

GrillCam: 食事動作認識によるリアルタイム食事記録システム

岡元 晃一^{1,a)} 柳井 啓司^{1,b)}

1. はじめに

近年、世間の健康的思考の高まりにより食事記録を取る人が増えている．そこで、簡単に食事記録を取るために、食事画像認識を用いたスマートフォンアプリケーションが提案されている．例えば、河野らによる食事記録支援システム“FoodCam” [1] や、foo.log による“FoodLog App” [2] では、ユーザが食事開始前に料理の写真を撮影し、システムが認識、カロリー推定を行う．こうしたシステムでは、事前に食べる食事がすべて食卓上に準備されていることが前提となっており、大皿料理や鍋、焼き肉など、事前に食べる量が決まっていないスタイルの食事に対しては、対応することが不可能であった．

そこで、本研究では、事前に用意された食事を静止画で撮影するのではなく、食事シーンを動画で撮影し、ユーザの口に運ばれた食事の種類を認識することによって、実際にユーザが摂取した食事の総カロリーをリアルタイムに推定する自動食事記録システムの実現を目指す．これにより、事前に食事量が決まっていない、大皿料理や鍋、焼き肉などの料理についても自動食事記録が可能となる．

本発表では、特に、「焼き肉」シーン限定したプロトタイプシステムについて発表を行う．具材を「肉」「ご飯」「かぼちゃ」「ピーマン」「にんじん」の5種類に限定し、「焼き肉」シーンにおける摂取具材名およびカロリーの推定を実現した．実装は、スマートフォン上に行い、図1に示すように食事のユーザに向けて置くことによって利用する．食事動作はスマートフォンのインナーカメラで撮影し、認識結果は即時にスクリーン上に表示される．ユーザは累積の摂取カロリーを確認しながら食事を行うことが可能となる．

2. 関連研究

食事記録支援システムでは“FoodLog App” [2] や、河野らの“FoodCam” [1] が挙げられる．

FoodLog App は食べる前の食事画像を撮影、FoodLog サーバーへとアップロードし、サーバー側で食事画像認識を行い食品名やカロリーなどの結果をユーザに返している．しかし、これには通信を行わなければならないのでいつでも手軽に使えるというわけではない．

そこで河野らの FoodCam ではスマートフォン上での食事画像認識を可能にしており、サーバーとの通信を必要としないシステムを作成した．最新バージョンでは 100 クラスを対象とし、Fisher Vector と線形 SVM で分類を行い、上位 5 位までの結果で 79.2 % の結果を示している．

しかし、これらは二つとも食事前の画像を撮影するもの



図 1 本システムの使用時のイメージ図

なので、食べる量のわからない大皿料理や鍋、焼き肉には対応できない．そこで、本研究では食事シーンの動画を用いて口元に近づいた食べ物を認識することで、大皿料理や鍋、焼き肉などでも、摂取具材名およびカロリーの推定をスマートフォン上で行えるシステムを作成する．

3. システムの概要

本システムでは、食べ物を食べたという判断には、口元に食べ物を運んだ瞬間とする．これには動画内から口を検出して、その付近に食べ物が近づいたことを箸などの食器を検出することで判断する．そして画像を切り出し、その切り出した画像から具材を判別し、その具材の名前、カロリーの計算などを行う．

つまり本システムでは以下の手順で処理を行う．

- (1) 食事シーンの動画を撮影する
 - (2) 食べ物を口元に運んだ瞬間の画像を切り出す
 - (3) 切り出した画像から具材の判別をする
 - (4) 食べた量やカロリーの計算を行い記録する
- システム中の関数には OpenCV ライブラリを使用している．

4. システムの詳細

4.1 口検出

本研究では口元に食べ物が来た瞬間をものを食べた瞬間として認識するのでまずは口の検出を行う必要がある．しかし、ただ口の検出を行っても背景や顔でも眼が口として検出されてしまっている場合がある．

そこで口を検出する前に顔を検出し、その領域内でのみ口を検出した．これにより背景に口が検出されることがな

¹ 電気通信大学大学院 情報理工学研究所 総合情報学専攻

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

b) yanai@cs.uec.ac.jp

くなり、更に顔の下部分のみを探索対象とすることで眼が口として検出されることがなくなった。結果を図 2 に示す。



図 2 顔検出を用いた前(左)と後(右)の口検出

4.2 箸検出

箸は基本的に直線なので画像中の直線部分を検出し、認識する。直線の検出には確率的ハフ変換を用いて線分として検出できるようにした。

しかし、実際に検出を行うと背景部分にも大量の線分が検出されてしまいどれが正しい箸が分からず、総当たりで箸かどうかの認識を行うのも非効率的である。そこで動画中の動いている前面部分のみを切り出し、そこでのみ直線の検出を行い、箸の検出を行った。これにより背景に検出された線分を無視することができ非常に良い結果となった。結果を図 3 に示す。



