An Analysis of the Relation between Visual Concepts and Geo-locations using Geotagged Images on the Web

IEEE ICME WS on Internet Multimedia Search and Mining
July 3rd, 2009

Hidetoshi Kawakubo *and* **Keiji Yanai**
The University of Electro-Communications, Tokyo, Japan

1. Objective & Background
2. Related Work
3. System & Methods
4. Experimental Results

1. Objective & Background

Background: geotagged photos

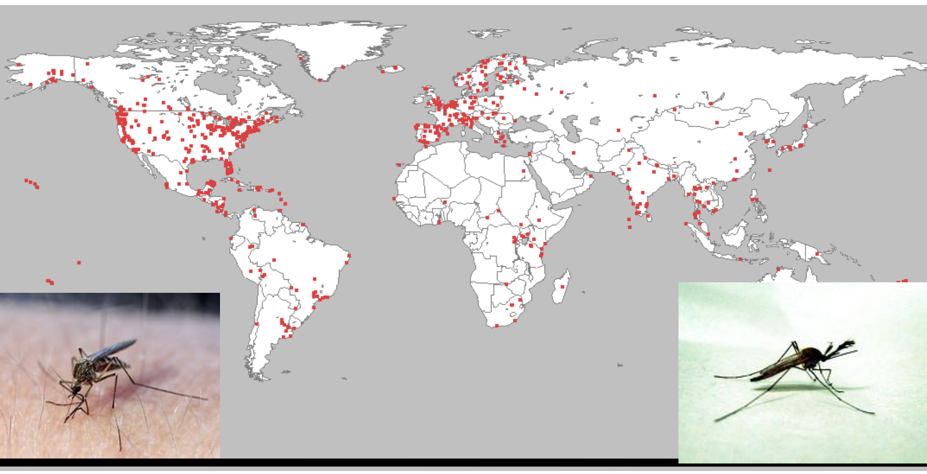
- The number of geo-tagged photos on the Web grows rapidly: Flickr, panoramio
  - Flickr has 100,000,000 geotagged photos.

(Feb. 2009)

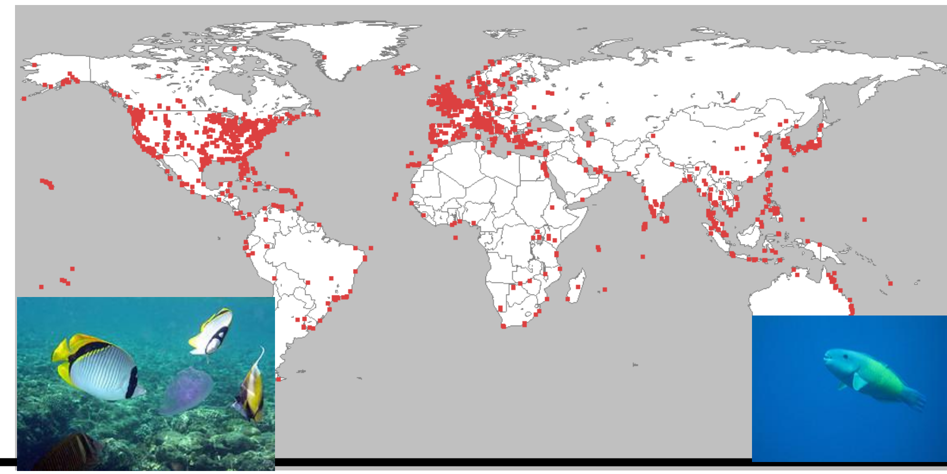
A "geo-tag" represents the coordinates (latitude,longitude) of a location where a photo are taken.

Distributions is different

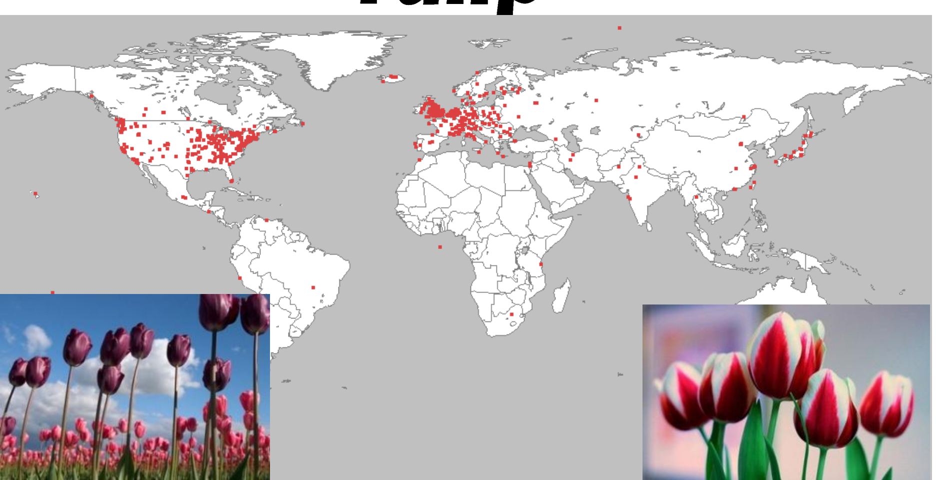
mosquito



fish



tulip



Deutschland

Objective

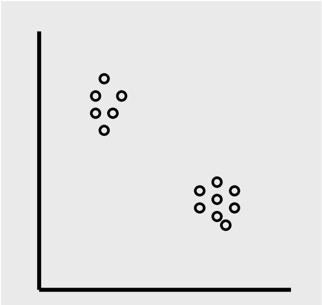
Examine the relation between distributions of visual features and geo-locations regarding many concepts (words)

1. Entropy-based measure of visual features
2. Entropy-based measure of geo-locations
3. Analysis the relation between both distributions
   - For 230 nouns and 100 adjectives

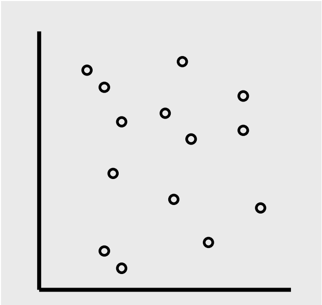
2. Related Work

Image region entropy [Yanai et al. 05]

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



Biased / uneven:
low entropy
having "visualness"



Random/uniform:
high entropy
not having "visualness"

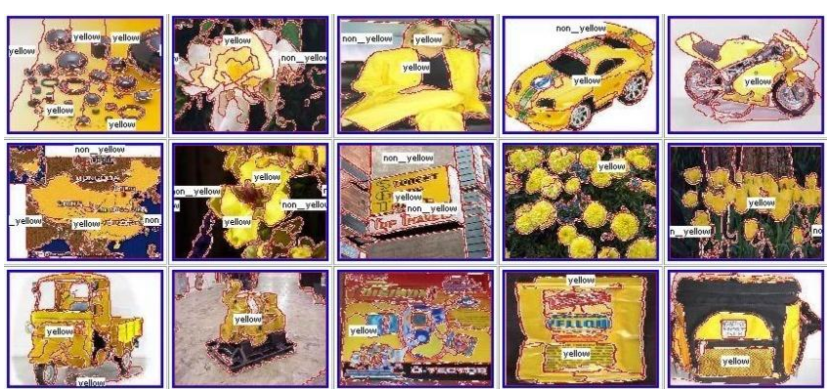
**"Low entropy" means the concept has visual property,
"High entropy" means the concept has less visual property.**

Image region entropy [Yanai et al. 05]

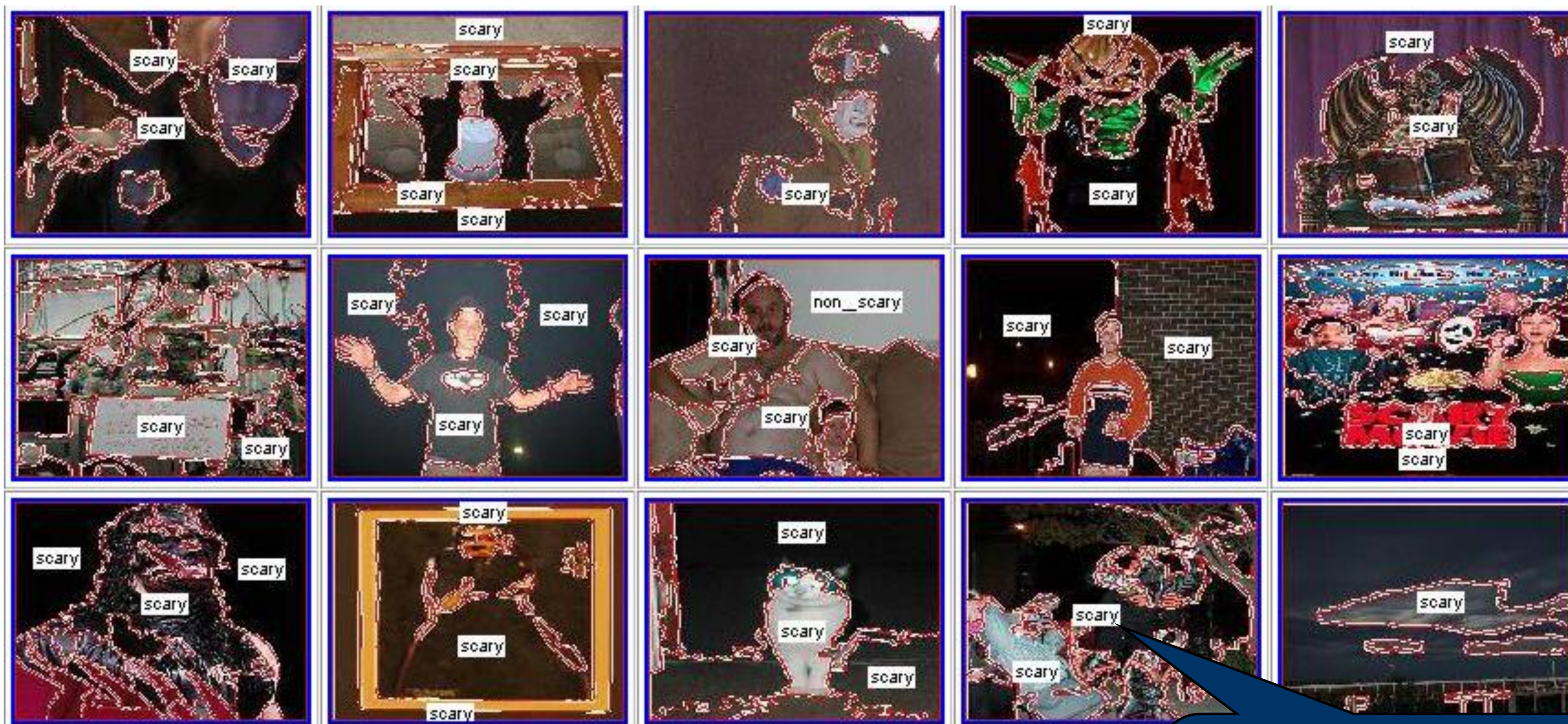
- Entropy-based analysis on "visualness" for 150 adjectives using Web images
  - Use Color, texture and shapes of regions
  - Select relevant regions to the given concepts and calculate entropy with only relevant ones

Translation model-based

- Color names tends to have low entropy.



