

Multiple Kernel Learningを用いた 食べ物画像の認識

電気通信大学 情報工学専攻

上東 太一, 甫足 創, 柳井 啓司

2009年7月21日MIRU

研究の背景

近年、健康管理への関心が高まってきている

特に、『食事』に関する健康管理が注目されている



簡単に食事内容が記録できるシステムが望まれている

研究の目的

- 実用を目指した多種類(50種類)に対応した食事画像の認識

- 食べ物画像の認識

- Multiple Kernel Learningによる多種特徴の統合

- システムのプロトタイプの実成

- 携帯カメラと添付メール

50種類もの大規模な食事分類は従来行われていない

関連研究

- マルチクラス分類の一般物体認識
 - Caltech101/256 ベンチマークが有名
 - 様々なカテゴリーを含む. 動物, 人工物, 顔, マーク...
 - 多種類特徴の統合による手法が最高性能
 - Multiple Kernel Learning による統合 [Varma et al. 2007]
 - 101種類分類 89.6% 256種類分類 60.3%
 - 特定ジャンルのカテゴリ認識
 - **互いに似ているので, より困難な問題.**
 - Flower 102種類 72.8% [Nilsback et al. 2008]
 - MKLによる統合 Caltech-101 に比べて, 16.8%精度低下
 - 室内シーン分類 67種類 25 % [Quattoni et al. 2009]

画像認識方法

- Multiple Kernel Learning(MKL) [Lanckrietら,2004]
を用いた特徴統合によるマルチクラス分類手法
- 統合する画像特徴
 - 局所特徴のBag-of-Keypoints(6通り)
 - カラーヒストグラム
 - ガボール特徴のヒストグラム(2通り)

局所特徴

- SIFTを使用
- 特徴点検出は3タイプ
- 画像はBag-of-keypoints表現を用いて局所特徴の出現頻度のヒストグラムで表現
 - 1000次元と2000次元で表現



