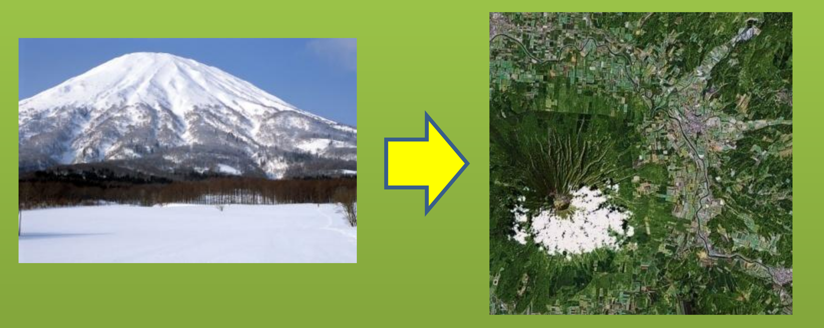
WeBCT 9.37

Geotagged Photo Recognition using Corresponding Aerial Photos with Multiple Kernel Learning



The University of Electro-Communications, Tokyo, JAPAN Keita Yaegashi and Keiji Yanai

Background & Objective

◆ Geotagged Photos are easy to obtain.

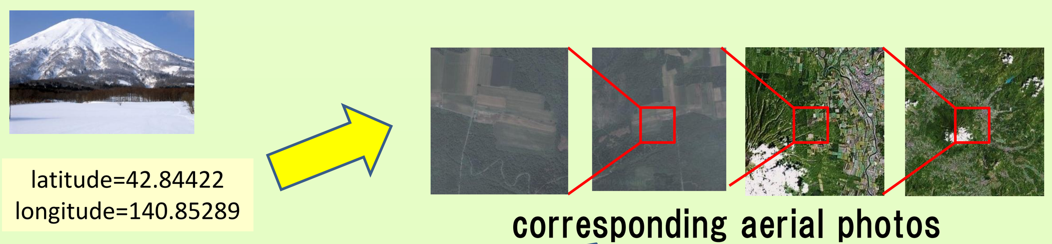
geotag = (latitude, longitude)

Can a geotag help image recognition ?

No ! when using them as 2D vectors.

Yes ! It can help image recognition by converting it into aerial photos.

shown by the following two papers.



corresponding aerial photos

Represent geographical context

[4] J. Luo, J. Yu, D. Joshi, and W. Hao, Event recognition: Viewing the world with a third eye, In *Proc. of ACM International Conference Multimedia, 2008*.

Only events such as "baseball" and "on beach"

[7] K. Yaegashi and K. Yanai, Can geotags help image recognition ? In *Proc. of Pacific-Rim Symposium on Image and Video Technology, 2009*.

