

Recognition of Multiple-Food Images by Detecting Candidate Regions

**Yuji Matsuda, Hajime Hoashi, and Keiji Yanai
The University of Electro-Communications, Tokyo
JAPAN**

Background

- Recording of food habits has become popular.
 - Users can become aware of own diet, and evaluate nutrition.
- To record food items in every meal is a quite troublesome task.
- It is desired to make recording of food items more easier and quickly.

Objective

- Recognition of “multiple food images”
 - Contain two or more food items



food image



soup

funny-side up

cutlet

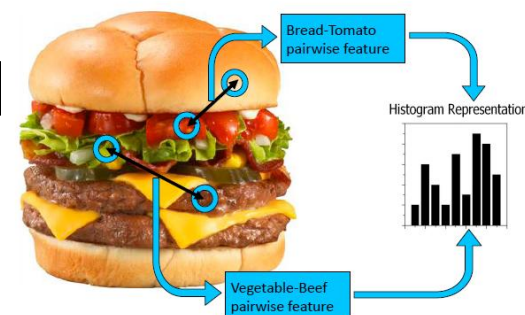
on meuniere



10. Tom bread

Related Work

- [S. Yang et al. CVPR 2010]
 - specialized for American fast food
 - defined 8 basic food materials
 - classify 61 food categories using detected materials and their relative position
- [Z. Zong et al. ISM 2010]
 - the same fast food dataset
 - SIFT detector + LBP



Related Work

- Our previous work [Hoashi et al. ISM 2010]
 - 85 kinds of food categories
 - fusing various kinds of image features using MKL

Existing methods

- Recognize only **single** food-item in one image.



Proposed method

- Recognize **multiple** food-items in one image **at the same time.**

Recognition Flow

Input image

Candidate Region Detection

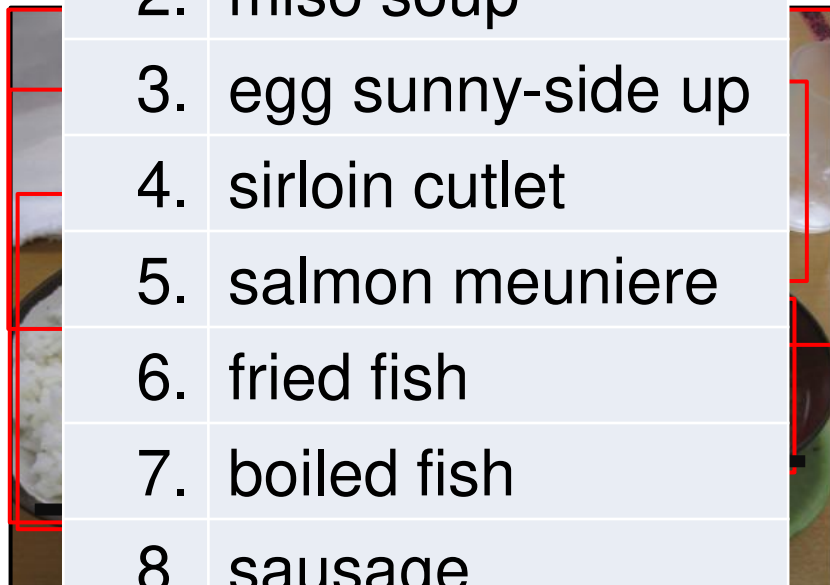
Whole DPM Circle JSEG

Coding Image Feature Vector

Color SIFT CSIFT HoG Gabor

Classification

Output results



1. rice
2. miso soup
3. egg sunny-side up
4. sirloin cutlet
5. salmon meuniere
6. fried fish
7. boiled fish
8. sausage
9. sandwiches
10. roll bread

