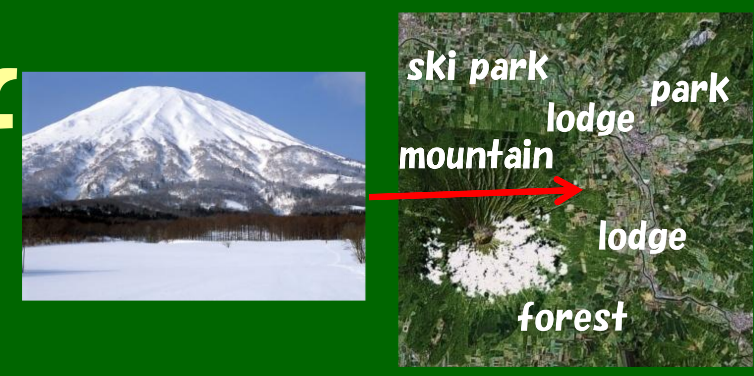
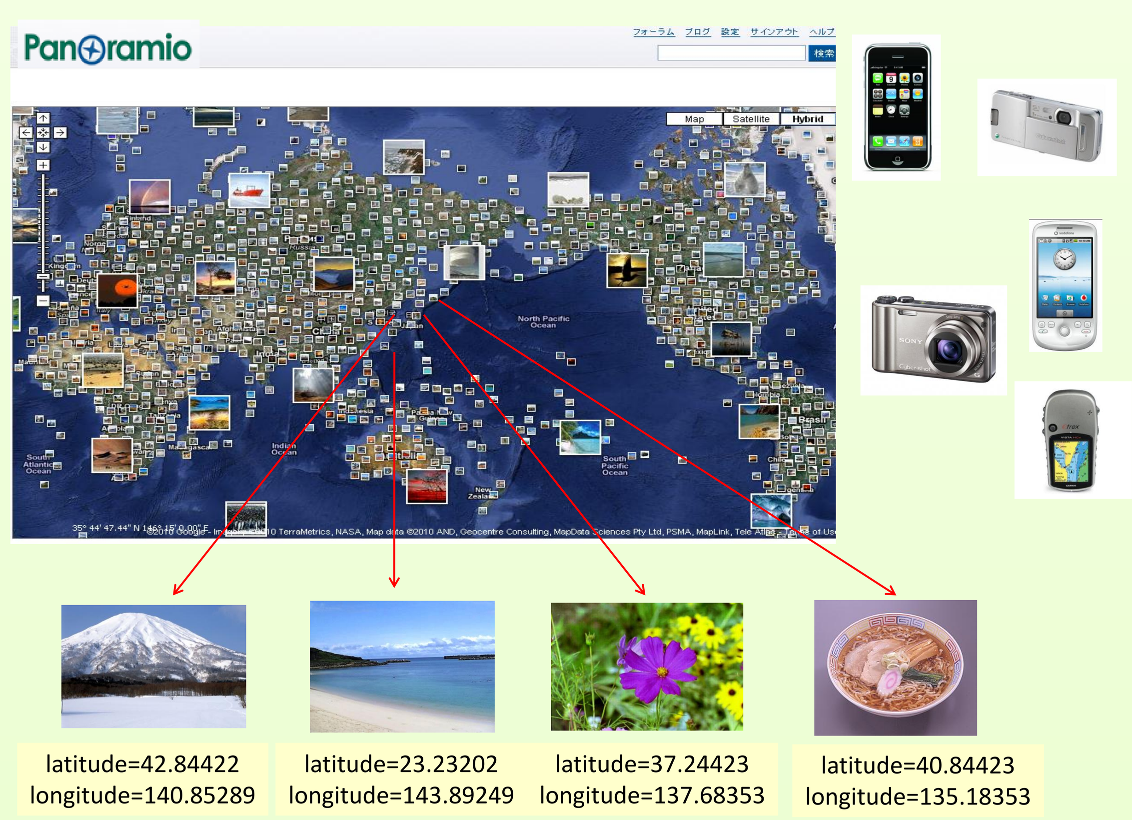
Geotagged Image Recognition by Combining Three Different Kinds of Geolocation Features



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

Background & Objective

◆ Geotagged Photos are easy to obtain.

