GeoVisualRank: A Ranking Method of Geotagged Images Considering Visual Similarity and Geo-location Proximity

Hidetoshi Kawakubo and Keiji Yanai
Department of Computer Science, The University of Electro-Communications
kawaku-h@mm.cs.uec.ac.jp, yanai@cs.uec.ac.jp

Categories and Subject Descriptors: H.3.3 Information Search and Retrieval: Miscellaneous

General Terms: Algorithms, Experimentation, Measurement

Keywords: geotagged images, pagerank, visualrank

1. INTRODUCTION

Recently, the number of geotagged images are increasing explosively on the Web due to the spread of on-line Web albums with maps such as Panoramio and Flickr, and GPS-equipped cellular phones and cameras such as iPhone and Android smart phones. In fact, Flickr announced they had more than one hundred million geotagged photos in 2009, and more than one hundred thousand geotagged photos are being uploaded per day.

To search such huge geotagged image databases, effective ranking methods for geotagged images are needed. As ranking methods for normal images, "VisualRank"[3] has been well known. Although VisualRank can estimate representativeness of images based on visual similarity between images, it cannot take account of geo-location information attached to geotagged images. Then, in this paper, we propose a new method to rank geotagged images "GeoVisualRank", which is an extension of VisualRank for considering geo-locations. The proposed method generates ranking of geotagged images considering both visual similarity and proximity of locations where photos are taken. This method can be considered as selecting representative geotagged images regarding the given reference places. For example, if we have a large number of "cake" photos geotagged over the world, we can get to know typical "cakes" regarding various regions or countries over the world by applying the proposed method. The proposed method can answer the questions such as "what do cakes in France look like ?" and "what do cakes in China look like ?".

To do the same as this with the existing methods, we have to employ two-step processing; the first step is selection of geotagged images around the reference locations, and the second step is selection of representative images from selected images in the first step. To select representative images, they carry out image clustering and selection of representative images from large clusters (e.g. [1]).

On the other hand, in the proposed method, we extend "VisualRank"[3] for geotagged image ranking naturally. VisualRank is a image ranking method based on "PageRank" which is based on Markov chain. In PageRank, the transition matrix is computed based on links between Web pages, while the transition matrix of the Markov chain in VisualRank is computed based on visual similarity between images. The rank of Web pages or images are estimated according to the probability of the steady state distribution of the Markov chain. In the proposed "GeoVisualRank", a transition matrix is constructed based on visual features in the similar way as VisualRank, and a bias vector is generated based on the degree of the proximity of images to the given reference points, while a bias vector is set as a uniform vector in VisualRank. This bias vector setting makes GeoVisualRank take into account the proximity to the given reference locations by giving larger bias values to the images taken at the places close to the reference locations.

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0632-4/11/03.

2. METHOD

Before describing the proposed method, firstly, we explain about VisualRank briefly, and then we describe GeoVisualRank.

2.1 VisualRank

VisualRank is an image ranking method based on PageRank. PageRank calculates ranking of Web pages using hyper-link structure of the Web. The rank values are estimated as the steady state distribution of the random-walk Markov-chain probabilistic model.

VisualRank uses a similarity matrix of images instead of hyperlink structure. Eq.(1) represents an equation to compute VisualRank.

$$\boldsymbol{r} = \alpha S\boldsymbol{r} + (1-\alpha)\boldsymbol{p}, \quad (0 \leq \alpha \leq 1) \tag{1}$$

S is the column-normalized similarity matrix of images, $\boldsymbol{p}$ is a damping vector, and $\boldsymbol{r}$ is the ranking vector each element of which represents a ranking value of each image. α plays a role to control the extent of effect of $\boldsymbol{p}$. The final value of $\boldsymbol{r}$ is estimated by updating $\boldsymbol{r}$ iteratively with Eq.(1). Because S is column-normalized and the sum of elements of $\boldsymbol{p}$ is 1, the sum of ranking vector $\boldsymbol{r}$ does not change. Although $\boldsymbol{p}$ is set as a uniform vector in VisualRank as well as normal PageRank, it is known that $\boldsymbol{p}$ can plays a bias vector which affects the final value of $\boldsymbol{r}$ [2].

2.2 GeoVisualRank

We propose using a geo-location-based bias vector in calculation of VisualRank instead of a uniform damping vector. We call this as "GeoVisualRank". GeoVisualRank computes ranking of geotagging images considering both visual similarity and geo-based bias. We compute a geo-location bias vector based on distances between images and given reference locations.

Since GeoVisualRank uses the iterative computation shown in Eq.(1) in the same as VisualRank, we describe how to construct a visual similarity matrix S and a geo-location bias vector $\boldsymbol{p}$ in the rest of this section.

Jing et al. used the number of identical SIFT local descriptors between two images as visual similarity in VisualRank [3]. This approach is effective for concepts have specific figure such as commercial products. In this study, we use color histogram and Bag-of-Features representation of SIFT features as visual features in order to apply the proposed method to various kinds of concepts including nouns and adjectives. These kinds of visual features are widely used in the object recognition research, since they have high ability to express various visual concepts.

