

料理画像認識を用いたモバイル食事記録システム

河野 憲之¹ 柳井 啓司¹

概要：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，スマートフォンの計算資源のみを用い，スマートフォン上でリアルタイムに料理画像認識を行い，ユーザの食事記録の補助を行うシステムを提案する．それは，カラーヒストグラムと Bag-of-SURF を fast χ^2 kernel SVMs で 50 種類の料理を分類し，また，認識する領域をユーザが与えるとバックグラウンドで GrabCut によるその領域の補正と，認識結果が誤った場合を考慮し，SVM の評価値に基づき認識結果が向上する方向を写すように指示するシステムである．実験では正しい料理領域が与えられ料理の候補を 5 つ提示した場合，81.55% の分類率を達成した．また，ユーザによるシステム評価を行い提案システムの有用性を検証した．

キーワード：スマートフォン，料理認識，食事記録，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から利用できる食事記録システムが多く現れるようになった．しかしながら，一般的な階層型メニューより選択する方法は，毎日複数回使用することを考えると手間が大きく，継続した利用が難しい．また，画像認識を用いて食事記録を取ることのできるシステムも提案されているが，いずれのシステムにおいてもサーバに画像を送信し，そこで画像処理を行うため，ユーザ数の増加に対して多数の計算機が必要になる，デバイスとサーバ間での通信コストがかかる，ネットワーク環境によっては利用できない，という問題が存在する．

また，近年クアッドコアの CPU を搭載したスマートフォンが一般的になり，世代前の PC と同程度のスペックにまでなっている．そのため，従来はリアルタイムに動作させることができなかった処理も複数コアを用いることにより高速に処理させることが可能になった．

そこで本研究では，近年のスマートフォンの性能向上に着目し，スマートフォン上でリアルタイムに，また χ^2 カーネルの近似線形化手法を適用し高精度に食事画像認識を行い，食事記録をとる補助を行うシステムを提案する．さら

に提案システムはリアルタイムに画像認識を行うと共に，認識結果が誤っていた場合を考慮し，次にどの方向を写すと認識結果が期待されたものになるか提示する．また，計算コストの高い，認識する領域の決定はユーザに入力してもらおうが，バックグラウンドでその領域の補正を行う，ユーザインタラクティブなシステムである．

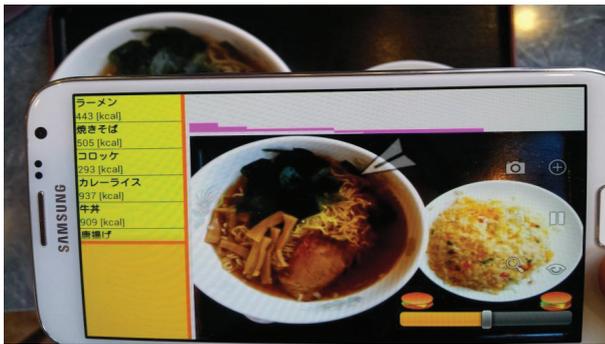
図 1 は提案システムのイメージである．

2. 関連研究

料理画像認識は料理に決まった形はなく，同じ料理であっても視覚的变化が大きいため難しいタスクである．松田ら [1] は複数検出手法により料理領域を推定後，SIFT，HOG，Gabor，カラーヒストグラム特徴を MKL-SVM により 100 種類の食事画像に対して分類を行った．Yang ら [2] は，画素間の距離や角度等，材料の位置関係の特徴量とする手法により，ファストフードの分類を行った．本研究でも，料理認識を行い，料理を分類して結果上位をユーザに提示する．

食事記録システムとしては，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記録システムには，食事画像からバランス推定を行い，その結果を返す FoodLog[3] や，チェックボードとともに食事を撮影し，食事の分類と量の認識

