局所特徴・色特徴に基づく静止画像からの料理領域検出

宮野瑠衣子† 植松 裕子† 斎藤 英雄†

† 慶應義塾大学大学院理工学研究科 〒 223–8522 神奈川県横浜市港北区日吉 3–14–1
E-mail: †{rui,yu-ko,saito}@hvrl.ics.keio.ac.jp

あらまし 近年，カメラで料理の写真を撮るユーザが増加しており，それに伴い料理画像の研究への関心が高まっている．その一例として，料理画像から種類や分量を推定する研究が近年盛んに行われている．これらの研究では，料理が画像全体に写った画像を使用している．しかし，実際に撮影される画像では料理が中心に大きく写っているとは限らず，その位置が未知である場合が多いため，料理領域を画像処理の入力とする場合には，料理の位置を推定する必要がある．そこで，本論文では 1 枚の静止画像中から自動で料理の写っている領域を検出することを目的とする．そのために，画像の局所特徴を考慮した Bag-of-Features 表現，画像の大域的な色情報を考慮した色特徴を用いて料理の認識を行い，料理であると認識された領域を結合する手法を提案した．最終的に，2 種類の特徴の有効性を確認する実験と実画像を用いて料理領域を検出する実験を行い，提案した手法で料理領域の検出が行えることを示した．

キーワード 料理領域検出，一般物体認識，BOF，色特徴，SURF

1. はじめに

近年，デジタルカメラやカメラ付き携帯電話で料理を撮影するユーザが非常に増えている．その理由としては，ブログなどネット上で公開するために料理を撮影するユーザの増加や，健康管理のために料理画像を記録するサービスが普及したことなどが考えられる．

そして，近年画像処理の分野では，料理画像に関する研究が盛んに行われている．例えば，上東ら [1]，Yang ら [2]，Puri ら [3] は，ユーザの健康管理という目的のために料理画像の認識・分類を行っており，料理画像に関する研究は日常生活と密接した非常に重要な分野であると言える．

一方，顔検出 [4] や人検出 [5] など，画像内から特定の物体を検出するオブジェクト検出に関する研究も盛んに行われている．現在販売されているデジタルカメラにはほとんど顔検出機能が搭載されており，最近では顔検出以外に笑顔検出・ペット顔検出などの機能も開発されている．これらの機能は，検出した領域に対してオートフォーカスや明るさの調整などの処理を行うために非常に有効であり，顔検出機能はデジタルカメラ以外にカメラ付き携帯電話にも搭載され，広く普及している．

そこで，これらのニーズを踏まえ，本研究では静止画像内から料理領域を検出するシステムを提案する．料理の位置が検出できれば，食事に関する他の研究 [1]〜[3] などへの適用や，オートフォーカスなどの画像処理への応用が期待できる．

2. 関連研究

本研究では，関連研究と同様に一般物体認識を用いて料理領域検出を行う．一般物体認識とは，画像に含まれる物体を一般的な名称で認識する技術で，画像から抽出した特徴を用いて認識を行う [6], [7].

例えば，上東らは 50 種類の料理画像を分類する手法を提案した [1]．この手法では，画像から抽出した局所特徴・色特徴・ガボール特徴の 3 種類の特徴を統合することにより分類を行っている．実験データには Web から集めた画像を使用しており，手動でバウンディングボックスを与えて領域を指定し，背景と前景 (料理) を分離して使用している．Yang らはファーストフードの 7 種類の料理を分類する手法を提案した [2]．この手法では，画像内の 2 つの画素間の距離や角度といった幾何学的な関係を特徴として用いている．実験データには既存のデータセットを使用している．このデータセットも，背景と前景 (料理) が事前に分離されているものである．Puri らは携帯電話で撮影した料理画像から料理の種類の認識と分量の推定を行った [3]．この手法では，色特徴・テクスチャ特徴を利用している．事前にセグメンテーションされた画像は必要なく，円検出により皿の位置を推定して認識を行っている．