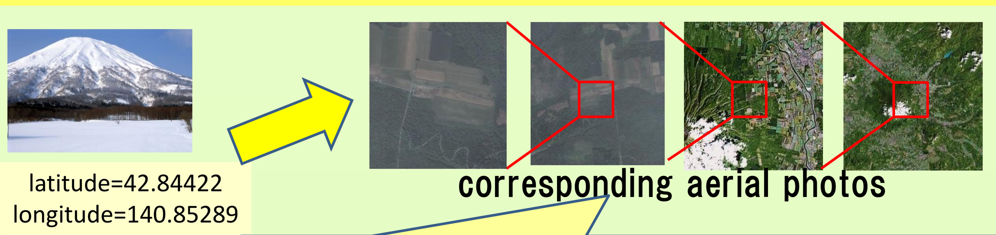


geotag = (latitude, longitude)

Can a geotag help image recognition ?

No ! when using them only as 2D vectors.

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



Represent geographical context

[1] 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.

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

Yes ! It can also help image recognition by reverse geo-coding.

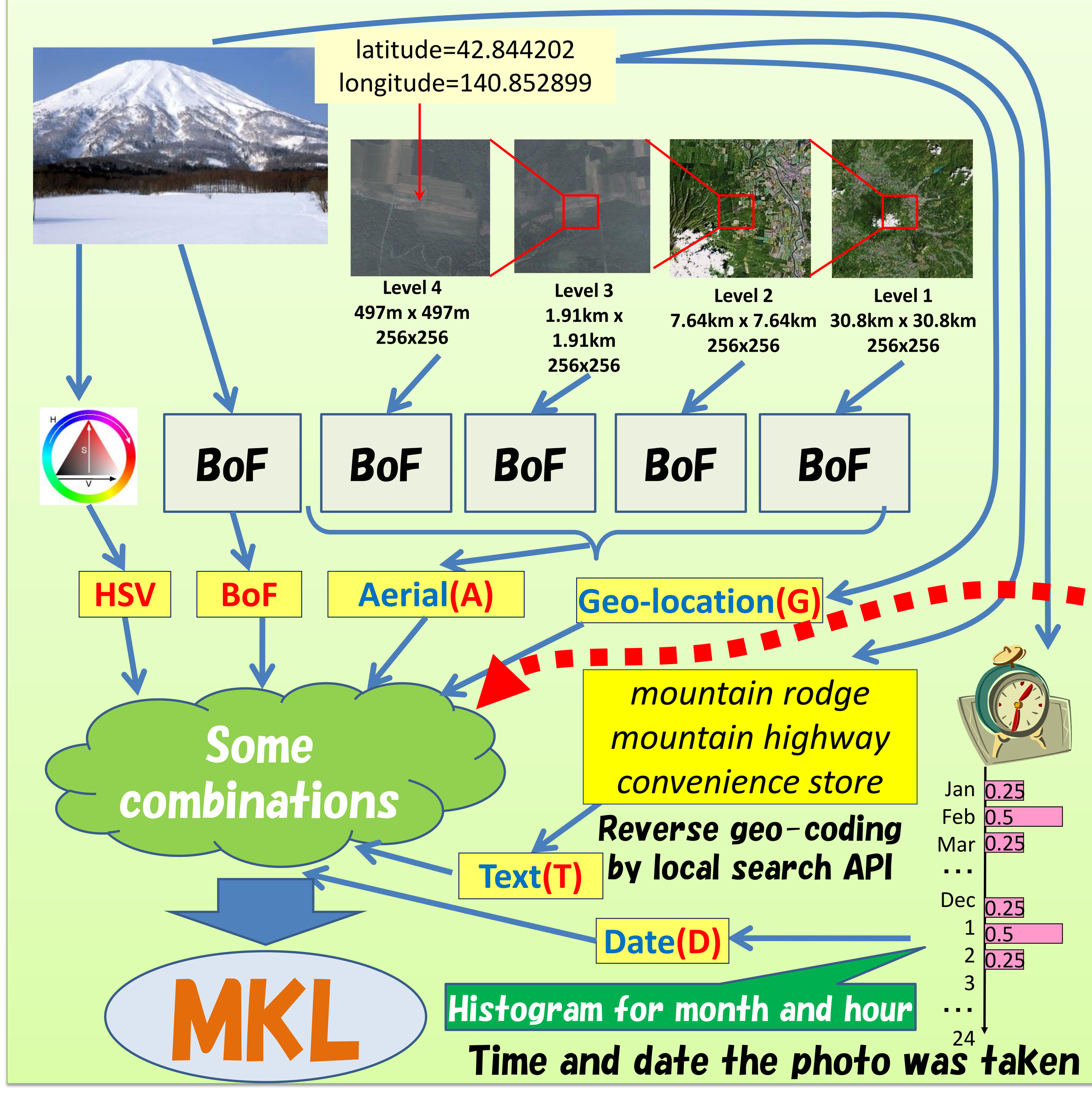
[2] Joshi, D., Luo, J.: In: Proc. of Inferring generic activities and events from image content and bags of geo-tags ACM International Conference on Image and Video Retrieval.

