldsymbol{I}_{\mathrm{first}} = \{I_1,\cdots,I_N\}$ が与えられた時，以下の手順で手と頭の見えを取得する．

(1) 人が $\boldsymbol{I}_{\mathrm{first}}$ から適当な画像 I_i を選択し，人が手と頭の領域を選択して，手と頭のテンプレート画像 I_{a_j}，I_{b_j} と，その時の手と頭の位置の差分ベクトル $\boldsymbol{v}_j$ を取得する．これを，教師データ $r_j = \left\{I_{a_j}, I_{b_j}, \boldsymbol{v}_j\right\}$ とする．
(2) 確率モデルに従い，$\boldsymbol{I}_{\mathrm{first}}$ に含まれる他の画像の手と頭の位置を検出する．位置が検出できた画像 I_i を $\boldsymbol{I}_{\mathrm{first}}$

(a) Step 1 (b) Step 2 (c) Step 3 (d) Step 4 (e) Step 5

図 3 各摂食動作段階のサンプル
Fig. 3 Examples of the eating steps

表 1 摂食動作段階の定義
Table 1 Definition of eating steps

動作段階	定義	選択する画像	画像例
1	皿の上で箸を使って食べ物を把持している	箸先が皿から離れ始める直前	図 3(a)
2	食べ物を箸で把持したまま口に近づけている	箸先が皿から離れて口に触れるまでの中間	図 3(b)
3	食べ物を口に入れている	最も箸先を口の中に挿入している瞬間	図 3(c)
4	手を摂食前の状態まで戻すために移動している	箸先が口に触れている状態から次の動作に移るまでの中間	図 3(d)
5	手が摂食前の状態に戻っている	次の動作に移る直前	図 3(e)

から取り除く．確率 $p(\boldsymbol{a},\boldsymbol{b}|I,r_j)$ が閾値 T_{init} を超える $\boldsymbol{a}$，$\boldsymbol{b}$ が存在する画像 I は検出ができた画像とみなす．

(3) $\boldsymbol{I}_{\text{first}}$ が空集合 ϕ になるまで (1), (2) の処理を繰り返す．

(2) の手と頭の位置の検出のために判定モデルを用いる．確率 $p(\boldsymbol{a},\boldsymbol{b}|I,r)$ の各項は，式 (4)〜(6) のように定義する．ただし，s_a は $\boldsymbol{a}$ のテンプレートマッチングの類似度（正規化差分相関で計算するため小さいほど良い），T_v^{init} はテンプレート作成時のベクトルのマッチングの閾値を表す定数とする．

$$p(I|\boldsymbol{a},r) = \exp(-s_a) \tag{4}$$

$$p(I|\boldsymbol{b},r) = \begin{cases} 1 & \left(\boldsymbol{b}=\hat{\boldsymbol{b}}\right) \\ 0 & (\text{otherwise}) \end{cases} \tag{5}$$

$$p(\boldsymbol{a},\boldsymbol{b}|r) = \begin{cases} 1 & \left(|(\boldsymbol{a}-\boldsymbol{b})-\boldsymbol{v}_j| < T_v^{\text{init}}\right) \\ 0 & (\text{otherwise}) \end{cases} \tag{6}$$

ここで，式 (5) の確率 $p(I|\boldsymbol{b},r)$ はテンプレート r によって検出された頭の位置 $\hat{\boldsymbol{b}}$ のもっともらしさである．$p(I|\boldsymbol{b},r)$ はテンプレート r を用いた頭のテンプレートマッチングで記述でき，頭の位置の検出結果 $\hat{\boldsymbol{b}}$ が正確である場合には 1，誤検出が生じた場合には 0 と定義する．また，式 (6) の確率 $p(\boldsymbol{a},\boldsymbol{b}|r)$ はテンプレート r が与えられた時の手と頭の位置関係のもっともらしさである．$p(\boldsymbol{a},\boldsymbol{b}|r)$ は，手と頭の位置関係を表すベクトル $(\boldsymbol{a}-\boldsymbol{b})$ と，テンプレート r に含まれるベクトル $\boldsymbol{v}_j$ とのマッチングで記述でき，ベクトルの差分の大きさが閾値 T_v^{init} よりも小さい場合には 1，大きい場合には 0 と定義する．

