# A Visual Analysis of the Relationship between Word Concepts and Geographical Locations

*ACM International Conference on Image and Video Retrieval (ACM CIVR)*
*Island of Santorini, Greece, July 9th, 2009*

**Keiji Yanai**, Hidetoshi Kawakubo and Bingyu Qiu
Department of Computer Science,
The Univ. of Electro-Communications, Tokyo, Japan

# "Sushi" in Caltech 256

- (Probably) Collected in English Keywords

# "Sushi" in our own dataset

- Collected in Japanese Keywords

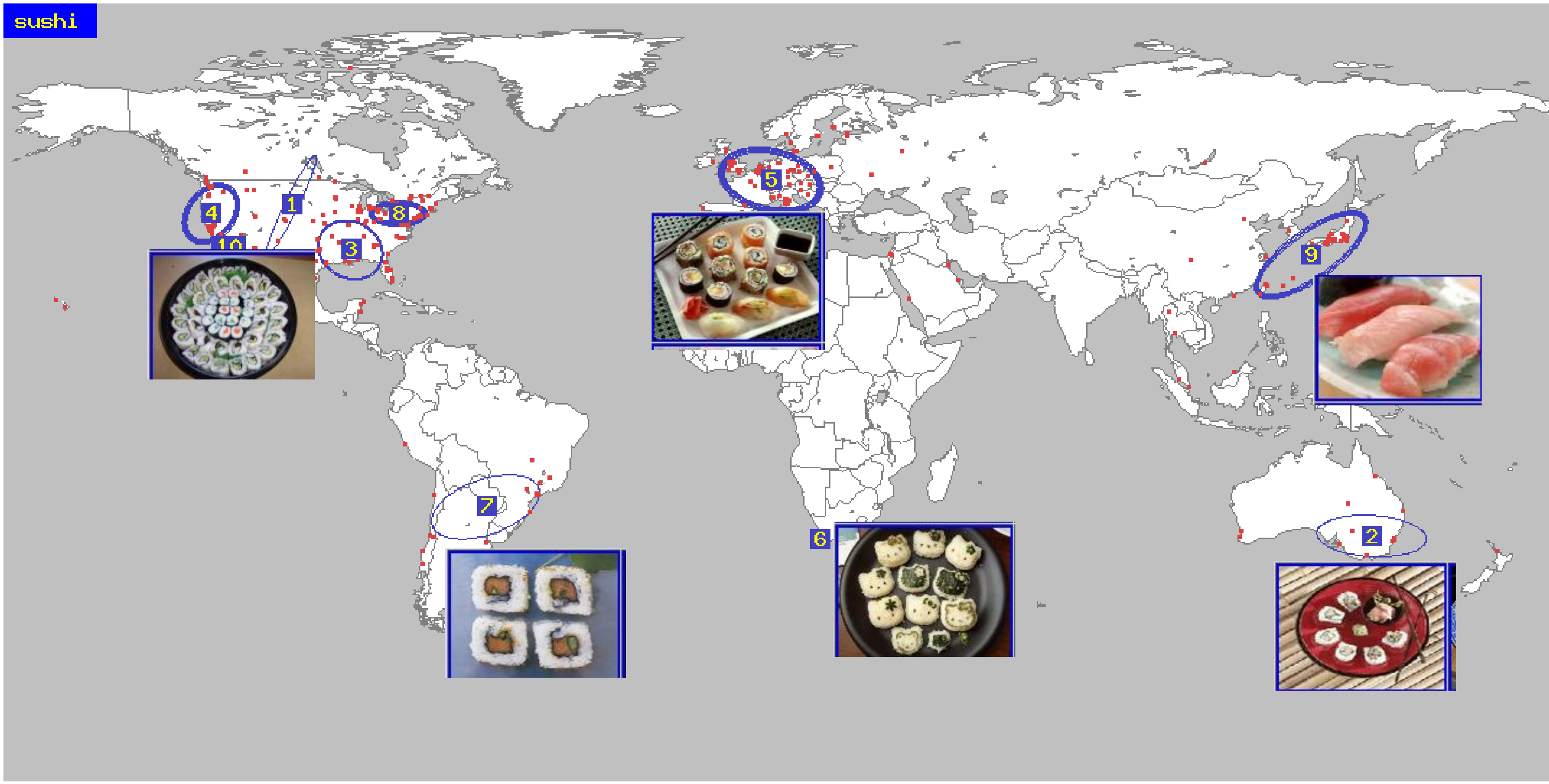
# Which do you like to eat ?

These two "sushi" image sets are surely different, although both are image sets associated with the "sushi" concept !

Caltech "sushi"

Japanese "sushi"

# "Sushi" over the world




- Image sets corresponding to the same concept are changing depending on locations or cultures.

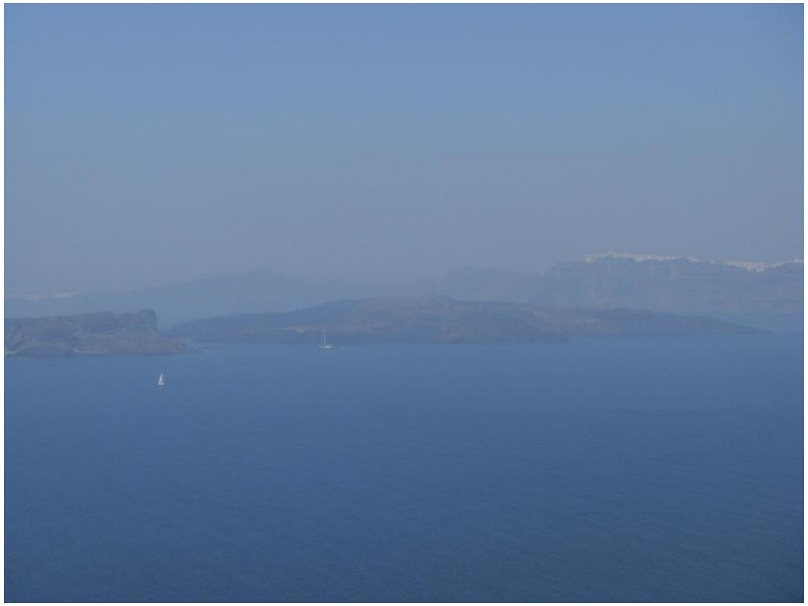
# Questions about concepts and locations (or culture)

- From this observation, representative image sets associated with a given concept might change (slightly) depending on locations or cultures.

Questions?

- Which concepts are location-dependent ?
- Which concepts are global (unchanged) over the world ?　(e.g. "sea", "sky")
- How concepts change depending on locations ?

# "Sea": global concept



**The Aegean sea** **The Japanese sea**

# 1. Objective, Background & Related work

# Objective of this paper

- Analyze the relationship between word concepts and locations using geotagged photos on Flickr

- Consist of two parts:
  1. Entropy-based analysis
  2. A system to detect "cultural differences"

  They are relatively independent.

# Background: *geotagged photos*

- The number of **geotagged photos** on the Web grows rapidly: Flickr, panoramio
  - Flickr has 100,000,000 geotagged photos.

(Feb. 2009)


***A "geo-tag" represents the coordinates (latitude,longitude) of a location where a photo are taken.***

# Related work: Geotagged image

Many work used geotags to organize *landmark photos*.

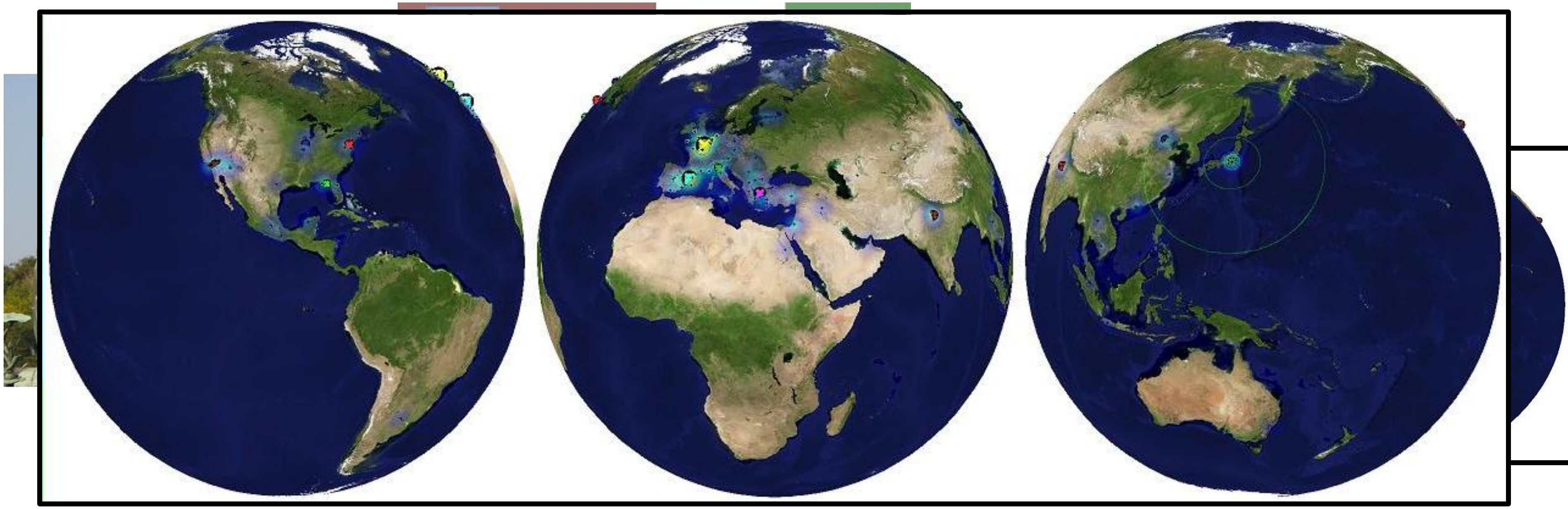
- Toyama et al. (2003)
- Jaffe et al. (2006)
- Simon et al. (2007)
- Kennedy et al. (2008)

…… and other many works.

