食事認識を用いたモバイル食事管理システム

河野　憲之†　　柳井　啓司††

† 電気通信大学 電気通信学部 情報工学科　〒 182-8585 東京都調布市調布ヶ丘 1-5-1
†† 電気通信大学 大学院情報理工学研究科 総合情報学専攻　〒 182-8585 東京都調布市調布ヶ丘 1-5-1
E-mail: †kawano-y@mm.inf.uec.ac.jp, ††yanai@cs.uec.ac.jp

あらまし　近年、スマートフォンの性能が大きく向上している。そこで、本研究では、通常サーバにデータを送り、画像処理をする部分をスマートフォン上でリアルタイムに実行することにより、通信コストのかからない、ネットワークに依存しない食事管理システムを提案する。50 種類の料理に対して、背景を含まない料理の領域が与えられたとき、候補を 5 つ提示し 81.4%の認識精度であった。また、バックグラウンドでは料理の領域の補正を行い、さらに、認識を誤った場合を考慮し、ユーザに料理のある方向を提示する。ユーザに提示する料理の方向は、認識する領域から料理が 15%ずれていた場合、角度差 ±20° 以内に 31.8%、±40° 以内に 50.3%の精度で、25%ずれていた場合、角度差 ±20° 以内に 34.5%、±40° 以内に 54.2%の精度であることを確認した。

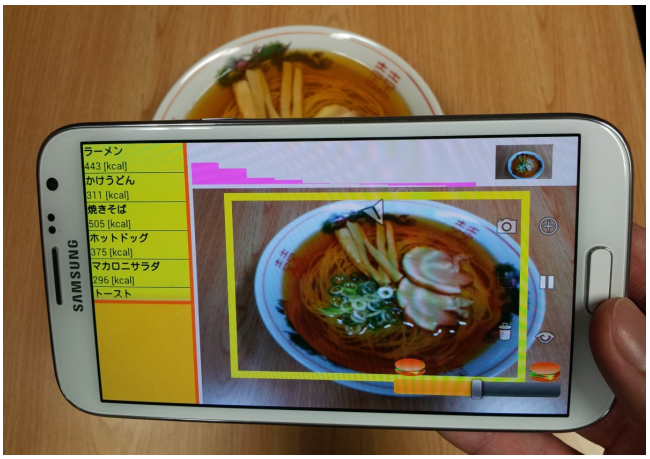
キーワード　モバイル、食事認識、食事管理、ユーザインタラクティブ

1. はじめに

近年、健康志向の高まりによりスマートフォンなどのモバイルデバイスから食事記録をとることのできるシステムが多く現れるようになった。一般的な食事管理システムの記録方法は、テキスト入力や階層型メニューによる選択などが挙げられるが、入力に手間がかかり、継続した利用が難しい。

また、スマートフォンの普及により、それに伴いスマートフォンの性能も大きく向上し、スマートフォン上で以前より計算コストの高い処理をすることが可能になった。スマートフォンから画像処理を利用する一般的なシステムは、スマートフォンを通信手段として利用することが多いが、通信コストがかかり、ネットワークにも依存する。

そこで本研究では、スマートフォン上で食事認識をリアルタイムに行うことにより、ネットワークに依存しないモバイル食事管理システムを提案する。図 1 は提案システムのイメージである。



(a) システムを食事にかざす

(b) システムの認識画面

図 1　提案システムのイメージ

2. 関連研究

食事認識は食べ物に決まった形はなく、同じカテゴリ内であっても視覚的変化が大きいため、難しいタスクである。松田ら [1] は、円検出、JSEG、DPM により食事領域推定後、SIFT、HOG、Gabor、カラーヒストグラムにより 100 種類の食事画像に対して複数品の分類に取り組んだ。Yang ら [2] は、画素間の距離や角度等、材料の位置関係を特徴量とする手法により、ファストフードの分類に取り組んだ。本研究でも、食事認識を行い、料理を分類して結果上位をユーザに提示する。

食事管理システムとしては、一般的な画像認識を用いない場合は、料理データベースの拡大が容易であるが、手動による入力のため手間が多く、継続した利用が難しい。画像認識を用いた食事管理システムでは、食事画像からバランス推定をし、そ

の結果を返す FoodLog [3] や、なチェッカーボードとともに食事を撮影し、食事の分類と量の認識を行う TADAproject [4] がある。しかし、いずれもサーバに画像データを送り、画像処理しているため通信コストが高く、認識を誤った場合は、ユーザが後から手動で直すことになる。本研究でも食事認識を用いた食事管理システムを構築するが、スマートフォン上でリアルタイムに認識、記録することにより、簡単に正確な食事記録をとることのできるシステムを提案する。なお、本研究では、量はユーザに入力してもらい、食事の種類の認識のみになっている。

