Automatic Construction of A Folksonomy-based Visual Ontology

Hidetoshi Kawakubo, Yuuta Akima and Keiji Yanai

Department of Computer Science,
The University of Electro-Communications, Tokyo, Japan

December 13, 2010

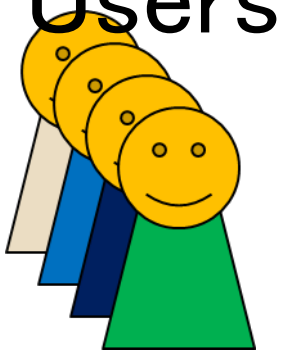
Outline

- Objective
- Related Work
- Proposal method
- Experimental Results
- Conclusions & Future Work

What is Folksonomy？

- Folk + taxsonomy = Folksonomy
  - Folksonomy is Contents taxonomy by Users of social media
  - Users add tags to online contents for all users
  - Online photo sharing site “Flickr” is an famous social media.

Users add tags to



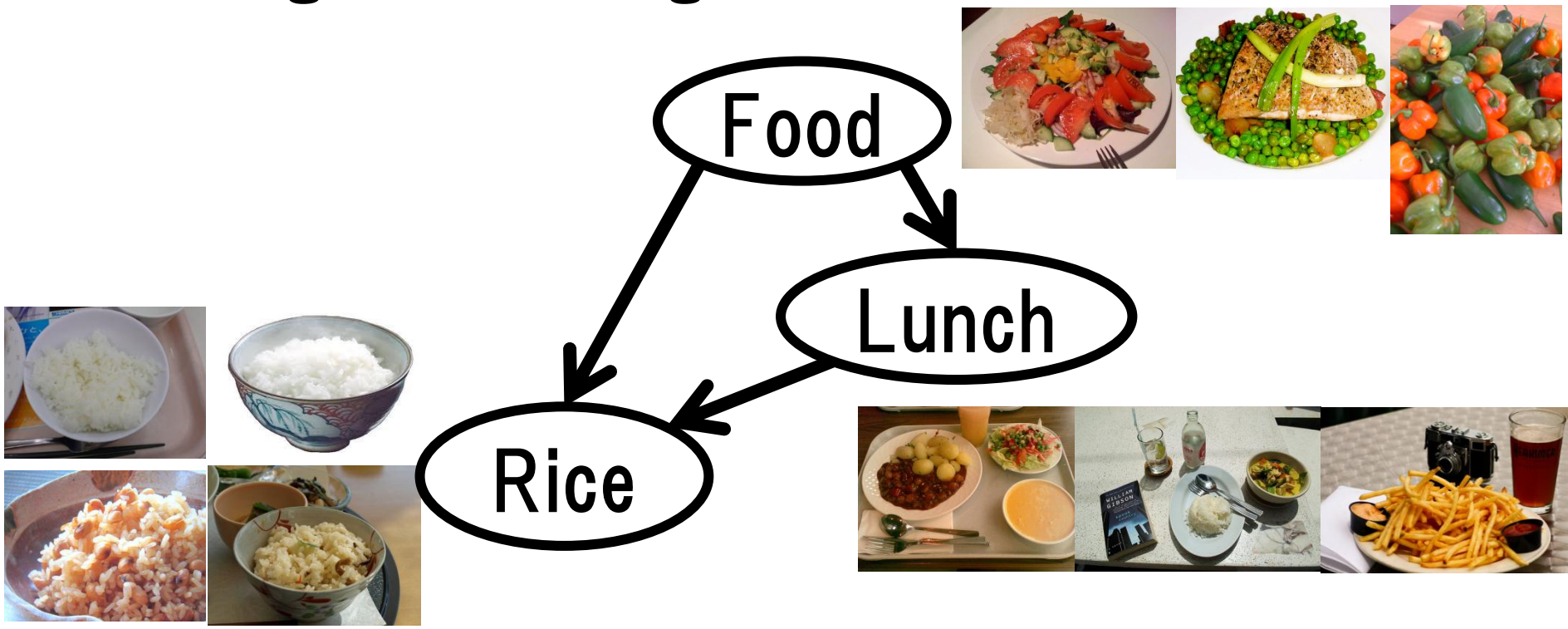
online contents

Background

- A large quantity of Folksonomy images are available on the Web.
- More efficient image retrieval is necessary.
- Ontologies for image retrieval has been studied.

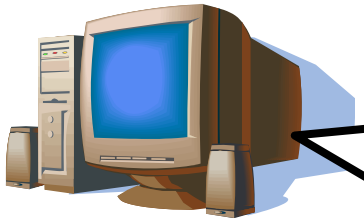
Objective

- Automatic construction of hierarchical image datasets (Visual Ontology) using Flickr images



About “Ontology”

- Ontology is a database to let computers understand relationships between things in the real world.

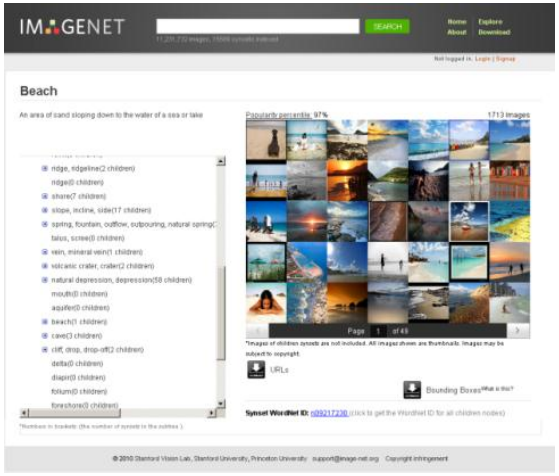


Generally,
Food ⊃ Rice

- We use the term “Ontology” as structures to show relationships of concepts.
- Our “Visual Ontology” contains images of 2657 concepts.