◆Question ?

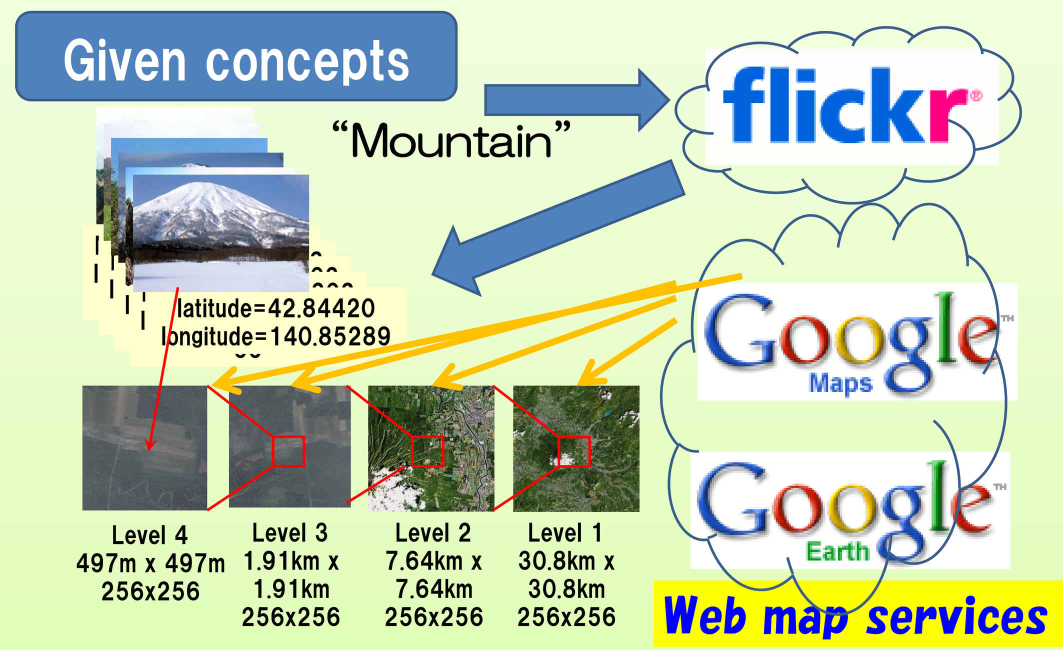
1) To what extent can geotags help ?
2) What kinds of categories are geotags effective for ?

