

G-09

ユーザ調理環境の自動登録に向けた調理道具の識別手法の検討

Study of identification method of cooking tools for the automatic registration of the user cooking environment

吉藤 大貴† 宮森 恒†
Hiroataka Yoshifuji Hisashi Miyamori

1. はじめに

近年、インターネット環境の普及とともにウェブサイト上で Cookpad(クックパッド)[1]のような数多くの料理レシピ検索サイトが存在し、日常的に利用されている。それに伴い、料理レシピの検索においても様々な手法が提案され、ユーザが目的に応じた料理レシピを効率よく検索することが出来るようになり、大変便利なものとなってきている。

現在の料理レシピ検索の多くは、使いたい食材名などをキーワードとして入力することで、その食材を含むレシピを検索結果として返すが、ユーザ個別のニーズに合わせた、より高度な検索を実現しようとする、ユーザ属性をシステムに何らかの手段で入力する必要があり、ユーザにとって大きな手間がかかる点が問題となる。例えば、各家庭の冷蔵庫にある食材を考慮した検索[2]を実現するには、買い物直後や料理をした後に、その時点での冷蔵庫内の食材情報を登録・更新する必要がある。また、各ユーザの調理環境を考慮した検索[3]を実現するには、各ユーザが台所で使用しているフライパンや鍋などの調理道具を少なくとも一度はシステムに登録する必要がある。これらの登録作業を全て手入力で行うのは、ユーザに過剰なシステム操作や時間的な負担を要求するものであり、自動化をはじめとする何らかの効率化を図る手段が必須となる。

一方、近年では個人でのスマートフォンやタブレットをはじめとするカメラ付き携帯情報通信端末の保有率が増加しており、身近な風景を手軽に撮影し、ネットを介したシステムへの登録作業が容易に行える環境となってきている。

そこで、本稿では、ユーザの調理環境の自動登録の課題を取り上げ、カメラ付き携帯情報通信端末等で撮影されたユーザの調理道具画像からその調理道具名を識別する手法を検討する。

本稿の構成は以下の通りである。2 章では関連研究について、3 章では提案システムについて説明する。4 章では実験により、調理道具を対象とした画像についての識別精度を調査し、それに伴う結果と考察について述べる。5 章では、今後の課題について述べ、6 章でまとめる。

2. 関連研究

料理レシピ検索の従来研究として、冷蔵庫食材を考慮した料理レシピ検索の研究がある[2]。冷蔵庫食材の分量や賞味期限を考慮したレシピのランキング手法が検討されているが、ユーザによる冷蔵庫食材の登録や更新が必要であり、その負担を軽減する効率的な登録・更新方法が課題となっている。

また、ユーザの調理環境を考慮した料理レシピを検索する研究がある[3]。調理道具と調理動作のペアに難易度を付与し、ユーザの調理技能や調理環境に、よりマッチした料理レシピ検索について検討されている。ここでも、ユーザの所持する調理道具を登録する手段が、現状では、チェッ

クボックスによる手動入力となっており、ユーザの負荷を軽減する何らかの自動登録手段の実現が有用であると期待される。

一方、画像中に映っている各物体を認識し、その一般的な名称を認識結果として出力する一般物体認識は、コンピュータビジョンの重要な研究課題の一つとしてこれまでに数多くの研究がおこなわれている[4][5]。現在、「顔」「自動車」「ライオン」などの様々な対象に対して、大量の学習データを用意し、領域ごとのラベリング結果や局所的な特徴の組合せに対して、統計的機械学習を適用する手法が数多く提案されている。特に、料理に関連する画像を対象とした認識についても盛んに研究が行われている。食堂の定食のように、複数品目の料理が複数の皿に盛り付けられた食事画像を対象とした認識手法が提案されている[6]。料理部分を候補領域として検出し、複数の視覚的特徴を統合して利用した SVM による分類を行うことで、85 種類の料理について 60%程度の分類率を達成している。