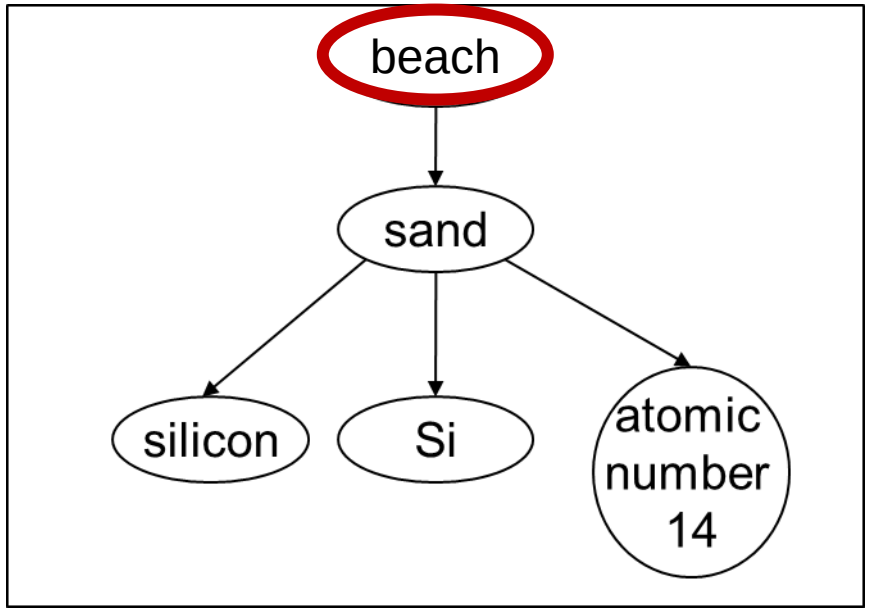
Related Works

- WordNet
  - It is an ontology database built by hand and contains part-of and instance-of relations.
- ImageNet [J. Deng et al. CVPR 2009]
  - It is an image database according to WordNet

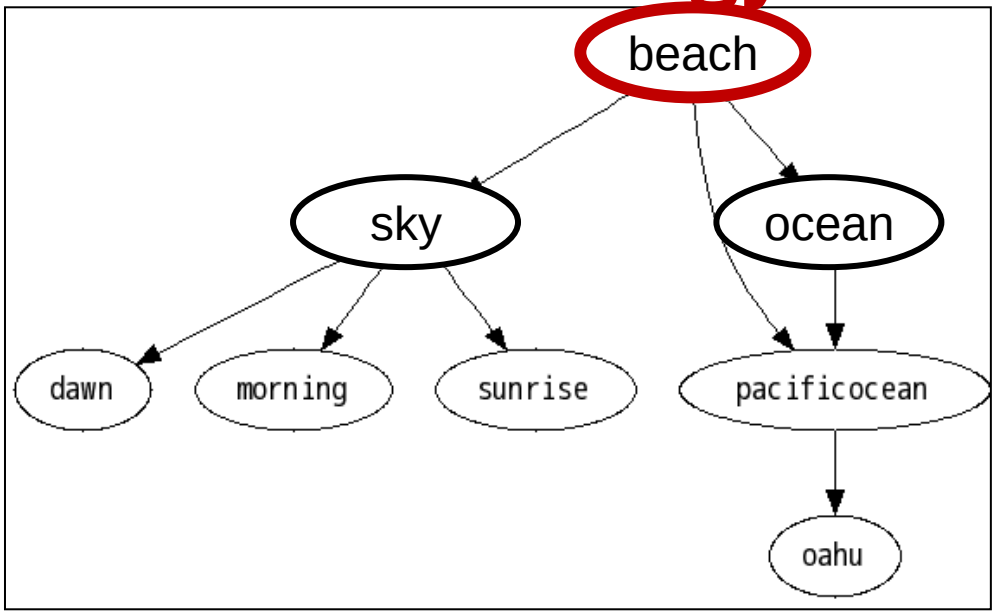


Subordinate structure of “Beach”

WordNet



Our ontology



It includes technical knowledge

It contains more visual concepts and predictably-effective for image retrieval.

Flow of Proposal Method

Generate visual, tag and combined feature vectors, and compute pLSA topic vectors.

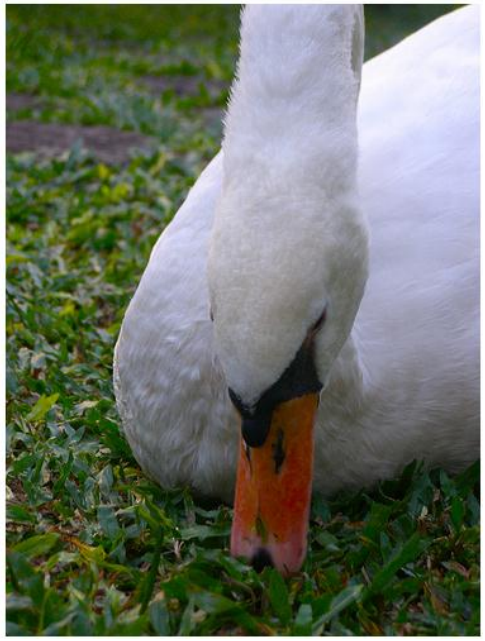
Remove noise images for each concept.