¹ 電気通信大学 情報工学科．現在，同大学院 情報理工学研究科 総合情報学専攻所属



(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

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、画像処理をしているため通信コストが高く、一度のみの認識となり認識を誤った場合は、ユーザが後から手で直すことになる。本研究でも、画像認識を用いた食事記録システムを構築するが、スマートフォ上でリアルタイムに認識、その場で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、料理の量はユーザに入力してもらい料理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。物体認識システムとして有名な Google Goggles*1 は、ロゴや建造物など特定物体を認識するアプリケーションである。また、Kumarら [5] の一定条件下で葉の画像をサーバに送り、葉独自の湾曲具合から特徴量を抽出し、葉を認識するアプリケーション Leaf snap や、Maruyama[6] らの色特徴を抽出し 30 種類の食材を認識しレシピを返すアプリケーション、Leeら [7] のビデオからテンプレートを学習しそれを強度と方向の記述子に分解し、テスト時には複数スケールでのテン

*1 <http://www.google.com/mobile/goggles/>

プレートマッチングによりリアルタイムに物体検出や追跡を行うアプリケーションが提案されるなど近年モバイルと画像認識の研究が盛んに行われている。本研究では視覚的变化の大きい料理に対してスマートフォンの計算資源のみを用いてリアルタイムに料理認識を行う。

ユーザインタラクティブなシステムとしては Yu ら [8] の研究が挙げられる。Yu らはモバイル位置検索で認識を誤った場合を考慮し、次にどの視点を撮影すればよいかを、オフラインで求めた顕著性とオンラインで画像マッチング、Gist+SVM により求め、それをユーザに返す Active Quert Sensing (AQS) を提案した。本研究でも、処理に時間のかかる部分はユーザに補助してもらい、また料理があると推定された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) ユーザはシステムを起動し、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。入力後バックグラウンドで領域補正が行われる。
- (3) 料理認識を行う。料理領域が入力されていれば逐次料理領域に対して認識を行い、入力されていなければフレーム全体で料理認識を行う。
- (4) 一定時間後、認識結果上位を提示する。この一定時間の間料理認識を繰り返し、各領域の各料理のスコアはそれぞれの認識結果の平均とし、最終的に全スコアをソートし、評価値の高い料理名から順にリストに表示する。
- (5) ユーザは認識結果上位から料理を選択する。結果が誤っていた場合は、料理性の高い方向が提示されているので、参考にスマートフォンを動かし見えを変化させる。
- (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。ユーザは、登録した食事のバランスを閲覧可能である。

記録にはメモや位置情報も登録でき、サーバにアップ

スマートフォンを食事にかざす



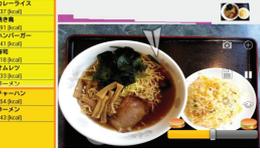
認識を始める



料理領域を入力する



料理を選択する
食事画像を保存する



食事記録を登録する



図 2 使用の流れ



(a) 食事記録



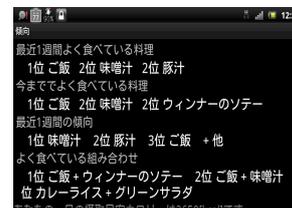
(b) 栄養素詳細



(c) 食事画像



(d) Map



(e) 食の傾向



(f) 食事記録 (Web)

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

記録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する。食事記録閲覧画面において、1日の食事の評価を新 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 [9] より 4 群点数、またユーザ情報設定により設定された性別、年代、運動習慣から日本人の食事摂取基準*2より摂取目安カロリーを参照し、5 つ星で行い表示している。
- (2) Google Maps 上で閲覧する。記録した食事画像に位置情報を付与することで Map 上で閲覧可能である。
- (3) 最近の食の傾向の確認。最近食べた回数の多い料理や組み合わせ、過去の履歴から各料理の頻度を求めそれと最近の食べた料理の回数から食の傾向を表示している。
- (4) アップロードされた食事記録の閲覧。記録した食事画像をサーバにアップロードすることにより、食事記録を共有閲覧可能である。



(g) Map (Web)