Candidate Region Detection

- Whole image
- Sliding window search
- Circle detector
- Region segmentation

Whole Image

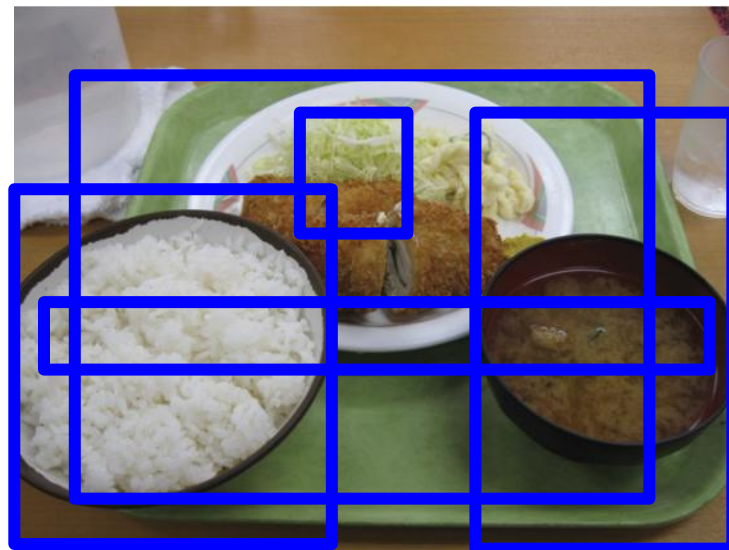
- **Advantage:** suitable for larger dish
- **Disadvantage:** unsuitable for small dish



Sliding Window Search (Deformable Part Model)

- **Advantage**: can obtain region with a high evaluation value
- **Disadvantage**: is based on only gradient-based features

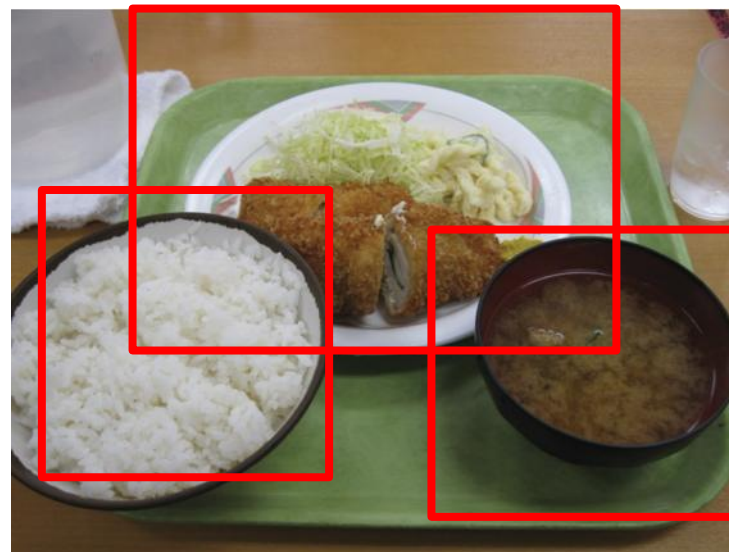
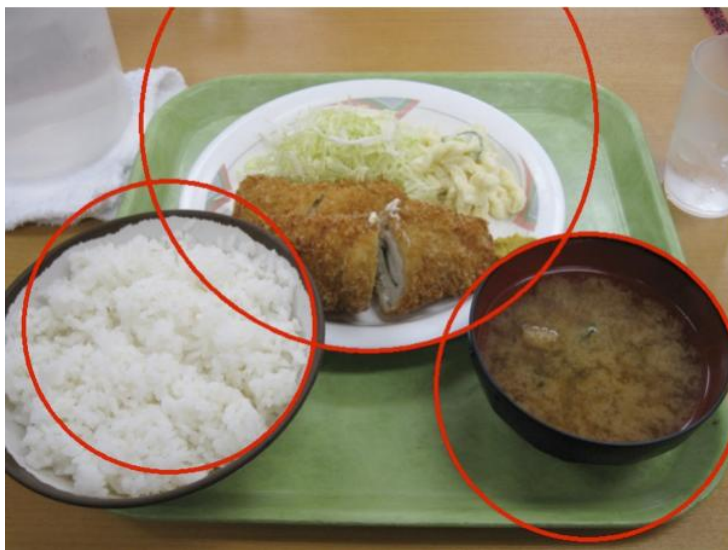
region is 100 in total