スマートフォンと画像認識の研究では、近年のスマートフォンの普及により、スマートフォンから利用できる画像認識システムが多く現れるようになった。物体認識アプリケーションとして有名な Google Goggles(注1) は、ロゴや芸術品、建造物などを認識し、その情報を返すアプリケーションである。また、Kumar ら [5] の一定の環境条件下 (照明、背景) で撮影した葉の画像をサーバに送り、葉独自の形状特徴を抽出し、その葉から種を認識して結果を返すアプリケーション Leaf snap や、Maruyama ら [6] の、30 種類の食材を認識し、レシピを返すアプリケーションがある。スマートフォン上でリアルタイム性に重点をおいたアプリケーションとして、Lee ら [7] の研究がある。複数スケールでのテンプレートマッチングを提案し、ユーザが登録した物体に対して物体検出や追跡をリアルタイムに実現した。本研究では、視覚的変化の大きい食事に対してスマートフォンの計算資源のみで認識を行う。

ユーザインタラクティブなシステムとして Yu ら [8] の研究である。モバイル位置検索で認識を間違えた場合を考慮し、次にどの視点を撮影すればよいかを、オフラインで場所ごとに求めた顕著性、オンラインで画像マッチング、Gist 特徴と SVM により認識しやすい視点を求め、それをユーザに返す Active Query Sensing(AQS) を提案した。本研究においても、処理に時間のかかる部分はユーザに補助してもらい、また料理のある方向を提示することによりユーザインタラクティブなシステムを構築する。

3. システム概要

本システムの目的は画像認識技術を利用してユーザの食事記録をとる補助と食事記録を見直すことで食生活を確認できるようにすることである。

3.1 食事記録登録

本システムの食事記録登録の基本的な使用の流れを以下に、例を図 2 に示す。

（1） 食事にスマートフォンをかざす
（2） ユーザは料理領域を入力する
（3） 食事認識を行う
（4） 一定時間後、認識結果上位を提示する
（5） ユーザは認識結果上位から料理を選択する
（6） 未選択の料理があれば、2 もしくは 3 に戻る
（7） 食事画像を保存する
（8） 食事記録を登録する

2 では、ユーザによる料理の領域 (以後、料理領域) 入力は対角線を引くことで矩形領域とし、バックグラウンドで領域推定による料理領域の補正が行われる。料理領域は 4 つまで入力可能であり、料理領域を入力しなければフレーム全体から食事認識を行う。4 では、一定時間食事認識を繰り返し、最終的な出力値は各出力値の平均としている。また、結果リストの更新の間隔はユーザ調査により決定した。5 では、料理があると考えられる方向を提示することにより、認識結果上位に目的の料理が現れなかった場合の対処をする。記録にはメモや位置情報も登録でき、サーバにアップロードすることにより食事記録をユーザ間で共有可能である。

3.2 食事記録閲覧

管理システムとしては以下のように閲覧できる。例を図 3 に示す。

（1） 日ごとに食事記録を閲覧
（2） Google Maps 上で閲覧
（3） 最近の食の傾向の確認
（4） アップロードされた食事記録の閲覧

また 1 日の食事の評価を [9] より 4 群点数を用いて 5 つ星で行い、Web から記録を閲覧できるように Web サイトを作成した。

4. 認識手法

本システムはモバイル上で認識するため、高次元特徴量などを多数使用することができない。本研究では、カラーヒストグラム、カラーモーメント、カラーオートコレログラム、Gabor、HOG、PHOG、SURF 記述子からその性能を比較し、よい結果となったものを使用した。

食事画像認識の研究における特徴量の性能を示したものでは、Bosch ら [10] は色特徴が、Hoashi ら [11] は局所特徴が、李ら [12] は、テクスチャ特徴がそれぞれ分類器に SVM を用いて最もよいという結果になった。しかし、実験環境の違いもあり、同一特徴量が最高の性能を示しているわけでないことからも、特徴量の比較検討を行った。なお、モバイルで扱うために次元数が大きくなりすぎないようにそれぞれ設定し、画像サイズが大きい場合には、総ピクセル数が 3 万になるようにアスペクト比を保ったままリサイズした。

4.1 特徴量

4.1.1 カラーヒストグラム

カラーヒストグラムは色分布を表現した特徴量である。画像を 3×3 に分割し、RGB 色空間、HSV 色空間、La*b*色空間それぞれ 4 色ずつに減色し、合計 64 次元特徴ベクトルを各領域から抽出することで 3 種類各 576 次元特徴ベクトルを得た。

4.1.2 カラーモーメント

カラーモーメントは、カラーヒストグラムと同様に色分布を表現した特徴量であるが、一般にカラーヒストグラムよりも少ない情報量で表現することが可能である。画像を 5×5 に分割し、RGB 色空間と HSV 色空間、RGB 色空間と La*b*色空間の各チャンネルの平均と分散を算出し、合計 12 次元特徴ベクトルを各領域から抽出することで 2 種類各 300 次元特徴ベクト

