

k-meansによる局所特徴量抽出と 皿検出器による食事画像認識の改良

PRMU研究会 (2013/03/15)

松田 裕司 柳井 啓司 (電通大)

食事記録を取るサービスの普及



毎回の食事を記録することで健康管理を行う



カロリー, 脂質
過剰摂取注意など

昨日の懇親会...



食事記録を取るサービスの普及



毎回の食事を記録することで健康管理を行う

一般的な記録方式

テキスト入力, リスト選択

毎回の食事を記録するには手間が大きい

より簡単に記録を取ることが求められる

画像から自動的に料理を認識

食事記録のサポートを行う

複数品目の料理を含む画像も対象

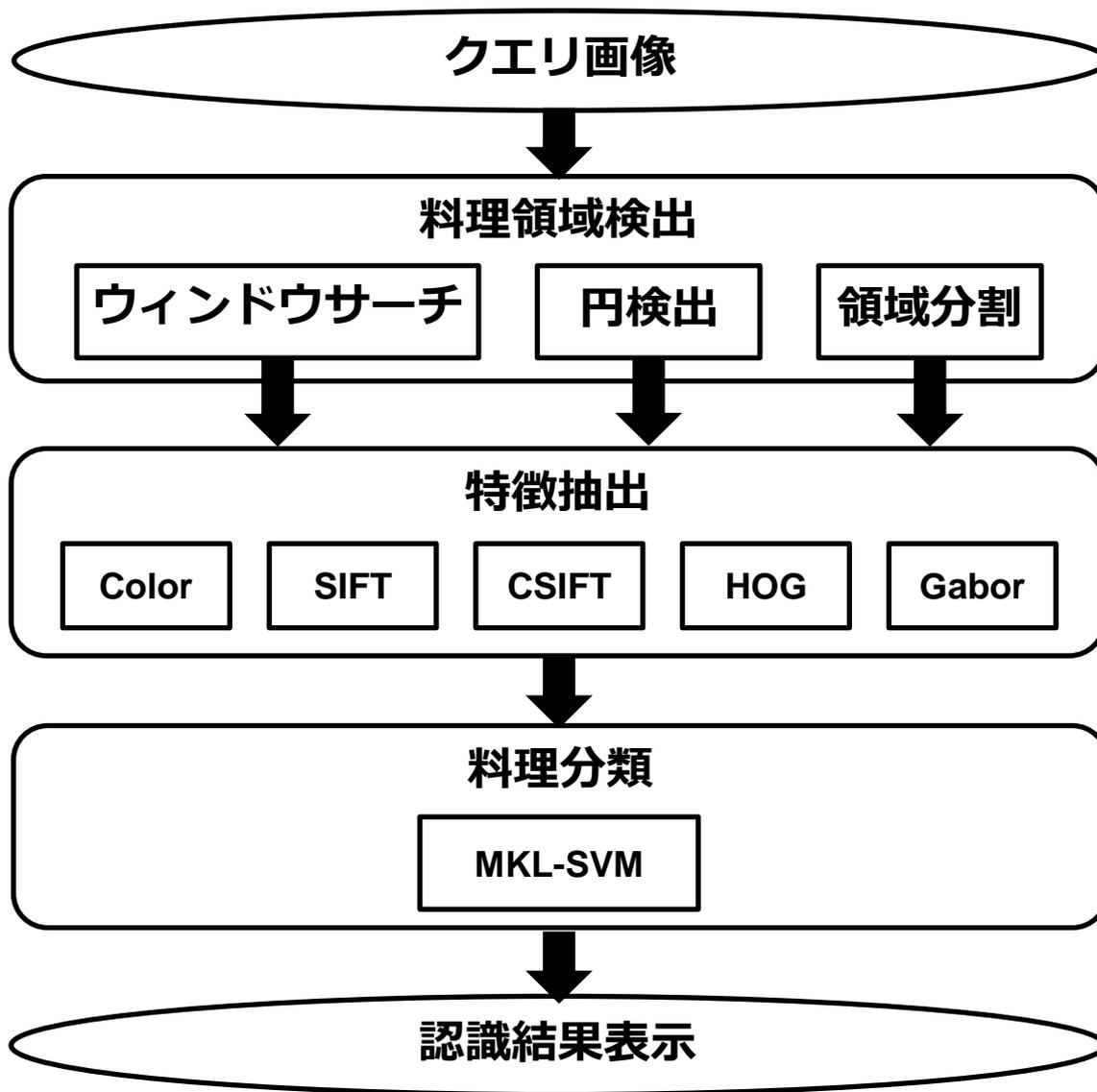


食事画像

認識

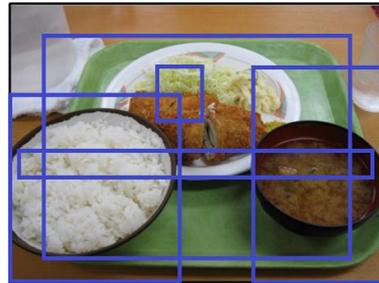
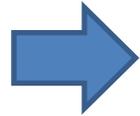
候補料理

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. ウィンナーソテー
9. サンドイッチ
0. ロールパン