DoG検出



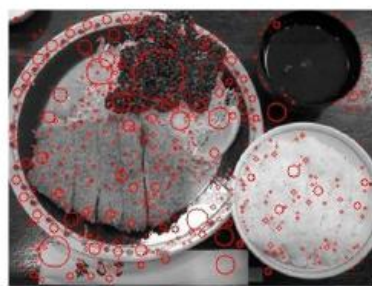
グリッド検出



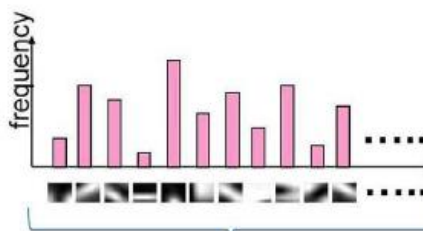
ランダム検出



SIFT



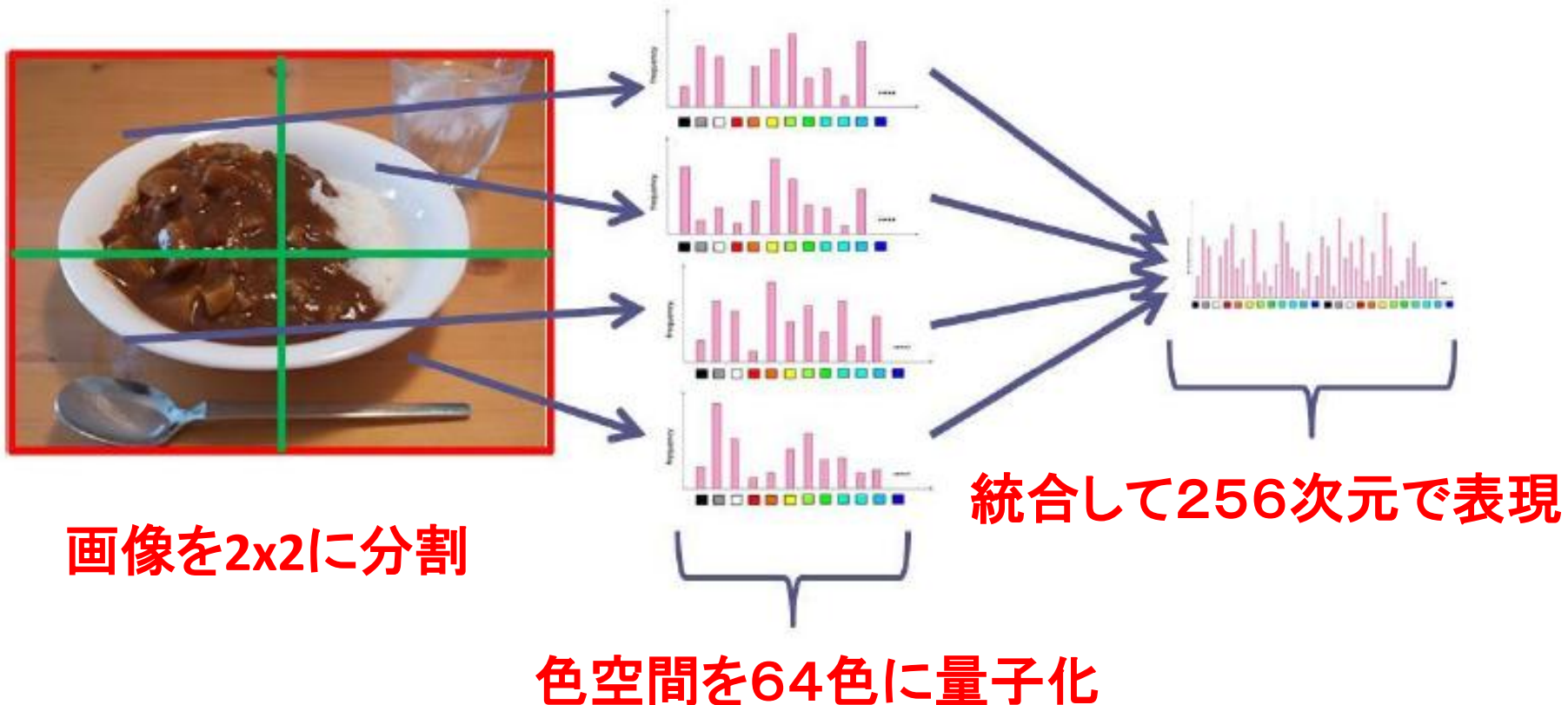
Bag-of-keypoints表現



ヒストグラム

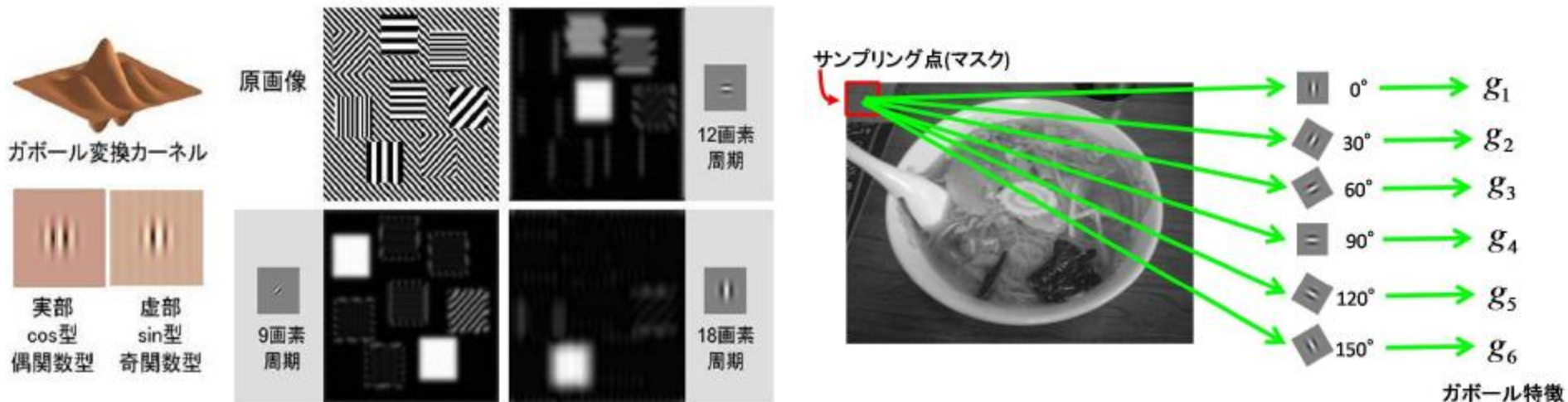
カラーヒストグラム

- RGB色空間を使用



ガボール特徴

- 画像化から局所的な濃淡情報の周期と方向を表現した特徴
- 6方向、4周期のガボールフィルタで特徴抽出
- 色特徴と同様に画像を3x3と4x4に分割
 - 216次元と384次元のベクトルで表現