また、スマートフォンを用いて撮影された食材画像を対象とした認識を行うことで、その食材を使ったレシピを推薦するシステムが提案されている[7]。30 種類の食材に対して 45%程度の認識率を達成している。さらに、食器洗浄作業の自動化を想定し、上部から撮影した単一の皿やカップ等が映った食器画像を対象として、輪郭特徴量の累積ヒストグラムによりマッチングする手法が提案されている[8]。9 種類の食器について 90%以上の認識率が得られたとされている。

さらに、ロボットによる一般的な家事支援として調理行動をとりあげ、ロボットによる調理道具・野菜の発見と状態認識を試み研究が報告されている[9]。「ボウル」や「包丁」といった調理道具やまな板上の食材を、画像中の色や特定の幾何形状といった基礎的な画像特徴の組み合わせで認識する手法が報告されている。

本稿では、スマートフォンで撮影された調理道具画像を対象として、該当する調理道具名を識別することを目的とする。

3. 提案手法

本稿で検討するユーザ調理環境の登録システムの概要を図 1 に示す。ユーザが所持している調理道具を、カメラ付き携帯情報通信端末で撮影することで、調理道具が識別され、その結果が、ユーザ毎の調理道具環境として登録される。登録された調理道具環境は、他のレシピ検索や推薦と適宜連携、より高度な検索や推薦を実現するために活用することができる。

本稿では、上記登録システムのうち、調理道具を撮影した画像を入力として、調理道具名を識別する手法について検討する。

3.1 調理道具の識別手順

調理道具画像の識別手順を図 2 に示す。

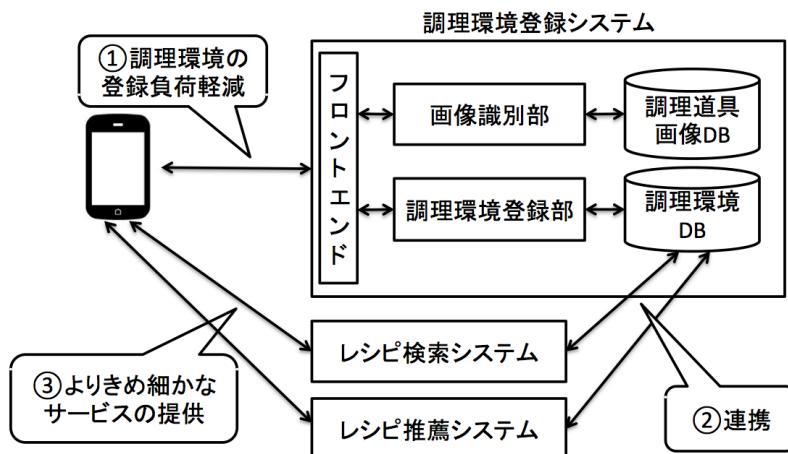


図1 システム概要

まず、Web より、画像検索 API を利用して、調理道具名が既知の調理道具画像を収集する。次に、各調理道具画像に対して、画像特徴量を計算し、調理道具名とともに調理道具画像 DB に登録する。

ユーザによる調理道具識別を行う際には、名前が未知の調理道具を撮影した画像が入力され、上記と同様の手順で画像特徴量が計算される。得られた画像特徴量を用いて、調理道具画像 DB 内の画像群を k 近傍法で探索し、該当する調理道具名を識別結果として出力する。

3.2 調理道具画像の収集

本稿で対象とする調理道具は、野田ら[3]の研究を参考にした。野田らは、クックパッドのレシピ約 43 万件から調理道具名 966 種類を抽出し、表記ゆれや打ち間違いを修正した 202 種類の調理道具名を使用している。本研究で検討するユーザ調理環境の登録システムにおいても、この 202 種類の調理道具名を識別し登録できるようにすることを目標とする。ただし、本稿においては、第一段階として、このうち、最低限必要とされる調理道具 23 種類を、一般ユーザの生活の知恵・知恵が掲載された nanapi(ナナピ)[10]を参考に選択した。表 1 に、本稿が対象とした調理道具名の一覧を示す。

