

食事画像からのカロリー推定

— 複数の低次特徴に基づく辞書照合と重回帰分析によるアプローチ —

宮崎 達也[†] デシルヴァ, ガムヘワゲチャミンダ^{††} 相澤 清晴^{†††,††}

[†] 東京大学学際情報学府 〒 113-8656 東京都文京区本郷 7-3-1

^{††} 東京大学情報理工学系研究科 〒 113-8656 東京都文京区本郷 7-3-1

^{†††} 東京大学情報学環 〒 113-8656 東京都文京区本郷 7-3-1

E-mail: †{zakimiya,chamds,aizawa}@hal.t.u-tokyo.ac.jp

あらまし 食事画像から抽出した特徴量のみを用いて画像内の食事のカロリー値を推定する。特徴量は色のヒストグラムなどを使用し、6 種類の方法で推定を行った。さらに、それらの推定方法を用いた重回帰分析により、最終的な予測値を求めた。食事画像は、一般的な食事画像 1500 枚を使用した。以上のような条件のもと実験を行い、正解値 $\pm 20\%$ 以内に推定値が含まれる量が 30%、正解値 $\pm 40\%$ 以内に推定値が含まれる量が 67% という精度で推定を行うことができた。

キーワード 料理メディア, 画像認識, 画像検索, 特徴抽出

1. まえがき

生活の重要な要素である“食事”に関する関心が高い。そのことはネット上のメディアにもよく現れている。例えば、写真アップロードサイトである Flickr [1] には、「food」というタグの付いた写真が 2011 年 3 月現在で 800 万枚もアップロードされている。1 年前の 2010 年 3 月には約 350 万枚がアップロードされていたので、この一年間で「food」タグのついた画像が倍以上に増えているということになる。

また、Web 上にも食に関連したサービスが多数存在する。例えばレストランの口コミサイトである食べログ [2] にレストランなど飲食店の店舗情報が 600 万件以上集まったり、レシピ投稿サイトであるクックパッド [3] にレシピが 95 万件以上登録されたりと、Web 上でも食に関連するサービスが盛り上がりを見せている。

さらに、画像処理を使用した食に関連した研究も徐々に行われ始めている。これらについては関連研究の項目で詳しく述べるが、ほとんどのもので、入力する食事画像に対して撮影の際にマーカーが必要となるなど、何らかの前提が必要となっている。また、既存のカロリー推定を行うサービス [4] では、食事情報の入力をユーザーが細かく行っていたり、カロリーの推定を裏で栄養士が行っていたりと手間や時間が多くかかるものとなっている。そこで、本論文では Kitamura らによる食事バランス推定技術 [5] [6] をシステムに組み込んだ Web サービスである FoodLog [7] を使用している一般ユーザーが撮影、アップロードした一般的な食事画像に対するカロリー推定を試みた。

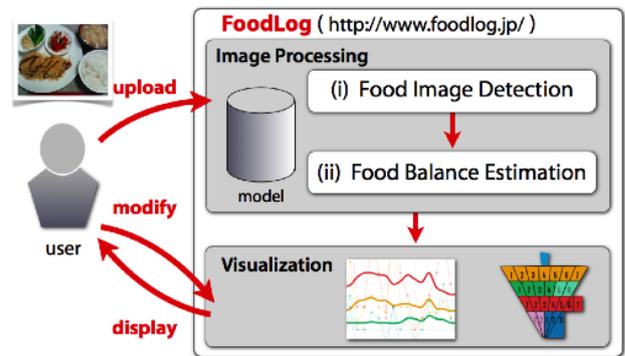


図 1 フードログシステムの概要

2. 食事画像の画像処理に関する関連研究

前項でも述べたように、画像処理を使用した食に関する研究も多数行われている。Kitamura ら [5] は、局所特徴量、大域特徴量両方を使用した食事バランスの推定を行った。食事バランスとは、食事を主食、主菜、副菜、果物、乳製品の五つに分類し、それらがどの程度食事内に存在しているか、というものである。さらに、丸山らにより個人の食傾向を反映する手法 [8] が提案され、精度が大きく改善している。

