

Finding Visual Concepts by Web Image Mining

Keiji Yanai

Department of Computer Science,
The University of Electro-Communications
Chofu, Tokyo, 182-8585 JAPAN
yanai@cs.uec.ac.jp

Kobus Barnard

Computer Science Department,
University of Arizona
Tucson, AZ, 85721 USA
kobus@cs.arizona.edu

ABSTRACT

We propose measuring “visualness” of concepts with images on the Web, that is, what extent concepts have visual characteristics. This is a new application of “Web image mining”. To know which concept has visually discriminative power is important for image recognition, since not all concepts are related to visual contents. Mining image data on the Web with our method enables it. Our method performs probabilistic region selection for images and computes an entropy measure which represents “visualness” of concepts. In the experiments, we collected about forty thousand images from the Web for 150 concepts. We examined which concepts are suitable for annotation of image contents.

Categories and Subject Descriptors: I.4 [Image Processing and Computer Vision]: Miscellaneous

General Terms: Algorithms, Experimentation, Measurement

Keywords: Web image mining, image recognition, probabilistic method

1. INTRODUCTION

We propose novel use of image data on the Web. So far, much work regarding Web image search has been proposed as well as commercial services. However, most of them focused on only “search”. On the other hand, our purpose is “Web image mining” [3, 2], which means searching the Web for images and then using them as visual knowledge for some applications. In this paper, we describe a new method and experiments to measure “visualness” of word concepts with Web images as a case study of “Web image mining”.

Regarding text data, there are many studies about how to gather data from the Web and use it as “knowledge” effectively. While such Web text mining is an active research area, mining image data on the Web poses additional challenges and has seen less research activity. The problem with mining images for knowledge is that it is not known how to reliably automatically determine semantics from image data. This has been referred to as the semantic gap. To solve it, it is indispensable to introduce sophisticated image recognition methods into Web mining regarding images.

In [3, 2], we proposed gathering a training data set for generic image recognition from the Web automatically, and have revealed that we could use Web images as “visual knowledge”. In this paper, we propose yet another new application of “Web image mining”.

In this paper, we describe how to measure “visualness” of word concepts using Web images, that is, what extent concepts have visual characteristics. We have many words to annotate images with. However, not all words are appropriate for image annotation, since some words are not related to visual properties of images. To know

which word concept has visually discriminative power is important for image recognition task, especially automatic image annotation by generic image recognition systems. For example, “animal” and “vehicle”. They are not tied with the visual properties represented in their images directly, because there are many kinds of animals and vehicles which have various appearance in the real world. Generic image recognition systems should first recognize the concepts which have much visual properties.

So far, most of the work related to image annotation or image classification has either ignored the suitability of the vocabulary, or selected concepts and words by hand. The popularity of “sunset” images in this domain reflects such choices, often made implicitly. We propose that increasing the scale of the endeavor will be substantially helped with automated methods for selecting a vocabulary which has visual correlates.

Web images are as diverse as real world scenes, since Web images are taken by a large number of people for various kinds of purpose. This property is completely different from commercial or personal photo collections built by one or a few persons. It can be expected that such diverse images on the Web enable us to measure general “visualness” of a concept by analyzing Web images associated with the word concept.

To gather images associated with a certain word concept, we use Google Image Search as the first step. However, the precision of the results are not good in general, since Google Image Search rely on text associated with the images as determined from surrounding HTML data by a variety of heuristics. To select only relevant images to the given word concept from output of Google Image Search, we perform a probabilistic region selection method, and then we compute a measure of the entropy of the selected regions based on a Gaussian mixture model for regions. Intuitively, if such an entropy is low, then the concept in question can be linked with image features. Alternatively, if the entropy is more like that of random regions, then the concept has some other meaning which is not captured by image features.

To estimate “visualness” of concepts, we can use precision and recall diagram. However, to compute precision and recall, we need a ground truth set, namely, labeled images. In general, no labeled images are available for general “X”, while “image region entropy” do not need labeled images or annotation of images by human at all. In that sense, “image region entropy” is a very useful measure to examine many kinds of concepts and compare their “visualness”.

To investigate these ideas, we collected forty thousand images from the World-Wide Web using the Google Image search for 150 adjectives. We examined which adjectives are suitable for annotation of image contents.

2. METHOD

To compute the “image region entropy” associated to a certain concept, we begin by gathering images related to the concept. While

it is difficult to manually collect large numbers of images related to one concept, we can gather images likely associated to a certain concept using Web image search engines such as Google Image Search. Of course, raw results from the Web image search engines, usually include irrelevant images. Moreover, the images usually include backgrounds as well as objects associated with a concept. Therefore, we need to eliminate irrelevant images and pick up only the regions strongly associated with the concept in order to calculate the image entropy correctly. We use only the regions expected to be highly related to the concepts to compute the image entropy. To select regions associated with concepts, we use a probabilistic method.