図 3 記録閲覧

4. 認識手法

4.1 領域補正

特徴抽出する領域はユーザにより正しく与えられることが理想であるが、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ことは手間であり、また背景を多く含む場合認識精度は一般に低下する。

そこで、本研究ではユーザは料理全体を含むように領域を入力するという制約を与え、領域分割手法により料理領域を推定し、認識する領域の補正を行う。領域分割手法には、矩形内の画素値を前景と背景に分離する GrabCut を適用することにより料理領域の補正を行う。

4.1.1 GrabCut

GrabCut[10] は、与えられた矩形領域内は前景と背景が混在するモデルとして色分布から GMM (Gaussian Mixture Model) を作成する。そして、作成した GMM を基に各画

*2 厚生労働省 「日本人の食事摂取基準 (2010 年度版)」 p.61

素の前景らしさ、背景らしさの尤度を求めそれにより求めた前景、背景ラベルを GrabCuts のシードラベルとすることにより領域分割を行う。

認識を行うごとに GrabCut を適用するにはコストが大きく、ユーザが料理領域を入力後、料理認識を行うと同時にバックグラウンドで一度のみ認識領域の補正を行う。また、動画でリアルタイムに認識を行うためカメラの位置は固定ではない。そのため、領域補正を行う入力画像と領域補正後の画像はずれがあることが想定される。従って、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した。

4.2 特徴量

本システムは、モバイルデバイス上で画像処理を行うため、高次元特徴量などを多数使用することはできない。そこで本研究では、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、ガボール、HOG、PHOG、Bag-of-SURF より有効であったカラーヒストグラムと Bag-of-SURF を画像特徴量として使用する。また、計算コストを抑えるために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りサイズした。

4.2.1 カラーヒストグラム

画像を 3×3 に分割し、それぞれの領域から 64 次元 RGB カラーヒストグラムを抽出し結合することで 576 次元特徴ベクトルを得た。

4.2.2 Bag-of-SURF

SURF[11] は照明変化、スケール変化、回転に対して頑健な 64 次元特徴量である。SURF を 8 ピクセルごとにスケールを 12, 16 ピクセルとして dense sampling し、Bag-of-Features 表現で 500 次元の特徴ベクトルを得た。

また、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 [12] を行い、最近傍コードワード探索は、近似最近傍探索により行った。

4.3 分類器

分類器に線形 SVM を用い、one-vs-rest 法により多クラス分類を行う。線形カーネルは $K(\mathbf{x}, \mathbf{z}) = \langle \mathbf{x}, \mathbf{z} \rangle$ で表される。線形 SVM は、入力ベクトルを \mathbf{x} 、出力値を $f(\mathbf{x})$ 、サポートベクトルを \mathbf{x}_i 、サポートベクトルの重みを α_i 、サポートベクトルの数を M 、クラスラベルを $y_i \in \{+1, -1\}$ 、バイアス値を b としたとき、

$$f(\mathbf{x}) = \sum_{i=1}^M y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (1)$$

$$\begin{aligned} &= \sum_{i=1}^M y_i \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle + b \\ &= \left\langle \sum_{i=1}^M y_i \alpha_i \mathbf{x}_i, \mathbf{x} \right\rangle + b \\ &= \langle \mathbf{w}, \mathbf{x} \rangle + b \end{aligned} \quad (2)$$

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、必要なメモリ量も $\mathcal{O}(N)$ である。

SVM の学習には、LIBLINEAR[13] 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。

4.4 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少しだけ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 [14] は、Hellinger, χ^2 , intersection, Jensen-Shannon (JS) の任意の additive 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。

$$\phi_{\omega} = \kappa_{\omega} \sqrt{x} e^{-i\langle \omega, \log x \rangle} \quad (3)$$