次に、表 1 の各調理道具名に該当する画像を収集する。画像収集には、検索エンジン Bing の検索 API を利用し、表 1 の各調理道具名をクエリとして与え、それぞれ約 1000 件ずつ画像を収集した。写っている調理道具の向きや角度に関しては特に制限は設けていない。収集した画像には、調理道具を学習させる上で必ずしも適当とはいえない画像が含まれており、それらを人手で判断し除外した。具体的には、複数の調理道具が写り込んでいる画像、文字やイラストが混在する画像、調理道具の一部が他の物体に隠され、全体の形を捉えられない画像、調理道具の一部が画像から見切れている画像を除外した。除外した画像の例を図 3 に示す。

以上より、調理道具画像は、各調理道具につき、約 60~420 件取得した。

表 1 調理道具名と収集した画像件数

調理道具名	件数[件]
片手鍋	422
オーブドースタ	370
フライパン	340
ボウル	331
ピーラー	277
フードプロセッサ	255
炊飯器	240
計量カップ	220
圧力鍋	219
包丁	213
おろし金	196
スプーン	182
タジン鍋	144
泡立て器	143
計量カップ	138
フライ返し	130
鍋	125
フォーク	101
菜箸	95
ざる	80
おたま	75
杓文字	75
キッチンばさみ	60

3.3 特徴量抽出

収集した調理道具画像から識別に有用と思われる特徴量を抽出する。本稿では、類似画像検索において用いられる以下のいずれかの特徴量を使用し、それぞれどの程度の識別性能が得られるかを調べることにする。

- Color histogram

画像中の色分布を表すヒストグラム。RGB 色空間を 64 ビンに離散化することで生成する。2 つの色ヒストグラムの類似性を測る尺度には、Jensen-Shannon divergence を用いる[11]。

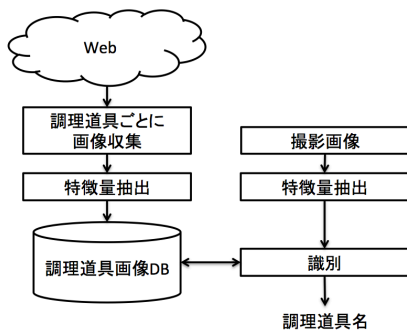


図2 処理手順



図3 除外した画像の例

• Color layout

高速な検索のため、解像度に依存しないよう色をコンパクトに記述した特徴量[12]. 画像を64個のブロックに分割し、各ブロックにおける色の平均を代表色として取得する. 2つのColor layoutの類似度計算は、MPEG-7で定められた方法に準じる[12].

• Scalable color

HSV色空間での色ヒストグラムにHaar変換を適用し、コンパクトに記述したもの[12]. 256ピンの色ヒストグラムを4ビットで非線形量子化し、得られたヒストグラムにHaar変換を適用する. 2つのScalable colorの類似度計算には、L1-ノルムを用いる.

• Edge histogram

画像の局所的なエッジ特徴の分布を表す[12]. 画像を4×4ブロックに分割し、各ブロックでエッジを検出し、垂直、水平、斜め45度、斜め135度、エッジなしの5タイプのいずれかに分類する. 2つのEdge histogramの類似度計算には、L1-ノルムを用いる.

• Tamura texture features

人間の視覚に基づいたテクスチャ特徴[13]. 本稿では、coarseness, contrast, directionalityの特徴を用いてテクスチャヒストグラムを生成する. 2つのTamura texture featuresの類似度計算には、L2-ノルムを用いる.

• CEDD

色とエッジの方向性を記述したヒストグラム[14]. 画像はいくつかのブロックに分割され、ブロックごとに、異なる色を表現するヒストグラムと、エッジに関するヒストグラムが抽出される. 2つのCEDDの類似度計算には、Tanimoto係数を用いる[14].

