

食材画像認識を用いたレシピ推薦システム

丸山 拓馬[†] 秋山 瑞樹[†] 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 東京都調布市調布ヶ丘 1-5-1

E-mail: †{maruya-t,akiyama-m}@mm.inf.uec.ac.jp, ††yanai@cs.uec.ac.jp

あらまし 本論文では、モバイルデバイスでの画像認識を利用したレシピ推薦システムを提案する。提案システムはスマートフォン側で食材の一般物体認識をリアルタイムに行う。食材にスマートフォンをかざすだけで次々とレシピを推薦するので、従来のキー入力のみシステムよりも直感的で簡単なレシピ検索が可能となる。一般物体認識手法としてカラーヒストグラム、SURF を用いた Bag-of-Features 表現を採用している。実験では 30 種類の食材を対象にしてタッチのみで操作するレシピ検索システムとの比較を行い、ユーザによる評価を実施した。また画像認識では 44.9% の精度で目的の食材を認識でき、上位 5 位までを考慮すると 80.9% を達成することを確認した。

キーワード モバイル, スマートフォン, 画像認識, 一般物体認識, レシピ

Takuma MARUYAMA[†], Mizuki AKIYAMA[†], and Keiji YANAI^{††}

[†] Graduate School of Informatics and Engineering, The University of Electro-Communications

1-5-1 Chofugaoka, Chofu, Tokyo 182-8585 Japan

E-mail: †{maruya-t,akiyama-m}@mm.inf.uec.ac.jp, ††yanai@cs.uec.ac.jp

Key words mobile, smartphone, object recognition, generic object recognition, recipe

1. はじめに



図 1: 提案システムのイメージ

近年、画像認識技術の進歩がめざましく、特に対象の見え方に違いがある一般物体を対象にした研究分野では Bag-of-Features 手法を始めとして認識率の大幅な改善を促す画期的な手法が提案されている。これにともない OpenCV をはじめとした画像認識のためのライブラリも普及して、高度な最先端の画像認識技術が容易に利用できる機会が増えるようになった。またスマートフォンの世界的な普及と共に、モバイル端末での処理能

力が向上しておりスマートフォン上で画像認識を行うことが可能になった。モバイル上で画像認識を行う研究はさまざまな存在するが、本研究では買い物中のレシピ検索の操作に着眼点を置く。

買い物中は買い物かごを持っていたり子供をひき連れていたり、ちょっとした動作がユーザの負担になってしまうことがある。特にスマートフォンでレシピを確認・検索する際には、従来は複雑なキー入力が必要とすることが多く直感的な操作をすることが出来なかった。

そこで本研究ではスマートフォンを食材にかざすだけでレシピを直感的に容易に選択できるシステムの構築を目指す。図 1 にそのイメージを示す。提案システムでは画像認識によって食材を認識してレシピを検索できるため、従来のタッチだけのシステムに比べてより直感的で容易な操作が可能になっている。システムはスマートフォンアプリケーションとして実装し、画像認識手法としてカラーヒストグラム、SURF による Bag-of-Features 表現を採用するものである。またスマートフォン上で高速に実行可能なシステムを実現するために、線形 SVM を採用して認識時間の高速化を果たしている。

2. 関連研究

ここでは主にモバイルデバイスでの画像認識、食メディアに

関する画像認識、レシピ推薦という観点から関連する研究を紹介する。

モバイルデバイスでの画像認識サービスとして、Google Goggles [1] がある。Google Goggles はスマートフォン上のアプリとして Android と iPhone で動作し、カメラで撮影した文字、美術作品、ロゴ、本、ランドマークなどを対象に画像認識した結果から web 上の情報検索ができる。またユーザが物体情報を登録して認識にフィードバックが可能なサービスとなっている。ただし、Google Goggles は対象を広く認識できるように特定物体の認識や類似画像検索によるところが大きく、見た目の異なるような一般物体に対しては認識が困難である。またスマートフォン上でリアルタイムに特定物体認識を実施する研究として、Lee らの研究 [2] が挙げられる。Lee らは iPhone 上で特定物体認識をリアルタイムに行い、ユーザがその場で認識物体を登録して特定物体を検索可能にするシステムを提案している。モバイル上でリアルタイムに認識を実行するために、独自の方向と強度に関する記述子を定義して、スケールごとのテンプレートによるマッチングを実施している。これら研究とサービスは特定物体に限定した認識であり、本研究では一般物体である食材画像に対応しているという点で大きく異なる。