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 $e^{i\omega \log x} \sqrt{x \operatorname{sech}(\pi\omega)}$ で表される。また、特徴ベクトルの次元数を d とすると、写像後の次元数は近似のオーダー n に対し、 $d \times (2n + 1)$ 次元になる。本研究では $n = 1$ とした。 $n = 1$ のとき写像の一つは以下のように表現できる。

$$\phi(x) = \sqrt{x} \begin{bmatrix} 0.8 \\ 0.6 \cos(0.6 \log x) \\ 0.6 \sin(0.6 \log x) \end{bmatrix} \quad (4)$$

そして、正規化が正しく行われるため L1 正規化したヒストグラムと相性がよいのでカラーヒストグラム、Bag-of-SURF とともに適用した。

4.5 方向提示

料理が正しく認識されない場合、料理の見え方を変更しなければ評価値は変わらず、目的の料理は一向にリストに現れない。そこで、料理を認識する際、料理を認識し易い方向を推定し提示することにより、次の認識ではその領域を写すようにユーザに促す。結果として、認識結果が変わり目的の料理がリストに提示されることが期待される。

本研究では、Bag-of-SURF を直接線形 SVM に適用した評価値による効率の良い Window 探索を用いる。BoF と線形 SVM の評価値を用いて物体検出を行う手法には ESS (Efficient SubWindow Search) [15] が挙げられる。BoF を直接線形 SVM に適用した評価値の場合、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 の要素を累積することに相当する。また、式 2 の w は、

$$w = w^+ + w^- \quad (5)$$

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、 $O(1)$ の計算量で得ることができる。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 Window 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であり少なくとも互いの 1 辺が重なるように Window 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た。ここで、 $B = 2x$ ($3 \leq x \leq 6, x \in \mathbb{N}$) とした。Window ごとに各カテゴリの評価値が得られるが、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 や図 2 の認識中の画像において料理領域や全体画像の上部に矢印が提示されている。

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

本来は本システムはスマートフォン上で画像処理を行うため、実際にスマートフォン上で評価することが望ましいが、[1] で使用されているデータセットから各 100 枚以上ある 50 種類の画像、合計 6,781 枚からなるデータセットを構築し精度評価実験を行った。このデータセットには、正解領域となるバウンディングボックスとそれに対応する



図 4 50 種類の料理のサンプル

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。

5.1.1 実験設定と評価方法

精度評価実験は、以下の項目に対して行う。

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

1 では、正解領域が与えられた 50 種類の料理について検証、評価に各 20 枚、残りを学習に使用して、データを入れ換えて 5 回実験した平均値により評価した。

初めに、使用する特徴量を他のいくつかの特徴量について性能を比較し、次に特徴量を結合時の評価を行う。比較手法としては、 χ^2 カーネルの近似線形化手法を適用した特徴量単体 (Color (fast chi2), SURF (fast chi2)), 非線形 χ^2 -RBF カーネルをサブカーネルとした MKL-SVM (Multiple Kernel Learning) (Color+SURF (chi2-RBF)), feature map を適用せずに直接線形 SVM に適用した場合 (Color+SURF (linear)) とする。

2 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した。1 で使用した 5,000 枚の評価画像のうち、料理領域の幅が高さを一方に拡大した際に、背景を 25% 以上含む合計 1,912 枚の画像を使用した。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した。

