

CalorieCam360: 全方位カメラによる複数人同時食事カロリー量推定システム

寺内 健人^{1,a)} 山本 耕平^{1,b)} 柳井 啓司^{1,c)}

概要

近年、食事画像認識を用いた食事管理アプリが人気を集めているが、既存アプリでは、複数人の食事を同時に記録することはできない。本研究では、全方位カメラを用いて食卓全体の食事を一度に全て取得することで、手軽に食卓を囲む全員の食事を同時にリアルタイムに記録するシステムを提案する。システムはすべて iPhone アプリとして実装したため、サーバが一切不要となり、iPhone と全方位カメラを持ち運ぶだけで、どこでも利用可能なシステムを実現した。

1. はじめに

近年、世間では健康志向が高まるとともに、食事管理の重要性が高まっている。食事管理アプリでは、取った食事や摂取カロリー量を記録することで自らの食習慣をよりよく把握し、食生活の改善に役立てることができる。既存手法では、1 人分の食事のみを記録する、もしくは摂食動作のみを記録するものが存在する。しかし、複数人同時に食事推定することはできない。全方位カメラでは、2 つの広角カメラの映像を合成することでデバイス一つでテーブル、周りの人物など食卓全体を捉えることができる。そこで、全方位カメラを用いて食卓全体の食事を一度に全て取得することで手軽に食卓を囲む全員の食事を記録することで、簡素化を目指す。全方位カメラは近年 VR の普及や、低価格化などによってより身近なものになっている。しかし、全方位カメラでリアルタイムに深層学習を適用する研究はまだ少ないため工夫が必要である。また、全方位カメラは持ち歩くことを想定するためより手軽に持ち歩けるスマートフォンを用いる。スマートフォンによる深層学習モデルの実装は、近年 CoreML, Pytorch Mobile などのライブラリの発達によって簡単にできるようになっている。

本研究では、全方位カメラとスマートフォンのみで食卓全



図 1 食卓全体の食事を記録するシステムの概要

体の食事を図 1 のように記録するシステム, CalorieCam360 を構築する。食卓全体の食事について、食事カロリー量を推定し、食卓全体の人物それぞれが摂取した食事とそのカロリー量についても記録する。食事カロリー量は、食事カテゴリと食事の実寸面積により推定できる。実寸面積はリファレンスの矩形物体を検出し、物体の面積をユーザーが入力することで求める。その後、物体認識で検出された食事カテゴリ、領域分割と組み合わせることでカロリー量推定を行う。全方位カメラから取得できる画像は多くの場合正距円筒図法で保存されており、特有の歪みがあるので、平面投影で補正する。また、食卓全体の人物について、それぞれの摂取カロリー量について記録するため、さらに人物を検出し、検出された料理と対応付ける。対応付けられた料理は前フレームとの分割領域の差分を各人物の摂取量として摂取カロリー量を計算する。

2. 関連研究

2.1 食事認識アプリ

食事管理アプリには、料理を撮影し、カロリー量を推定するもの [2], [5], [9], [12], [13], [14], [18], 摂食動作を認識し、摂食カロリー量を推定するもの [1], [11] が存在する。しかし、既存のアプリケーションでは、一人分の食事しか認識せず、複数人で食卓を囲んだ場合を想定しておらず、複数人が一つの机についているにもかかわらずそれぞれアプリケーションを使う必要がある。本研究では、全方位カメラを用いることで食卓全体の食事を一度に認識することを試みる。

¹ 電気通信大学

a) terauchi-k@mm.inf.uec.ac.jp

b) yamamoto-k@mm.inf.uec.co.jp

c) yanai@mm.inf.uec.ac.jp

2.2 全方位物体認識

全方位画像の物体認識では、既存の研究では、一般的な画像に用いるものと同じの物体認識手法を用いるが、物体検出手法の適用にはいくつかの方法が存在する。正距円筒図法の画像に直接アノテーションしたもので学習する手法 [4]、学習画像に正距円筒図法の歪みを加えて学習する手法 [17]、学習は一般的な画像を用いて行い、推論時に正距円筒図法の歪みを補正する手法 [16] 等がある。提案手法は正距円筒図法の歪みを投影を用いて補正する手法を用い、食事画像に適した平面投影を用いて検出する。