Joutou ら [9] は、50 種類のメニューの中からのメニュー推定を行った。特徴量として、Bag of Features (BoF) 表現、色ヒストグラム、ガボール特徴を用い、学習器として Multiple Kernel Learning (MKL) を使用した推定を行った。その結果最大 61.34% の精度を出した。使用されている食事画像は、対象となる食事がズームされている以外は一般的な食事画像である。

Zhu ら [10] は、食事前、食事後の写真を利用してど



図 2 食事バランスガイド

れくらの量の食べ物 (food item) が摂取されたのかを 93.745%の精度で推定した。食べ物は 13 種類しかなく、食事の撮影の際には白黒のチェッカーボード上に白い皿を乗せ、その上に食事を乗せなければいけないという制約がある。

Wuら [11] はファストフードに限って摂取したカロリーを推定した。Scale-invariant Feature Transform (SIFT) 特徴量を用いたマッチングと、Web から収集したファストフードのカロリーデータとを使用している。また、食事食べている様子を撮影した 10 分程度の動画を使用しており、正解率は 40~73%であった。

Yangら [12] は、画像中の 1 対のピクセルの局所特徴量を使用し、栄養素の推定を行った。Pittsburgh fast-food image dataset (PFID) [13] という食事データベースに登録されている食事画像を 61 個のグループに分け、そのどれに分類されるかで実験を行い、28%の精度で推定を行った。

これらの研究は食事バランスを扱う [5], [6] 以外のもの全てで、画像内のメニューやフードアイテムを推定している。カロリーをはじめその他の栄養素を推定したい場合は認識したメニューを元に算出することとなる。そのため、メニュー認識の精度に全てが依存する。それに対して本論文では異なるアプローチを提案する。本提案では、入力食事画像からメニューというセマンティクスを介すことなく直接カロリーの推定を行うという手法をとっている。

3. FoodLog について

3.1 FoodLog の概要

FoodLog [7] は日々の食事を手軽に記録する Web アプ

リケーションである [6]。一般に公開されており、毎日の食事を写真に撮ってアップロードすることで食事記録の可視化を行うことができる。FoodLog のシステムの概要を図 1 に示す。ユーザーが画像をアップロードするだけで食事バランスの推定結果が出るため、ユーザーが負担を感じることなく継続しやすいシステムとなっている。また、推定結果が良くない場合には結果の修正を手動で行うこともできる。

内部で行っている処理は大きく分けて二つあり、まずアップロードされた写真が食事であるかどうかの判定を行う。そして次に食事のバランス推定を行う。食事バランスというのは農林水産省などが提唱している食事バランスガイド [14](図 2) というものを使用し、それに基づき Serving(SV) という単位を用いる。SV はおよそ一皿を表しているため、専門的知識を持っていなくても簡単に予測することがきでる。また、そのおおよその指標は図 2 内に示されている。食品は図 2 のように主食、主菜、副菜、果物、乳製品に大きく分類され、それぞれについて一日何 SV をとるのが理想的かも示されているため、健康のための指標とすることもできる。

FoodLog のインターフェースは図 3 のようになっている。図 3a は一ヶ月分の食事を載せた「食事カレンダー」である。また、図 3b はそのうちの一つの食事について詳しく表示したもので、各食事要素の SV 値の推定結果を示している。また、この画面から推定値の修正も行うことができる。図 3c は食事の時間分布を表したもので、自分の食事の一週間分の摂取時間を見ることができる。図 3d は食事バランスの推移グラフで、自分の健康管理のために適正な食事バランスで食事摂取することを維持できているかなどが確認できる。