**The exception is "IM2GPS" [Hayes et al. 2008]**
**among works on geotagged photos using image analysis.**

# Related work: "IM2GPS" [Hayes et al. 08]

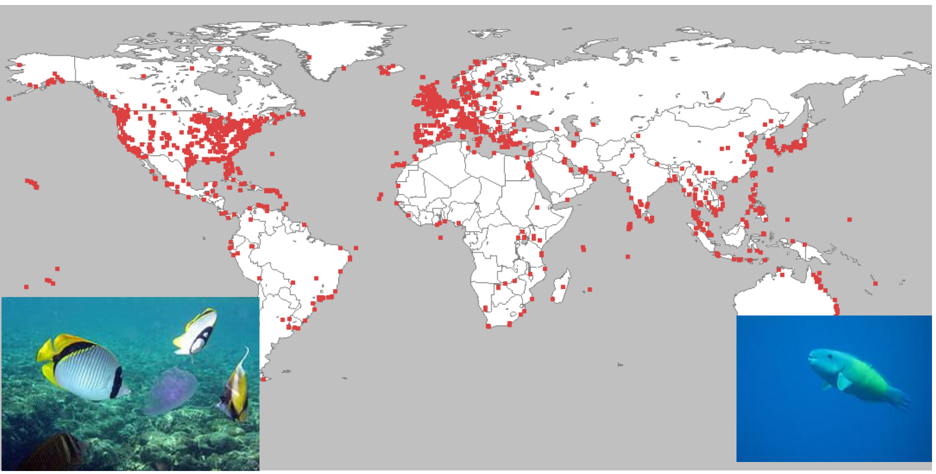
- Estimate the probability distribution over the world by nearest neighbor search for large-scale geotagged image DB. (ignoring "concepts")



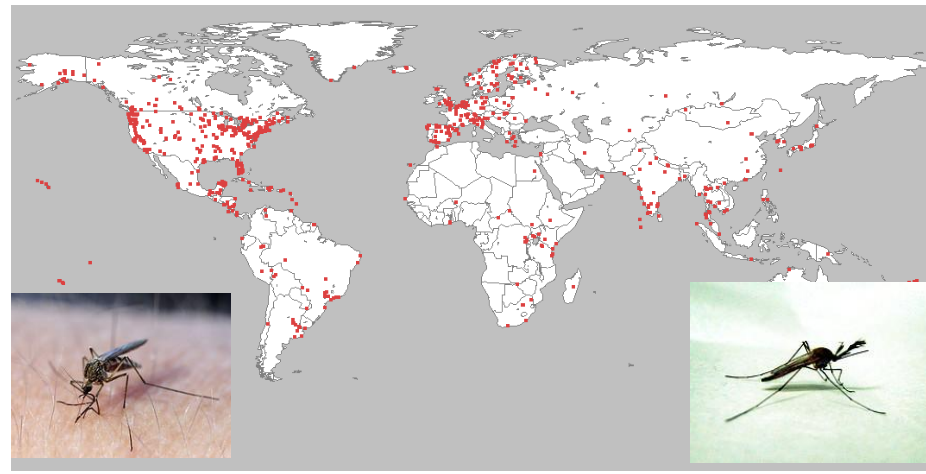
- This work suggests there exists the relation between visual features and geo-locations.

# There are many non-landmark geotagged photos in Flickr !

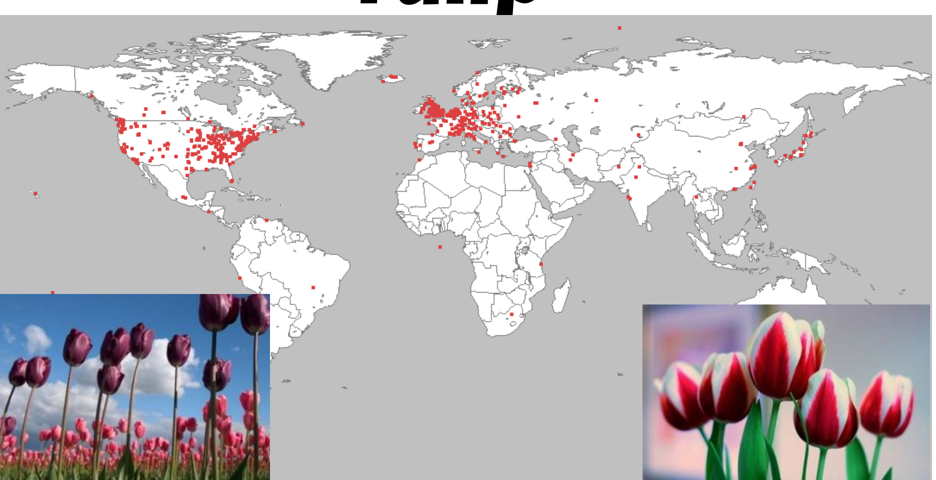
## fish



## mosquito



## tulip



## Deutschland

# 2. [Part 1]
# Entropy-based analysis

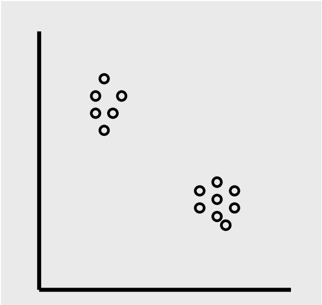
# Entropy-based analysis

**Examine the relation between distributions of visual features and geo-locations for many concepts**

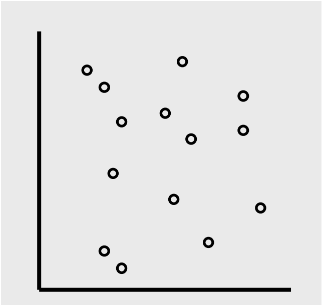
1. **Entropy-based measure of visual features (Modified method of** [Yanai and Barnard 05] )
2. **Entropy-based measure of geo-locations**
3. **Analysis the relation between two kinds of entropy**
   - **For 230 nouns**

# Image region entropy [Yanai and Barnard 05]

- **A measure of "visualness" of words** (concepts)
- **Represent the property of the distribution of image region features**



Biased / uneven:
low entropy
having "visualness"



Random/uniform:
high entropy
not having "visualness"

**"Low entropy" means the concept has visual property, "High entropy" means the concept has less visual property.**

# Low entropy: "scary" [Yanai and Barnard 05]


"Visual" concept

Detected "scary" regions

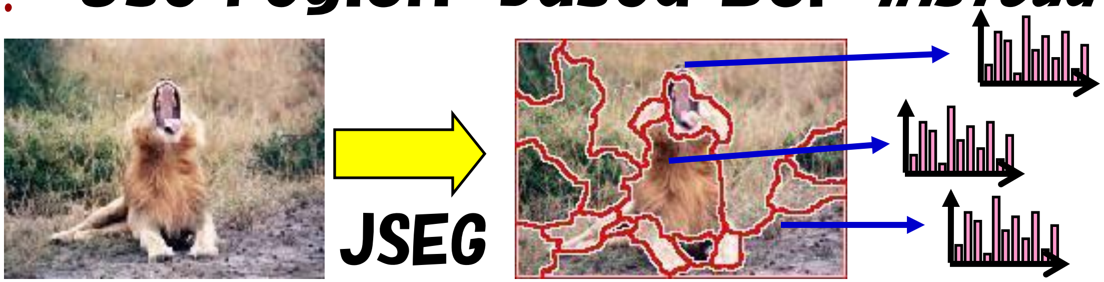
# High entropy: “famous”

[Yanai and Barnard 05]


**“Non-visual” concept**

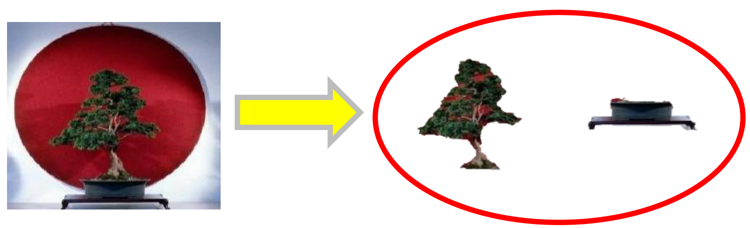
# Modified image region entropy

- **Follow "image region entropy"** [Yanai et al. 05]
  1. **Use region-based BoF** ***instead of color, texture***




Randomly-sampled SIFT-based bag-of-features(BoF) (1000-dim)

