社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

k-means による局所特徴量抽出と皿検出器による食事画像認識の改良

松田　裕司†　　柳井　啓司†

† 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 東京都調布市調布ケ丘 1-5-1
E-mail: †matsuda-y@mm.cs.uec.ac.jp, ††yanai@cs.uec.ac.jp

あらまし　我々は，以前にスライディングウィンドウサーチや領域分割，円検出を用いて料理領域を推定し，推定された領域を複数の特徴量を用いて分類を行う食事画像認識エンジンを提案した．本研究では，この認識エンジンに k-means を用いた局所領域の画素値に基づく特徴量を加え，さらに料理領域推定に料理の種類に依らない皿検出器を用いることで改良を行う．実験では，k-means による特徴量は，料理領域が既知の場合ではあるが，同次元数である BoF-SIFT と比較して，分類率が 2.4 ポイント向上し，69.3%，皿検出器は，従来の料理ごとに学習した DPM と比較して，分類率が 3.8 ポイント向上し，66.4%となり，それぞれの有効性を示した．
キーワード　食事画像認識, 皿検出

Improvement of food image recognition using local feature extraction by k-means and dish detector

Yuji MATSUDA† and Keiji YANAI†

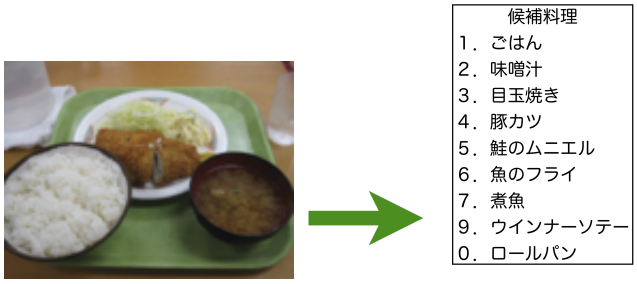
† Department of Informatics, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585 Japan
E-mail: †matsuda-y@mm.cs.uec.ac.jp, ††yanai@cs.uec.ac.jp

Abstract　In our previous work, we propose a method to recognize multi-food images by detecting candidate regions with several food region detectors including a circle detector, the JSEG region segmentation and sliding window search by the Deformable Part Model. In this paper, we improve food image recognition using feature extraction by k-means and dish detector.
Key words　Food image recognition, dish detector, k-means

1. はじめに

近年，スマートフォンなどの携帯端末を利用して，食事記録を取るサービスが普及しつつある．食事の記録を行うことで，栄養素の評価や食生活の見直しができる．一般的な記録方式には，リストから階層的に選択するものや，検索によるものが挙げられるが，毎食すべての料理について記録を行うには非常に手間が大きい．そこで，より手軽に食事の記録を取る方法が望まれている．

我々はこれまでの研究 [1] で，食事内容を簡単に記録するために画像認識によって食事画像中の料理の候補を提示する認識エンジンを構築した (**図** 1)．本研究では，k-means による教師なし学習で得られた特徴量，料理領域検出に皿モデルによるウィンドウサーチを用いることで認識エンジンの改良に取り組んだ．



(a) 入力画像　　(b) 結果表示

図 1　認識例．入力画像から料理の候補推定し出力する．

2. 関 連 研 究

食事画像の認識に関する関連研究として，FoodLog(注1)では，画像から得られる画像特徴を用いて，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，複数品目を含むような画像にも対応し，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最終目標としている．

Yang ら [2] 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，米国でよく食べられている 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 [3] も同様のファーストフードデータセット [4] 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，これらの米国の食生活に基づくデータセットとは異なり，日本でよく食べられている物を中心にデータセットを構築している．

我々は以前から食事画像認識の研究を行なっている [1]．従来手法では，複数品目の料理を含む画像に対して，100 種類の料理分類を行ったところ，10 個の候補を提示した時に，55.5%の分類率であった．

3. 従来手法概要

従来手法の概要を**図** 2 に示す．従来手法 [1] では，料理領域検出として，円検出，JSEG による領域分割，DPM によるウィンドウサーチが用いられていた．次に，料理領域検出によって得られた候補領域から特徴抽出を行い，それぞれ分類を行う．画像特徴は，SIFT, CSIFT, Color, HOG, Gabor の 5 つを用いており，SIFT, CSIFT, Color は Bag-of-Features 表現により特徴ベクトルとした．分類器には，SVM を用いて，複数の特徴を統合するため，multiple kernel learning(MKL) により学習を行う．複数の候補領域から各料理の分類器の評価値が得られるため，最終的な認識結果には，料理ごとに評価値の最大値をその料理の評価値とみなして，上位 N 個の料理を候補として出力する．

4. k-means による局所特徴量抽出

従来手法で用いられている SIFT などは，人手により設計された特徴量であった．それに対して，画像を表現するのに最適な特徴量を，学習データから教師なしで設計する手法が提案されている．