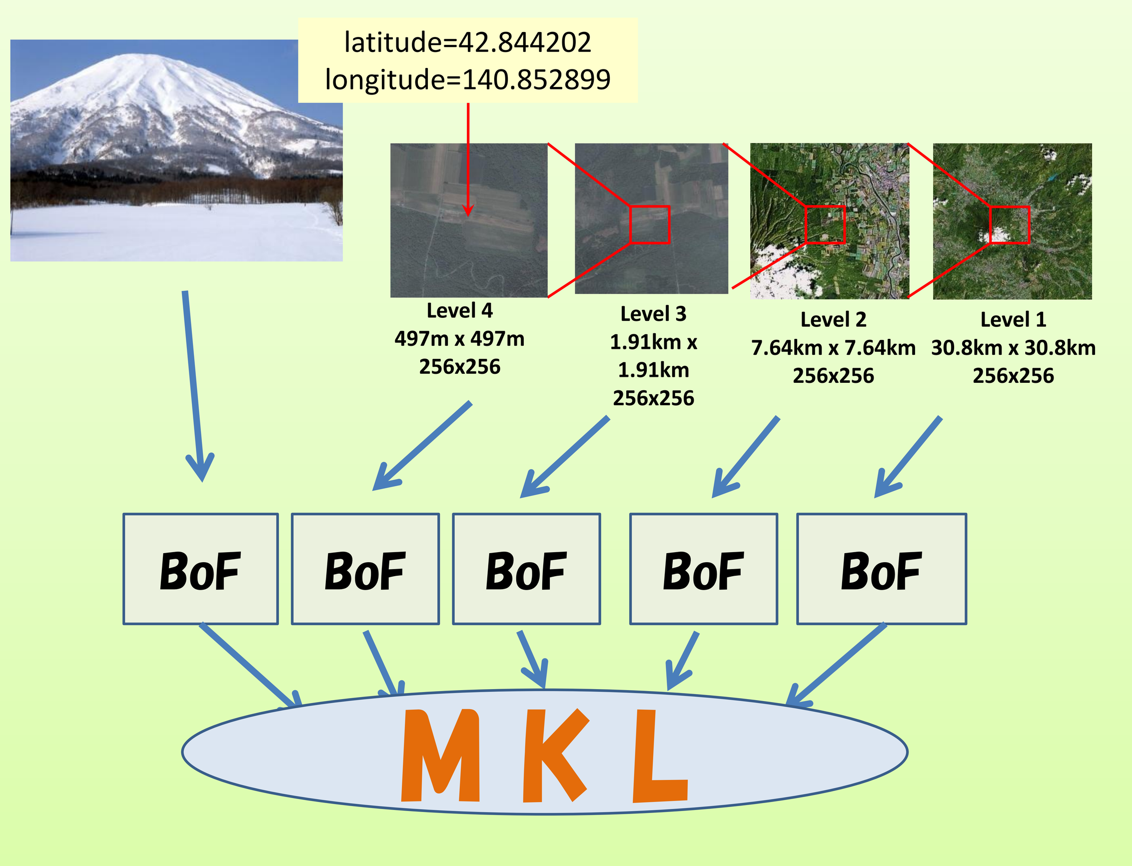
10 categories.
Fusion by concatenating both feature vectors.

◆ **Questions ?**

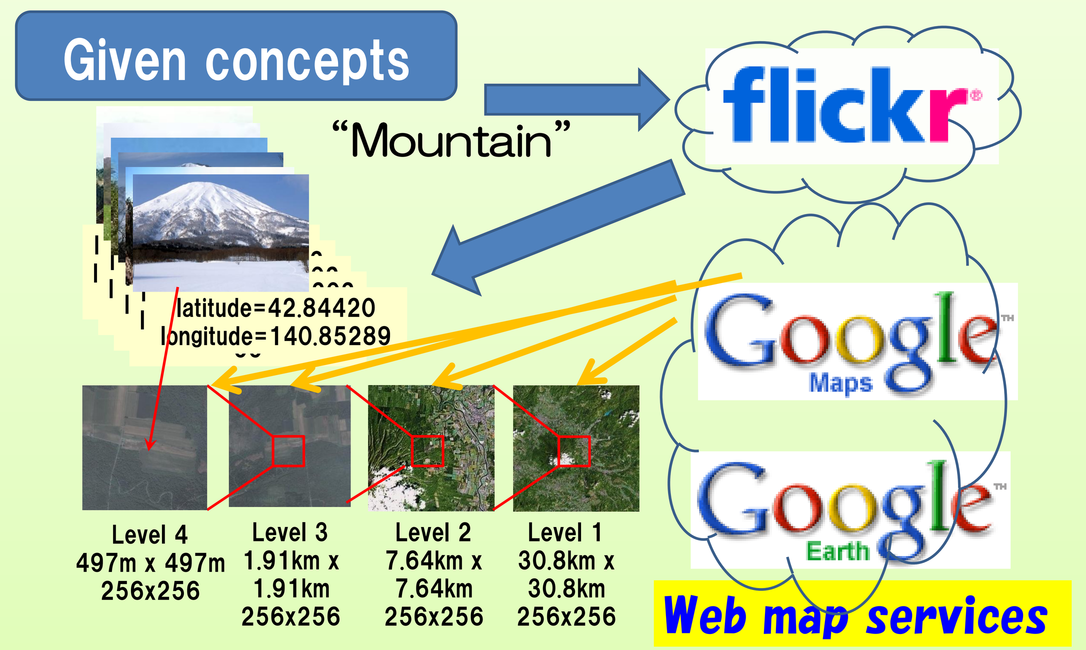
1) To what extent can geotags help ?
2) What kinds of categories are geotags effective for ?