(注1)：http://www.googles.com/mobile/goggles/

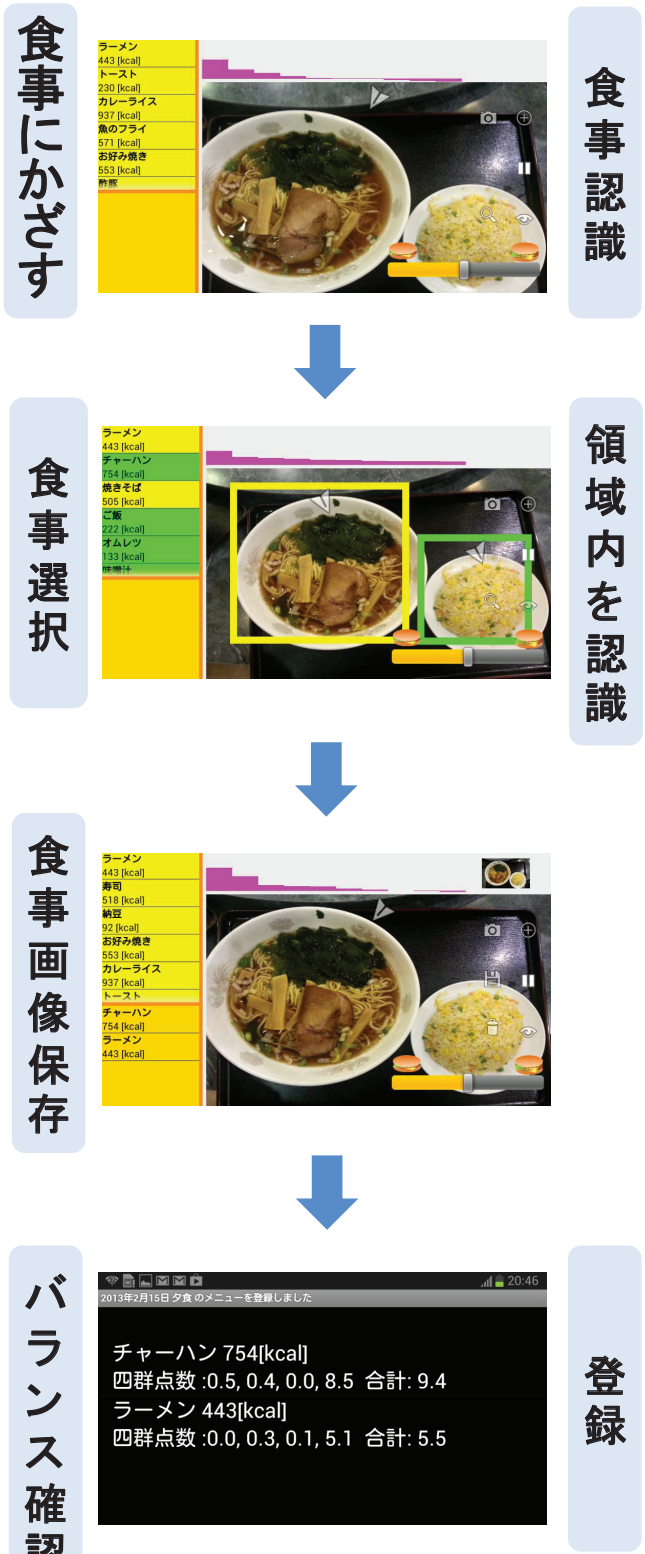


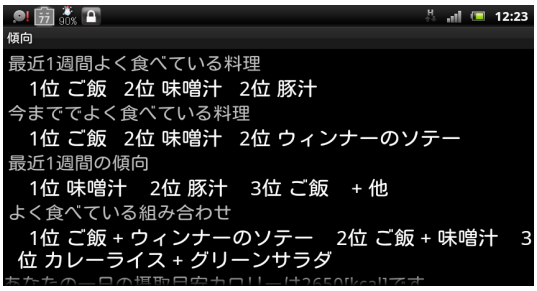
図 2　使用の流れ

(a) 閲覧 (1 日ごとに閲覧、4 群点数に基づき 5 つ星で評価する。)

(b) 詳細情報 (1 食分の栄養素を表示する。)

(c) 食事画像 (記録した食事画像を表示する。)

(d) Map(位置情報を付けると Google Maps 上に表示する。)



(e) 食の傾向 (ユーザの食事の傾向を表示する。)

(f) 詳細情報 (Web にアップロードして表示する。)

図 3　記 録 閲 覧

ルを得た。

4.1.3 カラーオートコレログラム

カラーオートコレログラム [13] は、隣接する色の類似度を表現した特徴量である。対象を同色の画素同士としているため、カラーコレログラムよりも少ない情報量で表現することができる。RGB 色空間を 4 色ずつに減色し、距離を $2k+1(0 \leqq k \leqq 16, k:$ 整数) として 1024 次元特徴ベクトルを得た。

4.1.4 Gabor

Gabor は、局所的な濃淡情報の周期と方向を表現した特徴量である。画像を 4×4 に分割し、各領域から 6 方向、4 周期のガボールフィルタにより得られる 24 次元、合計 384 次元特徴ベクトルを得た。

4.1.5 HOG

HOG [14] は、輝度の勾配方向をヒストグラム化した特徴量である。画像を 6×6 に分割し、1296 次元特徴ベクトルを得た。

4.1.6 PHOG

PHOG [15] は、画像をいくつかのレベルに分割し、各領域で勾配ヒストグラムを作成する。ピラミッドレベルは 3 を使用し、合計 680 次元特徴ベクトルを得た。