Low entropy: "scary"



"Visual" adjective

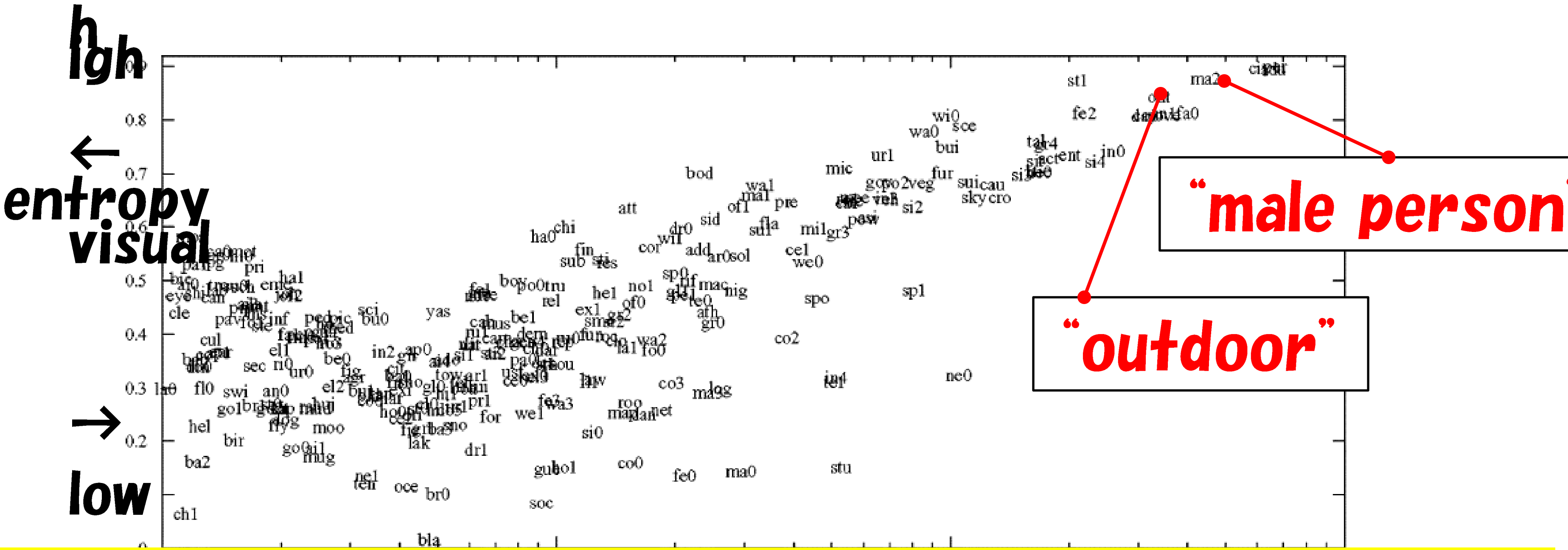
Detected "scary" regions

High entropy: "famous"

"Non-visual" adjective

Concept analysis [Koskela et al. 07]

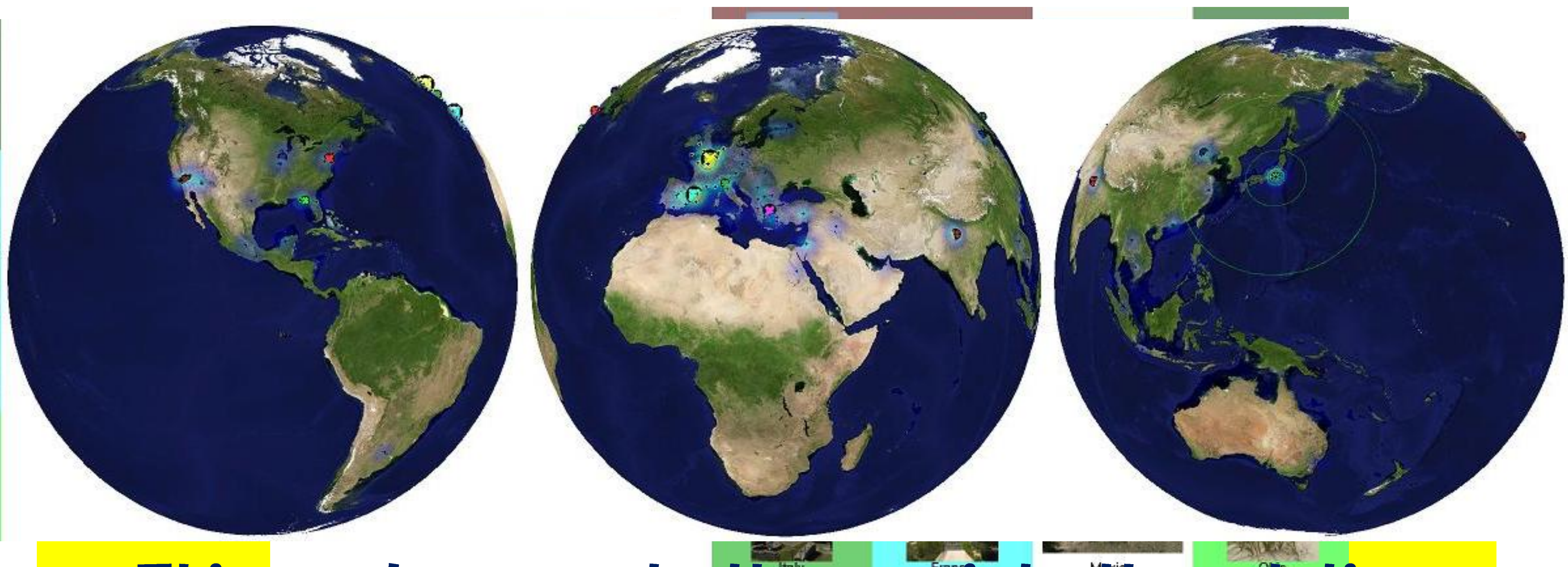
- Entropy-based analysis for 280 LSCOM concepts.
  - Including compound words such as "Asian people"
  - Use color, edge and textures as image features



The frequent concept tends to have high entropy.

"IM2GPS" [Hayes et al. 09]

- Estimate the probability distribution over the world by nearest neighbor search for large-scale geotagged image DB.



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

In this paper

- Entropy-based analysis of the relation between visual features and geo-locations

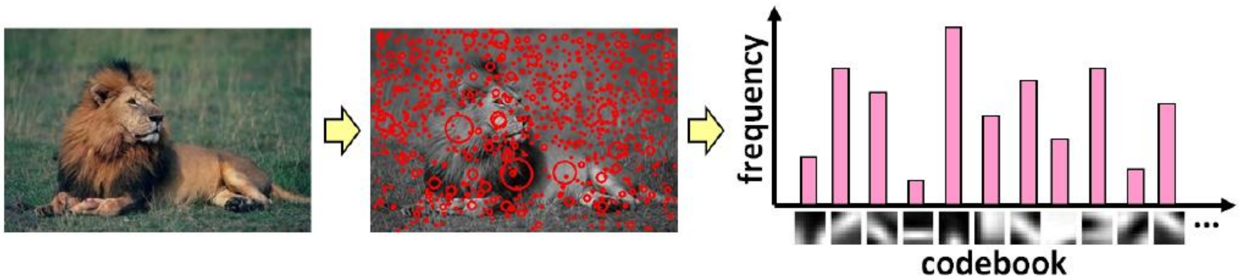
  which was inspired by the following works:
  - Entropy-based visual feature analysis [Yanai et al. 05]
  - Comparison between visual entropy and frequency [Koskela et al. 07]
  - Estimation of geo-location probability by only visual features [Hayes et al. 08]

No work having the same objective so far

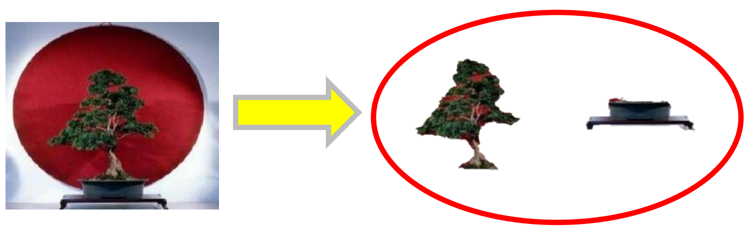
3. Methods

Overview (1): image entropy

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



  - Use mi-SVM to select relevant regions



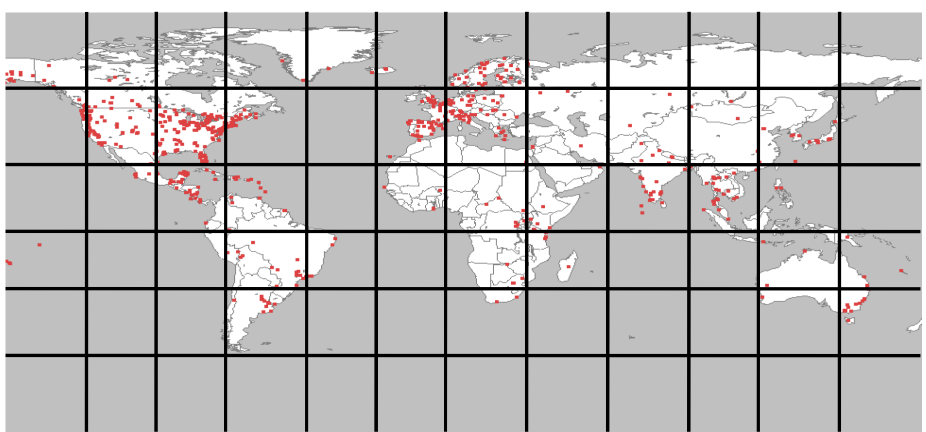
For excluding background and noise regions

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

Overview(2): geo-location entropy

- Collect geotagged images from Flickr via Flickr API

- Calculate geo-location entropy based on region grids of the world map



Method: compute vis-entropy

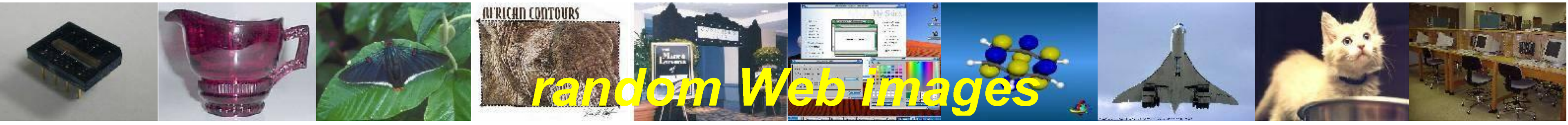
- Collect geotagged images associated with a given word "X" from Flickr using Flickr API
- Carry out region segmentation (JSEG)
- Extract a BoF vector from each region
- Select relevant regions to the given word by mi-SVM
- Estimate the distribution of the BoF vectors of the selected with pLSA
- Calculate entropy of the estimeted distribution with respect to the generic base distribution.

