FoodCam: スマートフォン上でのリアルタイム食事画像認識による食事記録アプリケーション

電気通信大学　総合情報学専攻
河野 憲之, 柳井 啓司

はじめに

- 日々の食事記録をとるサービスの増加
  - 携帯端末から、PCから利用
  - レコーディングダイエット

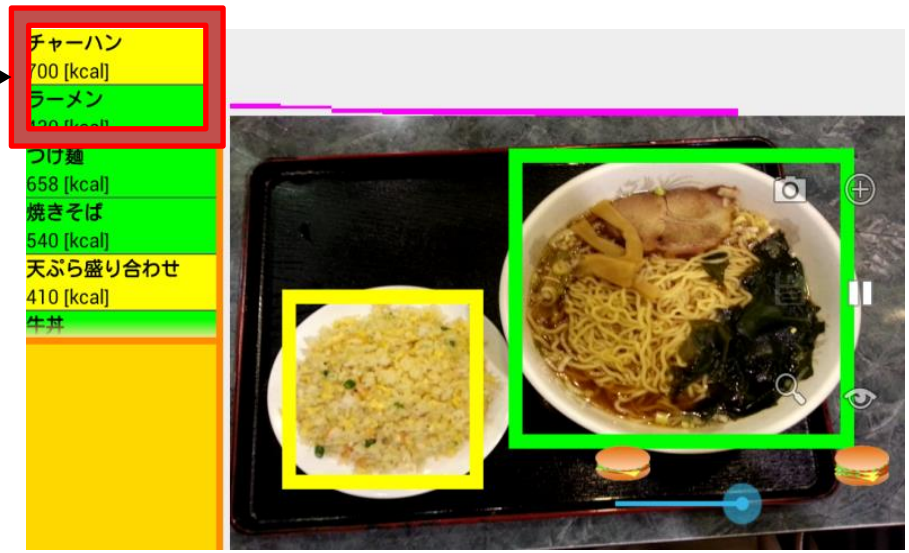
- スマートフォンの普及と性能向上
  - スマートフォン上で画像認識が可能に

目的

- 一般的な食事記録システム
  - 全てが手入力、大きい手間、利用に飽きる恐れ

食事記録を手軽にとる

チャーハン
ラーメン



関連研究

- 松田らの研究 (International Conference on Multimedia and Expo ‘12)
  - 100種類の食事を認識
  - 実行はクラスタマシン

食事画像

認識エンジン

候補料理
1．ごはん
2．味噌汁
3．目玉焼き
4．豚カツ
5．鮭のムニエル
6．魚のフライ
7．煮魚
8．ウインナーソテー
9．サンドイッチ
0．ロールパン

関連研究

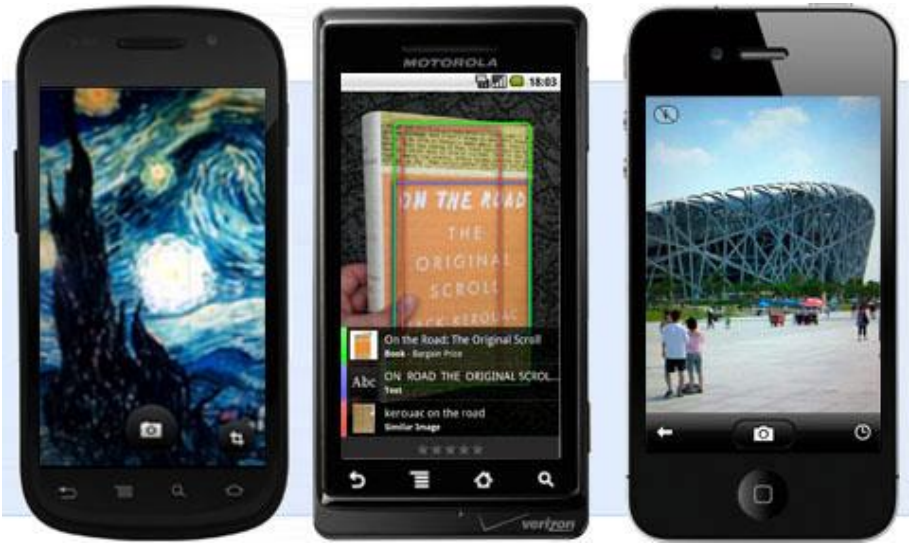
- FoodLog (東京大学相澤研究室) [Kitamura et al. CEA ‘09][高松ら MIRU ‘13]
  - 画像からカロリーや栄養を直接推定
  - どんな種類の料理でも栄養推定が可能
  - 料理の種類の記録には不向き

推定

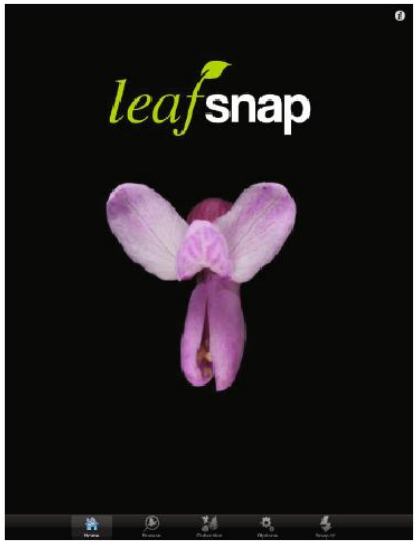
http://www.foodlog.jp

関連研究

- Google Goggles
  - ロゴや芸術品など
    特定物体を認識
  - 食事の認識不可

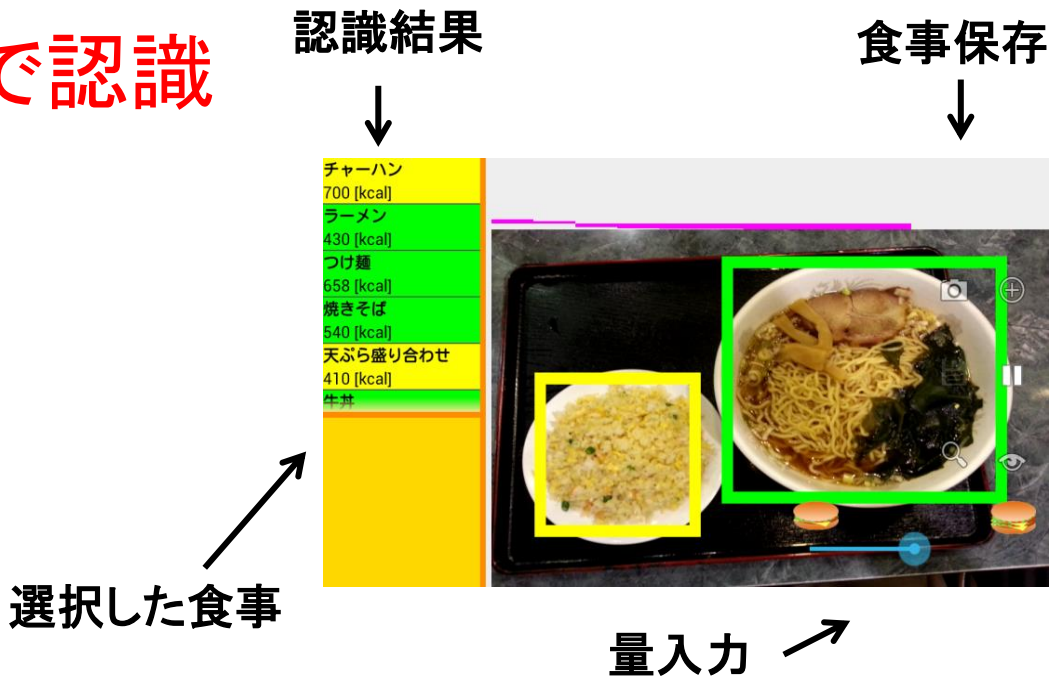


- Leaf snap (ECCV ’12)
  - 葉を認識する



既発表システム (DEIM'13)