4.1.7 SURF

SURF [16] は、スケール変化、回転、照明変化に頑健な 64 次元特徴ベクトルである。8 ピクセルごとにスケールを半径 12、16 ピクセルとして dense sampling し、Bag-of-Features 表現で 500 次元と、1000 次元の特徴ベクトルを得た。

soft 割り当て [17] は、複数コードワードに割り当てることにより再現性を高めることができる。本研究では、最近傍 3 つにコードワードまでのユークリッド距離の逆数を割り当てることで soft 割り当てと通常の hard 割り当てを行った。また、最近傍コードワード探索には kd-tree に基づく近似最近傍探索により行った。

4.2 分類器

本研究では、分類器に線形 SVM を用い、1-vs-rest 法により多クラス分類を行う。線形カーネルは $K(x,z) = x \cdot z$ で表される。線形 SVM は、入力ベクトルを x、出力値を $f(x)$、サポートベクトルを x_i、サポートベクトルの重みを α_i、バイアス値を b としたとき、

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{1}$$

$$= \sum_{i=1}^{N} \alpha_i x \cdot x_i + b$$

$$= \sum_{i=1}^{N} \alpha_i x_i \cdot x + b$$

$$= w \cdot x + b \tag{2}$$

と展開、変形できるため、あらかじめ、サポートベクトルとサポートベクトルの重みの積の総和を計算しておくことで、特徴次元数 N だけの乗算にバイアス値を足すだけで出力値を得ることができる。

SVM の学習には、liblinear [18] を使用した。

BoF 表現した特徴ベクトルなどを直接線形 SVM で識別すると、識別性能が悪いことが知られている。そのため、特徴ベクトルを非線形写像して、高次元空間で線形識別を行うカーネルトリックにより識別性能が大きく向上するが、同時にスケーラビリティも低下する。そこで、explicit embedding 手法により、線形識別機での適用を可能にする。本研究では、kernel feature maps を用いる。

4.3 Kernel feature maps

Kernel feature maps は直接線形 SVM を適用できるように、得られた特徴ベクトルをあらかじめ高次元空間に写像しておくことで、線形 SVM を適用しても非線形 SVM と同等の性能をだすことが可能である。Vedaldi ら [19] は、Hellinger、χ^2、intersection、Jensen-Shannon(JS) の任意のカーネルの写像 ϕ が以下に近似表現できることを示した。

$$\phi_\omega = \kappa_\omega \sqrt{x} \mathrm{e}^{-i\langle \omega, \log x \rangle} \tag{3}$$

本研究では、χ^2 カーネルの写像を利用する。χ^2 カーネルの写像 ϕ は $e^{i\omega \log x}\sqrt{xsech(\pi\omega)}$ で表される。写像後の次元数は元の特徴ベクトルの 3 倍になるようにし、L1 正規化した特徴ベクトルと相性がよいとされる [19] ので、L1 正規化したカラーヒストグラム、PHOG、SURF に対して使用した。

また、Perronnin ら [20] は特徴ベクトルの各要素の平方根をとったベクトルが Hellinger カーネルに対する正確な写像になることを示したので、同様に L1 正規化した特徴ベクトルに対して比較した。

4.4 領域推定

特徴抽出する領域はユーザによって正確に与えられればよいが、実際にシステムを使用する上で正確に料理領域を与えるのは手間がかかり、また、背景を多く含むと認識精度は一般に低下する。そこで、領域分割手法により料理領域を推定し、料理領域の補正を行う。本研究では、ユーザは、少なくとも料理を含むように領域を入力する、という制約を与える。この制約を与えたとき、矩形内の画素値を前景と背景に分離する GrabCut を適用することで、料理領域を推定する。

4.4.1 GrabCut

GrabCut [21] は、矩形領域を与えると、矩形領域内は前景と背景が混在するとして、矩形領域外は背景として色分布から GMM を作成し、各画素の前景らしさ、背景らしさの尤度を求め、領域分割を行う。

認識ごとに毎回 GrabCut を適用するにはコストが大きいため、ユーザが料理領域を与えると認識を開始すると同時に、バックグラウンドで GrabCut による料理領域の補正を行う。また、動画でリアルタイムに認識、結果を提示するので、カメラの位置は固定でない。そのため、最終的な領域は前景領域を全て含む最小の矩形領域で表現することにした。そして、入力された料理領域と重心が重なり、高さと幅をそれぞれ 2 倍した矩形領域を、大きい場合には総ピクセル数を 6 万にリサイズした領域に対して、元の矩形領域に対応する領域を領域分割するように GrabCut を適用した。

4.5 方向提示

ユーザがシステムを使用する際、正しく認識できない場合、料理の見え方を変更しなければ評価値は変わらず、認識させたい料理はリストに一向に現れない。そこで、ユーザインタラクティブな要素として、料理を認識すると同時に料理があると考えられる方向を指示することにより、認識結果がよくなる領域を写すように促す。

