

# 大量の画像付き料理レシピからの学習による料理に含まれる素材推定

松長 大樹<sup>†</sup> 横井 聡<sup>††</sup> 林 泰宏<sup>††</sup> 道満 恵介<sup>†††</sup> 井手 一郎<sup>††</sup>  
出口 大輔<sup>††††</sup> 村瀬 洋<sup>††</sup>

<sup>†</sup> 名古屋大学工学部 〒464-8603 愛知県名古屋市千種区不老町

<sup>††</sup> 名古屋大学大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町

<sup>†††</sup> 中京大学情報理工学部 〒470-0393 愛知県豊田市海津町床立101

<sup>††††</sup> 名古屋大学情報連携統括本部 〒464-8601 愛知県名古屋市千種区不老町

**あらまし** 本報告では、入力された料理画像からその料理に含まれる素材を推定する手法を提案する。近年、日々の食事を記録することで個人の食事の傾向や好みを分析し、健康状態の改善を支援するサービスが増えている。このようなサービスの実現を目指した従来の研究では、食事画像から料理名を推定することで料理のカロリーを推定しようとしている。しかし、実際には同じ種類の料理でも使われる素材の種類が異なる場合も多く、料理の種類を認識するだけでは不十分である。そこで本研究では、個人が撮影した料理画像から、その料理に含まれる素材を推定する手法を提案する。料理には素材の色や形状の特徴が現れることが多いため、料理画像から画像特徴量を抽出して学習し、識別器を構築する。評価実験の結果、すべての識別器で識別性能がランダムな識別を超え、有効性を確認した。

**キーワード** 料理画像, 素材検出, 画像特徴

## Ingredients Estimation in a Dish by Learning from a Large Number of Recipes with Images

Hiroki MATSUNAGA<sup>†</sup>, Satoshi YOKOI<sup>††</sup>, Yasuhiro HAYASHI<sup>††</sup>, Keisuke DOMAN<sup>†††</sup>,

Ichiro IDE<sup>††</sup>, Daisuke DEGUCHI<sup>††††</sup>, and Hiroshi MURASE<sup>††</sup>

<sup>†</sup> Faculty of Engineering, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8603 Japan

<sup>††</sup> Graduate School of Information Science, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

<sup>†††</sup> School of Information Science & Technology, Chukyo University

Tokodachi 101, Kaizu-cho, Toyota-shi, Aichi, 470-0393 Japan

<sup>††††</sup> Information and Communications Headquarters, Nagoya University

Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan

**Abstract** In this report, we propose a method for estimating the ingredients contained in the dish in an input food image. Recently, the number of services that analyzes the preferences of an individual by recording daily meals, for improving of health, is increasing. Conventional works that try to realize such services, estimate the calorie from the name of the dish detected from the image. However, in practice, different ingredients may be used in the same kind of dishes, so only recognizing the type of a dish is insufficient. Therefore, we propose a method for estimating the ingredients in the dish in an input food image. Since the color and the shape features of an ingredient tend to appear in a dish, the proposed method extracts image feature from the food image and builds classifiers. As a result of an evaluation experiments, we confirmed that the performance of the classifier exceeds random classification.

**Key words** Food image, ingredient detection, image feature

## 1. まえがき

日常生活を送るうえで健康の維持は重要であり、そのための最も基本的な方法は、日々の活動を記録、確認することである。今日では、スマートフォンなどの携帯型情報端末の普及により、テキスト、画像、映像など様々なメディアで我々の日々の活動を容易に記録できるようになった。また、活動記録を分析することで、個人に合わせた健康管理を提供するサービスが増えている。このようなサービスの1つとして、日々の食事を記録することで、個人の食事の傾向や好みを分析し、健康状態の把握、改善を支援するサービスも提供されている。例えば、日々の食事画像からカロリーや料理名を推定し、これらを記録することで健康管理をするサービスである“goo からだログヘルスアシスト” [1] は、10万人以上が利用している。また、“Food-Log” [2] では、毎日の料理画像から食事バランスの推移を記録して、健康管理をするサービスを提供している。このようなサービスの実現には、料理画像から正確にカロリー、料理名を推定する技術が必要である。しかし、システムによる自動推定の精度は低いのが現状である。

