Visual Analysis of Tag Co-occurrence on Nouns and Adjectives

Yuya Kohara and Keiji Yanai

Department of Informatics, The University of Electro-Communications,Tokyo
1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585 Japan,
kohara-y,yanai@mm.cs.uec.ac.jp

Abstract. In recent years, due to the wide spread of photo sharing Web sites such as Flickr and Picasa, we can put our own photos on the Web and show them to the public easily. To make the photos searched for easily, it is common to add several keywords which are called as "tags" when we upload photos. However, most of the tags are added one by one independently without much consideration of association between the tags. Then, in this paper, as a preparation for realizing simultaneous recognition of nouns and adjectives, we examine visual relationship between tags, particularly noun tags and adjective tags, by analyzing image features of a large number of tagged photos in social media sites on the Web with mutual information. As a result, it was turned out that mutual information between some nouns such as "car" and "sea" and adjectives related to color such as "red" and "blue" was relatively high, which showed that their relations were stronger.

1 Introduction

In recent years, due to the wide spread of digital cameras and mobile phones with camera, the amount of images on the Web has increased explosively. At the same time, because photo sharing sites such as Flickr have become common where users post their images with tags, there are so many tagged images on the Web. These tag information are used as a keyword on image search. However, most of the tags are added one by one independently without much consideration of association between the tags. This sometimes causes irrelevant results when we do AND-search with multiple keywords. For example, we obtain a photo showing blue sky and a red car for the query with "blue AND car". To remove such a photo and to obtain only the photos including blue cars, simultaneous image recognition of multiple tags such as "blue cars" is needed. If we can automatically eliminate irrelevant images, search results for multiple keywords become more correct. In addition, we can create dataset easily with less noise. In order to perform more accurate image acquisition and image search, it is necessary to take into account the relationship between the tags and to focus more on contents of images. Then, in this paper, we analyze visual relationship between nouns and adjectives using a large number of tagged images in social media sites on the Web such as Flickr. To do that, we use entropy and mutual information based on visual features extracted from image regions. Moreover, the result of analysis in this paper can be used in simultaneous recognition where we recognize a certain object by a noun and the state of the object by an adjective further like "there is a car and the color of the car is red".

In the rest of this paper, we describe related work in Section 2. We explain the overview in Section 3 and the detail of the proposed system in Section 4. We show experimental results in Section 5. In Section 6, we conclude this paper.

2 Related work

In the community of object recognition, recently, recognition of attributes of objects such as adjectives are paid attention to. In this paper, we are inspired by the current trends of the object recognition research, and focus on "attributes" in the tags of social media photos.

T. L. Berg et al.[1] focused on attributes on color, shape, and texture. They extracted words associated with the attributes from the texts which were listed in the shopping site, and labeled local regions represented by attributes corresponding text description.

D. Parikh et al.[2] focused on the attributes from the perspective of "nameable". "Nameable" means whether human can understand and represent attributes automatically extracted from images by language. They discovered attributes of "nameable" by an interactive approach by using Amazon Mechanical Turk.

A. Farhadi et al.[3] described images by a set of attributes. They recognized not only "dog" but "spotty dog". By using attributes, we are able to describe "dog" which has "spots" as "spotty dog", when we have no knowledge about the subordinate categories of "dog". In addition, it has become possible to mention also distinctive attributes by using this description. That is, if attributes which "dog" has but "sheep "does not have exists, they will be the attributes which discriminate a dog from a sheep. Moreover, discovering distinctive attributes enables us to mention the attributes which should exist or shouldn't exist in each of the given classes. Therefore, a car whose doors are not visible can be recognized as a car with doors. In addition, they also estimated bounding box regions to which the given attributes correspond.

S. Dhar et al.[4] focused on the particular attributes of aesthetics and interestingness. They dealt with two types of attributes about composition and contents of objects in a given image as a guideline of aesthetics and interestingness. The attribute of composition contains saliency, location, and color of objects. The attributes of contents contain type, place, and scene where objects are shown. In this research, they mainly focused on subjective attributes.