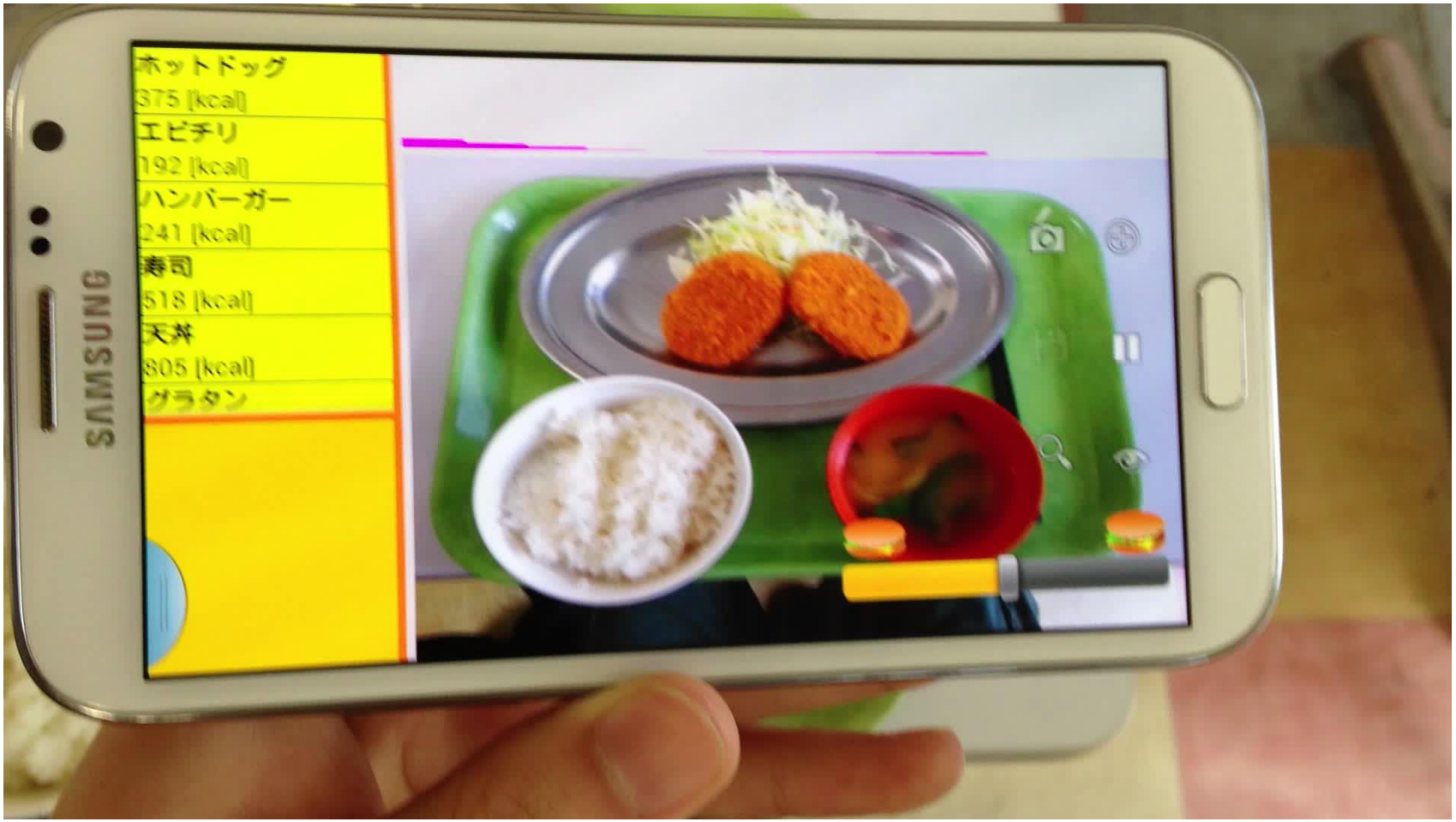
- FoodCam (河野ら DEIM'13)
  - 認識結果上位より選択
  - 50種類に対応
  - モバイル上で認識



システム使用例1


カレーライス
937 [kcal]
エビチリ
192 [kcal]
ホットドッグ
375 [kcal]
グリーンサラダ
79 [kcal]
寿司
518 [kcal]
きんぴらごぼう
SAMSUNG

システム使用例2


ホットドッグ
375 [kcal]
エビチリ
192 [kcal]
ハンバーガー
241 [kcal]
寿司
518 [kcal]
天丼
805 [kcal]
グラタン
SAMSUNG

既発表システムの問題点と解決法

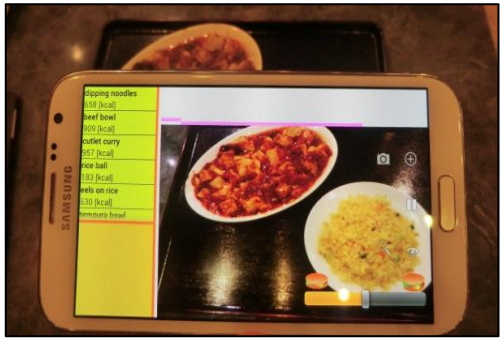
- 認識精度が十分でない
  - Bag-of-Featuresの代わりに
    最新手法のFisher Vectorを導入