上東ら [1]，Yang ら [2] の研究では，事前にセグメンテーションされた画像が必要である．また，Puri ら [3] の研究では特別な画像は必要ないが，円検出により料理の位置を推定しているため，皿の形によっては対応できない．また，高精度に認識を行うため，2 種類のチェッカーボードを用いてカメラをキャリブレーションする必要がある．更に，入力データとして 3 視点からの料理画像とユーザの音声データが必要となる．

そこで，本研究では 1 枚の静止画像から料理の写っている領域を検出することを目的とする．従来研究が，画像中に何らかの料理が写っているという前提で，その料理の種類をいくつかの候補から選ぶというのに対して，

本研究では，未知の画像中に対して料理画像であるか否かを判定する．目的は異なるが，関連研究と同様に画像から抽出した特徴を用いるのが有効と考え，上東ら [1] の研究で使用されている 3 種類の特徴のうち，局所特徴・色特徴を使用する．これらの特徴を，料理画像の特性を考慮して改善する．また，料理の有無だけでなく，位置を推定するために画像内の小領域を切り出して料理領域を検出する手法を提案する．

3. 提 案 手 法

図 1 に提案手法の流れを示す．提案手法は，学習フェーズと検出フェーズから成る．

学習フェーズでは，学習データの画像から局所特徴と色特徴という 2 種類の特徴を抽出する．それぞれの特徴は，512 次元・192 次元のヒストグラムで表現される．詳細は 4. 章で述べる．

検出フェーズでは，まずクエリ画像から小領域を切り出し，クエリ領域とする．そのクエリ領域から学習フェーズと同様に 2 種類の特徴を抽出し，学習フェーズで求めたヒストグラムと比較することにより，クエリ領域が料理か否かを判定する．クエリ領域を移動させながら画像全体にこの処理を行い，最終的に料理であると判定されたクエリ領域のみを結合して結果画像を得る．

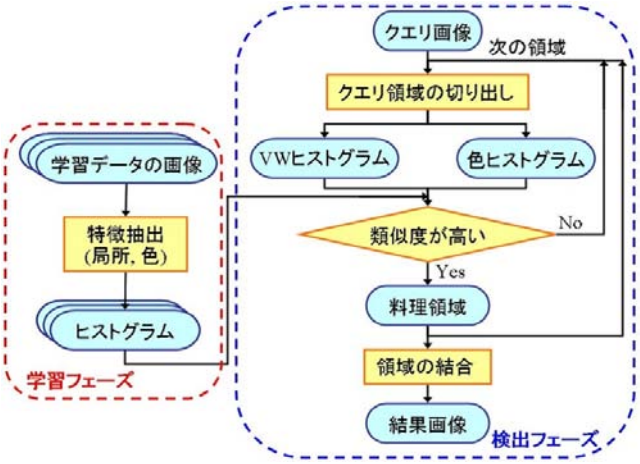


図 1　提案手法の流れ

3.1 画像特徴による料理判定

本章では，クエリ領域を料理か否か判定する手法を述べる．

提案手法では，学習データとクエリ領域で 2 種類の特徴を比較し，料理か否か判定する．2 種類の特徴とは，512 次元の Visual Word ヒストグラムで表される局所特徴と 192 次元の色ヒストグラムで表される色特徴である．

まず，局所特徴・色特徴それぞれ個別に，(1) 式を用いてヒストグラムインターセクションを求める [8].

$$H(I,M) = \frac{\sum_{i=1}^{n} min(I_i, M_i)}{\sum_{i=1}^{n} M_i} \qquad (1)$$

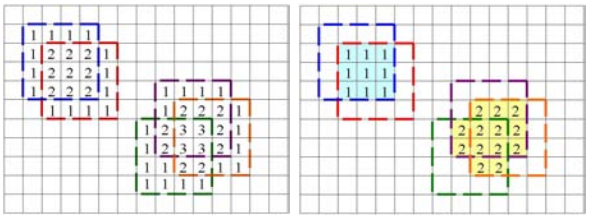
ここで，I はクエリ領域のヒストグラム，M は学習データ内画像のヒストグラム，n はヒストグラムの次元数 (局所特徴では 512 次元，色特徴では 192 次元) を表す．このヒストグラムインターセクションは，0〜1 の値を取る．また，ヒストグラムが類似している程大きくなり，完全に一致すれば 1 となる．