料理領域検出

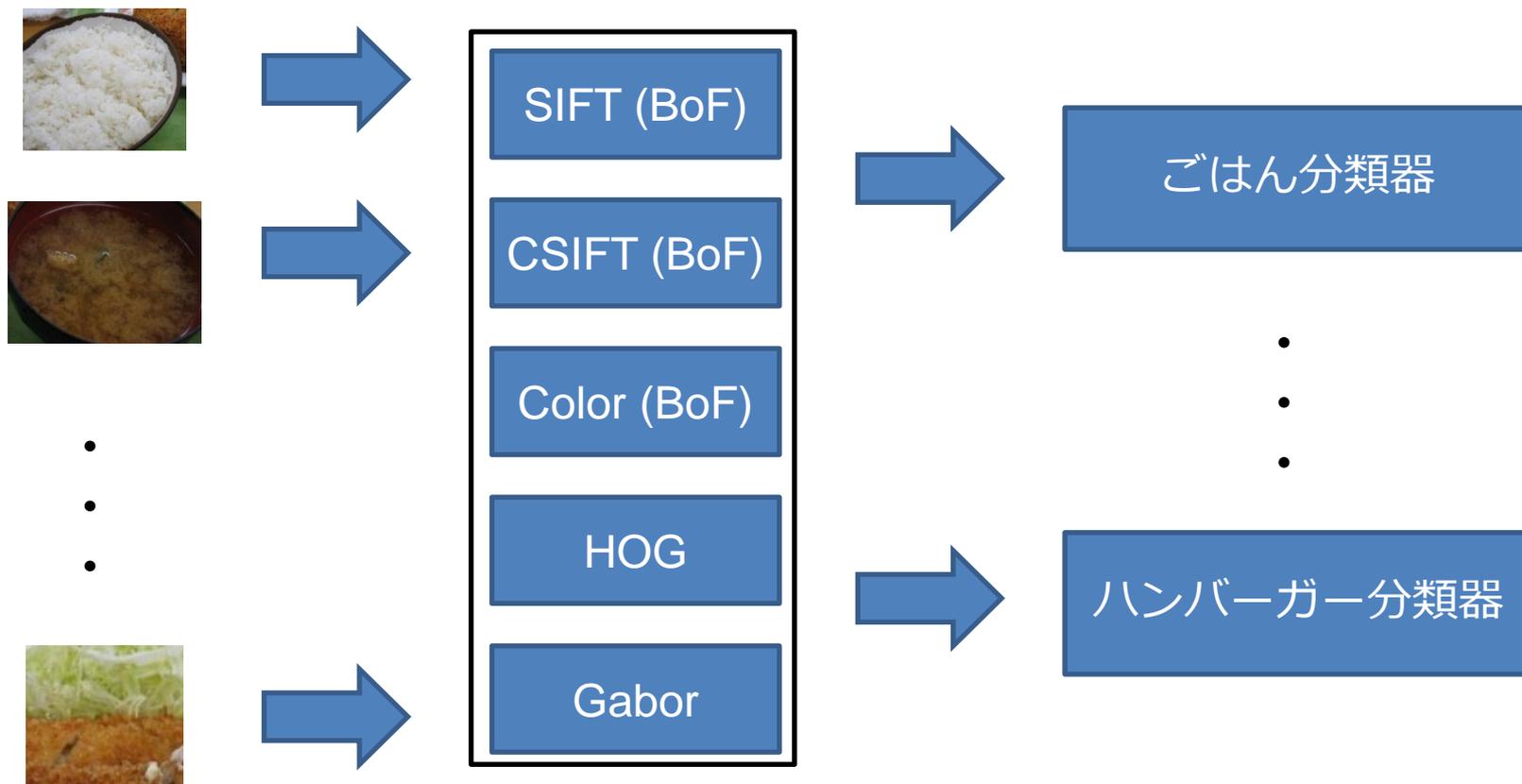
円検出, ウィンドウサーチ, セグメンテーション



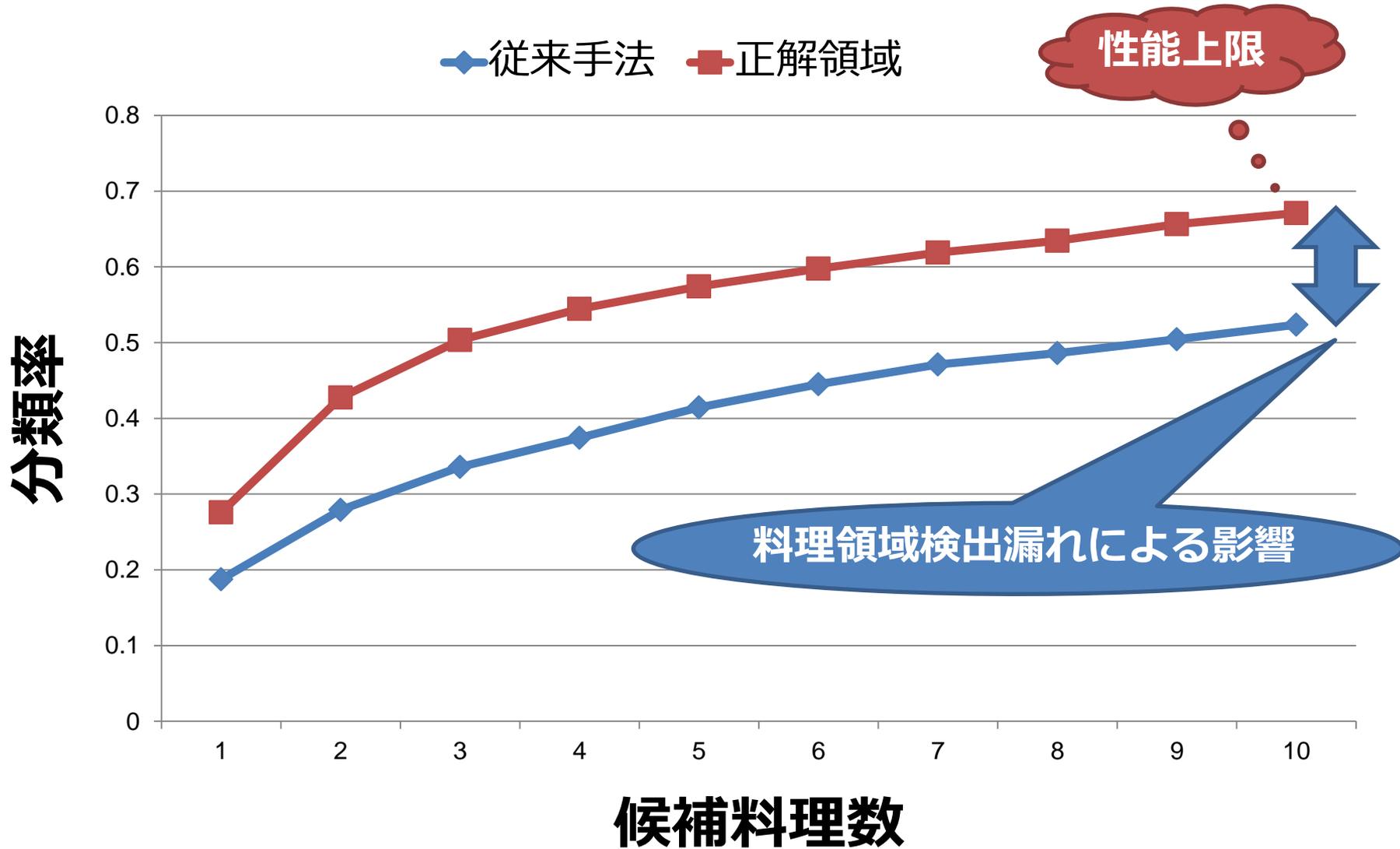
特徴抽出, 分類

特徴量 : SIFT, CSIFT, Color, HOG, Gabor

分類器 : Multiple Kernel Learning SVM



従来の認識性能 (参考)