これらの自動化を目指した研究として、料理画像からカロリーを推定する研究 [3] や複数の料理が写った画像からそれぞれの料理名を認識する研究 [4] がある。文献 [3] では、専門家によりカロリーが計算された料理画像から画像特徴を抽出し、料理の類似画像検索をすることで、カロリーを推定している。文献 [4] では、円検出などで料理候補領域を検出し、候補領域から画像特徴を抽出することで、料理の分類をしている。これらの研究では、料理画像から料理名を推定することで、料理のカロリーを推定するという目的を達成しようとしているが、実際には同じ種類の料理でも使われている素材が異なる場合が多い。例えば、同じミートスパゲティであっても、ピーマンが入っている場合も、入っていない場合もある。素材ごとにカロリーや栄養素が異なるため、料理の種類を認識するだけでは、カロリーや栄養素の推定には不十分である。そのため、料理の種類だけでなく含まれる素材も推定する必要がある。

そこで本研究では、個人が撮影した料理画像からその料理に含まれる素材を推定することを目的とする。これにより、素材による違いを考慮して料理を分類できるようになることが期待される。

本論文の構成を以下に示す。2. では、提案手法である料理画像付きレシピを用いた素材の推定方法について詳しく述べる。3. では、提案手法の有効性を評価するための実験とその結果を示し、4. で考察する。最後に、5. で本研究のまとめと今後の課題について述べる。

## 2. 大量の画像付き料理レシピの学習による料理の素材推定

### 2.1 手法概要

本手法は、学習段階と識別段階の2段階の処理により、料理画像からその料理に含まれる素材を推定する。図1に提案手法の処理の流れを示す。画像付き料理レシピは一般に、料理画像、

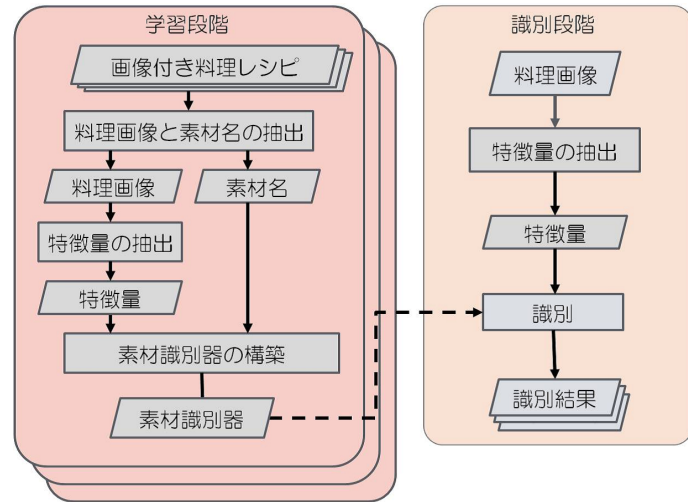


図1 素材推定手法の手順

素材一覧、調理手順などから構成される。そこで、事前に行う学習段階では、まず料理画像付き料理レシピから料理画像と料理に含まれる素材名を抽出する。次に、抽出した料理画像から画像特徴量を抽出する。そして、素材が含まれているか否かをラベルとして、それぞれの素材識別器を学習する。識別段階では、それぞれの素材識別器を用いて、入力された料理画像に各素材が含まれているか否かを推定する。

以降に各段階の処理について詳しく述べる。

### 2.2 学習段階

料理画像に特定の素材が含まれているか否かを推定する素材識別器を、素材ごとに学習する。学習処理の流れを図1の左側に示す。まず、料理画像付き料理レシピから料理画像を抽出し、また、素材一覧を調べることで素材が含まれているか否かを判定する。次に、各料理画像の中央部分を切り出し、特徴量を求める。そして、素材名の有無をラベルとし、得られた特徴量を用いて、SVMにより素材識別器を学習する。この処理を各素材に対して行い、各素材に対する素材識別器を構築する。

