

DS-20 Fisher Vectorを用いたスマートフォン上でのリアルタイム高精度物体認識システム FoodCam

電気通信大学大学院 情報理工学研究科 総合情報学専攻 河野憲之, 柳井啓司

FoodCam 特徴

- スマートフォン上で動作 (4コア推奨)
- 100種類の食事を認識可能
- 1回の認識に0.065秒
- 上位5位以内に約80%の認識精度

FoodCam 概要

ダウンロード: <http://foodcam.jp>



全ての処理はスマートフォン上で実行

- ### 食事認識
- 画像の局所的な色・勾配
 - 100種類の料理識別器を構築
- ### 食事領域の修正
- 入力と同時にバググラウンドで実行
 - 一度のみ実行
- ### 記録閲覧
- 4群点数に基づいて評価

認識手法

局所特徴量

- HOG - 輝度勾配 (8方向, 2x2)
- Color - 色モーメント (1,2次, 2x2)

それぞれ 16x16, 24x24 のパッチを 6ピクセル毎のグリッドサンプリング

画像表現 Fisher Vector

GMM: K=32, base feature: 24dim(PCA)
Color-FV(1536dim), HOG-FV(1536dim)

識別器

線形SVM (late fusionで統合)

実装

画像サイズの縮小

200x150 (30K) ピクセルに縮小

HOG特徴抽出の高速化

先に画像全体ノ勾配方向と強度を算出

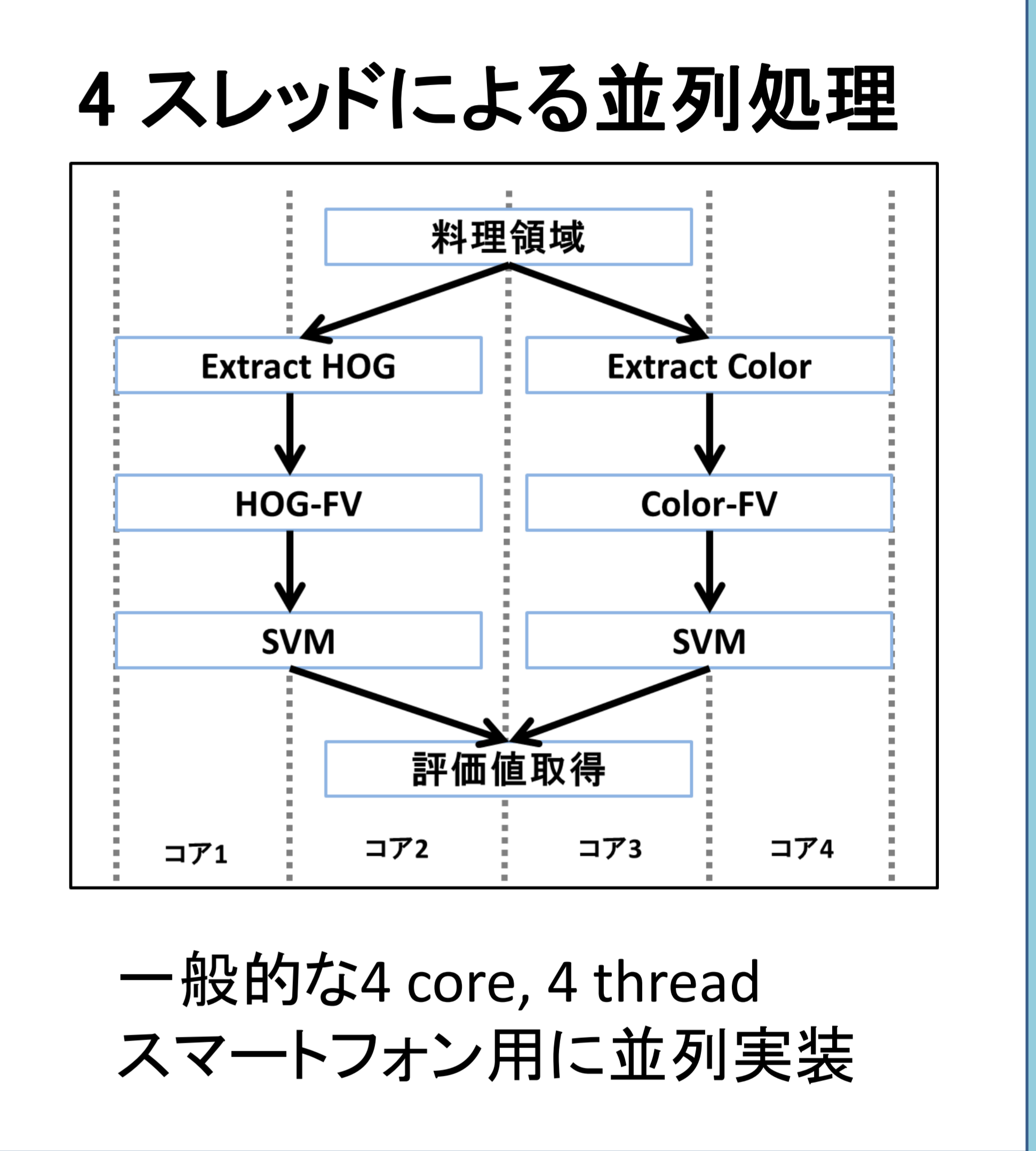
Fisher Vector計算の高速化

$$G_{\mu,i}^X = \frac{1}{T\sqrt{\pi_i}} \sum_{t=1}^T \gamma_t(i) \left(\frac{x_t - \mu_i}{\sigma_i} \right)$$

