AN ANALYSIS OF THE RELATION BETWEEN VISUAL CONCEPTS AND GEO-LOCATIONS USING GEOTAGGED IMAGES ON THE WEB

Hidetoshi Kawakubo and Keiji Yanai

Department of Computer Science, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 JAPAN

ABSTRACT

Recently, a large number of geotagged images are available on photo sharing Web sites such as Flickr. In this paper, we propose image region entropy and geo-location entropy for analyzing the relation between visual concepts and geographical locations using a large-scale geotagged image database. Image region entropy represents to what extent concepts have visual characteristics, while geo-location entropy represents to what extent concepts are distributed over the world. In the experiment, we analyzed relations between image region entropy and geo-location entropy in terms of 230 nouns and 100 adjectives, and we found that the concepts with low image entropy tend to have high geo-location entropy and vice versa.

Index Terms— geotag, entropy, Flickr, PLSA, bag-of-features

1. INTRODUCTION

Recently, consumer-generated media (CGM) on the Web has become very popular. Especially, photo sharing sites such as Flickr and Picassa are representative CGM sites, which store a huge number of consumer-generated photos people uploaded, and make them accessible via the Web for everyone. Photo sharing sites collect not only photos but also metadata on uploaded photos. Regarding metadata for photos, textual information such as keywords and comments is common. Recently, in addition to texts, some users attach "geo-tags" to their uploaded photos. Note that a "geo-tag" means metadata which represents a location where the corresponding photo was taken, which is expressed by values of latitude and longitude.

Our objective is exploring relationship between word concepts and geographical locations by using a large number of geotagged images on the photo sharing Web sites such as Flickr. In this paper, we propose bag-of-feature-based image region entropy and geo-location entropy and using both of them to analyze relations between location and visual features. In the experiment, we analyzed relations between image region entropy and geo-location entropy in terms of 230 nouns and 100 adjectives, and we found that the concepts with low image entropy tend to have high geo-location entropy and vice versa.

2. ENTROPY ANALYSIS

In this section, we propose a new method to analyze relations between location and concepts in terms of image features. We compute both image region entropy [1] and geo-location entropy for many concepts using geotagged images gathered from the Flickr.

2.1. Image Region Entropy

"Image Region Entropy" is a measure of "visualness" of concepts, that is, to what extent concepts have visual characteristics [1]. In the original method to compute image region entropy, they perform probabilistic region selection for regions that can be linked with concept "X" from images which are labeled as "X" or "non-X", and then they compute a measure of the entropy of the selected regions based on a Gaussian mixture model for regions. By introducing a probabilistic region selection method, they can separate foreground regions from background regions, and compute the entropy using only the foreground regions. Intuitively, if such an entropy is low, then images associated with the concept have typical appearances, and the image features of the concepts are relatively concentrated. Alternatively, if the entropy is larger, the image features of the concepts are distributed, and the concept has no typical images.

In this paper, we modify the original method by using the bag-of-features representation (BoF) [2] and the probabilistic latent semantic analysis (PLSA) [3] instead of color and texture features and the Gaussian mixture model (GMM), since it is regarded that the BoF representation has more semantically discriminative power than other representations [2] and PLSA is more appropriate for the BoF vectors, which is usually high-dimensional and sparse, than GMM. In addition, while the original method employed the probabilistic generative methods to select foreground regions which are used for computation of the region entropy, we use mi-SVM [4], which is a discriminative method, to select positive regions by taking account of the multiple instance learning setting, since discriminative method is superior to generative method in general in case that much training data is available.

We used the following iterative procedure based on mi-SVM [4] to select foreground regions:

1. Prepare a positive image set gathered from Flickr and a random background image set, carry out region segmentation with JSEG [5], and construct the region-based BoF vector for each region.
2. Sample one third of positive images and negative background images. Train SVM with them.
3. Classify all the regions of positive images with the trained SVM.
4. Select one third of regions in the descending order of the output values of the SVM. The selected regions can be regarded as positive regions.
5. If the number of iteration is more than the pre-defined value r, finish the selection of positive images. In the experiment, we set r as 5.
6. Otherwise select one sixth of positive regions in the ascending order of the output values of the SVM as nega-

tive samples. Sample one sixth of negative background images, and add them to negative samples.
7. Train SVM, and jump back to (3).

The steps from Step 3 to 7 is iterative steps. It is the same way as [1] to employ an iterative process for selecting positive regions.

