Image Recognition of 85 Food Categories by Feature Fusion

University of electro-Communications, Tokyo, Japan
Hajime Hoashi, Taichi Joutou and Keiji Yanai

Background & Objective

◆Healthful eating habit is important to avoid obesity and diseases.

◆If there is a food recommendation system, it is work to keep people in good health.

◆A food recognition engine is needed to build a automatic food recommendation system.

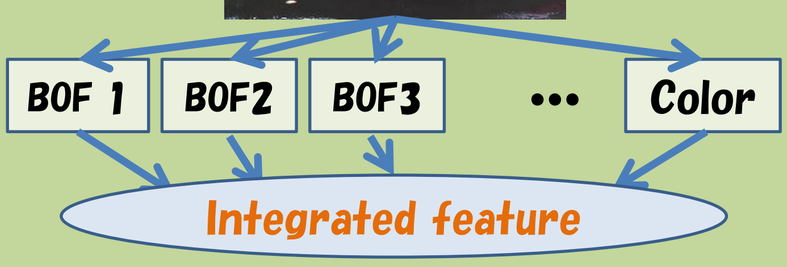
We propose a food recognition engine using MKL in this paper

◆It can recognize many kinds of food with a high classification rate.
- ◆85 food categories

◆Using MKL(Multiple Kernel Learning)
- ◆Achieved the highest performance for Caltech-101/256 [Varma et al. ICCV2007]
- ◆Integrate many kinds of image features.

Method

◆Feature fusion by Multiple Kernel Learning (MKL)
- ◆ Bag-of-Features (BOF) : 6 kinds
- ◆ Gabor Features : 2 kinds
- ◆ Gradient Features : 8 kinds
- ◆ Color histogram : 1 kinds



◆Multiple Kernel Learning
- ◆Is an extension of a SVM .
- ◆Can handle "a combined kernel" which is a linear combination of kernels.
- ◆Can estimate kernel weights and SVM model parameters simultaneously.
- ◆Can integrate features by assigning one feature to one kernel.

Combined kernel

BoF1 kernel..... Gabor1 kernel..... Color kernel

Combined kernel

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

Kernel weight (to be estimated by MKL)

Food Image Dataset

◆We built a food image set.
- ◆Includes 85 kinds of food categories.
  - ◆ They are common in Japan.
- ◆Has 100 images for each category.

Features

◆Bag-of-Features(BOF)

(1) Sample points in three ways (DoG Keypoints, Random sampling, Grid sampling)

(2) Describe local patterns around the sampled points with SIFT [Lowe 2004]

(3) Generate codebooks by k-means (*size of a codebook: K=1000, 2000*)

(4) Convert images into BoF vectors by voting to nearest codewords

DoG Keypoints | Random sampling | Grid sampling

◆Gabor Features

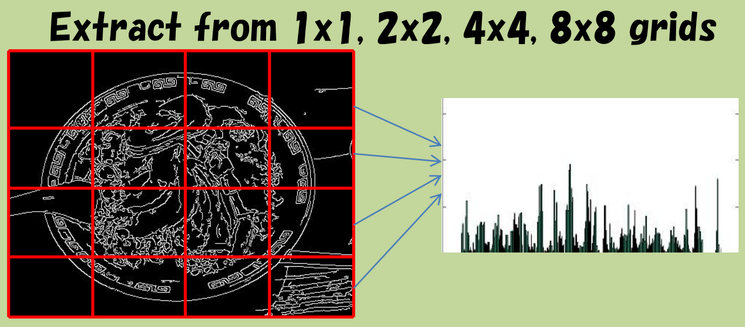
6 directions * 4 frequencies
Extract from 3x3 and 4x4 grids
Totally 216dim. & 384dim.

◆Color histogram

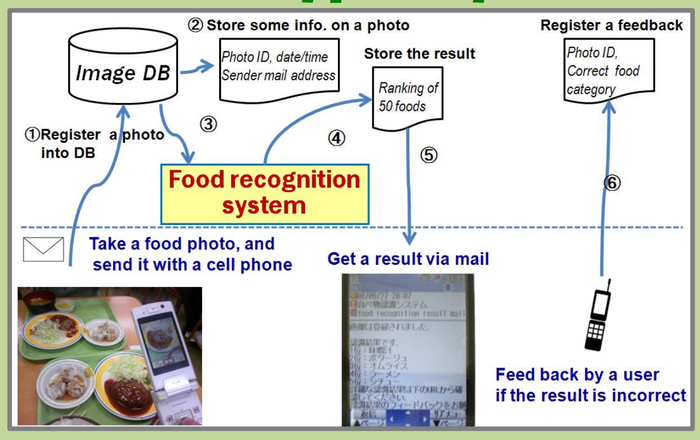
64 colors
Extract from 2x2 grids
Totally 256dim.

◆Gradient Features

(1)Extract edges by Canny edge detector
(2)Calculate gradient on edges
(3)Build histograms by voting gradient orientation
Extract from 1x1, 2x2, 4x4, 8x8 grids



Prototype System



◆◆You can try it !◆

◆Please Send a food photo to food@mm.cs.uec.ac.jp

◆ And you will get a recognition results in a Japanese mail.

