情報処理学会研究報告
IPSJ SIG Technical Report

候補領域推定による食事画像の複数品目認識

甫足　創[†1]　柳井啓司[†1]

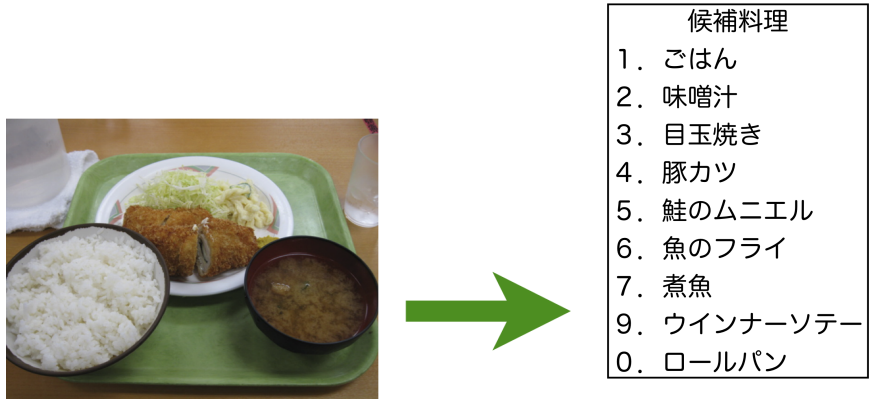
本研究では食事内容を少ない手間で記録するために，画像認識技術を用いて，画像中に含まれると推測される料理名の候補を表示する認識エンジンを構築した．我々は以前，Multiple Kernel Learning を用いて，色特徴や局所領域特徴等の複数の特徴を統合して学習・分類を行なう認識エンジンを構築した．本研究では食事画像認識手法の改良を行う．高速スライディングウィンドウ探索や領域分割，円検出を用いて，画像中の料理の位置候補を推定し，その部分に対して従来の手法による分類を行うことで，画像中に複数の料理がある場合に対応した．実験では，85 種類の料理について分類を行い性能を評価を行った．その結果，複数の領域検出法を組み合わせて用いることで，10 個の候補を表示したとき，単品を含む画像では，従来手法と比べ 2.6 ポイント向上し，71.6%，複数品を含む画像では，従来手法と比べ 39.0 ポイント向上し 60.2%の分類率を達成し，特に複数品を含む食事画像の認識において提案手法は有効であることが示された．

Recognition of multi-food images by detecting candidate regions

Hajime Hoashi[†1] and Keiji Yanai[†1]

In this paper, we propose a method to recognize multi-food images by detecting candidate regions with several methods. The proposed method is based on our previous work on food-image recognition which assumes that one image has only one food. We detect several candidate regions by fusing output of several region detectors including the efficient sliding window search (ESS), a circle detector and the JSEG region segmentation.

In the experiments, we estimated ten food categories for one multi-food image in the descending order of confidence. As results, we have achieved the 60.2% classification rate, which improved our previous method by 39.0 points. This demonstrates that the proposed method is effective for recognition of multi-food images.



(a) 入力画像　(b) 結果表示

図 1　本研で構築した認識エンジンは入力画像から，料理の候補推定し出力する．

1. はじめに

近年，携帯電話やスマートフォン等の情報端末を利用して食事記録をとるサービスが普及しつつある．食事情報を記録することで，食生活について意識したり，栄養分の評価を行うことができる．食事情報を記録する際の一般的な方法として，ユーザーがテキストを入力し，サービスに登録してある食べ物を検索する方法や，サービスに登録してある食べ物から階層的なリンクを用いて選択する方法が挙げられる．それらは，摂取した食品毎に登録をする必要があり，複数品目の料理を毎食記録するのは特に手間が大きい．そこで，より手軽に，より短時間で食事の記録をとる方法が望まれている．

本研究では，食事内容を少ない手間で記録するために，画像認識技術を用いて，画像中に含まれると推測される料理名の候補を表示する認識エンジンを構築した (**図 1**)．