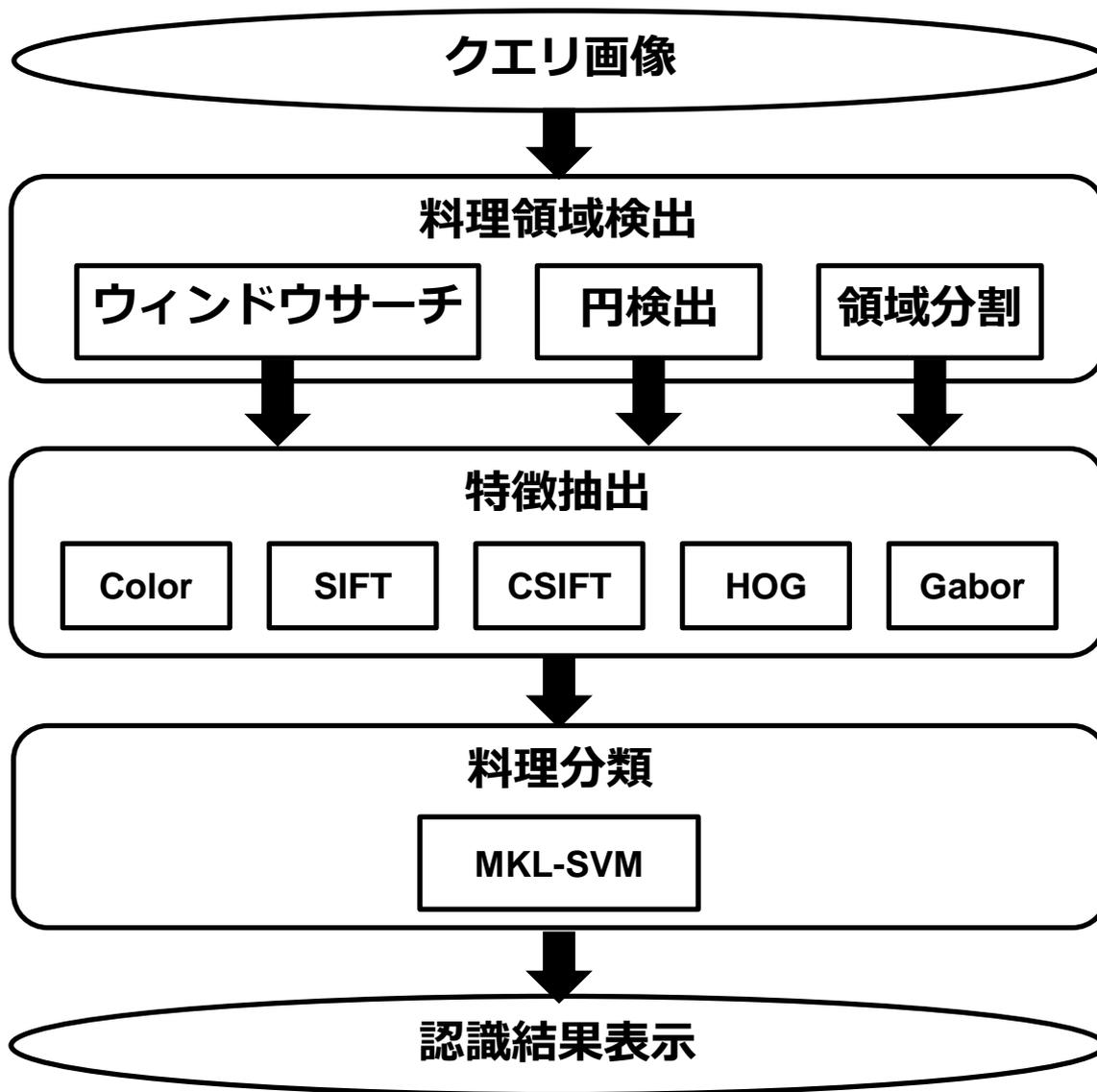
料理領域検出および料理分類の双方の観点で改良

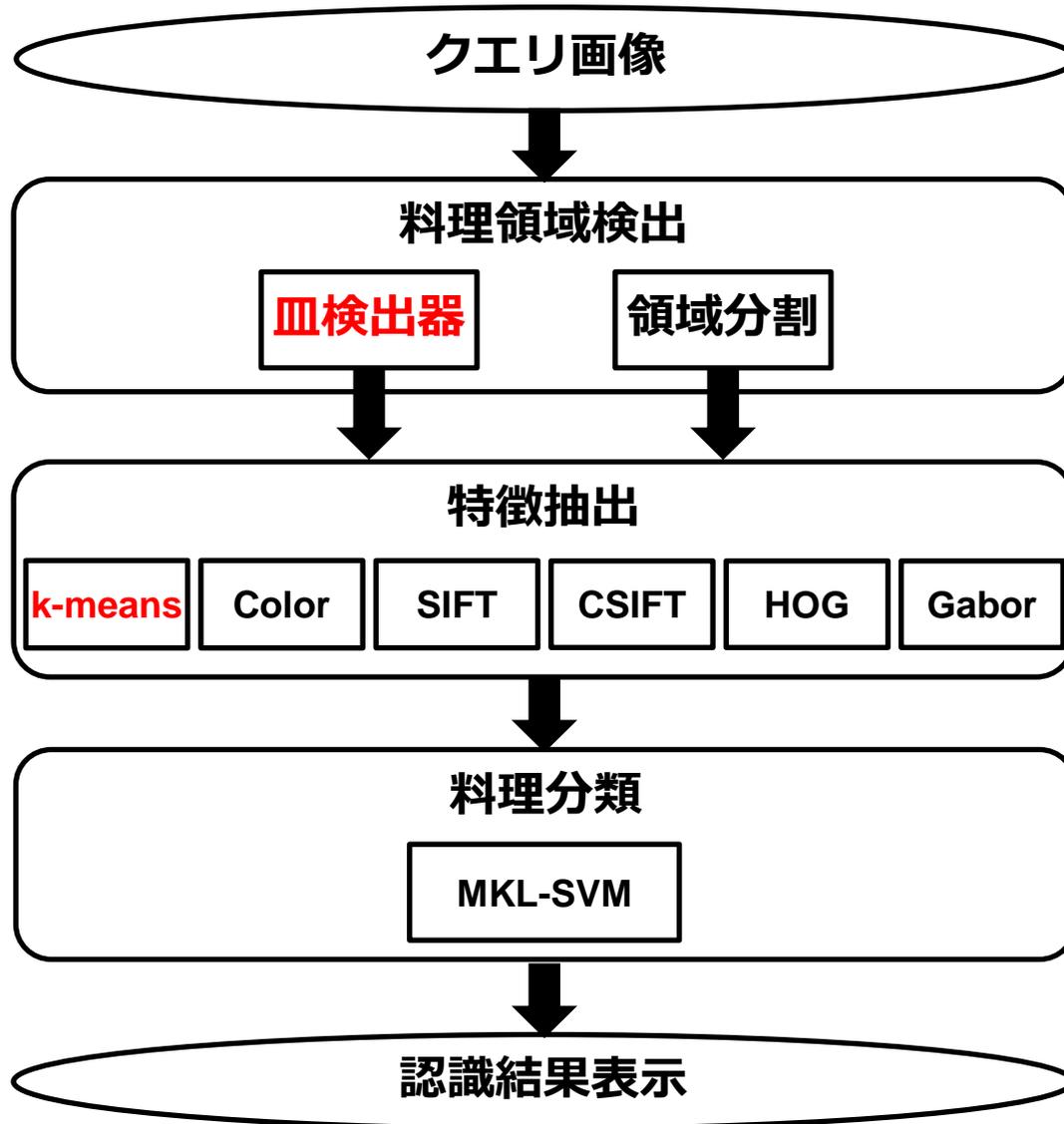
料理領域検出

料理の種類に依らない形状を学習

料理分類

教師なし学習による特徴量の導入





FoodLog (<http://www.foodlog.jp>)