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

Method



Data Collection

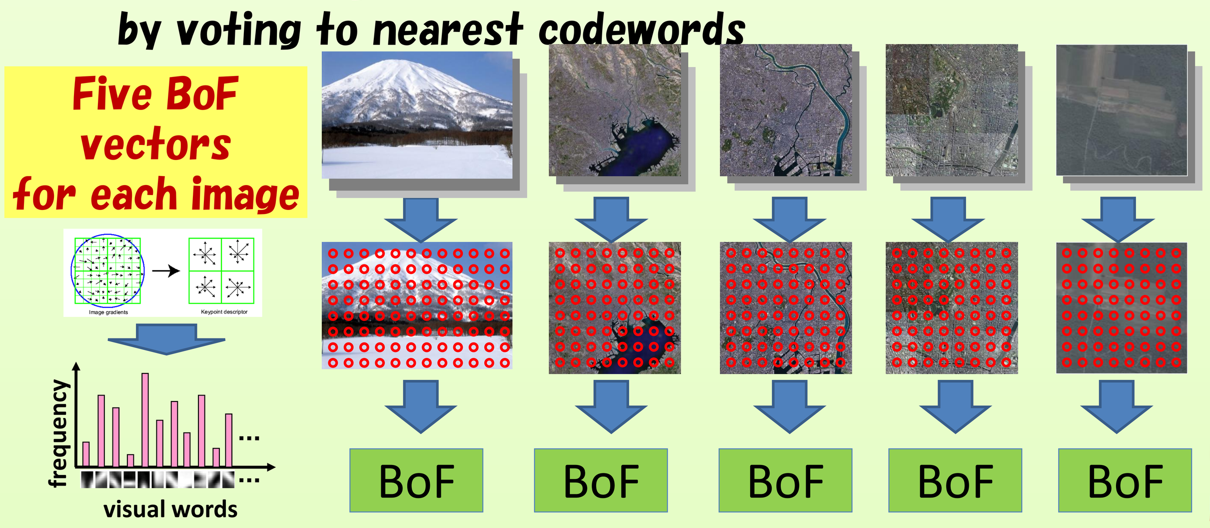


◆ Gather 4 levels of aerial photos for each geotagged photo.

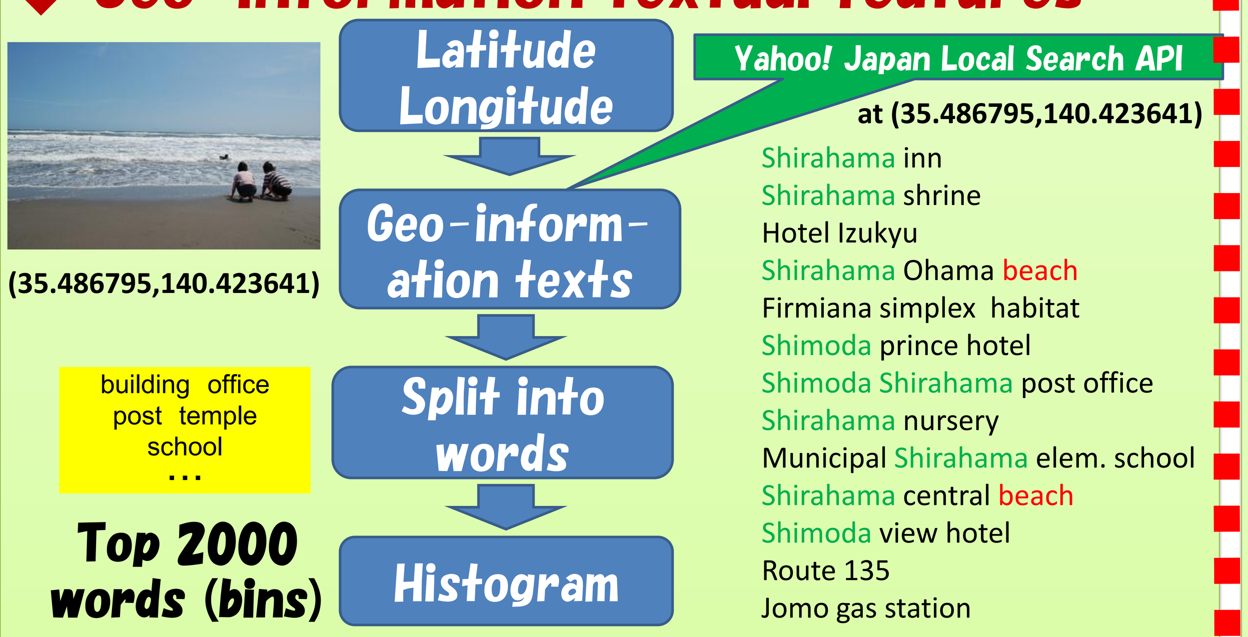
Features

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

(1) Sample points by grid sampling (*every 10px*) and describe local patterns around the sampled points with SIFT [Lowe 2004]
(2) Generate codebooks by k-means (*size: 1000*) and convert images into BoF vectors by voting to nearest codewords



◆ **Geo-information textual features**



Learning & Classifying

◆ **M**ultiple **K**ernel **L**earning

- Is an extention 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

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

Kernel weight (to be estimated by MKL)