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

Circle Detector

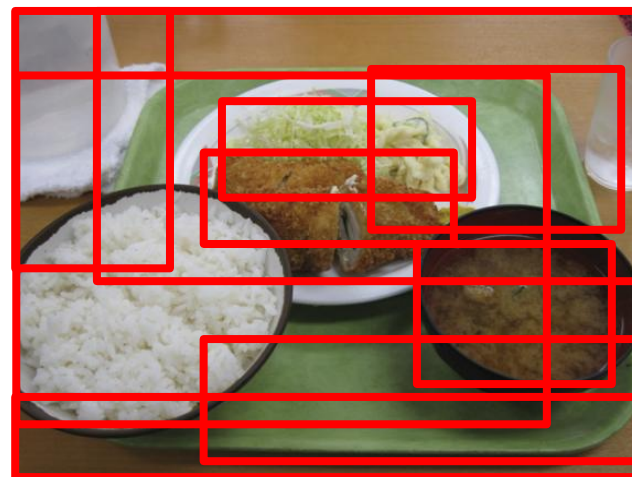
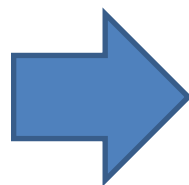
- **Advantage:** can detect food by circular plate
- **Disadvantage:** dishes are not always circular
region is 4 on average



Region Segmentation (JSEG)

- **Advantage**: detect dishes by segmentation
- **Disadvantage**: does not always success

region is 10



Y. Deng and B. S. Manjunath: Unsupervised segmentation of color texture regions in images and video, PAMI 2001

Region Segmentation (cont.)

- One food item is sometimes divided into several regions.
 - combine regions based on *circularity*

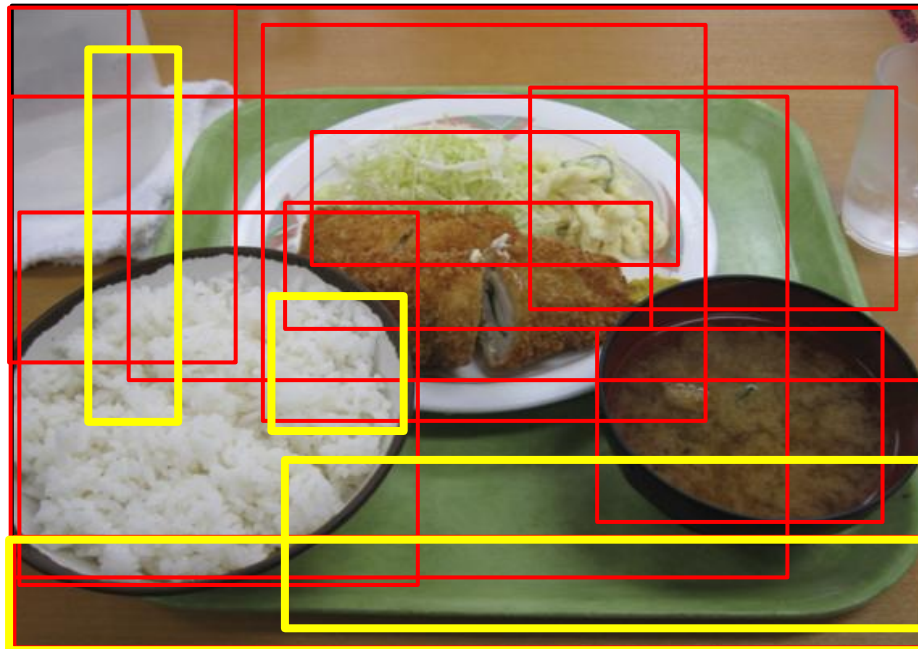


$$Circularity = \frac{4\pi(Area)}{(perimeter)^2}$$

combined region is 4
on average

Irrelevant Region Removal

- Irrelevant region:
 - less than 60 pixels (shorter side)
 - apart from the average aspect ratio



Classification

- Image Features:
 - Color histogram
 - SIFT, CSIFT (Spatial-pyramid bag-of-features)
 - Frequency of the pattern of local patch
 - HOG
 - Rough shape of the object
 - Gabor
 - texture patterns

