

## 食事画像における類似画像検索

北村 圭吾<sup>†</sup> デシルヴァ, ガムヘワゲチャミンダ<sup>‡</sup> 山崎 俊彦<sup>†</sup> 相澤 清晴<sup>††</sup>

東京大学大学院 <sup>†</sup>情報理工学研究科電子情報学専攻 <sup>‡</sup>情報学環

### 1. はじめに

パーソナルな生活情報を取得・処理するライフログ技術が広く認知されつつある。その中で実際に注目を浴びているのは、従来から研究開発されてきた汎用指向のライフログシステムではなく、運動やつぶやきなど特定応用を指向したシステムである。汎用指向のシステムが理想ではあるが、ライフログの応用性・受容性を追求する目的での特定指向のシステムの意義も大きい[1]。

特に近年食事に特定し、その画像の解析を扱う研究が増えてきている[2-5]。食事の映像からファーストフードのメニューを推定するグループ[2]やオムライス、そばといった 50 種類のメニューを分類するグループ[3]など多様な研究が世界中で見られる。その中で我々は食事に特定したライフログシステム FoodLog を構築し、Web アプリケーションとして公開している[1][6][7]。本論文ではこのシステムの公開後の画像解析の精度低下の原因を解析し、こういった推定の誤りやユーザの入力の負担を低減することを目指した食事画像に特化した類似画像検索について述べる。

### 2. FoodLog システム概要

システム全体の流れを図 1 に示した。システムは Flickr[8] とのマッシュアップという形態を用いており、ユーザは携帯電話やデジタルカメラで食事画像を撮影し Flickr にアップロードすれば良い。画像を受け取ったシステムは 3 つの処理を行う。まずたくさんの日常画像から食事画像を見つけ出す。次に、食事画像に写る食事バランスを食事バランスガイド[9]に則って推定する。最後に集められた食事画像とそのバランスを可視化する。ユーザは、食事画像を撮影することと、時折画像解析の間違いを訂正することでマルチメディア食事記録を作ることができる。このシステムを 2009 年 3 月より公開し、2010 年 1 月の段階でユーザ数 288 人食事画像 5695 枚が集まっている。

### 3. 画像解析

我々のシステムでは 2 つの画像解析、食事画像検出機能と食事バランス推定機能がある。

#### 3.1. 手法

食事画像検出では、大域特徴量として Hough 変換による円の検出情報と色モーメント情報を用い、局所特徴量として Bag of Keypoints (BoK) モデル[10]による表現を用いている。そしてそれら 2 種類の特徴量を Support Vector Machine (SVM) で学習することで実現している。

#### Contents-based image retrieval on food images

<sup>†</sup>Keigo Kitamura <sup>‡</sup>Gamhewage Chaminda De Silva <sup>††</sup>Toshihiko Yamasaki <sup>††</sup>Kiyoharu Aizawa

<sup>†</sup>Dept. of Information and Communication Eng., Graduate School of Information Science and Technology,

<sup>††</sup>Interdisciplinary Information Studies, The University of Tokyo

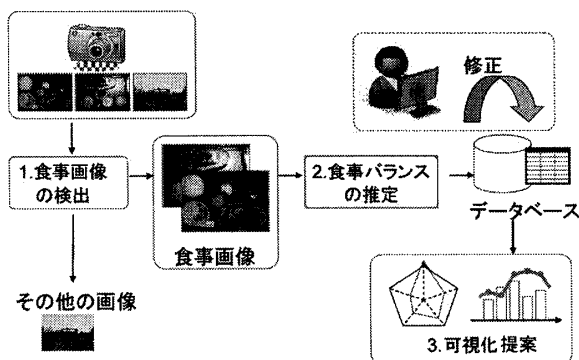


図 1. FoodLog 概要図

表 1. 画像解析結果の比較 (画像解析モデルを作成する際に利用した画像の枚数)

|       | 食事画像検出       | 食事バランス推定    |
|-------|--------------|-------------|
| 事前実験  | 93% (500 枚)  | 50% (100 枚) |
| 一般公開後 | 85% (500 枚)  | 37% (100 枚) |
| 拡張後   | 92% (8000 枚) | 38% (800 枚) |

バランス推定は 2 ステップからなる。300 個の矩形ブロックに分割、300 ブロックの分類情報を用いてバランス値を推定、という手順である。矩形ブロックの分類は色モーメントと DCT 係数を SVM によって学習させたモデルによって行う。

#### 3.2. 精度

以上の手法を用いて解析した結果を表 1 に示した。最上段は事前実験として、筆者自身の画像と無作為に web から集めた画像によって行った実験結果である。中段は事前実験の際に作成したモデルによって実際の一般画像 1000 枚に対して食事画像検出を、食事画像 300 枚に対して食事バランス推定を行ったものである。事前実験より精度が落ちていることがわかる。下段では実際に集まった画像によりデータベースを拡張した場合の結果を示した。食事画像検出に関しては、データベースを大きくすることで多様な画像を学習でき、精度が維持することが可能であった。一方で食事バランス推定では精度の改善は見られなかった。実際関連研究でもテスト画像の多様性を認めたあとに大きく精度を落とす例が見られた。

### 4. 類似画像検索

食事バランス推定に加えて類似画像検索をもとにした栄養情報の付加を検討している。

#### 4.1. 推定補助の機能

画像解析機能は完全に正解するものではなく、単純なアルゴリズムでは大きくバランス推定の精度を上げることも難しい[7]。自動解析にユーザの手入力をいかにうまく利用できるかが重要である。例えば、食事画像に含まれる料理に似た画像を次々に表示させ、ユーザがその中から実際に含まれているものに近い画像を選択する機能