Entropy : how much the distribution of region features is biased compared to the generic distribution of region features

Method: prepare generic model

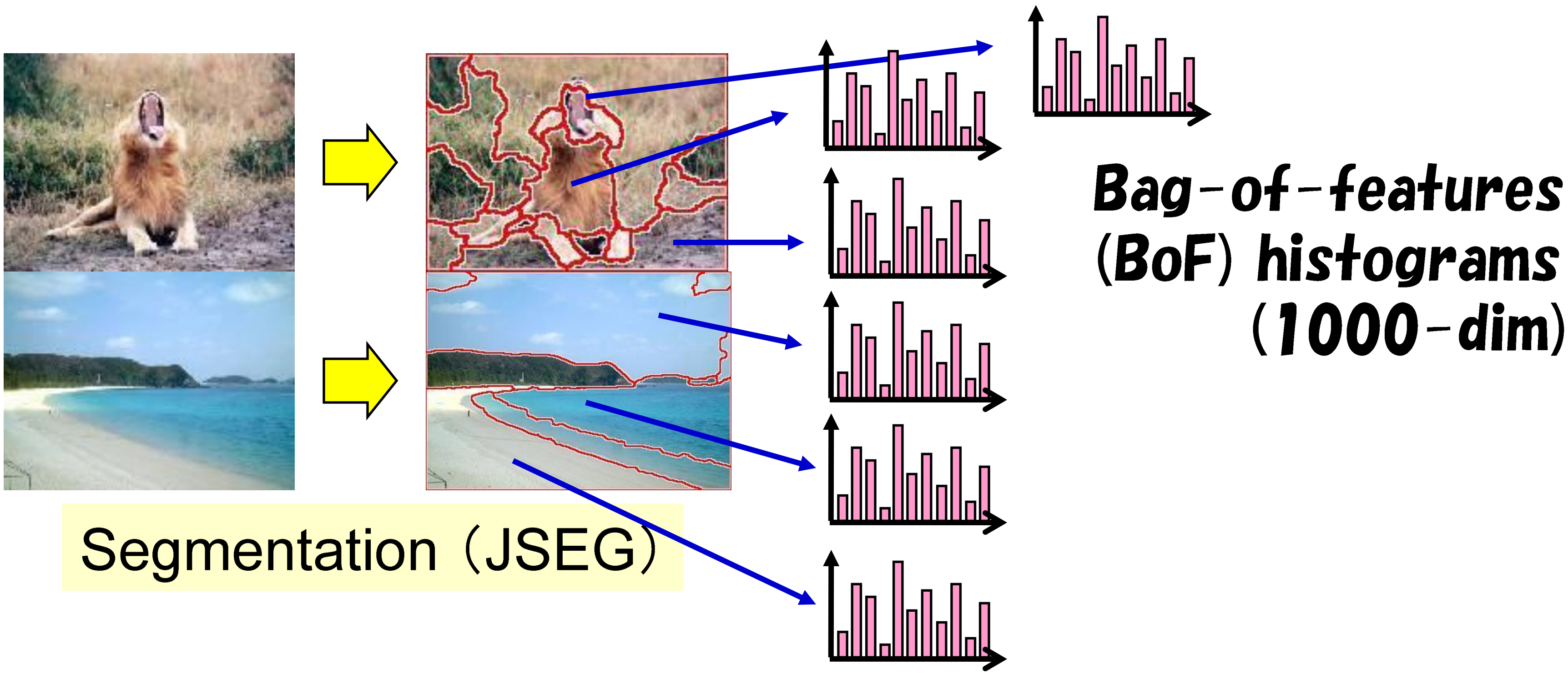
- Calculate the entropy of the "X" regions with respect to the generic base distribution
  - Build a generic distribution model of the region features of randomly collected Web images in advance
  - Use pLSA to model distributions
    - Probabilistic Latent Semantic Analysis [Hofmann 99]

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



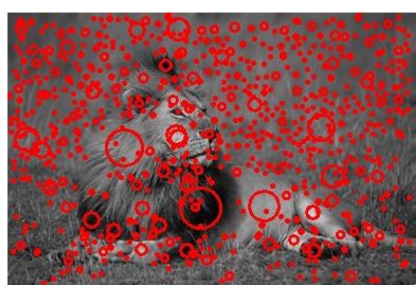
[image representation (1)] Region segmentaion by JSEG

- Divide each image into regions by JSEG (8 regions on the average)
- Extract a BoF vector from each region



[image representation (2)]
Region-based bag-of-features

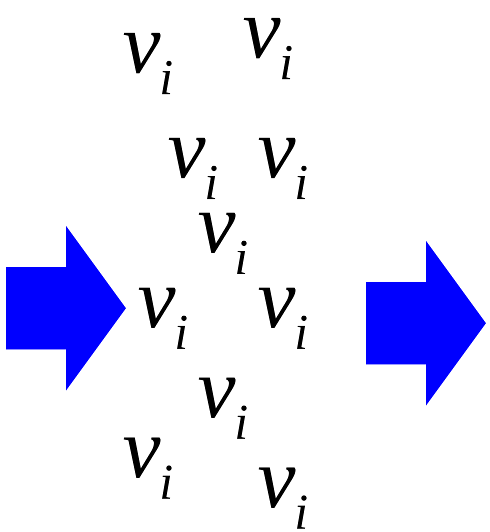
- Represent an image as sets of features
  1. Sample 2000 points randomly
  2. Represent local patterns around sampled points with SIFT descriptor
  3. Vector-quantize SIFT vectors based on pre-computed visual words (codebook(300))



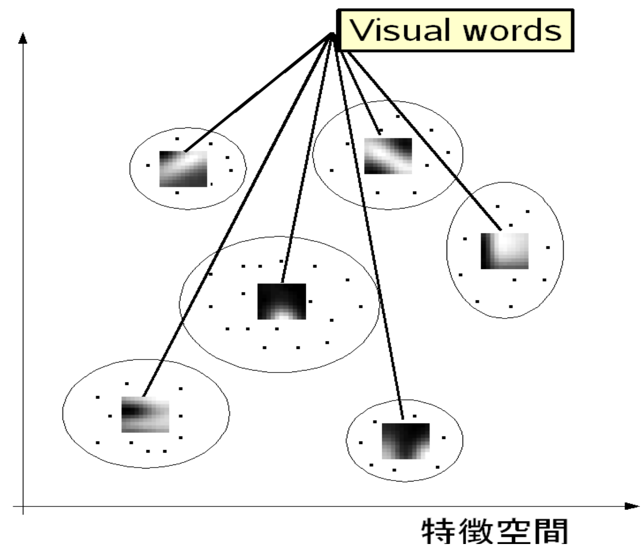
How to obtain visual words

- Extract many SIFT vectors from positive and negative training samples
- Perform k-means clustering

center of clusters are “visual words”.



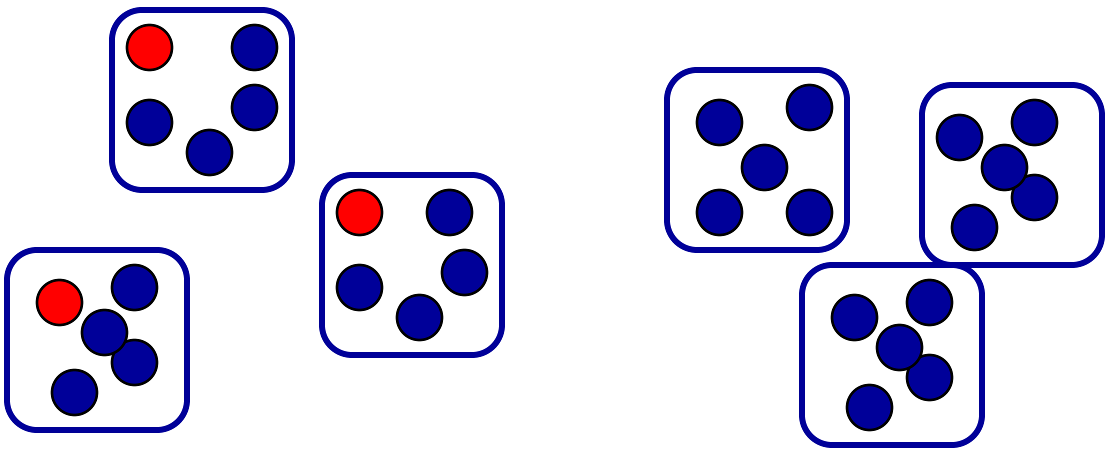
SIFT vectors



“Visual words” are representative local patterns.