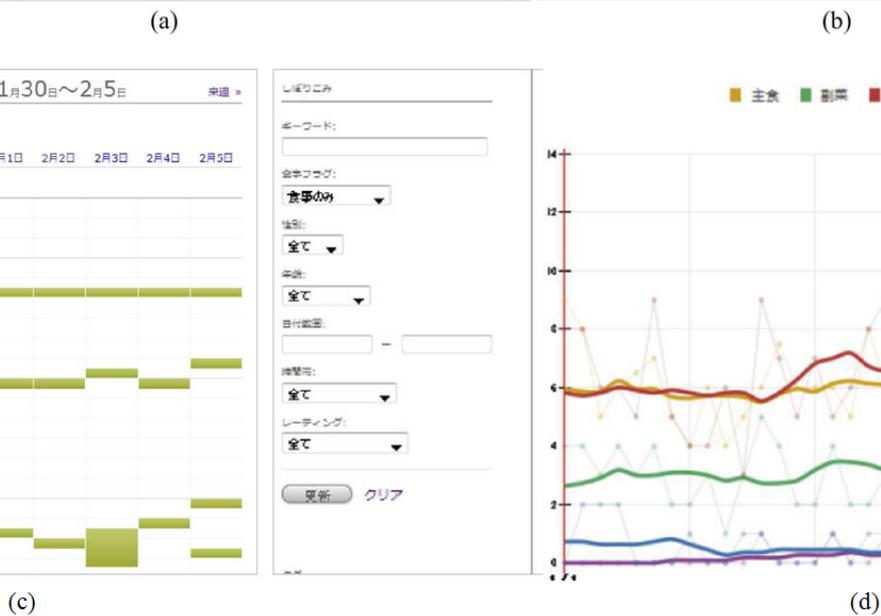
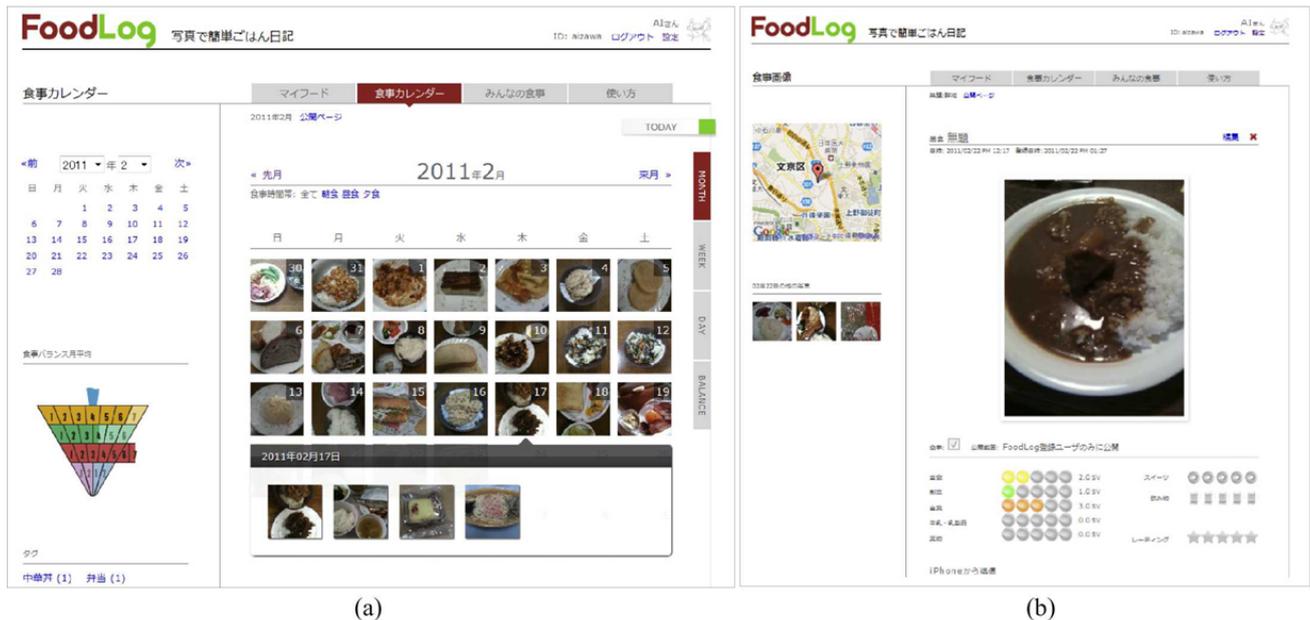