2. 関連研究

食事画像の認識に関する関連研究としてとして，FoodLog[⋆1]では，画像から得られる画像

†1 電気通信大学
The University of Electro-Communications
⋆1 http://www.foodlog.jp

特徴を用いて，栄養を直接推定している．この方法は，どのような種類の料理でも認識対象にすることもできるが，認識結果が本当に正しいかどうかは，知識のないユーザーには理解しづらい．それに対して，本研究では料理の種類を認識してユーザーの記録のサポートを行い，その後栄養を計算するというアプローチを採っている．

Yang ら[1] は，野菜やパンや肉などの材料の位置関係を特徴ベクトルとする事で，61 種類のファーストフードの分類に取り組み，28.2%の精度で分類する事ができた．また，Zong ら[2] も同様のファーストフードデータセット[3] に対して，SIFT 特徴点検出と Local Binary Pattern 記述子を用いた分類で，ベンチマークよりも良い分類精度を出した．我々の研究では，ファーストフードだけでなく，日本でよく食べられている物を中心に認識が行われるように認識対象の料理の種類を調節した．

我々は以前から食事画像認識について研究をしている[4]．この研究では，9 種類の視覚的特徴を Multiple Kernel Learning(MKL) を用いた方法で効率良く特徴を統合することで，50 種類の料理について 61.34%の割合で正しく分類する事が出来た．この研究で対象にした画像は，画像全体に一品の対象の料理が大きく写っている物に限定されていた．本研究では，複数の料理が写った画像や，対象の料理が大きく写っていない物も対象として考慮した．

Vedaldi ら[5] は，計算量の少ない線形カーネルを用いた SVM(Support Vector Machine) によってウィンドウサーチを行い，候補領域を絞り込む方法と，高い精度が得られる非線形カーネルを用いた SVM の組み合わせにより，高い精度で画像中のオブジェクトの位置を特定した．本研究においても，画像中にある料理の位置の候補を検出し，その検出領域に対して非線形カーネルを用いた SVM による精度の高い評価値を得る方法の組み合わせによって画像中にある料理の種類の候補を推定する．本研究では線形カーネルを用いた SVM によるウィンドウサーチでの候補領域に加えて，円検出や領域分割アルゴリズムによっても候補領域を検出した．

3. 画像認識手法

本研究では，画像認識の手法を用いて，画像データからその画像中に含まれる料理の候補を出力する認識エンジンを作成した．認識の流れの概要を**図 2** に示す．認識対象の画像が与えられたら，まず，画像中に含まれる料理の候補領域を検出する．本研究では，候補領域検出手法として，画像全体，線形 SVM によるウィンドウーチ，円検出，領域分割の 4 つの手法を用いた．次に，各手法で検出された候補領域を統合する．このとき，候補領域としてふさわしくない形の領域を除外する．選定された領域に対して，画像特徴ベクトルを生成し，非線形カーネルを用いた SVM による分類を行う．最終的に分類の評価値を降順にソートして，上位 N 個の結果を返す．