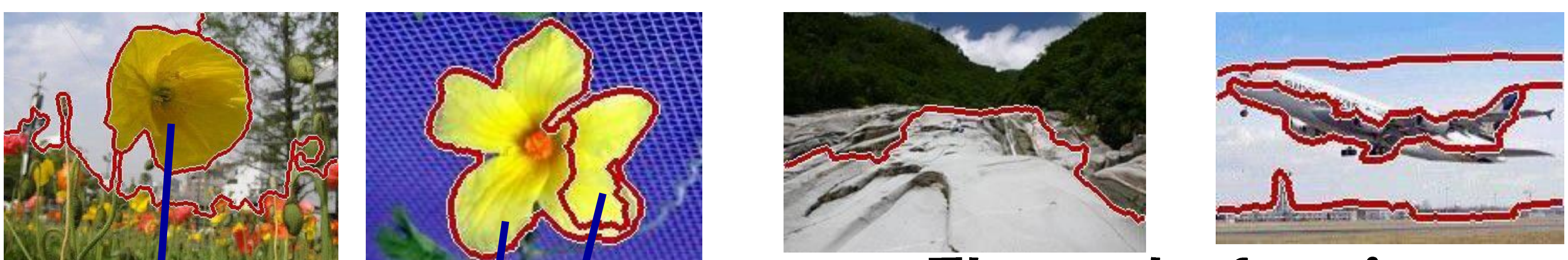
Multiple Instance Setting

- Positive bags / Negative bags



positive ins. (foreground)

negative ins. (background)



Positive instances of "flower"

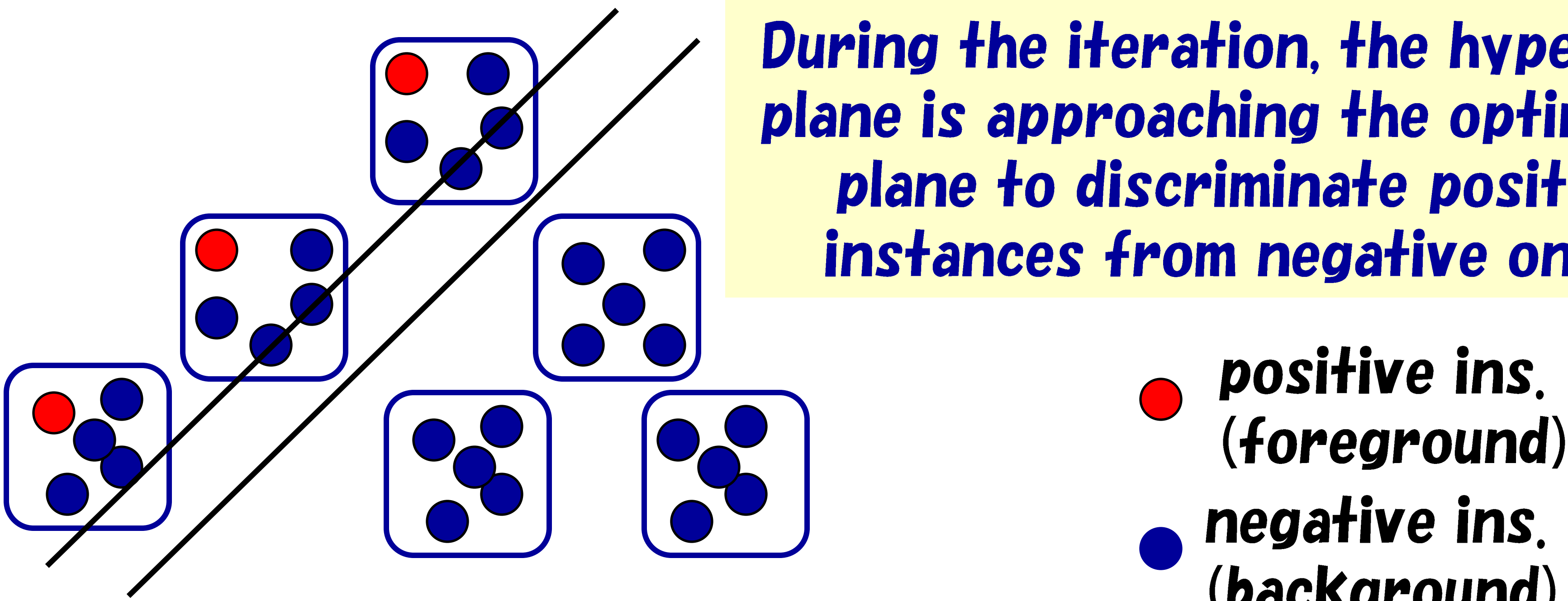
The rest of regions are negative regions.

pseudo-training images

random images

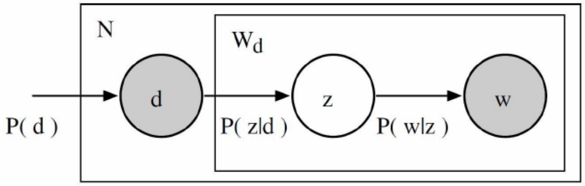
mi-SVM [Andrew et al. NIPS 03]

- 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



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

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

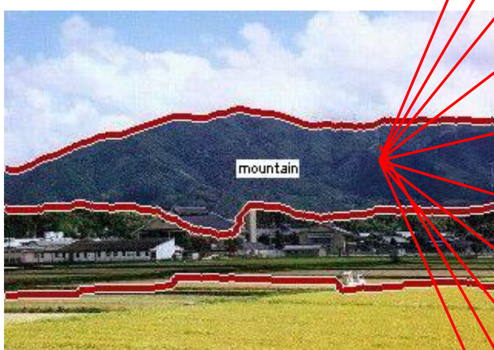
① **Apply PLSA for all the regions of all the random images in advance**

⇨ **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]

Examples of PLSA

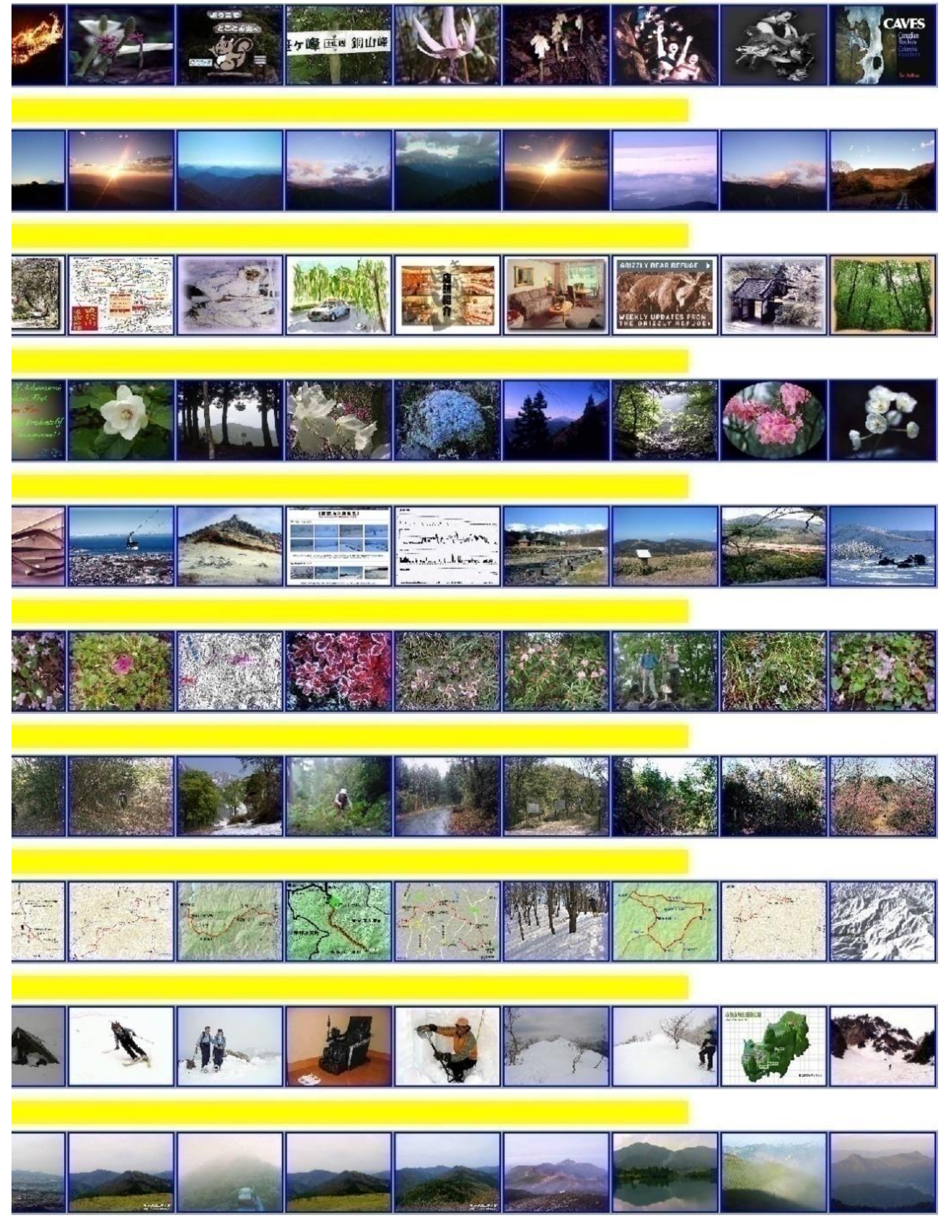
- A region of "Mountain"



mountain

P(topic|region)

0.15
0.24
0.03
0.05
0.10
0.01
0.01
0.01
0.05
0.30



W set the topic number as 300 In the experiment.

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 calculate from each iteration of mi-SVM

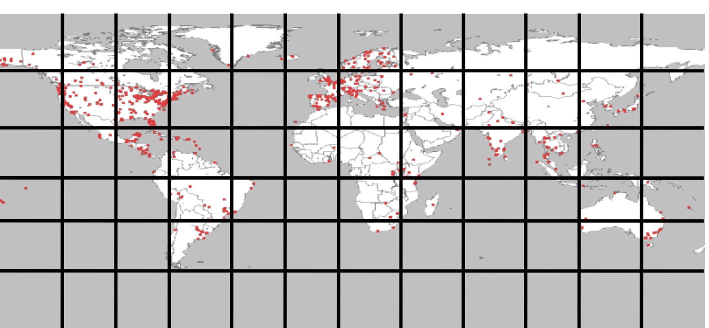
- Regard the minimum H(X) during 5 iterations as the final entropy H(X)

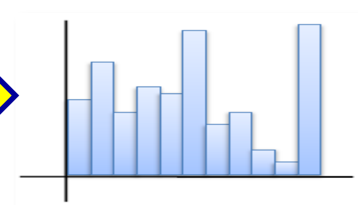
Calculate geo-location entropy

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 geotags of the selected regions
3. Calc entropy
4. Select minimum one