As the next step, to estimate the entropy of the image features of selected regions with respect to a generic distribution of image futures. To represent a generic model, we use the probabilistic Latent Semantic Analysis (PLSA) [3] which is the probabilistic method to identify latent topics with the given number of topics. PLSA was originally proposed as a probabilistic model to extract latent topics from text documents.

We need to obtain generic base topics in advance by the PLSA for computing the entropy. To obtain the generic base, we used about ten thousand images randomly picked up from the images gathered from the Web.

The PLSA model is represented as the generative model of each word w in a document d:

$$P(w,d) = P(d) \sum_{z \in Z} P(w|z) P(z|d) \tag{1}$$

where $z \in Z = (z_1, .., z_k)$ is a latent topic variable, k is the number of topics, $d \in D = (d_1, ..., d_N)$ is an image region expressed by the bag-of-features vector, and $w \in W = (w_1, ..., w_M)$ is one element of the BoF vector, which corresponds to a "visual" word. The joint probability of the observed variables, w and d, is the marginalization over the k latent topics Z. The parameters are estimated by the EM algorithm. In the experiments, we set 300 to the number of base topics k. We carry out this estimation of $P(w|z)$ in advance which is regarded as training process of the PLSA. For full explanation of the PLSA model refer to [3].

For each positive region i for the concept "X", we estimate $P(z|d_i^X)$ employing "fold-in heuristics" [3]. The entropy for the concept "X" $H_{img}(X)$ is given by

$$H_{img}(X) = -\sum_k P(z_k|X) \log_2 P(z_k|X) \tag{2}$$

$$where \;\; P(z_k|X) = \frac{1}{|I_{selected}|} \sum_{i \in I_{selected}} P(z_k|d_i^X) \tag{3}$$

and $|I_{selected}|$ is the number of selected positive regions. We summarize the procedure described above in Figure 1.

2.1.1. Image Representation

In this subsection, we describe about the region-based bag-of-features (BoF) representation [2] we use as an image representation to estimate image region entropy.

The main idea of the bag-of-features representation [2] is representing images as collections of independent local patches, and vector-quantizing them as histogram vectors. Before constructing the bag-of-features vector, we apply region segmentation for all the images. To obtain the region-based BoF vector, we extract the BoF vector from each region. As a region segmentation method, we use JSEG [5] after adjusting the parameters so as to generate about eight regions per image on average.

The main steps of the method are as follows:

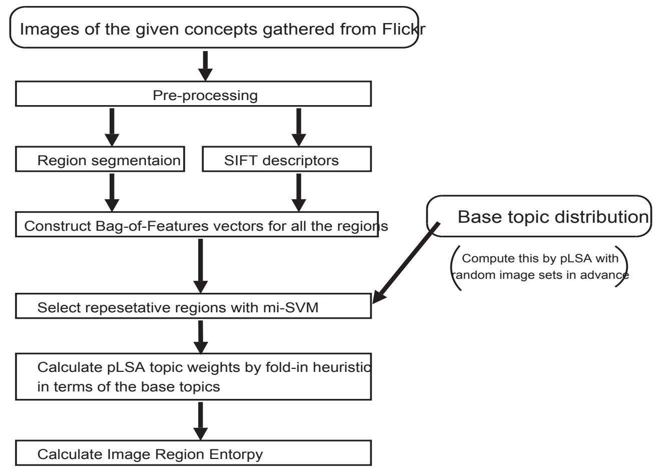

Fig. 1. Flow diagram of the procedure to obtain the "image region entropy".

1. Carry out region segmentation with the JSEG algorithm.
2. Sample 3000 patches per image randomly in the same way as [6].
3. Generate feature vectors for the sampled patches by the SIFT descriptor [7].
4. Construct a codebook with k-means clustering over all the extracted feature vectors. A codebook is constructed for each concept independently. We set k as 300.
5. Assign all SIFT vectors to the nearest codeword of the codebook, and convert a set of SIFT vectors for each region into one k-bin histogram vector regarding assigned codewords. In addition, background images which are prepared as negative training samples in advance are also divided into regions and converted the sets of SIFT vectors extracted from regions into k-bin histograms based on the same codebook.

