SA.PC.1a

A Food Image Recognition System with Multiple Kernel Learning




University of Electro-Communications, Tokyo, JAPAN
Taichi Joutou and Keiji Yanai

Background & Objective

◆ Health care on foods is important to avoid obesity and diseases.
◆ Object recognition techniques is growing greatly these days.

An automatic food advisory system become possible now !!

A food recognition engine for a food advisory system is needed .

[requirements for a food recognition engine]
◆ Can recognize many kinds of food (>50) with a high classification rate.
◆ Can output a result within 30 seconds.

Multiple Kernel Learning (MKL)

◆ Achieved the highest performance for Caltech-101/256 (89.6% / 60.6%) [Varma et al. ICCV2007]
◆ Integrate many kinds of image features.

We propose a food recognition engine using MKL in this paper.

Method

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

BOF 1 | BOF2 | BOF3 | … | Color

M K L

◆ Multiple Kernel Learning
- Is an extenstion 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 histogram kernel

Combined Kernel

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

β_k: Kernel weight (to be estimated by MKL)

50 food image set

◆ We built a Japanese food image set.
- Includes 50 kinds of food categories.
- Has 100 images for each category.

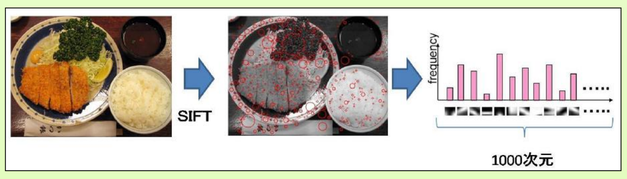
Features (9 kinds)

◆ Bag-of-features (BoF) (local pattern)

(1) Sample points in three ways

DoG Keypoints | Random sampling | Grid sampling

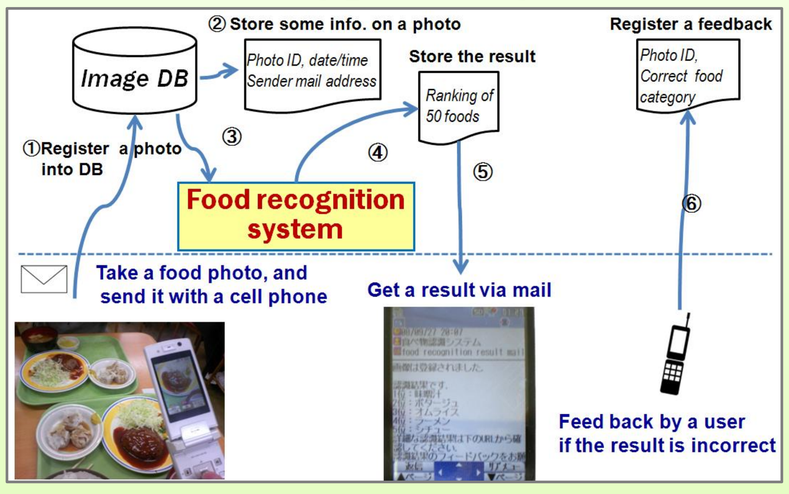
(2) Describe local patterns around the sampled points with SIFT [Lowe 2004]
(3) Generate codebooks by k-means
(size of a codebook: K=1000, 2000)
(4) Convert images into BoF vectors by voting to nearest codewords



- Totally generate six kinds of BoF vectors ([1000,2000]*[DoG, rand, grid])

◆ Color histograms (color) 64dim.
◆ Gabor features (texture)
- 6 directions * 4 frequencies
- Extract from 3x3 and 4x4 grids
- Totally 216dim. & 384dim.

Prototype System



◆You can try it !◆

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

◆ 50 category classification by one-vs-rest (5-fold cv)
- Classification rate for 50 categories.

Table 1. Results from single features and fusion by MKL

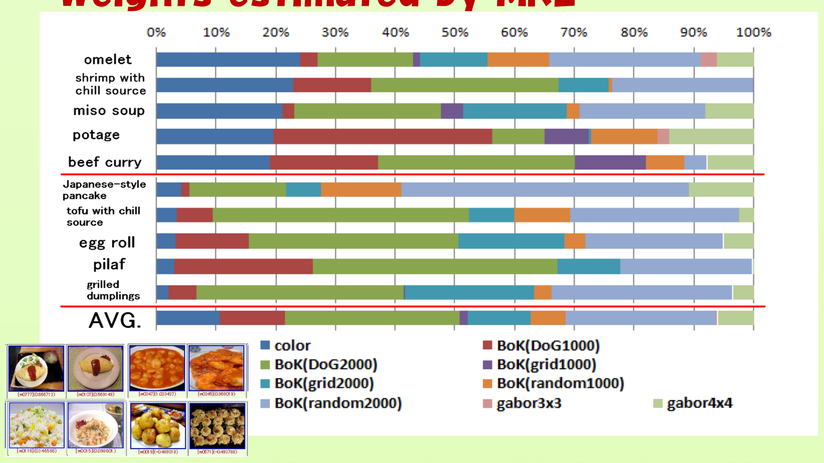
image features	classification rate
color	38.18%
BoF (dog1000)	26.52%
BoF (dog2000)	27.48%
BoF (grid1000)	26.10%
BoF (grid2000)	27.68%
BoF (random1000)	28.42%
BoF (random2000)	29.70%
Gabor3x3	31.28%
Gabor4x4	34.64%
MKL (after fusion)	61.34%

- The best five and the worst five

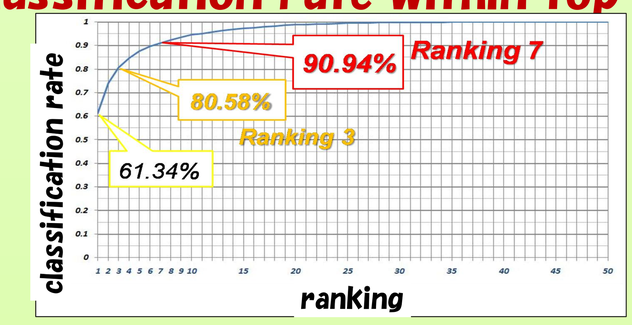
Table 2. The best five and worst five categories in the recall rate of the results by MKL.

top 5	category	recall	worst 5	category	recall
1	miso soup	97%	1	simmered pork	18%
2	soba noodle	94%	2	ginger pork saute	28%
2	eels on rice	94%	3	toast	31%
4	potage	91%	4	pilaf	39%
5	omelet with fried rice	87%	4	egg roll	39%

- Weights estimated by MKL



- Classification rate within top N



◆Built a system for cell phones and run it for one year
-40.09%, 60.00% (within 3rd candidates)

Some ill-conditioned cellular photos

out of focus, after eating half, taken partly

Conclusions

◆Propose a food recognition engine with MKL-based future fusion
- Achieved 61.34% classification rate
- 80.05% when allowing three candidates

◆Future work
- More than 100 categories
- More features (e.g. shape context, PHoG)
- Other features (e.g. date/time, GPS info.)
- Implement a food advisory system