The paper [1, 2] recognized a single attribute, and the paper [3] dealt with attributes as parts which should exist in the corresponding objects like "a door is a part of a car", while we limit attributes to only adjectives, and also focus on the relationship between nouns and adjectives. The paper [4] analyzed the specific adjectives such as aesthetics and interestingness, while we dealt with more general adjective.

Next, we introduce related works on visual concept analysis. Here, we cite the papers of Yanai et al. [5], Akima et al. [6], and Kawakubo et al. [7]. Yanai et al. proposed the entropy as a way to quantify the relation of visual concept, and referred to the visual relation about 150 adjectives [5]. We use the method of quantifying the visualness of the word by entropy and calculation of entropy. Akima et al. built a database with hierarchical structure of between concepts from distance relationship and hierarchical relationship [6]. They used entropy and calculate the distribution of the images to determine the hierarchical relationship. In addition, tag information which is given to the images was also used. Kawakubo et al. analyzed the visual and geographical distribution in word concepts [7]. In this research, they calculated the image distribution of the class of concepts such as a noun or an adjective to evaluate visualness by using calculation of entropy and region segmentation. The difference between this research and the above-mentioned works is that we define concept classes with the combinations of two words, and pursuit the visual relation of the combinations of nouns and adjectives.

3 Overview

In this paper, the visual relationship between a nouns and an adjective is evaluated by the distribution of the image features of the corresponding image regions. We judge that there is a high relation if visual distribution is narrow enough. The extent of the distribution is quantified using the concept of entropy. Entropy is used to calculate the local features obtained from a set of image regions. Mutual information is the difference between entropy of a noun and entropy of combination with an adjective and the noun. Mutual information becomes higher, when visual relation between a noun and an adjective becomes higher.

Processing procedure of the experiment in this paper is shown below.

Procedure

1. Image acquisition from the Flickr by tag-based search
2. Image segmentation
3. Feature extraction and creating BoF for each region
4. Positive region detection
5. Calculation of feature distribution in each positive region by PLSA
6. Calculation of entropy and mutual information

In this paper, we also calculated similarity by co-occurrence of tags by the Normalized Google Distance (NGD) for comparison with the visual relation by entropy.

4 Proposed method

In this section, we describe the methods used in the experiment. In this experiment, we calculated the entropy to refer to visual relation between an adjective and a noun. In addition, we calculated the similarity by co-occurrence of tags for comparison.

4.1 Image acquisition

We collect 200 positive tagged images for each class by using the Flickr API. At the time, we use AND-search with a noun word and an adjective word. In addition, we prepare the 800 negative images which have neither of the tags. Note that we collected only one image from the same Flickr contributor for one query, since the same user tends to upload many near-duplicated photos, which sometimes causes irrelevant bias on feature distribution.

4.2 Region segmentation

To select regions directly related to the given words and remove background regions, we perform region segmentation with JSEG [8]. In the experiment, we set the number of the maximum regions as 10, and carry out post-processing to unify relatively smaller regions into larger regions.

4.3 Feature extraction

As feature representation of each region, we use Bag-of-Features (BoF) [9] with Color-SIFT [10].

At first, we extract the Color-SIFT feature as local features. To extract Color-SIFT, we extract the SIFT [10] features from each of the color channels such as R, G and B, regarding each keypoint, and create new feature vectors by concatenating SIFT vectors of the three channels. Therefore, the dimension size of this feature vector is 128×3. When extracting Color-SIFT features, we use dense sampling where all the local features are extracted from multi-scale grids.

Bag-of-Features (BoF) [9] is a standard feature representation to convert a set of local features into one feature vector. To convert a set of local feature vectors into a BoF vector, we vector-quantize them against the pre-specified codebook. After that, all the BoF vectors are L1-normalized. In the experiments, we built a 1000-dim codebook by k-means clustering with local features sampled from all the images. Note that we construct a BoF vector for each region by using the local features extracted inside the corresponding region.