2.3.2 各動作段階の代表画像における手と頭の位置関係

画像セット $\boldsymbol{I}_{\text{first}}$ に対し手と頭の位置を検出した結果を人に提示し，摂食における代表的な 5 つの状態を表す画像を選択させる．ただし，人には手，頭の位置が正しく検出されていない結果を選択しないように指示する．

2.3.1 項の結果を用いて，最終的な手と頭の位置 $\hat{\boldsymbol{a}}$，$\hat{\boldsymbol{b}}$ は，すべてのテンプレートに対して確率 $p(\boldsymbol{a},\boldsymbol{b}|I,r)$ が最大となる $\boldsymbol{a}$，$\boldsymbol{b}$ によって決定する．検出された手と頭の位置を人に提示して，動作段階 d を決定する．これにより，動作段階 d における手と頭の位置関係を表すベクトル $\boldsymbol{u}_d$ が得られる．**表 1** に各動作段階の定義を，**図 3** に選択する画像の例を示す．

2.4 教師データを用いた摂食判定

摂食判定は残りすべてのフレームに対し，手と頭の位置の検出と，動作の段階の判定を繰り返すことで行う（**図 4** 参照）．摂食判定の流れは以下の通りである．

(1) 教師データを使って，時刻 t における画像 I_t に対して $\boldsymbol{a}$，$\boldsymbol{b}$ の位置を検出し，確率 $p(d|\boldsymbol{a},\boldsymbol{b})$ が最大となる動作 d を選択

(2) 時刻 t と $t-1$ で判定された動作段階 d_t，d_{t-1} から摂食したかどうかを判定

教師データを用いた判定において，式 (2) の $p(\boldsymbol{a},\boldsymbol{b}|r)$ は，式 (7) のように分解される．

$$p(\boldsymbol{a},\boldsymbol{b}|r) \propto p(\boldsymbol{b}|\boldsymbol{a},r)p(\boldsymbol{a}) \tag{7}$$

式 (2) の確率モデルの各項を式 (8)〜(11) のように定義して推定する．ただし，位置 $\boldsymbol{a}$ の画像中での横方向の位置を a_x，画像の幅を l，テンプレートマッチングの類似度の閾値を T_{s_a}，ベクトルの類似度の閾値を T_v とする．

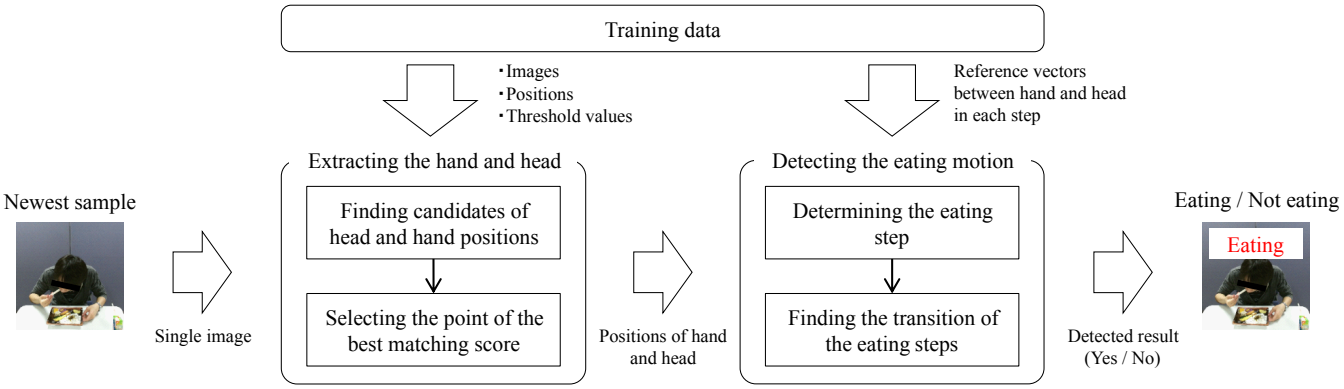


図 4　教師データに基づく摂食判定の流れ

Fig. 4　Process of eating scenes detection based on training data

$$p(I|\boldsymbol{a},r) = \exp(-s_a) \tag{8}$$

$$p(I|\boldsymbol{b},r) = \begin{cases} 1 & \left(\boldsymbol{b}=\hat{\boldsymbol{b}}\right) \\ 0 & (\text{otherwise}) \end{cases} \tag{9}$$