以下で、各処理について説明する。

#### 2.2.1 画像と素材名の抽出

画像付き料理レシピから料理画像と素材一覧を抽出する。

抽出した料理画像は個人が撮影したものであり、器ごと写した料理画像(図2.)や、料理のみを写したもの(図3.)など様々である。そのため、画像を9分割し、その中央部分から特徴量を抽出する。

また、構築しようとしている素材識別器の素材が素材一覧に含まれているか否かを調べる。

#### 2.2.2 画像特徴量の抽出

料理には使用する素材の色、形状が現れることが多い。例えば、トマトを使った料理は赤色であることが多い。そのため、提案手法では料理に含まれる素材推定のために、各料理画像から色特徴や形状特徴、テクスチャ特徴を抽出する。

##### a) 色特徴: HSV色ヒストグラム

本研究では、色特徴としてHSV色ヒストグラムを用いる。本節では、この特徴量の性質と求め方手順について説明する。



図 2 器ごと映した料理画像の例



図 3 料理のみを大きく映した料理画像の例

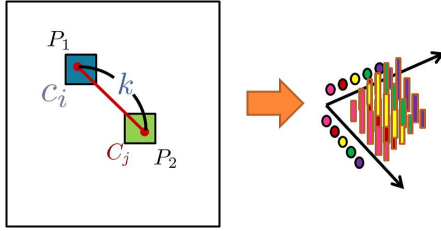


図 4 色コリログラム

HSV 色ヒストグラムは、HSV 色空間で表現した画像の各画素値の出現頻度を表したものである。HSV 色空間とは、色相 (hue)、彩度 (saturation)、明度 (value) で構成される色空間であり、人間の感覚に近い色空間である。本研究では、画像の大域的な性質を把握するために、位置情報を扱わない HSV 色ヒストグラムを色特徴とする。色ヒストグラムは画像全体から各画素値の出現頻度を抽出したもの [8] なので、位置情報が失われた特徴量である。

HSV 色ヒストグラムの抽出手順は以下のとおりである。まず、入力された画像を HSV 色空間に変換する。その後、照明変化の影響を軽減するため、式 (1) に示すように、明度を除いて色相と彩度のみで画素値を表現する。また、次元削減のために画素値を量子化する。そして、得られた画素値のヒストグラムを作成する。なお、画像ごとに大きさが異なるため、ヒストグラムの各値を画像のピクセル数で除算して、正規化する。以上から得られるのが HSV ヒストグラムである。

$$c_{ij} = (h_i, s_j) \quad (0 \leq i \leq m, 0 \leq j \leq n) \quad (1)$$

b) テクスチャ特徴: 色コリログラム

本研究では、色コリログラムをテクスチャ特徴として用いる。本節では、この特徴量の性質と抽出手順について説明する。

色コリログラムは、Huang らによって提案された画像特徴量であり、画像中の異なる位置の色同士の空間的な相関を算出するものである [9]。

色コリログラムの抽出手順は以下のとおりである。まず、入力された画像を HSV 色空間に変換する。前述の理由から、式 (1) のように各画素を色相と彩度のみで表現する。そして、量子化された各画素値  $c_i$  ( $0 \leq i \leq m$ ) の画素  $p_1$  に対して、距離  $k$  にある画素  $p_2$  の画素値が  $c_j$  である確率を色コリログラムとする。

c) 形状特徴: HOG

本研究では形状特徴として HOG (Histograms of Oriented Gradients) を用いる。以下では、この特徴量の性質と抽出手

順を説明する。

HOG は Dalal らによって提案された画像特徴量であり、画像の輝度勾配をヒストグラムとして表す特徴量である [10]。物体の輪郭などのエッジ情報の抽出に適しており、料理に含まれる素材形状の表現に有効であると考えられる。