1 つのクエリ領域と 1 枚の学習データの画像間で Visual Word ヒストグラム，色ヒストグラムのヒストグラムインターセクションを求め，2 つの値の平均を取って 2 画像間の類似度を算出する．この類似度をしきい値と比較し，しきい値よりも大きければその 2 つの画像は類似画像とみなされる．学習データには数種類の料理画像が含まれており，料理ごとに類似画像の枚数を計算する．類似画像の枚数がしきい値よりも多い料理が 1 つでもあれば，そのクエリ領域は料理が写っていると判定される．

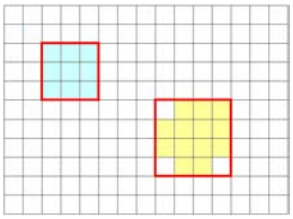
3.2 クエリ領域の結合

本章では，3.1 章において料理が写っていると判定されたクエリ領域を結合する手法を述べる．

まず，3.1 章で料理と判定されたクエリ領域について，領域内に含まれる全画素に 1 票ずつ投票をする．図 2(a) において，実線の正方形が画素を表し，破線の正方形が料理が写っていると判定された領域を表す．画素内の数字は投票された数を表している．次に，図 2(b) のように投票数がしきい値以上の画素に関してラベリングを行う．(図 2 の例では，しきい値は 2 である．) 最後に，図 2(c) のようにラベリングされた領域が収まるように四角形を描画し，これが最終的な結果となる．



(a) 領域への投票　　(b) ラベリング



(c) 結合結果

図 2　クエリ領域の結合

4. 特徴による画像表現

提案手法では，3.1 章のように事前に用意した学習データ内の各画像とクエリ画像から切り出したクエリ領域それぞれの特徴を比較して料理判定を行う．料理画像は，人間の顔の画像などと違い，全ての料理に共通するような特徴的なパターンを持たない．そこで局所特徴と色特徴という 2 種類の特徴を料理ごとに抽出してヒストグラムで表現し利用する．

4.1 局所特徴

局所特徴は一般物体認識の分野では非常に重要であり，頻繁に用いられている．局所特徴を一般物体認識で活用する場合，Bag-of-Features (BOF) 表現という手法が有効である [6], [7], [9], [10]．BOF 表現では，画像が局所特徴の分布として表される．

画像の BOF 表現を行う際には，まず大量に存在する局所特徴の特徴ベクトルを，値が似ているベクトル同士でクラスタリングして Visual Word という代表ベクトルを作成する必要がある．この処理は学習フェーズで行われる．

学習フェーズでは，学習データの各画像から SURF (Speeded Up Robust Features) [11] を抽出し，任意の数のクラスタにクラスタリングする．最も一般的なクラスタリング手法として，K-means 法という手法がある [7], [10], [12]．しかし，K-means 法では代表ベクトルの選び方に精度が依存するため，提案手法では特徴ベクトルの次元ごとの平均・分散に注目する．学習フェーズにおいて，学習データから抽出した全ての SURF の特徴ベクトルについて，次元ごとの分散・平均を求めて，分散が大きい n 次元とその平均を記憶する．

次に，学習フェーズで求めた分散・平均を元に SURF を量子化し，画像を Visual Word ヒストグラムで表現する方法について述べる．学習フェーズでは学習データの各画像，検出フェーズではクエリ領域の画像についてこの処理が行われる．それぞれの画像から SURF を抽出し，まずは学習フェーズにおいて最も分散が大きかった次元に注目する．注目する次元の値と学習フェーズで求めた平均値を比較し，図 3 のように平均値よりも小さかったら左の子ノードへ，大きかったら右の子ノードへ降りる．これを n 次元で繰り返す．最終的に，特徴ベクトルは 2^n 個に量子化される．今回は 9 次元を使用したため，$2^9 = 512$ 次元である．対象画像内の各特徴ベクトルを量子化し，Visual Word の ID に投票を行う．全ての特徴ベクトルを投票し，画像内の SURF の個数で正規化をし，Visual Word ヒストグラムを得る．