Geo-location entropy

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





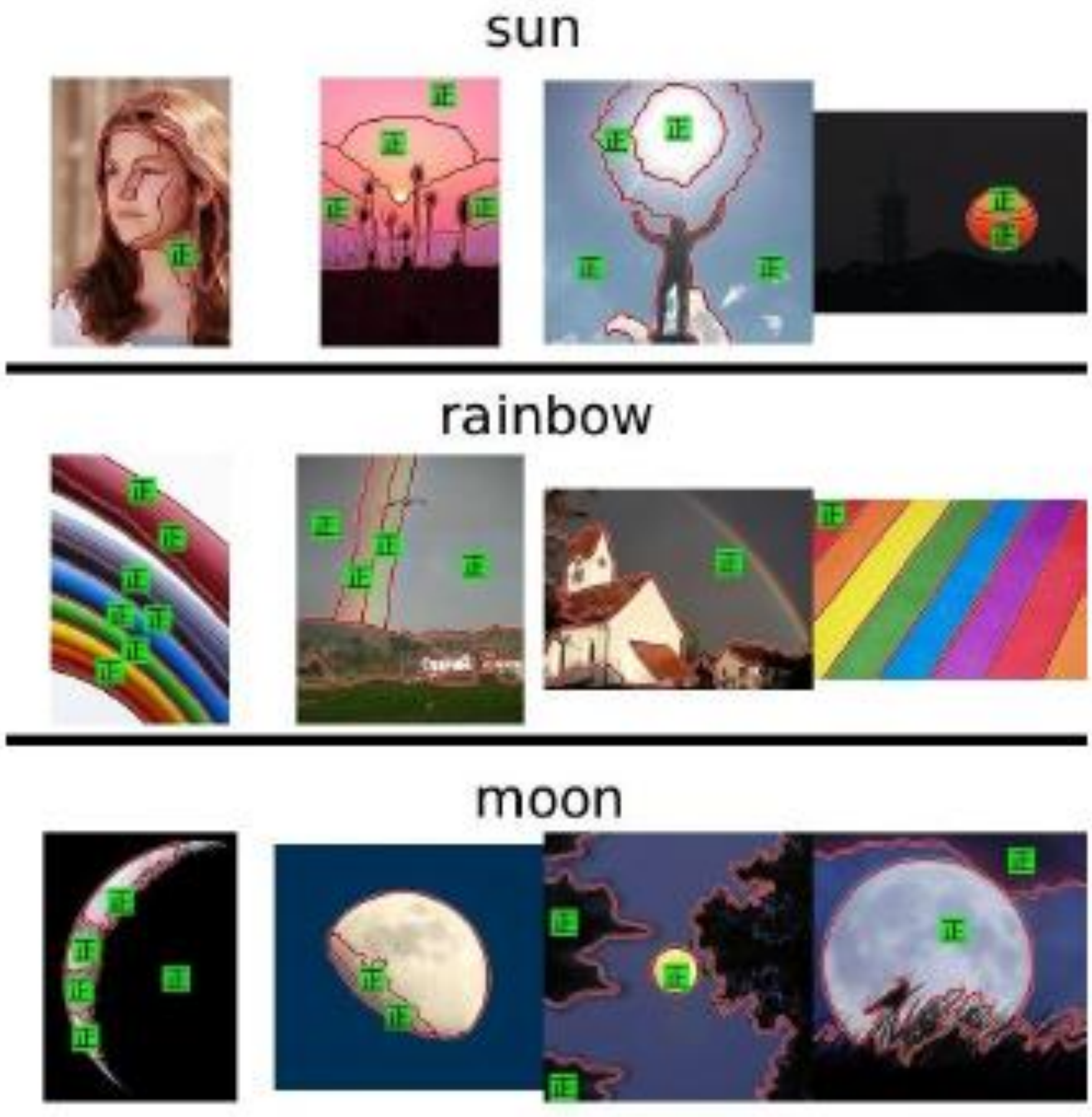
4. Experimental results

Experiments

- Data
  - 230 nouns and 100 adjectives including various kinds of words
  - 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

Results of selection of regions



Results of selection of regions

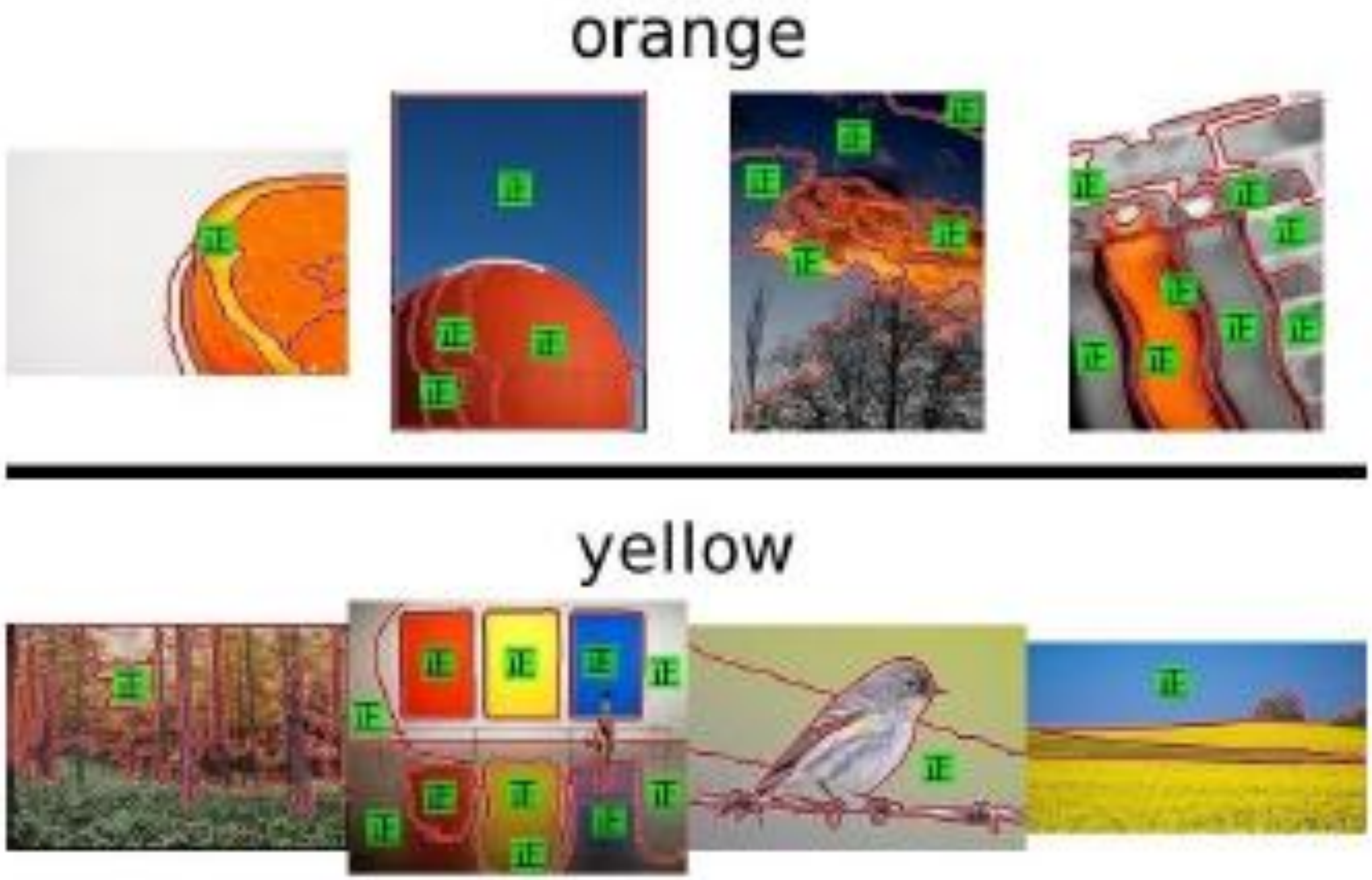


Image region entropy $H_{vis}(X)$ nouns

sun	3.6497	
rainbow	4.5538	
moon	4.6686	
mozart	7.8349	

Image region entropy $H_{vis}(X)$ adj.

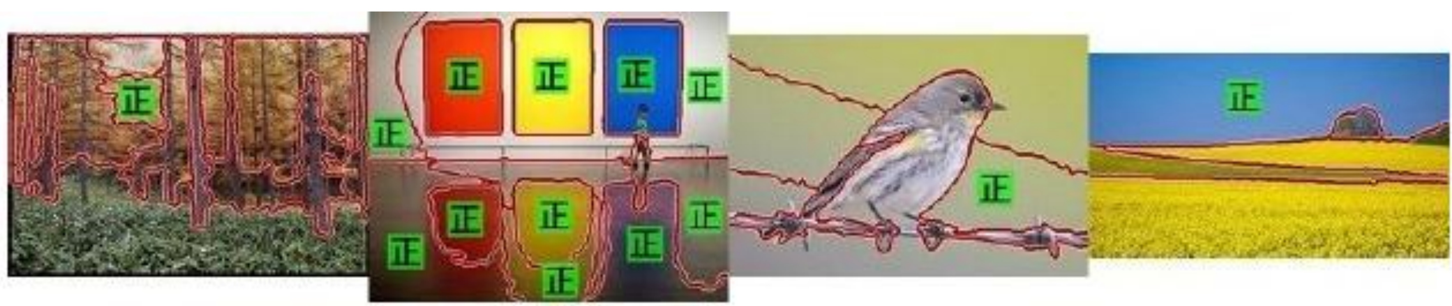
名詞	H(X)	画像の一部
orange	5.5608	
yellow	5.6780	
dark	5.7451	
historic	7.6279	

Results on adjectives

- Adjectives are more abstract than nouns. Their entropy tends to be high.

- ***The entropy of color adjectives are relatively low among the adjectives.***

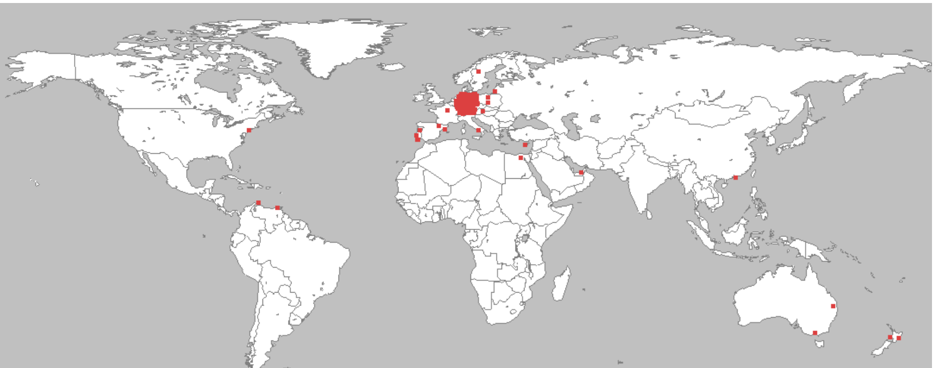
yellow



- ***Since we do not use color features, selected relevant regions are not correct.***