画像から栄養を直接推定

Technology Assisted Dietary Assessment project

Purdue 大学の研究

チェッカーボードを利用していたり、
料理は白い皿に乗っているなど制約あり

PFID (Pittsburgh Fast-food Image Dataset)

単品のファストフード中心

[Yang et al. CVPR 2010] : 食材の位置関係を特徴量化

[Zong et al. ISM 2010] : SIFT特徴点検出 + LBP特徴

血検出器

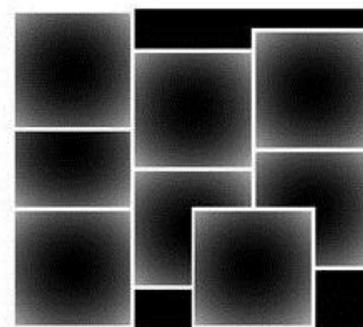
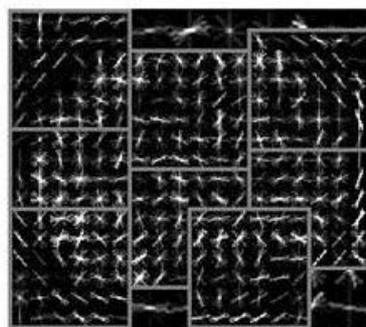
全体とパーツの組み合わせで表現 (*)

全体とパーツの形状およびパーツの位置で評価

複数のサブクラスモデルを持つ

特徴量：HOG

識別：線形SVM



料理の種類に依らない形状を学習

(*) P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan,
Object detection with discriminatively trained part-based models, PAMI 2010

料理ごとのモデル（従来）

学習・検出は料理クラス数（100）

正例：ある料理領域

負例：それ以外の料理を含む画像

検出された領域が正しいかSVMで再度評価

皿モデル

学習・検出は一度だけ

正例：全ての料理領域

負例：料理を含まない画像

検出された領域がどの料理であるかSVMで評価

教師なし学習による特徴量

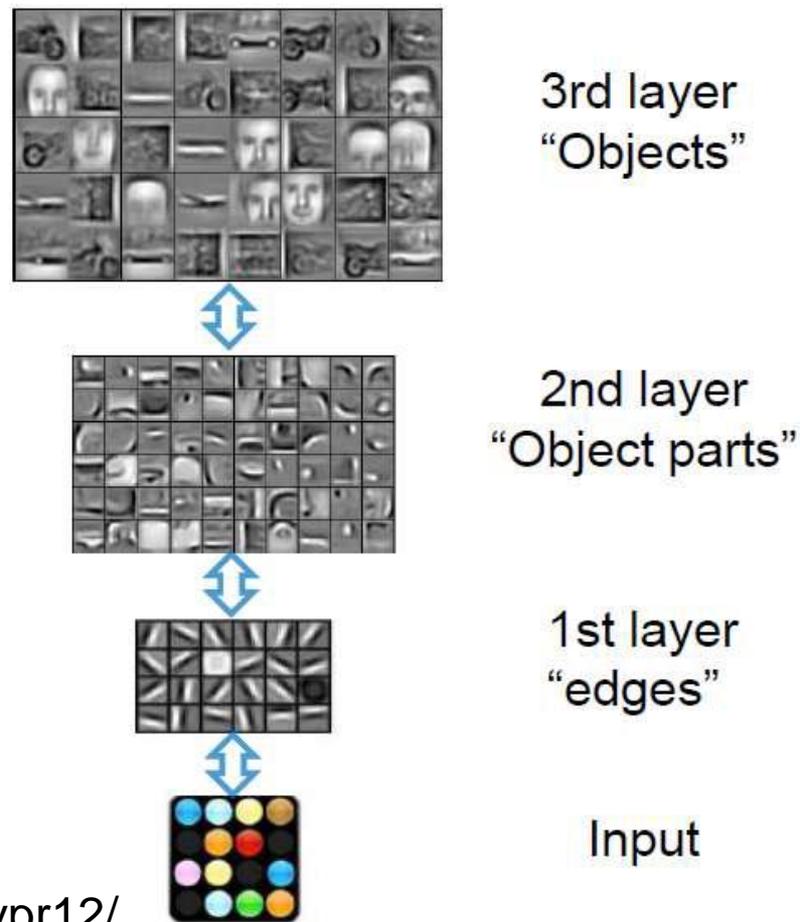
従来の特徴量：SIFT, HOG etc.

人手による設計 (hand-craft)

学習による特徴量設計の台頭

画素値からエッジなどを学習

特許の影響を受けない



CVPR 2012 Tutorialより引用

http://cs.nyu.edu/~fergus/tutorials/deep_learning_cvpr12/

k-meansによる手法(*)

小領域の画素値ベクトルをクラスタリング

sparse auto-encoder, sparse RBMよりも高精度

学習の流れ

1. 学習画像から大量の小領域（パッチ）を切り出す
2. 白色化（共分散行列の単位行列化）を行う
3. k-meansでクラスタリング
4. クラスタ中心をコードブックとする

(*) A. Coates, H. Lee, and A.Y. Ng.

An analysis of single-layer networks in unsupervised feature learning.

In Proc. of International Conference on Artificial Intelligence and Statistics, 2011.

学習された特徴量 (一部)