手法には、SURF-BoF を直接線形 SVM に適用した評価値による Window 探索を用いる。線形 SVM の評価値を用いて物体検出する手法には、ESS(Efficient Subwindow Search) [22] などがある。この線形 SVM の評価値の場合、BoF であれば特徴ベクトルをコードワードに割り当てることは、それに対応する式 2 の w の値を累積することに相当する。また、式 2 の w は、

$$w = w^+ + w^- \tag{4}$$

と表現できるため、ある矩形内の線形 SVM の評価値は w^+ と w^- それぞれについて積分画像を作成しておくことにより、$\mathcal{O}(1)$ の計算量で得ることができる。前述の soft 割り当ての場合は、スケーリングを考慮しなければ w とコードワードに割り当てられた値との積を累積することで可能である。

探索するウィンドウは入力されている矩形領域を $B \times B$ の領域と考え、各辺が $(B-2)$ の矩形領域を、与えられた矩形領域の内側かつ少なくとも互いの 1 辺が重なるようにウィンドウを 8 ピクセルずつスライドさせ 1 周するまでそれぞれ評価値を得る。ここで、$B = 2x$ $(3 \leqq x \leqq 6, x:$ 整数$)$ とした。ウィンドウごとにカテゴリ数分評価値が得られるが、得た全ての評価値の中で最も評価値が高かった矩形領域の重心の方向を最終的な方向として、矢印でユーザに提示する。図 1 に矢印提示の例が示してある。

各 SVM の学習には負例に他カテゴリの前景領域と全画像の背景画像から抽出した特徴を加えてオフラインで学習した。

5. 実　　験

5.1 精度評価実験

5.1.1 データセット

本来は本システムはスマートフォン上で認識するため、実際にスマートフォン上で評価することが望ましく、実際と異なるが、[1] で使用されているデータセットから 50 種類の画像が 100 枚以上、合計 6,781 枚のデータセットを構築し使用する。このデータセットには、Ground Truth となるバウンディングボックスとそれに対応する料理名がラベル付けしてある。50 種類の料理は、食事のバランスがよくなるように選択した。図 4 は、本研究で対象にした 50 種類の料理のサンプルである。

5.1.2 実験設定と評価方法

精度評価は、以下の項目に対して行う。

（1） 各特徴量の分類性能の評価
（2） 特徴量結合時の評価
（3） 領域推定による分類性能の評価
（4） 料理のある方向提示の評価

1、2 では、前述のデータセットを用い、検証、評価用に各 20 枚、残りを学習に使用して、データを入れ換えて 5 回実験し、その平均値で評価した。2 により、特徴結合時の精度を示し、本システムの識別機を構築する。また、領域推定も適用した場合の精度を示す。

図 4　50 種類の料理のサンプル

3 では、ユーザが入力した領域が実際の料理領域より大きくとった場合で評価する。実験では、料理領域の幅か高さを一方に拡大し、背景を 25%含んだ場合で行う。評価用画像は、1 で分割した評価用サブセットに適用し、背景を十分含む合計 1,912 枚となった。ここでは、カテゴリごとの画像枚数が異なるので、いずれのカテゴリも背景情報による性能劣化は同程度であると仮定し、背景を含む場合での認識精度低下の程を評価する。

1、2、3 での評価方法は、以下に定義する分類率を用いて行った。

$$\text{分類率} = \frac{\text{候補 } N \text{ 位までに正解を含む画像枚数}}{\text{評価画像枚数}}$$

4 では、3 と同じ評価画像を使用する。料理を x%(x=10,15,20,25) 周囲にずらした領域に対して、料理のある方向を推定する精度を評価する。評価方法は、以下に定義する分類率を用いて行った。

$$\text{分類率} = \frac{\text{角度の差 } y \text{ 度以内になった画像枚数}}{\text{評価画像枚数}}$$

5.1.3 実験結果

（1） 各特徴量の分類性能の評価

正解領域に対して、各特徴量単体で分類した結果を図 5 に示す。

結果より、SURF-BoF は、soft 割り当てにより性能が向上し、特徴量は 900 から 950 個程度サンプリングしているが、この場合コードブックサイズを 500 から 1000 にするよりもコードブックサイズ 500 で χ^2 カーネルの写像を適用した方が性能がよい、カラーヒストグラムは各要素の平方根をとっただけでは、カラーオートコレログラムより性能が悪いが、χ^2 カーネルの写像を適用することで性能が大きく向上することがわかる。また、近年提案されているバイナリ記述子においても実験した

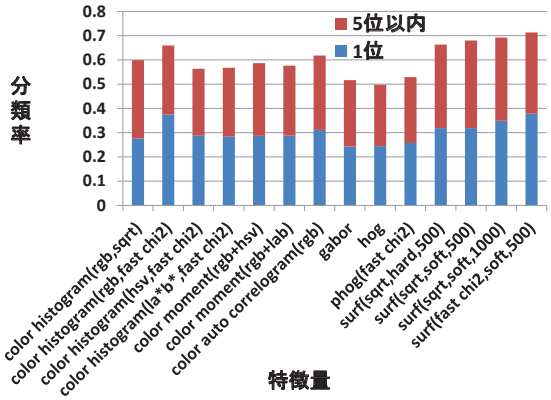


図 5　各特徴量の分類結果

が、精度が極めて悪かった。これは、BoF 表現にした際、情報量が少ないために量子化誤差が大きくなったことが考えられる。

結果を踏まえ本研究で使用する特徴量は、χ^2 カーネルの写像を適用した、カラーヒストグラムと SURF-BoF にした。

(2)　特徴量結合時の評価

特徴量の結合は、各識別器の出力値の重み付き和とする。カテゴリごとの識別器の重みは検証セットを用いて求めた。正解領域に対して、SURF のみ、カラーヒストグラムと SURF により分類した結果と、GrabCut による領域推定も行った場合で分類した結果を図 6 に示す。