次に食メディアに関する画像認識の研究を取り上げる。食材画像認識を取り扱った研究としては、京都大学美濃研究室のスマートキッチン（ユーザ主導型調理支援システム）[3] がある。スマートキッチンに関連した具体的な研究として、橋本らの食材のテクスチャ特徴より調理台上の食材を観測・識別する研究 [4]、外見の変化モデルを利用した調理中の食材追跡 [5]、森らの調理者の手の動きを時間制約として食材追跡 [6] などが存在する。これらスマートキッチンに関連する研究と本研究とでは調理動作ではなくレシピ選択過程においてのユーザ支援を目的としている点で異なる。

最後に食材情報からレシピ推薦する研究事例を紹介する。余剰食材の使い切りを考慮したレシピセットの提案に関する研究として木原 [7]、鄭 [8]、赤澤 [9] の研究、また高畑らの好き嫌いを考慮したレシピセットの推薦方法の研究 [10] などがある。とくに赤澤らの研究では冷蔵庫の食材を考慮してレシピセットを推薦するシステムを提案しており、分量、賞味期限、最終利用日について食材の情報を記録してそれに基づいたランキングを算出して表示している。本研究ではこれら研究のような食材の詳細な情報を用いてレシピ推薦を行うまでには踏み込んではいないが、今後システムを改良する際に参考にしたいと考えている。

3. 提案システム

本章では、提案システムの内容について説明する。

3.1 方針

本研究の目的は、食材画像認識によってユーザのレシピ決定動作を支援することである。そこでまず提案するシステムがどのような用途と目的で利用されるかを確認し、その方針を明らかにする。

ユーザが食事を作成することを決定して、料理を食べるまで

の過程を確認すると、図 2 のようになる。例えばレシピを決定せずにスーパーに行ってその場で購入すべき食材を決定するなど、人によっては食事を作ることを決めてから実際に調理するまでの過程に変更の余地があるのでこの過程は十分に厳密なものではないが、その要素は概ね現実に即した過程だと考えられる。ここでなぜ 4 つめのレシピの確認と変更の項目が必要になるのかというと、実際にスーパーに買い物に行った際にはその時の値段の食材を考慮してユーザの気分が変わる事や目的のレシピに必要な素材が販売されていないケースやレシピの購入項目を確認する場合などがよくあると考えられるからである。提案するシステムではこの 4 番目の行為に着目しており、ユーザがスーパーで目的のレシピを得るまでの操作を画像認識によって改善することに主題を置いている。

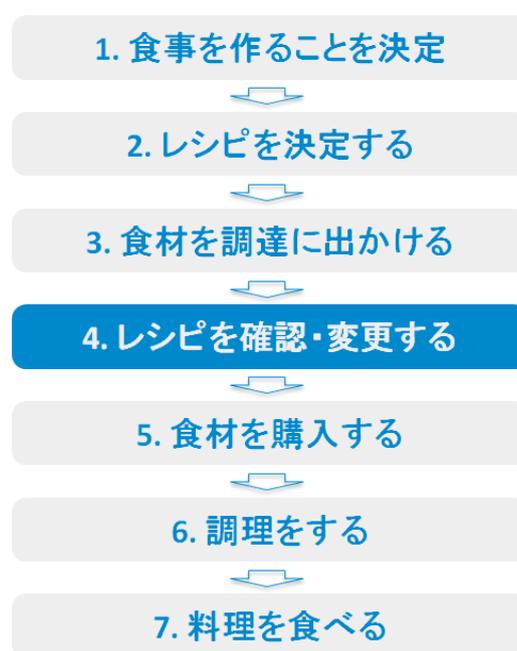


図 2: 食事作成の流れ

3.2 システムの流れ

システムが買い物中のレシピ選択に視点を置いていることを先に説明したが、ここではシステムの概要をレシピ選択の具体的な利用過程に沿って説明する。図 3 にその様子を示す。

手順 1 まずスマートフォンを食材にかざす。ここでシステムはバックグラウンドで逐次的にカメラ画像のフレームを取得する。

手順 2 各取得フレームごとに今何の食材が写っているのか画像認識を行う。認識結果の食材は食材候補としてランキング化されてシステムに表示される。この際、画面上に一度に 6 位までの食材候補が表示可能である。