We evaluate the contribution of aerial photos for image recognition by using Multiple Kernel Learning (MKL).

Method

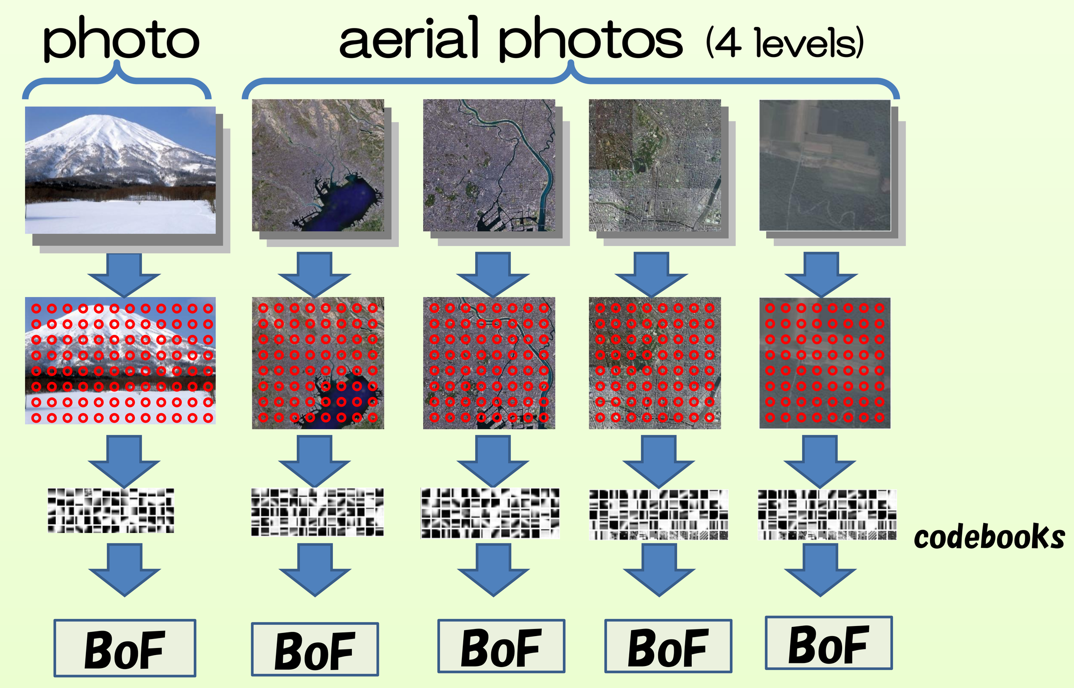


Data Collection



◆ **Gather 4 levels of aerial photos for each geotagged photo.**

Features



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

(1) Sample points by grid sampling (every 10px)

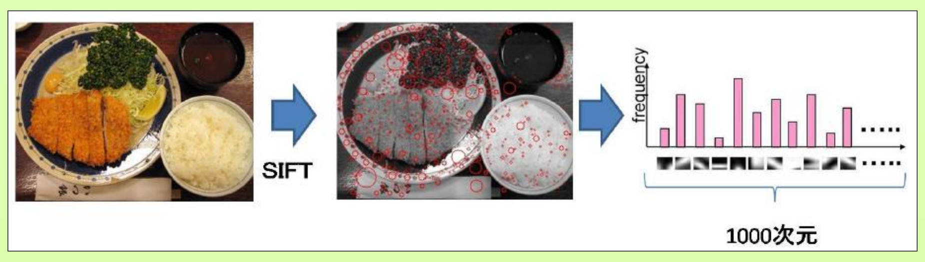
Grid sampling

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

(3) Generate codebooks by k-means

(size of a codebook: 1000)

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



Five BoF vectors for each image

Learning & Classifying

◆ **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

photo BoF kernel | aerial photo(Lv1) ker. ········ Aerial photo(Lv4) ker.