3. CalorieCam360

3.1 提案手法概要

全方位カメラ、スマートフォン用いて食卓全体の人物ごとの食事のカロリー量推定を目指す。全ての処理は、外部のサーバを用いず、全方位カメラ、スマートフォンのみで完結するようにシステムを実装する。全方位カメラは Insta 360 ONE X2 を用いる。Insta 360 ONE X2 では、幅 1024、高さ 512 の正距円筒図法の画像をリアルタイムに取得することができる。

CalorieCam360 は図 2 の 4 つの段階に分けられる。(1) リファレンスサイズの決定、(2) 料理物体の検出、(3) 料理物体のカロリー量推定、(4) 人物ごとのカロリー量推定となる。

(1) リファレンスサイズの決定では、ユーザーが面積が既知の矩形物体を選び、面積を入力することで 1 ピクセルあたりの面積を推定し、料理の実寸推定を可能とする。(2) 料理物体の検出では、YOLO v7[15] を UEC-FOOD100[7] を用いて学習し、料理物体の位置とカテゴリを検出する。(3) 料理物体のカロリー量推定では、DeepLab v3+[3] を用いて料理領域分割を行い、料理物体の実寸面積を計算し、面積からカロリー量推定を行う。(4) 人物ごとの食事カロリー量推定では、継続して料理領域分割を行い、面積の追跡をする。同時に人物の追跡も行い、人物と料理の対応付けを行うことで人物ごとに食事ごとの食べた割合を計算できる。食べた割合をもとに、人物ごとの食事カロリー量の計算が可能になる。それぞれの段階では、テーブルに置いた食事の情報を扱うため、平面投影を用いて歪みを補正したテーブルの画像を用いる。



図 2 提案手法概要

3.2 テーブルを対象とした平面投影

全方位カメラから送られてくる映像は、正距円筒図法の画像として毎フレーム取得することができる。料理はカメラ下方のテーブルに置かれると想定するため、正距円筒図法では歪みが強くなってしまい、検出が難しい。歪みを補正し、検出に向けた画像にするため、テーブルを対象として平面投影する。平面投影では、食事がカメラ下方の水平面にあると仮定し、球として表せる正距円筒図法の画像を水平面に投影する。投影式は式 1 のように表す。この時 θ は緯度、 ϕ は経度である。投影結果は図 3 のようになり、投影前画像の下部に写った歪んだテーブルや、その上に乗った食事が平面として投影後画像に表示されていることが分かる。

$$x = \frac{\sin \phi}{\tan \theta}, \quad y = \frac{\cos \phi}{\tan \theta} \quad (1)$$

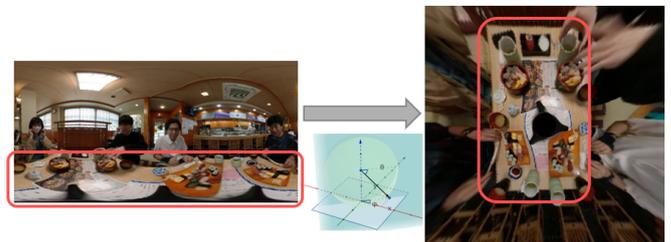


図 3 テーブル面に対する平面投影

3.3 リファレンスサイズの決定

料理のカロリー量を推定するために本研究では物体の面積を用いる。しかし、面積を求めるにはカメラに写った物体の実寸が必要になる。そこで、ユーザーが面積が既知な矩形物体を検出することで実寸を計算する。料理検出は平面投影後の画像で行うため、リファレンスサイズの計算も平面投影後の画像で行う。この際、平面投影後の画像にも歪みが残っているが、本研究では考慮しないものとする。アプリでは矩形検出をした後、ユーザーが既知の面積の矩形物体を選択し、面積を入力する。矩形検出では、Apple の Vision フレームワークの矩形検出機能を用いる。検出した矩形は左上、左下、右下、右上の 4 点で表され、その 4 点で囲まれたピクセル数を計算する。入力された面積を検出矩形のピクセル数で割ることで 1 ピクセルあたりの面積を求めることができる。