◆ (Sorry !! But you can see a result with photos by clicking a URL in the mail !)

Experiments

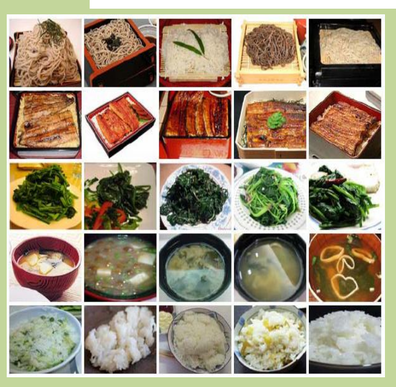
◆Classification rate for 85 categories
- ◆Classification by one-vs-rest

Table I
RESULTS FROM SINGLE FEATURES AND FUSION BY MKL

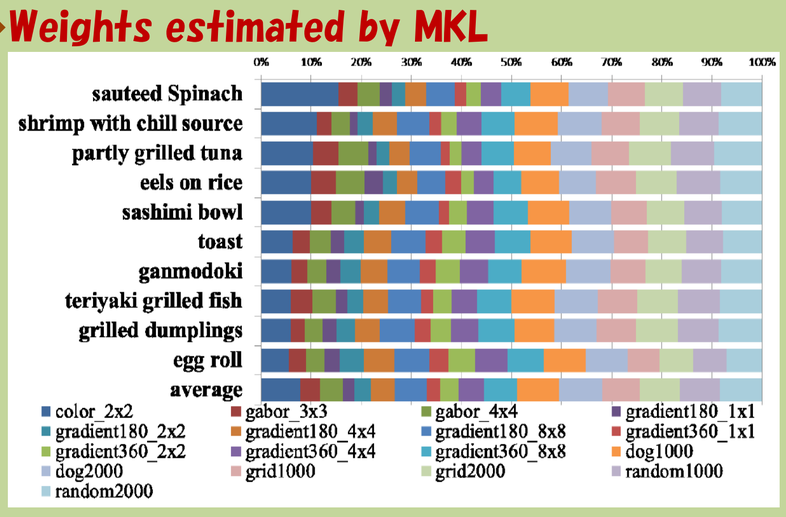
image features	classification rate
BoF (dog1000)	33.47%
BoF (dog2000)	33.42%
BoF (grid1000)	30.73%
BoF (grid2000)	32.21%
BoF (random1000)	29.61%
BoF (random2000)	30.36%
Color	27.08%
Gabor (3 × 3)	23.60%
Gabor (4 × 4)	25.35%
Gradient (180, 1 × 1)	3.87%
Gradient (180, 2 × 2)	10.12%
Gradient (180, 4 × 4)	17.04%
Gradient (180, 8 × 8)	19.44%
Gradient (360, 1 × 1)	5.67%
Gradient (360, 2 × 2)	13.15%
Gradient (360, 4 × 4)	20.87%
Gradient (360, 8 × 8)	21.84%
SVM (uniform)	60.87%
MKL (mean-χ^2distance)	62.52%

◆The best five and the worst five

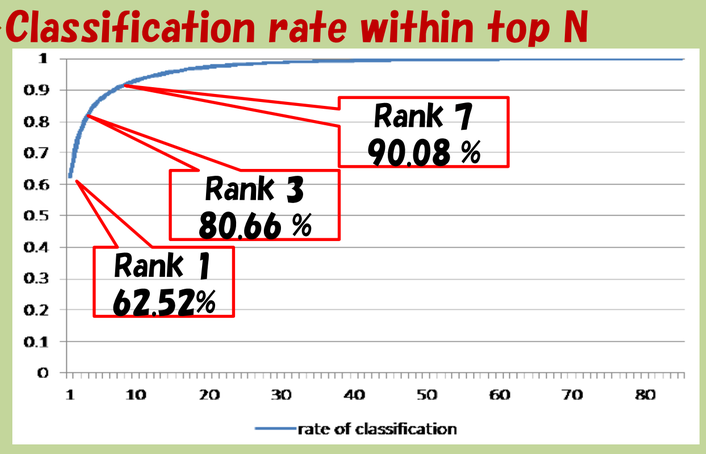
top 5	category	recall	worst 5	category	recall
1	soba noodle	95%	1	Ganmodoki	17%
2	eels on rice	94%	2	sandwich	24%
3	sauteed spinach	93%	3	toast	30%
4	miso soup	92%	4	grilled eggplant	30%
5	rice	90%	5	simmered pork	31%



◆Weights estimated by MKL



◆Classification rate within top N



Conclusions

◆Propose a food recognition engine with MKL-based future fusion
- ◆Achieved 62.52% classification rate
- ◆More than 80.66% when allowing three candidates

◆Future work
- ◆More than 100 categories
- ◆More features (e.g. shape context, CSIFT)
- ◆Other features (e.g. date/time, GPS info.)
- ◆Implement a food recommendation system
- ◆Multi-object recognition