Word2Image: Towards Visual Interpretation of Words

Haojie Li
National University of Singapore

Table of Contents

- Introduction
- Discovering Semantic and Visual Diversity
- Experiment and result
- Conclusion

Introduction

- Motivat
  - A pict
  - Tradit
    - Contai
      illustr
  - Onlin
    - Merria
    - The Vi
    - Serve
      - Manually strictly picked image
      - Not all concepts have corresponding images
      - Images are usually not real world images

lrawing to

Introduction

- Our goal
  - To de nerate
    sets o ven
    word
    - Le ltimedia
      co

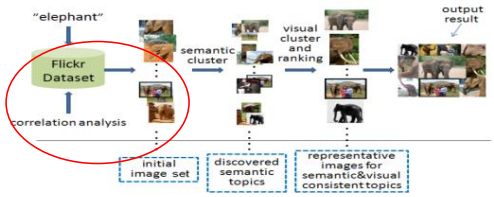
- Challe
  - Corre
  - Diversity
  - Representativeness

Table of Contents

- Introduction
- Discovering Semantic and Visual Diversity
- Experiment and result
- Conclusion

Approach

- Framework



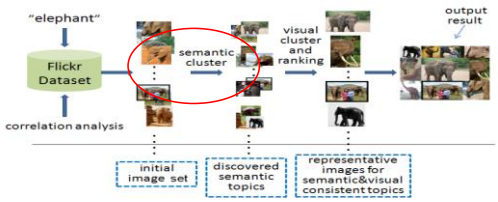
Generating diverse and precise image set

- Heuristic for diversity
  - Sample images from different groups, different users, different time
- Correlation analysis for precision
  - Flickr related tags: high co-occurrence

$$CorrScore(J,w) = \#\{w' \mid w' \in RT_w \ \& \ w' \in (Tag_J \cup Title_J)\}$$

  - Image J is accepted as relevant if CorrScore(J,w) > Th

Discover diversity using semantic clustering



- Computing salience score of keywords
- Text-based clustering of images

Discover diversity using semantic clustering

- Computing salience score of keywords
  - Different keywords in the image's tags contribute differently to the discovery of topics (unique characteristics or semantic aspects)
    - Eg. "trunk" dominates over "water" in finding interested topics for "elephant"
  - Statistical and knowledge-based measure
    - Flickr TFIDF of tags
    - Wordnet
      - **Hyponymy: eg.** the hyponyms of "athlete" include "acrobat", "baseball player", "tennis player","runner"
      - **Meronym: eg.** "tusk" and "trunk" are meronyms of "elephant"

$$Saliency(K) = TFIDF(K) \cdot (1 + HM(K)) \cdot \frac{1}{1 + HH(K)} \cdot (1 + RT(K))$$

  - Top-10 salient keywords for "elephant":
    african, tusk, wildlife, trunk, safari, zoo, Thailand, animal, nature, India
- Text-based clustering of images

Comparison

Before salient words discovering	After salient words discovering
zoo	African
animal	tusk
Africa,	wildlife
animals	trunk
safari	safari
London	zoo
wildlife	Thailand
Kenya	animal
nature	nature
Tanzani	India

African, tusk, wildlife, trunk are more representative than the original tag set. Using saliency words discovering, we can find more meanings given a certain concept.

Discover diversity using semantic clustering

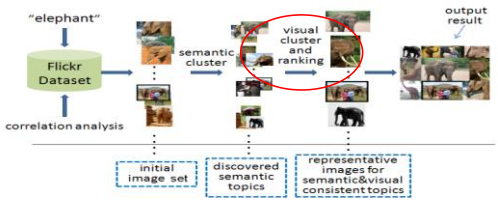
- Computing salience score of keywords
- Text-based clustering of images
  - Select top M salient tags as Keywords
    - KEYWORD={$K_1,K_2,...,K_M$}
  - Convert image tags into Keyword vector