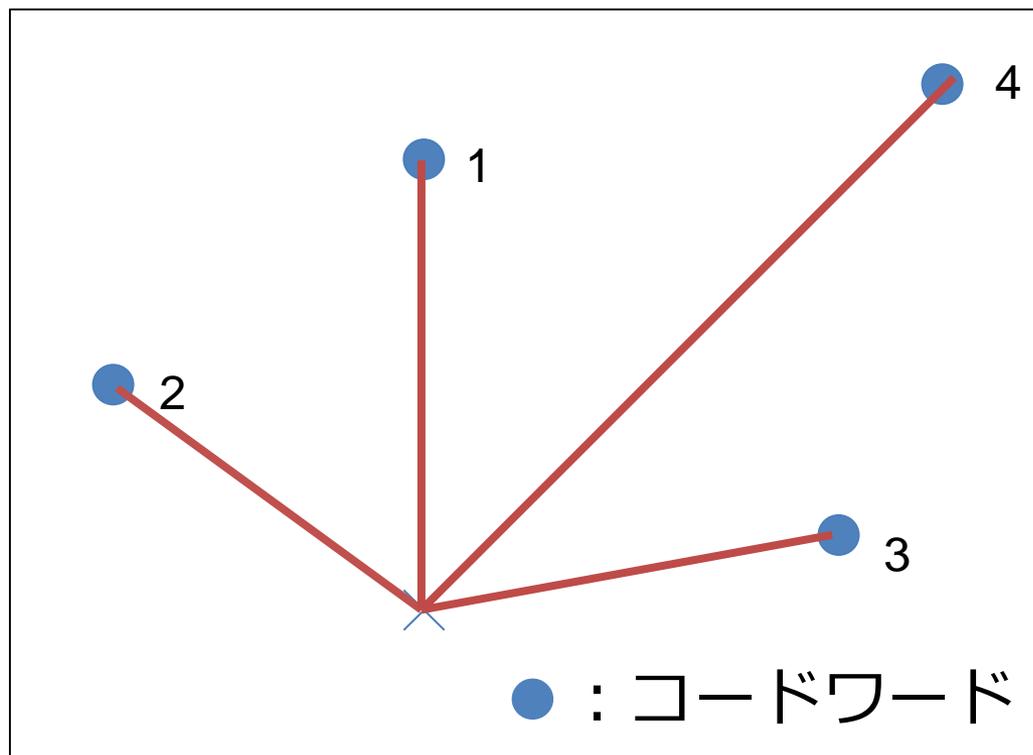


BoF: 最近傍コードワードの頻度

k-means: 各コードワードとの距離に応じた表現
(平均 - 距離)

BoF
[0, 1, 0, 0]

k-means
[0.85, 2.60, -1.76, -1.70]



内訳

100種類のカテゴリ（各100枚以上）

5分割したうちの4つを学習用，1つを評価用

以降の評価では複数の料理を含む画像のみを使用

評価方法

$$\text{分類率} = \frac{\text{第}N\text{候補までに挙げられた正しい料理品数}}{\text{認識対象の料理品数}}$$

1. 料理領域検出

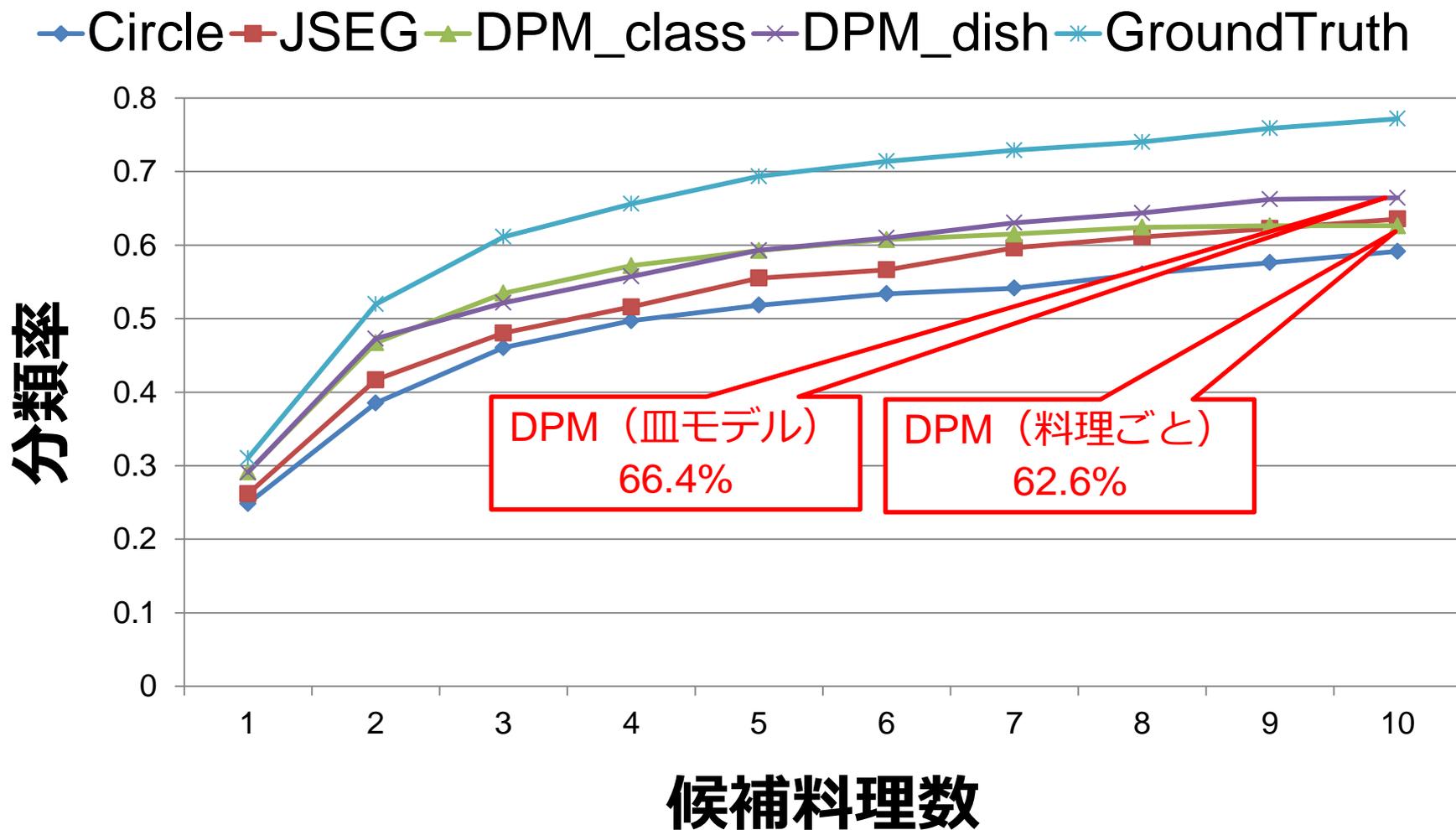
従来の領域検出手法（単体）との比較

他の領域検出手法と組み合わせた場合の性能

特徴量は従来のものを用いる

2. 料理分類

料理領域検出手法の比較



	円検出	領域分割	料理検出器	皿検出器
Avg. Recall	0.520	0.657	0.449	0.651
Avg. Precision	0.315	0.336	0.699	0.412