2.2. Geo-location Entropy

We can obtain location information of downloaded images which is represented by a set of values of latitude and longitude from Flickr with FlickrAPI. In this work, we calculate entropy regarding geo-location in addition to image region entropy. To estimate geo-location entropy, we build a histogram regarding location distribution on each concept by dividing latitude and longitude by every 10 degrees as shown in Figure 2. Geo-location entropy $H_{geo}(X)$ is calculated by the following equation:

$$H_{geo}(X) = -\sum_i b_i \log_2 b_i \tag{4}$$

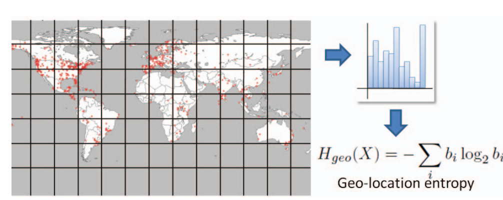

Fig. 2. The overview of "Geo-location entropy".

3. EXPERIMENTS

We examined image region entropy and geo-location entropy for 230 nouns shown in Table 1 and 100 adjectives shown in Table 2. For the experiments, we gathered 500 images for each concept from Flickr by providing the concept words as search keywords for FlickrAPI. Totally, we collected 115,000 geotagged photos. In the experiments, we used images including a given concept word as metadata as positive images, while images randomly selected from all the downloaded photos are used as negative images. In [1], they used only 150 adjectives and did not make an experiment with nouns. We imported 100 adjectives from the 150 adjectives, and for noun concepts we selected abstract nouns, person names and location names as well as nouns related to objects and scenes by hand in order to mix the words which are likely to be related to location and the words are not likely to be related to location.

Table 1. 230 noun concepts

africa	alexander	alligator	america	ant	asia	bach	banana	barbecue	battle	beach	beauty
beaver	bee	beer	beetle	board	boat	bob	book	box	bread	brother	bug
building	burger	bus	california	canada	candy	car	castle	cat	cedar	chair	chalk
cherry	chicken	china	circle	city	coffee	coke	color	computer	cookie	coral	crow
dandelion	daughter	desert	desk	dessert	deutschland	dish	doctor	dolphin	dragonfly	dream	eagle
edison	eel	egg	egypt	election	elevator	erica	europe	father	fern	field	fish
flea	flower	fly	fork	france	frog	fruit	game	gates	giraffe	goat	gorilla
grape	grass	grasshopper	gun	half	ham	hawk	height	hibiscus	hornet	house	icecream
india	insect	italia	ivy	japan	jellyfish	jump	kangaroo	killerwhale	lamp	lavender	lawn
leaf	lemon	level	library	light	lincoln	lion	lizard	locust	love	mangrove	manta
mantis	marriage	milk	mint	monkey	moon	mosquito	moss	moth	mother	mountain	mouse
mozart	museum	napoleon	newyork	octopus	owl	oyster	palm	paris	park	parrot	party
pen	penguin	phone	pine	pizza	plant	pool	pope	potato	pride	rabbit	rice
rome	rose	salad	salmon	santaclaus	school	sea	shakespeare	shark	ship	shrimp	sister
sky	snail	snake	son	sound	sport	square	starfish	steak	sun	sushi	sword
tea	teacher	temple	test	thomas	tiger	toad	tokyo	tool	town	tulip	tuna
turtle	usa	valley	village	whale	wine	worm	zoo	airplanes	backpack	bear	buddha
butterfly	cactus	cake	canoe	dice	dog	duck	eiffeltower	elephant	fireworks	goose	helicopter
horse	kayak	mars	mushroom	people	pyramid	rainbow	skyscraper	socks	spider	swan	tripod
watch	waterfall										

Table 2. 100 adjective concepts

aerial	ancient	antique	bad	beautiful	best	better	big	black	blue	botanical	bottom
bright	brown	cherry	classic	clean	clear	cold	colourful	concrete	cool	crazy	cute
dark	digital	dry	electric	empty	famous	female	first	general	good	grand	gray
great	green	happy	hard	heavy	high	historic	holy	hot	human	iced	interior
international	large	latest	long	male	medieval	military	mobile	modern	more	most	national
natural	nautical	new	nice	old	older	oldest	open	orange	outdoor	pink	present
public	purple	rainy	red	rural	rusted	scenic	second	sexy	short	small	special
sunny	sweet	top	traditional	tropical	twin	underwater	urban	vintage	warm	welcome	white
wide	wild	wooden	yellow								

We show the top 10 and bottom 10 results in terms of region entropy and geo-location entropy for noun concepts in Table 3 and Table 4, respectively. In the same way, we show the top 10 and bottom 10 results in terms of region entropy and geo-location entropy for adjective concepts in Table 5 and Table 6, respectively.