3.4 料理物体検出

物体検出では、全方位カメラで取得した画像を投影した後、UEC-FOOD100[7] で学習した YOLO v7[15] を用いて食事検出を行う。テーブル全体の画像を扱うため、料理が小さくなってしまふ。小さいサイズの物体に対応するために、学習画像のデータ拡張において画像スケールを 0.04 倍から 0.3 倍にリサイズする。また、投影後の全方位画像は、

料理が全方位から写るため、 -180° から 180° に画像を回転する。他の設定は標準の YOLO v7 の設定に従う。画像の解像度は学習時 640、推論時 1280 に固定する。投影後画像の中央は全方位カメラの直下であり、死角になり黒くなる。そのため料理と認識されることがあるため、真ん中のバウンディングボックスは取り除く。

3.5 料理領域分割

料理物体を検出後、検出バウンディングボックスを領域分割することで料理のピクセル数を求め、ピクセル数をもとに面積を推定する。料理領域分割は、料理物体検出同様に平面投影後の画像に対して行う。料理領域の分割では、セマンティックセグメンテーションモデルの Deeplab v3+[15] を用いる。Deeplab v3+は領域分割ツールセットである MMSegmentation[8] を用いて実装し、バックボーンが ResNet50[6] の標準設定を用いて学習する。データセットは UEC-FoodPIX Complete[10] を用いて学習する。UEC-FoodPIX Complete は 100 種類 1 万枚の食事画像データセットであり、それぞれの画像にはカテゴリごとに手動でピクセル単位のアノテーションがされている。データ拡張はクロップ、スケーリングの他に、回転、カットアウトを加え、食べかけの料理画像に対しても頑強になるようにする。

3.6 食事カロリー量推定

料理領域分割で求めた面積をもとに食事カロリー量の推定を行う。食事カロリー量の推定では岡元 [12] らの手法に従い、面積、カテゴリからカロリー量を計算する。會下ら [18] が作成したカロリー量の回帰曲線を用い、図 4 のように料理領域分割で求めた料理の面積、物体検出で検出したカテゴリでのカロリー量を求めることで料理のカロリー量とする。UEC-FOOD100 では、100 種類のカテゴリがあるが、全てのカテゴリの面積、カロリー量データがあるわけではないため、データがないカテゴリでは、カロリー量推定は行わない。

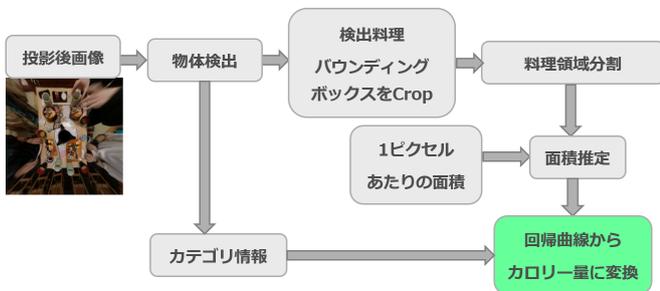


図 4 食事カロリー量推定の流れ

3.7 人物ごとの摂取カロリー量推定

テーブルを使っている人物一人ひとりについての摂取カ

ロリー量推定について考える。平面投影により、テーブル全体の料理を観測し、カロリー量を推定することはできるが、料理情報のみを扱うだけでは、テーブルを使っている人全員分の摂取量しか観測できない。したがって、テーブルを使用している人々の追跡も必要となる。