Recall

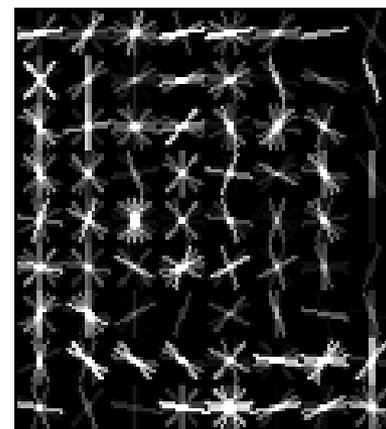
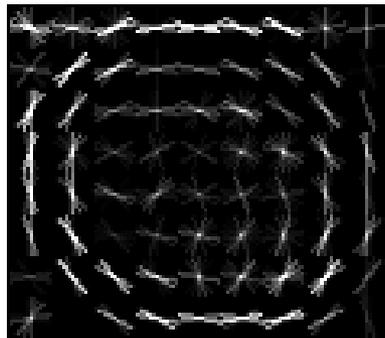
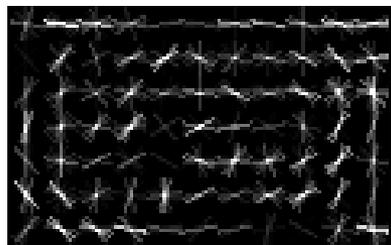
各正解領域と最も重なる候補領域とのオーバーラップの割合

Precision

各候補領域と最も重なる正解領域とのオーバーラップの割合

全体形状のHOGを可視化（サブクラス数3）

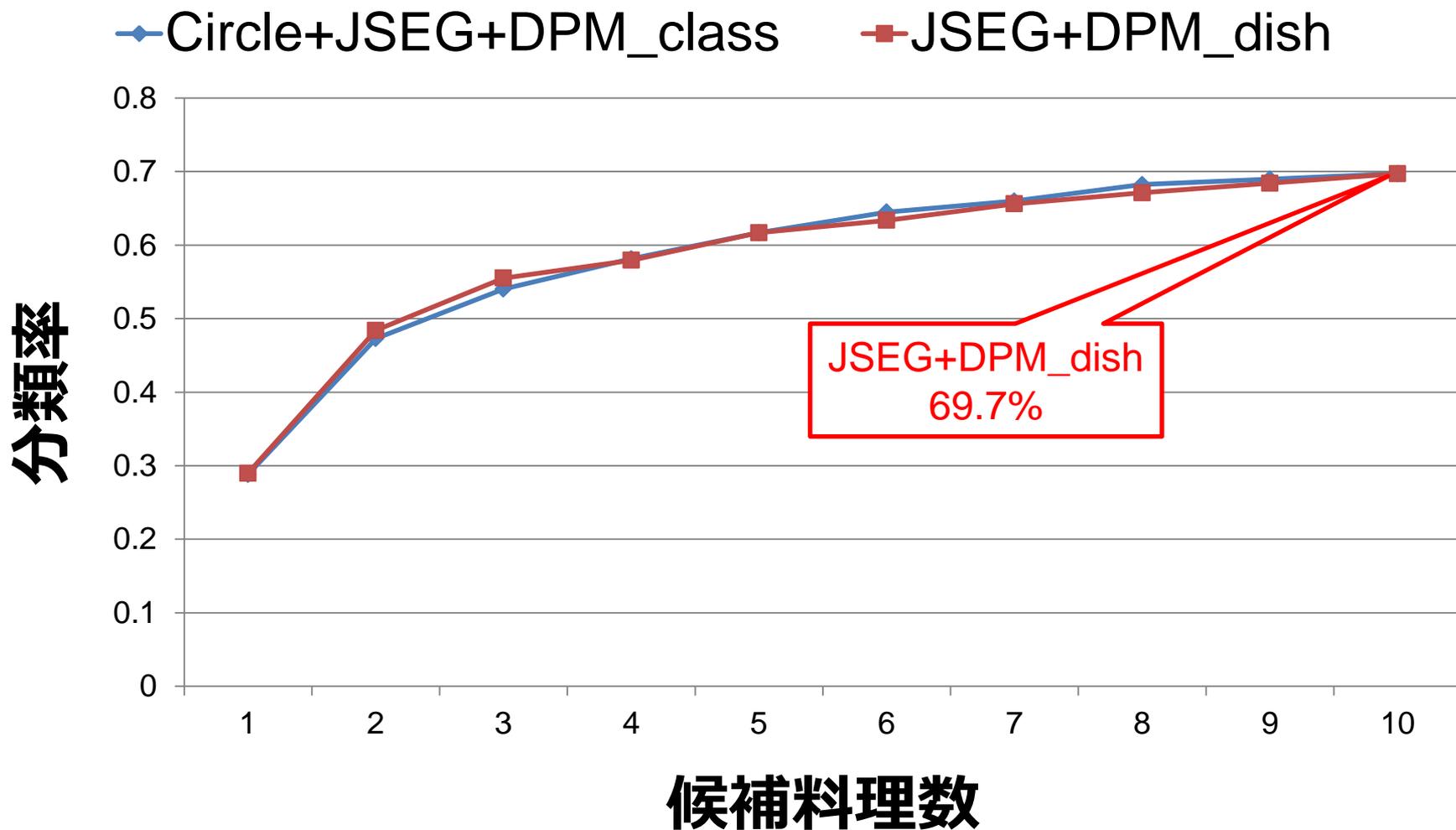
やや斜めから見た器の形状



真上から見た器の形状

?