Compute concept vectors, distance between concepts and entropy of each concept.

Build hierarchical structures.

Image collecting from Flickr

- We collected 2 million tagged images randomly from Flickr using Flickr API



tags

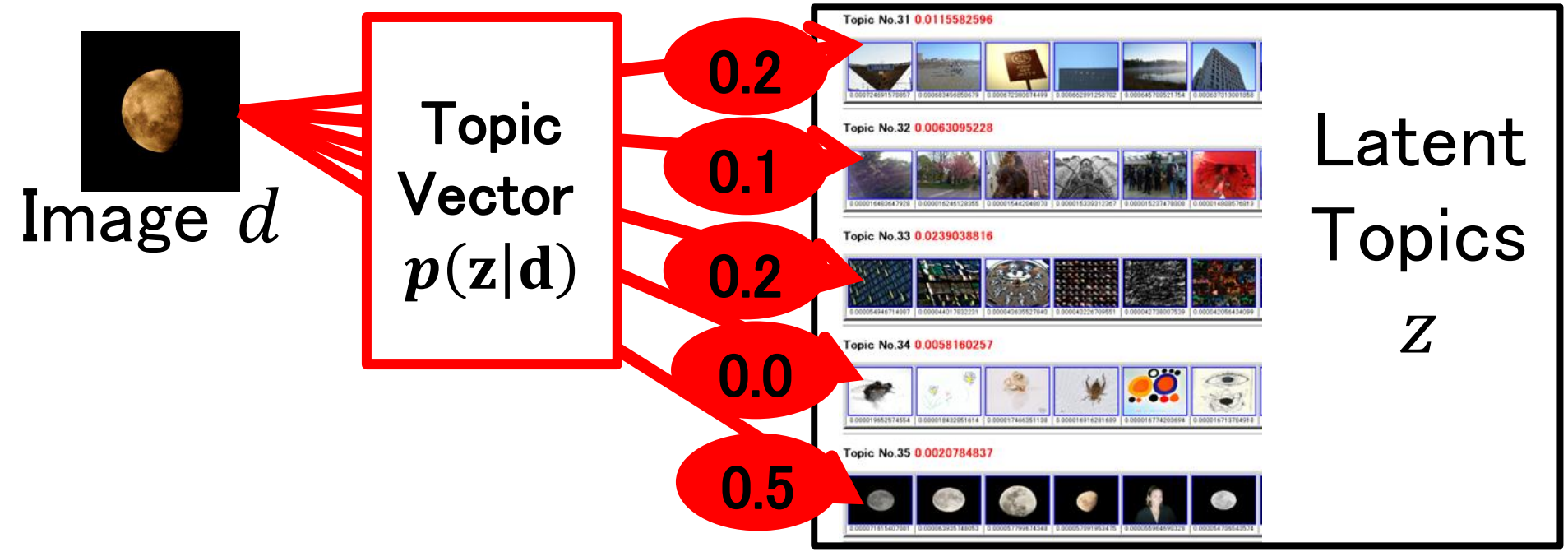
- **bird**
- **goose**
- **swanny**
- **Bali Bird Park**