1, 2 での評価方法は、以下に定義する分類率を用いて行った。

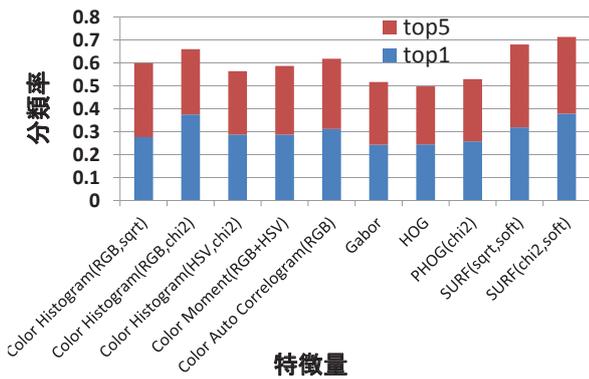


図 5 各特徴量での分類率

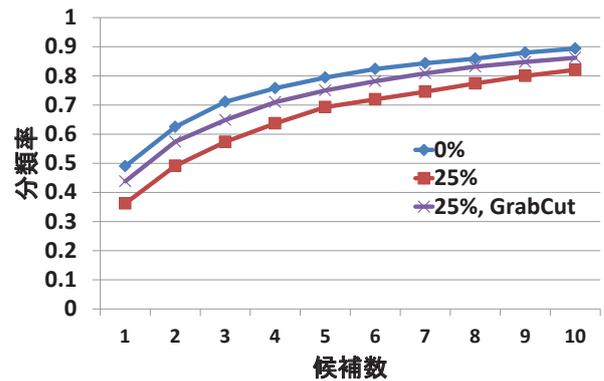


図 7 背景を含む場合の分類結果

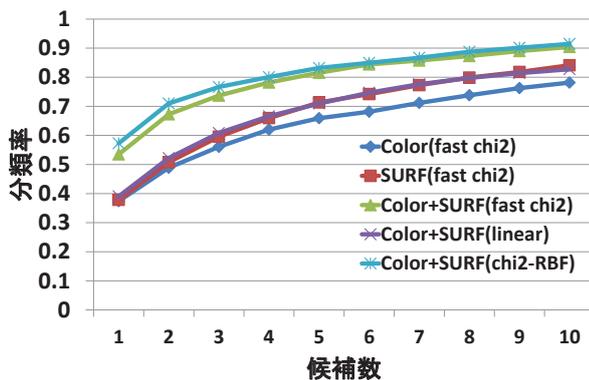


図 6 各手法による分類率

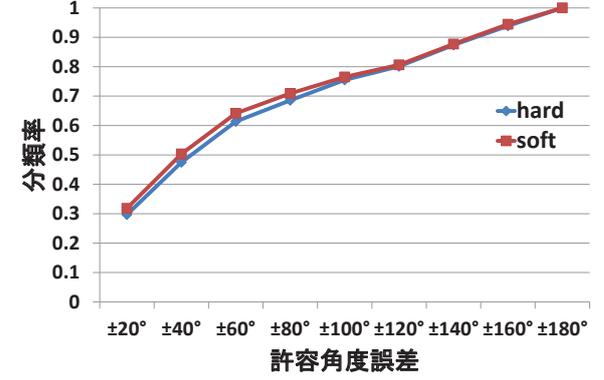


図 8 料理が 15%ずれた場合の hard 割り当てと soft 割り当てでの分類率

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

3 では、2 と同じ評価画像を使用した。料理領域を 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した。評価方法は、以下に定義する分類率を用いて行った。

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

5.1.2 実験結果

(1) 特徴量の分類性能の評価

各特徴量単体での分類結果を図 5 に示す。図 5 より、 χ^2 カーネルの近似線形化手法によりカラーヒストグラムと Bag-of-SURF の分類性能が大きく向上し、相補的な関係にあることから本システムではこれらの特徴量を採用した。

次に、特徴量結合時の分類結果を図 6 に示す。図 6 より、特徴量を結合することで精度が向上し、正解領域が与えられたとき 1 位に 53.5%、5 位以内に 81.6% の認識精度であった。また、非線形 χ^2 -RBF カーネルの場合は、1 位で 57.3%、5 位以内で 83.2% の認識精度であった。構築したシステムでは料理候補を 5 つまで表示可能であり、ユー

ザは提示されたりリストから選択するため、5 位以内の認識精度が重要であり、5 位以内と比較すると 1.6% の認識精度の差のみであり、速度とメモリの面で線形 SVM の方が大きく有利であることからモバイル上で画像認識を行う本研究の場合には適した手法であることがわかる。