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

Deutschland

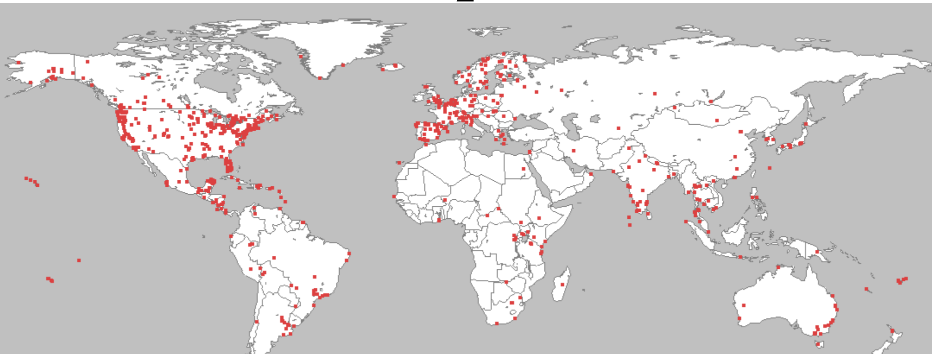


$H_{geo}(X) = 0.2602$

Rome

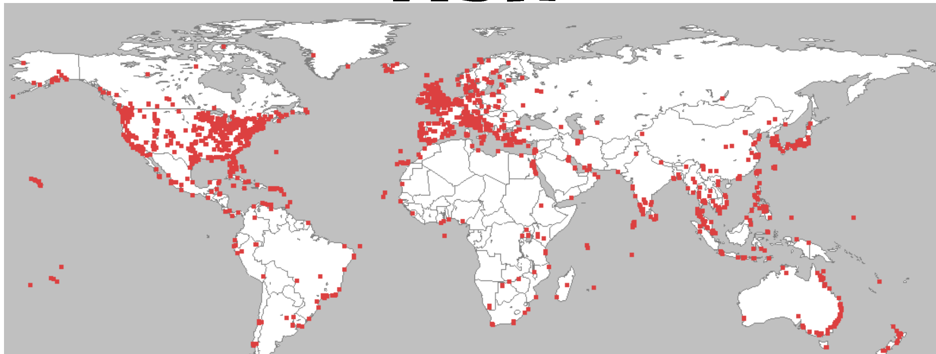
$H_{geo}(X) = 0.3843$

mosquito



$H_{geo}(X) = 5.9759$

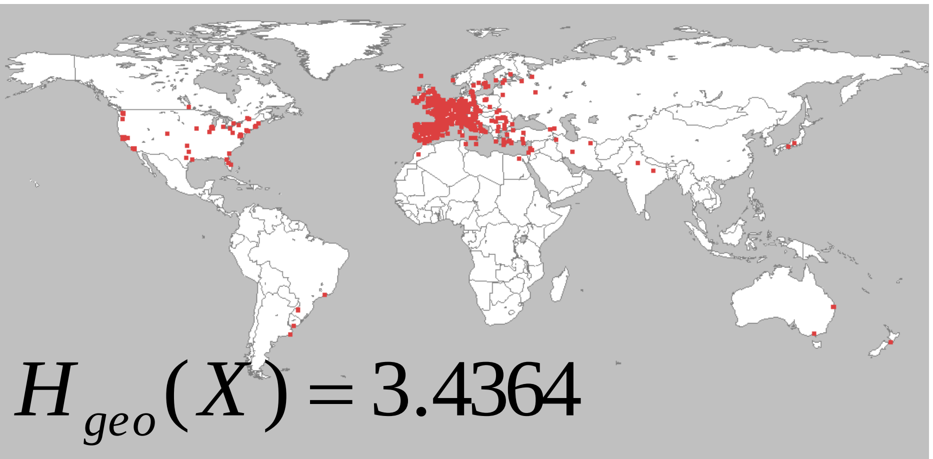
fish



$H_{geo}(X) = 5.7831$

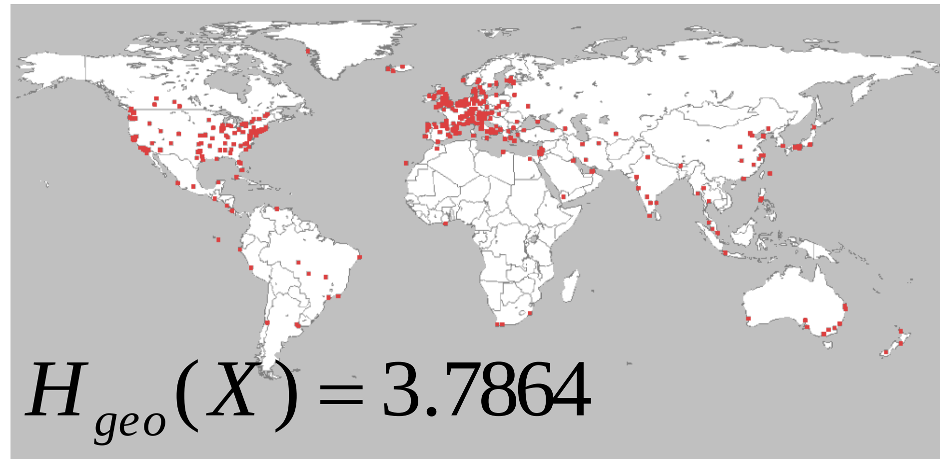
Geo-entropy $H_{geo}(X)$ adjectives

medieval



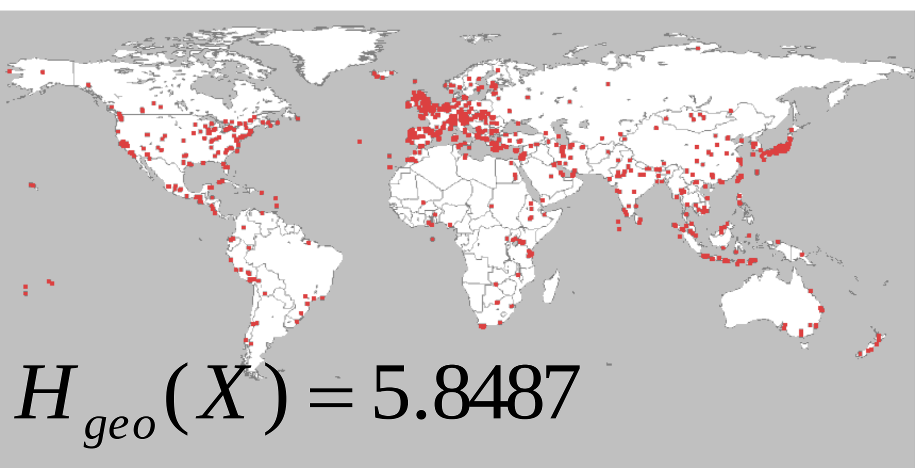
$$H_{geo}(X) = 3.4364$$

modern



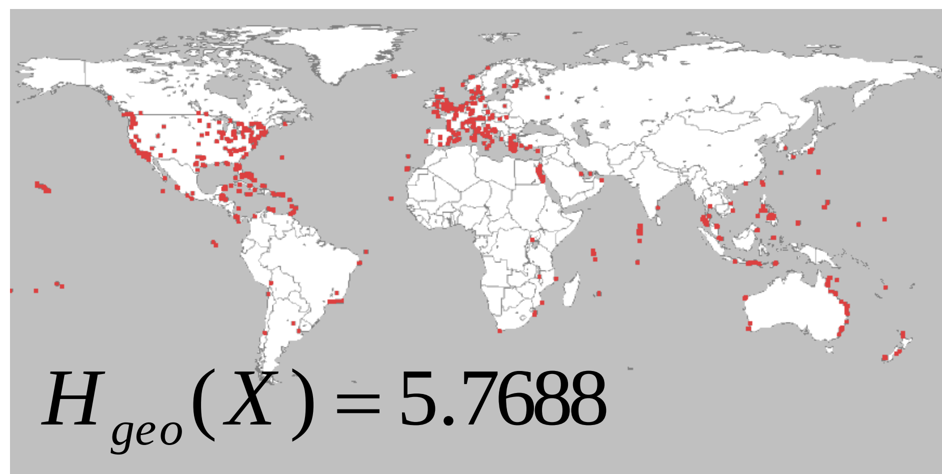
$$H_{geo}(X) = 3.7864$$

traditional



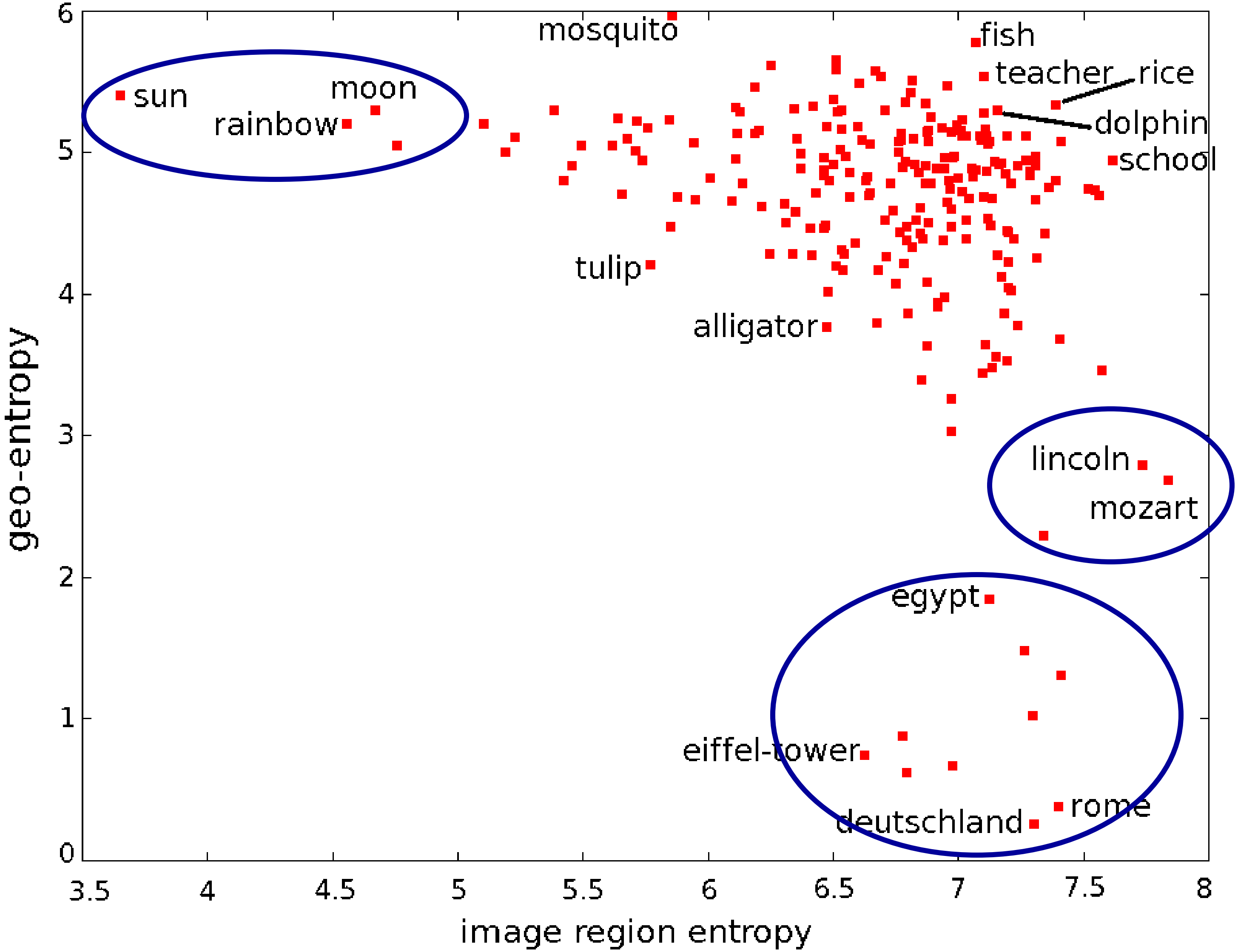