マルチクラス分類

- 1-vs-rest SVM分類器によって分類

学習

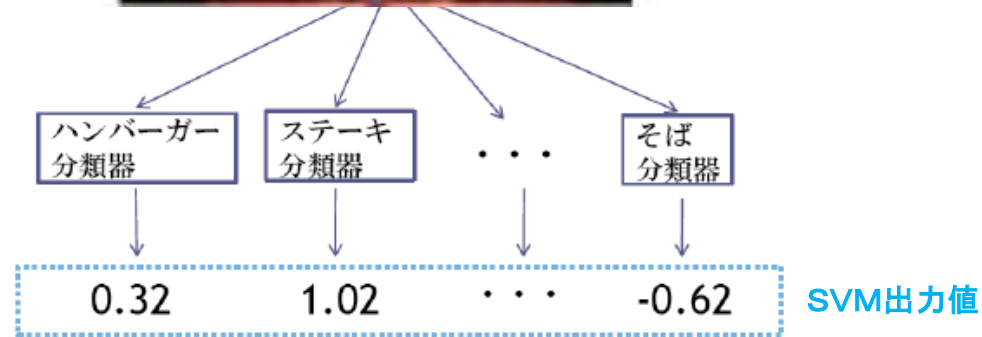
あるクラスを正例とし
その他のクラスを負例として
一つのクラスの分類を学習
それを、全クラスで学習

分類

学習したすべての分類器に
テストデータを入力して、
出力値の最も大きかった
クラスに分類



入力画像



出力値の最も大きいクラスに分類

Multiple Kernel Learningによる特徴統合

- 本研究では複数のSVMのカーネルを線形結合することにより特徴を統合する
- Multiple Kernel Learning 問題[Lanckrietら,2004]

$$K_{combined}(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^K \beta_j k_j(\mathbf{x}, \mathbf{x}') \quad \text{with } \beta_j \geq 0, \quad \sum_{j=1}^K \beta_j = 1.$$

- 最適な重みパラメータ β_j を求める
- 凸面最適化問題として解く
- MKL-SVMを用いる
 - SVMフレームワークで最適化問題を解く
[Sonnenburgら,2006]

Multiple Kernel Learning(MKL)

- 複数のカーネルに対し、最適な重みを学習する手法
- 画像認識では、特徴統合の手法として利用可能。
 - Varmaら(2007)はcaltech101/256で最も良い結果を出している

各サブカーネルをそれぞれの特徴と対応させ、
特徴統合に用いる

色特徴のカーネル テクスチャ特徴のカーネル

$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^K \beta_k \mathbf{k}_k(\mathbf{x}_i, \mathbf{x}_j)$$

最終的なカーネル 結合 特徴別のカーネル

MKLの主問題

- 2クラス分類に対するMKLの主問題

$$\begin{aligned} \min \quad & \frac{1}{2} \left(\sum_{k=1}^K \|\mathbf{w}_k\|^2 \right)^2 + C \sum_{i=1}^N \xi_i \\ \text{w.r.t.} \quad & \mathbf{w}_k \in \mathbb{R}^{D_k}, \xi \in \mathbb{R}^N, b \in \mathbb{R}, \\ \text{s.t.} \quad & \xi_i \geq 0 \text{ and} \\ & y_i \left(\sum_{k=1}^K \langle \mathbf{w}_k, \Phi_k(\mathbf{x}_i) \rangle + b \right) \geq 1 - \xi_i, \\ & \forall i = 1, \dots, N \end{aligned}$$

ここで, $\mathbf{w}_k = \beta_k \mathbf{w}'_k$ ($\beta_k \geq 0, \forall k = 1, \dots, K$), $\sum_{k=1}^K \beta_k = 1$
 $\mathbf{k}_k(\mathbf{x}_i, \mathbf{x}_j) = \langle \Phi_k(\mathbf{x}_i), \Phi_k(\mathbf{x}_j) \rangle$ であるとする.