    精度向上の実現
    100種類に倍増

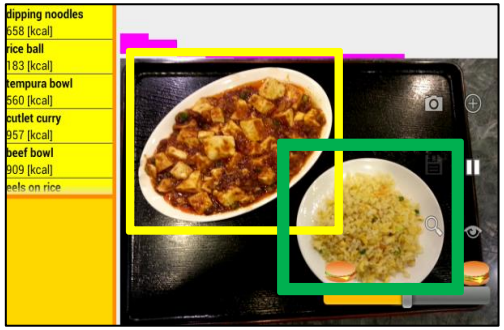
- 不要な機能があった
  - 方向提示機能の廃止

    処理時間の節約

システムの使用の流れ

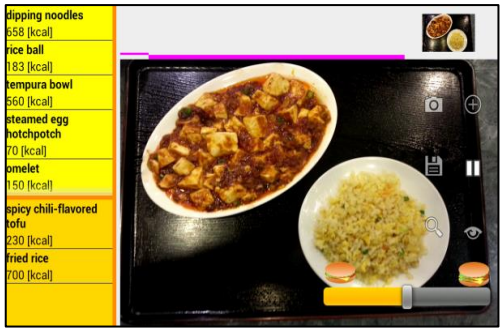


スマートフォンを
食事にかざす

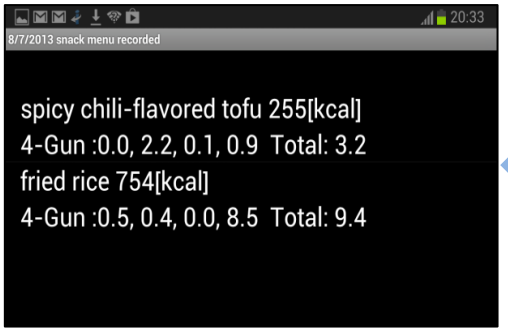


認識領域を与える

認識領域の自動修正



食事画像認識
結果リストから選択

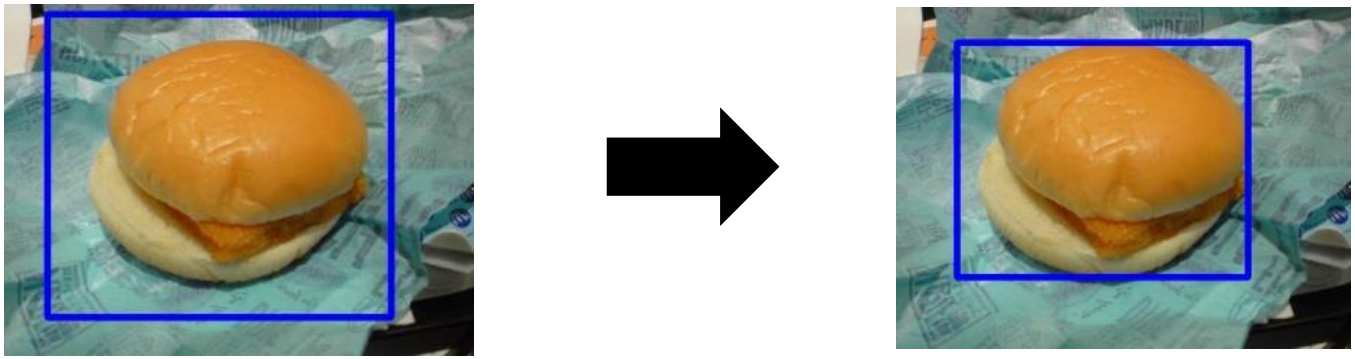


栄養を確認、登録

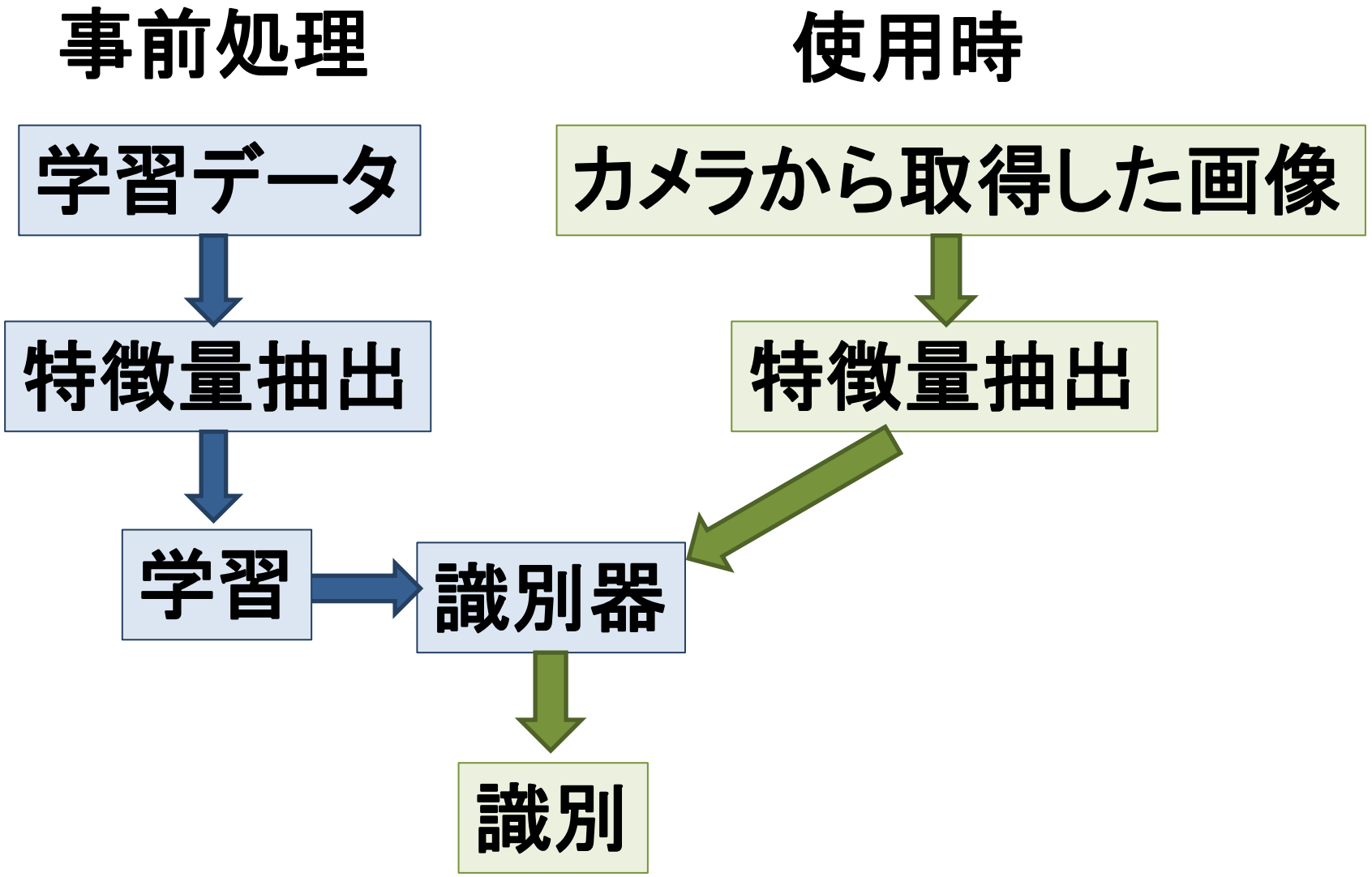
記録確認

認識する領域の修正 (変更なし)

- ユーザは料理より大きく矩形領域を与える
- GrabCut
  - 前景を全て含む最小の矩形領域に領域を補正
- 領域入力と同時にバックグラウンドで一度のみ実行



認識の流れ



従来手法との違い

従来

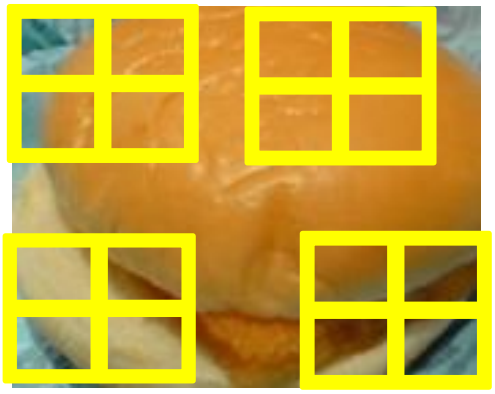
- 局所特徴量
  - SURF
    - 64次元
- 画像特徴量
  - SURF-BoF
  - Color Histogram
- Feature embedding
  - chi2 kernel feature map