があれば、より精度の高い栄養解析が可能になると期待できる。また、ユーザの入力・修正の手間を省略することや、カロリーの推定など追加的な栄養情報を付与することも可能となる。

(1)

#### 4.2. 類似画像検索手法

類似画像は BoK モデルの表現方法によって表されたヒストグラム間の距離によって定義する。つまり画像をベクトル量子化された局所特徴量 Visual Words(VW)の出現頻度によって表現し、その出現頻度が類似しているものを類似画像とする。ヒストグラム間の距離を測る手法として 3 種類を利用した。

- ・ ユークリッド距離
- ・ Histogram Intersection
- ・ 順位相関係数

ユークリッド距離は各画像を BoK による多次元ベクトルと見なしそのユークリッド距離を計算するものである。Histogram Intersection[11]は 2 つのヒストグラムの各要素の値のうち、小さい方の値を求めて、これを累積した値を類似度とする方法である。この方法では、2 つのヒストグラムのグラフを重ね合わせたとき、そのグラフが重なる部分の面積に相当する。順位相関係数は各ビンを出現頻度順に並び替え、その順位の分布を測るものである。BoK では 0 となるビンが多いため上位 50 位までの順位の順位相関係数を計算する。

またヒストグラムの重みとして、分布度を表す tf-idf によって求まる値によって、各 VW の値を重み付けする実験も行った。tf-idf による重み  $w_j^i$  は以下の(1)式によって求められる。

$$w_j^i = tf_j^i \log \frac{N}{df_j}$$

$tf_j^i$  は  $i$  番目の画像における  $j$  番目の VW の出現頻度、 $df_j$  は  $j$  番目の VW を含む画像の枚数、 $N$  は全画像数を表す。その画像に多く含まれ、かつ他の画像では少数しか含まれないようなその画像特有の VW の重みが大きくなるように計算される。

#### 4.2. 類似画像検索の実験結果

実験は 600 枚の画像によって VW と idf の値を算出し、100 枚の出力用画像のうちから 5 枚を類似画像として出力することによって行った。その精度は 10 枚のテスト画像の正解類似画像を各 5 枚合計 50 枚予め選んでおき、そのうちの出力画像 50 枚の中に含まれていた枚数による。実験画像は図 2 に示されたような画像である。

結果は図 3 のようになった。最も良い結果は tf-idf 処理を行い順位相関係数を計算したもので 50 枚中 11 枚が表れたものだった。

### 5. まとめ

本論文では一般公開を行っている食事ログシステムの画像解析の精度が事前実験より低下していることを報告した。また精度の改善を目指した類似食事画像検索機能を実装した。一般の食事画像に対して行った実験の結果、類似画像検索の精度は 22%であった。今後精度改善の為に類似検索の前に画像分類を行っておくなどの工夫が考えられる。また、この類似画像検索を実際にユーザに提供してバランス推定の精度変化の調査を行っていきたい。



図 2. 実験用画像 (a)(c)テスト用画像 (b)(d)出力用画像

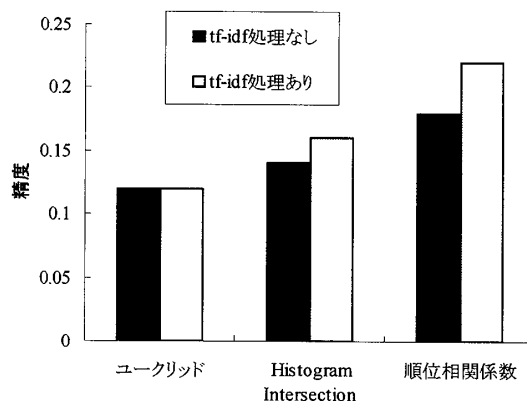


図.3 類似画像検索結果

#### 参考文献

- [1]相澤清晴, 実践的ライフログ: 食事ログからの展開, 情報処理学会誌, Vol.50 No.7, pp. 592-597, July 2009
- [2] Wen Wu and Jie Yang, "Fast food recognition from videos of eating for calorie estimation" In Proc. of IEEE International Conference on Multimedia & Expo, pp 1210-1213, June 2009
- [3] T. Joutou and K. Yanai, "A food image recognition system with multiple kernel learning", In Proc. of IEEE International Conference on Image Processing, pp. 285-288, September 2009
- [4] Corby K. Martin et al, "Quantification of Food Intake Using Food Image Analysis". In Proc of IEEE Engineering in Medicine and Biology Society, pp. 6869-6873, September 2009
- [5] Fengqing Zhua et al "Technology-assisted dietary assessment" In Proc. of SPIE, Vol. 6814, 681411, 2008
- [6]"FoodLog", <http://www.foodlog.jp/>
- [7] K. Kitamura, T. Yamasaki, and K. Aizawa, "FoodLog: Capture, Analysis and Retrieval of Personal Food Images via Web," Proc. of the 16th ACM international conference on Multimedia, workshop on Multimedia for cooking and eating activities, pp. 23-30, 2009.
- [8]"Flickr", <http://www.ickr.com/>
- [9]"食事バランスガイド", <http://www.ma.go.jp/foodguide/balance.htm>
- [10] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," in Proc. of ECCV Workshop on Statistical Learning in Computer Vision, 2004, pp. 59- 74.
- [11] Michael J.Swain, Dana H.Ballard : Color Indexing, International Journal of Computer vision, Vol.7, No.1, pp.11-32, 1991.