図 3 箸の認識、前面切り出し前(左)と後(右)

4.3 具材判別

前述の方法で口と箸を検出し、2点が生じた瞬間を食べたと認識し、箸先を中心に 200×150 ピクセルの領域で切り出し、その画像に対して具材判別を行う。

スマートフォンは PC に比べて、CPU の処理性能が十分ではないため、実行速度も非常に重要である。そこで現在画像認識で主流である局所特徴量から 1 種、カラーヒストグラムから 1 種の計 2 つを候補に挙げた。

局所特徴量では SIFT 特徴量や SURF 特徴量が主流である。予備実験で SURF 特徴量を用いた実装も行ったが、十分な処理速度が得られなかったため、今回はそれらの発展系である Rublee らの提案した ORB 特徴量 [3] を用いる。

前述の特徴量よりもバイナリコードでの記述なので精度が落ちるがハミング距離での計算が可能なので非常に高速であり、回転やスケール変換にも強いという利点がある。本システムでは、求められた ORB 特徴量を Bag-of-Features を用いてヒストグラムとして表現しそれを画像の値とする。

また、食品画像における特徴量では色特徴量が有力であると上東ら [4] の研究でも言及されている。そこで本システムでも HSV 色空間を用いたカラーヒストグラムを作成し、そのヒストグラムを特徴量として用いて認識を行う。

分類には線形 SVM を用いる。このとき Vedaldi ら [5] が提案している Explicit Feature Maps を用いて高次元写像

を行うことで線形 SVM を用いても非線形 SVM と同等の精度を出すことができ、なおかつ計算量とメモリの大幅な削減を実現している。

5. 実験

実験では食べる前に量がわからない大皿料理や鍋の中から家庭内や外出においても非常に多く食べられている焼き肉を対象にする。そして、焼き肉を行う上で具材として選ばれる可能性の高い以下の 5 種類に対して実験を行う。

(1) 肉 (2) ご飯 (3) かぼちゃ (4) ピーマン (5) にんじん
使用する特徴量は前述の 2 つの特徴量を組み合わせて使用したものがもっとも結果が良かったのでそれを用いる。

実験は精度評価とユーザ評価の 2 つを行う。以下精度評価実験を行い食材それぞれの再現率、適合率、全体の精度を示したものが表 1 である。

表 1 精度結果

具材	再現率 (%)	適合率 (%)
肉	76.0	74.5
ご飯	86.0	89.6
かぼちゃ	56.0	57.1
ピーマン	88.0	72.1
にんじん	68.0	82.9
全体の精度 (%)	74.8	

またユーザにおける評価でも食べる毎に画面上のボタンをタッチするシステムよりも使いやすいという評価を得た。

6. まとめ

本研究では食事シーンをスマートフォンで撮影することでリアルタイムに食事記録をつけられるシステムを作成した。結果としては 74.8 % の精度を得て、ユーザ評価においてもベースシステムよりも使いやすいという評価を得た。

今後はスマートフォンアプリの UI の見やすさや、ユーザによる誤認識の修正機能、食事記録の履歴を閲覧できる機能の追加などを考えたい。

また今回データセットにおいては食品サンプルと実際の食品を混合させたものを作成したが、これがより改善されれば精度の向上に貢献すると考えられるのでデータセットの構築方法を考えるとともに焼き肉以外の鍋などの大皿料理にも対応したい。

参考文献

- [1] Y. Kawano and K. Yanai. FoodCam: a real-time food recognition system on a smartphone. *Multimedia Tools and Applications*, 2014. (in press)
- [2] "Foodlog App" <http://app.foodlog.jp/>.
- [3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: an efficient alternative to SIFT or SURF. *Proc. of IEEE International Conference on Computer Vision*, 2011.
- [4] 上東太一, 甬足創, 柳井啓司. Multiple kernel learning による 50 種類の食事画像の認識. 電子情報通信学会論文誌. D, 情報・システム, Vol. 93, No. 8, pp. 1397-1406, 2010.
- [5] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 480-492, 2012.