MKLの双対問題[Bashら,2004]

- MKLの双対問題

$K = 1$ のとき, 標準のSVM
の双対問題と等価になる

$$\begin{array}{ll} \min & \gamma \\ \text{w.r.t.} & \gamma \in \mathbb{R}, \alpha \in \mathbb{R}^N \\ \text{s.t.} & 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^N \alpha_i y_i = 0 \end{array}$$

$$S_k(\alpha) = \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{k}_k(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^N \alpha_i \leq \gamma, \\ \forall k = 1, \dots, K$$

単一カーネルとの違い

拘束条件 $S_k(\alpha) \leq \gamma$

全カーネルで共通の上限値の γ を最小化

MKLのアルゴリズム

単一カーネルのSVM学習の反復による解法[Sonnenburg,2006]

1. 最初に、重み β_j を均等にする
2. β_j を固定し、統合カーネルを単一カーネルとみなし、通常のSVM学習を行い $\alpha_i (i = 1..N)$, b を求める
3. 求めた α_i を固定して、 $\sum_{k=1}^K \beta_k S_k(\alpha)$ が増加するように β_j を微小に変化させる
4. 終了条件に達するまで(2)-(3)を繰り返す

評価実験

- **データセット**: 50種類食べ物画像セット
- **画像特徴**: 3タイプの画像特徴(計9種類)
- **手法**: 各特徴単独のSVM

全特徴によるMKL-SVM

– MKLの実装は機械学習ライブラリ SHOGUNを利用

- **カーネル**: χ^2 カーネル

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{1}{2\sigma^2} \sum_i \frac{\|x_i - y_i\|^2}{x_i + y_i}\right) \gamma$$


– スケールパラメータ γ の設定方法

- ① Cross-validationにより, 最適な値を探索
- ② すべての学習データ間の χ^2 距離の平均の逆数

[Zhang, 2007]