Coates ら [5] は，k-means を用いて学習された特徴量が，

(注1)：http://www.foodlog.jp

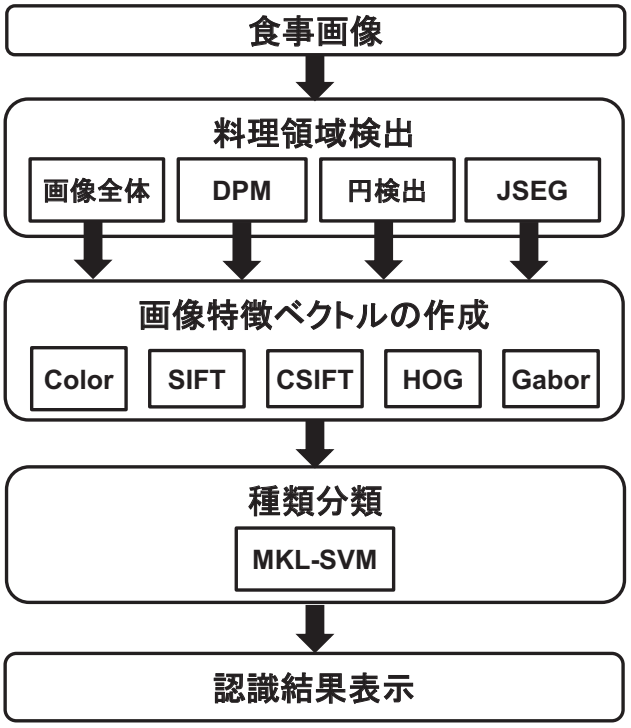


図 2　認識の流れ

sparse auto-encoder や sparse Restricted Boltzmann Machine(RBM), 混合ガウスモデルによって学習された特徴量よりも高い分類精度を得られることを示した．

以下に，特徴量学習の流れを示す．

（1） 学習画像から一定の大きさのパッチをランダムに切り出す．

（2） 得られたパッチに対して，正規化および白色化を行う．

（3） 教師なし学習アルゴリズムを用いて特徴量を学習する．

ここで正規化とは，各ピクセルの RGB について，全体の RGB の平均を引き，標準偏差で割る操作である．これにより，平均が 0，分散が 1 となり，明度とコントラストの正規化が行われる．また，白色化とは，パッチベクトルの共分散行列が単位行列になるような変換であり，Zero-phase whitening と呼ばれる．**図** 3 に k-means によって得られたクラスタ中心の例を示す．

k-means による学習では，クラスタリングの結果得られたクラスタ中心をコードブックとして利用する．しかし，Coates らの実験では，BoF のように最近傍のコードワードのヒストグラムにするのではなく，全クラスタ中心との距離の平均値との差に基づいて特徴ベクトルを作成することで良い結果が得られるとされる．

特徴抽出の流れを以下に示す (**図** 4)．パッチのサイズを $w \times w$, パッチを取得する間隔を s, パッチベクトルを x, コードブックサイズを K で表す．

（1） 画像から $w \times w$ のパッチを s pixel おきに切り出す．

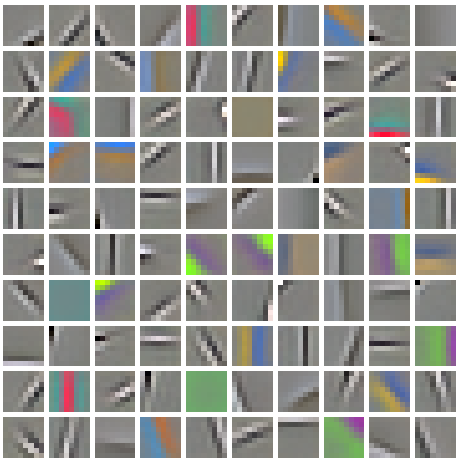


図 3　k-means による学習で得られたクラスタ中心 ([5] より引用)

（2） 学習の際と同様にして，正規化および白色化を行う．

（3） 各パッチを**式** (1) によって符号化する．

（4） 画像を 2 × 2 に分割し，領域ごとに (3) の総和をとる．

（5） (4) で得られた特徴ベクトルをパッチと同様の手法で正規化する．

$$f(x) = [f_1(x), f_2(x), \cdots, f_K(x)] \tag{1}$$

$$f_k(x) = \max\left(0, \mu(z) - z_k\right) \tag{2}$$

ここで，$z_k = ||x - c^{(k)}||_2$ (k 番目のクラスタ中心 $c^{(k)}$ とのユークリッド距離)，$\mu(z)$ は全クラスタ中心との距離の平均値を表す．