4.2 色特徴

色特徴は，画像の大域的な色情報を利用したもので，対象の形状変化などに対してロバストであることから，

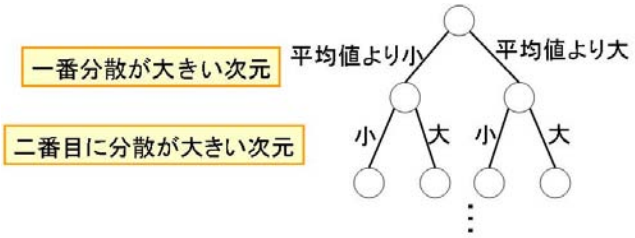


図 3 量子化

物体検出・物体認識などの分野で幅広く用いられている [8], [13]．料理画像でも，色情報は非常に重要である．

提案手法では，RGB それぞれの値を 4 階調に量子化し，ビンが $4^3 = 64$ 個のヒストグラムを作成する．上東らの手法では，に 0〜255 で表される明度値を 4 等分に 4 値化することで量子化を行っていた [1]．しかし，料理画像では赤色が強く現れることが多く，緑・青はそれほど強く現れないという特性がある．そのため，提案手法では学習データから RGB 値それぞれの分布を学習する．学習フェーズにおいて，学習データ内の全ての画像の全画素の明度値から RGB それぞれのヒストグラムを作成し，ヒストグラムの極小値をしきい値として手動で 3 つ定める．極小値が 3 つ以上存在しない場合は，平均値を取る．RGB それぞれにつき 3 つのしきい値を定め，そのしきい値と明度値を比較して 4 階調に量子化を行う．

また，料理画像は皿など円形の物体が写っている可能性が高いために，図 4 のように同心円状の 3 つの領域に区切る．最終的に $3 \times 4^3 = 192$ 次元のヒストグラムで表現する．

学習フェーズでは，学習データからしきい値の設定を行う．そして，色ヒストグラムの作成は学習フェーズでは学習データの各画像，検出フェーズではクエリ領域の画像について行う．画像の全画素の RGB 値を学習フェーズで設定したしきい値に基づいて 4 階調に量子化する．そして，各ビンの投票数を画素数で正規化して色ヒストグラムを得る．

図 4 色特徴の領域

5. 実験と結果

提案手法の有効性を示すために以下の 2 つの実験を行った．

- 画像特徴の有効性確認
- 実画像実験

また，本実験に用いた画像は以下の通りである．画像は写真共有サイトである Flickr [14] から収集したものと，自分で撮影したものを使用した．

- 学習データ
  - 色量子化，ヒストグラム作成：料理 730 枚
  - Visual Word 作成：料理 100 枚，非料理 100 枚
- クエリ画像
  - 画像特徴の有効性確認：料理 100 枚，非料理 50 枚
  - 実画像実験：料理 135 枚，非料理 15 枚

図 5 に学習データの一部を載せた．今回は 8 種類の料理を対象としている．

5.1 画像特徴の有効性確認

本実験では，クエリ画像に料理画像と非料理画像を用い，局所特徴・色特徴の有効性を確認した．クエリの料理画像は画像全体に料理が写っている画像を選び，非料理画像には，皿・人物の顔など，料理と共に写っている可能性が高い対象を選んだ．そして，クエリ領域を切り出さずに画像全体から 2 つのヒストグラムを作成して料理か否か判定した．

表 1 に 2 種類の特徴を用いた判定の結果を載せた．局所特徴と色特徴を組み合わせた場合の正解率は 90.7%となっており，精度良く判定できていることがわかる．

表 1　各特徴の精度

	正解数		
	料理	非料理	正解率
局所特徴	93 / 100	37 / 50	84 %
色特徴	99 / 100	30 / 50	86 %
局所＆色	98 / 100	38 / 50	90.7 %

図 6 に，2 種類の特徴の組み合わせにより，正しく料理と判定された料理画像の一部を載せる．これらの画像は，局所特徴のみを用いた場合は料理ではないと判定されたが，色特徴のみを用いた場合には料理と判定された．カレーやオムライスなどは濃淡変化が少なく局所的な特徴パターンが現れにくいが，画像によって色の変化が少ないため，色特徴では正しく判断されたと考えられる．