HOG 特徴量の抽出手順は以下のとおりである。式 (2)、式 (3) により点  $(x, y)$  における勾配強度  $m(x, y)$  と勾配方向  $\theta(x, y)$  を計算する。次に画像を等間隔に分割し、これらの領域 (セル) ごとに勾配方向のヒストグラムを作成する。ヒストグラムのビンは量子化された勾配方向とし、投票の際に勾配強度で重み付けをする。注目セル周辺の  $n \times n$  セルの集まりをブロックとし、ブロック内にある各セルのヒストグラムを連結して特徴ベクトルを作成する。その後、特徴ベクトルを L2 正規化する。 $\mathbf{x}$  を特徴ベクトル、 $x_i^2$  を  $\mathbf{x}$  の  $i$  番目の特徴量、 $\epsilon$  をゼロ除算を防ぐための定数 ( $\epsilon \ll \sqrt{\sum_i x_i^2}$ )、 $\mathbf{z}$  を正規化した特徴ベクトルと置くと、L2 正規化は式 (6) となる。すべてのブロックで作成して連結したものが HOG 特徴量である。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (2)$$

$$\theta(x, y) = \tan^{-1} \frac{f_x(x, y)}{f_y(x, y)} \quad (3)$$

$$f_x(x, y) = L(x+1, y) - L(x-1, y) \quad (4)$$

$$f_y(x, y) = L(x, y+1) - L(x, y-1) \quad (5)$$

$$\mathbf{z} = \frac{\mathbf{x}}{\sqrt{\sum_i x_i^2 + \epsilon}} \quad (6)$$

### 2.2.3 素材識別器の学習

本研究では、2.2.2 で説明した特徴量を使い、素材識別器を SVM (Support Vector Machine) [11] で学習する。ここで、SVM は 2 クラス判別のための教師あり学習によるパターン認識手法である。

SVM の学習手順は以下の通りである。料理レシピから抽出した素材一覧で、料理画像に識別したい素材が含まれるか否かを調べる。調べた結果からその素材が含まれているか否かを料理画像のラベルとする。料理画像から算出した特徴量とラベルを学習データとして SVM で料理画像中に含まれる各素材の識別器を構築する。

### 2.3 識別段階

学習で構築した識別器を用いて料理画像に各素材が含まれるか否かを識別する。識別処理の流れを図 1 の右側に示す。まず料理画像から特徴量を算出する。そして、識別器を用いて料理画像中の各素材を識別する。

#### 2.3.1 特徴量の抽出

素材識別器の学習と同様に、料理画像の中央部分を切り出し、各特徴量を求める。

#### 2.3.2 素材の識別

各素材の素材識別器を用いて、入力された料理画像に対して含まれると識別された素材の一覧を出力する。

表 1 識別対象の素材と学習・評価にポジティブサンプルとして用いるレシピ数

素材名	学習用データ (28 万件中)	評価用データ (7 万件中)
タマゴ	43,916	10,884
ニンジン	19,461	8,768
タマネギ	18,745	4,658
ニンニク	16,343	5,640
ジャガイモ	15,083	3,730
トマト	14,832	4,793
キュウリ	11,681	2,692
ピーマン	9,991	2,453
モヤシ	9,432	2,321
ホウレンソウ	7,676	1,851

### 3. 実験

#### 3.1 実験条件

本節では、使用したデータセット、特徴量と識別器のパラメータ、実験方法、評価方法について述べる。

##### 3.1.1 実験データ

データセットとして、楽天レシピ [12] から提供された料理レシピ 44 万件から、一般に料理とはやや異なるカテゴリである「おやつ」、「飲み物」、「お弁当」を除いた 35 万件のデータを使用した。本実験で用いる料理レシピには、必ず料理画像と素材一覧が含まれている。このデータセットを 2 つに分け、学習用に 28 万件、評価用に 7 万件を用いた。本来は様々な種類の素材について識別器を作成すべきであるが、本実験ではデータセットに多数出現する代表的な素材 10 種類を識別対象とした。識別対象の素材名とデータセットに含まれるレシピ数を表 1 に示す。各素材識別器を学習するときのネガティブサンプル数は、各素材のポジティブサンプル数と同数にした。本実験で扱うデータセットには同じ素材でも漢字表記、かな表記、カナ表記など表記のゆれが見られた。そのため、これらを人手で統合して、同一の素材として扱った。

##### 3.1.2 特徴量と識別器のパラメータ

SVM の実装は、Chang らが開発した LIBSVM [13] を利用した。SVM のカーネルは RBF (Radial Basis Function) を利用した。