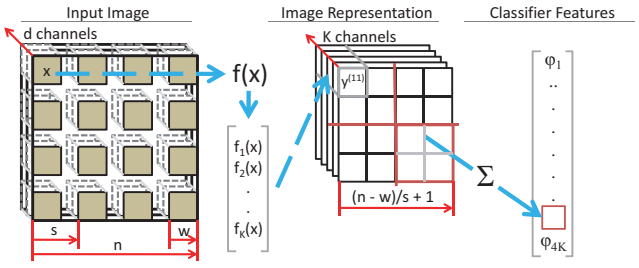


図 4　特徴抽出の流れ ([5] より引用)

本研究では，Coates ら [5] と同様にして，パッチのサイズを 6 × 6，パッチの間隔を 1pixel, コードブックサイズを 1600 とした．ただし，画像全体ではなく，領域ごとに特徴抽出を行うため，特徴ベクトルは領域の大きさ (幅 × 高さ) で割ることで，スケールを統一している．

5. 皿検出器

皿検出器には，従来手法と同様に，Felzenszwalb らによって提案された Deformable Part Model(DPM) [6] を利用する．DPM は，全体の形状を表す global root filter と複数の part filter の 2 層でオブジェクトを定義し，各 filter の妥当性とパーツの位置関係で評価を行う．また，学習の際にアスペクト比でクラスタリングを行い，ひとつのクラス中に複数のモデルを学習することで，クラス内の変化にも対応している．

図 5　100 種類の料理サンプル

従来手法では，DPM による料理領域検出は，料理ごとに学習を行い，入力画像に対して全ての料理の検出を行なっていた．しかし，この手法では認識対象の料理数だけ探索を行う必要があり，認識対象の増加によって計算コストが増大してしまうことが問題である．

また，食器には円形のものが多いということから円検出による料理領域検出も行なっていた．この手法では，当然であるが円形以外の皿領域は検出できないという問題がある．

そこで，本研究では DPM を用いて料理の種類に依らない形状を学習することで皿検出を行う．学習には，複数品目画像中の Ground Truth 領域全てを正例として用いて，それ以外の料理を含まない画像を負例として学習する．

6. 実　　験

実験では，複数品目の料理を含む画像を対象とし，皿検出器による料理領域検出および k-means で学習を行った特徴量の評価を行う．

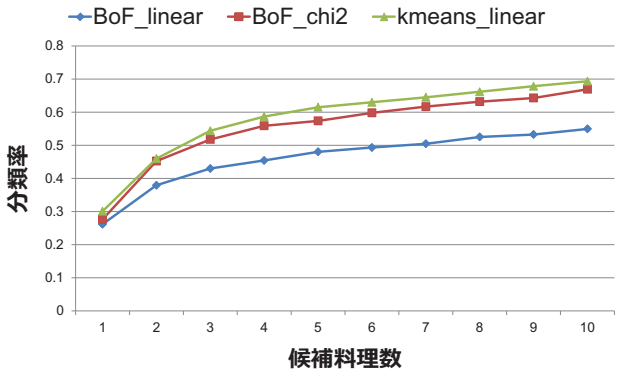


図 6　GroundTruth 領域における特徴量ごとの分類率の比較

6.1　データセット

実験には，我々が以前から構築している食事画像データセットを用いる．データセットには，普段食べるであろう料理を中心にした 100 種類の料理が含まれる．100 種類の料理のサンプルを**図** 5 に示す．この中から，学習用に 8547 枚 (10014 品を含む)，評価用に 207 枚 (535 品を含む) の画像を用いる．

6.2　評 価 方 法

分類結果の評価に用いる基準として，以下で定義する分類率を用いる．

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

6.3　k-means による特徴量の評価

ここでは，単独の特徴量での分類性能の比較を行う．比較対象には，Bag-of-Features 表現の SIFT を用いる．

BoF のコードブックサイズは 1600 とし，2×2 に分割した領域ごとに BoF を作成して連結することで，6400 次元の特徴ベクトルにする．SVM のカーネルには，BoF は線形カーネルおよび χ^2−RBF カーネル，k-means で学習した特徴量は，Coates ら [5] と同様に線形カーネルを用いる．

ここでは，単純に特徴量ごとの分類性能を比較するために，料理領域は既知であるものとして，各領域がどの料理であるか分類するものとする．

特徴量ごとの分類率の比較を**図** 6 に示す．**図** 6 中の BoF_linear は線形カーネルを用いた BoF，BoF_chi2 は χ^2−RBF カーネルを用いた BoF，kmeans_linear は線形カーネルを用いた k-means で学習した特徴量による分類の結果である．

k-means で学習した特徴量による分類率は，BoF-SIFT の線形カーネルを用いた場合よりも 14.3 ポイント高く，χ^2−RBF カーネルを用いた場合よりも 2.4 ポイント高い結果となり，k-means で学習した特徴量の有効性が確認できた．