Feature combinations of input for MKL

	BoF	HSV	Aerial	Geo	Text	Date
Vis (BoF + HSV)	✔	✔				
Vis + G (VG)	✔	✔		✔		
Vis + T (VT)	✔	✔			✔	
BoF + A (BA)	✔		✔			
Vis + A (VA)	✔	✔	✔			
Vis + A + T (VAT)	✔	✔	✔		✔	
Vis + G + T + D (VGTD)	✔	✔		✔	✔	✔
Vis + A + T + D (VATD)	✔	✔	✔		✔	✔
All (All)	✔	✔	✔	✔	✔	✔

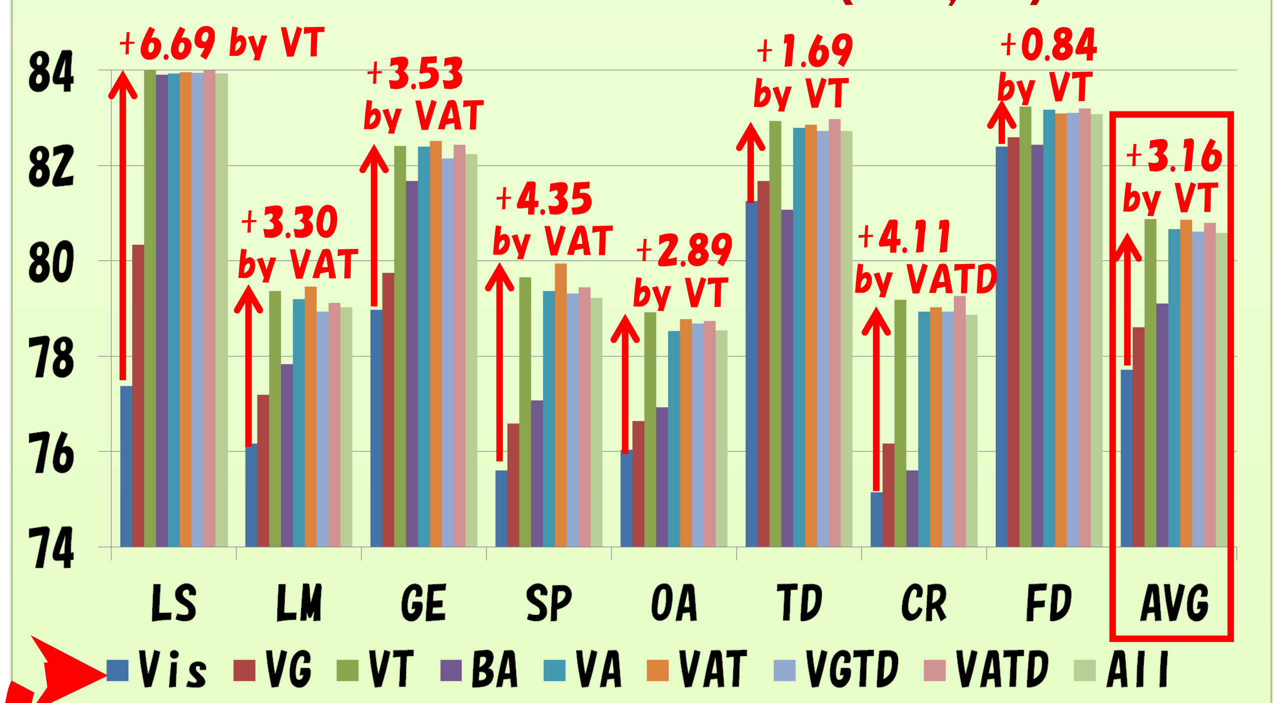
Experiments

◆ **1/0 classification for 28 classes by 5-fold cross validation**

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

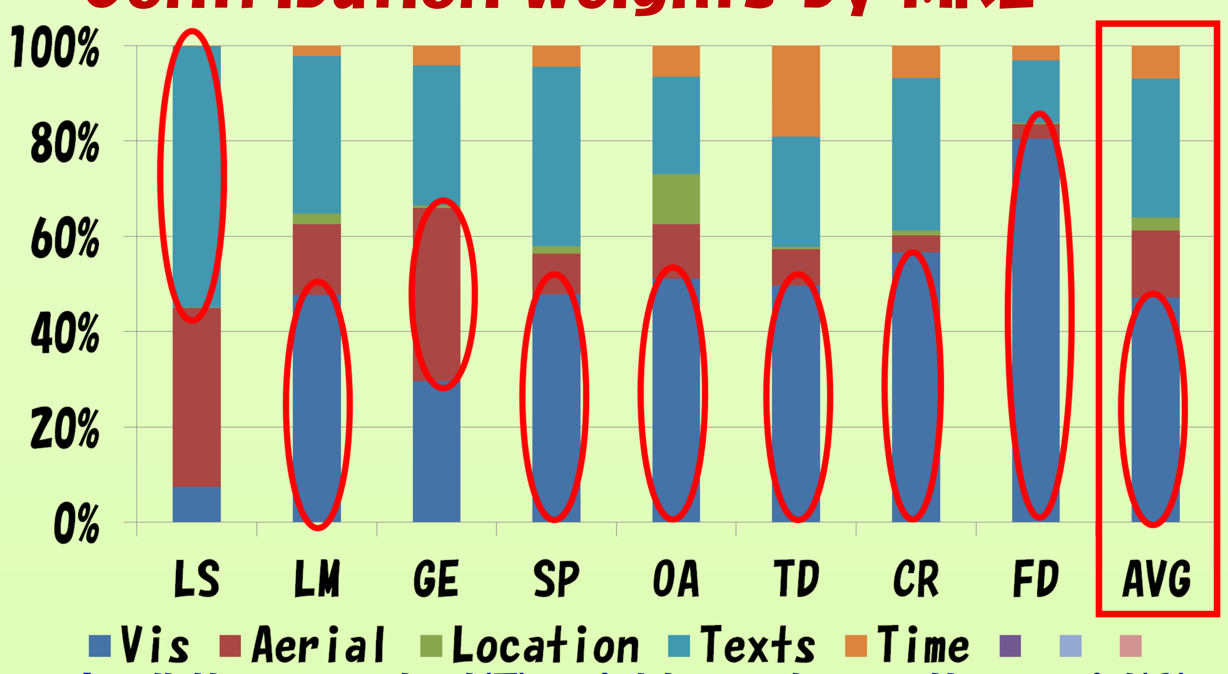
(LS) Location specific: Disneyland, Tokyo Tower
(LM) Landmark: bridge, shrine, building, castle, railroad
(GE) Geographical: lake, river, beach
(SP) Space: park, garden, landscape
(OA) Outdoor artifact: statue, car, bicycle, graffiti, vending machine
(TD) Time dependent: sunset, cherry blossom, red leaves, festival, costume-play festival
(CR) Creature: cat, bird, flower
(FD) Food: sushi, ramen noodle

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



Geo-text(T) is more helpful than aerial(A) in many cases.

- **Contribution weights by MKL**



In all the cases, text(T) weights are larger than aerial(A) weights. In LS and GE, aerial(A) weights are relatively large.

- Examples of geo-text features (top 5 elements of the average word vector over the concept)

Concepts assigned with large weights

costume-play		Disneyland		castle	
Ariake	0.0391	parking	0.1170	Himeji	0.0324
building	0.0388	garage	0.1074	city	0.0252
Tokyo	0.0250	Tokyo	0.0649	company	0.0201
parking	0.0223	Maihama	0.0424	building	0.0199
Harajuku	0.0222	resort	0.0416	school	0.0173

Concepts assigned with small weights

flower		vending machine		ramen noodle	
building	0.1089	building	0.0286	building	0.0522
embassy	0.0395	company	0.0200	company	0.0173
hall	0.0255	post office	0.0187	post office	0.0169
Roppongi	0.0211	nursery	0.0168	city	0.0117
Kamiyacho	0.0199	Toyama	0.0154	center	0.0115

Location words

Conclusions

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

- AP was improved by 3.16 % on average
- Geo-textual features and aerial features improved AP for most of the concepts
- Less helpful for "food" concepts
- Texts are more helpful than aerial photos for our dataset (limited within Japan) (depending on availability of rev-geocoding)

◆**Future work**

- More large-scale experiments with much more categories
- multi-class experiments