$$p(\boldsymbol{b}|\boldsymbol{a},r) = \begin{cases} 1 & (|(\boldsymbol{a}-\boldsymbol{b})-\boldsymbol{v}_j| < T_v) \\ 0 & (\text{otherwise}) \end{cases} \tag{10}$$

$$p(\boldsymbol{a}) = \exp\left(-T_{s_a}\frac{|2a_x - l|}{l}\right) \tag{11}$$

手と頭の位置の検出結果 $\hat{\boldsymbol{a}}$，$\hat{\boldsymbol{b}}$ は式 (12) で表される．

$$\left(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}}\right) = \text{argmax}_{\boldsymbol{a},\boldsymbol{b}}\text{max}_r\left(p(\boldsymbol{a},\boldsymbol{b}|I,r)\right) \tag{12}$$

式 (3) の $p(d|\boldsymbol{a},\boldsymbol{b})$ を式 (13) と定義して，動作段階 d を判定する．

$$p(d|\boldsymbol{a},\boldsymbol{b}) = \exp\left(-|(\boldsymbol{a}-\boldsymbol{b})-\boldsymbol{u}_d|\right) \tag{13}$$

ただし，確率 $p(d|\boldsymbol{a},\boldsymbol{b})$ が手と頭の位置関係の類似度を表す閾値 T_u を超えない場合には，動作段階 d は割り当てない．

ここで，時刻 t における動作段階 d_t は摂食行動の前後で対称性があるため，$\{d_t\}$ に対して $d_{t-1}=2$（動作段階 2 に対応．以下同様），$d_t=3$ または，$d_{t-1}=4$，$d_t=3$ となる t を摂食状態として選択する．

3. 実装

摂食判定の正確性を向上するために，以下に示す処理を導入した．

- 手と頭の誤検出の除外
- 動作段階の判定基準の拡張

3.1 手と頭の誤検出の除外

明らかな手と頭の位置の誤検出を防止するために，以下の処理を適用する．

(1) あらかじめ用意した背景画像を用いて背景領域を除外するマスクを設定し，人以外の領域に該当する検出結果を除外

(2) 肌色領域のみを抽出するマスクを設定し，手や頭の領域以外の検出結果を除外

(3) 前のフレームからの移動量が明らかに大きい位置検出結果を除外

(1)〜(3) の処理のために判定モデルを確率関数で記述する．判定モデルは，画像中の手や頭の位置を $\boldsymbol{q}$，入力画像 I の位置 $\boldsymbol{q}$ における画素値を $I(\boldsymbol{q})$，背景画像を I_{bg}，肌色を表す画素値を c，手や頭が最後に観測された位置を $\boldsymbol{q}_{\text{last}}$ とすると確率 $p(\boldsymbol{q}|I,I_{\text{bg}},c,\boldsymbol{q}_{\text{last}})$ は式 (14) で定義する．

$$p(\boldsymbol{q}|I,I_{\text{bg}},c,\boldsymbol{q}_{\text{last}}) = p(\boldsymbol{q}|I,I_{\text{bg}})p(\boldsymbol{q}|I,c)p(\boldsymbol{q}|\boldsymbol{q}_{\text{last}}) \tag{14}$$

ここで，検出された手と頭の位置 $\boldsymbol{a}$，$\boldsymbol{b}$ における確率 $p(\boldsymbol{a}|I,I_{\text{bg}},c,\boldsymbol{a}_{\text{last}})$，$p(\boldsymbol{b}|I,I_{\text{bg}},c,\boldsymbol{b}_{\text{last}})$ が 0 となる位置は誤検出であるとする．式 (14) の各項は，背景差分法の画素値の閾値 T_{bg}，肌色抽出の画素値の閾値 T_{sk}，フレーム間での手や頭の移動量を表す閾値 T_{mv} を用いて，式 (15)〜(17) で定義する．