6.4　皿検出器の評価

料理領域検出手法ごとの分類率の比較を**図** 8 に示す．**図** 8 中の DPM_class は料理ごとのモデル，DPM_dish は皿モデル，

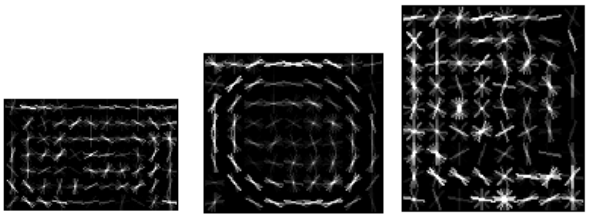


図 7　DPM で学習した皿モデルのエッジ強度

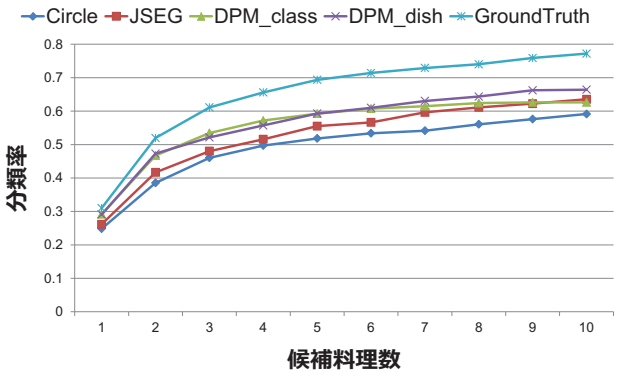


図 8　領域検出手法ごとの分類率の比較

表 1　各領域検出手法で得られた領域の平均適合率，平均再現率

	円検出	JSEG	DPM(料理ごと)	DPM(皿モデル)
平均再現率	0.520	0.657	0.449	0.651
平均適合率	0.315	0.336	0.699	0.412

GroundTruth は人手で付けられた Bounding box の領域での結果である．

上位 10 位まで考慮した場合の各手法の分類率は，円検出が 59.2%，JSEG が 63.6%，DPM の料理ごとのモデルが 62.6%，皿モデルが 66.4%となり，単独では DPM の皿モデルが一番良い結果となった．

次に，各領域検出手法で得られた領域の平均適合率および平均再現率を**表** 1 に示す．ここでの適合率は，各候補領域と最も重なる Ground Truth 領域とのオーバーラップの割合であり，再現率は，Ground Truth 領域と最も重なる候補領域とのオーバーラップの割合である．

従来の手法では，円検出や領域分割によって検出された領域は 100 種類のどの料理であるか SVM で評価するのに対して，ある料理クラスの DPM によって検出された領域は，その料理であるかのみを SVM によって評価していた．そのため，JSEG のように適合率が低い場合でも再現率が高いため，最終的な分類率は高くなる．しかし，料理ごとのモデルでは，適合率が高い一方で，再現率が低く，検出漏れが他の手法よりも多いことがわかる．それに対して，料理の種類に依らない皿のモデルでは，JSEG と同程度の再現率を保ちつつ，円検出，JSEG よりも高い適合率が得られたため，今回の実験では，最も良い結果になったと考えられる．

7. まとめと今後の課題

本研究では，k-means による特徴量と皿検出器を用いて食事画像認識の改良を行った．k-means による特徴量は，料理領域が既知の場合ではあるが，同次元数である BoF-SIFT と比較して，分類率が 2.4 ポイント向上し，69.3%となった．DPM による皿モデルでは，従来の料理ごとにモデルを作成した場合と比較して，分類率は 3.8 ポイント向上し，66.4%となった．DPM による皿モデルと他の料理領域検出手法，k-means で学習した特徴量と他の特徴量の組み合わせについて，実験を行う予定である．

本研究では，k-means による特徴量学習を用いたが，他の手法による特徴量の学習を検討している．また，DPM の皿モデルの学習でのクラス内モデルをいくつにするかや，今回は Felzenszwalb らの手法をそのまま用いてアスペクト比でクラスタリングを行ったが，形状に基づいてクラスタリングを行うなど検討する必要もある．

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,” Proc. of IEEE Computer Vision and Pattern Recognition, pp.2249–2256, 2010.

[3] Z. Zong, D.T. Nguyen, P. Ogunbona, and W. Li, “On the combination of local texture and global structure for food classification,” IEEE International Symposium on Multimedia, pp.204–211, 2010.

[4] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, “Pfid: Pittsburgh fast-food image dataset,” Proc. of IEEE International Conference on Image Processing, pp.289–292, 2009.

[5] A. Coates, H. Lee, and A.Y. Ng, “An analysis of single-layer networks in unsupervised feature learning,” International Conference on Artificial Intelligence and Statistics, pp.215–223, 2011.

[6] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.32, no.9, pp.1627–1645, 2010.