##### 3.1.3 実験方法

3.1.1 で紹介した学習用データセットを用いて素材ごとに識別器を構築した。そして、構築した素材識別器で評価用データセットを識別した。提案手法の有効性を調べるために、HSV 色ヒストグラム、HSV 色コリログラム、HOG のそれぞれを特徴量とした識別器とこれらの特徴量を連結した特徴量による識別器の結果を比較した。

##### 3.1.4 評価方法

実験の識別結果から適合率、再現率、F 値を求め、識別器の性能を評価した。以下で適合率、再現率、F 値について詳細を述べる。



図 5 誤ってトマトを含んでいると識別した例 図 6 誤ってタマゴを含んでいないと識別した例

#### a) 適合率

正解の識別結果が正しいものを  $N_t$ 、識別器が正解と識別した総数を  $N_s$  とすると、適合率 (precision) は式 (7) の通りである。

$$\text{precision} = \frac{N_t}{N_s} \quad (7)$$

つまり、識別器により料理がその素材を含むと正しく識別できた割合を表す。

#### b) 再現率

正解の識別結果が正しいものを  $N_t$ 、実際の正解の総数を  $N_m$  とすると、再現率 (recall) は式 (8) の通りである。

$$\text{recall} = \frac{N_t}{N_m} \quad (8)$$

つまり、識別器が素材を含む料理を認識できた割合を表す。

#### c) F 値

一般に、適合率、再現率は各々が高いほど性能が高いことを表すが、一方の評価だけでは実質的な性能はわからない。そのため、性能評価をするためには両者のバランスを考えた評価指標として F 値が用いられる。F 値は適合率と再現率の調和平均であり、式 (9) で求められる。

$$\text{F 値} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (9)$$

#### 3.2 実験結果

実験結果を図 7 にまとめる。各素材をランダム ( $N_t = 0.5 \times$  ポジティブサンプル数,  $N_s = 0.5 \times$  入力データ数,  $N_m =$  ポジティブサンプル数) に識別した結果を超える性能の識別器があったものの、素材推定精度には改善の余地がある。

ランダムに識別した結果に比べて HSV 色ヒストグラムによる識別結果の F 値が高い素材として、トマト、ニンニクがあった。トマトは一般に料理に素材の色が現れやすいものである。そのため、素材の色が現れている料理画像が識別結果に多く見られた。しかし、図 5 のようにトマトに近い色のものが誤識別される例もあった。そのほかにも、タマゴは黄身の特徴が現れる料理が多く、そのような料理の多くは正しく識別されたが、図 6 白身の特徴が現れる料理は正しく識別されなかった。