手順 3 認識された食材をもとに、レシピサイトから該当の食材名をクエリとして、レシピリストを取得する。

手順 4 食材の認識結果の最上位のものが常にレシピリストに反映され、2 位以下の食材はユーザのタッチ動作で反映される。

手順 5 レシピリストよりユーザは目的のレシピを選択する。



図 3: システムの流れ

手順 6 選んだレシピに対して、作り方や必要食材などの詳細ページを表示する。この時システムはブラウザモジュールを利用しており、内容はもとのレシピサイトの内容を表示するものである。

以上の手順 1 から 6 をもってユーザはシステムを利用する。具体的なシステムの画面は図 4 に示す。手順 2 の食材候補リストは認識が行われると逐次更新されレシピリストに反映される。また実験時に手で選択する場合との比較を行えるように手動選択が可能になっている。

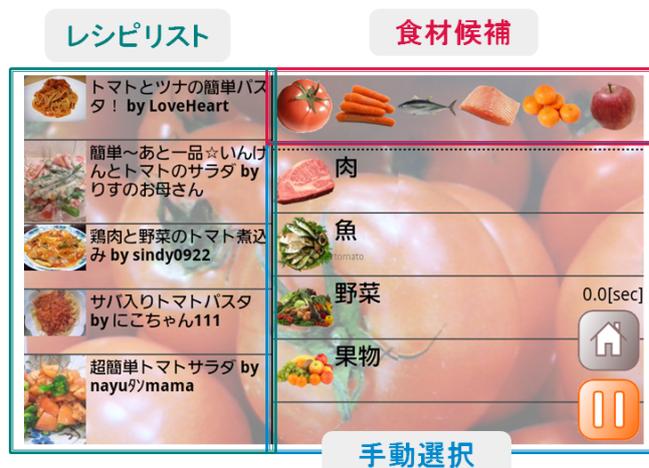


図 4: 画面の様子

a) レシピの取得について

レシピ情報は COOKPAD [11] より取得したものを新着順に表示する。ここでレシピのよりよい提示方法が先に述べた関連

研究でも多数言及されているが、本研究ではその部分についての工夫を行っておらず、単品の食材名のみからレシピリストを取得して表示するにとどまっている。

4. 画像認識手法

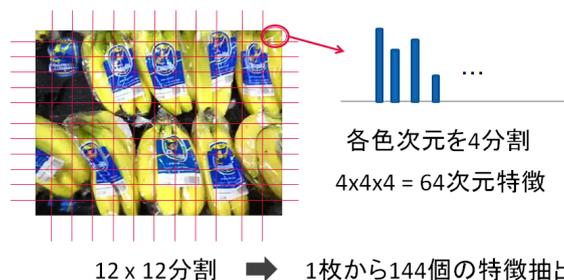
提案システムでは画像認識を用いて食材を認識しレシピ情報を取得するが、ここでは提案システムに利用した画像認識手法について説明する。

4.1 画像特徴

モバイル上で画像認識を行う際に精度と速度の点で重要となるのが画像特徴量である。本研究ではモバイルデバイスでの認識を実施しているために、特に速度とメモリ使用量が重要となる。近年は局所特徴量の改良が頻繁に行われ、BRIEF [12]、ORB [13] を始めとして高速で省メモリなバイナリ特徴が複数提案されている。しかしこれらバイナリ特徴は特定物体のマッチングに用いられる場合が多く、一般物体認識に用いると、Bag-of-Features ベクトルを作成する際に特徴量の情報劣化が懸念されるため本研究ではバイナリ特徴でなく高速な SURF [14] を採用した。

SURF とは照明変化、スケール変化、回転に対して頑健な局所特徴量である。本研究では 64 次元の SURF を採用した。また SURF のキーポイントの検出を利用した場合と、マルチスケールのグリッドで検出した場合を利用している。

また食材の認識を考える際に色による情報が重要となることが考えられるため、グリッドで検出したカラーヒストグラムを採用する。カラーヒストグラムは各式空間の次元を 4 分割した計 64 次元の特徴量を採用した。図 5 にその様子を示す。本研究では画像を 12×12 分割をして 1 枚から 144 個のカラーヒストグラムを作成して Bag-of-Features ベクトルを作成している。