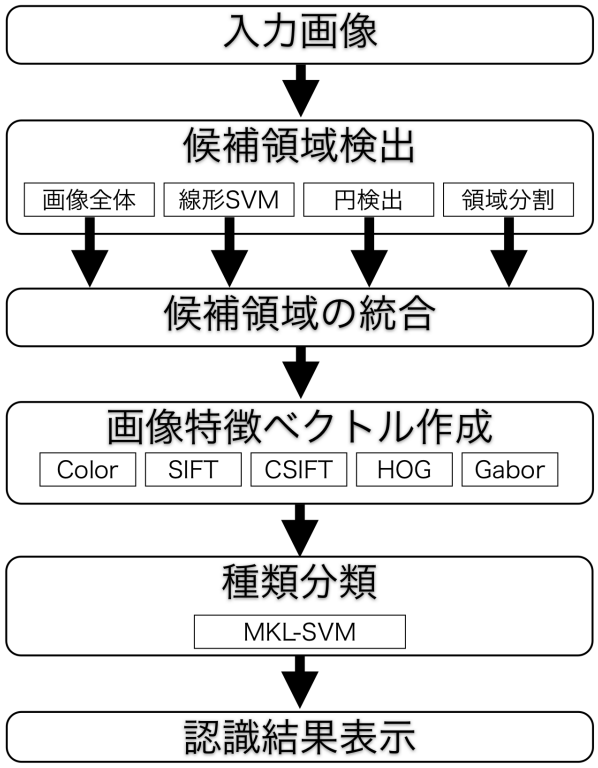


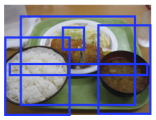
図 2 認識の流れ

本節では，画像認識手法の詳細について記述する．

3.1 候補領域検出

空間ピラミッド表現による特徴ベクトルを用いる手法や，以前の研究で用いた，Multiple Kernel Learning による特徴統合を利用した SVM による分類は背景情報が少ない時，効率よく働く事が知られている．本研究では，分類の精度を上げるために，特徴ベクトルを作成

表 1 本研究で用いる候補領域検出法

	画像全体	線形 SVM	円検出	領域分割
				
候補領域数	1 つ	各料理クラスにつき 1 つ	4 つ程度 (平均)	14 つ程度 (平均)
長所	大きいオブジェクトに有効	評価値の高い矩形領域が得られる.	皿の領域が捉えられる.	領域で料理を捉えられる.
短所	小さいオブジェクトに不向き	線形 SVM での認識の精度はあまり良くない.	円形以外の皿はうまく捉えられない.	似た色が同じ領域になってしまう.

する前処理として，候補領域の推定を行い，背景情報の少ない矩形領域を得る．本研究で用いる候補領域検出法を**表 1** にまとめる．

3.1.1 画 像 全 体

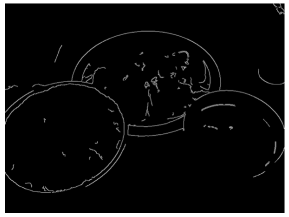
単純な候補領域として，画像全体を候補の一つとする方法が挙げられる．これは，画像に一つの料理が大きく写っている事を期待した候補領域である．そのため，大きく写っている料理には有効だが，小さく写っている料理には不向きである．以前の研究[4] では大きく写っている料理を対象にしていたので，この領域のみを利用していた．

3.1.2 線形 SVM によるウィンドウサーチ

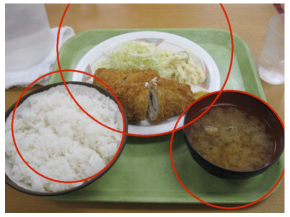
位置検出の手法として，有名な手法に Sliding Window 法があげられる．Sliding Window 法は，入力画像中の矩形領域に着目し，その位置と大きさを変化させながら，評価関数を用いて探索を行う．結果として評価関数の値が最大の矩形領域を得る．しかし，すべての矩形領域を探索するので，一辺の長さを n とする時，評価回数は $O(n^4)$ となってしまう．そこで，本研究では，分枝限定法を用いた効率的な物体の位置検出アルゴリズムである Efficient Subwindow Search(ESS)[6] を用い，効率的に評価関数が最大となる領域を求める．

ウィンドウサーチに用いる評価関数は，線形カーネルを用いた Support Vector Machine(SVM) を利用し，各種類の料理につき，評価関数が最大となる領域を検出する．線形カーネルを用いた SVM は，非線形カーネルを用いた SVM と比べ高速に評価できるが，最も高い評価値を出す領域であるとは限らない．

(a) 入力画像



(b) 輪郭抽出



(c) 円検出

図 3 円検出の流れ

3.1.3 円 検 出

画像から円形の輪郭を抽出する事で，皿の領域を検出し，それを候補領域とする．

まず，入力画像をグレースケール画像に変換し，Canny Edge Detector により，輪郭を抽出する．抽出された輪郭に対して，Hough 変換による円検出を行うことで，画像から円形の輪郭を抽出する (**図 3**).