Classification

- Classifier:
 - Multiple Kernel Learning (MKL-SVM)
 - 1-vs-rest

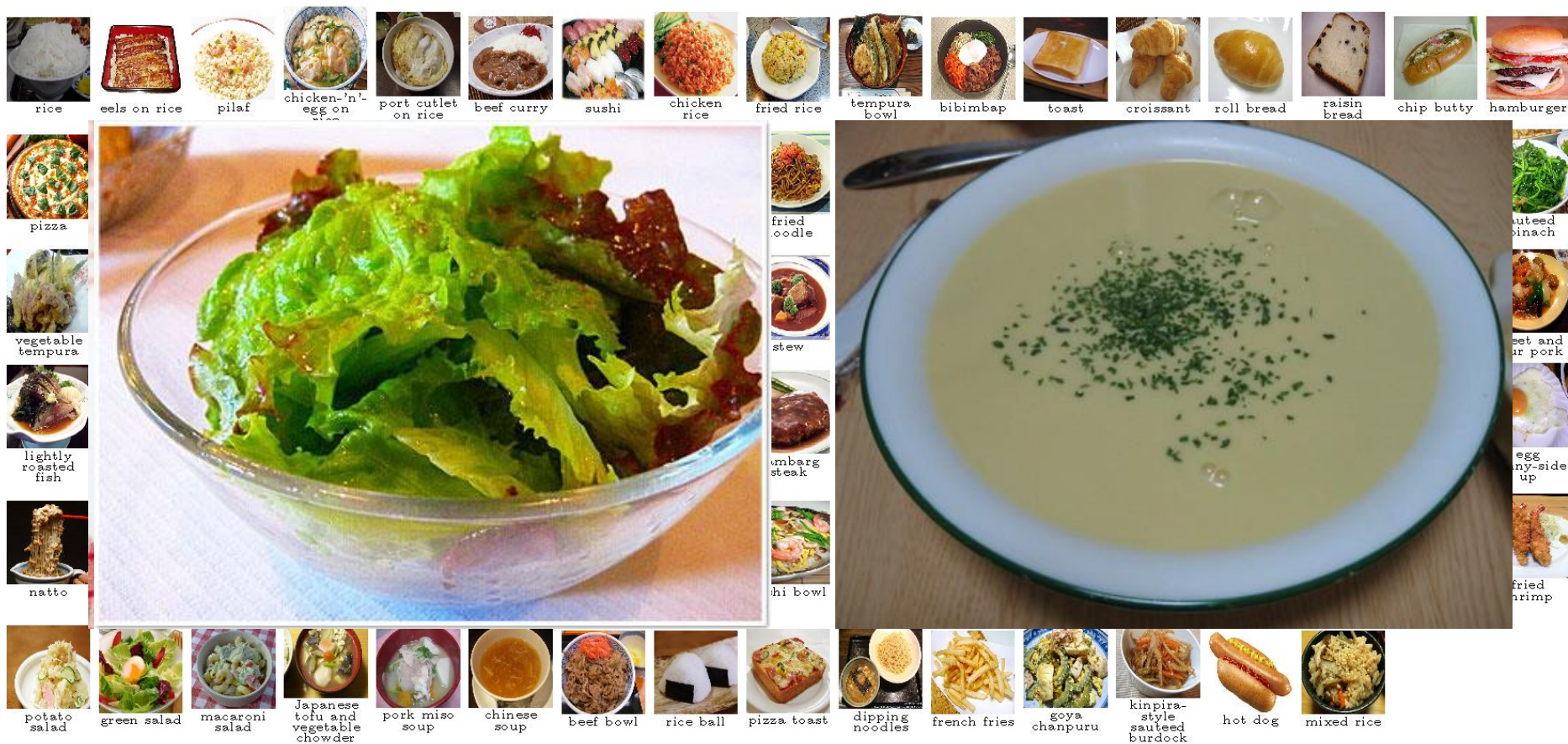
$$K_{MKL}(x, x') = \sum_{j=1}^K \beta_j K_j(x, x')$$

- Results:
 - A list of top N food categories

Experiments

- Dataset:
 - includes 100 kinds of food categories
 - has about 100 images for each category
- For test:
 - 500 multiple food-item images (contain 1200 items)

100 food category database



Experiments

- Performance:

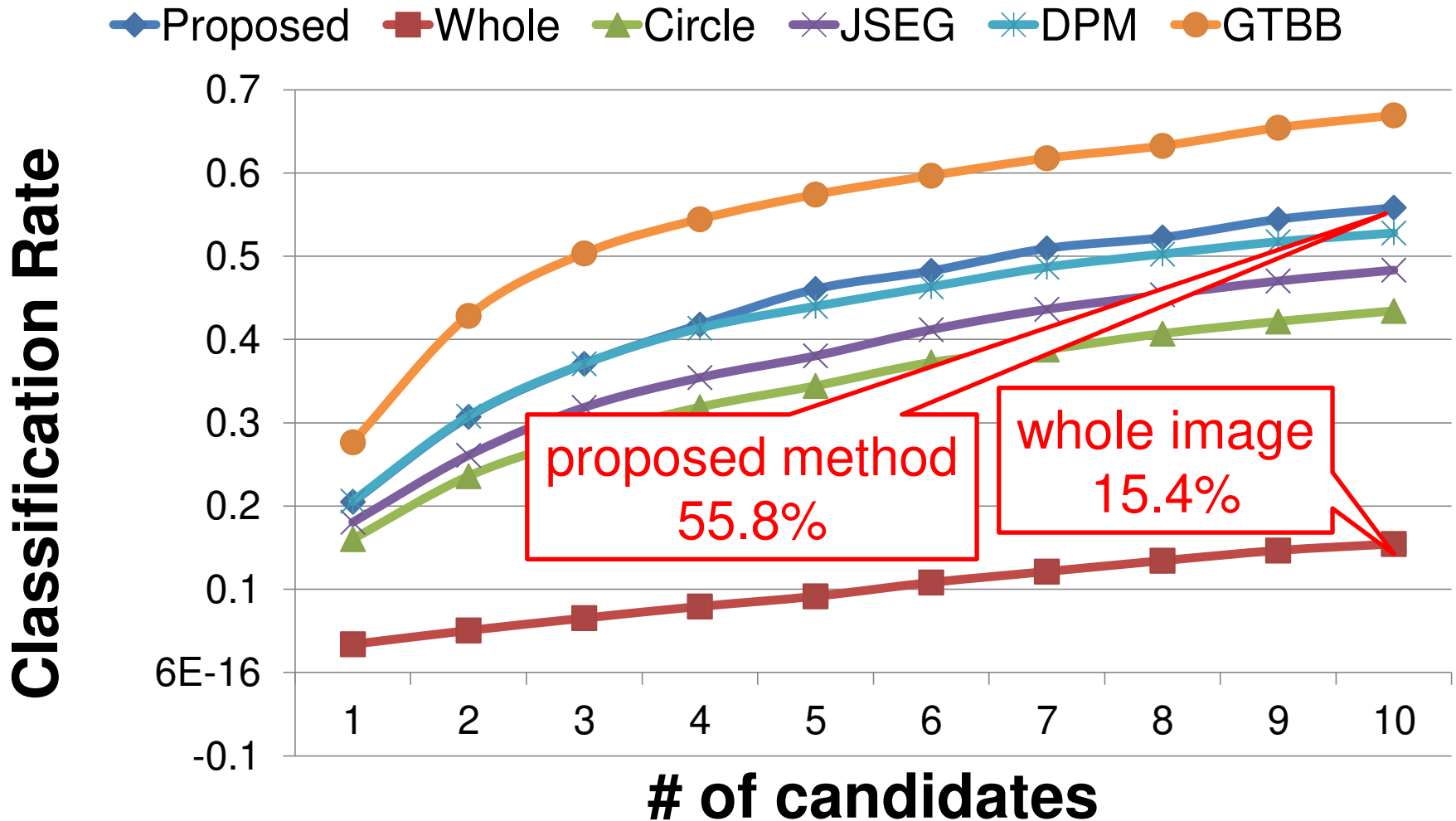
Classification Rate

$$= \frac{\# \text{ of correctly detected food items in } \mathbf{Top\ } N}{\# \text{ of all the food items}}$$

- Comparison:

1. Proposed Method
2. Only single detector (includes previous method)
3. Ground truth bounding-box region

Multiple Food-Item Images



Conclusions

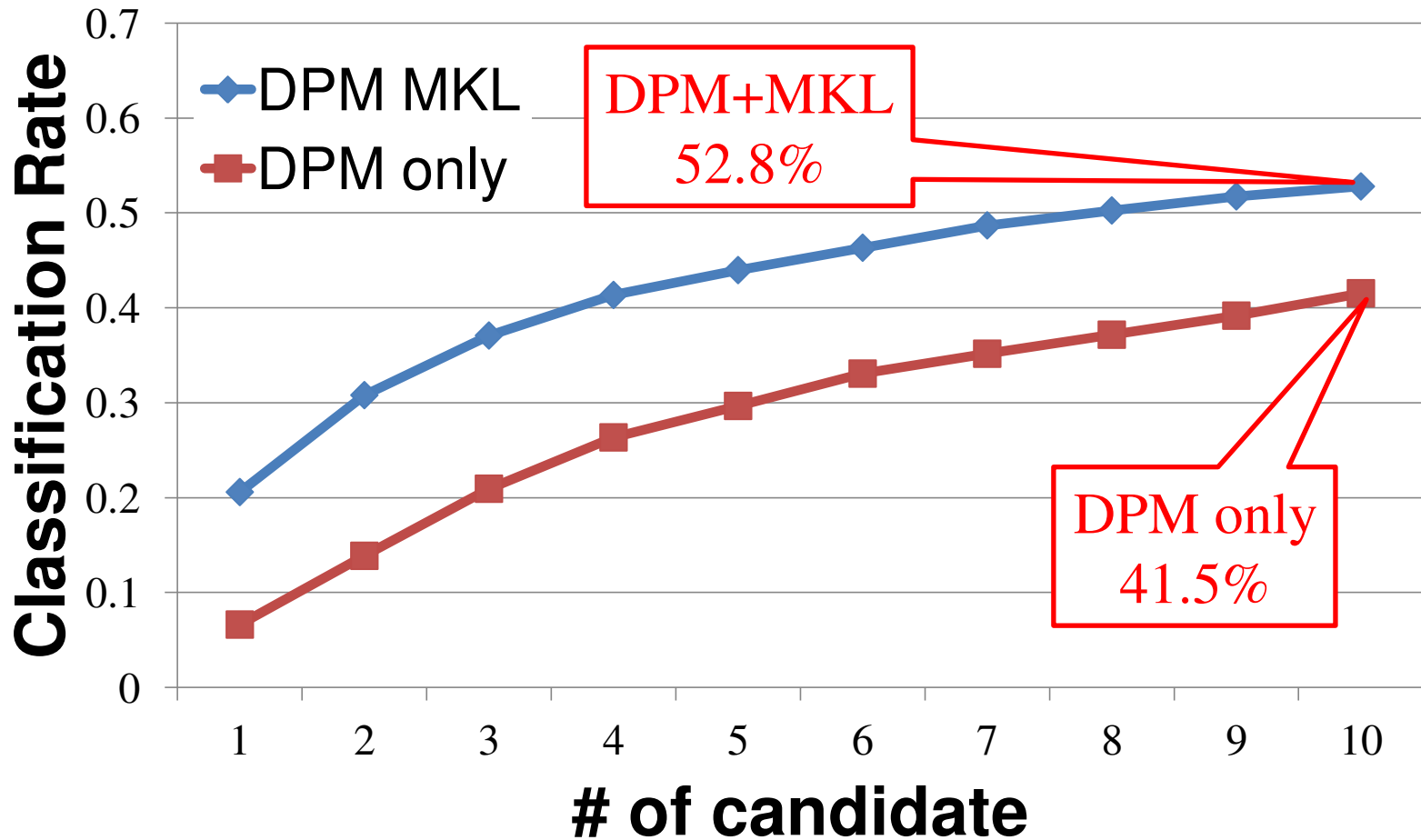
- Proposed two-step food recognition method
 1. candidate region detection
 2. classification for candidate region
- Achieved 55.8% classification rate (top 10)
 - improved by 40.4 points
- Future Work:
 - introduce co-occurrence probability
 - estimate calories



**Thank you !
Questions?**



DPM only vs. DPM+MKL (multiple food-item images)

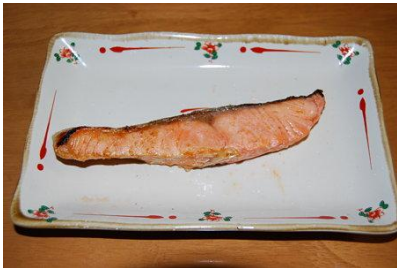


Processing Time

- Total processing time is about 2 minutes.
 - Candidate Region Detection : about 15 sec
 - Feature encoding : about 90 sec
 - Classification : about 20 sec

Easy categories

grilled salmon (100%) sausage (100%)



rice (92.3%)



egg sunny-side up (91.7%)



miso soup (89.2%) hamburger (86.7%) toast (85%) sandwiches (81.8%)



Difficult categories

macaroni salad
(0%)



pork miso soup
(6.3%)



Chinese soup (6.7%)



French fries (10%)



omelet (10%)



Hamburg steak
(11.1%)



jiaozi (11.1%)



udon noodle
(11.1%)