  2. **Use mi-SVM to select relevant regions**

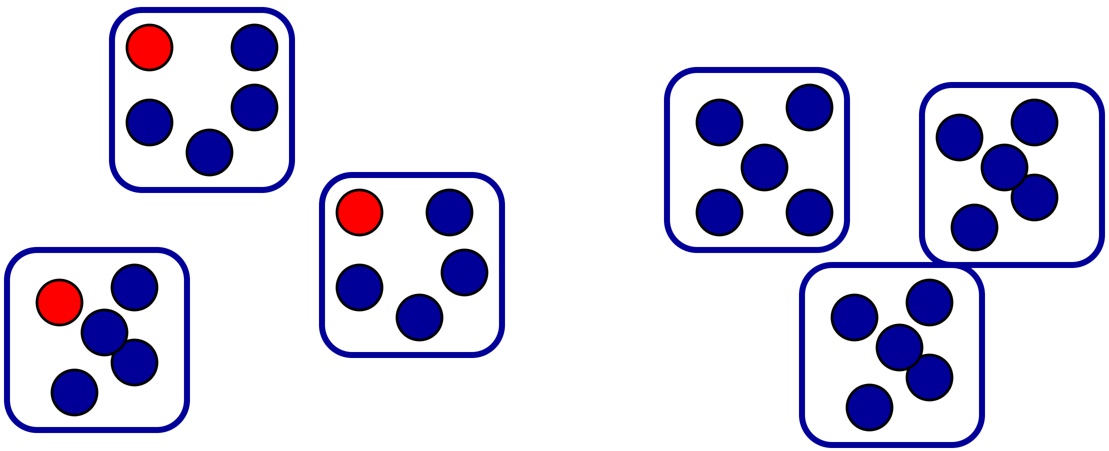


*For excluding background and noise regions*

  3. **Model the distribution of region-based BoF vectors with pLSA** ***instead of GMM***
  4. **Calculate entropy based on pLSA vectors**

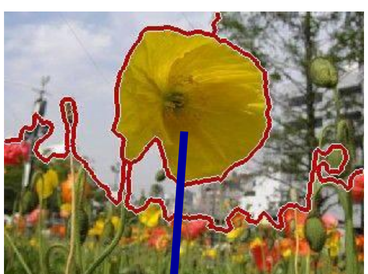
# Multiple Instance Setting

- Positive bags / Negative bags



positive ins. (foreground)

negative ins. (background)



Positive instances of "flower"

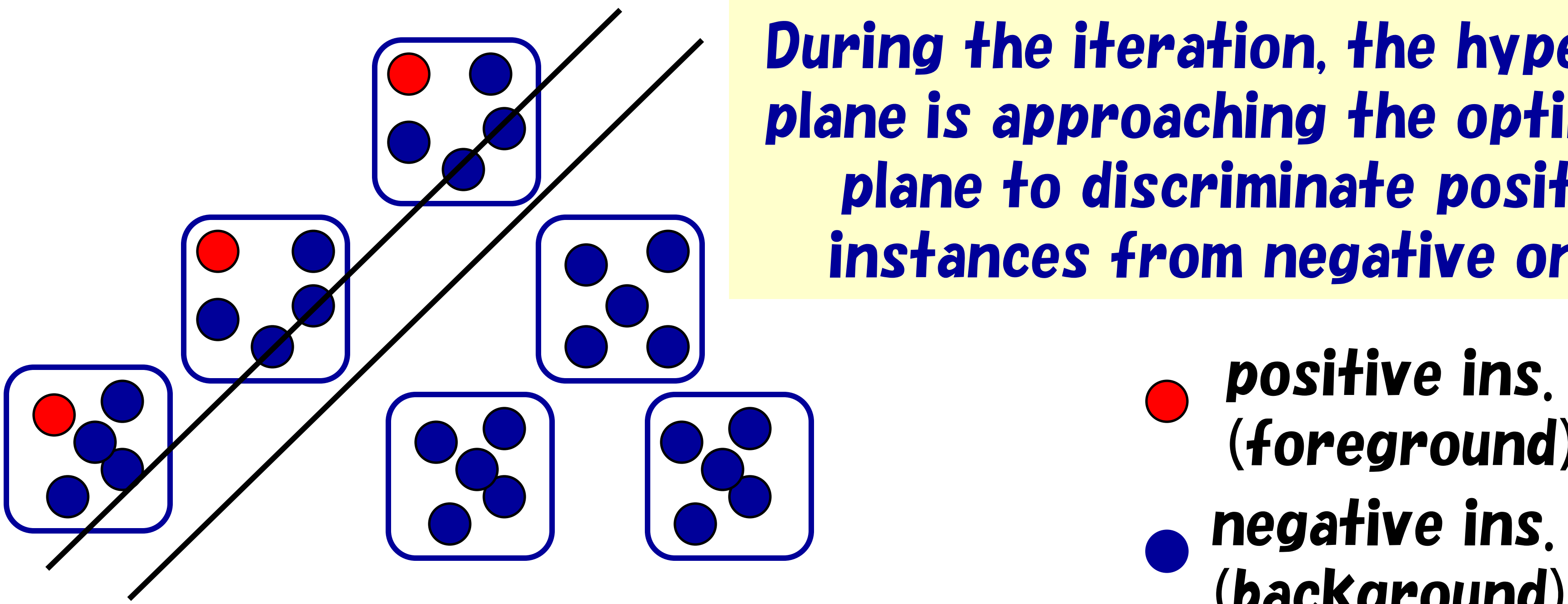
The rest of regions are negative regions.

**pseudo-training images** | **random images**

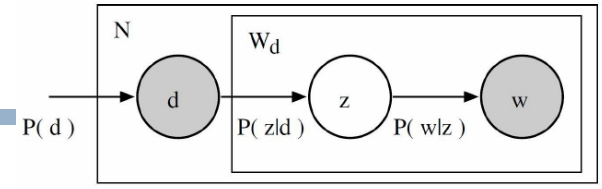
# mi-SVM *[Andrew et al. NIPS 03]*

- **Apply soft-margin SVM iteratively**
  - Training → classifying → training → classifying → …… (repeat 5 times)

**During the iteration, the hyper-plane is approaching the optimal plane to discriminate positive instances from negative ones.**



# Distribution modeling with the PLSA topic mixture




$$P(w,d) = P(d)\sum_{z} p(w \mid z)P(z \mid d)$$

w: visual words, d: regions, z: topic

① **Apply PLSA for all the regions of all the random (background) images in advance**

⇒ **Obtain P(w | z) and fix it (based distribution)**


② **Estimate P(z | d) for each regions with fixed P(w | z) using fold-in heuristic** [Hofmann 09]

# Calculate image region entropy

- H(X): entropy of the given word "X"

$$H(X) = -\sum_k P(z_k|X)\log_2 P(z_k|X)$$

$$P(z_k|X) = \frac{1}{I}\sum_i P(z_k|d_i^X)$$

- H(X) can be calculated from each of 5 iterations of mi-SVM

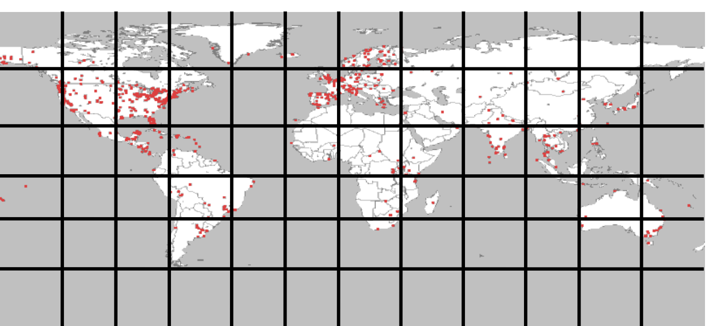
- Regard the minimum H(X) during 5 iterations as the final entropy H(X)

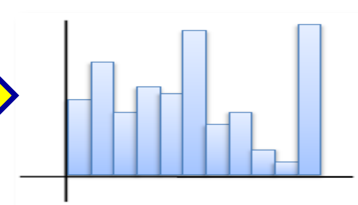
# Calculate geo-location entropy

1. Divide the world into 4 kinds of grids with every 10 degrees by shifting 5 degrees in terms of both latitude and longitude
2. Build histograms regarding the geotags of the selected regions
3. Calc entropy
4. Select minimum one

Geo-location entropy

$$H_{geo}(X) = -\sum_i b_i \log_2 b_i$$



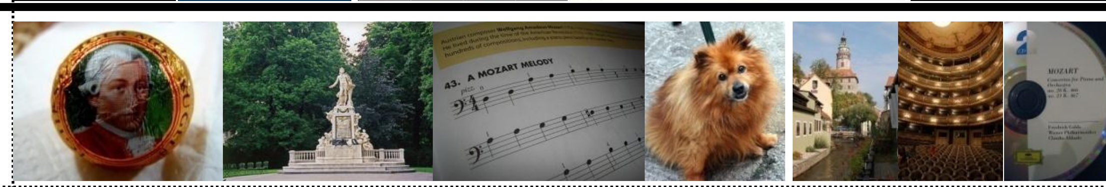
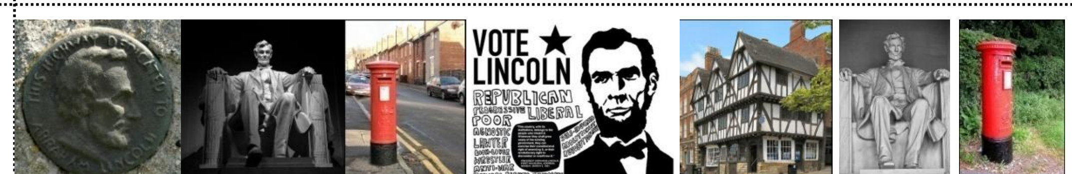