3.1.4 領 域 分 割

領域分割とは，似た色を持つ領域に画像を分割する事である．本研究では，領域分割アルゴリズムとして JSEG[7] ⋆1を用いた．JSEG では，色空間の量子化を行い，カラークラスマップを作成することで，空間分割を行う．JSEG では，パラメータとして分割後の領域数を設定することができる．本研究では，画像をおよそ 10 個の領域に分割し，候補領域とした．

また，領域分割によって得られた領域の 2 つを結合した時の円形度が，結合された 2 つの領域より大きくなる場合，結合した領域も料理の候補領域とする (**図 4**)．円形度とは，領域がどの程度円に近いかを示す指標である．円形度は領域の面積を S，領域の周囲長を L とした場合，$(4\pi S)/L^2$ で求められ，この値は最大 1 となり，大きいほど円形に近い．

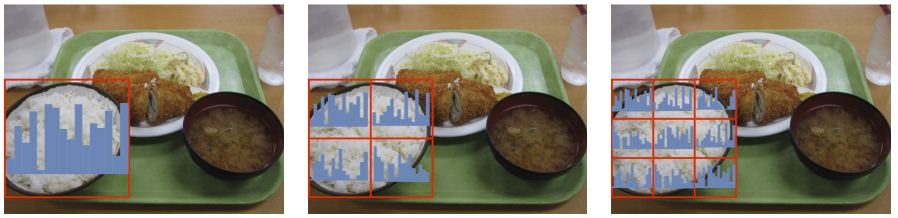
3.2 候補領域の統合

それぞれの手法で検出した候補領域は以後の処理で同等に扱うために，検出領域を含む bounding box として統合した．このとき，各手法で検出した候補領域に対して，領域の形を調べ，明らかに間違っている候補領域を除去する事で，分類にかかる計算コストを削減し，かつ，ノイズとなる評価値も削減する．本研究では，検出された候補領域の短辺が 60

⋆1 http://vision.ece.ucsb.edu/segmentation/jseg/software/

(a) 入力画像　(b) 領域分割の結果　(c) 領域の結合

図 4　領域分割での候補領域検出



(a) level 1　(b) level 2　(c) level 3

図 5　各ピラミッドレベルで Bag-of-keypoints での特徴ベクトルの作成

ピクセル以下の物は小さすぎる領域として候補領域から除外する．さらに，学習画像から各種類の料理の縦横比の平均と標準偏差を計算しておき，縦横比の値が平均値を中心として標準偏差の ±2 倍以内の範囲から外れている，縦横比が極端なものを候補領域から除外する．

3.3　画像特徴ベクトルの生成

料理画像を視覚的特徴から認識するためには，色特徴や SIFT 特徴を単純に利用するだけでは良い結果は得られない[1),4)]．そこで，本研究では以前の研究[4)] と同じく，複数の視覚的特徴を効果的に統合して利用する．本節では本研究で利用した画像特徴について記述する．

3.3.1　Dense sampling による局所特徴

局所特徴とは，画像中の小領域の特徴ベクトルである．本研究では，dense sampling によって，10 ピクセル毎に，半径 8 ピクセルと 16 ピクセルの二つのサイズに対して局所領域を選択した．各局所領域に対して，カラーヒストグラム，局所画像パターンである SIFT，照明変化に頑強な色空間に対して SIFT を得る CSIFT[8)]，の 3 種類の特徴を抽出した．

3.3.2　空間ピラミッド表現

抽出された局所特徴は空間ピラミッド表現[9)] を用いて，各領域の画像特徴ベクトルとした．空間ピラミッド表現では，認識対象とする領域を階層的にグリッドで分割し，それぞれのグリッドに対して，Bag-of-keypoints 表現を用いて，局所特徴の出現頻度の特徴ベクトルを作成する事で，局所特徴の空間情報も考慮した特徴ベクトルを得ることができる．ピラミッドレベル l では，画像を $l \times l$ のグリッドに分割する (**図 5**)．本研究では，Bag-of-keypoints 表現では，各局所特徴は 1000 種類の代表パターンのいちばん近い物に割り当て，ピラミッドレベル 1 〜 3 を使用した．これにより，ピラミッドレベル 1 では 1000 次元，ピラミッドレベル 2 では 4000 次元，ピラミッドレベル 3 では 9000 次元の画像特徴ベクトルが得られる．

3.3.3　Histograms of Oriented Gradients

一般物体認識の為の gradient-base の視覚的特徴として，Dalal ら[10)] は Histograms of Oriented Gradients（HOG）を提案した．HOG は SIFT と同様に，輝度の勾配方向をヒストグラム化した特徴量である．SIFT は特徴点の周りに対して特徴量を記述するのに対し，HOG は一定領域に対する特徴量の記述を行う．そのため，物体の大まかな形状を表現することが可能である．