改良後

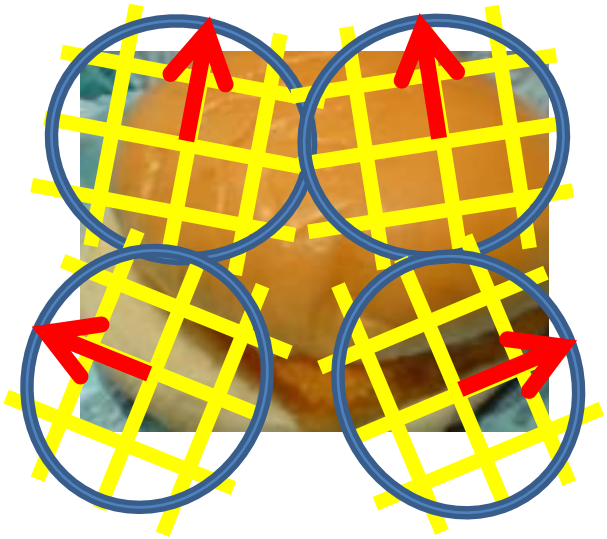
- 局所特徴量
  - HOG Patch
    - SIFTと類似、だが高速
    - 32次元
  - Color Patch
    - RGBの1^{st}, 2^{nd}モーメント
    - 24次元
- 画像特徴量
  - HOG Patch Fisher Vector
  - Color Patch Fisher Vector

認識手法詳細

- 特徴量
  - HOG Patch
    - 回転不変でない
    - 高速



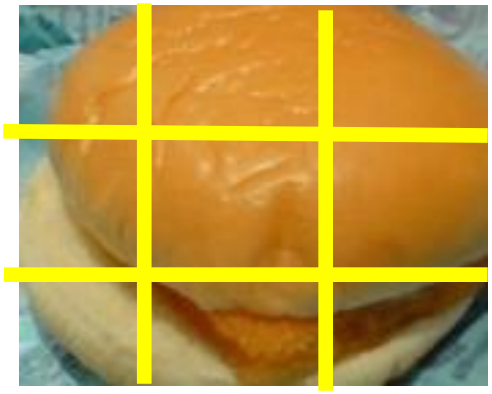
  - SIFT
    - 回転不変
    - 重みづけ



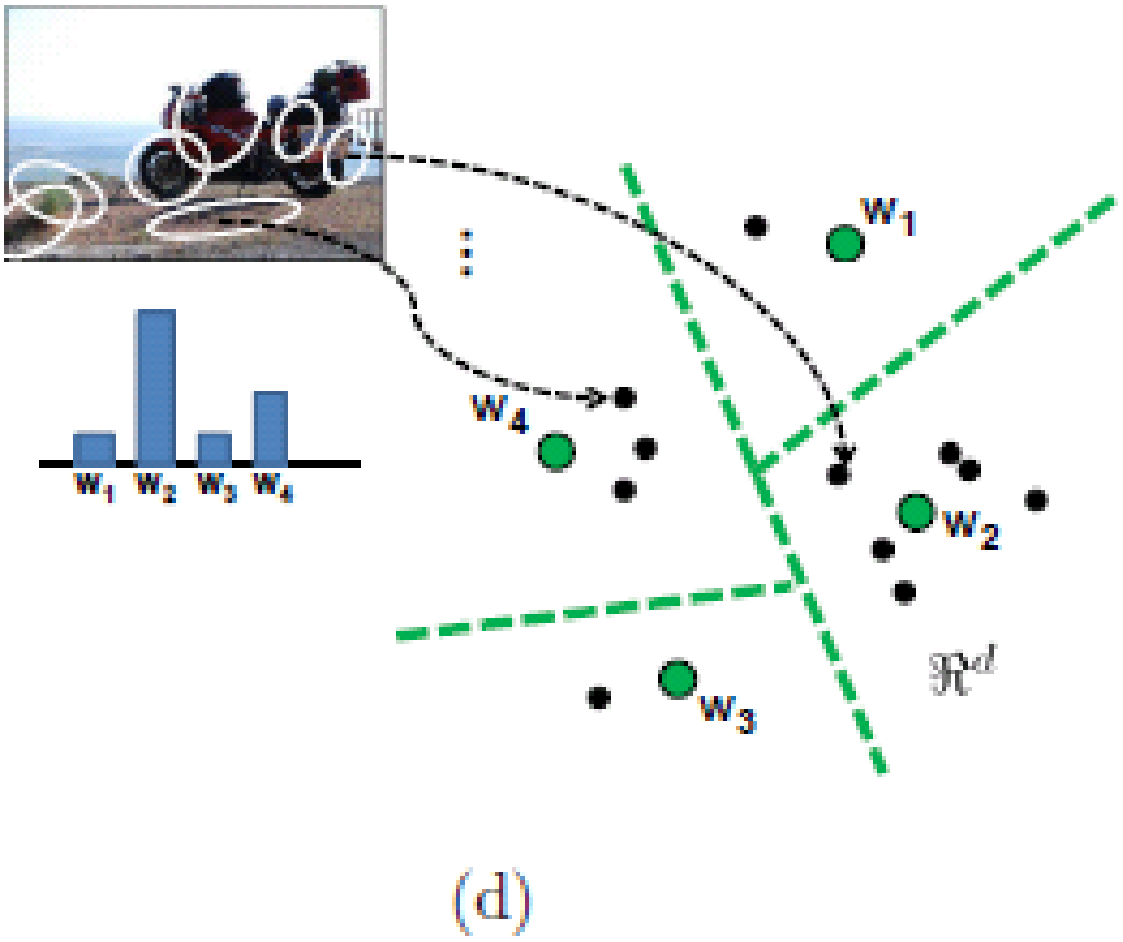
認識手法詳細

- 特徴量
  - Color Patch
    - 局所特徴量

  - カラーヒストグラム
    - 大域特徴
    - 極めて高速



Bag of Features



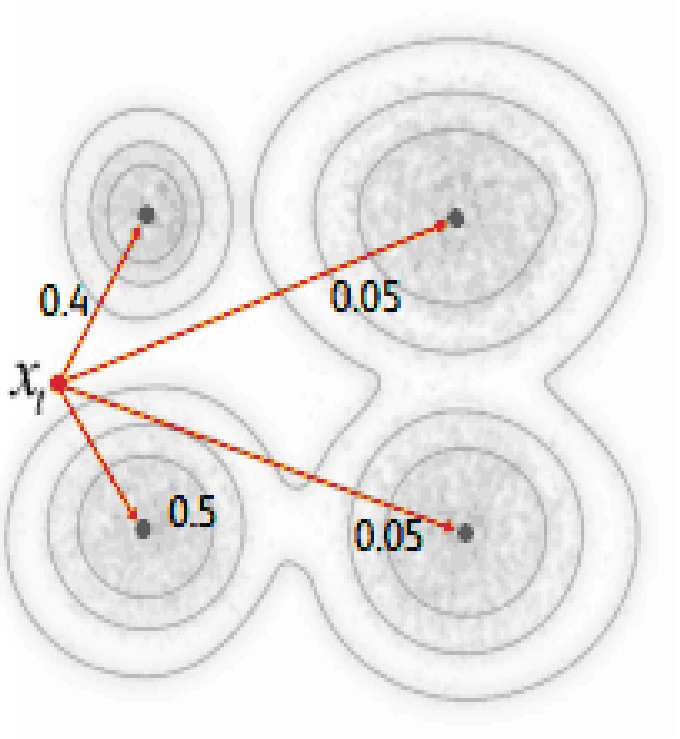
http://www.cs.utexas.edu/~grauman/courses/fall2009/papers/bag_of_visual_words.pdfより引用

Fisher Vector

- gradient wrt to μ and σ

$$\mathcal{G}^X_{\mu,i} = \frac{1}{T\sqrt{w_i}} \sum_{t=1}^{T} \gamma_t(i) \left(\frac{x_t - \mu_i}{\sigma_i} \right)$$

$$\mathcal{G}^X_{\sigma,i} = \frac{1}{T\sqrt{2w_i}} \sum_{t=1}^{T} \gamma_t(i) \left[\frac{(x_t - \mu_i)^2}{\sigma_i^2} - 1 \right]$$