人物ごとの摂取カロリー量を推定する手法は図 5 のようになり、人物を検出し、人物と食事の対応付けをし、食事領域の減少量から人物ごとの食事の摂取量を求め、摂取カロリー量に変換する。人物の検出には、Vision フレームワークの骨格検出機能を用いる。骨格検出機能では、検出した人物ごとに首、頭、手、足などのランドマークの位置の情報を得ることができる。検出された人物は、前フレームで検出された最も近い人物と同一人物として扱うことで人物の追跡を可能にすることができる。この時、離れすぎている、もしくは前フレームより検出人物数が多いなどの理由で対応が見つからなかった人物は新しく検出した人物として扱う。

料理ごとに、現在食べている人物を対応付けることで、人物ごとの食べた料理の量を推定することができる。料理ごとに、最も近い手首、肘の人物を対応づける。食事の摂取量は食事領域の減少量によって求める。最初全ての人物の全ての料理に対する摂取量を 0 とし、各フレームにおいて、それぞれの料理ごとに、対応する人物の料理の摂取量にそのフレームでの食事領域の変化量を足す。このようにすることで、人物ごとに料理ごとの摂取量が記録できる。摂取カロリー量は人物ごとの摂取量から 3.6 節と同様の方法で求める。

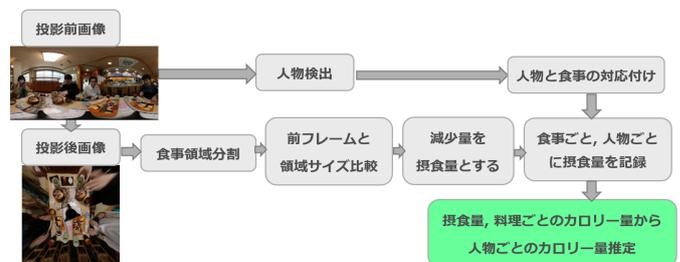


図 5 人物ごとの摂取カロリー量推定の流れ



図 6 人物と料理の対応付け

4. CalorieCam360 の使用例

提案システムの使用例を図 7~11 に示す。

まず、準備として、図 7 のように (a) テーブルにカメラを設置し、(b) リファレンスサイズ決定画面で A4 用紙を映して A4 用紙を囲う矩形を検出する。(c) 検出できたことが確認出来たら A4 用紙の矩形を選択し、A4 用紙の面積である $624cm^2$ を入力し、リファレンスサイズ決定を終了する。

次に、料理物体検出画面が表示されるので、図 8 のように (a) テーブルに料理を置き、置いた料理を検出する。(b) 検出出来たらボタンを押し、実際に検出が出来ているかを確認する。料理が検出出来たことが確認出来たら、人物ごとの摂取カロリー量追跡へ進む。

実際に料理を食べ進めると、図 9 のようにパーセント表記の食事の残りが減ることが分かる。

料理を食べ終わったら図 10 のように (a) それぞれの料理の残量が少なくなっていることが確認できる。(b) 食事結果閲覧ボタンを押すと食事結果閲覧画面に進み、人物ごとの料理ごとの摂取量が分かる。カロリー量が検出できる料理カテゴリであった場合は摂取カロリー量も表示される。

結果として、牛丼は焼きそばと検出されてしまい、検出、追跡がうまく働かなかった。スパゲッティは検出は出来たが、追跡はある時点から領域分割が不安定になっている。餃子については検出、追跡がうまく働いた。食事結果閲覧画面と人物の対応は図 11 のようになり、正距円筒図法の画像に写った右側の人物が“人物 1”，中央の人物が“人物 2”，左側の人物が“人物 3”として表示され、最も近い料理を摂取したとして対応付けられていることが分かる。