図 3 FoodLog のインターフェース (a) 月間表示 (b) 食事バランス表示 (c) 週間の食事時間表示 (d) 食事バランスの推移

3.2 本論文で用いるデータセット

本論文で用いたデータセットは、FoodLog 上にアップロードされた食事画像である。そのため、すべての画像は食事画像であるという以外に特殊な前提条件がない画像である。図 4 は、データセットに含まれる画像のサンプルである。

使用した画像の枚数は合計で 1500 枚で、そのうち 1400 枚は辞書データセット、残りの 100 枚をテストデータセットとして使用した。辞書データセットは入力したテストデータがどの辞書データに似ているかを測るために使用した。ここで用いた画像には元々正解値がついていないため、栄養学の専門知識を持った複数の専門家によりカロリーの正解値を設定した。

図 5 は辞書データセットのカロリー分布を表すヒスト

グラム、図 6 はテストデータセットのカロリー分布を表すヒストグラムである。辞書データセットとテストデータセットとの間にカロリー分布の大きな差異がないことがわかる。

4. 辞書データセットを用いた低次特徴からのカロリー推定

本論文では低次の特徴量と辞書データを利用したカロリー推定を行う。その手続は以下ようになる。

- ・ 画像からの特徴量の抽出
 - ・ 特徴量に基づく辞書データの画像類似評価
 - ・ 類似性の高い上位 n 位によるカロリー値の推定
- なお、上位 n 位のカロリーデータに付いては平均を求めることで推定を行う。



図 4 データセットに使用した画像の例

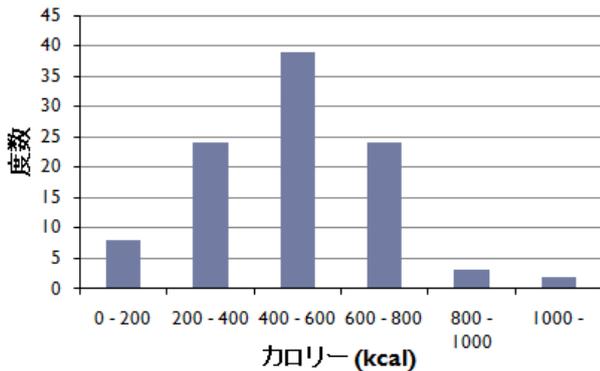


図 5 辞書データのカロリー分布

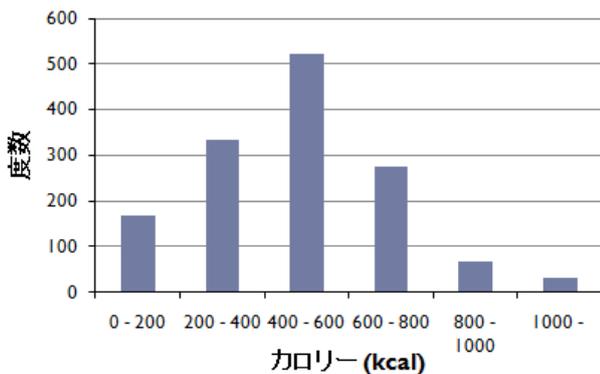


図 6 テストデータのカロリー分布

4.1 使用した特徴量

本論文で使用した特徴量を以下に記す．また，使用した特徴量と，使用した上位 n 件の n の値の一覧を表 1 に示す．特徴量の選定については，食事画像に置いて色が重要であることから，RGB，HSV のヒストグラム，さらにカラーコレログラムを用いた．また，輝度変化の評価に関わる局所特徴量も合わせて用いた．

RGB ヒストグラム

RGB カラーモデルは，色を三原色である赤 (Red)，緑 (Green)，青 (Blue) の加法混色法で表すモデルである．

本論文では，各色を 4 色ずつに減色し，合計 $4^3 = 64$ 色でヒストグラムを作成した．

HSV ヒストグラム

HSV カラーモデルは，色を色相 (Hue)，彩度 (Saturation)，明度 (Value) の 3 つの指標で表すモデルである．人間が色を感知するのに近い表現方法であるとされ，色相は色の波長を，彩度は色の鮮やかさを表している．本論文で用いた HSV ヒストグラムは HS をそれぞれ 4 段階に，V を 6 段階に減らし，合計 $4 \times 4 \times 6 = 96$ 色でヒストグラムを作成した．