• FCTH

Fuzzy colorとテクスチャ特徴を記述したヒストグラム[15]. 画像はいくつかのブロックに分割され、ブロックごとに、3つのファジー処理を順に施され、192ピンのヒストグラムが抽出される. 2つのFCTHの類似度計算には、Tanimoto係数を用いる[15].

4. 実験

4.1 実験目的

調理道具画像に対する各特徴量での識別精度を検証する. 被検索対象となる23カテゴリの調理道具画像に対し、テストで識別する調理道具名のカテゴリ数を1,3,5,10,23と変化させて調べる. 例えば、テストカテゴリ数が1の場合、23カテゴリの調理道具画像を対象として、特定の1カテゴリの調理道具名に該当する画像であるかどうかをテストすることになる. これを5-fold交差確認で調べることにする. さらに、k近傍法での識別におけるパラメータkについて、1,3,5,7,9と変化させ、特徴量については、3.3節で説明した7種類の各特徴量について識別精度を調べる.

4.2 実験結果

実験結果を図4~図8に示す. 図中の識別精度は、各テストカテゴリ(の組合せ)による識別精度のマクロ平均を示している. 今回は、時間の都合上、テストカテゴリ数が1,3,5,10の実験については、23カテゴリからランダムに選んだ3通りのテストカテゴリあるいはその組合せによる識別精度のマクロ平均を算出している.

実験結果から、どの特徴量を用いた場合でも、テストカテゴリ数が大きくなるにつれて識別精度が低下することが確認できた. テストカテゴリ数1の場合、最高精度が約0.69となり、テストカテゴリ数が23の場合、最高精度で約0.47となった.

Edge histogramを用いた場合、カテゴリ数に関係なく、今回の特徴量の中で最も高い精度が得られた.

次に、Color layout, CEDDが比較的良好な精度を示した. テストカテゴリ数が小さい場合はCEDDがよくなる場合も見られたが、テストカテゴリ数が大きくなると総じてColor layoutの方がよい結果を示した.

その次に良い精度となったのは、テクスチャ特徴を用いたTamura, FCTHであった. テストカテゴリ数が小さい場合はTamuraの方がよい結果を示したが、テストカテゴリ数が大きくなるとFCTHの方が良好な結果となった.

最も精度が低かったのは、Color histogram, Scalable colorであった. 特に、Scalable colorによる識別結果では、ほとんどの条件において、精度が0.1を下回った.

k近傍法におけるkの値が1,3,5,7,9と5つの場合で実験を行ったが、結果として特に大きな差異は見られなかった. ただし、テストカテゴリ数1の場合は、ほとんどの特徴量で、kの値が大きくなるにつれて識別精度が上がる傾向にあることが確認された.

4.3 考察

実験結果より、今回用いた特徴量の中では、Edge histogramが最も精度が高い結果となり、次に良好な精度を示したColor Layout, CEDDと比べ、ほぼ全ての場合で0.1~0.2の精度差が見てとれることがわかる. これは、

調理道具の識別においては、色の空間的配置や、色とエッジ特徴の組合せといった特徴量も、ある程度は有効に機能するものの、局所的なエッジ特徴のみの特徴量を用いる方がはるかに効果的であることを示している。

一方、テクスチャ特徴については、今回用いた特徴量の中では中程度の識別精度となり、色とテクスチャを組み合わせた場合もあまり違いがないことが分かった。これは、調理道具のテクスチャが、各調理道具を十分に区別できるほど多種多様ではないためであると考えられる。

色情報のみを用いる Color histogram, Scalable color の場合、識別精度は条件によらず最も低い識別精度となった。これは、一般に、各調理道具に対して固有の色というものがないこと、撮影条件により背景を含めて色が大きく変化し得ることが主な要因であると考えられる。現在の調理道具にはカラフルなものが多いことを考えると、調理道具の識別に色情報を利用することは適切ではないことが分かる。