4.4 Positive region selection

To select positive regions with the region-based BoF vectors, we use mi-SVM [11] which is a method of multiple instance learning. The mi-SVM is a support vector machine modified for multiple instance setting, and it is carried out by iterating a training step and a classification step using a standard SVM.

Under the multiple instance setting, training class labels are associated with a set of instances instead of individual instances. A positive set, which is called as a "positive bag", has one positive instance at least, while a negative set, which is called as a "negative bag", has only negative instances. This multiple instance setting fits well with the situation where an image consists of several foreground regions and background regions. Since we can regard foreground and background regions as positive and negative instances, respectively, by using multiple-instance learning methods we can classify regions into either foregrounds or backgrounds without explicit knowledge on foregrounds.

The process of positive region selection is shown below.

The process of positive region selection

1. Initially, regard all the regions in the positive images as positive instances, and regard all the regions in the negative images as negative instances.
2. Train a standard SVM using the positive and negative instances.
3. Classify all the instances in the positive images with the trained SVM.
4. Regard the instances assigned the higher scores by the SVM as positive instances in the next step, and regard the instances having the lower scores as negative instances in the next step.
5. Repeat from 2 through 4 several times.

4.5 Calculation of feature distribution

To calculate the feature distribution, we perform probabilistic clustering of feature vectors using the Probabilistic Latent Semantic Analysis (PLSA) [12].

pLSA The calculation of pLSA is performed as follows: First, the joint probability of an image and an element of the BoF vector, which corresponds to "visual words", are represented as

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

where $d_i (i = 1, 2, \ldots, I)$ is an image, $w_j (j = 1, 2, \ldots, J)$ is an element of BoF feature vectors (visual word frequency), and $z_k (k = 1, 2, \ldots, K)$ is a latent topic variable. Then, the probability that the word will be generated within the document is given by

$$P(w_j|d_i) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i) \tag{2}$$

using latent topic variable z_k. In addition, if the number of the word w_j within the document d_i is defined as $n(d_i, w_j)$, the log-likelihood of the data is represented as the following expression:

$$L = \sum_{i=1}^{I} \sum_{j=1}^{J} n(d_i, w_j) \log P(d_i, w_j) \tag{3}$$

We determine $P(z_k)$, $P(d_i|z_k)$, and $P(w_j|d_i)$ such as to maximize this log-likelihood by the iterative EM algorithm.

4.6 Calculation of entropy and mutual information

The entropy was calculated using the probability obtained by the pLSA. The value of the entropy increases, when the distribution of the BoF feature vectors corresponding to the positive regions becomes wider. On the other hand, the entropy value decreases, when the distribution becomes narrower. Therefore, the value of the entropy represents the size of the image distribution belonging to a given class concept which is the combination of a noun and an adjective. That is, calculating the entropy leads to searching for visual relation on the combinations of nouns and adjectives.

We calculate the entropy based on $P(z_k|d_i)$ which is estimated by pLSA. First, we calculate the probability of z_k over a given concept X by:

$$P(z_k|X) = \frac{\sum_{d_i \in X} P(z_k|d_i)}{|X|}, \tag{4}$$

where X represents a set of the images corresponding to a given concept, for each latent topic variable. Then, we calculate the entropy of a given concept X by

$$H(X) = -\sum_{k=1}^{K} P(z_k|X) \log P(z_k|X). \tag{5}$$

The mutual information is a value represented by the difference between the entropy and the conditional entropy, which indicates the relevance between the tags. We calculate the mutual information as

$$MI(X;Y) = H(X) - H(X|Y), \tag{6}$$

where $H(X)$ is the entropy of one class, and $H(X|Y)$ is the entropy of the class combined two classes. If image distribution becomes narrow by combining the tag X with the tag Y, we judge the visual relevance become higher from the increase of mutual information.