12 x 12分割 ➡ 1枚から144個の特徴抽出
図 5: グリッドでカラーヒストグラムを抽出

4.2 Bag-of-Features 表現

Bag-of-Features [15] は、統計的言語処理のための手法である Bag-of-Words [16] を画像処理に適用したものである。Bag-of-Features は、幅広いカテゴリの画像を対象とした場合の表現性の高さから、一般画像認識の分野で頻繁に用いられる。

本研究では食材をカメラを動かしながら撮影した 1 秒から 6 秒未満の動画を各カテゴリ毎に 10 本づつ用意して、それからランダムに 1000 枚のフレームを選出してコードブックを作成した。コードブックの要素数は 1000 として、画像は 1000 次元の Bag-of-Features 表現ベクトルとなる。特徴量のベクトルの

1-norm は 1 になるように正規化して利用した。

また本研究では認識に動画を用いておりフレームごとに画像を取得できるため、BoF ベクトルは複数の連続フレームで抽出した特徴をまとめてひとつの BoF ベクトルを作成する、マルチフレーム認識を利用して認識率の向上を図った。

4.3 線形 SVM

SVM においてベクトル間の類似度を決定するのがカーネルであり、一般に非線形のカーネル計算を用いることで SVM 分類能力を向上させることが多い。しかしそのかわりに非線形カーネルでの計算は線形カーネルよりも計算コストがかかるため、線形カーネルでの分類を本研究では採用した。ここでは線形カーネルの計算方法について詳しく述べる。

線形カーネルでは式 1 によってサポートベクトルと、テストベクトルの内積を計算する。

$$K(x, y) = x \cdot y \quad (1)$$

入力ベクトルを x 、出力値を $y(x)$ 、サポートベクトルを x_i 、サポートベクトルの重みを w_i 、バイアス値を b としたときに、

$$y(x) = \sum_{i=1}^N w_i K(x, x_i) + b \quad (2)$$

$$= \sum_{i=1}^N w_i x \cdot x_i + b$$

$$= x \cdot \sum_{i=1}^N w_i x_i + b \quad (3)$$

と変形できるため、あらかじめ、サポートベクトルと重みをかけ合わせた

$$v = \sum_{i=1}^N w_i x_i \quad (4)$$

を計算しておくことで、1 回の内積にバイアス値を足すだけで線形分類ができ高速になる。この方法を用いて認識過程におけるさらなる高速化を図った。

5. システムの評価

ここではシステムの評価として、まず実際に販売されている食材とその認識結果の例を示す。また実験においては食材の認識率についての評価とユーザによる評価を示す。

5.1 動作例

システムを Android2.2、Galaxy S2 上で動作させたところ、30 食材を対象にカラーヒストグラム特徴のみで一回の分類に要する時間は 0.15 秒であった。

5.2 実験と考察

実験では東京都内のスーパーマーケット 10 箇所以上で撮影した食材動画を用いた。食材の種類は表 1 に示す 30 種類の食材を利用している。

30 種類の各食材について 10 本ずつ動画を用意して、動画単位の 10-fold cross validation を行った。各動画はすべて 1 秒以上 6 秒未満の VGA サイズの動画になっており、カテゴリ毎

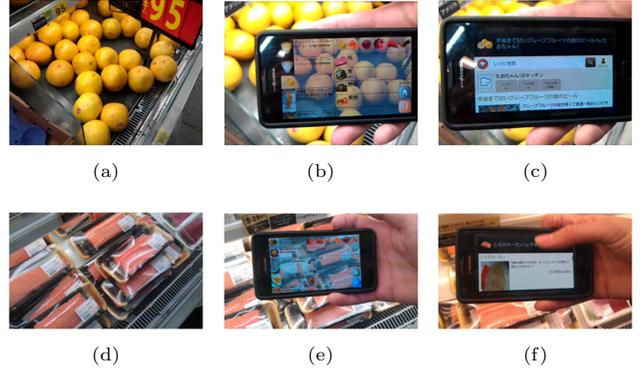


図 6: 動作例：上段はグレープフルーツ、下段は鮭が対象。グレープフルーツでは認識食材の 1 位にグレープフルーツがランキングされ、レシピリストに食材のレシピが即座に反映されている。鮭では 3 位に認識されたので手動で食材を選択した後レシピを表示している。

表 1: 認識する食材の種類

食材の種類	食材
魚	マグロ、イカ、たこ、エビ、鮭
肉	牛肉、豚肉、鶏肉、ひき肉、ウインナー、ハム
野菜	しめじ、じゃがいも、なす、ニンジン、大根、 トマト、きゅうり、キャベツ、ネギ、白菜、レタス、 しいたけ、玉ねぎ
果物	りんご、いちご、パイナップル、みかん、バナナ、 グレープフルーツ