カラーコレログラム

カラーコレログラム [15] は，隣り合うある色の画素同士がどの程度相関関係を持っているかを表した特徴量である． $n \times n$ の画像 $I(x, y)$ に対し，その画素 $p = I(x, y)$ の画素を色を c とし， c を c_1, c_2, \dots, c_m と m 色に量子化する．その時，色が c_i である画素から k 離れた位置にある画素の色が c_j である確率を求める．

$$\gamma_{C_i, C_j}^{(k)} \equiv Pr_{p_1 \in I_{c_1}, p_2 \in I} [p_2 \in I_{c_j} \mid |p_1 - p_2| = k]$$

SURF (Speeded Up Robust Features) 特徴量

SURF 特徴量 [16] はまず周りの点と比べて輝度が最大となる点を抽出し，その点の輝度勾配の大きさ $m(x, y)$ と方向 $\theta(x, y)$ からなる特徴量である．方向 $\theta(x, y)$ は，点の周囲を 4×4 の 16 ブロックに分割し，各ブロック内で 8 方向の勾配を求めるため， $4 \times 4 \times 8 = 128$ 次元からなる特徴量となる．SURF 特徴量は，近似処理を行っているため，似たような特徴量である SIFT 特徴量に比べて名前のとおり計算速度が速い．本論文では SURF 特徴量を後述の BoK 表現と合わせて使用した．

BoK (Bag of Keypoints) 表現

BoK [17] は，テキストをいくつかのクラスタに分類する手法である Bag of Words 表現を画像認識に応用させた手法で，画像から抽出した特徴量を元にクラスタリングを行う．本論文では事前に求めた SURF 特徴量に対して k-means 法を適用することにより分類を行った．k-means 法は合計 N 個の点 x_1, x_2, \dots, x_N を k 個のクラスタ S_i にクラスタリングする手法であり，下記のように表される．

$$\operatorname{argmin} \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2$$

はじめに各クラスタの重心 μ_i を決定し， x_i を一番近い μ_i のクラスタに所属しているものとして重心を再計算する．そして再び x_i がどのクラスタに所属するかを計算する．この操作を x_i の所属が計算前も計算後も変わらないまで繰り返す．その結果， N 個の点が k 個のクラスタに分類される．本論文では $k = 50$ でクラスタリングを行った．

表 1 使用した特徴量

No.	特徴量	n
1	RGB ヒストグラム	5
2	RGB ヒストグラム	10
3	HSV ヒストグラム	5
4	HSV ヒストグラム	10
5	カラーコレログラム	5
6	カラーコレログラム	10
7	SURF + BoK 表現	5
8	SURF + BoK 表現	10

4.2 単一特徴からの辞書照合による推定結果

使用した特徴量が RGB ヒストグラム, $n=10$ (表 1 の No2) の場合の結果が図 7a である。図 7a の結果によると正解値 $\pm 20\%$ の誤差範囲内に 42%, $\pm 40\%$ の誤差範囲内に 70% が推定値として入っている。一見すると良い結果に見える。しかし、グラフを見ると弱い相関関係は見受けられるものの、ほとんどの推定値が 400 ~ 600kcal で推定されて十分な推定になっていない。このことから、推定器をひとつしか用いなかった場合には十分に推定が行えないということが言える。図 7b は、HSV ヒストグラム, $n=10$ (表 1 の No4) の場合、図 7c は、カラーコレログラム, $n=10$ (表 1 の No6) の場合、図 7d は、SURF+BoK 表現, $n=10$ (表 1 の No8) の場合の結果である。

