Mining Text Snippets for Images on the Web

Anitha Kannan, Simon Baker, Krishnan Ramnath, Juliet Fiss, Dahua Lin, Lucy Vanderwende, Rizwan Ansary, Ashish Kapoor, Qifa Ke, Matt Uyttendaele, Xin-Jing Wang, Lei Zhang
Traditional img2txt

• Image -> Human Readable Text
• Computer vision combined with NLP
  • People/Object/Scene recognition etc..

• Lots of Recent Work:
  • [Farhadi et al., 2010] [Ordonez et al., 2011]
    [Yang et al., 2011]
  • [Kulkarni et al., 2011] [Kuznetzova et al., 2012]
  • [Gupta and Mannem, 2012] [Mitchell et al., 2012]
Focus of this work

• Image -> Snippets of text about the image
  • A snippet is a piece of text that conveys something interesting about the image
  • May not be visually obvious

• For all “interesting” images on the web
  • Redundant at least x times on the web

On Einstein's 72nd birthday on March 14, 1951, UPI photographer Arthur Sasse was trying to persuade him to smile for the camera, but having smiled for photographers many times that day, Einstein stuck out his tongue instead.
The Maldives experiences two monsoon periods; The Maldives southwest monsoon lasts from May to October and The Maldives northeast monsoon lasts from November to April.

Although The Maldives officially covers 90 thousand square kilometres, the actually land area of The Maldives sitting above the surface of the water is only 300 square kilometres.
James Marshall Hendrix (born Johnny Allen Hendrix; November 27, 1942 – September 18, 1970) was an American guitarist, singer, and songwriter whose guitar playing continues to be a considerable influence on rock music.

In 2006, his debut US album, Are You Experienced, was inducted into the United States National Recording Registry, and Rolling Stone named Hendrix the top guitarist on its list of the 100 Greatest Guitarists of All Time.
The Web is big:

At 1 result per second, it would take years to show all results...
Application 1: Image Browsing

Web Image Search = Browsing (almost all time)

Typical “Asks” from Image Search Users:
- Interesting Information About Image
- Pointers to Related Interesting Images

See Paper for links to videos showcasing the applications
Through the photosynthetic process, green plants absorb sunlight in their leaves and convert the photonic energy into chemical energy that is stored as sugars in the plants’ biomass.
Algorithm for Mining Snippets
Snippet Mining Algorithm (Key Insights)

- Redundancy
  - If an image is interesting, multiple webpages will contain that image

- Diverse contexts
  - Same image can be used in different contexts

- Mining these webpages can enable obtaining diverse set of relevant and interesting snippets about that image
Snippet Mining Algorithm

- Technique is run at web scale over the entire index of web images
- Entire pipeline takes about 45 hours on 100 cores
Snippet Mining Algorithm

Image Sets generation

Each image set

Candidate Snippets

Snippet Subset Selection
Relevance, Interestingness, Diversity

Top K snippets
Image Sets generation

• Form clusters of images that are near duplicate of each other
  • Keep track of source webpages

• Image near-duplicate discovery over the entire index of web images

• Two-step clustering
  • Obtain seed clusters using global features
  • Grow the seed clusters using local features

• Scalable implementation using hashing

Candidate Snippets

Image Sets generation

Each image set

Candidate Snippets

Snippet Subset Selection
Relevance, Interestingness, Diversity

Top K snippets

All snippets (sentences) extracted from all webpages associated with the image set
Candidate Snippets

Image Sets generation

Each image set

Candidate Snippets

Snippet Subset Selection
Relevance, Interestingness, Diversity

Top K snippets
Snippet Subset Selection

\[ \mathcal{F}(\mathcal{T}|I) = \lambda \sum_{s \in \mathcal{T}} \text{rel}(s) \text{int}(s) + (1 - \lambda) H_0(\mathcal{T}) \]

- Relevance of snippet s to image
- Interestingness of s
- Diversity of chosen subset of snippets (measured using entropy)

\[ \mathcal{T}^* = \arg \max_{\mathcal{T} \in 2^{|S|}, |\mathcal{T}| = k} (\mathcal{F}(\mathcal{T}|I)) \]

- Sub-modular function (diminishing returns as we add more snippets)
- Greedy solution with (1-1/e) guarantees
- Can be solved independently for each image set
Learning Relevance and Interestingness

Regularized Linear Regression

\[ w^* = \arg \min_w \lambda \|w\|^2 + \sum_{i} (w^T \phi(x_i) - y_i)^2 \]
Empirical Evaluation

[Quick Overview – More results in the paper]
Baseline Algorithms

Query-By-Image/Webpage Summarization
1. Find all occurrences of image on the web
2. Use webpage summarization
E.g. images.google.com

Ordonez, Kulkarni, and Berg 2011 – “Describing images using 1 million captioned photos”
1. Create DB of captioned Flickr images
2. Match image to the DB
3. Transfer captions from top matches

Comparison with this algorithm is not really fair:
• It is solving a slightly different problem
• DB size is far smaller than the entire web
• DBs are mismatched: web vs Flickr
Our Results:
- Towering over Liberty Island in New York, the Statue of Liberty is recognized world-wide as a symbol of the United States.
- The statue, a gift to the United States from the people of France, is of a robed female figure representing Libertas, the Roman goddess of freedom, who bears a torch and a tabula ansata (a tablet evoking the law) upon which is inscribed the date of the American Declaration of Independence.

Query-By-Image/Webpage Summarization:
- Located in New York Harbor, the Statue of Liberty was a gift of international friendship from the people of France to the people of the United States.
- The Statue of Liberty (Liberty Enlightening the World; French: La Liberté éclairant le monde) is a colossal neoclassical sculpture on Liberty Island in the middle.

Ordonez, Kulkarni, and Berg 2011:
- Kenaf flower against background of deep blue sky in Kelantan.
- The radio tower is said to be the tallest wooden structure in the world, at 110 meters high.
Our Results:

• Banana is the common name for any of the very large, tree-like, herbaceous plants comprising the genus Musa of the flowering plant family Musaceae, characterized by an above-ground pseudostem (false stem) with a terminal crown of large leaves, and hanging clusters of edible, elongated fruit.

• These bananas, grown right along side the railroad tracks, would benefit him and the United Fruit Company in the future when the railroad would help transport the bananas to the United States (Landmeier).

Query-By-Image/Webpage Summarization:

• I am not always entranced by new ways of doing things, or new technology. I am resistant to the lure of novelty. I am often fatalistic about ...

• Ways to Use a Banana Peel: The banana peels are not only good to make people fall, but they are also very useful for the following: Relieves ...

Ordonez, Kulkarni, and Berg 2011:

• I put extra challenge on still life photography by taking on a white teapot and cup on a white background. It was fun!

• a broken sphere plastic ball under two regular tungsten bulbs
Our Results

Text Snippet Preference

- All Votes
- Per Image

Percentage

- Our Results
- No Preference
- QbI/WS
Coverage on top 10K Image Search Queries

Percent of Queries with At Least Given Number of Results Images with K Text Snippets

- K = 3
- K = 4
- K = 5
“Failure” Cases: No Text Above Threshold

(a) Personal Photos
(b) Foreign Language
(c) Commercial
(d) Icons
Summary and Future Work

• Web mining approach to associate textual snippets with images.
  • Based on exploiting redundancy of images on the web
  • Snippets are relevant, interesting and diverse, but may not be visually apparent

• Single framework to combine multiple modalities

• A number of interesting applications
  • Augmentation of the web
  • Interesting browsing experience based on user interests