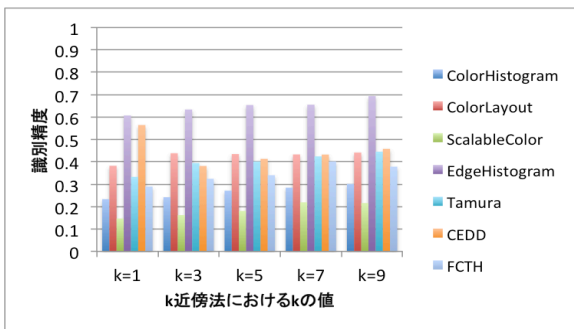


図4 各特徴量での識別精度 (テストカテゴリ数 1)

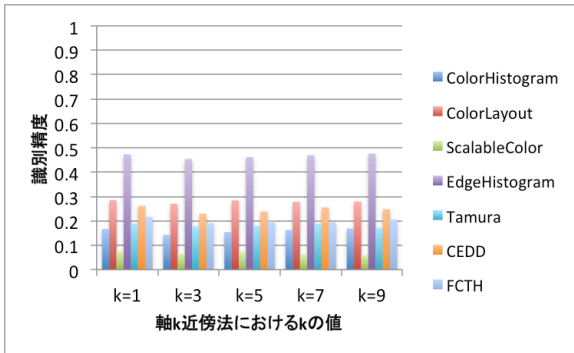


図5 各特徴量での識別精度 (テストカテゴリ数 3)

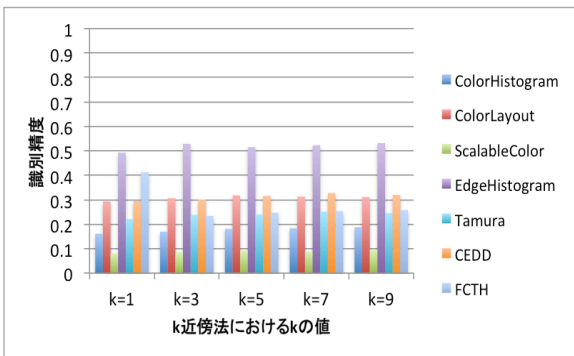


図6 各特徴量での識別精度 (テストカテゴリ数 5)

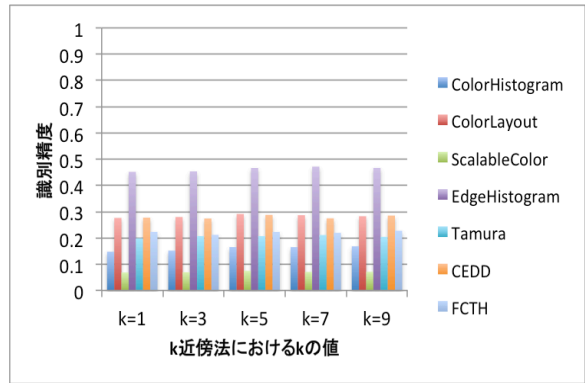


図7 各特徴量での識別精度 (テストカテゴリ数 10)

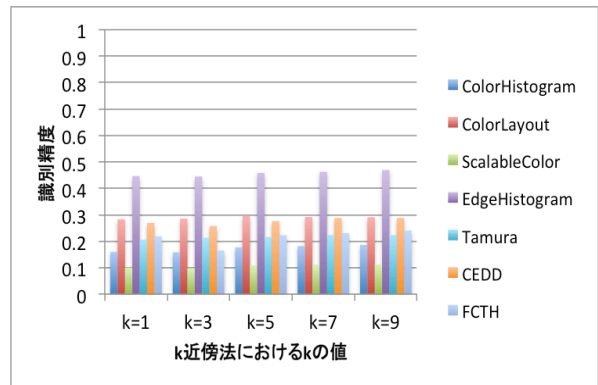


図8 各特徴量での識別精度 (テストカテゴリ数 23)