(2) 領域補正による分類性能の評価

食事領域に背景を含むとき、領域補正を行わなかった場合と行った場合の分類率を図 7 に示す。

図 7 より、領域補正を行わなかった場合は、背景を 25% 含む場合 1 位で 12.8%、5 位以内で 10.1% の精度低下が見られたが、領域推定を行うことにより 1 位で 5.2%、5 位以内で 4.4% の精度低下で済み、有効性が示された。背景を含まないとき、領域補正を行った場合は分類率に変化はあまり見られなかった。これは、GrabCut 適用後、矩形領域で料理領域を表現していることが考えられる。

(3) 料理のある方向提示の評価

料理領域を 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 より、hard 割り当てでは $\pm 20^\circ$ 以内に 29.7%、 $\pm 40^\circ$ 以内に 47.5% の精度で、soft 割り当てでは $\pm 20^\circ$ 以内に 31.8%、 $\pm 40^\circ$ 以内に 50.3% の精度で料理のある方向を提示可能であるこ

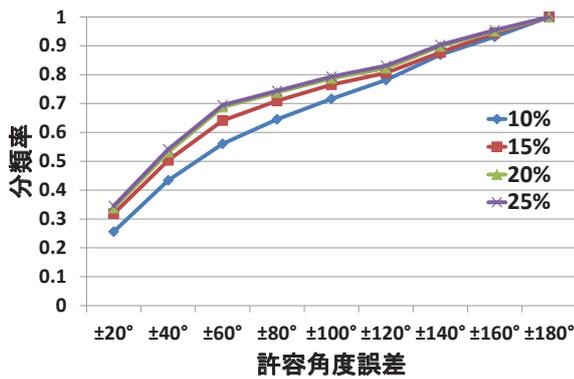


図 9 料理が 10,15,20,25%ずれた場合の分類率

とがわかる．従って，本システムでは高い精度となった soft 割り当てを使用した．

次に，料理領域を 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．

図 9 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．

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．

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．

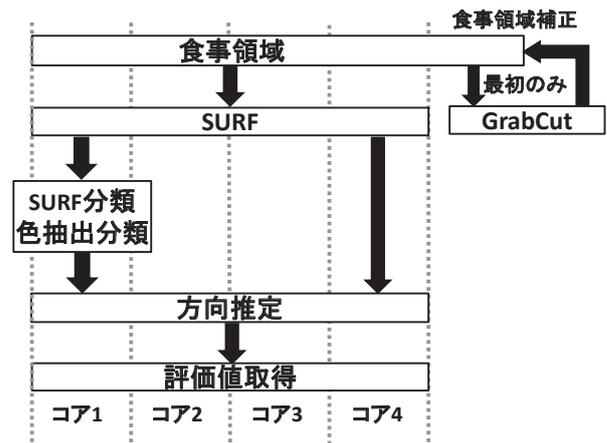


図 10 画像処理の流れ

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

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

処理の大部分は SURF 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．表 1 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を更新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．

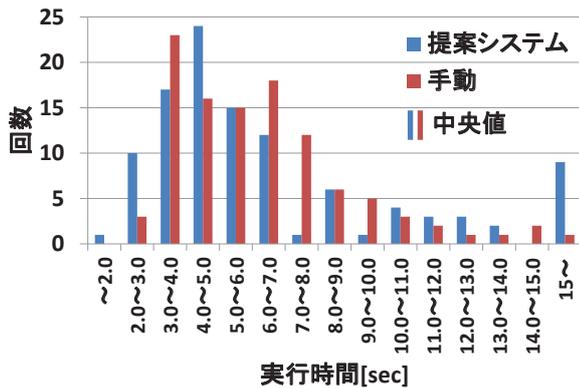


図 11 食品選択に要する時間

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。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、精度が現状では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記録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、正しい料理領域が与えられたとき、候補を 5 つ提示し 81.6% 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インタラクティブなシステムである。

今後は、現在はユーザ情報を使用していないので、本システムを使用するごとにユーザ情報を収集し、それを動的に識別機に反映させることにより、ユーザ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ため、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。そして、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、ユーザが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。

参考文献

- [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] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *Proc. of ACM SIGGRAPH*, pp. 309–314, 2004.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.