Features of images

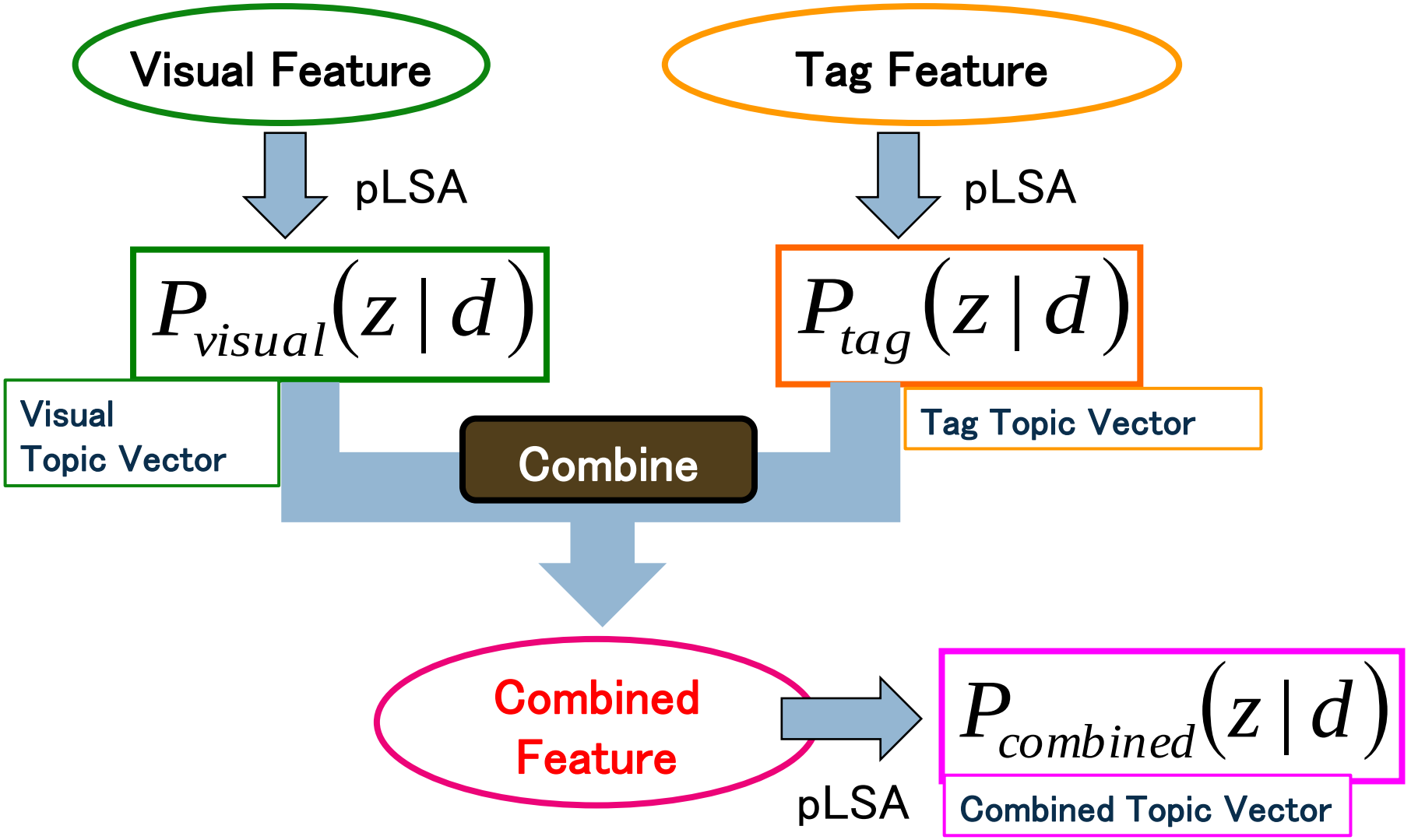
- Visual Feature
  - SIFT + Bag-of-Features, 1000 dimension
- Tag Feature
  - Bag-of-Words (Bag-of-Tags), 4345 dimension
- Combined Feature (Visual + Tag)
  - as we will describe later

probabilistic Latent Semantic Analysis

- We use pLSA to represent distribution of images
  - Each image are represented as a mixture of latent topics.
  - pLSA is a kind of soft clustering.



Feature Combine Method



Topic vectorization of a concept

"moon" concept

$$P(z|Concept) = \frac{\sum_{d \in Concept} p(z|d)}{|Concept|}$$

Average

Distance between concepts

- We use JS divergence between topic vectors as measurement of dis-similarity.

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

$$D_{JS}(P||Q) = \frac{D_{KL}(P||(P/2+Q/2))}{2} + \frac{D_{KL}(Q||(Q/2+P/2))}{2}$$

P, Q are topic vectors of concepts

Hierarchical relations of concepts

- We use entropy of concepts' topic vector to estimate broadness of concepts

$$H(Concept) = -\sum_{z \in Z} P(z|Concept) \log(P(z|Concept))$$

Animal
(Broader Concept)

Bird
(Subordinate Concept)

Construction of hierarchical structures

- We use DAG (Directed Acyclic Graph)

For each concepts,

Step1. select 20 neighborhood concepts.

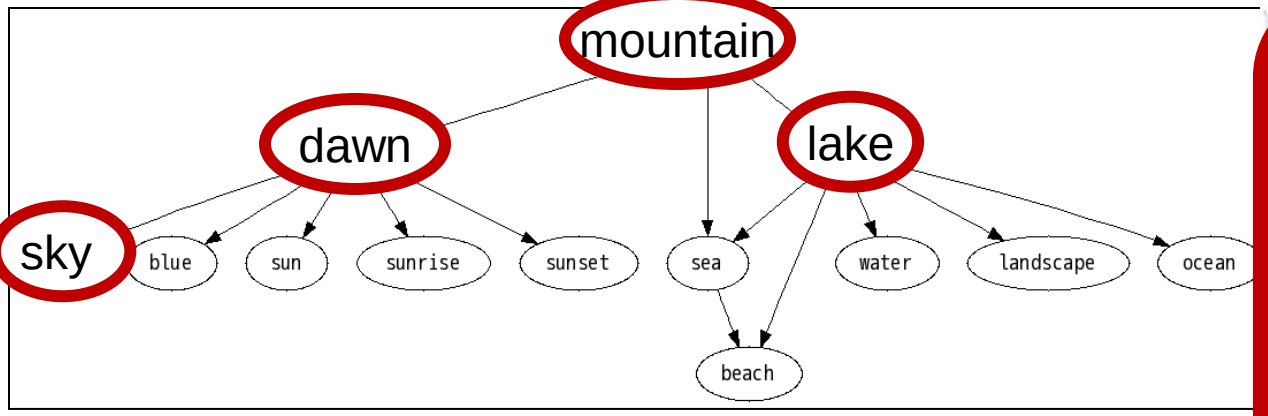
Step2. compare entropy value and classify
the concepts with smaller entropy as child nodes
the concepts with larger entropy as parent nodes.