$V_J=(v_1,v_2,\ldots,v_m)$, where v_i is defined as

$$v_i = \begin{cases} Saliency(K_i), & K_i \ is \ in \ image \ J's \ tags \ or \ title \\ 0, & otherwise \end{cases}$$

  - Agglomerative algorithm

Clustering in Visual Space



- Correlation analysis: *Flickr* tags provide additional information
- Semantic Clustering: given the salience score of key words, cluster images
- Visual Clustering: discover visual consistent sub-clusters after visual clustering

Generating representative images using visual clustering

- K-means on the visual (grid of color moments) space of each semantically consistent cluster
- Ranking each sub-cluster
  - the sum of saliency score of keywords in the cluster
  - the number of images in the cluster
  - the coherence of the cluster
- Select representative image from top sub-clusters

Table of Contents

Experimental results

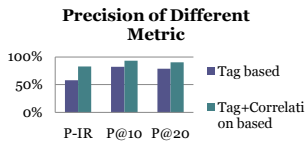
- 2 types of evaluation
  - Objective evaluation on the precision
  - Subjective evaluation on the diversity & representativeness
- 25 Concepts
  - *elephant, camel, buildings, athlete, pyramid, holi days, temple, flower, bridge, …*

Experimental results

- Example: elephant
- Top-10 salient keywords
  - African, tusk, wildlife, trunk, safari, zoo, Thailand, animal, nature, India
- Discovered topics
  - "india- wildlife- pachyderm- temple"
  - "animal-art- sculpture-Asia"
  - "zoo-London- trunk- tusk"
  - .....

Experimental results

- Precision evaluation
  - To validate the effectiveness of correlation analysis in improving the accuracy of retrieval and generating representative images
  - Baseline: tag-based
  - Metric:
    - the precision for image retrieval (P-IR) of 1000 images
    - the precision at generating top-10 (P@10) and top-20 (P@20) representative images



Experimental results

- User study
  - To highlight the system's usability and performances on discovering diversity and representativeness

Table 2. Survey results from students on *Word2Image*

Assessment Questions \ Score	1	2	3	4	5
1) Do you think this system is useful in explaining the meaning of a word?	0	0	0	13	8
2) The coverage of the discovered topics (The topics are explored with the superposed keywords shown on the representative images when mouse is placed over them)	0	0	1	16	4
3) The representativeness of the representative images	0	0	5	14	2
4) Overall satisfaction with the system	0	0	3	16	2

Experimental results

- More examples: pyramid

- Discovered topics:
  - "France- Paris- museum- louvre
  - Africa- Egypt- cairo- desert
  - Mexico- yucatan- maya- temple
  - history- architecture- giza- spohinx

Experimental results

- More examples: holidays

- Discovered topics:
  - "Winter-december-happy-xmas"
  - "Beach-sea-ocean-sun"
  - "Disneyland-disney-california-travel"
  - "Vacation-travel-hotel-happy"
  - …

Experimental results

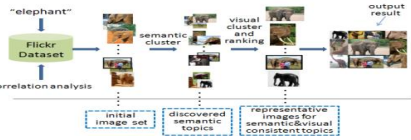
- More examples: athlete

- Discovered topics:
  - "Run-marathon-race-track"
  - "Run-swim-ironman-bike"
  - "Soccer-girl-ball-woman"
  - "Basketball-ball-people-high"
  - …

Table of Contents

- Introduction
- Discovering Semantic and Visual Diversity
- Experiment and result
- Conclusion

Conclusions



•Introduce a framework which attempts to leverage the web image collection to translate a word into its visual counterpart

•Generate high quality, precise, diverse and representative images given a certain concept

•The preliminary experimental results have demonstrated its usability and effectiveness.

Future works

- More experiments to evaluate the use of social media for auto-visual dictionary task
- Multimedia dictionary
  - ***To build a large-scale multimedia dictionary, where multi-modality information including image, video, audio and text are integrated to explain concepts.***

Q and A