Combined kernel

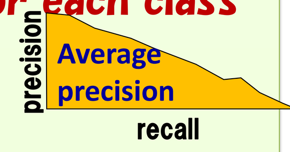
$$\mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{K} \beta_k \mathbf{k}_k(\mathbf{x}_i, \mathbf{x}_j)$$

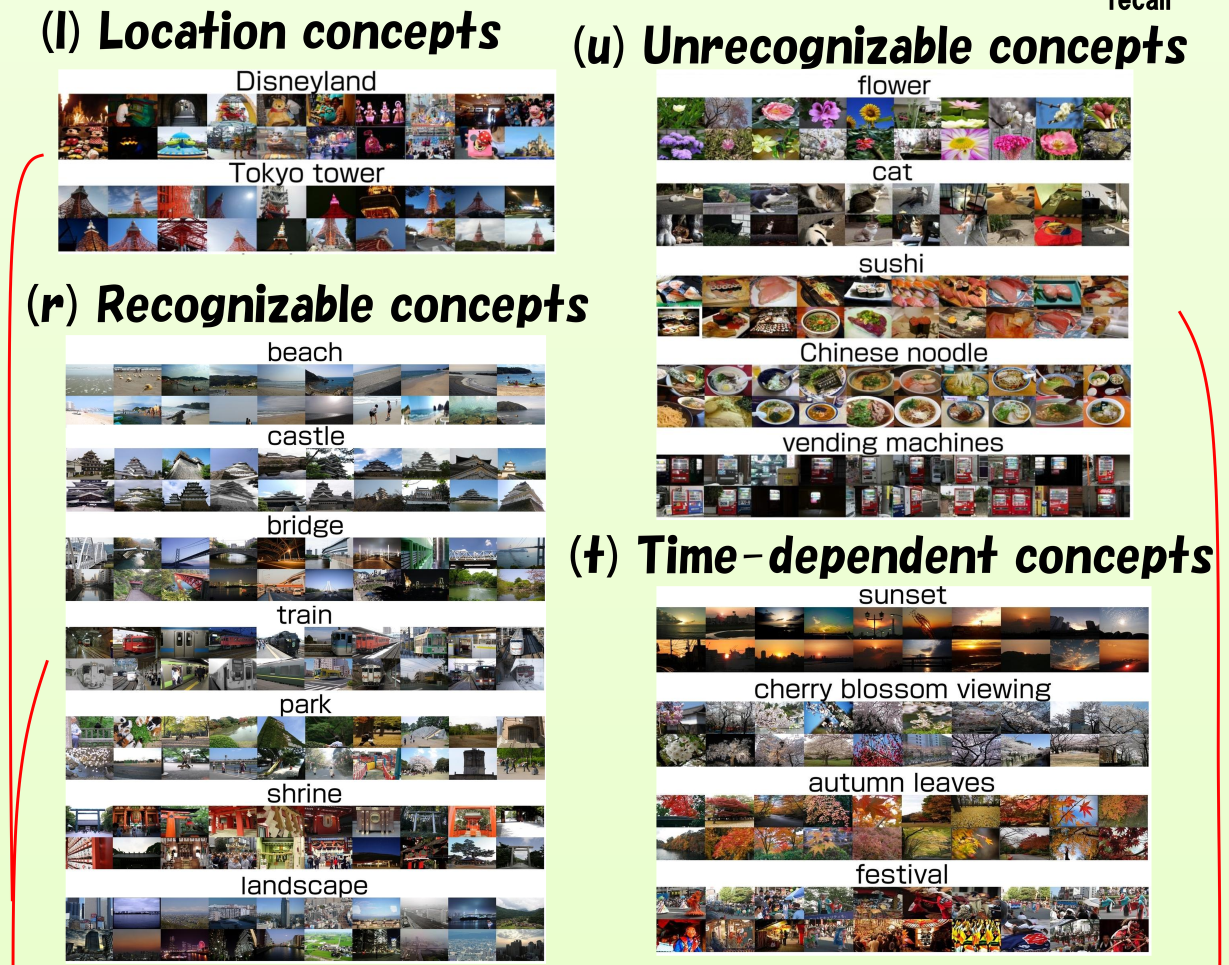
Kernel weight (to be estimated by MKL)