# 3. [Part 1] Results of entropy-based analysis

# Experiments

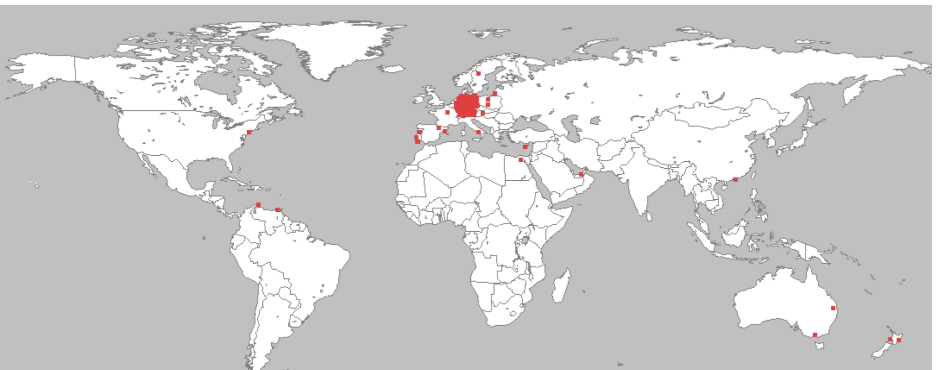
- Data
  - 230 nouns including various kinds of words
    - Gathered photos including the given nouns as their tags
  - 500 geotagged photos at least / each tag
    from Flickr
    (limiting 5 photos for each tag per user ID)
- After selecting relevant regions for each tag, calculate the two entropy:
  Image region entropy $H_{vis}(X)$
  Geo-location entropy $H_{geo}(X)$
- Analyze relation between them

# Image region entropy $H_{vis}(X)$

| | | |
|---|---|---|
| sun | 3.6497 |  |
| rainbow | 4.5538 |  |
| moon | 4.6686 |   |
| mozart | 7.8349 |   |
| lincoln | 7.7327 |   |
| school | 7.6173 |   |