Figure 3 and 4 show the relations between image region entropy (x-axis) and geo-location entropy (y-axis) regarding 230 nouns and 100 adjectives, respectively. Table 7 represents the cross table between image region entropy and geo-location entropy regarding some concepts picked up from Figure 3. This table shows concepts which has relatively larger or smaller image region entropy and larger or small image geo-location entropy at the same time. In this table, a concept with larger entropy means the concept is included in top 46 concepts in terms of entropy ranking, and a concept with smaller entropy means the concept is included in bottom 46 concepts in terms of entropy ranking.

3.1. Discussion on Nouns

As a prominent tendency, while geo-location entropy of location concepts such as “Rome”, “Africa” and “Japan” and name concepts of historical persons such as “Mozart” and “Napoleon” were small, image region entropy of locations and person names were larger. “Sox” also belongs to this category, since the “Sox” image set gathered from Flickr includes many “Red Sox” photos which is a popular baseball team in US.

Table 3. Image region entropy $H_{img}(X)$ of top 10 and bottom 10 of 230 nouns

10 smallest		10 largest	
concepts	$H(X)$	concepts	$H(X)$
sun	3.6497	usa	7.4020
rainbow	4.5538	backpack	7.4086
moon	4.6686	italia	7.4111
dragonfly	4.7550	town	7.5177
sky	5.1049	santa-claus	7.5431
mantis	5.1897	house	7.5598
airplanes	5.3851	napoleon	7.5704
egg	5.2288	school	7.6173
bee	5.4210	lincoln	7.7327
light	5.4524	mozart	7.8349

Table 4. Geo-location entropy $H_{geo}(X)$ of top 10 and bottom 10 of 230 nouns

10 smallest		10 largest	
concepts	$H(X)$	concepts	$H(X)$
deutschland	0.2602	sea	5.4936
rome	0.3843	mother	5.5114
tokyo	0.6253	teacher	5.5417
paris	0.6730	lizard	5.5448
eiffel-tower	0.7461	fruit	5.5779
california	0.8776	hibiscus	5.5856
new-york	1.0264	ant	5.6147
italia	1.3105	coral	5.6565
france	1.4833	fish	5.7831
egypt	1.8476	mosquito	5.9759

Table 5. Image region entropy $H_{img}(X)$ of top 10 and bottom 10 of 100 adjectives

10 smallest		10 largest	
concepts	$H(X)$	concepts	$H(X)$
orange	5.5608	wooden	7.3469
yellow	5.6780	oldest	7.3728
dark	5.7451	concrete	7.3827
latest	5.8305	older	7.3905
white	5.9033	international	7.3994
clean	5.9623	traditional	7.4268
pink	5.9857	general	7.5015
botanical	6.0937	public	7.5253
happy	6.1351	vintage	7.5972
nautical	6.2085	historic	7.6257

Table 6. Geo-location entropy $H_{geo}(X)$ of top 10 and bottom 10 of 100 adjectives

10 smallest		10 largest	
concepts	$H(X)$	concepts	$H(X)$
medieval	3.4364	yellow	5.2554
modern	3.7842	blue	5.2709
new	3.8983	cute	5.2842
rainy	4.1315	small	5.3409
grand	4.1568	beautiful	5.3897
cherry	4.1782	dry	5.4144
big	4.2196	colourful	5.4545
nice	4.2972	tropical	5.5687
public	4.3603	underwater	5.7688
historic	4.4289	traditional	5.8487

As shown in Figure 5, geo-location of location names and person names are strongly tied with the concepts themselves, while images related to them includes various appearance since they are relatively abstract concepts rather than physical concepts.

We found that for the concepts related to sky such as “sun” (Figure 6) and “rainbow” their image region entropy were smaller, while geo-location entropy were larger. This is because appearances related to such concepts tends to be very similar or almost the same everywhere over the world. So geo-location entropy became high, and image region entropy became low.