### 4. 考察

全体的に精度が低かった原因として、いくつかのことが考え

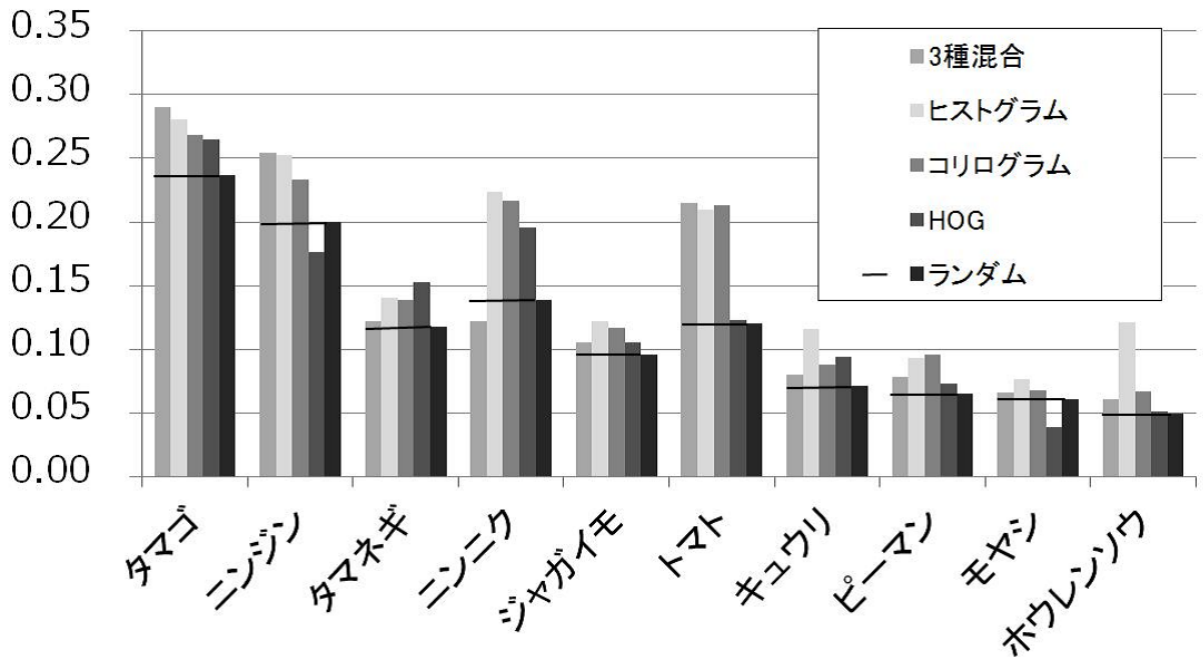


図 7 F 値の比較

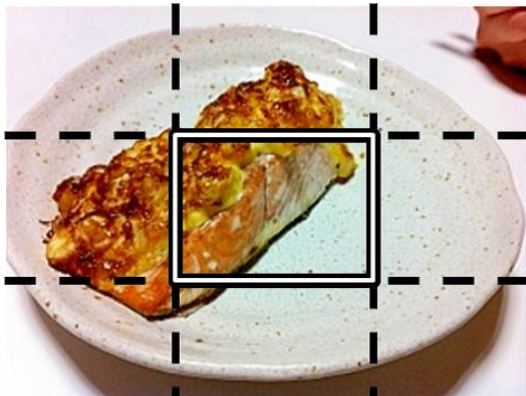


図 8 器の一部が写った例

られる。本研究では器から得られる画像特徴の影響を低減するために、料理画像からの特徴抽出を画像の中央のみから行った。しかし、実際には料理を入れた器の一部が写ってしまうこと(図 8)があり、器の領域除去は不十分であった。また、中央以外にも料理が写った部分は存在する。そのため、料理画像から、より正確な料理の画像特徴の抽出が必要であり、今後、料理領域の切り出し手法の検討が必要である。

また、素材が隠れていたり、色が現れにくいなど、そもそも見た目から識別できない素材もある。そのような素材は料理画像から識別することは困難であるため、識別可能な素材の調査が必要である。

次に識別器ごとの違いについて考察する。図 7 から、用いた特徴量に関わらず、素材によって識別器の性能に大きな差が見られた。また、素材によって有効な特徴量が異なる。そのため、素材ごとに異なる特徴量による識別器を利用する必要がある。また、混合方法として 2 種混合や、各識別器の識別結果を事後統合する方法を検討する必要がある。さらに、本実験で用いた

以外の特徴を検討する必要がある。

3 つの特徴量を混合した特徴量で構築した識別器の精度が、必ずしもほかの特徴量の識別器を上回る結果でなかったことについて考察する。色の特徴量が有効なものとしてニンニクがあったが多く料理で用いられるため、素材の色や形状などの特徴が現れにくい。これはニンニクを含む料理に偏りがあるため、ニンニク自身の特徴ではなく、料理の見た目の特徴を学習することで識別可能になったと考えられる。精度自体は高かったニンジンとトマトの色は似ているため、どちらか片方を含む場合はもう一方も含まれると判断される割合が高かった。この 2 つを識別するために素材の共起性を用いることが考えられる。料理には決まった種類の素材が入っていることが多い。例えば、一般に肉じゃがにはニンジンとジャガイモが入っている。前述のように、ニンジンとトマトが同時に識別された場合、ジャガイモも含まれると識別されていたら、ジャガイモとニンジンは同時に料理に使われることが多く、ジャガイモとトマトが使われることが少ないという知識から、識別結果をニンジンだけに修正できる。