にデータ量に大きな差が出ないように注意している。動画は 25fps でサンプリングした各フレームを認識に用いている。

また実験では以下のようなパラメータを設定して BoF ベクトルを得た。

パラメータの設定

SURF でのグリッド検出

12 × 12, 24 × 24, 48 × 48, 96 × 96 pixel の 4 スケール

カラーヒストグラムのグリッド検出

画像を 12 × 12 分割して BoF ベクトル作成

カラーヒストグラムの種類

RGB, La*b*, HSV の 3 パターン

マルチフレーム

1, 2, 3, 4, 5 フレームで実施

5.2.1 各手法と認識精度

ここではまず 1 フレームに固定して各特徴ごとの精度評価を行い、適用したどの特徴量が食材の認識に適合していたかを検討する。図 7 に各特徴ごとの 30 カテゴリでの分類率を示す。

手法ごとの結果を比較すると SURF 単体での分類率は通常
のキーポイント検出で 26.38%、グリッドでの検出で 19.65%
と全体として高い分類率を得ることができなかった。それに対してカラーヒストグラムでの分類では RGB で 42.53%、HSV
で 34.00%、La*b*で 30.54% と勾配特徴よりも良い結果が得ら

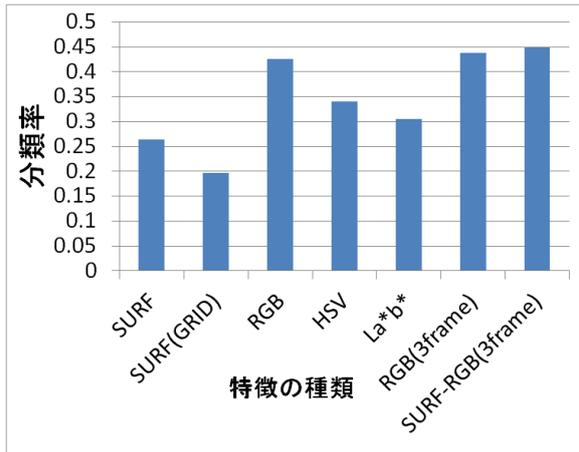


図 7: 各手法ごとの認識結果

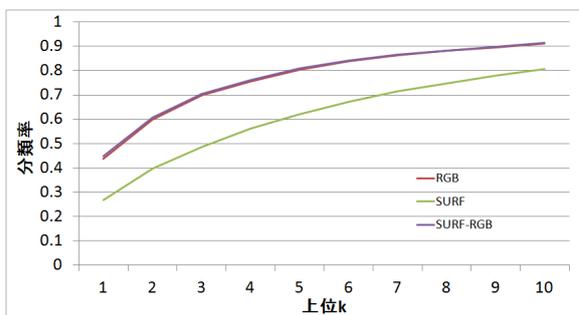


図 8: 上位 k 位まで考慮した際の分類率

れた。

全体として勾配特徴が色特徴に比べてうまくいかなかったのは、図 9 のように袋の反射やカメラのブレが大きな対象が含まれているため、安定したキーポイントと勾配表現が得られなかったためだと考えられる。



図 9: 袋づめで光の反射がある対象やカメラのブレが大きな対象

また、単体でもっとも分類率の高かった RGB カラーヒストグラムでは 3 フレーム特徴を統合した際に 43.78% に上昇し、さらに SURF と組み合わせた場合には 44.92% に精度向上した。また提案システムでは現在認識した結果から上位 6 位まで食材候補枠に表示できるため、上位 k 位まで考慮した分類結果について図 8 に示す。

もっとも分類率の良かった RGB と SURF を 3frame で組み合わせた場合の食材ごとの認識結果を図 10 に示す。特にエビでの結果が悪くこれは冷凍のエビや生のエビなど見た目にはかなりの差がある場合が多かったためだとと思われる (図 12)。一方で単純な見た目のみかんはほとんど変化がなく、色での分類が