“Tulip” (Figure 7) was the only concept which has low image entropy and low geo-location entropy. “Tulip” was mainly concentrated on the United States and Europe, espe-

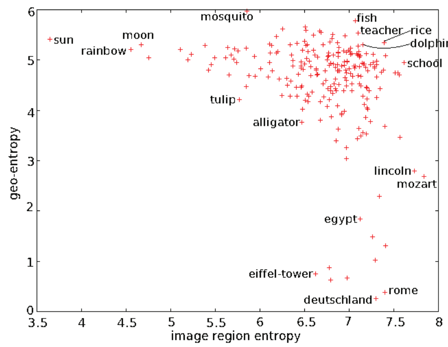

Fig. 3. Relations of visual and geo- entropy for 230 nouns.

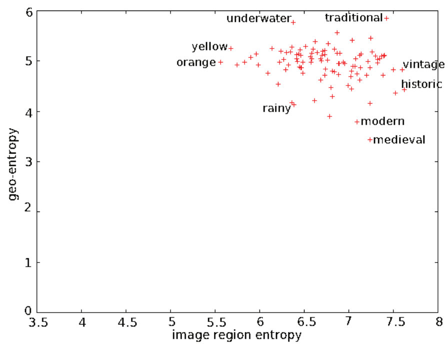

Fig. 4. Relations of visual and geo- entropy for 230 nouns.

Table 7. Cross table between image region entropy and geo-location entropy for nouns

		Image region entropy (IE): Smaller IE	Image region entropy (IE): Larger IE
Geo-location entropy (GI)	Smaller IG	tulip	france china new-york deutschland africa japan rome usa Italia napoleon lincoln morzart pope killer-whale chalk socks shakespeare thomas pride
	Larger IG	sun rainbow sky moon airplane bug dream insect beach mosquito beetle beauty banana mangrove ant flower	dolphin rice

cially , Holland, and most of the "tulip" photos included tulip flowers and tulip farms.

Both image and geo-location entropies on "Rice" (Figure 8) were large. Although "Rice" is a food concept which are common everywhere over the world, the way to cook is different depending on countries greatly. Moreover it also means the name of the ex-spokesperson of the US government.

3.2. Discussion on Adjectives

Compared Table 3 with Table 5 and compared Table 4 with Table 6, both image region and geo-location entropies on adjectives were higher than entropies on nouns on the average. Since adjectives are abstract and general concepts, there are no typical visual appearances and they do not depends on locations as much as nouns in general. In Table 5, the top 10 adjectives with the smallest entropy include many color concepts. This is because images related to color concepts contained many uniform regions filled with the corresponding color.

Fig. 5. "Mozart" images over the world.

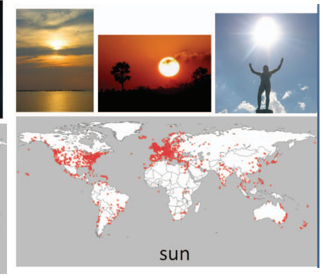

Fig. 6. "Sun" images over the world.

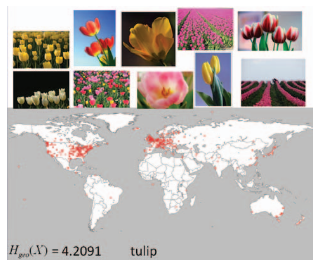

Fig. 7. "Tulip" images over the world.

Fig. 8. "Rice" images over the world.

4. CONCLUSION AND FUTURE WORK

In this paper, we proposed a novel method to analyze relationship between word concepts and geographical locations by using a large number of geotagged images on the photo sharing Web sites such as Flickr, and we proposed using both image region entropy and geo-location entropy to analyze relations between visual concepts and their locations. In the experiment, we analyzed relations between image region entropy and geo-location entropy in terms of 230 nouns and 100 adjectives, and we found that the concepts with low image entropy tend to have high geo-location entropy and vice versa.

For future work, we plan to investigate more deep relations between locations and concepts. We like to propose a new method to discover differences of a concept depending on locations like Western-style house is different from Japanese-style house. Finally we will conduct methods to discriminate concepts which have larger cultural differences from concepts with low image entropy.

5. REFERENCES

[1] 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, pp. 420–423.

[2] 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*, 2004, pp. 59–74.

[3] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Machine Learning*, vol. 43, pp. 177–196, 2001.

[4] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support Vector Machines for Multiple-Instance Learning," in *Advances in Neural Information Processing Systems*, 2003, pp. 577–584.

[5] 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.

[6] E. Nowak, F. Jurie, W. Triggs, and M. Vision, "Sampling strategies for bag-of-features image classification," in *Proc. of European Conference on Computer Vision*, 2006, pp. IV:490–503.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.