$$p(\boldsymbol{q}|I,I_{\text{bg}}) = \begin{cases} 1 & (|I(\boldsymbol{q}) - I_{\text{bg}}(\boldsymbol{q})| < T_{\text{bg}}) \\ 0 & (\text{otherwise}) \end{cases} \tag{15}$$

$$p(\boldsymbol{q}|I,c) = \begin{cases} 1 & (|I(\boldsymbol{q}) - c| < T_{\text{sk}}) \\ 0 & (\text{otherwise}) \end{cases} \tag{16}$$

$$p(\boldsymbol{q}|\boldsymbol{q}_{\text{last}}) = \begin{cases} 1 & (|\boldsymbol{q} - \boldsymbol{q}_{\text{last}}| < T_{\text{mv}}) \\ 0 & (\text{otherwise}) \end{cases} \tag{17}$$

閾値 T_{mv} は処理画像のサイズやフレームレートなどから想定される手と頭の 1 フレームあたりの最大の移動量を用いて決定する．さらに，トラッキングの失敗が発生した際に $\boldsymbol{q}_{\text{last}}$ からの再トラッキングを可能にするために，T_{mv} の値を定数倍して定義する．

3.2 動作段階の判定基準の拡張

2.4 節で述べた時刻 t での動作段階 d_t の変化による摂食判定では時刻 t，$t-1$ に基づいた摂食判定をしている．しかし，実際には 1 回の摂食場面は複数のフレームで構成さ

れており，同一の動作段階を継続する．具体的には，1 回の摂食場面において，箸を口に挿入している動作は連続して複数フレーム存在するため，その区間全体を摂食状態として判定する必要がある．ここで，動作段階 2 から動作段階 3 への遷移，動作段階 4 から動作段階 3 への遷移が摂食時であるとする．本手法では以下の処理を行う．

(1) 時刻 t における動作段階を判定
  (a) 時刻 t での動作段階 d_t が 3 である場合，振り返るフレーム数 n を 1 として (2) へ遷移
  (b) (a) 以外の条件である場合，時刻 t は摂食状態ではないと判定

(2) 時刻 $t-n$ における動作段階を判定
  (a) 時刻 $t-n$ での動作段階 d_{t-n} が 2 か 4 である場合，時刻 t は摂食状態であると判定
  (b) 時刻 $t-n$ での動作段階 d_{t-n} が 3 である場合もしくは，動作段階が割り当てられていない場合，n を $n-1$ に変更し (2) へ再度遷移
  (c) n が決められた最大回数 N に達した場合，その時刻 t は摂食状態ではないと判定
  (d) (a)〜(c) 以外の条件である場合，その時刻 t は摂食状態ではないと判定

ここで，最大回数 N は 1 回の摂食動作の長さとして想定されるフレーム数とする．この値は，明らかに摂食状況ではない長さの摂食動作に似た動作が検出されることを防ぐために設定する．

4. 実験

摂食判定手法の評価のために，実際の食事場面で撮影した映像に対して摂食判定を行った．

4.1 実験条件

実験では，**図 5** のような環境で被験者が一般的な弁当を食べる場面で摂食判定した．カメラには Kinect v2 [10] を使用し，RGB 画像機能を使用して 30 fps の映像を撮影した．処理の軽減のために被験者が写った 200 × 200 pixel の画像領域を切り出した．基準摂食場面としては，人が選択した 50 フレームで構成された摂食映像を用いた．

4.2 教師データの作成

初回の摂食映像に，2.3 節で述べた手法を適用し教師データを作成した．結果として，7 セットの手と頭のテンプレート画像群が作成された．作成されたテンプレート画像群と，各テンプレート画像の切り出し位置とテンプレート同士の位置関係を表す画像群を**図 6** に，各摂食動作段階の分類を図 3 に示す．ただし，図 6 上段の緑色矩形領域は手部のテンプレート画像 I_{a_j} を抽出した位置，赤色矩形領域は頭部のテンプレート画像 I_{b_j} を抽出した位置，水色線分は手と頭の位置関係 $\boldsymbol{v}_j$ を表す．