5. 複数の特徴に基づく辞書照合と重回帰分析によるカロリー推定

5.1 重回帰分析を用いたカロリー推定の枠組み

単一の特徴だけでは不十分な推定にしかならない。それを改善する方法として、複数の照合を行い、推定値を算出し、それらを統合する手法を考える。前項で求めた入力画像に対するカロリー推定を以下のようにして行った。図 8 はその工程を図で表したものである。図 8 内の X (特徴量) が前節で求めた推定値である。複数の特徴量それぞれの推定値を用いて重回帰分析を行い、最終的な推定値 y を求めた。 a_i は重み, X_i は単一特徴からの推定値である。

$$y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$$

5.2 結果

図 9 は最終的な推定結果を示したグラフである。正解値 $\pm 20\%$ の誤差範囲内に 30%, $\pm 40\%$ の誤差範囲内に 67% が含まれている。また、図 7d の相関係数 0.31 に比べて図 9 の相関係数は 0.33 と数字の上ではあまり改善が見られないが、正解カロリー値 ± 40

図 9 のグラフの中には誤差が非常に大きい物がいくつかある。そのなかでも誤差の大きい事例を図 10 に示す。図 10a はサラダ、図 10b はうな井、味噌汁など 4 皿である。図 10a は正解値が 58kcal であるのに対し、525kcal

と、図 10b は正解値が 1027kcal であるのに対し、319kcal と推定された。

原因として挙げられるのが、見た目だけを見ると似ている食事が存在するということである。例えば図 10a には緑色の菜っ葉と赤いトマト、つけあわせが写っており、色の取り合わせとしては肉と付け合せに似ている。そのため、野菜だけよりも多いカロリーが推定されたと考えられる。これはテクスチャを表す特徴量など、色以外の特徴量も多く利用することで解決できる可能性がある。ただし、これらの事例は多くあるものではない。

また、図 10b には複数の皿が映っているためうまく推定ができなかったのではないかと考えられる。一皿ごとの場合には似たような食事も多く存在するが、それが何皿も組み合わせさせた食事の場合、すべてが似た食事は殆ど存在しないと言ってもいい。このように、完全に正確な予測は非常に難しいということがわかる。

5.3 まとめ

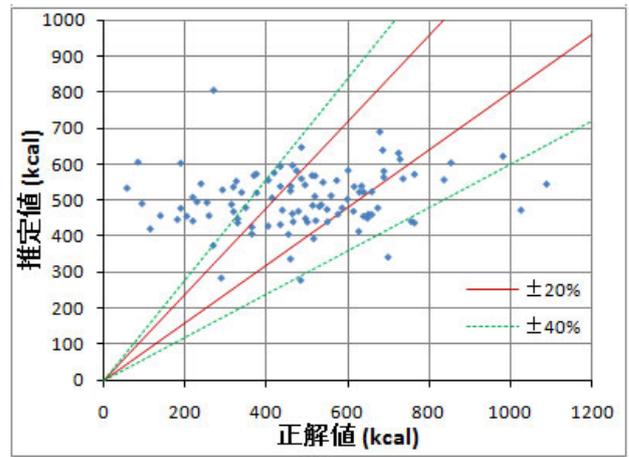
本論文では、食事画像からのカロリー推定について論じた。食事記録を行うアプリケーションである FoodLog の紹介を行い、FoodLog にアップロードされている一般的な食事画像から構築したデータセットを利用した。

画像から複数の特徴量を抽出し、個別に推定値を出した後、それらの重回帰分析を用い、食事画像からのカロリー推定を行った。既存の手法では多くの場合に食事画像や映像からメニュー、フードアイテムの推定を経てカロリーや栄養素を求めていたが、本論文ではそれらの予測は行わず、食事画像から直接カロリーを求めた。その結果正解値 $\pm 20\%$ の誤差範囲内に 30%, 正解値 $\pm 40\%$ の誤差範囲内に 67% の精度でカロリー推定を行うことができた。