本研究では与えられた領域を 8×8 セルに分割し，1 ブロックを 3×3 とし，$6 \times 6 = 36$ ブロックを取る．よって，与えられた領域全体で 2916 次元のヒストグラムを得た．

3.3.4　ガボール特徴

ガボール特徴は，画像から局所的な濃淡情報の周期と方向を表した特徴量である．カーネルの形を固定し，それを周期を変えて伸び縮みさせたり，回転させて方向を変えたりして，様々な周期や方向のカーネルフィルタカーネルを作成する．周期的濃淡変化を抽出する．解像度 m，方向 n のガボールフィルタは次式で表される．

$$g_{m,n}(x,y) = \frac{k_m^2}{\sigma^2} \exp\left\{-\frac{k_m^2(x^2+y^2)}{2\sigma^2}\right\} \times \left[\exp\left\{jk_m(x\cos\theta_n + y\sin\theta_n)\right\} - \exp\left(-\frac{\sigma^2}{2}\right)\right] \tag{1}$$

ここで，式 1 の k_m および θ_n は，以下のように表される．

$$\begin{aligned} k_m &= a^m \quad (0 \le m \le S-1) \\ \theta_n &= \tfrac{n\pi}{K} \quad (0 \le n \le K-1) \end{aligned} \tag{2}$$

K は方向の数，S は解像度の数，a は拡大率を表す．式 1 で表されるフィルタを用いて，それぞれに対応した空間周期の特徴を抽出 (パターン強度を数値化) する．ガボールフィルタ

は，特定の向きのエッジと特定の幅のエッジを抽出する．最後に，各フィルタ毎に強度の平均を求め，それをヒストグラムとする．本研究では与えられた領域を 8×8 セルに分割し，それぞれ 6 方向，4 周期のガボール変換カーネルについて特徴を抽出することで 1536 次元の特徴ベクトルを得た．

3.4 候補領域に対する種類分類

3.4.1 分 類 器

対象の領域から特徴ベクトルの抽出をしたら，事前に学習された特徴ベクトルと比較して，どの種類の料理クラスに属するかを決定する．

本研究では，分類器として Support Vector Machine(SVM) を用いて，各料理クラスに対する評価値を計算する．

3.4.1.1 カーネル関数

線形識別器は 2 クラスが線形分離可能であるときには高い認識率を期待できるが，非線形で複雑な問題に対してはその限りではない．そこで，非線形な写像 Φ で写像される先での内積 $(\Phi(x) \cdot \Phi(x^{'}))$ は，元の空間で定義されるカーネル関数 $K(\mathbf{x_1}, \mathbf{x_2})$ の値と一致するものとする．本実験で用いるカーネル関数として，線形識別関数の線形カーネルと χ^2 距離に基づく χ^2RBF カーネルについての定義は以下のようになる．

$$K(\mathbf{x}, \mathbf{y}) = \exp\Bigl(-\gamma \sum_i \frac{\|x_i - y_i\|^2}{x_i + y_i}\Bigr)$$

ただし，$\gamma > 0$ となる実数であり，パラメータとして与える必要がある．

Zhang らは[11] は，χ^2RBF カーネルのパラメータ γ に，全ての学習画像のベクトルの組み合わせの χ^2 距離 $\sum_i \frac{||x_i - y_i||^2}{x_i + y_i}$ の平均の逆数を設定することによって，良い結果を報告している．本研究においても，γ のパラメータは同様の方法で設定する．

3.4.1.2 Multiple Kernel Learning による特徴統合

本研究では，より高精度に料理を認識するために，複数の画像特徴量のカーネルを線形結合することにより統合カーネルを作成し，それを SVM に適用して特徴統合による画像認識を実現する．

最適なカーネル (カーネルを重みつきで線形結合したカーネル) のサブカーネルに対する重み β_j を学習する．統合カーネルは以下の式のように表される．

$$K_{\text{combined}}(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^{K} \beta_j k_j(\mathbf{x}, \mathbf{x}')$$
$$\text{with} \;\; \beta_j \geq 0, \;\; \sum_{j=1}^{K} \beta_j = 1. \tag{3}$$