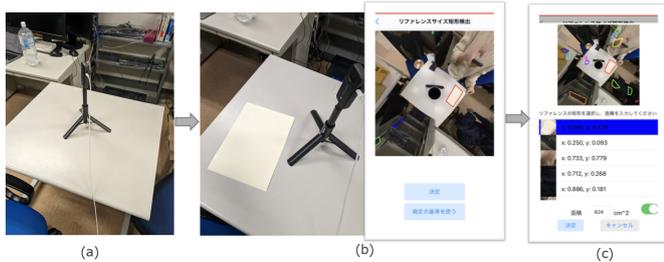


図 7 実際の使用例：準備



図 8 実際の使用例：食前

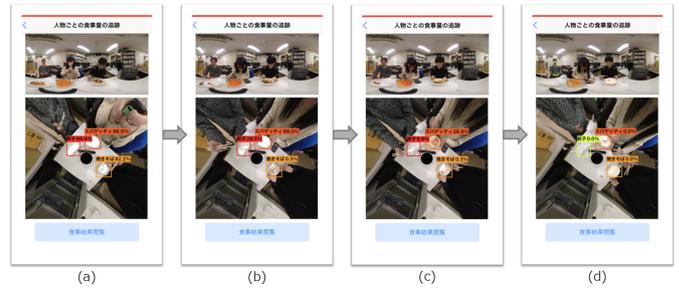


図 9 実際の使用例：食事中

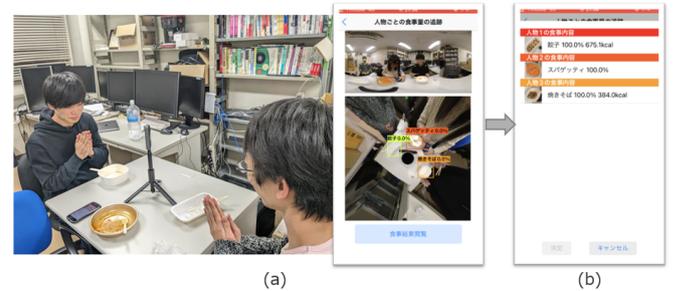


図 10 実際の使用例：食後

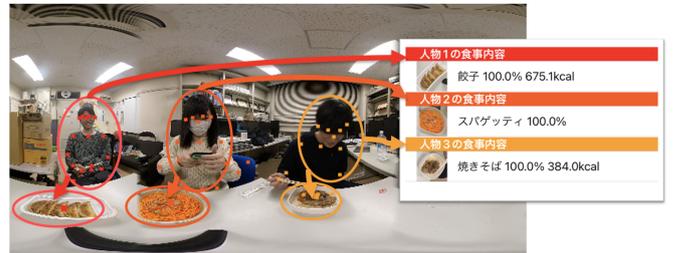


図 11 実際の使用例：人物の対応

5. おわりに

全方位カメラを用いて食卓全体の食事を一度に全て取得することで手軽に食卓を囲む全員の食事を記録するための、全方位カメラとスマートフォンを用いた食卓全体の食事認識アプリケーション CalorieCam360 を提案した。アプリではリファレンスサイズの決定、料理物体検出、料理カロリー量推定、人物ごとのカロリー量推定の機能により、食事が始まってから終わるまでを記録し、各人物の各料理に対する食事量を求める。アプリの各機能は矩形検出、物体検出、領域分割、骨格検出を組み合わせることで実現した。物体検出、領域分割は料理画像データセットを用いて学習し、CoreML モデルに変換してスマートフォンに実装した。アプリケーションは全てのモデルを Vision フレームワークを用いてスマートフォン上に実装し、全ての動作をサーバーなどを用いず全方位カメラとスマートフォンで完結するようにした。実験により、CalorieCam360 が実際に食卓を囲んだ複数人での食事に対して、食卓全体を一度に記録できることを示した。今後は、各段階のモデルの改良、顔検出による人物追跡、物体追跡による料理物体の追跡などの実装による使いやすさの向上に取り組んでいきたい。