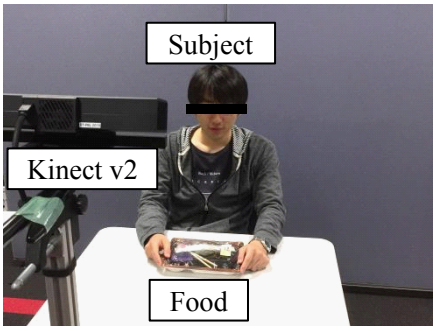


図 5 実験環境
Fig. 5 Experimental setup

表 2 摂食判定結果 (単位：フレーム)
Table 2 Result of detecting the eating scenes (Unit: frames)

		Ground truth	
		Eating	Not eating
Results	Eating	345	245
	Not eating	215	12618

4.3 実験結果

図 7 に初回摂食動作以降の摂食タイミングと本手法によって推定された摂食タイミングを示す．図 7 中，赤色区間は実際に摂食したタイミング，緑色区間は本手法によって摂食と判定されたタイミングである．なお，赤色区間で示した摂食タイミングは，人手により判定した．

動作単位での摂食判定の精度を確認するために，実際に被験者が摂食した動作の中で，本手法により摂食と判定された動作の回数を評価する．初回の摂食動作以降の摂食回数の真値は 55 回であった．この中で 41 回の摂食動作において，本手法により摂食状態と判定されたフレームが確認できた．摂食動作の約 75%を抽出できた．

各フレームの摂食判定の精度を確認するために，フレーム単位での摂食判定回数を評価する．**表 2** に，フレーム単位での摂食判定の結果を示す．評価基準として，適合率 (Precision) と再現率 (Recall) を用いる．ここではそれぞれを以下のように定義する．

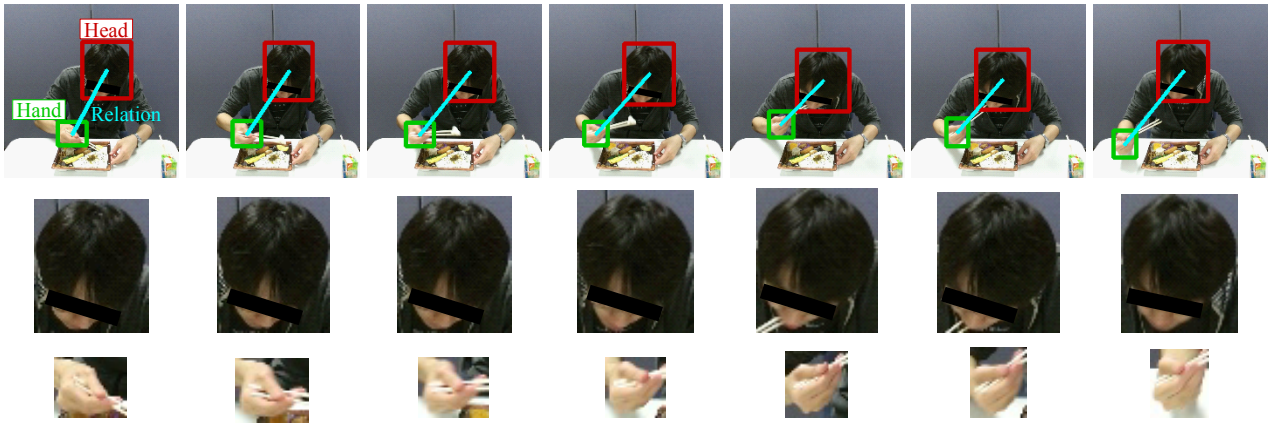
適合率 (Precision)：
  本手法によって摂食状態と判定されたフレームの中で，実際に摂食状態であったフレームが含まれる割合

再現率 (Recall)：
  実際に摂食状態であったフレームの中で，本手法により正しく摂食状態と判定されたフレームの割合

また，本手法を用いた摂食判定の F 値は式 (18) で計算する．

$$F = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{18}$$

表 2 ならびに式 (18) より，適合率は 0.58，再現率は 0.62，F 値は 0.6 であった．



(a) Template 1 (b) Template 2 (c) Template 3 (d) Template 4 (e) Template 5 (f) Template 6 (g) Template 7

図 6 テンプレート画像群 (上段：手と頭の位置関係，中段：頭部画像，下段：手部画像)
Fig. 6 Template images (Top: head-hand relation, Middle: head, Bottom: hand)