各サブカーネルをそれぞれの特徴と対応させることによって，MKL は特徴選択や特徴統合に用いることができる．

本研究では，以前の研究[4] と同様[12] に MKL の学習を行う．

3.4.2 候補の出力

それぞれの候補領域から得た，それぞれの料理クラスに対しての評価値を，降順にソートし，上位 N 個の料理を候補領域として出力する．上位から出力する際に，一度出力した種類の料理は再度出力はしない．

4. 実 験

4.1 データセット

本研究では，実験に使用するために大量の食事画像を利用する為に，大量の食事画像を保持するデータセットを構築した．以前の研究[4] で収集された 50 種類の料理に，新たに 35 種類の料理を加えた 85 種類について，それぞれについて，100 枚以上の画像を保持している．認識対象とした料理は,「食事バランスガイド」を基に，我々が普段食べるであろう物を選んだ.「食事バランスガイド」は農林水産省と厚生労働省が共同で，健康作りのために食生活指針を具体的な行動に結びつける物として制定した物である．図 6 は，今回対象にした 85 種類の料理のサンプル画像である．多くの画像は WEB 上の検索サービスを利用して集めた．また，以前の研究[4] で作られた認識システムに送られた画像等も利用した．

本実験では，このデータセットから，単品が写った画像 7178 枚，複数品が写った画像 43 枚を学習に利用し，また，単品が写った画像 1698 枚，複数品が写った画像 141 枚を対象に分類を行った．分類対象する画像の例を**図 7** に示す．

4.2 評 価 方 法

分類結果の評価に用いる基準として，分類率を用いた．本実験で使用する分類率を以下の式で定義する．

$$\text{分類率} = \frac{\text{第 N 候補までに挙げられた正しい料理の数}}{\text{分類されるべき全ての料理の数}}$$

図 6 85 種類の料理のサンプル

(a) 単品が写った画像例

(b) 複数品が写った画像例

図 7 料理画像のサンプル

4.3 認識エンジンの学習

本研究で構築する認識エンジンでは，ESS による位置検出と，MKL-SVM による分類を行う際に，事前に SVM の学習が必要となる．以下に，学習時に与えたデータの詳細を記す．

4.3.1 ESS の学習

本研究では ESS の評価値として用いる線形カーネルを用いた SVM に与える特徴として，色特徴，SIFT，CSIFT の 3 種類の視覚的特徴を用いた．それぞれの局所特徴は，それぞれの code book によって量子化した．本研究ではそれぞれの code book のサイズは 1000 とした．また，各特徴について，ピラミッドレベル 1 〜 3 について，特徴ベクトルを作成した．複数の特徴ベクトルを用いるため，学習は MKL を用いて行った．

本研究の ESS での領域検出では，各種類の料理についてそれぞれ候補領域の検出を行うので，各種類の料理について検出器を学習させ構築した．SVM の学習に与える正例として，学習用画像の対象種類の料理が含まれる領域の画像特徴を用いた．また，負例として，料理

の写っていない背景領域の画像特徴と，他の料理の前景領域の画像特徴を用いた．

4.3.2 候補領域に対する分類器の学習

本研究では分類器の評価値として用いる χ^2RBF カーネルを用いた SVM に与える特徴として，色特徴，SIFT，CSIFT，Gabor，HOG の 5 種類の視覚的特徴を用いた．色特徴，SIFT，CSIF の 3 種類の局所特徴は ESS の学習と同様に量子化した．また，3 種類の局所特徴について，ピラミッドレベル 3 を用いて，特徴ベクトルを作成した．複数の特徴ベクトルを用いるため，学習は MKL を用いて行った．

本研究では 1-vs-rest 法でマルチクラス分類を行うので，各種類の料理について分類器を作った．正例と負例に関しては ESS の学習と同様の領域を用いた．

4.4 比較手法

本研究で構築した認識エンジンの精度を評価するためにいくつかの手法を比較した．単純な手法との比較として，本研究で行った各候補領域の検出を単独で画像全体に対して用いた物と精度を比較する．本研究では，各候補領域検出方法に対して検出されたすべての領域を使用する．また，正しく候補領域が与えられた時の精度についても調べた．