# Geo-entropy $H_{geo}(X)$

## Deutschland

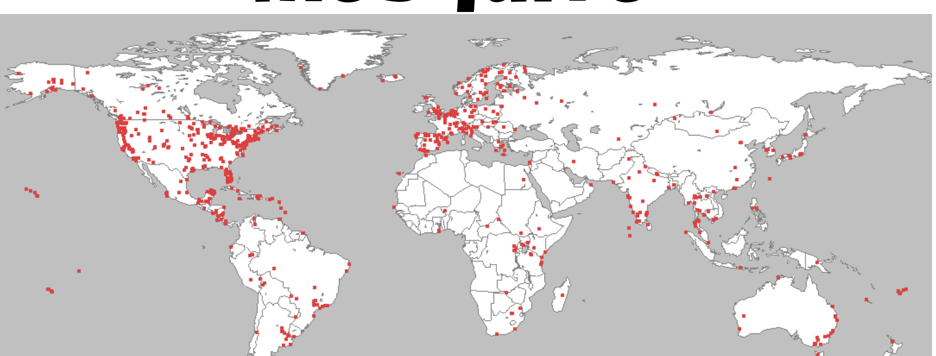


$H_{geo}(X) = 0.2602$

## Rome

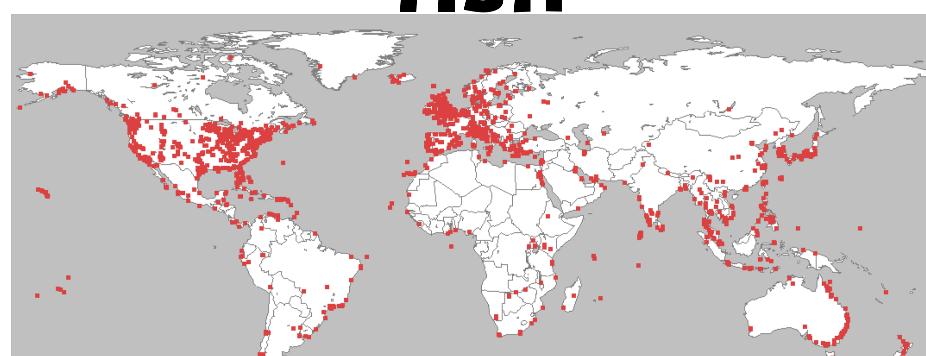
$H_{geo}(X) = 0.3843$

## mosquito



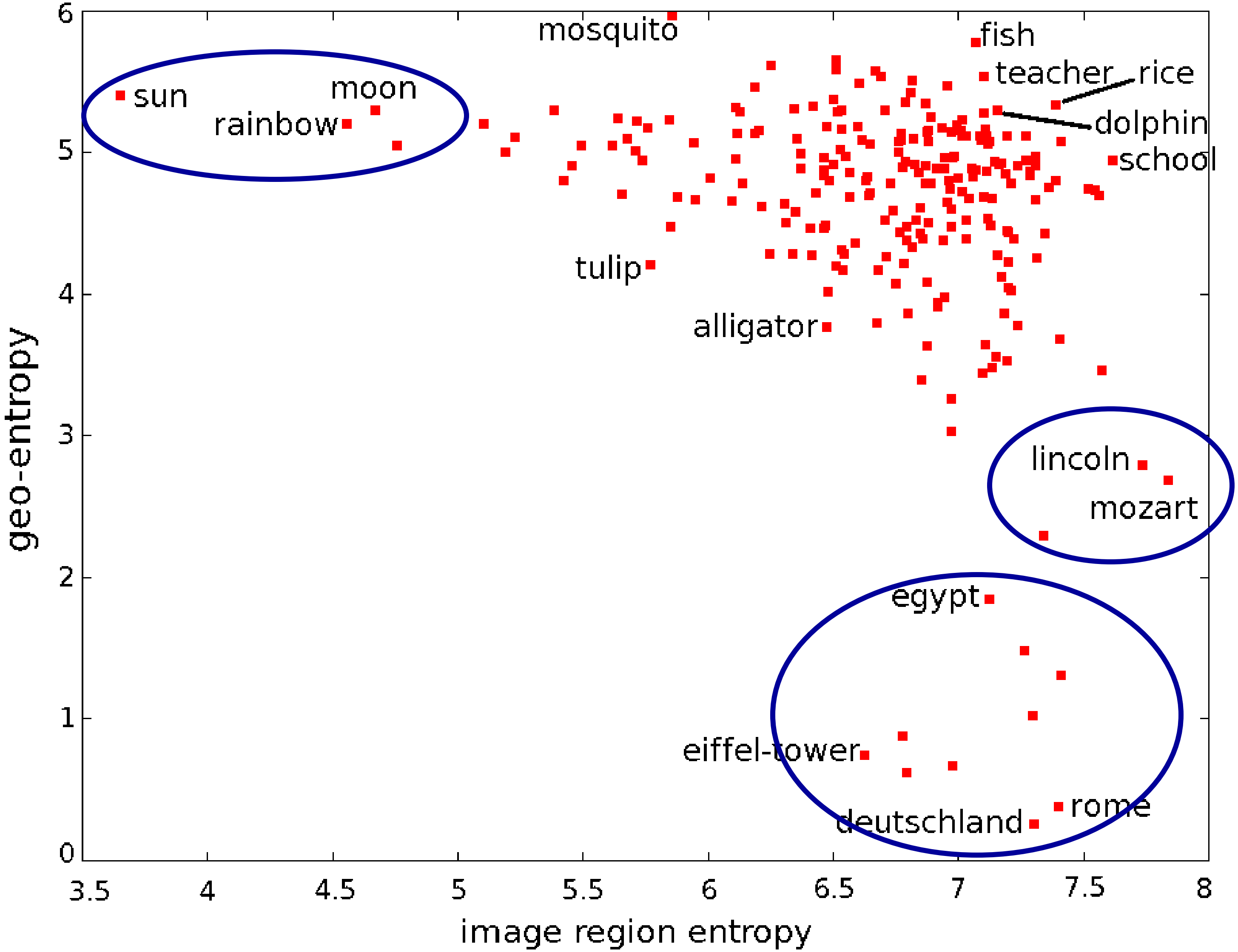
$H_{geo}(X) = 5.9759$

## fish



$H_{geo}(X) = 5.7831$

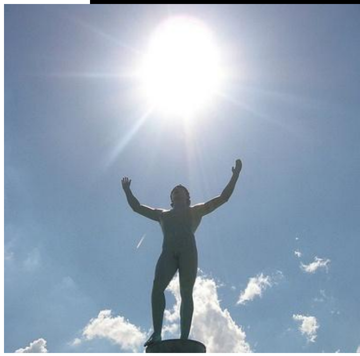
# Image entropy vs. geo-entropy

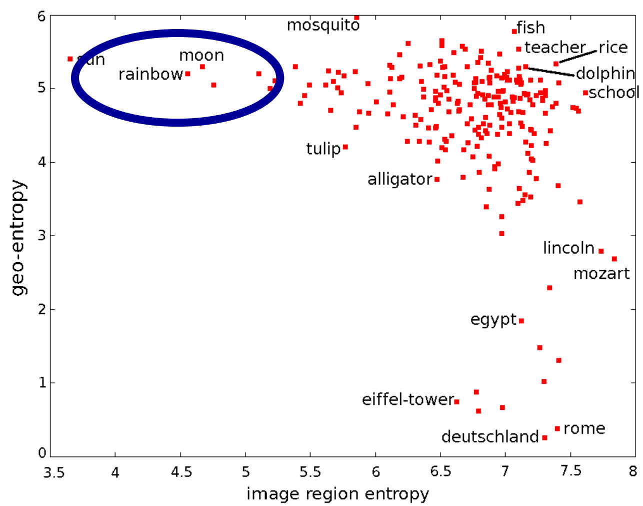


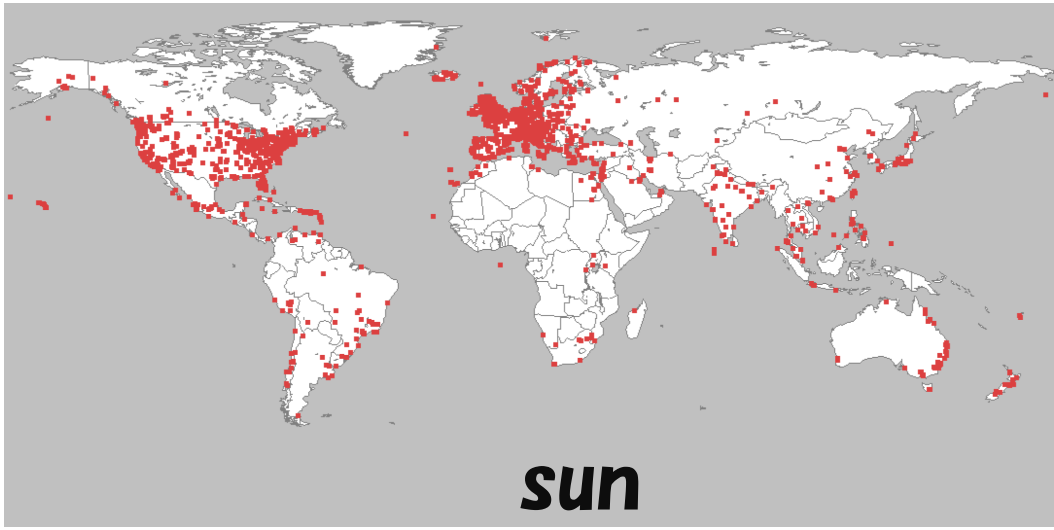
# Sun, rainbow, moon

- Comcepts related to sky
  - Image region entropy : low
  - Geo-location entropy : high

They exists everywhere in the world, and the apperances are similar.



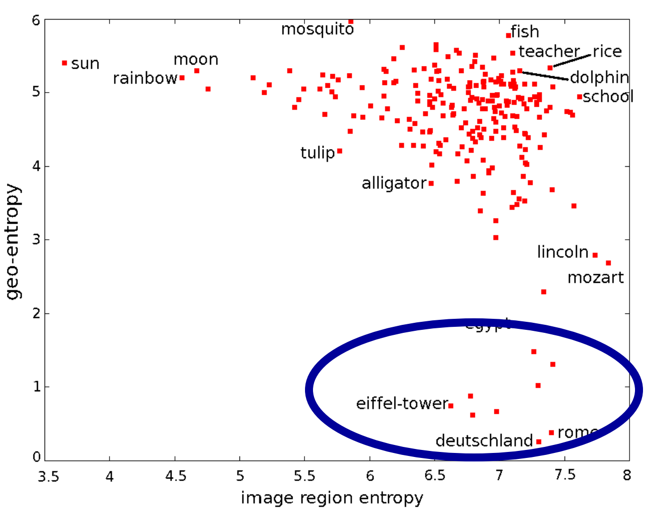


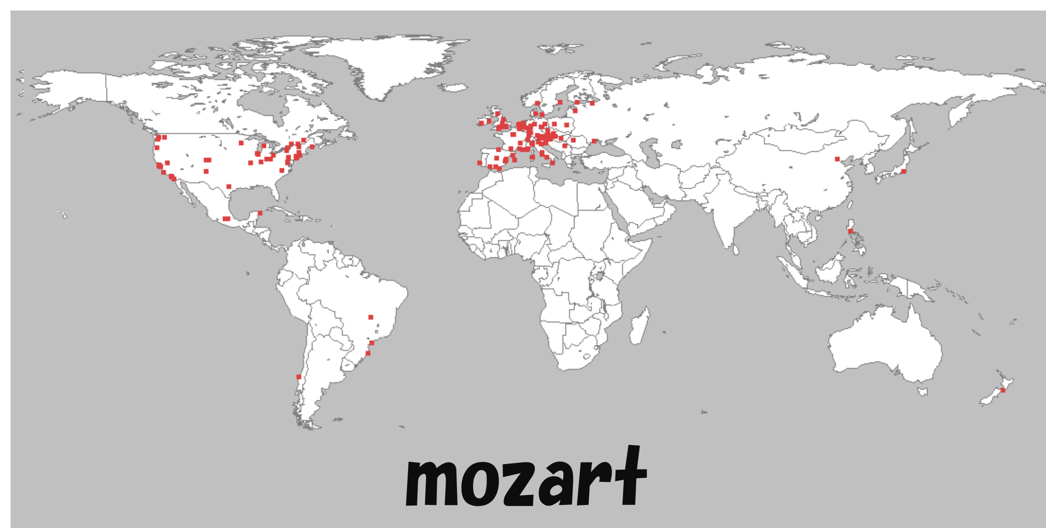



# Rome, Deutschland, Mozart

- Image region entropy: high
- Geo-location entropy : low

The geotags concentrates on specific areas. Their appearances are various.

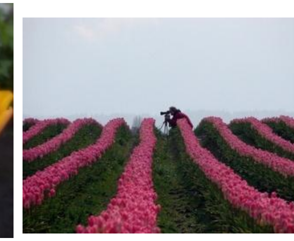
# tulip

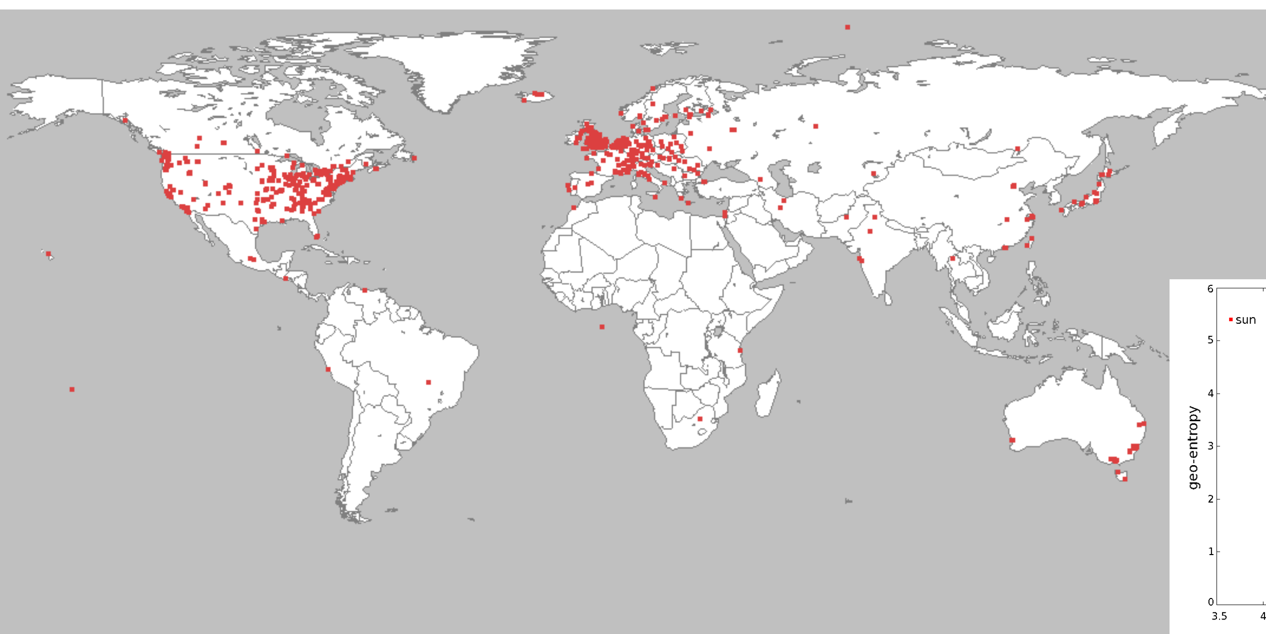
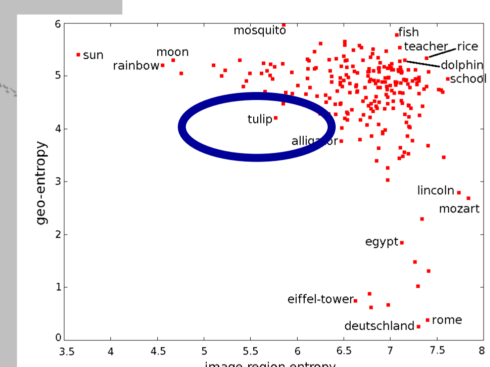
Image region entropy: low
Geo-location entropy : med.




- Variance of color did not reflect on image region entropy, since we use SIFT-based BoF representation.

- Holland and England are main areas.