Experiments

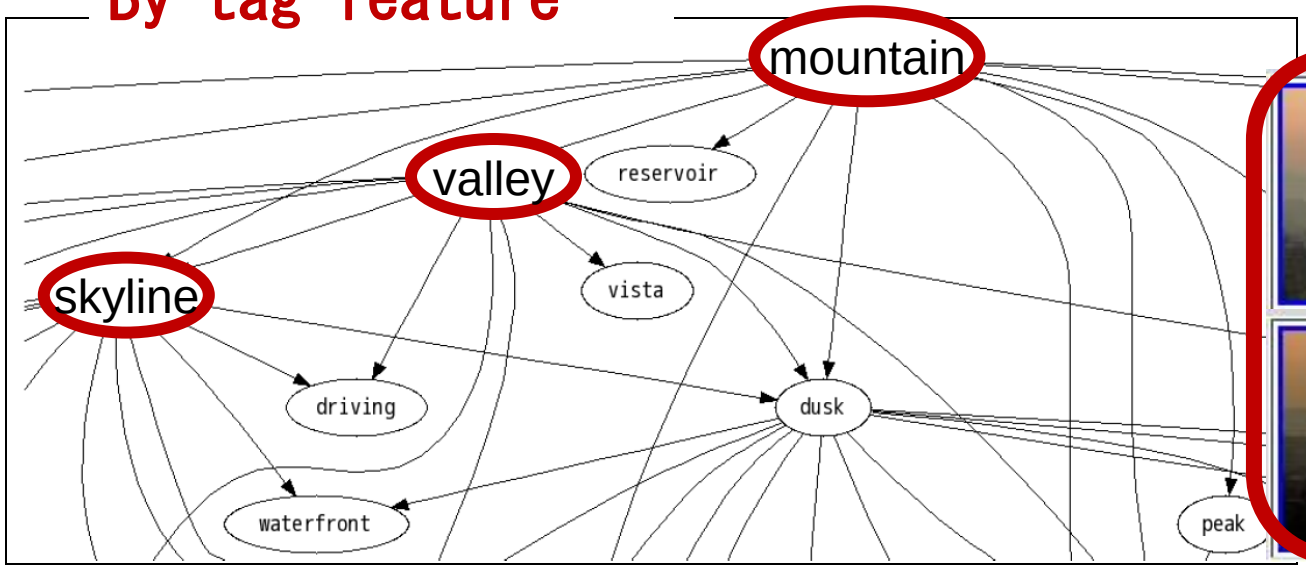
- We built hierarchical concept structures for 2,657 concepts.
- We show several results and consider differences about feature kinds.

Subordinate structure of “mountain”

By visual feature

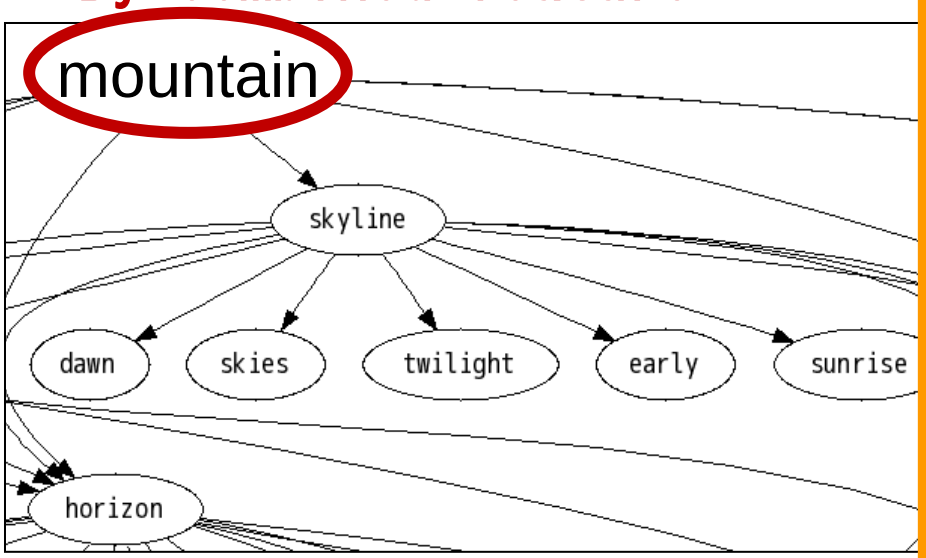


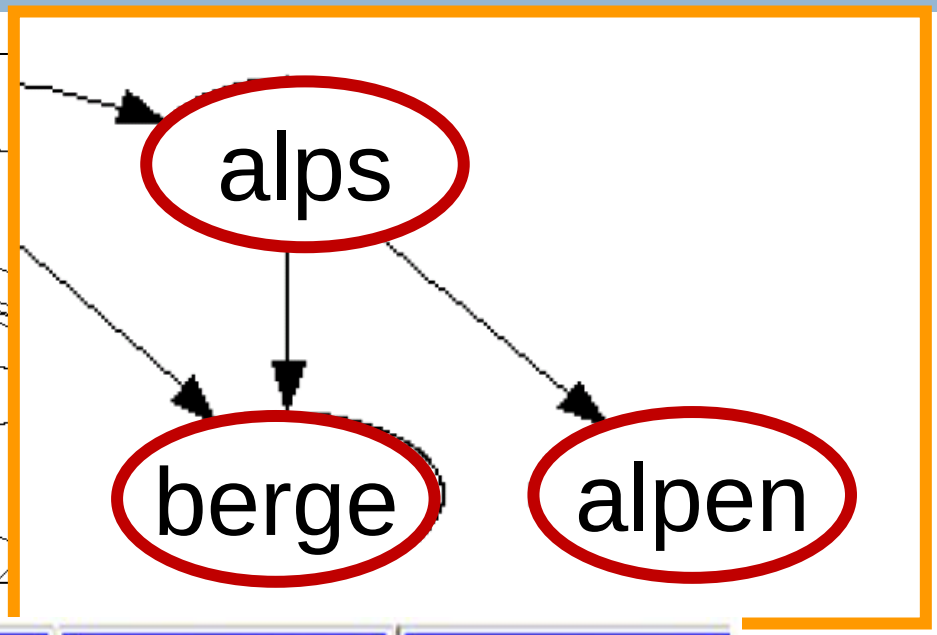
By tag feature



Subordinate structures of “mountain”