Table 1. The 20 nouns used in experiment

beach	bird	boat	bridge	car
cat	cloud	cup	dog	flower
fruit	house	people	sea	sky
snow	sun	tower	train	tree

Table 2. The 15 adjectives used in experiment

red	blue	green	black	white
circle	square	morning	night	winter
summer	new	old	beautiful	cool

4.7 Calculation of similarity by co-occurrence of tags

For comparison, we calculate similarity by co-occurrence of tags using the Normalized Google Distance (NGD) [13] as well. The formula is

$$NGD = \frac{\max\{\log f(x), \log f(y)\} - \log f(x, y)}{\log N - \min\{\log f(x), \log f(y)\}}, \tag{7}$$

where x is a noun, y is an adjective, $f(x)$ and $f(y)$ are the image number of tag search by a noun and an adjective in Flickr, and $f(x, y)$ is the image number of AND-search by combination of a noun and an adjective. Moreover, N is the number of all images in Flickr. However, we assume N is 50 billion since it is unable to get to know the exact number.

5 Experiments

5.1 Dataset

Images were collected using the API from Flickr. We collected 800 negative images and 200 positive images under the restriction that we obtained only one image from the same uploader. In addition, we retrieved positive images in order from the top in the search ranking of Flickr. Negative images were selected from among the images obtained at random from Flickr, which does not have the tags of nouns and adjectives of a particular class. In this experiment, we selected 20 nouns as shown in Table 1, and 15 adjectives as shown in Table 2. Thus, we calculate the entropy about 20×15 classes which are the combination of each noun and each adjective, as well as 20 classes which are only noun.

5.2 Experimental results

According to the procedure explained in the previous section, we calculated mutual information for each class. Figure 1 shows calculation result of the entropy values in the second columns and the mutual information values after the third columns. On the other hand, Figure 2 shows the calculated results on the similarity of NGD using tag co-occurrence. We summarized the combinations of nouns and adjectives which are judged to have high relation by mutual information values in Table 3 and by NGD in Table 4, respectively.

With these experimental results, we compare mutual information of each class. Mutual information decreases when the distribution of images in each class