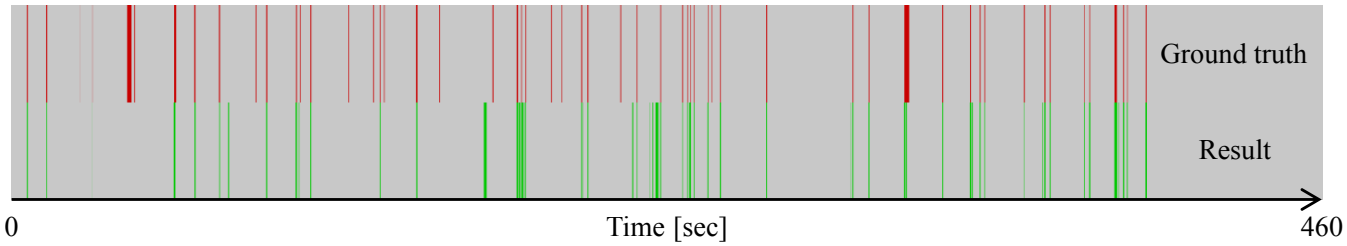


図 7 摂食タイミング (赤：真値，緑：摂食判定結果)
Fig. 7 Eating timing (Red: ground truth, Green: detection result)

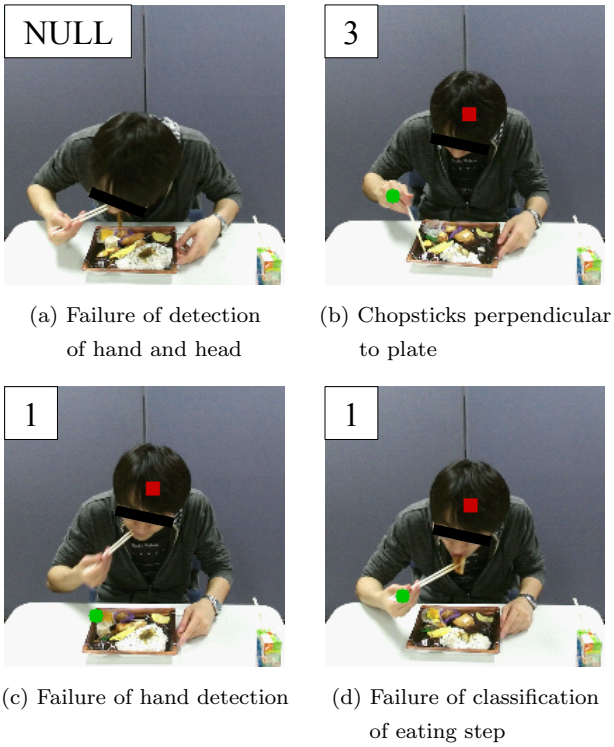
4.4 考察

図 7 より，摂食判定の結果を食事場面全体から大域的に見ると，人が実際に摂食しているタイミング（図中の赤色区間）と，本手法により摂食と判定されたタイミング（図中の緑色区間）がほぼ一致することが分かる．また，摂食動作単位での摂食判定の精度は約 75%，フレーム単位での摂食判定の精度として得られる再現率は約 62%であることから，摂食動作の区間をフレーム単位で推定する精度よりも，摂食動作が行われた回数を検出する精度の方が，良好な結果が得られることが分かる．このことから，本手法によって摂食動作を抽出するための手掛かりを検出することができるため，検出された摂食画像とその前後の数フレームの画像を抽出し，映像として再構成することによって映像要約が可能になると考えられる．

以下に，摂食判定に失敗した際の処理画像（**図 8**）とその原因，対策方法をまとめる．各画像中の赤色の四角点は頭の検出結果，緑色の丸点は手の検出結果，画像左上の数字は分類された動作段階を表す．

手と頭の検出の失敗（図 8(a)）

テンプレートマッチングにより手と頭を検出した際に，マッチングするテンプレートが存在しなかった．これにより，摂食判定に必要な手と頭の位置情報が得られなかった．この問題は，テンプレート群を増やすことや，閾値の緩和によって改善できると考えられる．



(a) Failure of detection of hand and head
(b) Chopsticks perpendicular to plate
(c) Failure of hand detection
(d) Failure of classification of eating step

図 8 摂食判定の失敗例（画像左上の数字は動作段階を表す）
Fig. 8 Examples of failure of eating detection. Eating step is shown in the upper left.