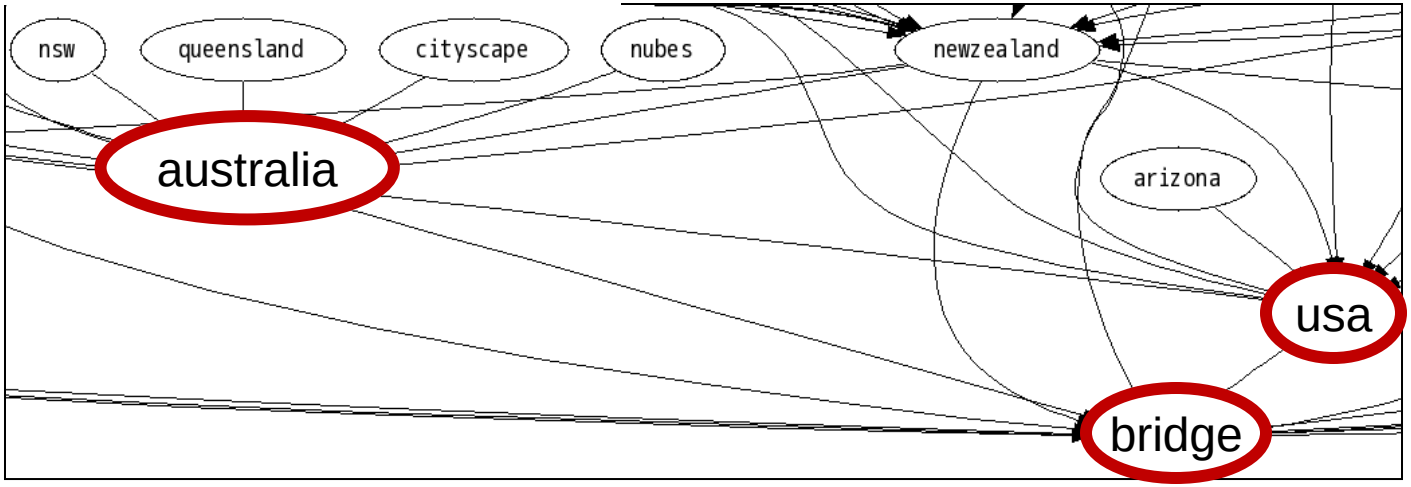
By combined feature





Superordinate structure of “bridge”