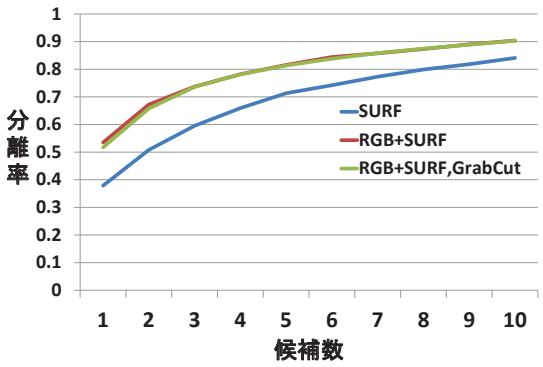


図 6　特徴量結合時の分類結果

結果より、正解領域が与えられたとき、領域推定をしない場合は 1 位に 53.5%、5 位以内に 81.6%の精度で、領域推定をした場合は 1 位に 51.7%、5 位以内に 81.4%の認識精度となり、領域推定すると少し下がる結果となった。本システムでは料理候補を 5 つまで表示可能なので、5 位以内の精度が重要となりこの場合 81.4%は認識可能であることがわかる。以後の実験では、カテゴリごとの評価画像枚数が異なるのでこの結果を本手法による領域が正しく与えられた場合の分類率とする。

(3)　領域推定による分類性能の評価

食事領域に背景を含むとき、領域推定をしない場合とした場合の分類性能の評価を図 7 に示す。

結果より、領域推定をしない場合は、25%背景を含む場合 1 位で 12.8%、5 位以内で 10.1%精度低下がみられたが、領域推定をすることにより 1 位で 4.1%、5 位以内で 3.1%の精度低下

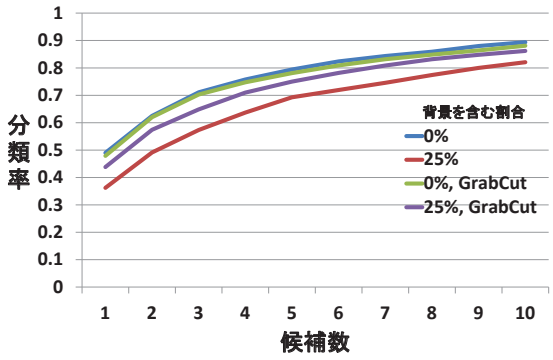


図 7　背景を含む場合の分類結果

で済むことがわかる。

(4)　料理のある方向提示の評価

まず、料理を 15%ずらしてその方向を提示する精度を hard 割り当てと soft 割り当てで評価した結果を図 8 に示す。

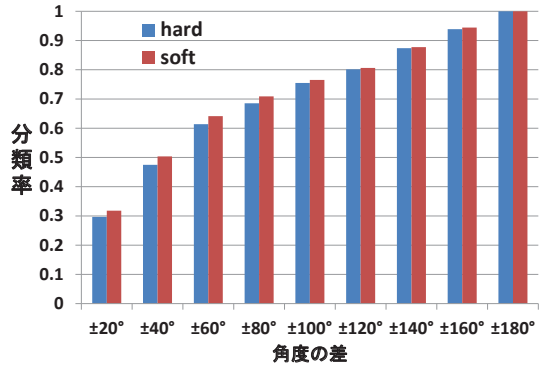


図 8　料理が 15%ずれた場合の hard 割り当てと soft 割り当てでの分類率

結果より、hard 割り当ては ±20° 以内に 29.7%、±40° 以内に 47.5%の精度で、soft 割り当ては ±20° 以内に 31.8%、±40° 以内に 50.3%の精度で料理のある方向を提示することが可能であることがわかる。±40° 以内までなら、ある程度正しく方向を提示できていると考えると、方向提示においても soft な割り当ての方が性能がよいことがわかる。そのため、soft 割り当てを使用することにする。

次に、料理を x%(x=10,15,20,25) ずらしてその方向を提示する精度を評価した結果を図 9 に示す。

結果より、料理のずれが大きいほど精度が高くなっていることがわかる。料理が 25%ずれた場合は、±20° 以内に 34.5%、±40° 以内に 54.2%の精度で料理の方向を示すことが可能である。

料理をずらして実験を行ったが、与えられた領域内にサブ領域を多数考え、それらの中で評価値最大となった方向を提示するので、料理がずれていなくても、料理らしさの高い領域を求め、その方向に動かすことにより認識結果が変化することを目的としている。