noun/adjective	–	red	blue	green	black	white	circle	square	morning	night	winter	summer	new	old	beautiful	cool
beach	5.383	0.198	0.099	−0.009	0.027	−0.059	0.018	0.181	0.338	0.305	0.101	−0.058	0.037	−0.045	0.075	0.011
bird	5.478	0.147	0.193	0.182	0.029	−0.045	−0.009	0.115	0.321	0.034	0.103	0.212	−0.023	−0.012	0.063	0.082
boat	5.398	0.193	0.123	−0.065	0.110	−0.034	−0.045	0.122	0.440	0.297	0.095	0.020	0.065	−0.050	0.197	−0.053
bridge	5.466	0.071	0.354	0.161	0.232	0.078	−0.018	0.151	0.336	0.143	0.003	0.042	0.085	−0.028	0.016	−0.022
car	5.486	0.139	0.105	0.003	0.130	0.118	0.131	0.035	0.101	0.129	0.049	0.044	0.150	−0.003	0.018	0.039
cat	5.521	0.003	0.061	0.046	0.145	0.117	0.061	0.092	0.032	0.083	0.069	0.064	0.046	0.044	0.070	0.048
cloud	5.334	0.078	0.066	−0.020	0.154	−0.024	0.030	0.217	0.220	0.135	0.063	−0.064	0.069	−0.005	0.086	0.014
cup	5.431	0.105	0.137	0.100	0.121	0.150	0.073	0.096	0.169	0.103	0.132	0.013	−0.027	−0.060	−0.015	−0.005
dog	5.522	0.027	0.024	0.069	0.120	0.124	0.144	0.086	0.137	0.211	0.069	0.048	0.050	0.038	0.066	0.008
flower	5.357	0.096	0.185	0.145	0.082	0.055	−0.040	0.175	0.153	0.088	0.011	0.077	0.106	−0.128	0.018	0.030
fruit	5.474	0.112	0.113	0.157	0.242	0.085	0.042	0.113	0.006	0.050	0.117	0.149	0.007	−0.046	0.061	0.018
house	5.536	0.114	0.170	0.163	0.161	0.060	0.040	0.091	0.224	0.078	0.129	0.033	−0.011	0.093	−0.003	−0.001
people	5.519	0.084	0.047	0.024	0.114	0.078	0.035	0.013	0.164	0.153	0.093	0.020	0.134	0.058	0.090	0.040
sea	5.368	0.211	−0.022	−0.038	0.198	−0.030	−0.032	0.108	0.439	0.237	0.198	−0.066	0.056	−0.021	0.077	−0.006
sky	5.387	0.188	0.108	0.016	0.146	0.030	−0.026	0.036	0.287	0.237	0.011	0.048	0.053	−0.022	0.006	−0.002
snow	5.490	0.036	0.261	0.038	0.084	0.044	−0.014	0.167	0.279	0.159	0.054	−0.009	−0.013	0.047	0.084	0.077
sun	5.380	0.278	0.044	0.027	0.069	−0.008	0.176	0.042	0.237	0.248	0.069	0.007	−0.016	−0.067	0.179	0.013
tower	5.473	0.113	0.234	0.051	0.151	0.046	0.022	0.063	0.443	0.101	0.044	0.012	0.043	0.037	0.015	0.015
train	5.535	0.056	0.133	0.054	0.242	0.128	0.149	0.071	0.036	0.145	0.028	0.016	0.040	0.045	0.050	0.023
tree	5.437	0.014	0.137	0.072	0.183	0.058	−0.022	0.173	0.376	0.186	0.164	0.056	0.046	0.056	−0.003	0.011

Fig. 1. Calculation result of mutual information (red: high relevance class, blue: low relevance class)

noun/adjective	red	blue	green	black	white	circle	square	morning	night	winter	summer	new	old	beautiful	cool
beach	0.678	0.492	0.650	0.646	0.630	0.802	0.983	0.620	0.639	0.669	0.445	0.669	0.715	0.539	0.717
bird	0.587	0.518	0.566	0.550	0.563	0.798	0.960	0.666	0.776	0.602	0.708	0.758	0.757	0.616	0.726
boat	0.618	0.516	1.556	0.683	0.647	0.676	0.954	0.606	0.629	0.725	0.575	0.690	0.593	0.618	0.710
bridge	0.646	0.579	0.616	0.623	0.619	0.665	0.798	0.600	0.487	0.613	0.680	0.584	0.567	0.640	0.728
car	0.508	0.556	0.613	0.523	0.573	0.766	0.918	0.747	0.573	0.704	0.683	0.623	0.425	0.666	0.542
cat	0.666	0.624	0.630	0.462	0.518	0.884	0.934	0.761	0.735	0.754	0.774	0.794	0.759	0.660	0.714
cloud	0.579	0.422	0.552	0.588	0.532	0.666	0.859	0.462	0.616	0.640	0.630	0.731	0.651	0.548	0.617
cup	0.659	0.721	0.720	0.711	0.679	0.671	0.943	0.628	0.853	0.853	0.858	0.700	0.734	0.831	0.770
dog	0.638	0.621	0.646	0.477	0.528	0.828	0.925	0.744	0.785	0.617	0.684	0.746	0.720	0.697	0.708
flower	0.405	0.480	0.379	0.579	0.408	0.724	0.878	0.666	0.730	0.709	0.523	0.739	0.765	0.517	0.707
fruit	0.508	0.687	0.534	0.694	0.663	0.667	0.890	0.699	0.809	0.779	0.647	0.812	0.735	0.707	0.671
house	0.594	0.583	0.555	0.604	0.543	0.722	0.895	0.689	0.597	0.618	0.649	0.521	0.434	0.623	0.657
people	0.597	0.589	0.600	0.525	0.524	0.763	0.788	0.700	0.506	0.640	0.527	0.625	0.576	0.474	0.604
sea	0.541	0.394	0.571	0.579	0.560	0.788	0.917	0.588	0.600	0.614	0.472	0.700	0.615	0.525	0.699
sky	0.463	0.226	0.415	0.535	0.444	0.696	0.806	0.489	0.450	0.504	0.480	0.645	0.599	0.498	0.635
snow	0.633	0.560	0.669	0.644	0.435	0.732	0.922	0.603	0.567	0.157	0.763	0.667	0.717	0.653	0.717
sun	0.495	0.408	0.468	0.530	0.496	0.673	0.871	0.420	0.606	0.515	0.416	0.664	0.582	0.405	0.585
tower	0.679	0.573	0.663	0.666	0.629	0.722	0.728	0.649	0.522	0.683	0.725	0.668	0.557	0.670	0.713
train	0.697	0.694	0.731	0.664	0.681	0.748	0.910	0.690	0.649	0.684	0.759	0.692	0.571	0.742	0.711
tree	0.483	0.447	0.376	0.536	0.483	0.709	0.826	0.541	0.565	0.447	0.601	0.691	0.574	0.558	0.654