ECCV2012: Part 4 Explicit embeddings (Ⅲ): high order representationより使用

識別器

- 線形SVM

$$
\begin{aligned}
f(x) &= \sum_{i=1}^{M} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \\
&= \sum_{i=1}^{M} y_i \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle + b \\
&= \langle \sum_{i=1}^{M} y_i \alpha_i \, \mathbf{x}_i, \mathbf{x} \rangle + b \\
&= \langle \mathbf{w}, \mathbf{x} \rangle + b
\end{aligned}
$$

SVをかけておくことで
サンプル数によらない

計算量 : $O(N)$

メモリ : $O(N)$

記録閲覧 (変更なし)

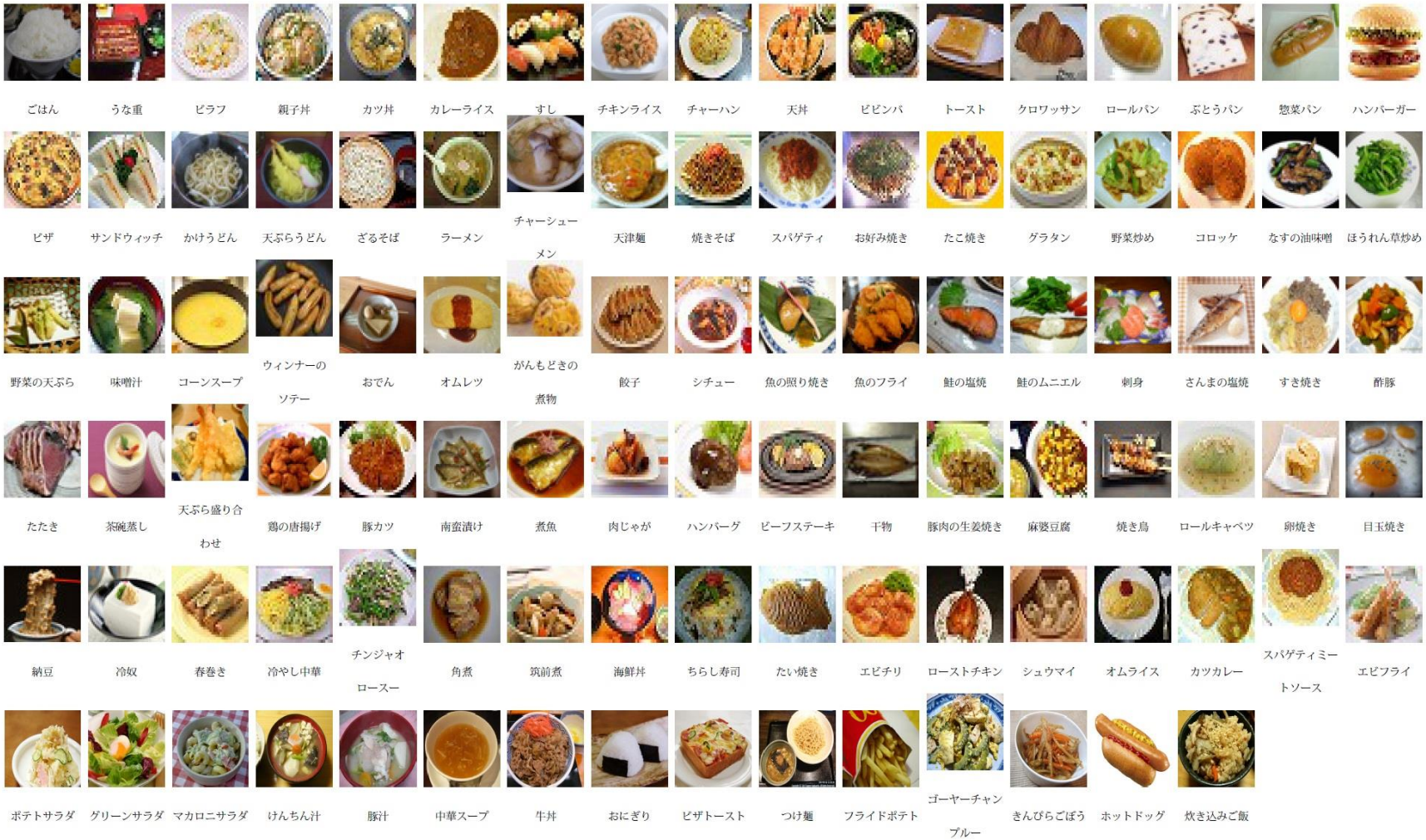
デバイス上で閲覧

サーバにアップロードしてWebで閲覧共有

実験

- 認識精度の評価
  - 比較
    - クライアントサイド(従来)
    - サーバサイド(松田ら) (PCクラスター版)
  - データセット
    - 100種類、各100枚以上
    - 人手でつけられた料理領域を使用
  - 評価方法、5-cross validation
    - 分類率 = $\frac{\text{候補}N\text{位までに正解を含む画像枚数}}{\text{評価画像枚数}}$
- 認識時間の評価