By visual feature

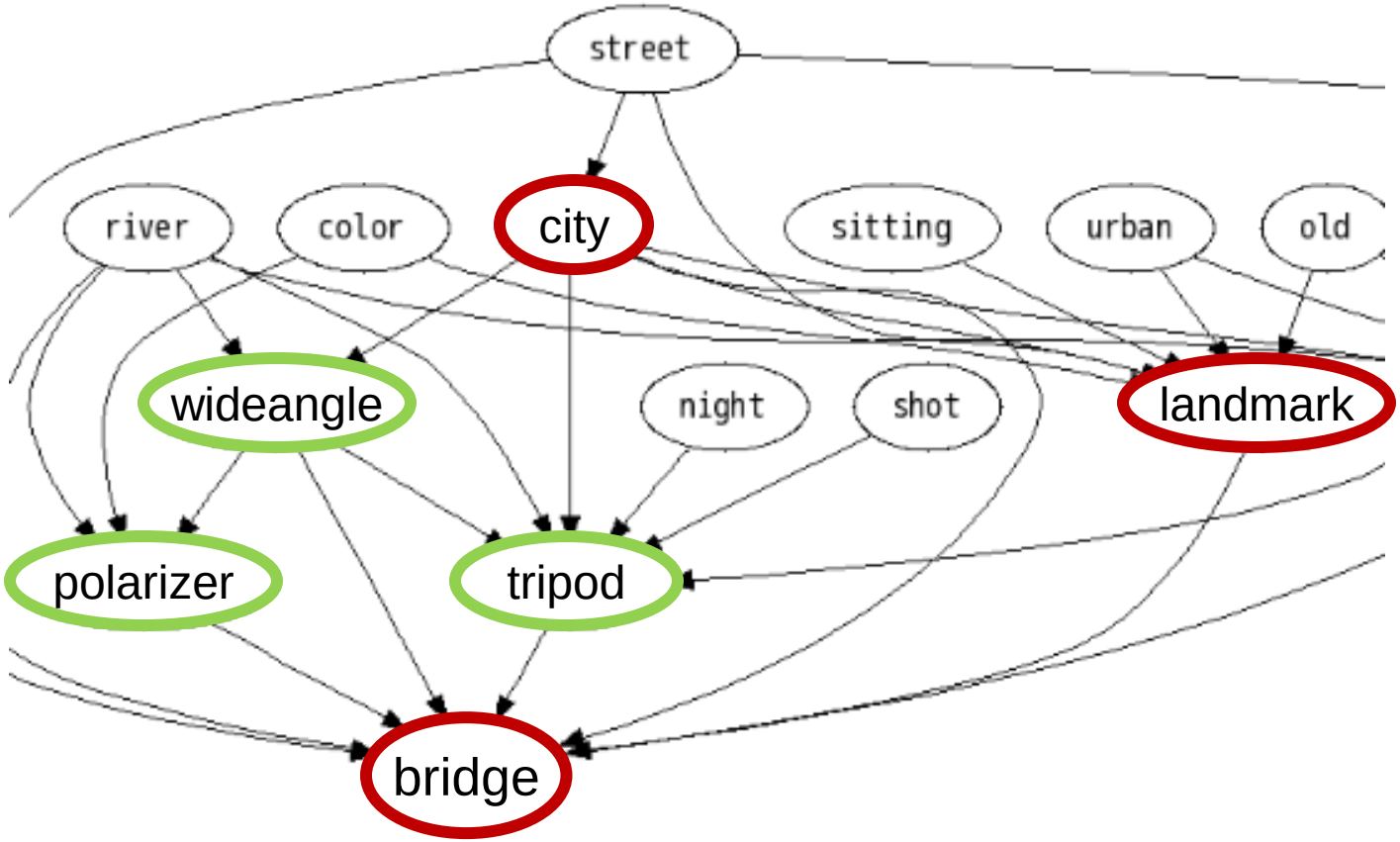


usa

bridge

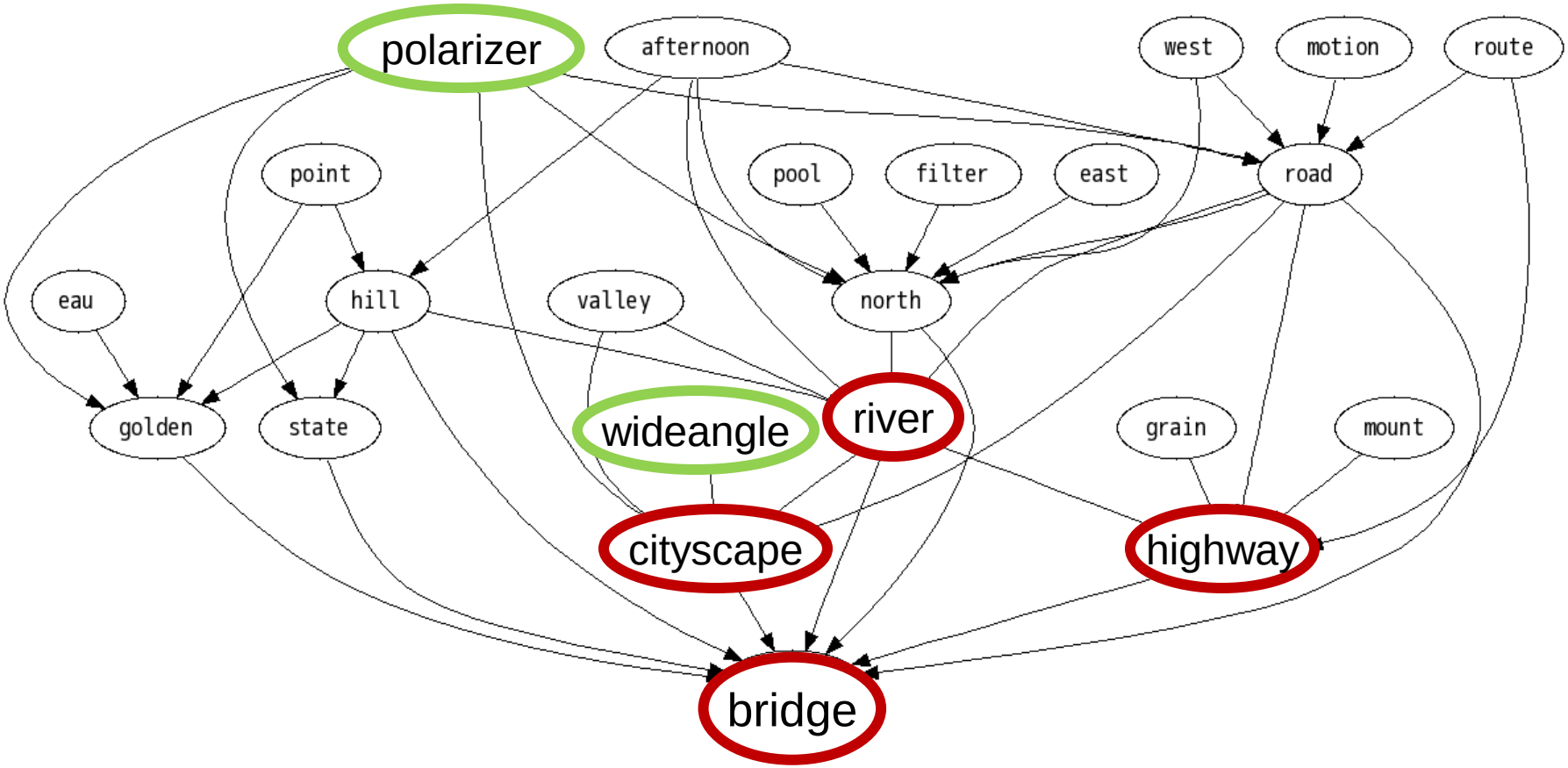
Superordinate structure of “bridge”