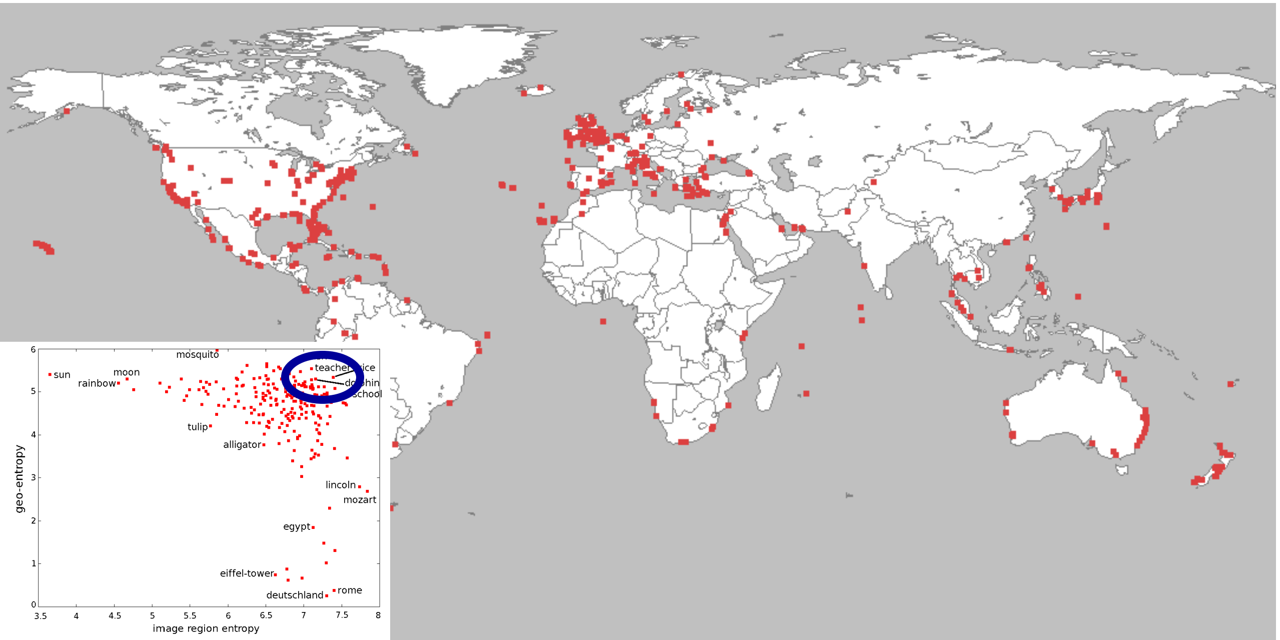




# dolphin

**Image region entropy: high**
**Geo-location entropy : high**

- Most of dolphins are taken in sea or aquarium
- In seaside areas over the world

# rice

## Image region entropy: high
## Geo-location entropy : high

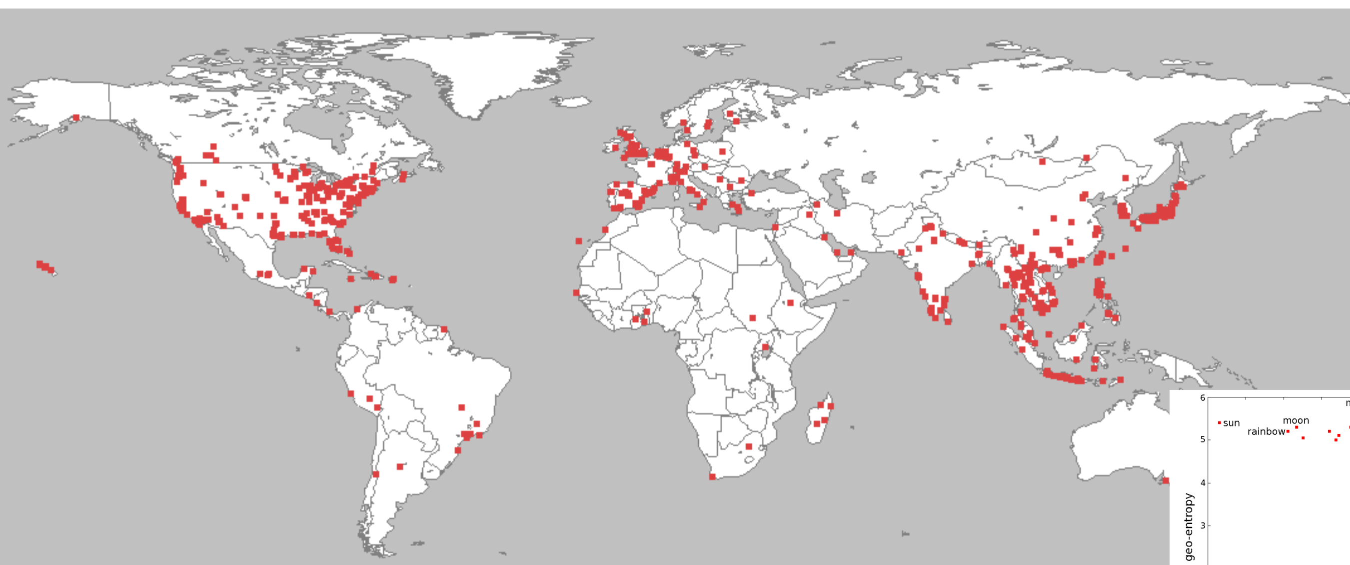
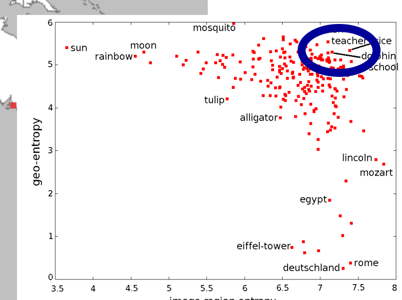








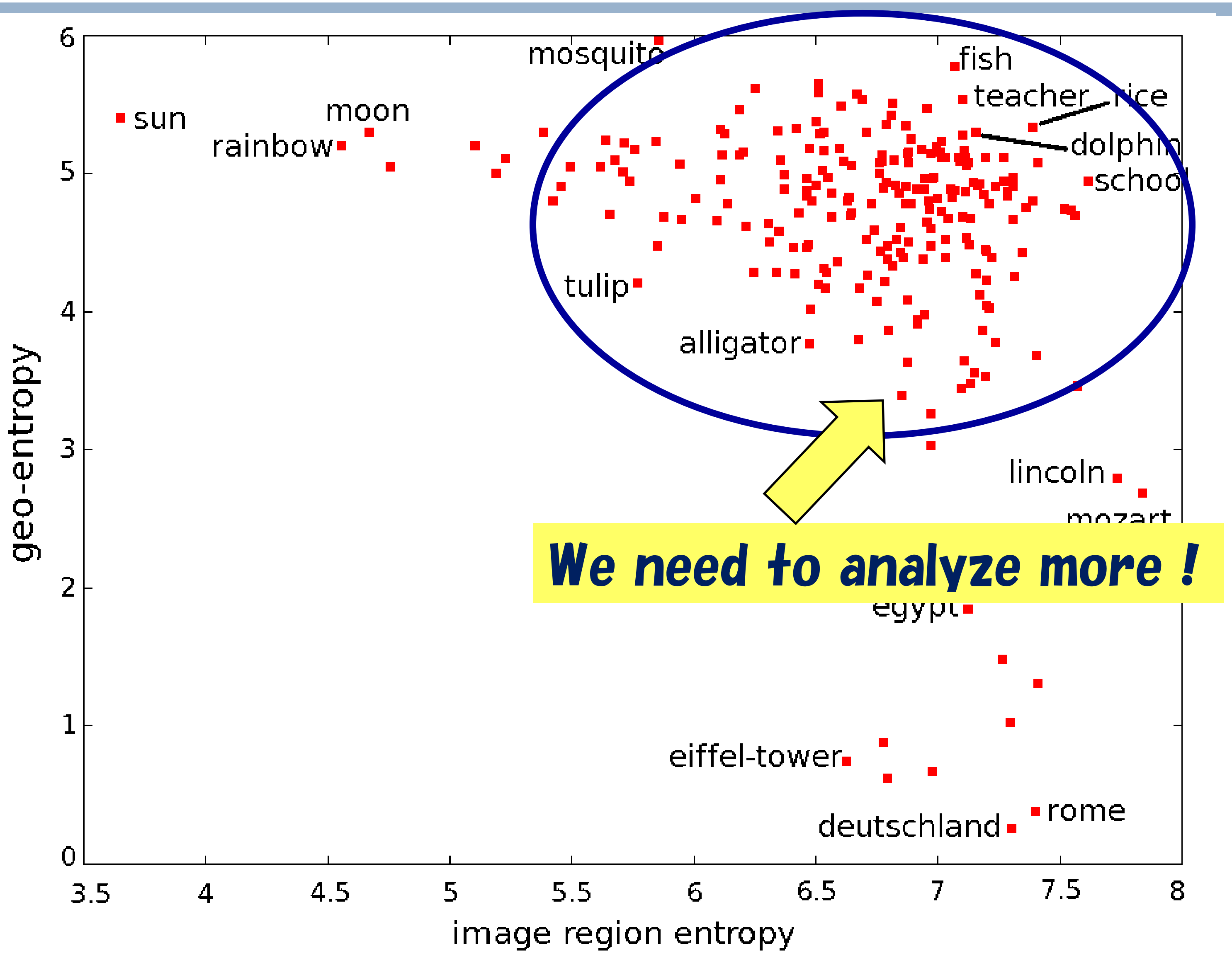



$$H_{geo}(X) = 5.3425$$

# 4. [Part 2]
# Discovering concept differences in terms of location

# Image entropy vs. geo-entropy



# Objective of the second part

Utilize the set of geo-tagged photos on Flickr

- A system to mine representative photos for representative areas or regions from geotagged photo DB.