認識対象の100種の食事



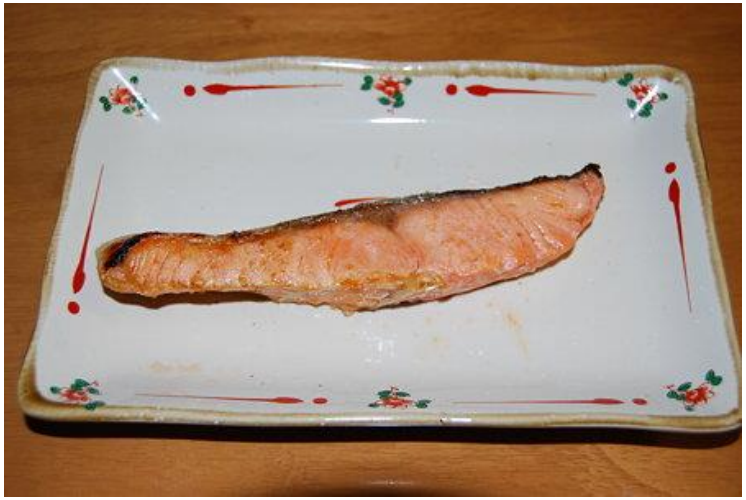
ご飯類：　18種類

麺類: 11種類

パン類: 11種類

肉類: 11種類

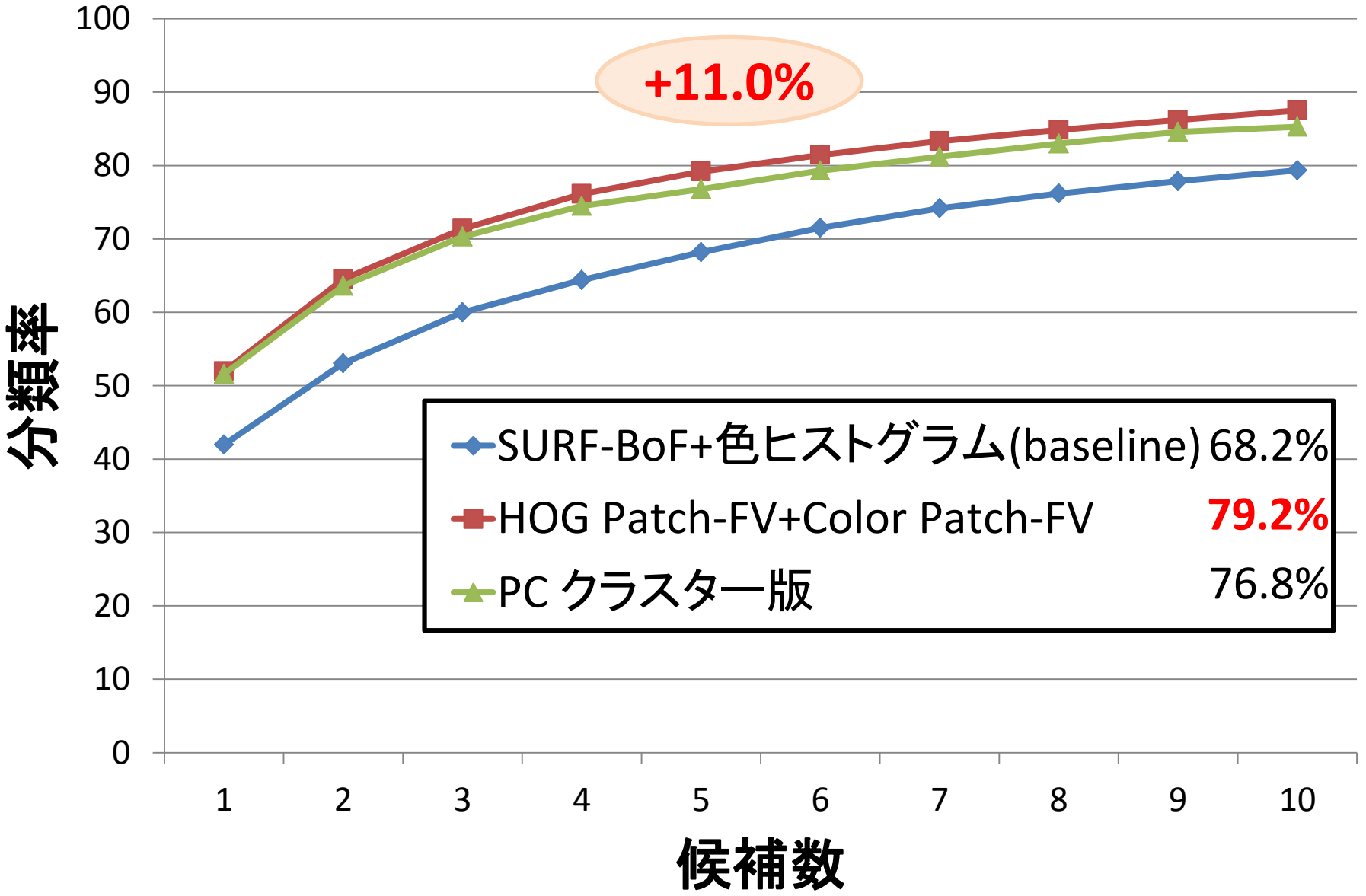
魚介類: 8種類



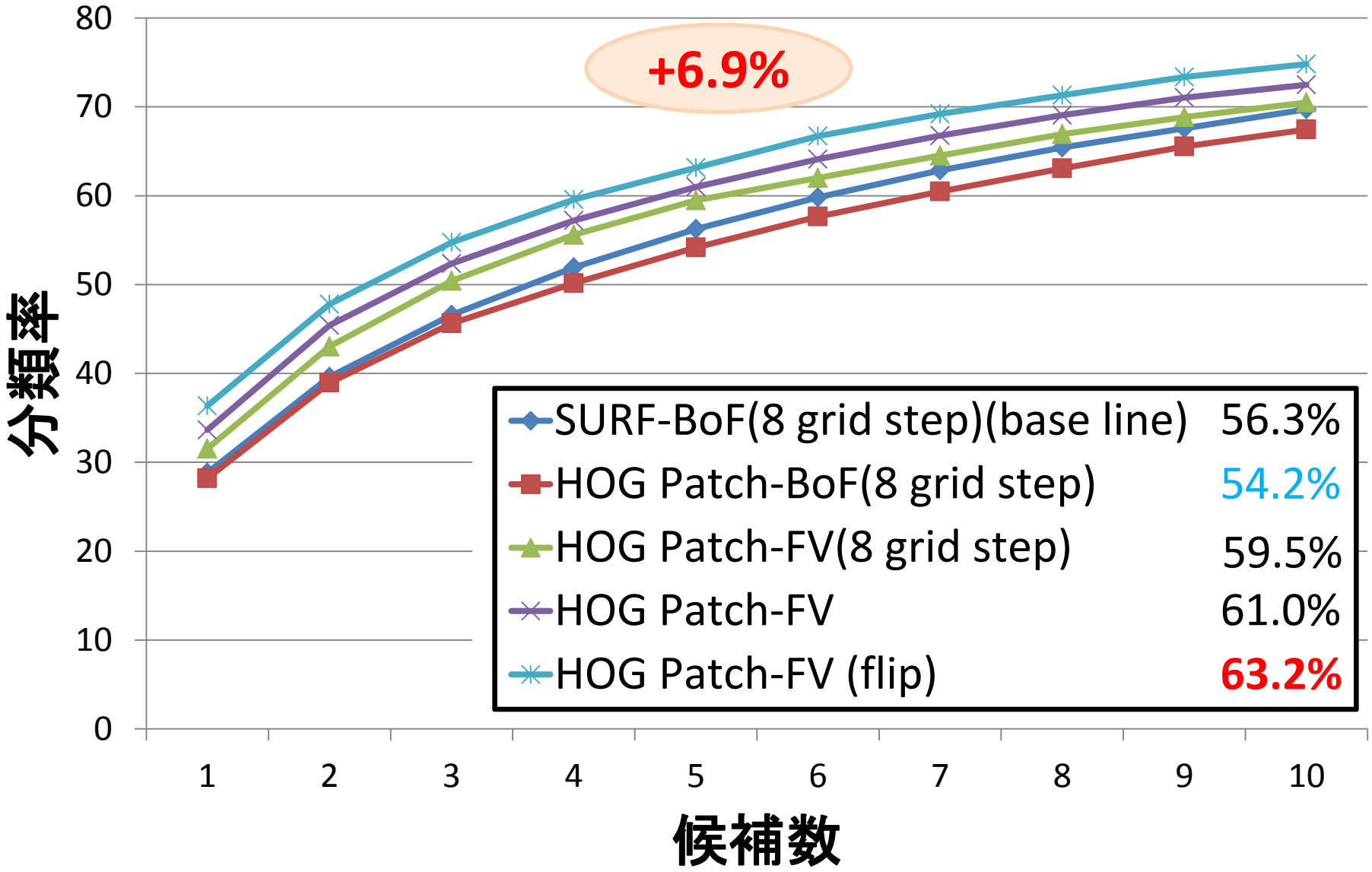
惣菜: 33種類

サラダ: 3種類、　汁物: 5種類

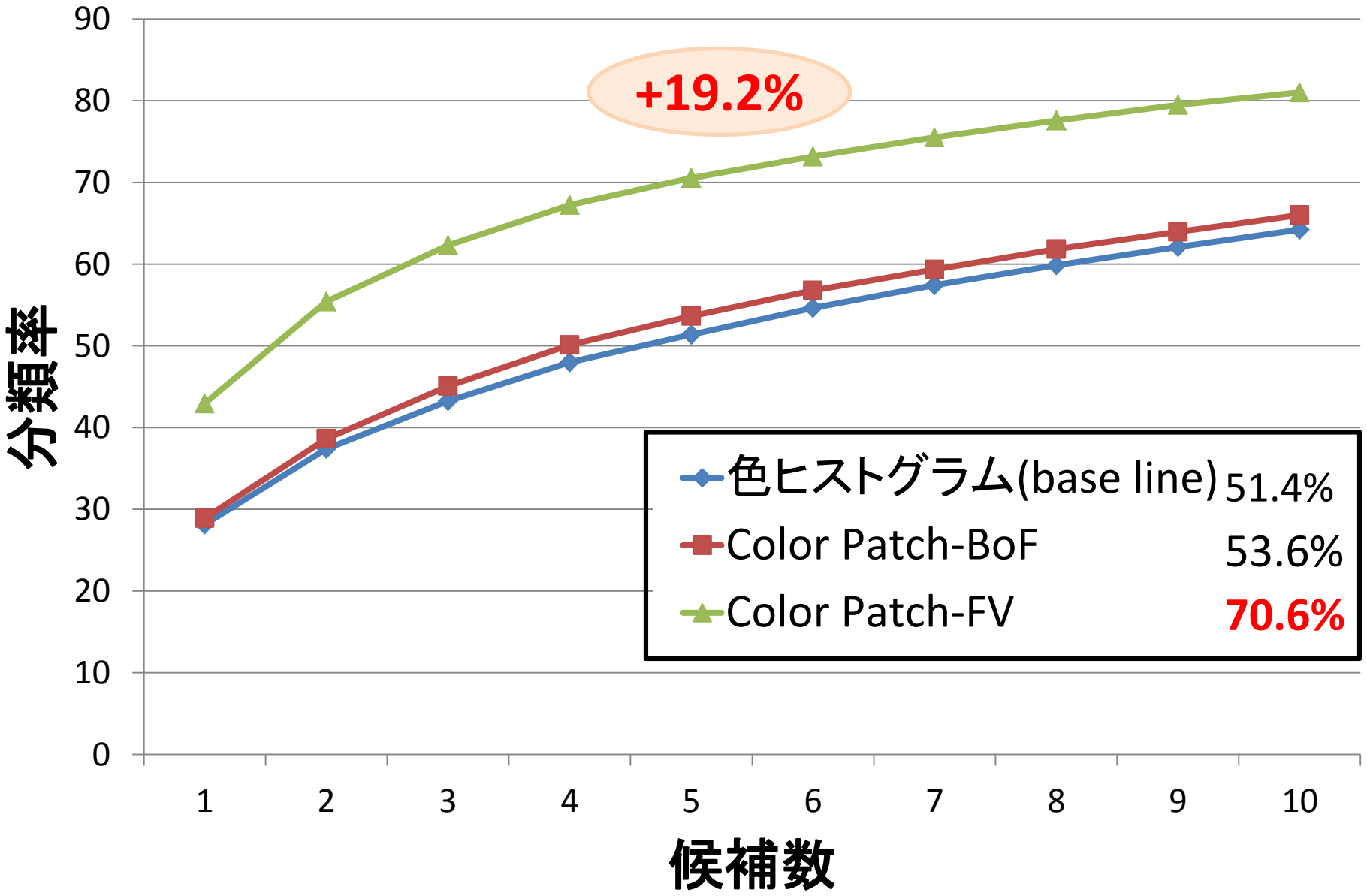
システムの認識率の比較



SURF BoFとHOG Patch BoF, FVの比較



色ヒストグラムとColor Patchの比較

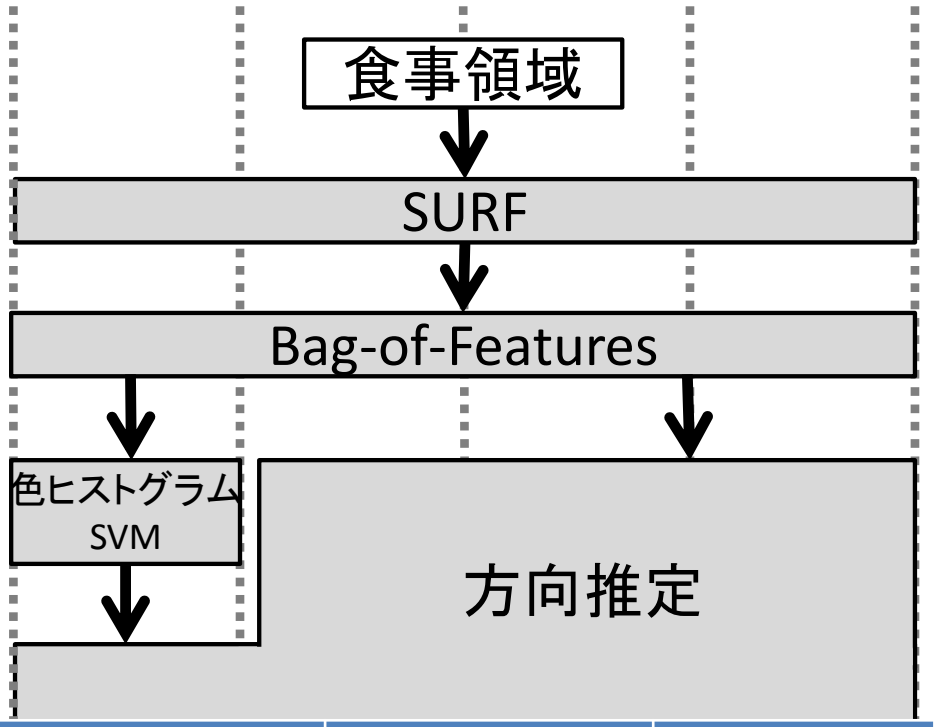


実行速度実験設定

- 速度評価
  - デバイス