食べ物に対して箸を立てた状態（図 8(b)）

食べ物の切り分けや迷い箸の動作で，被験者が食べ物に対して箸を立てた状態の姿勢をとった．これにより，手と頭が接近し，誤って動作段階 3 と判定され，誤検出が発生した．この問題は，手のテンプレートに対して箸の姿勢を加えることが有効であると考えられる．

手の誤検出（図 8(c)）

手の誤検出が発生し，食べ物が手として検出された．テンプレートマッチングの評価関数として正規化差分相関を用いているが，その他の評価関数を試す必要がある．

動作段階の分類誤り（図 8(d)）

手と頭の位置検出の誤差により，動作段階の判定に誤りが生じることがある．手と頭の検出位置の移動平均を計算することにより，突発的な検出誤差を平滑化することが有効であると考えられる．

5. おわりに

本研究では，日常生活を撮影した映像から発生頻度の低い行動を抽出することを目的として，食事中の映像に対して摂食場面の抽出を行った．本稿では，不特定多数の食事場面に対応した摂食判定ではなく，その回の食事に特化した認識器を構築する摂食判定手法を提案した．実験では，特定の被験者の食事場面に対して摂食判定を行い，55 回の摂食動作のうち 41 回の摂食動作において摂食と判定されたことを確認し，全ての摂食場面の約 75%の摂食場面を抽出できることを確認した．これにより，初回の摂食場面の映像のみから作成した教師データに基づく，摂食判定が可能になり，教師データを準備する手間を軽減できる．

今後は，摂食判定精度を向上するために，手と頭の検出精度の向上や，箸などの道具との関係を考慮した判定手法の開発，確率モデルの再設計，誤検出へのロバスト性を高めるための処理の導入が課題である．

参考文献

[1] 総務省統計局：統計局ホームページ/平成 23 年社会生活基本調査，総務省統計局（オンライン），入手先 ⟨http://www.stat.go.jp/data/shakai/2011/index.htm⟩（参照 2016-12-10）.

[2] 相澤清晴，前田一樹，小川　誠，佐藤陽平，笠松麻祐美：スマートフォン FoodLog のユーザビリティ評価，電子情報通信学会技術研究報告. LOIS, ライフインテリジェンスとオフィス情報システム， Vol. 113, No. 479, pp. 59–63 (2014).

[3] 雨宮寛敏，山岸勇貴，河合　純，金田重郎：導電性箸を用いた摂食行動の自動検出手法，電気学会論文誌 C (電子・情報・システム部門誌)， Vol. 134, No. 4, pp. 571–580 (2014).

[4] 大西杏菜，原　直，阿部匡伸：振り返り支援における効率的な映像要約のための自動収集ライフログ活用法，情報処理学会研究報告. SPT, セキュリティ心理学とトラスト， Vol. 2015, No. 4, pp. 1–6 (2015).

[5] 林　泰宏，道満恵介，井手一郎，出口大輔，村瀬　洋：料理レシピの記述に従った家庭内調理映像の要約，電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎， Vol. 112, No. 474, pp. 121–126 (2013).

[6] 佐藤琢磨，安田陽介，中井大輔，増田　彬，前川卓也：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，情報処理学会研究報告. UBI, ユビキタスコンピューティングシステム， Vol. 2015, No. 12, pp. 1–8 (2015).

[7] Wu, M.-Y., Chen, T.-Y., Chen, K.-Y. and Fu, L.-C.: Daily activity recognition using the informative features from skeletal and depth data, *in Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1628–1633 (2016).

[8] Okamoto, K. and Yanai, K.: GrillCam: A Real-Time Eating Action Recognition System, *International Conference on Multimedia Modeling*, Springer, pp. 331–335 (2016).

[9] Toshev, A. and Szegedy, C.: Deeppose: Human pose estimation via deep neural networks, *in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1653–1660 (2014).

[10] Microsoft: Developing with Kinect for Windows, Microsoft (online), available from ⟨https://dev.windows.com/en-us/kinect/develop⟩ (accessed 2016-11-25).