5. 今後の課題

本稿の実験では、収集した調理道具画像が数百枚程度と少ないという問題点がある。今後、識別精度を向上させるために、より多くの画像を効率よく収集する必要がある。

また、SIFT や SURF といった局所特徴量を用いた Visual Word による識別手法も検討する必要がある。

6. まとめ

本稿では、レシピ検索やレシピ推薦における、ユーザの調理環境の自動登録の課題を取り上げ、カメラ付き携帯情報端末等で撮影されたユーザの調理道具画像から、その調理道具名を識別する手法を検討し、様々な特徴量による識別精度を明らかにした。その結果、調理道具の識別においては、色の空間的配置や、色とエッジ特徴の組合せといった特徴量がある程度は有効に機能するものの、局所的なエッジ特徴のみの特徴量を用いる方がはるかに効果的であることが分かった。

今後、調理道具画像をより効率よく収集し、SIFT や SURF といった局所特徴量を用いた Visual Word による識別手法等を検討する予定である。

参考文献

- [1] レシピ検索 No.1/料理レシピ載せるならクックパッド <http://cookpad.com>
- [2] 赤澤康幸, 宮森恒. 冷蔵庫食材を考慮した料理レシピ検索システムの提案, DEIMForum2012E1-2, 2012年3月
- [3] 野田真, 宮森恒. 料理レシピにおける調理動作の道具別難易度付与の一検討, データ工学研究会, 2012, 信学技報, vol. 112, no.5, DE2012-3, pp. 13-18, 2012年6月
- [4] 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, Vol.48, No.SIG16(CVIM19)2007年11月
- [5] 秋山瑞樹, 柳井啓司. 特定物体認識手法による大量画像を用いた一般物体認識, 画像の認識・理解シンポジウム(MIRU2010), 2010年7月
- [6] 甫足創, 松田裕司, 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(MIRU2011), 2011年7月
- [7] 丸山拓馬, 秋山瑞樹, 柳井啓司. 食材画像認識を用いたレシピ推薦システム, 電子情報通信学会技術研究報告, 信学技報 111(478), 43-48, 2012年3月
- [8] 野村英祐, 林悠, 鏡慎吾, 橋本浩一. 画像上の輪郭特徴量を用いた食器認識システム, 計測自動制御学会東北支部第 236 回研究集, 資料番号 236-12, 2007年6月
- [9] 渡辺義明, 長濱虎太郎, 〇正, 山崎公俊, 岡田慧正, 稲葉雅幸. 一般調理道具を扱う等身大ヒューマノイドの認識行動システム統合とサラダ調理の実現, ロボティクス・メカトロニクス講演会講演概要集 2010, “2A2-A17(1)” - “2A2-A17(4)”, 2010
- [10] 生活の知恵があつまる情報サイト nanapi
<http://nanapi.jp/cook/1746>
- [11] Thomas Deselaers, Daniel Keysers, and Hermann Ney. 2008. Features for image retrieval: an experimental comparison. *Inf. Retr.* 11, 2, pp.77-107, 2008.
- [12] P. Salembier and T. Sikora. Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley& Sons, Inc., New York, NY, USA, 2002.
- [13] H. Tamura, S. Mori and T. Yamawaki: “Textural features corresponding to visual perception”, *IEEE Transactions on Systems, Man and Cybernetics*, 8, 6, pp. 460-473, 1978.
- [14] S. A. Chatzichristofis and Y. S. Boutalis: “Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval”, *ICVS*, pp. 312-322, 2008.
- [15] S. A. Chatzichristofis and Y. S. Boutalis. “FCTH: Fuzzy Color And Texture Histogram A Low Level Feature For Accurate Image Retrieval”, *WIAMIS 2008*, pp.191-196, 2008.