$$H_{geo}(X) = 5.8487$$

underwater



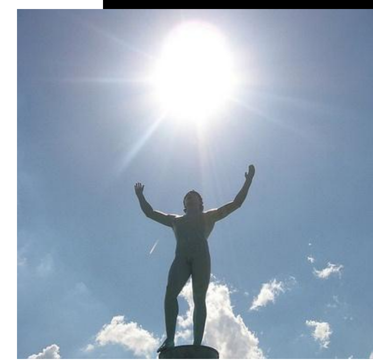
$$H_{geo}(X) = 5.7688$$

Visual entropy vs. geo-entropy [n.]

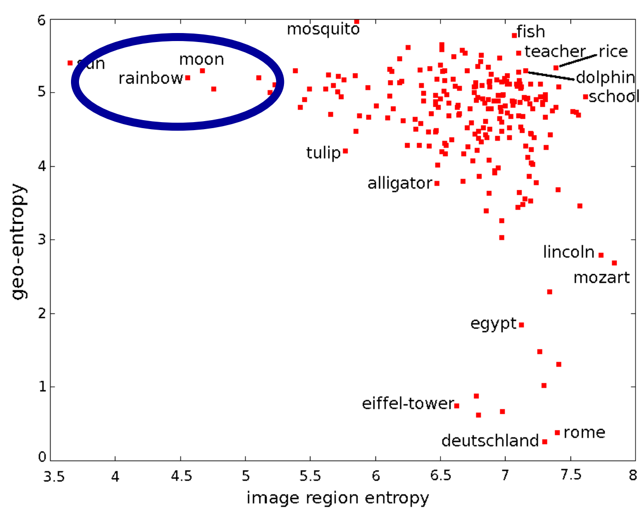


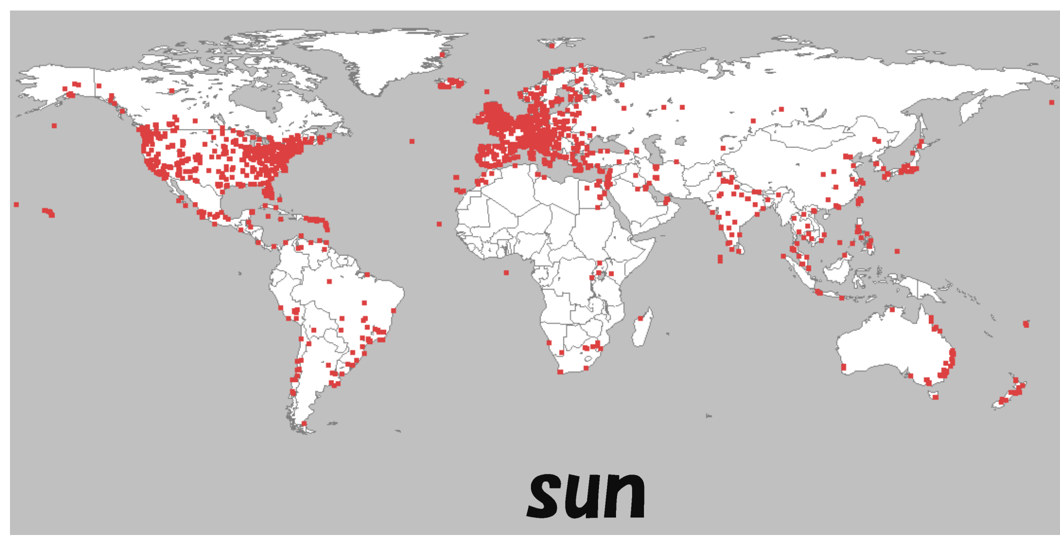
Sun, rainbow, sky

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



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



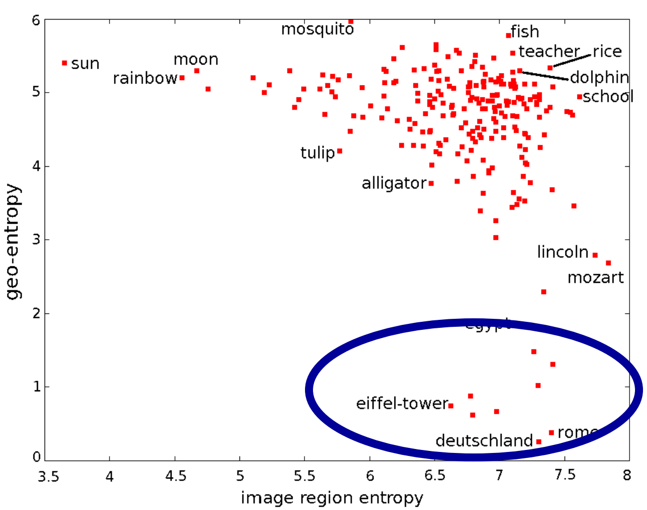


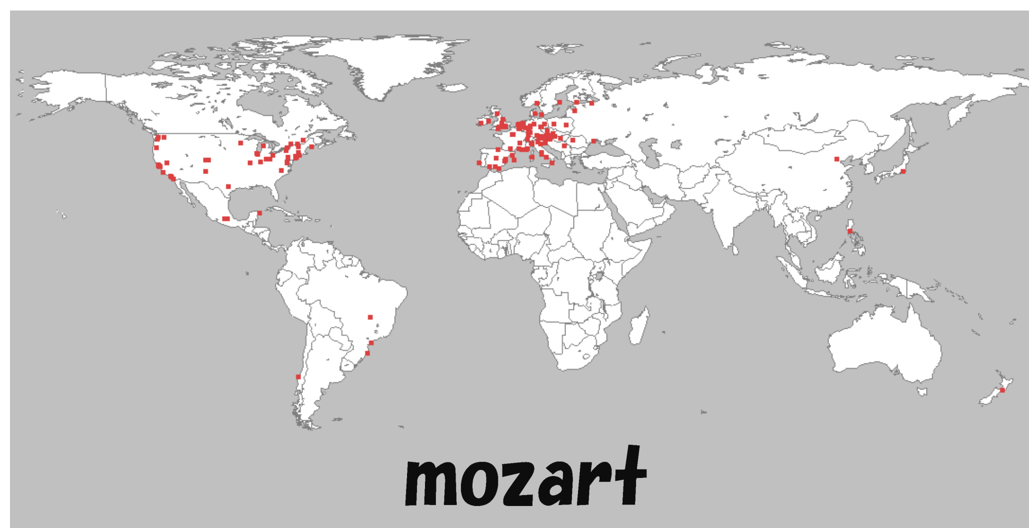
Rome, Deutschland, Mozart

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

Geotags are concentrated
to specific regions.

Appearances are various.