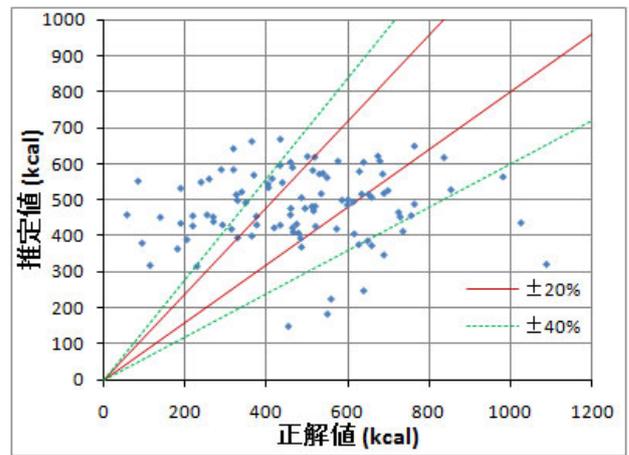
文 献

- [1] “ Flickr ” <http://www.flickr.com/> アクセス 2011 年 5 月 31 日 18:00
- [2] “ 食べログ ” <http://tabelog.com/> アクセス 2011 年 5 月 31 日 18:00
- [3] “ クックパッド ” <http://cookpad.com/> アクセス 2011 年 5 月 31 日 18:00
- [4] “ げんき! 食卓 コンシェルジュ ” <http://shoku365.com/> アクセス 2011 年 5 月 31 日 18:00
- [5] K. Kitamura, T. Yamasaki, and K. Aizawa “ Food log by analyzing food images. ”, Proceeding of the 16th ACM international.
- [6] 相澤清晴 “ ライフログの実践的活用: 食事ログからの展望 ” 情報処理, Vol.50, No.7 pp. 392-397, 2009
- [7] “ FoodLog ” <http://www.foodlog.jp/> アクセス 2011 年 3 月 25 日 18:00
- [8] 丸山悠斗, デシルヴァ ガムヘワゲ チャミンダ, 山崎俊彦, 相澤清晴 “ 食事ログシステムの解析の個人適応による食事バランス推定 ” HCG シンポジウム pp. 369-374, 2010
- [9] T. Joutou and K. Yanai. “ A food image recognition system with multiple kernel learning ”, ICIP, pp. 285.288, 2009.
- [10] F. Zhu, A. Mariappan, C. J. Boushey, D. Kerr, K.

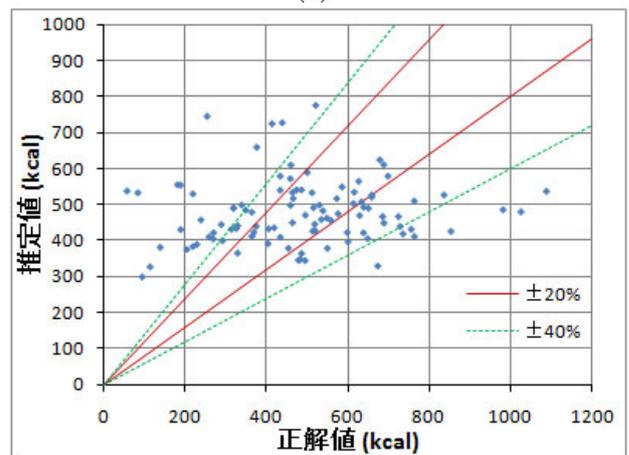
- D. Lutes, D. S. Ebert, and E. J. Delp " Technology-assisted dietary assessment ", Proc. SPIE, Vol.6814, No.681411. 2008.
- [11] W. Wu and J. Yang. " Fast food recognition from videos of eating for calorie estimation ", ICME, pp. 1210.1213, 2009.
- [12] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar " Food Recognition Using Statistics of Pairwise Local Features ", CVPR, No.27, 2010.
- [13] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, " PFID: Pittsburgh fast-food image dataset ", Proc. ICIP, 2009.
- [14] " 食事バランスガイド "
 http://www.maff.go.jp/j/balance_guide/ アクセス 2011年3月29日 18:00
- [15] J Huang, S R Kumar, M Mitra, W-J Zhu, R Zabih " Image Indexing Using Color Correlogram, "Proc. IEEE Conf. CVPR, pp. 762-768, 1997.
- [16] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, " SURF: Speeded Up Robust Features ", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346-359, 2008.
- [17] G. Csurka, C. Bray, C. Dance, and L. Fan, " Visual categorization with bags of keypoints ", European Conference on Computer Vision Workshop on Statistical Learning in Computer Vision, pp. 1-22, 2004