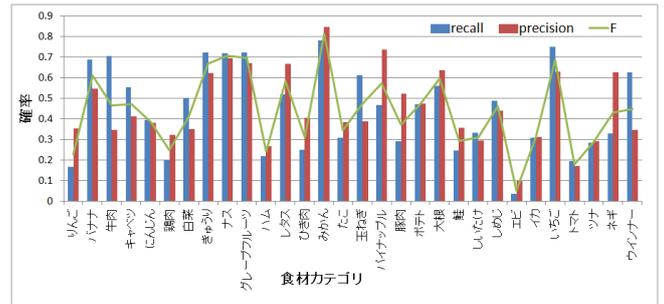


図 10: 各食材ごとの認識精度

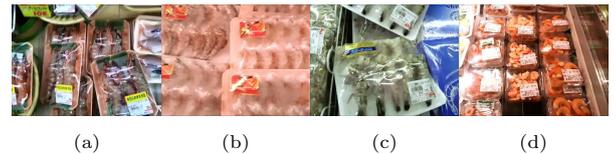


図 11: 間違いが多かった食材：エビ



図 12: 正解が多かった食材：みかん

うまく実行できている。

5.2.2 ユーザによる評価

ここでは提案したシステムの有用性を検証するためにユーザに対して行った評価を示す。実験では研究室のメンバー 5 人を被験者として、研究室にもってきた 3 食材に対して各々 3 レシピを指定してその検索にかかる時間をまず測定した。その後ユーザに対してアンケートを実施して手動選択と画像認識でどちらが利用しやすかったか、システムの使いやすさはどうかであったか、画像認識の精度はどうかであったかについて 5 段階で評価してもらった。

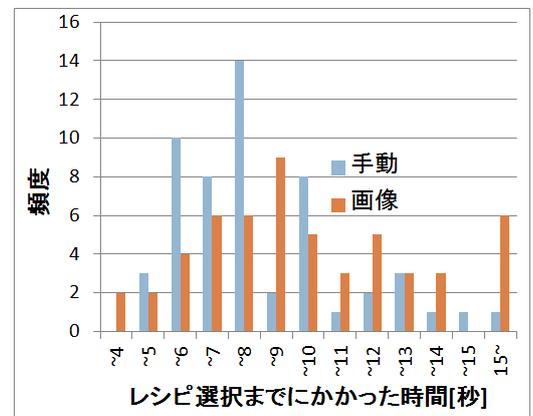
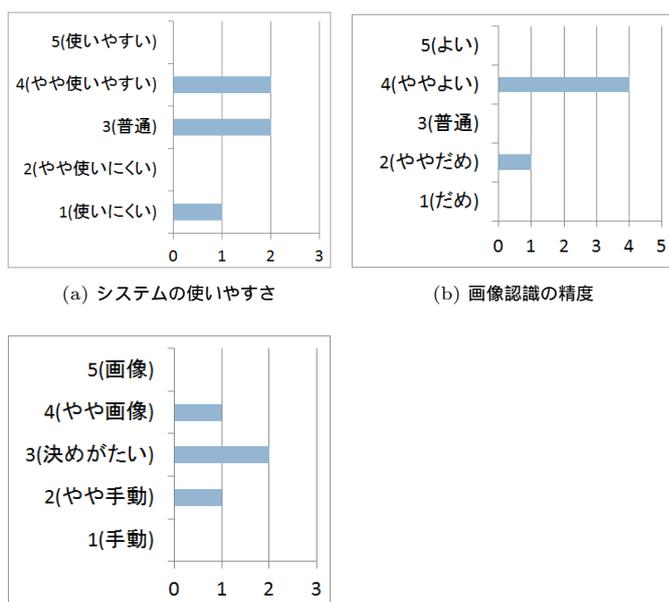


図 13: 手動選択と画像認識でのレシピ選択までの時間

まず手動と画像認識でのレシピ選択までの時間を図 13 に示す。手動での選択と画像認識を利用した場合は、中央値が手

動で 7.29 秒、画像で 8.51 秒となり手動での選択のほうが全体としてはよい結果になってしまった。これは画像での 15 秒以上の外れ値が影響しており、食材によってはなかなか認識できず選択までの時間を伸ばしてしまったためだと考えられる。だが画像認識を利用した場合には 4 秒未満で選択できる場合があり、手動よりも素早くレシピを表示できる結果も存在した。また今回の実験では単一の食材情報からレシピをボタンで選べたため手動選択との差が出なかったが、より複雑な複数食材でのレシピ検索時や条件付きのレシピ検索の際には画像認識を利用したシステムの優位性がでてくるのではないかと考えられる。