tulip

$H(X)$ = 5.7700 20th/230 in ascending order

$H_{geo}(X)$ = 4.2091 42th/230 in ascending order

- **Variance of color did not reflect on image region entropy.**

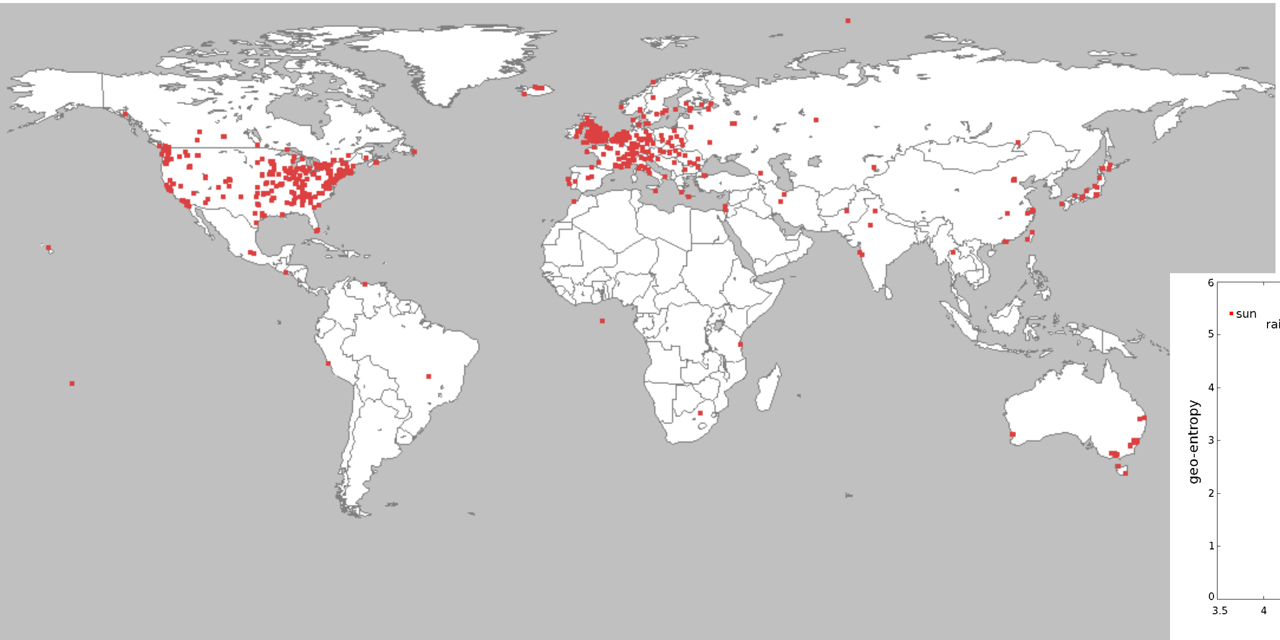
- **Holland and England is main parts**

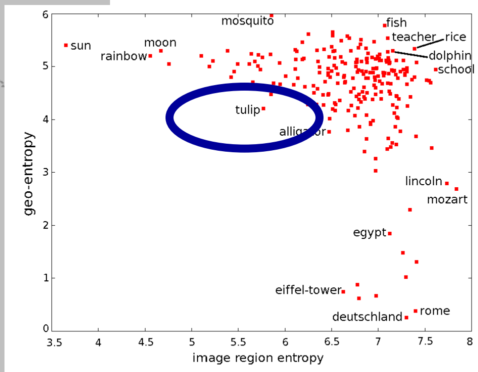








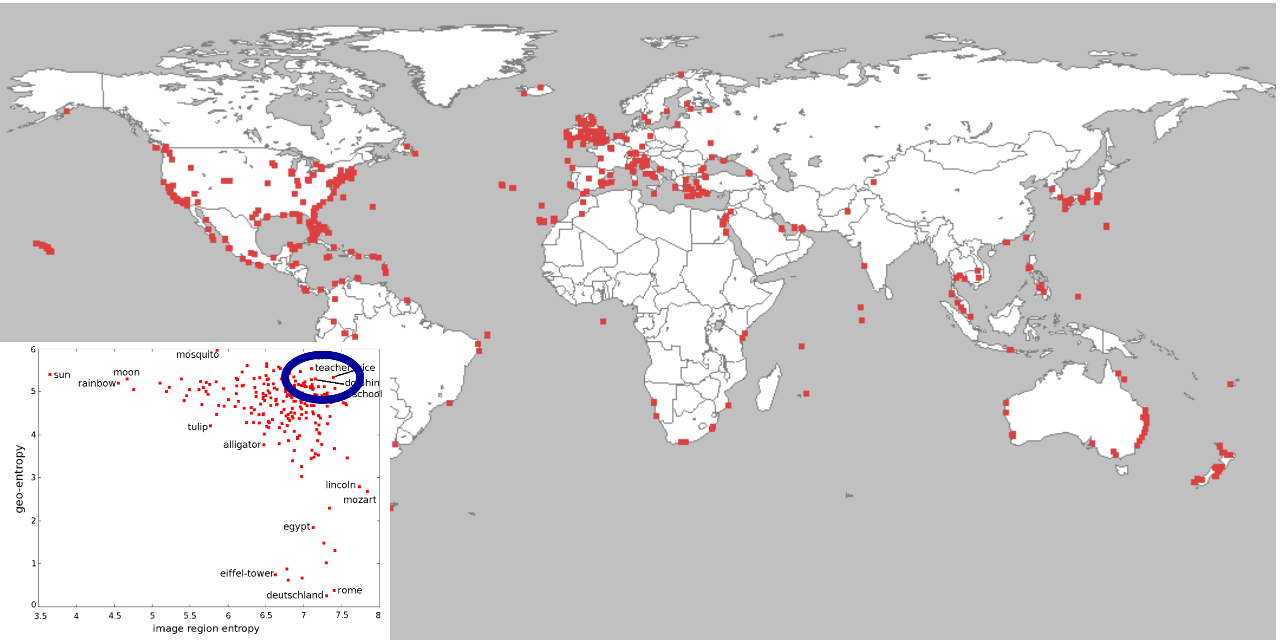


dolphin

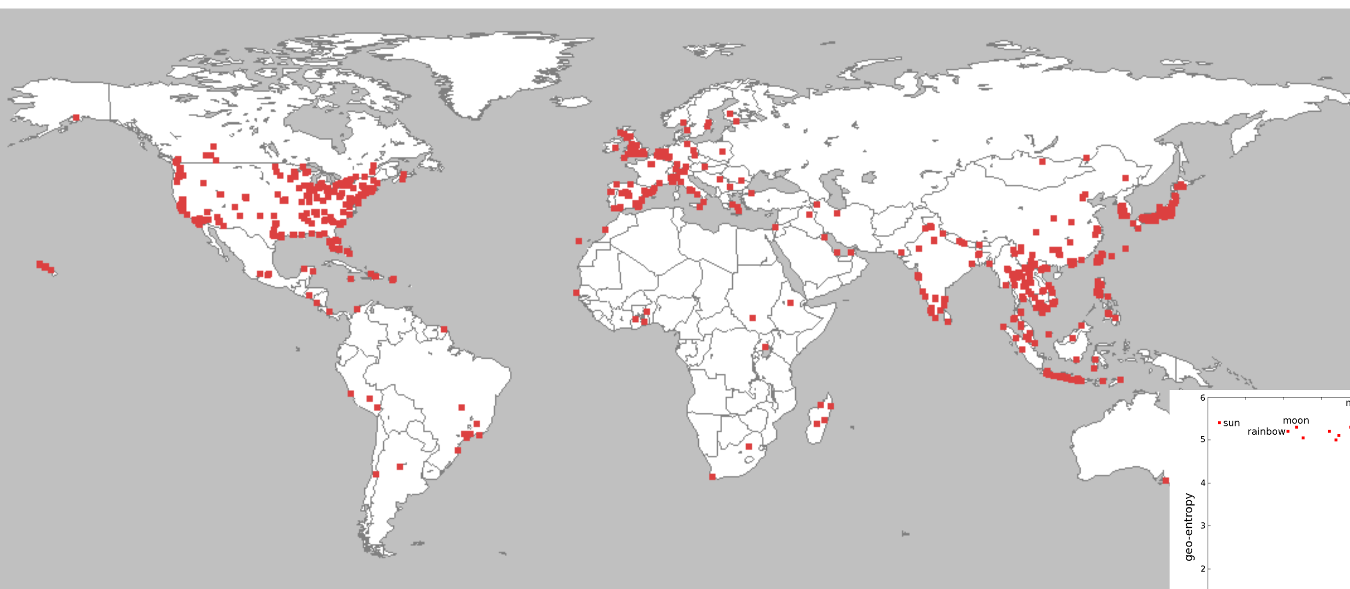
$H(X)$ = 7.1559 42th/230 in descending order

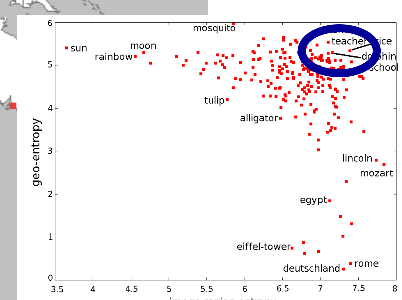
$H_{geo}(X)$ = 5.2981 25th/230 in descending order

- **Most of dolphins are taken in sea or aquarium**
- **In seaside areas**



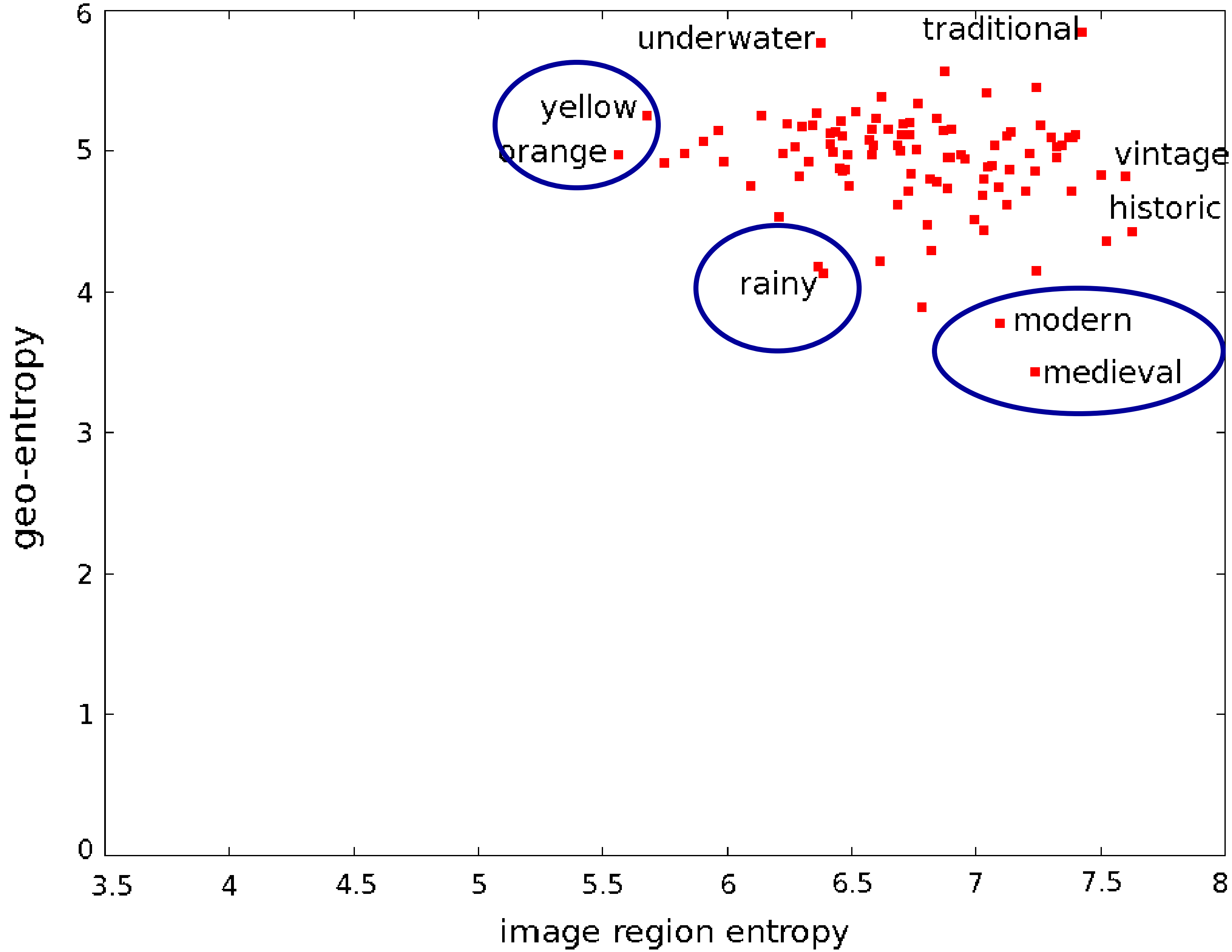
rice





$H_{geo}(X) = 5.3425$

Visual entropy vs. geo-entropy [a.]



Conclusions

Examine the relation between distributions of visual features and geo-locations regarding many concepts (words)

1. Entropy-based measure of visual features
2. Entropy-based measure of geo-locations
3. Analysis the relation between both distributions
   - For 230 nouns and 100 adjectives

Thank you
for your attention !