タマネギは HOG を用いた識別器による F 値が最も高かった。タマネギの形状はニンニクに似ているが、用いられる料理の種類がより多様なため、色による識別器の性能が低くなったと考えられる。しかし、様々な素材と一緒に混ぜることが多く、エッジが多いものが識別されたと考えられる。そのため、カレーのようにタマネギが溶けこんで形状を確認できない料理にはタマネギが含まれていると識別されなかった。一般には、調理によって素材の形状が変化するため、HOG は素材の識別にあまり有効な特徴ではなかった。そのため、3 つの特徴量を混合した特徴量で構築した識別器の精度が、HSV ヒストグラムを用いた識別器よりも低くなったと考えられる。

## 5. む す び

本原稿では、大量の画像付き料理レシピを用いて、入力された料理画像からその料理に含まれる素材を推定する手法を提案した。

提案手法では、画像付き料理レシピから料理画像と素材一覧を抽出し、料理画像から特徴量を算出した。そして、含まれる素材を教師信号とすることで素材識別器を構築した。まず、器ごと写した料理画像に対応するため、料理画像を9分割して切り出した中央部分から特徴量を抽出した。料理には使用する素材の色や形状が現れることも多いため、特徴量としてHSV色ヒストグラム、HSV色コリログラム、HOGの3つを用いた。抽出した特徴量からSVMにより学習し、素材ごとに識別器を構築した。

実験では、各素材の識別器を構築して性能を調査した。実験結果として、ほとんどの識別器のF値はランダムに識別した結果を上回り、有効性を確認した。また、素材毎に有効な特徴量が異なることも確認した。そのほかにも、素材識別器の精度には素材により大きな差があった。素材によっては、画像特徴から識別しやすいものとそうでないものが存在したことがわかる。本実験で対象とした素材数は10種類であったが、他の素材についても識別器を構築し有効な特徴量を調査する必要がある。

**謝辞** 本研究では楽天株式会社[12]から提供された楽天レシピデータを利用した。

## 文 献

- [1] エヌ・ティ・ティレゾナント株式会社, “goo からだログヘルスアシスト”, <http://karada.goo.ne.jp/help/app/> (2013/2/10 参照)
- [2] foo.log Inc., “Food-Log”, <http://www.foodlog.jp/> (2013/2/10 参照)
- [3] T. Miyazaki, G.C. de Silva, and K. Aizawa, “Image-based calorie content estimation for dietary assessment,” Proc. 2011 IEEE International Symposium on Multimedia, pp.363–368, Dec. 2011.
- [4] 甫 足創, 松田 祐司, 柳井 啓司, “候補領域推定による複数品目に対応した食事画像認識,” 画像の認識・理解シンポジウム (MIRU) 2011 論文集, IS1-24, Jul. 2011.
- [5] 紺谷 精一, 田辺 勝義, 内山 匡, “複数特徴量を統合した visual words による料理画像の分類,” 信学技報, 2010-LOIS-110, Jul. 2010.
- [6] 森本 雅和, 三好 卓也, 藤井 健作, “マイナー成分分析を用いたパンの画像識別,” 信学論 (A), vol.J94-A, no.7, pp.548–551, Jul. 2011.
- [7] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, “Food recognition using statistics of pairwise local features,” Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.2249–2256, Jul. 2010.
- [8] CG-ARTS 協会, デジタル画像処理, 日興美術株式会社, 2004.
- [9] J. Huang, S Ravi Kumar, M. Mitra, W. Jing, and Z. Zabih, “Image indexing using color correlogram,” Proc. 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.762–768, Jun. 1997.
- [10] N. Dalal and W. Triggs, “Histograms of oriented gradients for human detection,” Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.886–893, Jun. 2005.
- [11] V.N. Vapnik, “The nature of statistical learning theory,” Springer, 1998.
- [12] 楽天株式会社, “楽天レシピ”, <http://recipe.rakuten.co.jp/> (2012/2/10 参照)
- [13] C.C. Chang and C.J. Lin, “LIBSVM: A library for support vector machines,” ACM Trans. on Intelligent Systems and Technology, vol.2, no.27, pp.1–27, Apr. 2011.
- [14] itseez, “OpenCV”, <http://opencv.org/> (2013/2/10 参照)