Fig. 2. Calculation result of co-occurrence of tags by NGD (red: high relevance class, blue: low relevance class)

spreads, and increases when the distribution of images in each class is narrow. Then, we can judge that the classes which have amount of mutual information have high visual relation between nouns and adjectives. Moreover, we determine the classes which have small NGD have high visual relation between nouns and adjectives.

5.3 Discussion on visual relations

First, we discovered the tendency that the mutual information becomes large in the class where a lot of images have been greatly affected by color (see Figure 3). For example, "morning sea", "morning sky", and "blue bridge" class would be cited. Looking at the positive region of the image to be included in the "morning sky" class, there are the regions which have a lot of red region of the morning glow and blue region of the sunny sky. We consider that image distribution in

Table 3. Main classes of high similarity by mutual information

sea+morning	sky+morning
sun+red	sky+night
sun+beautiful	car+red
flower+blue	sun+circle

Table 4. Main classes of high similarity by co-occurrence of tag

snow+winter	sky+blue
tree+green	flower+green
sea+blue	sun+beautiful
flower+red	sun+blue

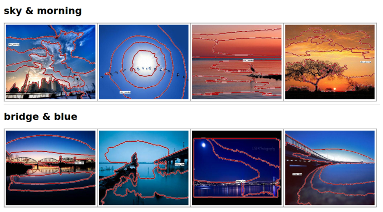

Fig. 3. Positive regions in the class where the color influenced greatly

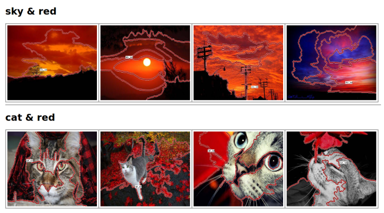

Fig. 4. Positive regions in the class combined with an adjective about color

these classes becomes narrow, and their mutual information increases for that reason. However, visual relevance in "blue bridge" class becomes high, although the class has few images of a blue bridge. We consider that visual relevance increases, because the positive regions include blue of sea, river, and sky around bridge in order to take the entire bridge in the photos.

Next, when we pay attention about the class in combination with the adjective about a color, it turns out that mutual information becomes relatively larger in the class where the adjective about color qualifies directly to the object being indicated by the noun (see Figure 4). We would mention "red sun" and "red car" class as examples of large mutual information, and "red cat" and "red dog" class as examples of small mutual information in the class which combined with the adjective about a color. In such classes where mutual information is greater, and positive regions of that class contains the particular color and object. Whereas, in the class where mutual information is smaller, positive regions of that class do not contain the particular color and object. Therefore, it can be thought that visual relation has been correctly calculated, which is consistent with our intuition.