Experiments

◆ **1/0 classification for 18 classes by 5-fold cross validation**

- 200 positive & 200 negative images for each class
- Evaluated by average precision





- **Classification results (AP, %)**

(type) concept	photo	MKL	gain
(l) Disneyland	68.00	**84.06**	+16.06
(r) park	67.43	**76.04**	+8.61
(r) shrine	72.79	**78.53**	+5.74
(t) festival	72.32	**77.75**	+5.43
(r) bridge	69.51	**74.89**	+5.38
(r) landscape	73.71	**78.37**	+4.66
(r) beach	80.10	**83.85**	+3.75
(t) red leaves	79.18	**82.45**	+3.27
(l) Tokyo tower	80.84	**83.84**	+3.00
(r) castle	81.28	**83.53**	+2.23
(u) sushi	80.11	**81.93**	+1.82
(r) railroad	74.70	**76.20**	+1.50
(u) flower	77.00	**78.48**	+1.48
(t) cherry blossom	80.94	**81.61**	+0.67
(u) ramen noodle	82.34	**82.70**	+0.36
(u) cat	73.98	**74.26**	+0.28
(u) vendor machine	83.17	**83.43**	+0.26
(t) sunset	83.01	**83.11**	+0.10
AVERAGE	76.69	**80.28**	+3.59

larger gain · gain · smaller gain

- **Contribution weights by MKL**

Photos · contribution · Aerial photos

(type) concept	photo	level1	level2	level3	level4
(u) ramen noodle	**0.873**	0.002	0.000	0.037	0.088
(u) vendor machine	**0.794**	0.058	0.009	0.074	0.065
(t) cherry blossom	**0.774**	0.038	0.006	0.093	0.090
(u) cat	**0.743**	0.028	0.008	0.063	0.158
(t) sunset	**0.729**	0.055	0.058	0.016	0.142
(u) flower	**0.658**	0.000	0.042	0.051	0.249
(r) railroad	**0.604**	0.106	0.014	0.052	0.224
(r) landscape	**0.604**	0.078	0.024	0.093	0.199
(u) sushi	**0.596**	0.062	0.015	0.062	0.266

(type) concept	photo	level1	level2	level3	level4
(r) bridge	**0.582**	0.077	0.044	0.070	0.226
(t) red leaves	**0.523**	0.141	0.006	0.062	0.269
(r) castle	**0.523**	0.166	0.004	0.099	0.208
(t) festival	**0.518**	0.058	0.001	0.185	0.238
(r) shrine	**0.507**	0.033	0.009	0.061	0.391
(r) park	**0.437**	0.073	0.012	0.045	0.433
(r) beach	**0.392**	0.115	0.173	0.055	0.265
(l) Disneyland	**0.384**	0.095	0.236	0.131	0.153
(l) Tokyo tower	0.364	0.008	0.002	**0.396**	0.231

Conclusions

◆ **Analyzed contribution ratios of aerial images for image recognition using Multiple Kernel Learning (MKL)**

- AP was improved by 3.59 % on average
- Much help to "location concepts" and "recognizable concepts"
- Less help to other kinds of concepts
- Detailed aerial photos are more helpful

◆ **Future work**

- More categories
- More features (e.g. color, HoG, Gabor)
- Other geo-info (e.g. geo-text: country name, area name, GIS-info: population…)