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 + taxonomy = 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.

- We use the term “Ontology” as structures to show relationships of concepts.

- Our “Visual Ontology” contains images of 2657 concepts.

Generally, Food ⊃ Rice
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**

- beach
  - sand
    - silicon
    - Si
    - atomic number 14

**Our ontology**

- beach
  - sky
    - dawn
    - morning
    - sunrise
  - ocean
    - pacific ocean
    - oahu

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.
We collected 2 million tagged images randomly from Flickr using Flickr API.
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
We use pLSA to represent distribution of images.

- Each image is represented as a mixture of latent topics.
- pLSA is a kind of soft clustering.
Feature Combine Method

- **Visual Feature**
  - $P_{\text{visual}}(z | d)$
  - pLSA
  - Visual Topic Vector

- **Tag Feature**
  - $P_{\text{tag}}(z | d)$
  - pLSA
  - Tag Topic Vector

- **Combine**
  - Combined Feature
  - $P_{\text{combined}}(z | d)$
  - pLSA
  - Combined Topic Vector
Topic vectorization of a concept

\[ P(z|\text{Concept}) = \frac{\sum_{d \in \text{Concept}} p(z|d)}{|\text{Concept}|} \]
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

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

Animal (Broader Concept)

Bird (Subordinate Concept)

Larger Entropy

Smaller Entropy
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

- mountain
- dawn
- lake
- sky
- blue
- sun
- sunrise
- sunset
- sea
- water
- landscape
- ocean
- beach

By tag feature

- mountain
- dawn
- valley
- reservoir
- vista
- driving
- waterfront
- dusk
- peak

Images:
- Mountain
- Dawn
- Skyline
- Lake
Subordinate structures of “mountain”

By combined feature

mountain

alps

berge

alpen

mountain

alps
Superordinate structure of “bridge”

By visual feature
Superordinate structure of “bridge”

By tag feature

[Diagram showing relationships between various tags and features related to bridge.]
Superordinate structure of “bridge”

By combined feature
Subordinate structure of “cute” by visual feature
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

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

\[ P(\text{Concept}|z_k) = \frac{P_{\text{Concept}}^{z_k}}{P_{\text{Concept}}^{z_k} + P_{\text{NotConcept}}^{z_k}} \]

Not “moon” Image set

\[ P_{\text{NotConcept}}^{z_k} = \sum_{d \in \text{NotConcept}} P(d|z_k) / |\text{NotConcept}| \]
Removal of Noise Images

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