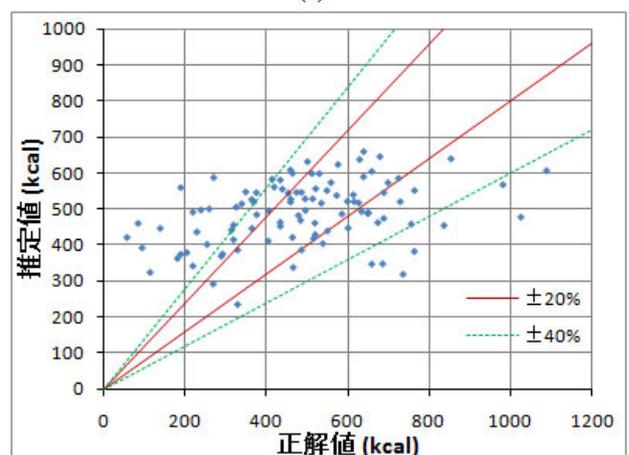
(a)



(b)



(c)



(d)

図7 単一特徴からの推定結果 (a)RGB ヒストグラム (b)HSV ヒストグラム (c) カラーコレログラム (d)SURF+BoK

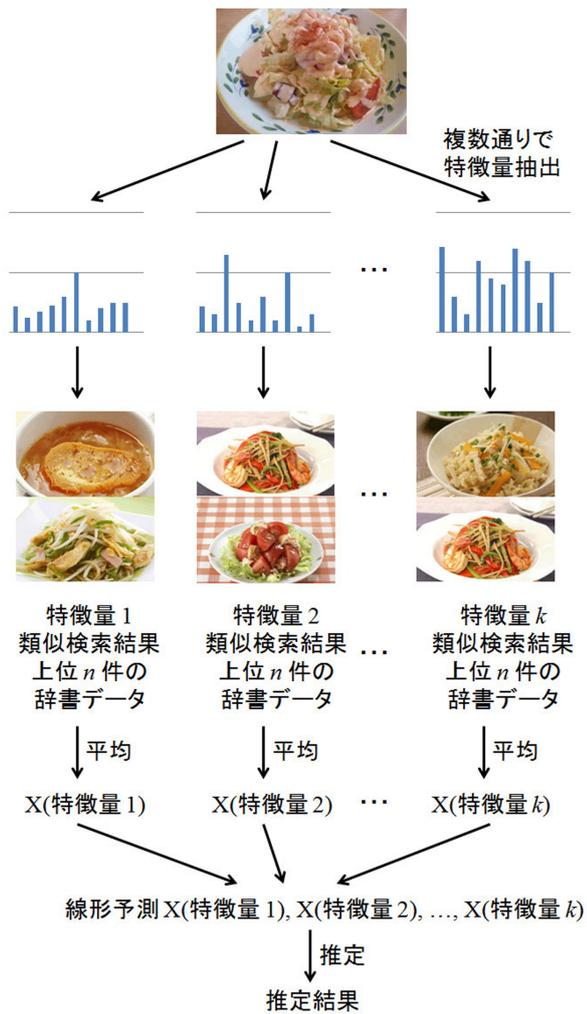


図 8 実験手順の概要



(a)



(b)

図 10 推定誤差が大きい画像の事例

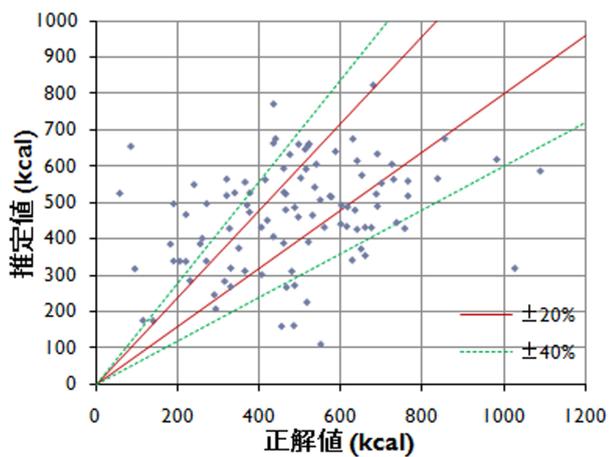


図 9 複数特徴からの推定結果