なお、会議当日は、ポスター（もしくはデモ）発表時に、疑似食材を用いて CalorieCam360 による複数人の同時食事動作認識のデモを行うことを予定している。

参考文献

- [1] Adachi, K. and Yanai, K.: DepthGrillCam: A Mobile Application for Real-Time Eating Action Recording Using RGB-D Images, *Proc. of the 7th International Workshop on Multimedia Assisted Dietary Management* (2022).
- [2] Ando, Y., Ege, T., Cho, J. and Yanai, K.: Depthcaloriecam: A mobile application for volume-based food calorie estimation using depth cameras, *Proc. of International Workshop on Multimedia Assisted Dietary Management* (2019).
- [3] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation, *Proc. of European Conference on Computer Vision* (2018).
- [4] Chou, S.-H., Sun, C., Chang, W.-Y., Hsu, W.-T., Sun, M. and Fu, J.: 360-Indoor: Towards Learning Real-World Objects in 360° Indoor Equirectangular Images, *Proc. of IEEE Winter Conference on Applications of Computer Vision* (2020).
- [5] Ege, T. and Yanai, K.: Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions, *Proc. of the on Thematic Workshops of ACM Multimedia* (2017).
- [6] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proc. of IEEE Computer Vision and Pattern Recognition* (2016).
- [7] Matsuda, Y., Hoashi, H. and Yanai, K.: Recognition of Multiple-Food Images by Detecting Candidate Regions, *Proc. of IEEE International Conference on Multimedia and Expo* (2012).
- [8] MMSegmentation Contributors: MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark, <https://github.com/open-mmlab/mms Segmentation> (2020).
- [9] Naritomi, S. and Yanai, K.: Hungry Networks: 3D Mesh Reconstruction of a Dish and a Plate from a Single Dish Image for Estimating Food Volume, *Proc. of ACM International Conference on Multimedia in Asia* (2021).
- [10] Okamoto, K. and Yanai, K.: UEC-FoodPIX Complete: A Large-scale Food Image Segmentation Dataset, *Proc. of ICPR Workshop on Multimedia Assisted Dietary Management* (2021).
- [11] Okamoto, K. and Yanai, K.: Realtime eating action recognition system on a smartphone, *Proc. of IEEE International Conference on Multimedia and Expo Workshops* (2014).
- [12] Okamoto, K. and Yanai, K.: An Automatic Calorie Estimation System of Food Images on a Smartphone, *Proc. of International Workshop on Multimedia Assisted Dietary Management* (2016).
- [13] Tanno, R., Ege, T. and Yanai, K.: AR DeepCalorieCam: An iOS App for Food Calorie Estimation with Augmented Reality, *Proc. of International Conference on Multimedia Modeling* (2018).
- [14] Tanno, R., Ege, T. and Yanai, K.: AR DeepCalorieCam V2: Food Calorie Estimation with CNN and AR-Based Actual Size Estimation, *Proc. of ACM Symposium on Virtual Reality Software and Technology* (2018).
- [15] Wang, C.-Y., Bochkovskiy, A. and Liao, H.-Y. M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *arXiv preprint arXiv:2207.02696* (2022).
- [16] Yang, W., Qian, Y., Kämäräinen, J.-K., Cricri, F. and Fan, L.: Object Detection in Equirectangular Panorama, *Proc. of International Conference on Pattern Recognition* (2018).
- [17] Zhang, Y., Xiao, X. and Yang, X.: Real-Time Object Detection for 360-Degree Panoramic Image Using CNN, *Proc. of International Conference on Virtual Reality and Visualization* (2017).
- [18] 會下拓実, Cho, J., 松平礼史, 柳井啓司: 米飯を基準とした CNN による食事画像からのカロリー量推定, 画像の認識・理解シンポジウム (2019).