    - Galaxy Note2(1.6GHz,4core,Android 4.1)
  - Quad コアで並列処理

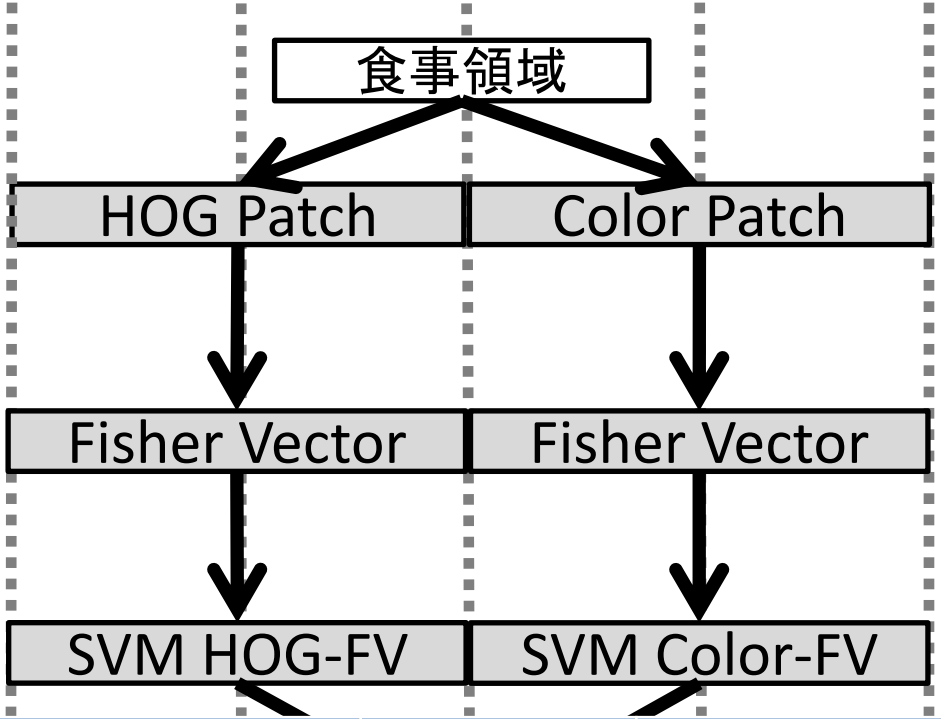
- 評価方法
  - 実時間
  - 20回計測し、その平均値

認識処理の比較



認識	+方向推定	認識対象
0.26 sec	0.34sec	50

コア1 コア2 コア3 コア4



認識	高速化	認識対象
0.065 sec	4倍	100

コア1 コア2 コア3 コア4

まとめ

- 提案システム: FoodCam
  - 料理100種類を認識
  - スマートフォン上で高速認識
    - 認識部分を改良
  - 料理領域が与えられると、バックグランドで領域補正

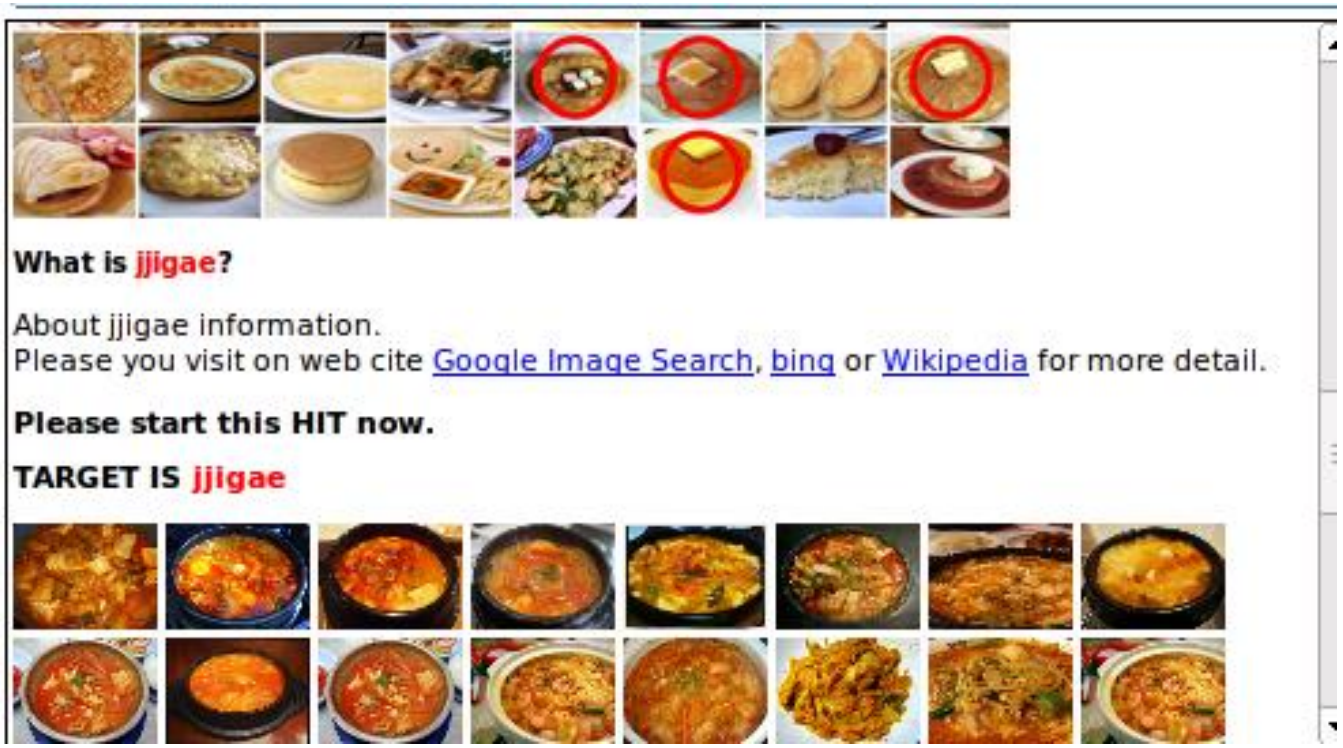
今後の課題

- 料理領域の検出

- ユーザ情報の利用
  - 識別器の出力は使い具合により、可変に