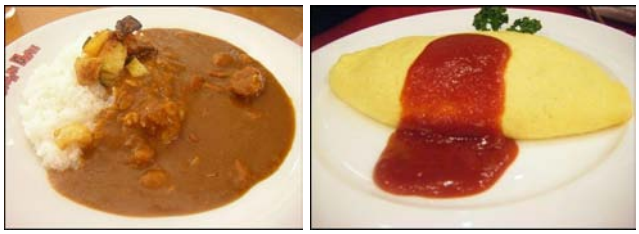


図 6　料理と判定された料理画像

図 7 に，2 種類の特徴の組み合わせにより，料理ではないと正しく判定された非料理画像の一部を載せる．これらの画像は，局所特徴のみを用いた場合は料理ではないと判定されたが，色特徴のみを用いた場合は料理と判定されてしまった．この結果から，赤色が強く現れる画像は料理画像とみなされやすいことがわかる．しかし，局所特徴と統合することにより，料理ではないと判定することができた．

図 7　料理と判定されなかった非料理画像

5.2 実画像実験

本実験では，クエリ画像から，提案手法を用いて料理領域を検出した結果を示す．今回は，200 × 200 画素のクエリ領域を横方向・縦方向に 30 画素ずつスライドさせて実験を行った．

図 8 に料理が写っている画像から料理検出した結果の一部を示す．これらの結果から，料理領域は正しく検出できていることがわかる．しかし，図 8(c) のパンや，図 8(h) の白米など，図 5 の 8 種類の学習データに含まれない料理は検出できていない．これらを検出するには，学習データの料理の種類を増やす必要がある．そして，図 8(a)，8(f) のサラダ，図 8(b) の切り分けられたピザ，図 8(g) の上部のハンバーガーのように，対象が小さいと検出できないことがわかる．この実験では多数のパラメータを経験的に定めて行ったため，サイズが小さい領域は検出できなかったと考えられる．

図 9 にクエリ画像内に複数の料理が写っており，複数の料理領域がつながってしまった失敗例を示す．図 9(a)，9(c) は 3.2 章で述べたクエリ領域への投票数の分布を表す画像，図 9(b)，9(d) は投票数の分布をラベリングした画像に基づき四角形を描画した結果画像である．図 9(a)，9(c) から，複数の料理同士が隣接しているとクエリ領域がつながってしまい，結果画像の料理領域もつながってしまうことがわかる．これは，ラベル画像の投票数の分布を解析し，2 つの料理間で投票数が少ない部分を分割するなどの処理を加えれば改善できると考えられる．

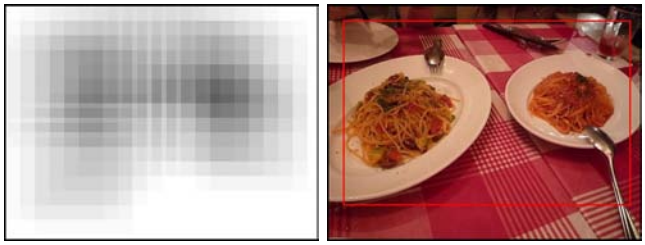
図 10 に，料理を含まないクエリ画像から料理検出を行った結果の一部を示す．図 10(a)〜図 10(c) から，料理を含まない画像からは料理領域が検出されないことが確認できた．図 10(d) に関しては，大部分が料理領域として検出されてしまった．これは，芝生に緑の領域が多いことと，人間の顔などから特徴点が多く検出されることが原因であると考えられる．