(c) 画像利用と手動選択どちらが利用しやすいか

次に利用してもらったユーザのアンケートの結果を図 14a、14b、14c に示す。全体として手動選択よりも完全に使いやすいという評価は得られなかった。ポジティブな意見として、「セール品で作れるレシピをその場ですぐに調べられていい」「認識率が上昇すればこのままでも使いやすい」などの意見が得られ、ネガティブな意見として、「食材候補の更新が激しく選択しづらい」「食材によって認識のしやすさにばらつきがある」という意見が得られた。その他、「複数品目の食材での検索機能がほしい」「レシピをメインディッシュやデザートなどで種類ごとに表示してほしい」「冷蔵庫の品を事前登録しておいて安い食材と組み合わせで推薦してほしい」「値札も認識してレシピを最安値で表示してほしい」などシステムのインターフェースや機能改善に関する意見が多数得られた。

6. まとめと今後

本研究ではスマートフォンを食材にかざすことでレシピを推薦してくれるシステムを提案した。30 種類の食材を用いて実験した結果、44.9% の分類率で目的の食材のレシピリストを得ることができ、また上位 5 個で推薦食材を提示すると目的の食材を 80.9% の精度で表示することが可能である。ユーザによる評

価では必ずしもシステムが有用であると示すことは出来なかったが、今後のシステムを改良する上での有用な意見を得ることができた。

今後はシステムと画像認識の両面で改良を行いたいと考えている。システムでは食材の組合せによる条件検索、レシピの価格情報・栄養情報を考慮したレシピリストの推薦、前菜やスープ、デザートなどのレシピカテゴリでの表示など、より実用的で価値のあるレシピ推薦を試みたい。画像認識部分では色以外の特徴で安定的な認識をできるようにして 1 位で 70%、5 位までに 90% を達成したい。

なお実装したシステムや使用した動画セットは <http://mm.cs.uec.ac.jp/maruyama/recipe/> で公開している。

文 献

- [1] G. Goggles, "Search the real world by taking a picture". <http://www.google.com/mobile/goggles/>.
- [2] T. Lee and S. Soatto, "Learning and matching multiscale template descriptors for real-time detection, localization and tracking," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1457-1464, 2011.
- [3] 京都大学美濃研究室, "スマートキッチン(ユーザ主導型調理支援システム)". <http://www.mm.media.kyoto-u.ac.jp/research/skitchen.html>.
- [4] 橋本敦史, 山肩洋子, 角所 考, 美濃導彦, "D-12-73 テクスチャに基づく食材識別の検討 (d-12. パターン認識・メディア理解, 一般講演)," 電子情報通信学会総合大会講演論文集, vol.2006, no.2, p.205, 2006-03-08.
- [5] 橋本敦史, 森 直幸, 船富卓哉, 山肩洋子, 椛木雅之, 角所 考, 美濃導彦, "把持の順序と外見の変化モデルを利用した調理作業における食材追跡," 電子情報通信学会和文論文誌 A, vol.J94-A, no.7, pp.509-518, 2011.
- [6] 森 直幸, 船富卓哉, 山肩洋子, 角所 考, 美濃導彦, "調理者の手の動きを時空間制約とした調理中の食材追跡," 電子情報通信学会技術研究報告 MVE, vol.107, no.454, pp.45-50, 2008.
- [7] 木原ひかり, 上田真由美, 中島伸介, "余剰食材の使い切りを考慮したレシピ推薦手法の提案," 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM), 2011.
- [8] 鄭 美玲, 井上悦子, 中川 勝, "食材の使い切りを考慮した期間的な料理レシピセットの推薦," 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM), 2011.
- [9] 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," 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM), 2011.
- [10] 高畑麻理, 上田真由美, 中島伸介, "食材に対する好き嫌いを考慮した料理レシピ推薦手法の提案," 第 3 回データ工学と情報マネジメントに関するフォーラム (DEIM), 2011.
- [11] クックパッド, "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/>.
- [12] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," European Conference on Computer Vision, pp.1-14, 2010.
- [13] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," International Conference on Computer Vision, Barcelona, 2011.
- [14] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," Proc. of European Conference on Computer Vision, pp.404-415, 2006.
- [15] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," Workshop on Statistical Learning in Computer Vision, ECCV, pp.1-22, 2004.
- [16] C.D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing, The MIT Press, 1999.