検出手法の組み合わせ



1. 料理領域検出

2. 料理分類

BoF(SIFT)との分類性能の比較

コードブックサイズ：1600

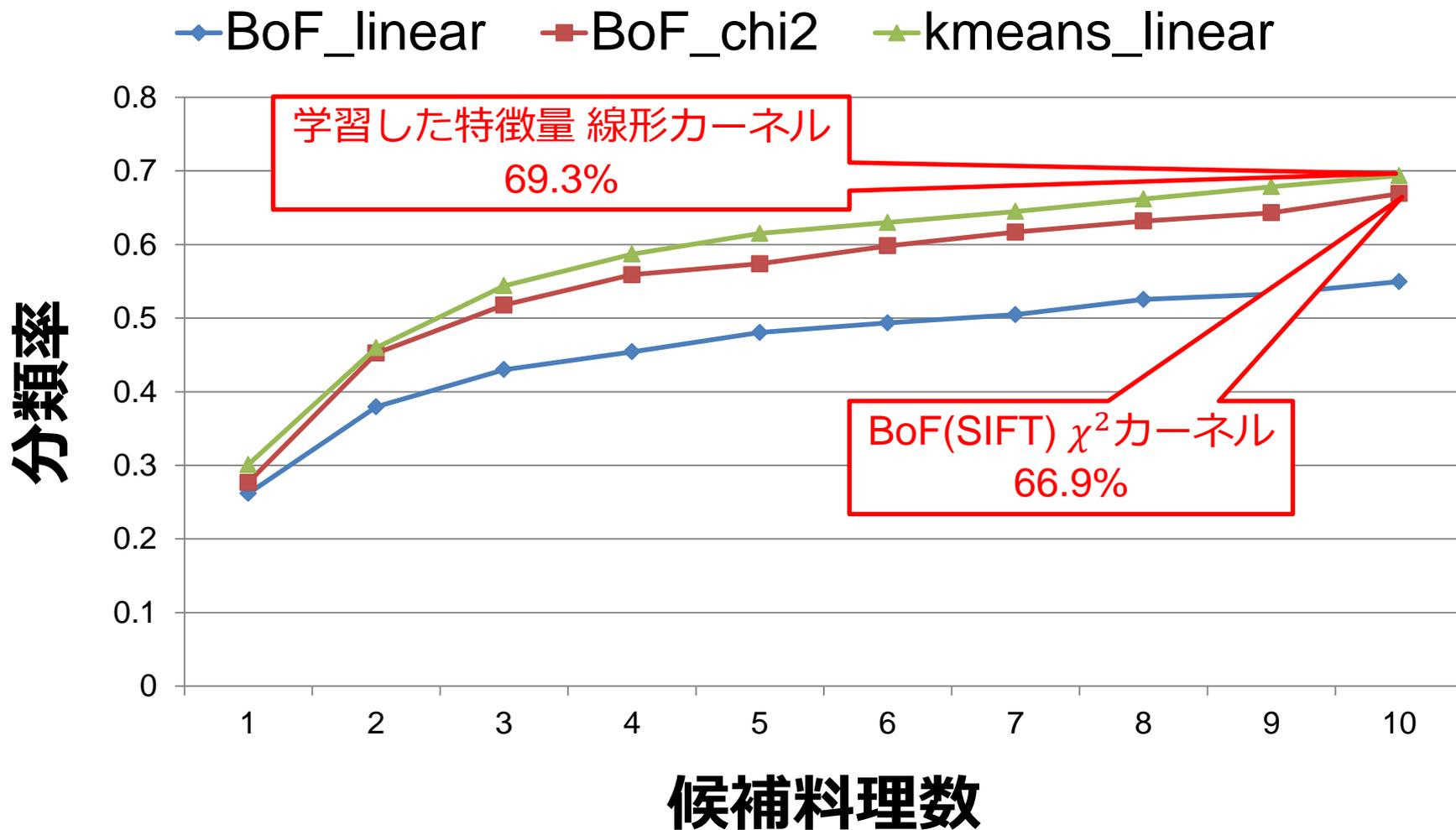
領域を2x2に分割し、それぞれで特徴ベクトルを作成
合計6400次元

MKLによる特徴統合による有効性

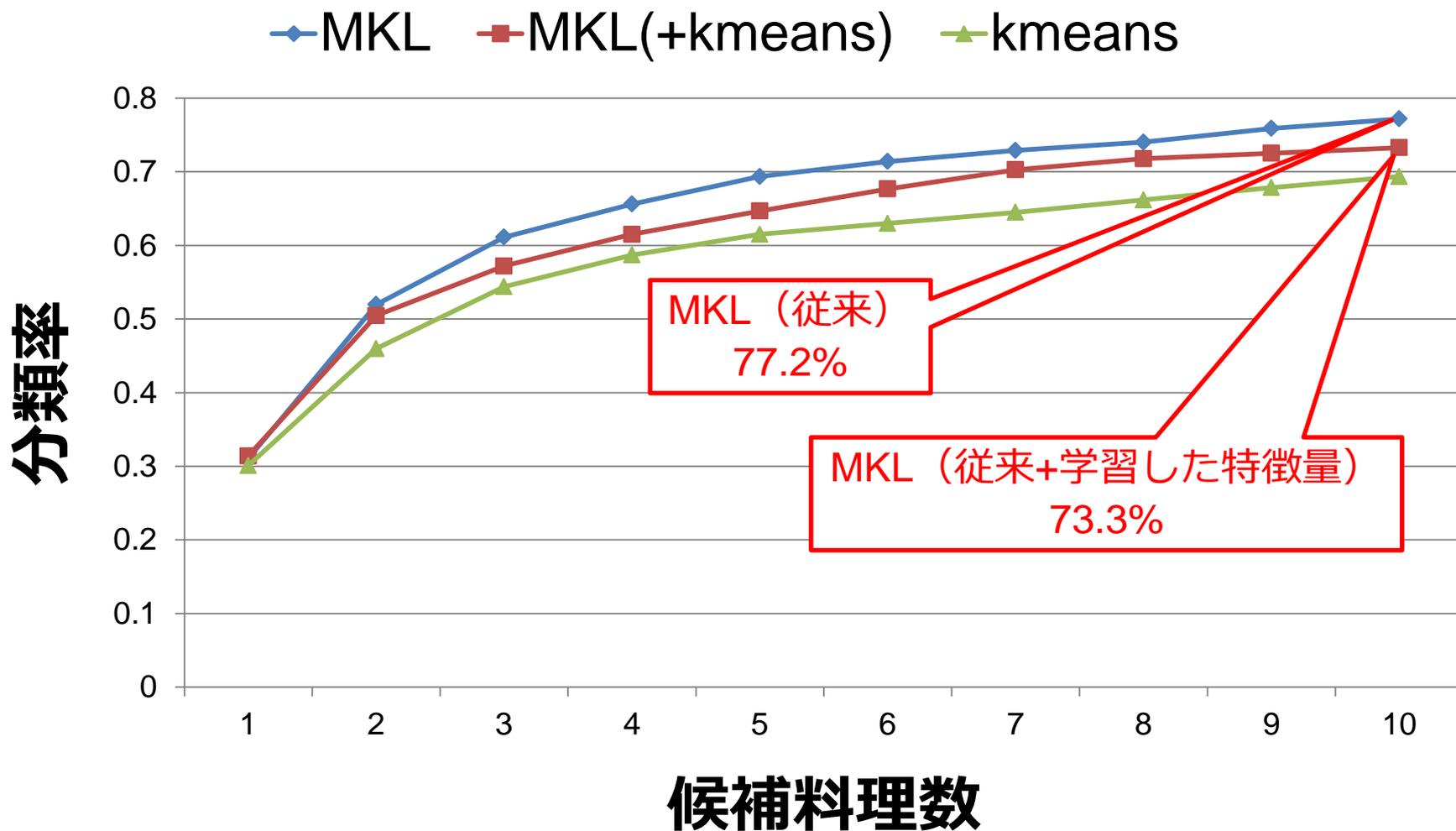
従来の5種類の特徴量+k-means learningによる特徴量

料理領域は既知とする

BoF(SIFT)との比較



MKLによる統合の有効性



皿検出器

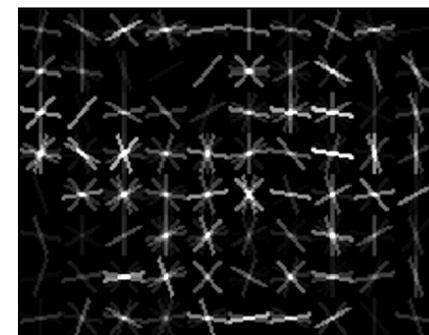
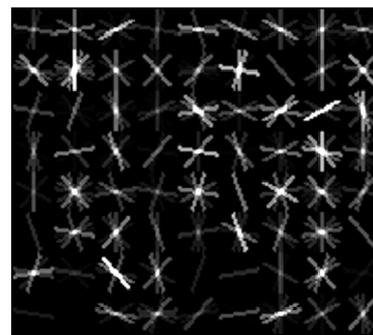
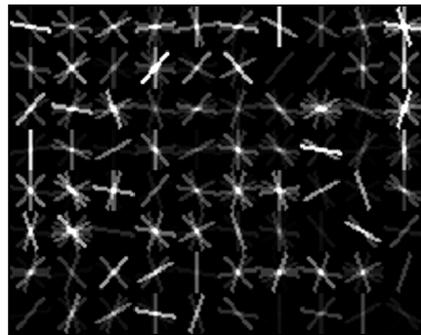