Raw geo-tagged photos on Flickr

Relevant photos after noise image removal


Representative photos for typical regions

# Motivation : Foods over the world

- So with such geotagged photos, we can **discover specific objects over the world.**
  - ✓ Do you know all kinds of famous "noodles" in the world?
    - ✓ "Ramen" and "Soba" in Japan, "Thai noodle" in Thailand, "Chinese noodles", "rice noodle" Taiwan, "Spaghetti" in Italy…







**As a result, we can discover cultural differences on specific concepts over the world !**

"clothes", "car", "sushi" ……

# Approach : three steps

## 1) Select relevant photos and remove noise

- ✓ Extract BoF vectors from all the images
- ✓ Visual clustering with k-means
- ✓ Select most relevant clusters based on the size of clusters

## 2) Detect representative regions

- ✓ Clustering based on geographic locations by k-means

## 3) Generate representative photo sets for representative regions

- ✓ Generate the PLSA topic vectors
- ✓ Aggregate photos according to the distribution of mixture topics and rank photos for each representative area

# Contributions

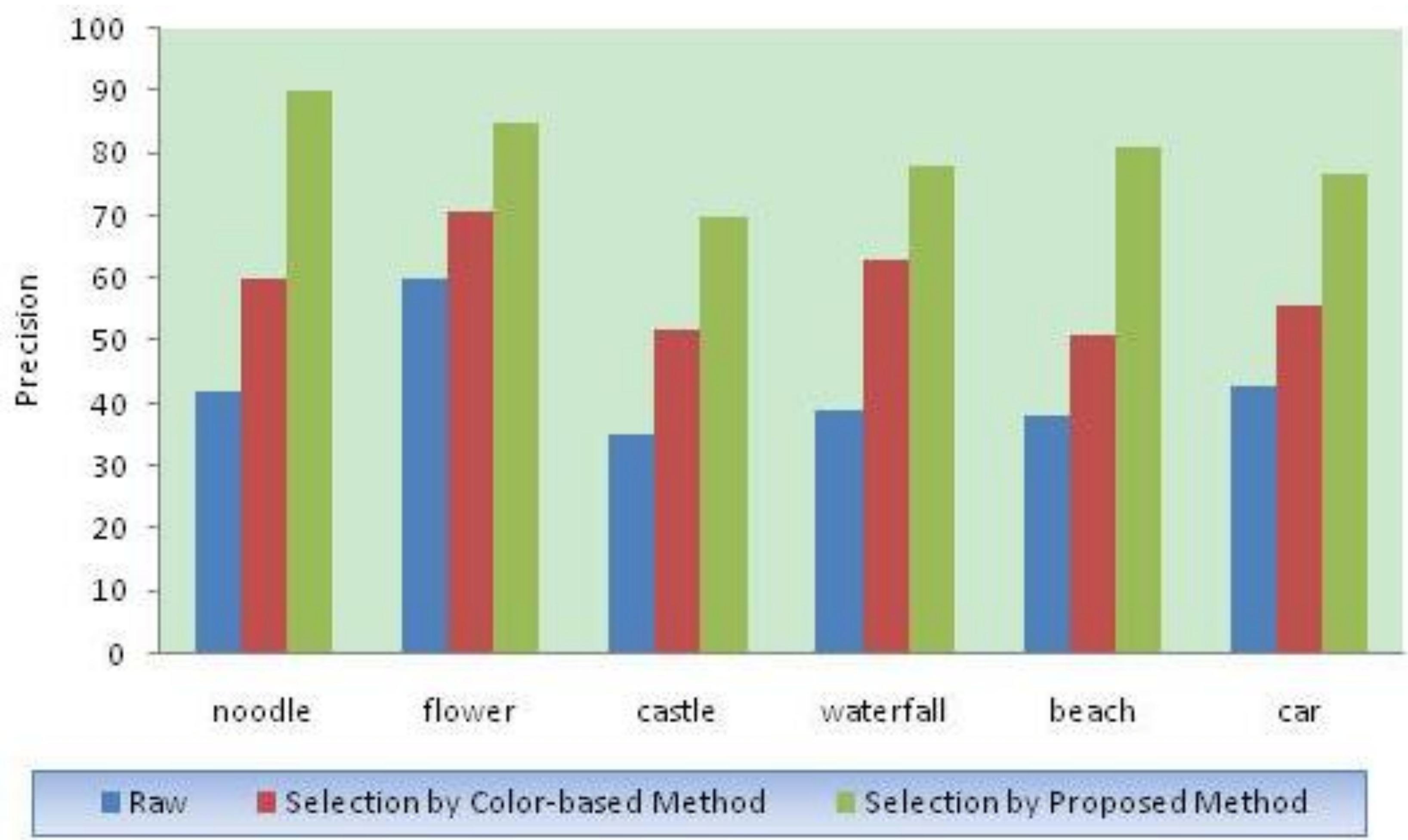
- Detect canonical photos of a specific object on each place over the world
  - Eg.) "noodle" : Chinese noodle in Asia,
    spaghetti in Europe
  - Do not limit to only scene of specific places or landmarks. Any objects are our targets !
- The method is not very novel, but the objective of the work is very novel.
  - Novel application for geotagged photo DB.

# 5. [Part 2] Results

# Experimental Results

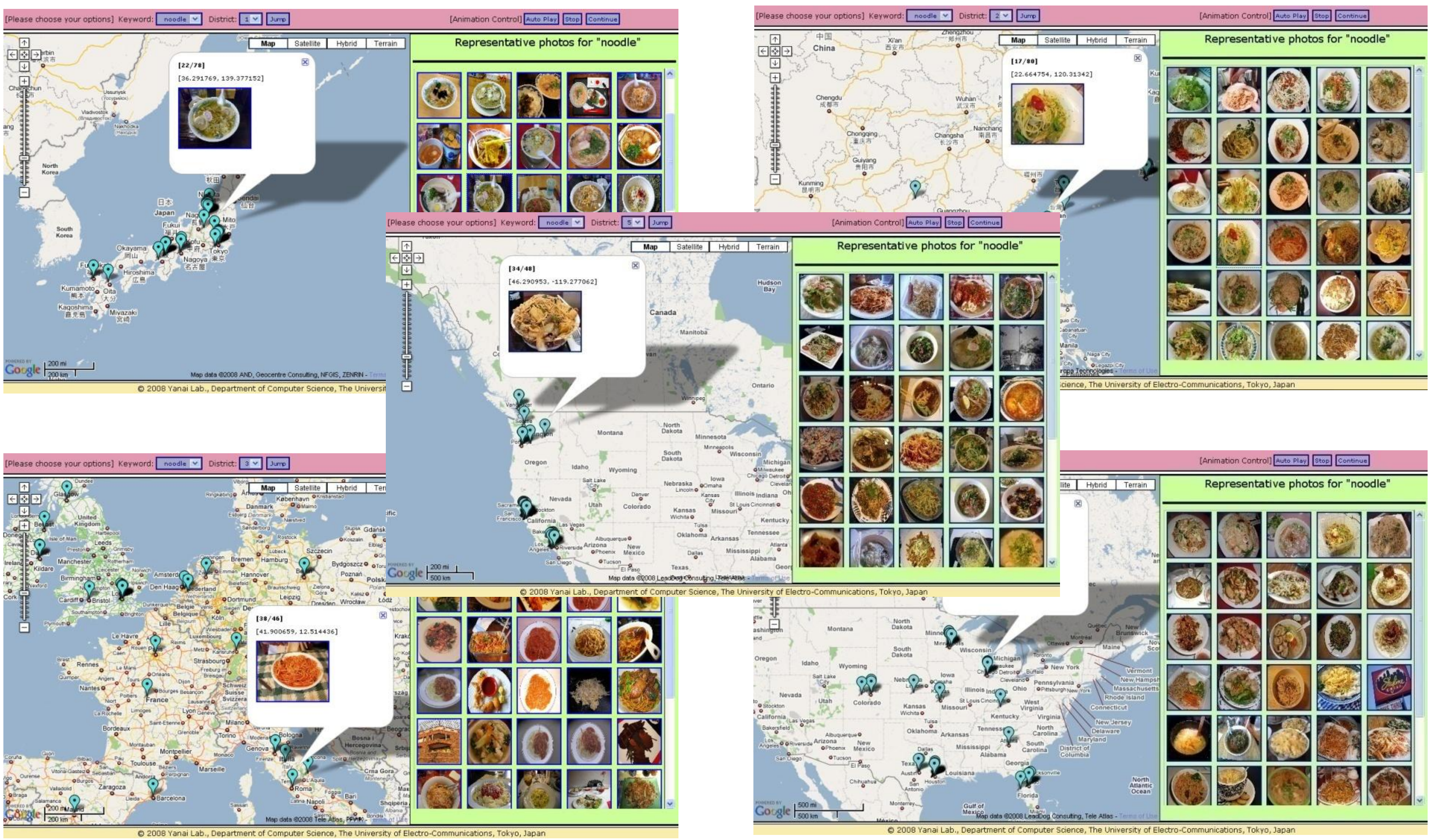
- "noodle", "flower", "castle", "waterfall", "beach", "car"
  - ✓ For each concept, collect about 2000 geo-tagged photos from Flickr distributed evenly in the world wide areas
- Quantitative evaluation for the 1st step
  - ✓ Evaluation on our proposed method for extracting the most relevant photos
  - ✓ Precision and Recall
  - ✓ Color-histogram-based method for comparison
- Examples of regional representative photos