6. 結　　論

本論文では，1 枚の静止画像から自動で食事の写っている領域を検出する手法を提案した．

提案手法では，画像内の特徴的な点の分布を利用した局所特徴と，画像全体の色を利用した色特徴の 2 つの特

(a) 投票数の分布　　(b) 結果画像



(c) 投票数の分布　　(d) 結果画像

図 9　失　敗　例

徴を利用した．これらの特徴を用いて静止画像内の小領域に関して料理か否かを判定し，判定された小領域を結合して料理の位置を検出した．

そして，本研究の有効性を示すために，2 つの実験を行った．1 つ目の実験では，まず，局所特徴・色特徴に基づき 1 枚の画像について食事判定を行った．2 種類の特徴を用いて画像全体に対して料理か否か判定する実験を行い，90.7%の精度で料理判定が行えることを確認した．2 つ目の実験では，静止画像内の小領域から同様に料理判定を行って，判定された小領域を結合することにより料理領域の位置を検出することができた．また，料理の写っていない画像からは検出されないことも確認できた．

今後の課題としては，以下の 6 点を考えている．

- 認識の精度を向上させるために他の特徴を導入
- SVM(サポートベクターマシン) などの導入により料理判定の手法を改善
- クエリ領域の結合手法を改善
- しきい値など手動のパラメータを減少
- 検出可能な料理の種類を増加
- 検出した領域に対して料理画像に関する他の研究を応用

謝　　辞

本研究の一部は，（独）情報通信研究機構の委託研究「革新的な三次元映像技術による超臨場感コミュニケーション技術の研究開発」の補助により行われたものである．

文　　献

[1] 上東, 甫足他：“Multiple Kernel Learning による 50 種類の食事画像の認識 (パターン認識応用, 特集 画像の認識・理解論文)”, 電子情報通信学会論文誌. D, 情報・システム, **93**, 8, pp. 1397–1406 (2010).

[2] S. Yang, M. Chen, D. Pomerleau and R. Sukthankar: “Food recognition using statistics of pairwise local features”, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference onIEEE, pp. 2249–2256 (2010).

[3] M. Puri, Z. Zhu, Q. Yu, A. Divakaran and H. Sawhney: “Recognition and volume estimation of food intake using a mobile device”, Applications of Computer Vision (WACV), 2009 Workshop onIEEE, pp. 1–8 (2010).

[4] P. Viola and M. Jones: “Robust real-time face detection”, International Journal of Computer Vision, **57**, 2, pp. 137–154 (2004).

[5] 山内, 藤吉, 山下：“Boosting に基づく特徴量の共起表現による人検出”, 電子情報通信学会論文誌, **92**, pp. 1125–1134 (2009).

[6] 柳井啓司：“一般物体認識の現状と今後”, 情報処理学会論文誌: コンピュータビジョンとイメージメディア, **48**, pp. 1–24 (2007).

[7] G. Csurka, C. Dance, L. Fan, J. Willamowski and C. Bray: “Visual categorization with bags of keypoints”, Workshop on statistical learning in computer vision, ECCV, Vol. 1Citeseer, p. 22 (2004).

[8] M. Swain and D. Ballard: “Color indexing”, International journal of computer vision, **7**, 1, pp. 11–32 (1991).

[9] J. Sivic and A. Zisserman: “Video Google: A text retrieval approach to object matching in videos”, Proceedings of the International Conference on Computer Vision, Vol. 2, pp. 1470–1477 (2003).

[10] D. Nister and H. Stewenius: “Scalable recognition with a vocabulary tree”, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 2IEEE, pp. 2161–2168 (2006).

[11] H. Bay, T. Tuytelaars and L. Van Gool: “Surf: Speeded up robust features”, Computer Vision–ECCV 2006, **3951**, pp. 404–417 (2006).

[12] J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman: “Object retrieval with large vocabularies and fast spatial matching”, 2007 IEEE Conference on Computer Vision and Pattern RecognitionIEEE, pp. 1–8 (2007).

[13] 村瀬, V.V.Vinod：“局所色情報を用いた高速物体探索–アクティブ探索法–”, 信学論 (D-II), **81**, pp. 2035–2042 (1998).

[14] “Flickr”. `http://www.flickr.com/`.

(a) パスタ　(b) サラダ　(c) カレー　(d) ハンバーガー

(e) ラーメン　(f) ピザ　(g) オムライス　(h) お好み焼き

図 5　学習データ

(a)　(b)　(c)　(d)

(e)　(f)　(g)　(h)

図 8　料理画像からの料理検出例

(a)　(b)　(c)　(d)

図 10　非料理画像からの料理領域検出例