50種類データベースのサンプル



 は人手で囲んだ食べ物の位置

各種類100枚ずつ 計5000枚

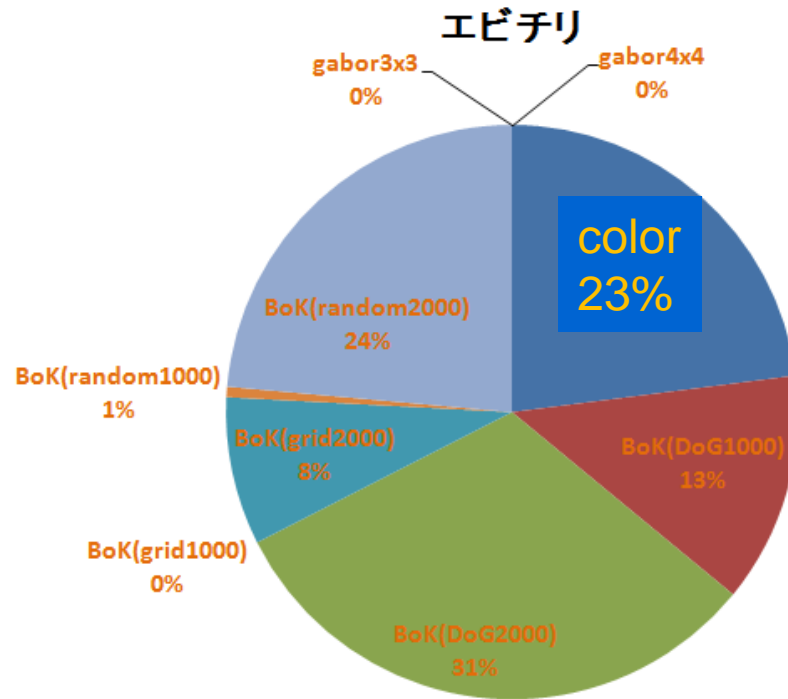
実験結果

MKLによる特徴統合方法
と
特徴単独で用いた手法
の50種類の平均分類率

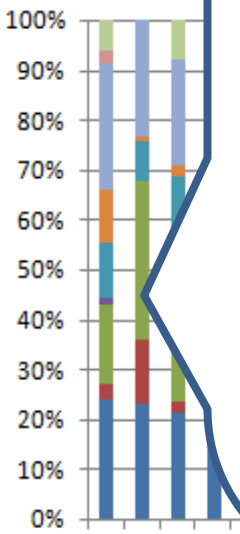
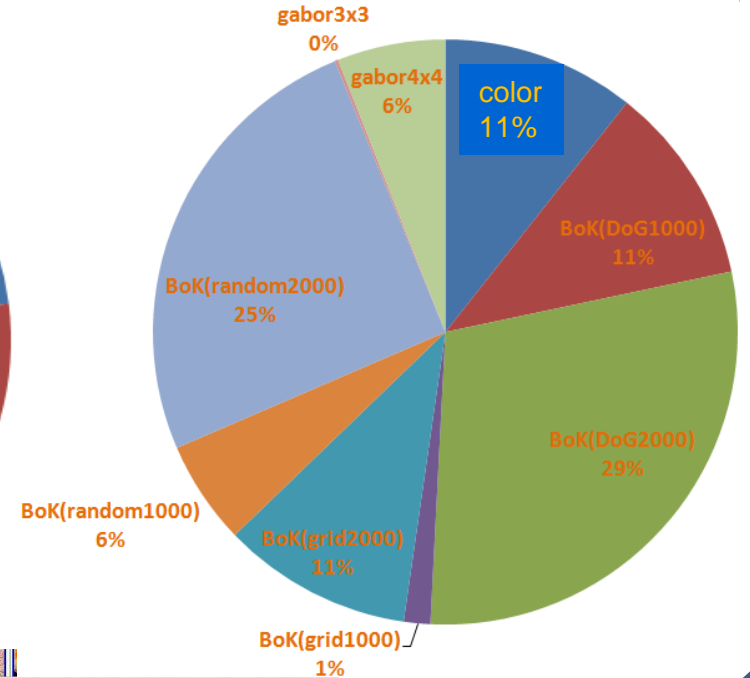
特徴	平均分類率
BoK(dog1000)	26.52%
BoK(dog1000)	27.48%
BoK(grid1000)	26.10%
BoK(grid2000)	27.68%
BoK(random1000)	28.42%
BoK(random2000)	29.70%
gabor3x3	31.28%
gabor4x4	34.64%
Color	38.18%
MKL(cross-validation)	53.16%
MKL(mean-x2 distance)	61.34%

MKLで学習した特徴の重み

エビチリの結果



平均の重み



- オムライス
- エビチリ
- 味噌汁
- ポターシユ
- ビーフカレー
- つな重
- ピンパ
- むそぼ
- ショアオロース
- やし中華
- マーキ
- 鮮丼
- エウマイ
- ハゲティ
- わし寿司
- ルキヤベツ
- い焼き
- バーガー
- 前煮
- サ
- ーメン
- 焼き
- 日ッサン
- ルパン
- しゃが
- エー
- 子丼
- ドウィッチ
- ツカレー
- ヤーハン
- 妻焼き
- 肉ヨロケ
- ースカツ
- 丹
- 子
- フライ
- ーストチキン
- きそば
- き鳥
- ン丼
- ラタン
- 煮
- ースト
- 好み焼き
- ーボー豆腐
- 春巻き
- ピラフ
- たに焼き
- 平均の重み

実験結果：混合行列

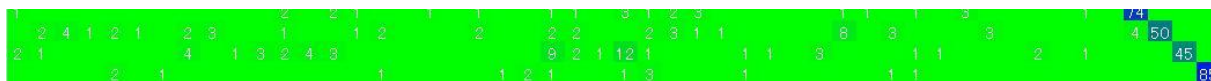
MKL(mean- χ^2 distance)