By tag feature

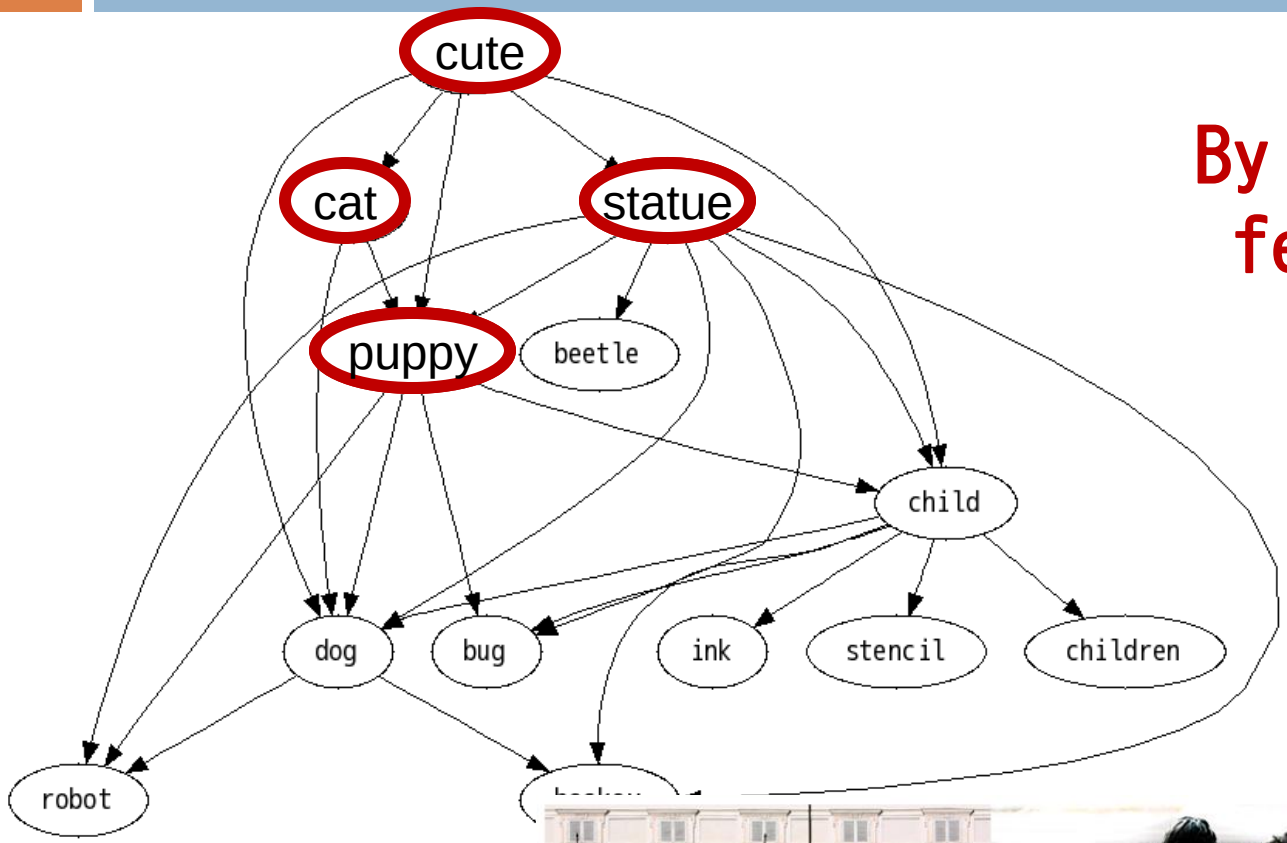


Superordinate structure of “bridge”

By combined feature



Subordinate structure of “cute”



By visual feature

statue

Conclusions & Future Work

- Conclusions
  - We proposed an automatic construction method of visual ontology using online tagged images.
  - Ontology by combined features is better than ontology by only visual features and by only tag features.
- Future Work
  - We plan to Improve the proposed method.
  - Quantitative evaluation of method is yet.

Thank you

APPENDIX

Removal of Noise Images

$$P_{Concept}^{z_k} = \sum_{d \in Concept} P(d|z_k) \ / \ |Concept|$$

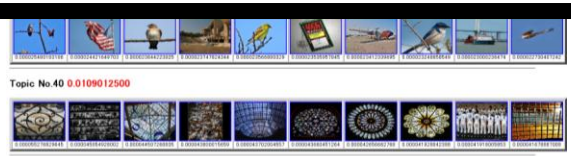
$P_{Concept}^{z_k}$

$$P(Concept|z_k) = \frac{P_{Concept}^{z_k}}{P_{Concept}^{z_k} + P_{NotConcept}^{z_k}}$$

pLSA

Not “moon” Image set

$P_{NotConcept}^{z_k}$



$$P_{NotConcept}^{z_k} = \sum_{d \in NotConcept} P(d|z_k) \ / \ |NotConcept|$$

Removal of Noise Images

$$P(Concept|d_i) = \sum_{k=1}^{K} P(Concept|z_k)P(z_k|d_i)$$