  - ユーザが認識対象を選択可能に

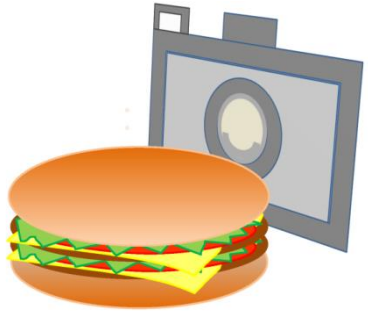
今やっていること

- Amazon Mechanical Turkによる
  食事画像データセット自動構築

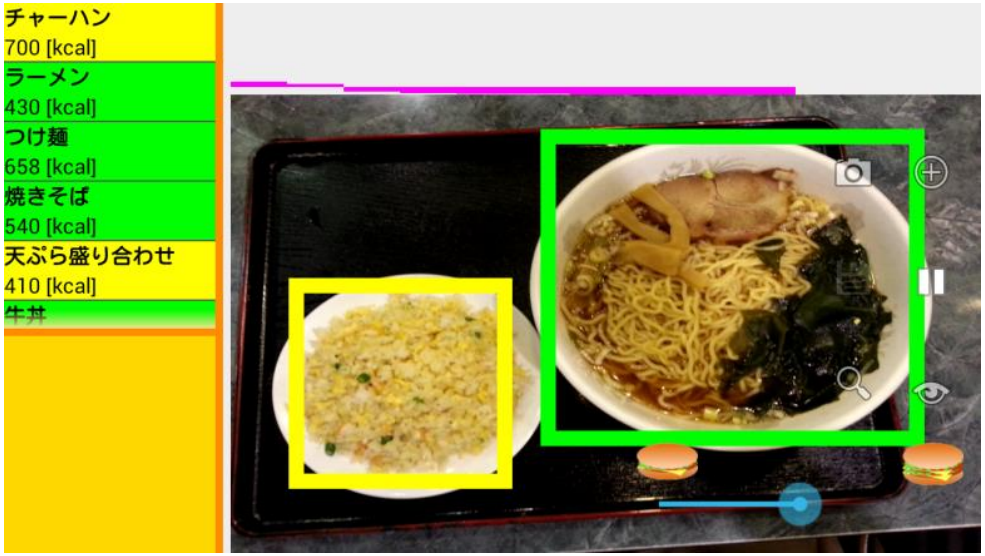


アプリ公開中

- FoodCam
  - http://foodcam.jp/



  - Android 3.1以上
  - Quad コア推奨、Dualでも動作可能



従来食事性の高い方向提示

- 認識領域内に多数領域を生成
- SVMの評価値最大の方向を提示

- ユーザによるシステム評価結果

認識のよさ	3.4
使いやすさ	4.2
方向提示のよさ	2.4
手動と比較してのよさ	3.8

従来システムとその違い

- 認識する領域の修正
  - 従来と同様
- 食事画像認識
- 食事性の高い方向提示
  - 使用しない