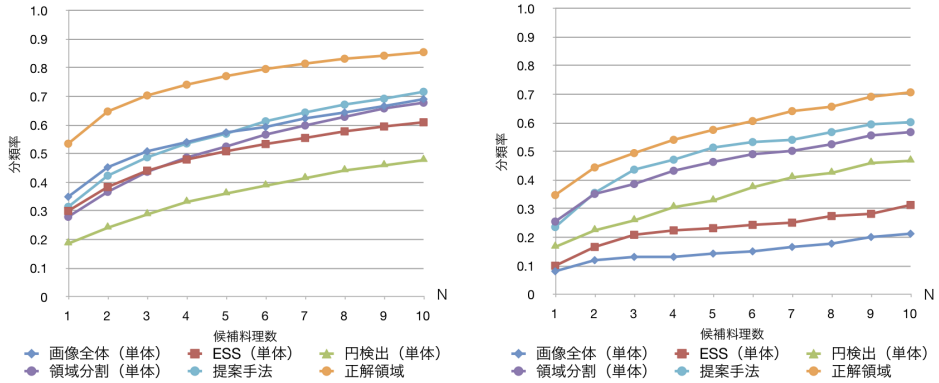
4.5 実験結果

単品が写った画像と複数品が写った画像それぞれについて，出力候補料理数を変えたときの分類率を図 8 に示した．この図では，各候補領域検出手法を単体で利用した場合，その 4 つの候補領域検出手法を全て利用する提案手法，正解データの bounding box を利用して領域が既知だった場合 (true region) の分類率を示した．

第 10 候補までの料理名を出力した場合，単品を含む画像での分類率においては，提案手法は従来手法から 2.6 ポイント向上し，71.6%の分類率を達成した．複数品を含む画像での分類率においては，提案手法は従来手法から 39.0 ポイント向上し，60.2%の分類率を達成した．複数品を含む画像での分類率が提案手法で大きく向上した．

単品を含む画像は，料理が大きく写っている傾向があり，画像全体を使って分類をする従来手法でも良い精度が出す事ができたが，複数品を含む画像での精度が良くないのは，画像中の小さな対象を認識する場合，背景や他の種類の料理の視覚的特徴も得てしまいい，認識が困難であった為であると考えられる．

ESS を用いた候補領域での認識精度がそれほど良くなかったのは，各料理クラスについて，候補領域を 1 つずつしか検出しなかった事が原因と思われる．線形 SVM での評価値の信頼度はあまり高く無いので，複数の候補領域を検出する事が望ましいが複数の候補領域を検出するためには，比較回数を大幅に増やす必要がある．



(a) 単品を含む画像での第 N 候補までの分類率　(b) 複数品を含む画像での第 N 候補までの分類率

図 8 候補料理数を変化させたときの分類率

本実験では，各候補領域検出法のうち，単体で検出を行ったときの認識精度が一番良いものは領域分割を利用した方法であった．これは，料理や皿の境界でうまく領域が分割できていた事と，候補領域の検出数が他の手法と比べて多い事が要因として考えられる．

正解領域が与えられた時の結果は，領域を検出して分類する手法の精度の上限と考えることができる．その分類率と，本研究で提案した手法の間にはまだ大きな差がある．よって，よりよい候補領域の検出手法を行うことで本研究の認識エンジンの精度上げる余地があると考えらる．

また，各領域検出手法を単体で用い場合より，全てを候補検出手法を統合した場合の方が良い精度となった．

これらの事より，複数品を含む食事画像の認識において提案手法は従来手法よりも有効である事が示された．

4.6 認識時間

本研究で構築した認識エンジンでは，位置検出や多数の特徴を用いた認識をするために，多くの処理時間を必要とする．以下に，認識時間について記載する．

局所特徴として用いた特徴の特徴抽出と code word 割り当てに，平均して色特徴はおよそ 38 秒，SIFT はおよそ 49 秒，CSIFT はおよそ 147 秒かかった．ESS による候補領域検