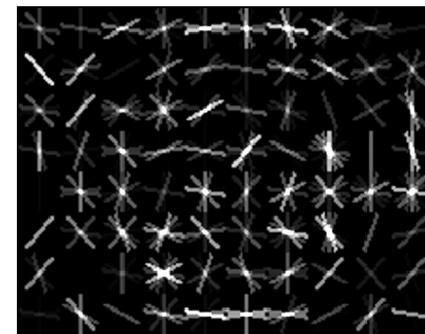
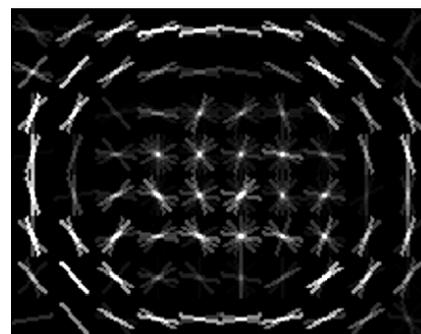
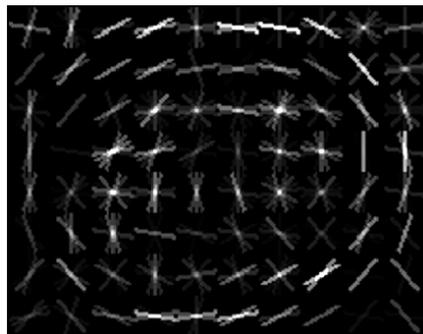
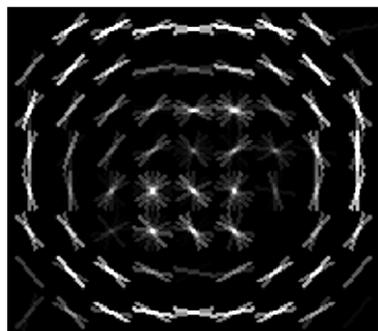
単独の検出手法では料理ごとに学習した場合よりも有効
検出コストは低

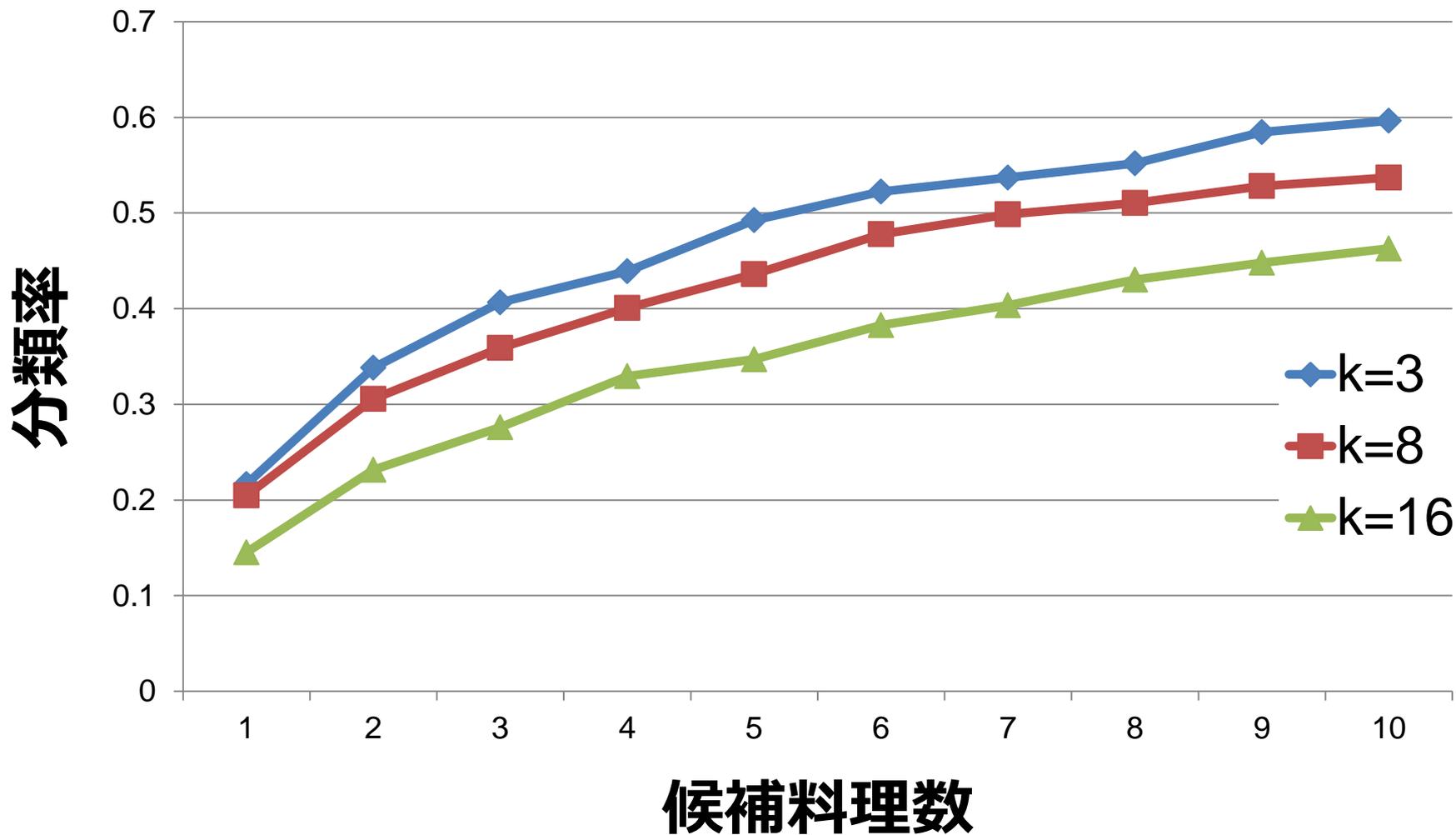
k-meansによる教師なし学習を用いた特徴量

BoF(SIFT)より高性能

従来の特徴量との組み合わせでは有効性は見られなかった
特徴量ごとの得意・不得意の検証が必要

k-meansクラスタリングの場合





Fisher vectorとの比較 (参考)