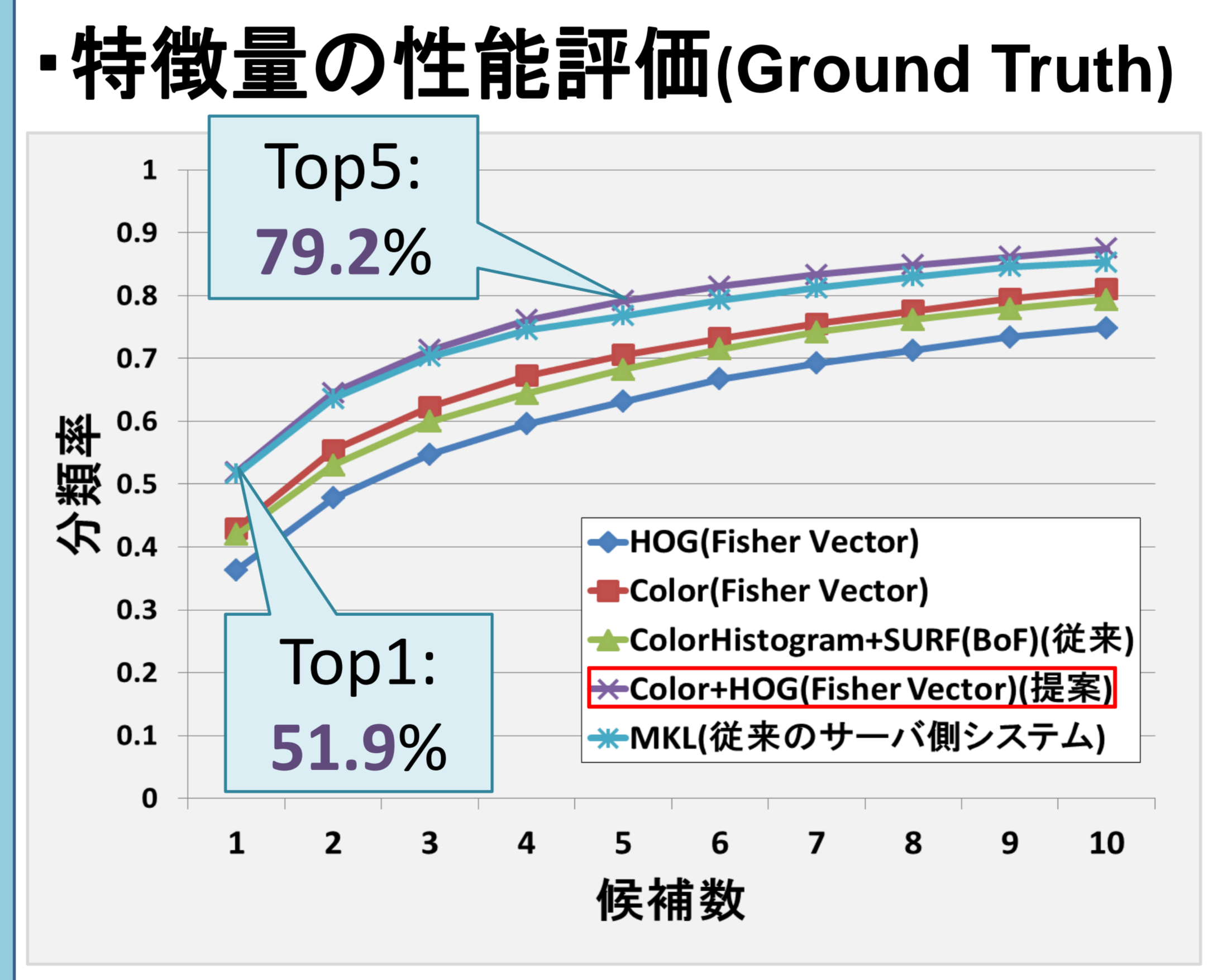
$$= \frac{1}{\sqrt{\pi_i}\sigma_i} \frac{1}{T} \sum_{t=1}^T \gamma_t(i) (x_t - \mu_i)$$

$$G_{\sigma,i}^X = \frac{1}{T\sqrt{2\pi_i}} \sum_{t=1}^T \gamma_t(i) \left[\frac{(x_t - \mu_i)^2}{\sigma_i^2} - 1 \right]$$

オフラインで一部計算
ルックアップテーブルの作成



実験



実行時間評価

デバイス: Galaxy Note 2 (1.6 GHz, 4 core, 4 Thread, Android 4.1)

項目	実行時間[秒]
領域補正	0.70
食事認識	0.065

食事認識システム公開中 (Android, Twitter)

食事認識Androidアプリ

ダウンロード・デモサイト:
<http://foodcam.jp>

食事認識bot

twitter @fooding_bot

- @fooding_bot宛に食事画像URLをtweet
- replyで認識結果が送られてきます