出は 85 種類全て行うと，平均しておよそ 107 秒かかった．円検出による候補領域検出には，0.06 秒程度と高速に実行可能であった．JSEG による候補領域検出には，平均しておよそ 24 秒かかった．検出された各領域に対して，特徴ベクトルを生成するのにかかった時間は平均して，Gabor はおよそ 12 秒，HOG はおよそ 0.05 秒かかった．色特徴，SIFT，CSIFT についてはすでに code word 割り当てまで行っているのでそれを利用してそれぞれおよそ 0.06 秒かかった．候補領域の数は各画像に対して平均 22 個検出された．分類には，85 種類全ての分類器を実行するのに平均しておよそ 1550 秒かかった．分類値をソートして重複した料理を除く処理には平均しておよそ 0.2 秒かかった．

これらの処理をすべて逐次的に処理すると合計して，およそ 2184 秒，すなわち，およそ 36 分かかるが，各特徴の抽出処理や，各候補領域検出処理，各候補領域に対する特徴ベクトルの生成，各料理の分類評価値を得る処理等，独立している並列に計算することが容易な処理が多い．それらの処理を理想的な状態で並列処理ができたとすると，候補領域検出におよそ 149 秒 (85 並列)，特徴ベクトル作成におよそ 12 秒 (23 並列)，分類におよそ 19 秒 (85 並列)，分類値のソートと結果表示におよそ 0.2 秒かかるので，合計およそ 180 秒，すなわち，およそ 3 分の処理時間がかかる計算になる．

なお，処理時間は，AMD Phenom II X4 3.0GHz の CPU を用いて，シングルコアで動作した時の実時間を計測した．

5. ま と め

本研究では，食事画像を分類する際に，候補領域を検出し，各候補領域について料理の分類をし，候補料理を出力する，食事画像認識エンジンを構築した．10 個の候補を表示するとき，単品を含む画像では，従来手法と比べ 2.6 ポイント向上し，71.6%，複数品を含む画像では，従来手法と比べ 39.0 ポイント向上し 60.2%の分類率を達成し，特に複数品を含む食事画像の認識において提案手法は有効であることを示した．

今後は，実用化のため，視覚的特徴以外の情報である撮影時間や料理同士の共起確率等を利用してさらに分類精度を高めていく必要がある．また，本研究で構築したアルゴリズムは多くの処理を行うので，分類にかかる時間が多い．GPGPU やより効率の良いアルゴリズムを使用することで，処理を高速化する事も重要な課題である．

5.1 食事画像認識システムの今後

本研究を行う目的は，食生活の情報を管理することで，より健康に生活をおくるためのヒントを得たり，管理されているという事を意識して自然と食生活が正されたりする事にある．そのためには，食事の量の認識を行うことができると，より発展するだろう．また，認識精度以外に，積極的にデータを集めたくなるような仕組みを考える必要がある．

参 考 文 献

1) 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.
2) Z.Zong, D.T. Nguyen, P.Ogunbona, and W.Li. On the combination of local texture and global structure for food classification. In *IEEE International Symposium on Multimedia*, 2010.
3) M.Chen, K.Dhingra, W.Wu, L.Yang, R.Sukthankar, and J.Yang. Pfid: Pittsburgh fast-food image dataset. In *Proc. of IEEE International Conference on Image Processing*, 2009.
4) 上東太一, 甫足創, 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, Vol. J93-D, No.8, pp. 1397–1406, 2010.
5) A.Vedaldi, V.Gulshan, M.Varma, and A.Zisserman. Multiple kernels for object detection. In *Proc. of IEEE International Conference on Computer Vision*, 2009.
6) 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.
7) Y.Deng and B.S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.23, No.8, pp. 800–810, 2001.
8) A.E. Abdel-Hakim and A.A. Farag. Csift: A sift descriptor with color invariant characteristics. In *Proc. of IEEE Computer Vision and Pattern Recognition*, Vol.2, pp. 1978–1983. IEEE, 2006.
9) S.Lazebnik, C.Schmid, and J.Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 2169–2178, 2006.
10) 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.
11) J.Zhang, M.Marszalek, S.Lazebnik, and C.Schmid. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study. *International Journal of Computer Vision*, Vol.73, No.2, pp. 213–238, 2007.
12) Manik Varma and Debajyoti Ray. Learning the discriminative power-invariance trade-off. In *Proc. of IEEE International Conference on Computer Vision*, 2007.