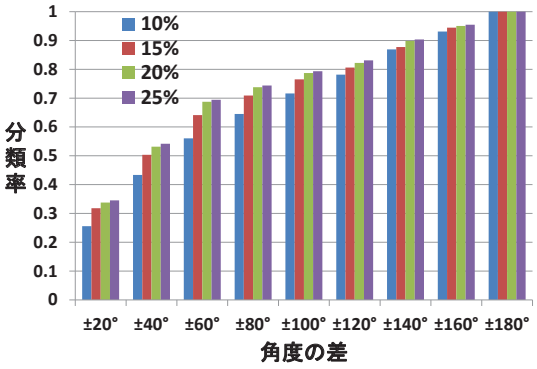


図 9　料理が x%ずれた場合の分類率 (x=10,15,20,25)

5.2　速度評価実験

5.2.1　評価デバイス

本研究では、スマートフォンの性能向上にも着目しているため、高性能なスマートフォンを用いて実験を行う。今回実験に使用したデバイスは Galaxy NoteⅡ(1.6GHz Quad Core Android4.1) である。

5.2.2　実　　装

本システムは、高速化のために今後一般になると考えられる 4 コアのデバイスを想定して、並列処理を行っている。特に、画像認識の場合は容易に並列可能な部分が多い。

そこで、本研究では、システムの画像処理をする部分の流れは図 10 のようにした。

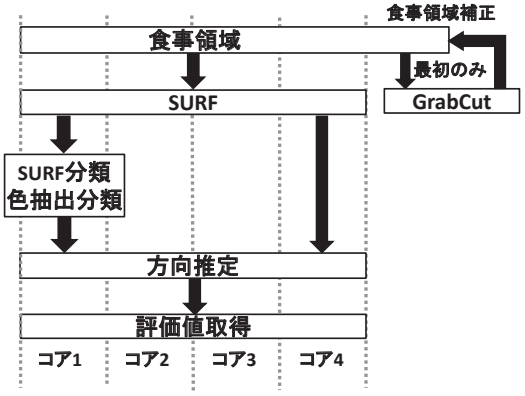


図 10　画像処理の流れ

最もコストの高い SURF の特徴記述と、コードワード割り当てを 4 コアで並列処理し、次に、コストの非常に小さい SURF を分類とカラーヒストグラム抽出分類はシングルコアで、また、それと方向推定を 4 コアで並列処理をした。GrabCut も初期モデルの作成を 2 並列で行っている。そして、SVM は、オフラインで学習しておき、近似最近傍探索のための kd-tree を構築した。

5.2.3　実 験 結 果

領域推定部分、認識部分、方向提示部分、認識と方向提示の部分の速度をそれぞれ 20 回計測し、その平均値を表 1 に示す。

なお、SURF-BoF の正規化と写像、分類、カラーヒストグラム抽出、分類は平均 0.003sec であり、処理の大部分は SURF

表 1　平均実行時間

	平均実行時間 [sec]
領域推定	0.70
認識	0.26
方向提示	0.091
認識+方向提示	0.34

の特徴記述と最近傍コードワード探索に要した。また、バックグラウンドで領域推定を行っている場合は、認識部分の平均実行時間は、0.31sec であった。結果より、複数領域が与えられても、それらの領域から評価値を得て、リアルタイムにリストに反映可能であることがわかる。

5.3　ユーザによる評価実験

5.3.1　実 験 設 定

被験者は学生 5 人である。1 食 3 品として 3〜4 食、各 2 回ずつ使用してもらい、システムの評価を得た。評価項目は、「認識の良さ」、「使いやすさ」、「方向提示のよさ」、「手動 or 本システム」として、5 をよい (本システム)、3 を普通とした 5 段階評価である。また、各食品選択までに要した時間を計測し、比較として階層型メニューによる手動での記録も同様に計測した。

5.3.2　実 験 結 果

各食品選択までに要した時間を図 11 に、5 段階評価の結果を表 2 にそれぞれ示す。

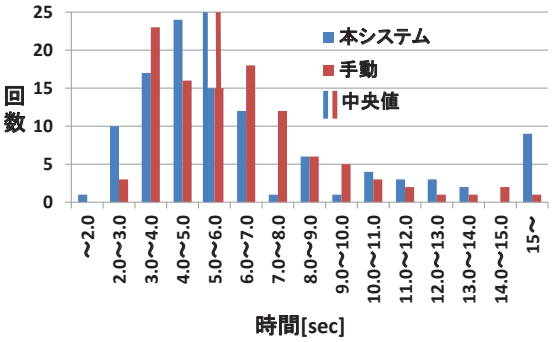


図 11　食品選択に要する時間

表 2　システム 5 段階評価

評価項目	平均点
認識のよさ	3.4
使いやすさ	4.2
方向提示のよさ	2.4
手動 or 本システム	3.8

食品選択に要する時間の本システムの中央値は 5.1 秒、手動は 5.7 秒であった。ユーザからのコメントは、「認識率が上がれば使ってみたい」、「不適当な結果の食品を除外する機能がほしい」、「認識対象を増やすか、別途登録できるようにしてほしい」などが挙げられた。

今回の場合は、手動よりも少し早く選択でき、使いやすいという評価を得た。しかしながら、本システムで認識できない食

品や時間が非常にかかる食品も存在し、この点に関して対処が必要であると考える。また、方向提示のよさも高い評価を得ることができなかった。これは、精度があまり高くないため示された方向に動かしても、期待する料理名がリストに提示されなかったことが考えられる。