To find regions related to a certain concept we use an iterative algorithm. Initially, we do not know which region is associated with a concept “X”, since an image with an “X” label just means the image contain “X” regions. In fact, with the images gathered from the Web, even an image with an “X” label sometimes contain no “X” regions at all. So at first we have to find regions which are likely associated with “X”. To find “X” regions, we also need a model for “X” regions. Here we adopt a probabilistic generative model, namely a mixture of Gaussian, fitted using the EM algorithm. In short, we need to know a probabilistic model for “X” and the probability of “X” over each region, $P(X|r_i)$, simultaneously. However, each one depends on each other, so we proceed iteratively. The detail of the algorithm to obtain $P(X|r_i)$ is described in [4].

After obtaining $P(X|r_i)$ for each region, we estimate the entropy of the image features of all the regions weighted by $P(X|r_i)$ with respect to a generic model for image regions obtained by the EM in advance. It is “image region entropy”, which corresponds to “visualness” of the concept. To represent a generic model, we use the Gaussian mixture model (GMM).

We need to obtain a generic base in advance by the EM for computing the entropy. To get a generic base, we used about fifty thousand regions randomly picked up from the images gathered from the Web.

The entropy for “X” is given by $H(X) = \sum_{j=1}^{N_{\text{base}}} -P(c_j|X) \log_2 P(c_j|X)$ where $P(c_j|X)$ is the probability of j -th component computed by the formula of the Gaussian, and N_{base} is the number of the components of the base. In the experiment, we set 250 to N_{base} .

3. EXPERIMENTS

As test images associated with concepts, we used the images gathered from the World Wide Web by providing 150 adjectives for Google Image Search. We obtained about 250 Web images for each adjective. Totally we obtained about forty thousand images associated with adjectives.

Table 1 shows the 10 top adjectives and their image entropy. In this case, the entropy of “dark” shown is the lowest, so in this sense “dark” is the most “visual” adjective among the 150 adjectives under the condition we set in this experiment. Figure 1 shows part of “dark” images. Most of the region labeled with “dark” are uniform black ones. Regarding other highly-ranked adjectives, “senior” and “beautiful” includes many human faces, and most of “visual” are not photos but graphical images such as screen shots of Windows or Visual C++.

We show the ranking of color adjectives in the lower part of Table 1. They are relatively ranked in the upper ranking, although images from the Web included many irrelevant images. This shows the effectiveness of the probabilistic region selection method we proposed. Initially we expected that all of them were ranked in the nearly top, but they weren’t. This is because all the images we used are collected from the Web automatically, and the test image sets always include some irrelevant images. So we could not obtain ideal

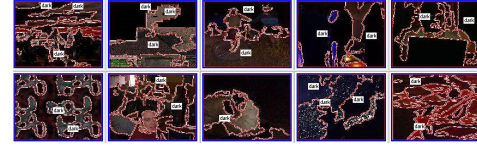


Figure 1: “dark” image.



Figure 2: “Religious” images.

Table 1: Top 10 entropy rank- Table 2: Bottom 10 entropy ing and results of color adjec- ranking.

rank	adjective.	entropy	rank	adjective.	entropy
1	dark	0.0118	141	elderly	2.5677
2	senior	0.0166	142	angry	2.5942
3	beautiful	0.0178	143	sexy	2.6015
4	visual	0.0222	144	open	2.6122
5	rusted	0.0254	145	religious	2.7242
6	musical	0.0321	146	dry	2.8531
7	purple	0.0412	147	male	2.8835
8	black	0.0443	148	patriotic	3.0840
9	ancient	0.0593	149	vintage	3.1296
10	cute	0.0607	150	mature	3.2265
(color adjectives)					
7	purple	0.0412			
8	black	0.0443			
36	red	0.9762			
39	blue	1.1289			
46	yellow	1.2827			

results in this experiment. Note that the ranking varies if the condition of the experiment such as some parameters, image features and image search engine to gather Web images are changed.

Table 2 shows the 10 bottom adjectives. In case of “religious” shown in Figure ??, which is ranked in the 145-th, the region selection did not work well and the entropy got relatively larger, since the image features of the regions included in “religious” images have no prominent tendency. So we can say that “religious” has no or only a few visual properties.

4. CONCLUSIONS

In this paper, we described a new application of “Web image mining” to measure “visualness” of concepts with images on the Web. The experiments showed that the method to select regions was effective and mostly “image region entropy” indicated “visualness” of concepts.

In fact, this result is not always consistent with our intuition, since region selection sometimes did not work well for some adjectives. As future work, we plan to improve the region selection method so that “image region entropy” represents “visualness” of concepts more precisely. As advanced work, we will develop an image annotation model to integrate nouns and adjectives by extending our image annotation models [1], and examine if adjectives improve image annotation task in which only nouns have been used so far.

5. REFERENCES

- [1] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.
- [2] K. Yanai. Generic image classification using visual knowledge on the web. In *Proc. of ACM International Conference on Multimedia 2003*, pages 67–76, 2003.
- [3] K. Yanai. Web image mining toward generic image recognition. In *Proc. of the Twelfth International World Wide Web Conference*, 2003.
- [4] K. Yanai and K. Barnard. Image region entropy: A measure of “visualness” of web images associated with one concept. In *Proc. of ACM International Conference on Multimedia 2005*, pages 420–423, 2005.