うな重

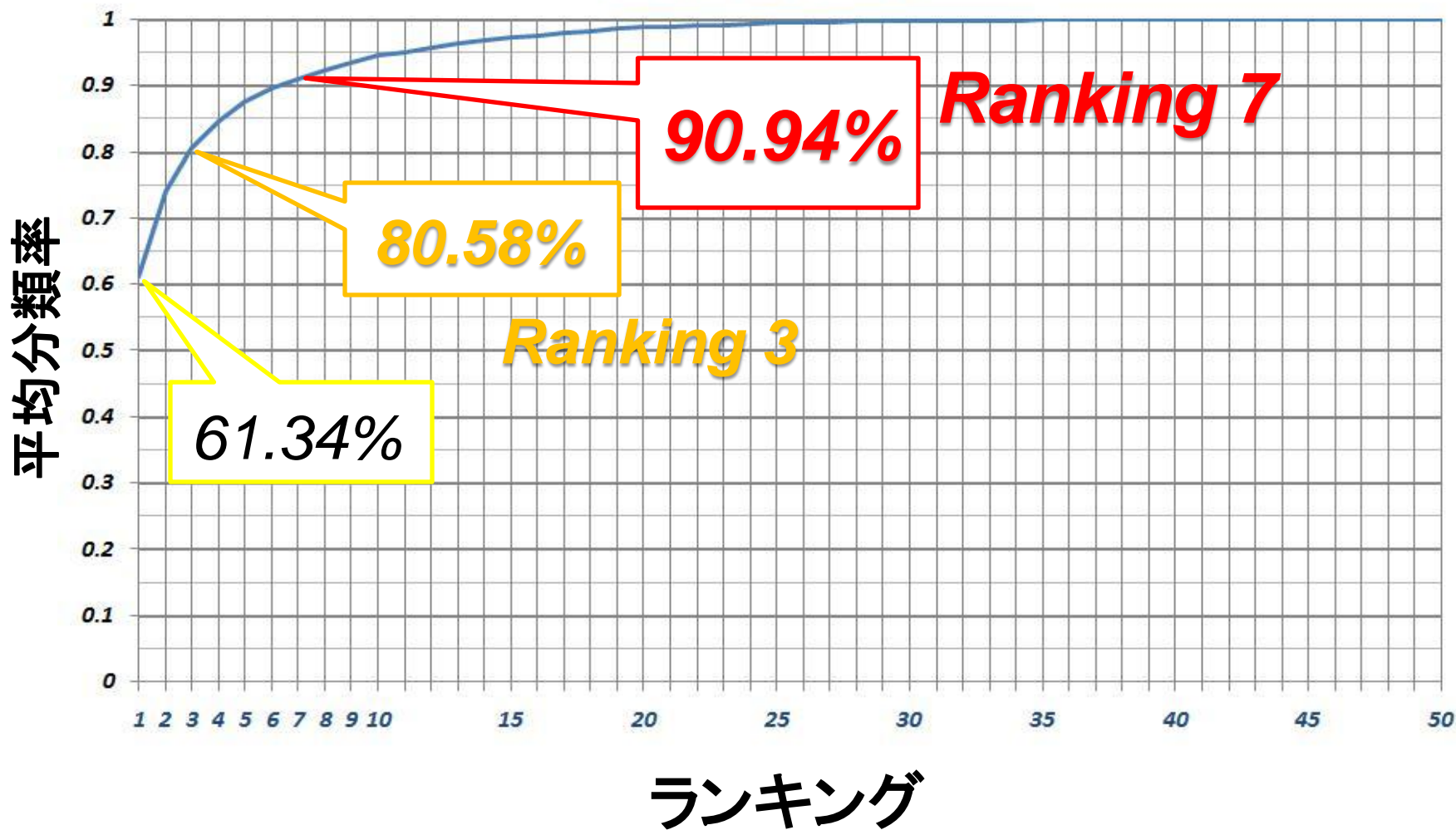
チャーハン

ピラフ



実験結果

許容クラス数を変化させたときの平均分類率の変化

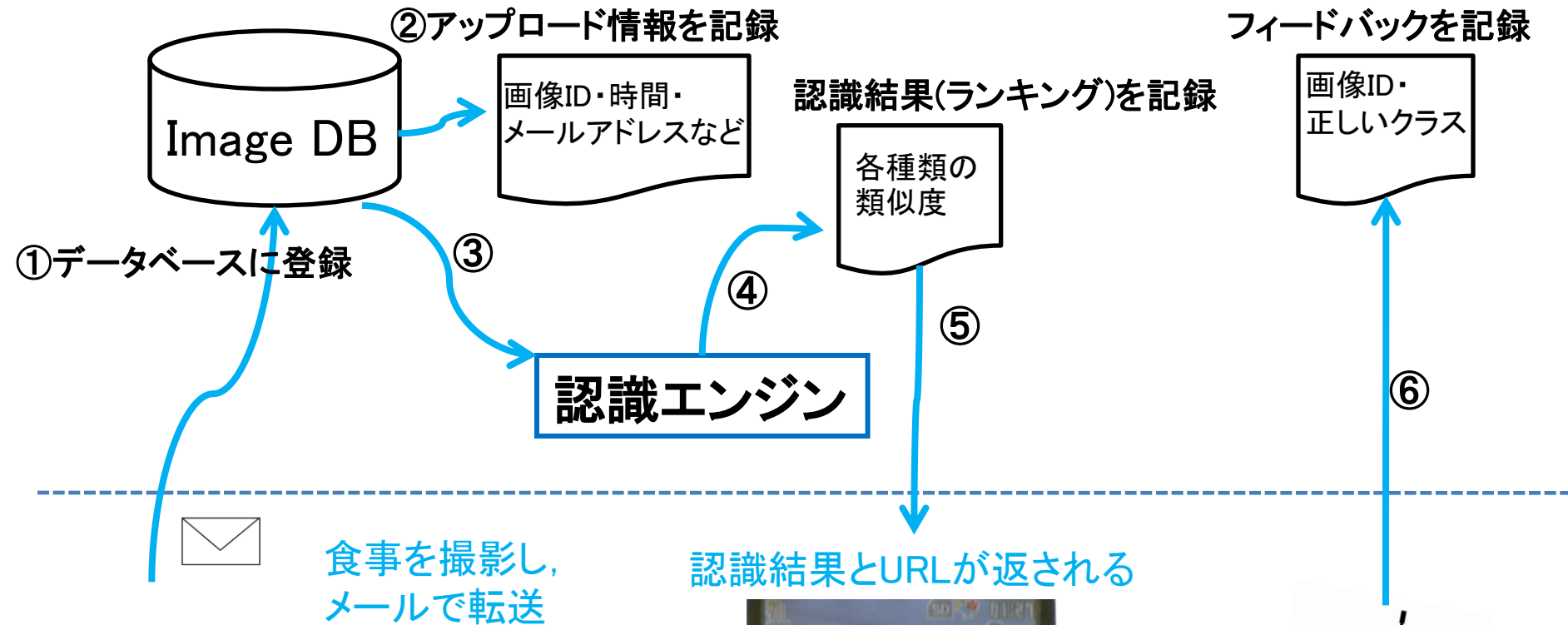


85種類の認識

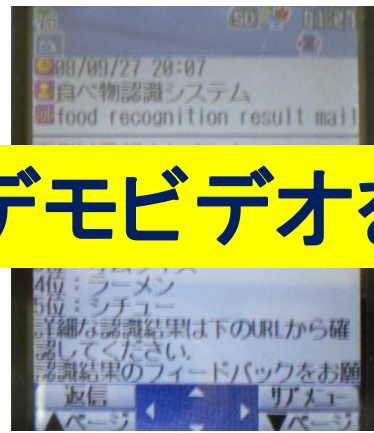
- 現在50種類から85種類に拡張実験中
 - 厚生労働省の「食事バランスガイド」より追加
 - 手法は50種類と同じ

ランキング	85種類DB	50種類DB
1位まで	57%	61%
3位まで	76%	81%
5位まで	83%	88%

システムのプロトタイプを作成： 携帯版認識システムの構成



実際の利用例は、デモビデオをご覧ください。



正しい食べ物の種類
を選択

システムの評価

- 約13ヶ月間試験的に運用
 - 300枚がアップロードされた
- システムの認識精度：
 - 1st 38.3%, 3rd 58.3%, 5th 64.0%

うまく認識された写真の例



うまく認識されなかった写真の例



まとめ

- 食べ物画像から料理を認識する手法を提案
 - 50種類で平均分類率61.34%を達成
 - 上位3位までの分類を許容すると80%を超える
- 13カ月間運用した携帯システムの認識精度
 - 1位まで38.3%, 3位まで58.3%

認識結果



オムライス



ざるそば



チャーハン

今後の課題

- カロリー計算のための食事の「量」の認識
 - おおよそのカロリーを自動計算可能にする
 - 箸など基準となる物体とのサイズの比較
- 認識精度向上
 - 現在使用している特徴とは異なる識別能力を持つ
新たな特徴量の追加
 - 皿の検出などの食事画像専用の処理の追加
- 食事カテゴリーの検討
 - 階層的な分類
 - 例) ラーメン → 味噌ラーメン、塩ラーメン、豚骨ラーメン、醤油
冷し中華、ごまだれ冷し中華、つけ麺、油そば、

ご清聴ありがとうございます