6. おわりに

本研究では、スマートフォン上でリアルタイムに食事認識をする、ネットワークに依存しない食事管理システムを提案した。提案システムは 50 種類の料理に対して、背景を含まない料理の領域が与えられたとき、候補を 5 つ提示し 81.4%の認識精度であった。また、バックグラウンドでは料理の領域の補正を行い、さらに、認識を誤った場合を考慮し、ユーザに料理のある方向を提示する。認識する領域から料理が 15%ずれていた場合、角度差 $\pm 20^\circ$ 以内に 31.8%、$\pm 40^\circ$ 以内に 50.3%の精度で、25%ずれていた場合、角度差 $\pm 20^\circ$ 以内に 34.5%、$\pm 40^\circ$ 以内に 54.2%の精度であることを確認した。

今後は、現在はユーザ情報を使用していないので、ユーザ情報を収集して、それを識別機に反映させることにより、ユーザごとに特化した識別機を構築する。また、方向提示は高い評価を得ることができなかったので、形状を考慮するなど他の手法について考察する。さらに、2 次元方向の提示でなく、傾き等考慮した 3 次元方向の提示への拡張を目指す。

認識精度に関しては、次元圧縮を適用し、使用する特徴量などを追加することなどが考えられる。また、認識する料理の数を増やす。単純に増やしただけでは、認識性能は一般に悪くなるので、ユーザに認識する料理を選択できるようにし、ユーザが食べるが、認識対象にないという料理が少なくなるようにする。

文　献

[1] 松田裕司, 甫足創, 柳井啓司. 候補領域推定に基づく複数品目食事画像認識. 電子情報通信学会論文誌 D, Vol. J95-D, No. 8, pp. 1554–1564, 2012.

[2] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar. Food recognition using statistics of pairwise local features. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2010.

[3] K. Kitamura, T. Yamasaki, and K. Aizawa. Foodlog: Capture, analysis and retrieval of personal food images via web. In *Proc. of ACM Multimedia Workshop on Multimedia for Cooking and Eating Activities*, pp. 23–30, 2009.

[4] A. Mariappan, M. Bosch, F. Zhu, C.J. Boushey, D.A. Kerr, D.S. Ebert, and E.J. Delp. Personal dietary assessment using mobile devices. In *Proc. of the IS&T/SPIE Conference on Computational Imaging VII*, Vol. 7246, pp. 72460Z–1–72460Z–12, 2009.

[5] N. Kumar, P. Belhumeur, A. Biswas, D. Jacobs, W. Kress, I. Lopez, and J. Soares. Leafsnap: A computer vision system for automatic plant species identification. In *Proc. of European Conference on Computer Vision*, 2012.

[6] T. Maruyama, Y. Kawano, and K. Yanai. Real-time mobile recipe recommendation system using food ingredient recognition. In *Proc. of ACM Multimedia Workshop on Interactive Multimedia on Mobile and Portable Devices*, pp. 27–34, 2012.

[7] T. Lee and S. Soatto. Learning and matching multiscale template descriptors for real-time detection, localization and tracking. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2011.

[8] F.X. Yu, R. Ji, and S.F. Chang. Active query sensing for mobile location search. In *Proc. of the 19th ACM International Conference on Multimedia*, pp. 3–12, 2011.

[9] 香川芳子. 新毎日の食事のカロリーガイドブック　外食編／ファストフード・コンビニ編／市販食品編／家庭のおかず編. 女子栄養大学出版部, 2002/05.

[10] M. Bosch, F. Zhu, N. Khanna, C.J. Boushey, and E.J. Delp. Combining global and local features for food identification in dietary assessment. In *Proc. of IEEE International Conference on Image Processing*, pp. 1789–1792. IEEE, 2011.

[11] H. Hoashi and K. Yanai. Image recognition of 85 food categories by feature fusion. In *Proc. of The second Workshop on Multimedia for Cooking and Eating Activities*, 2010.

[12] 李賀, 杉山春樹, 相澤清晴. 一般食事画像認識に対する特徴量・認識手法の比較検討. 画像の認識・理解シンポジウム (MIRU2012), 2012.

[13] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih. Image indexing using color correlograms. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 762–768, 1997.

[14] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of IEEE Computer Vision and Pattern Recognition*, Vol. 1, pp. 886–893. IEEE, 2005.

[15] A. Bosch, A. Zisserman, and X. Munoz. Representing shape with a spatial pyramid kernel. In *Proc. of the 6th ACM International Conference on Image and Video Retrieval*, pp. 401–408, 2007.

[16] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346–359, 2008.

[17] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[18] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, Vol. 9, pp. 1871–1874, 2008.

[19] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 3, pp. 480–492, 2012.

[20] F. Perronnin, J. Sánchez, and Y. Liu. Large-scale image categorization with explicit data embedding. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 2297–2304, 2010.

[21] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *Proc. of ACM SIGGRAPH*, pp. 309–314, 2004.

[22] C. H. Lampert, M. B. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2008.