# Quantitative Evaluation for 1st step




**Average recall : 73.0%**

# [Example of results] "noodle"



# [Example of results] “noodle”

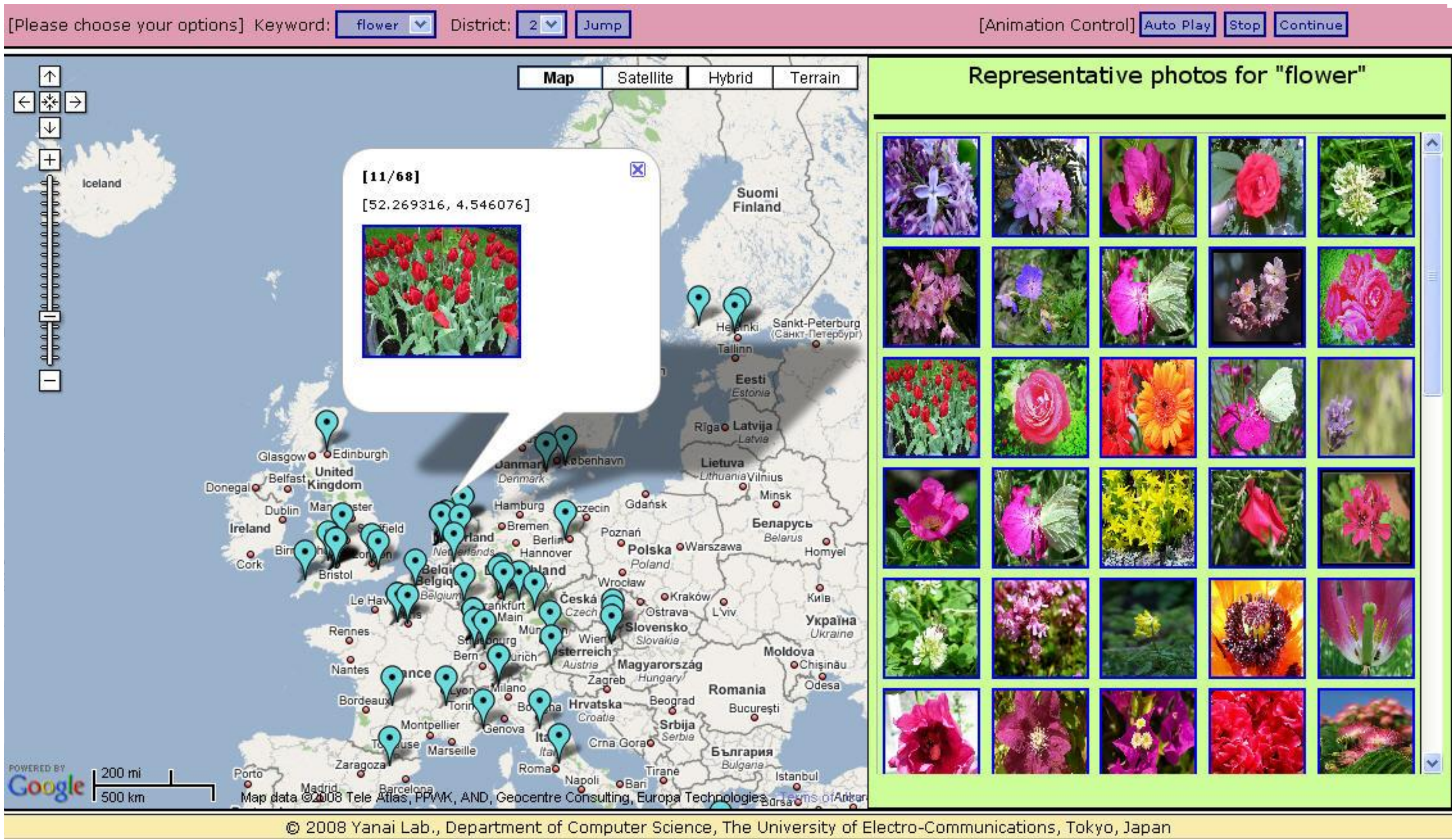

“spaghetti” photos in the European area

Taiwanese-style noodles and spicy Thai noodles

Many “ramen” photos in Japan

# [Example of results] "flower"

# [Example of results] "flower"

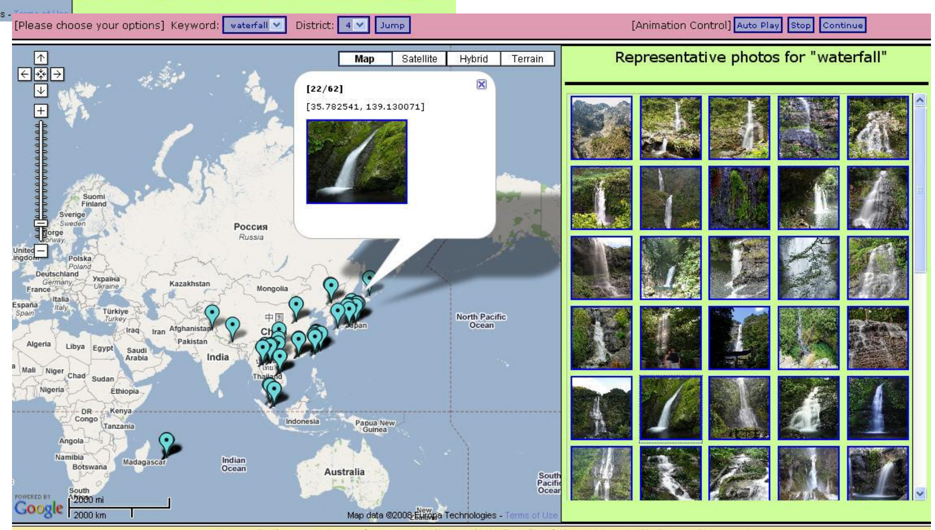



Netherlands national flower "Tulip"

Kansas State flower "Sunflower"

# [Example of results] "waterfall"


"Powerful " waterfalls in South America




"Beautiful" waterfalls in Asia

# Wedding cake !

Figure 12: "Wedding cake" in Mid US. Tall cakes are common. This is five-layered.

Figure 13: "Wedding cake" in Europe. They are much shorter and simpler than US.

# 6. Conclusions

# Conclusions

- In this paper, we pointed out that image sets associated with the same concepts are variable depending on locations
- To analysis that, we proposed to use geotagged photos on the Web.
  1. Entropy-based analysis
     - Image region entropy and geo-location entropy
  2. A system to help detect "cultural differences" by selecting representative photos for each location

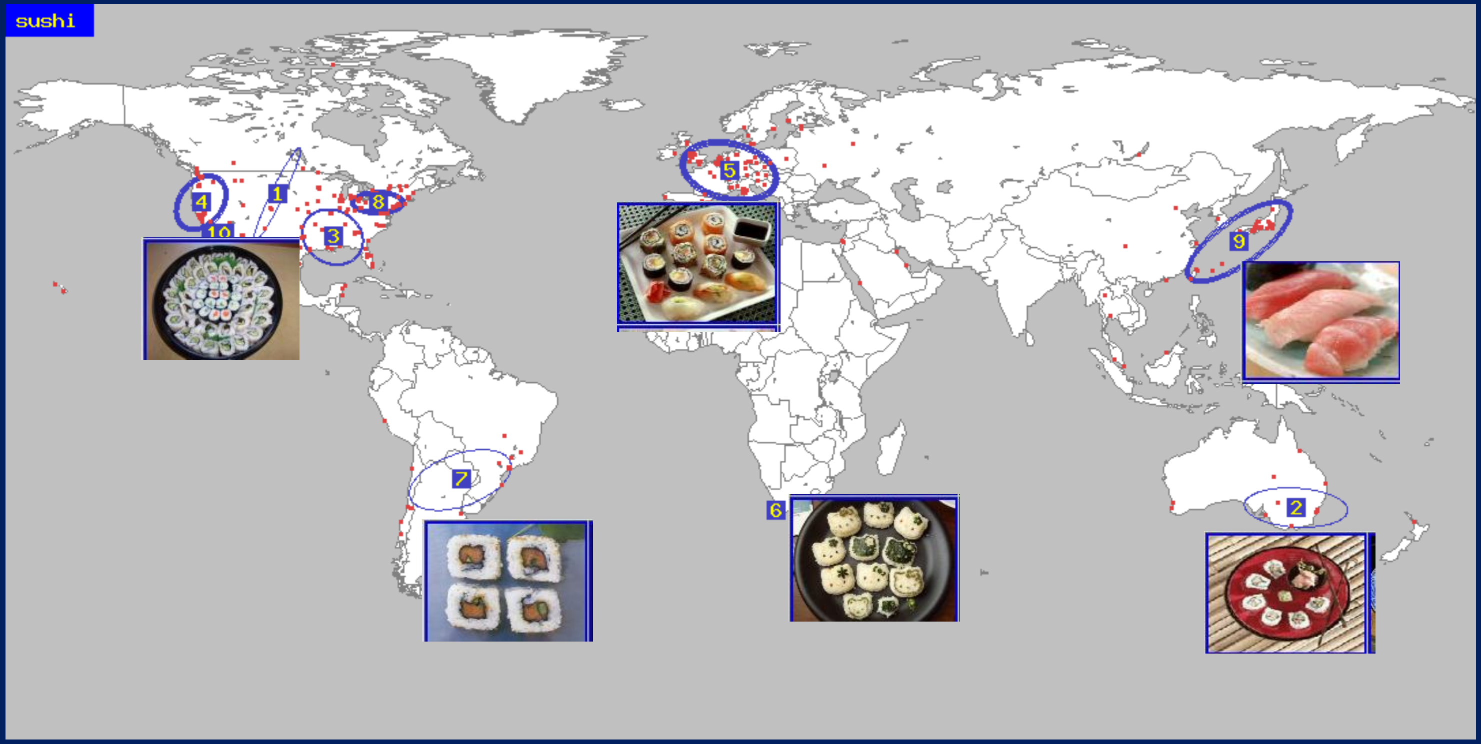
# Future work

- Use cross-language query to get images
  - Actually the obtained results are biased by English-culture, because we used only English words when gathering geotagged images from Flickr.
- Discover (subtle) cultural differences automatically *(hopefully hard for human)*
- Extensive study on cultural differences using a large-scale geotagged photo DB

# Towards location(culture)-specific image recognition

- This analysis will help build location (culture)-specific object/scene recognition systems.
  - For location-dependent concepts, specific training image sets are needed, while for global concepts global image sets are OK.
    - To build systems for Japanese people, a special training set for "Suishi" is needed.
    - For people in Santorini, specific image sets for "house" and "building" might be needed.

# Thank you for your attention !



# 3. Methods