5.4 Comparison with tag co-occurrence

Some classes have high visual relevance while their co-occurrence relevance by tag is low (see Figure 5). As an example, we cite "old people" class. Both visual relevance and relevance by tag are low in the class combined with the "old". However, there is a tendency that visual relevance of "old" becomes higher than other classes, when it combined with the noun in connection with artificial things and living things such as "house" and "people".

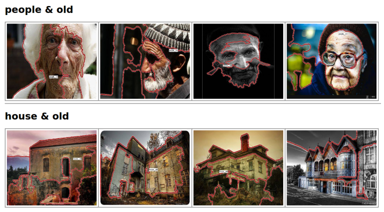

Fig. 5. Positive regions of the class which the relevance of co-occurrence is low, and visual relevance is high

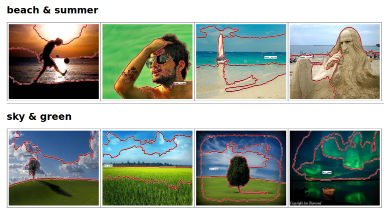

Fig. 6. Positive regions of the class which the relevance of co-occurrence is high, and visual relevance is low

On the other hand, there are classes which have low visual relation, although their relation by tag is high (see Figure 6). As an example, we cite "summer beach" and "green sky" class. It is thought that visual relation became low in "summer beach" class, because there are not only the image of a beach but many images of the people who are doing sea bathing. Meanwhile, it is thought that the relation by tag became high in "green sky" class, because the "green sky" class contains many images of grass, and the co-occurrence of "sky" and "grass" the color of which is green is higher. However, their visual relation is shown as being low.

6 Conclusion and Future work

In this paper, first, we collected images tagged with both particular nouns and adjectives from Flickr. Then, we extracted local features from images, and calculated the distribution of image as the numeric value by the entropy. Finally, we performed comparison and consideration about the visual relation between a noun and an adjective from the change in entropy for each class which combined a noun and an adjective.

As a result, we obtained the results that on mutual information represents intuitive visual similarity. Therefore, it turned out that there was a tendency that the pairs of nouns and adjectives related to color have the stronger visual relation. Regarding tag-based similarity, the degree of similarity by the co-occurrence of tag using NGD showed the results which fitted our intuition as well.

For future work, we plan to use other kinds of visual features than Color-SIFT BoF. In addition, we would like to utilize the results obtained in this paper to improve performance on simultaneous recognition of a noun and an adjective.

References

1. T. L. Berg, A. C. Berg, and A. J. Shih. Automatic attribute discovery and characterization from noisy web data. In *Proc. of European Conference on Computer Vision*, pp. 663–676, 2010.
2. D. Parikh and K. Grauman. Interactively building a discriminative vocabulary of nameable attributes. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2011.
3. A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1778–1785, 2009.
4. S. Dhar, V. Ordonez, and T.L. Berg. High level describable attributes for predicting aesthetics and interestingness. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1657–1664, 2011.
5. 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 Multimedia*, 2005.
6. H. Kawakubo, Y. Akima, and K. Yanai. Automatic construction of a folksonomy-based visual ontology. In *Proc. of International Symposium on Multimedia*, pp. 330–335, 2010.
7. K. Yanai, H. Kawakubo, and B. Qiu. A visual analysis of the relationship between word concepts and geographical locations. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2009.
8. Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800–810, 2001.
9. G. Csurka, C. Bray, C. Dance, and L. Fan. Visual categorization with bags of keypoints. In *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 59–74, 2004.
10. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
11. S. Andrews, I. Tsochantaridis, and T. Hofmann. Support Vector Machines for Multiple-Instance Learning. In *Advances in Neural Information Processing Systems*, pp. 577–584, 2003.
12. T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, Vol. 43, pp. 177–196, 2001.
13. R.L. Cilibrasi and P.M.B. vitanyi. The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 3, pp. 370–383, 2007.