We calculate histogram intersections as visual similarity between image features. As shown in Eq.(2), we generate a similarity matrix S by combining a color histogram similarity matrix S_{color} with Bag-of-Features similarity matrix S_{BoF} with a weighting constant β. In the experiment, we set β as 0.5.

$$S_{combine} = \beta\, S_{color} + (1-\beta)\, S_{BoF}, \quad (0 \leq \beta \leq 1) \tag{2}$$

In GeoVisualRank, reference locations need to be given to calculate a geo-location bias vector. GeoVisualRank gives higher PageR-

Figure 1: 'House' without location bias.

Figure 2: 'House' in Sydney ($\alpha = 0.95$).

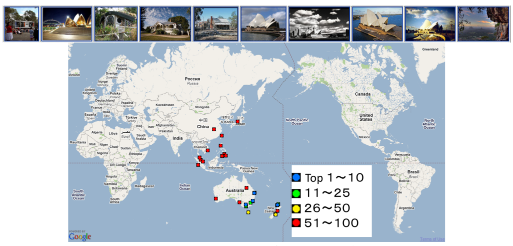

Figure 3: 'House' in Sydney ($\alpha = 0.85$).

ank values to geotagged images the geotag location of which are closer to the given reference locations. To do that, we give larger bias values to the image close to the given reference locations, while we give less bias values to the images far from the reference locations.

At first, we calculate a spherical distance between a reference location Ref_j and a geotag location of image i as shown in Eq.(5). Next, we obtain a unnormalized geo-location bias vector as shown in Eq.(3) and a L1-normalized bias vector as shown in Eq.(4).

$$p'_i = 1 - \min_j D_{i,j}/\pi \quad (3)$$

$$p_i = p'_i/||p'_i||_1 \quad (4)$$

$$D_{i,j} = \cos^{-1}\big(\sin(lat_i)\sin(lat_{\mathrm{Ref}_j}) + \cos(lat_i)\cos(lat_{\mathrm{Ref}_j})\cos(long_i - long_{\mathrm{Ref}_j})\big) \quad (5)$$

A geo-location bias vector can be constructed in other ways. For example, we can use a negative bias which is defined by $p'_i = \min_j D_{i,j}/\pi$ instead of Eq.(3). Using this, we can obtain images far from the reference points in the higher rank. Instead of physical distance, we can use political or cultural distance based country borders or regional borders to compute a geo-location bias vectors as well.

3. EXPERIMENTS

We gathered 2000 geotagged photos per concept for 350 concepts including 250 noun words and 100 adjective words from Flickr by providing concept words for Flickr WebAPI. The noun concepts include objects, scenes, animals, plants, landmarks and place names. After collecting geotagged images, we calculate GeoVisualRank for each concept by providing ten representative cities over the world including New York, San Francisco, Sydney, Paris, Cairo, Tokyo, Rio de Janeiro, Delhi, Beijing, and Cape Town. Although we show the results using just one city as a reference location due to space limitation, we can use multiple cities as reference locations. All the results can be seen at http://mm.cs.uec.ac.jp/geovisualrank/.


Figure 4: 'Pyramid' in Cairo.


Figure 5: 'Pyramid' in Paris.


Figure 6: 'Pyramid' in Rio de Janeiro.


Figure 7: 'Pyramid' <u>far from</u> Cairo.


Figure 8: 'Traditional' in Tokyo.


Figure 9: 'Traditional' in Sydney.


Figure 10: 'Traditional' in Rio de Janeiro.

Here we show some part of results. Fig.1 shows the top 10 images and a map for "house" using a uniform weight, which is the same result obtained by VisualRank. The map shows the distribution of the top 100 images in terms of GeoVisualRank values. Most of the selected image represents Western-style houses. Fig.2 and Fig.3 shows the results with Sydney as a reference point in case of setting 0.95 and 0.85 to α, respectively. The map in case that α is 0.85 shows that the top 100 images are distributed closer to Sydney than the map in case that α is 0.95. This shows that we can adjust the balance between visual similarity and geo-location proximity by changing the value of α.

Fig.4, Fig.5 and Fig.6 correspond to the top 10 images in terms of GeoVisualRank for "pyramid" on Cairo, Paris, and Rio de Janeiro with $\alpha = 0.85$. The result on Cairo includes many world-famous Giza pyramids, while the result on Paris contains the pyramid-structured building of the Louvre Museum. The result on Rio de Janeiro includes several Mexican pyramids in not Brazil but Mexico, since no typical "pyramid" exists in Brazil. These results indicate that GeoVisualRank can discover representative "pyramid" images regarding given reference locations. On the other hand, Fig.7 is the example of a negative geo-location bias, which shows "pyramid" far from Cairo. This result includes some pyramid-style buildings found in US East Coast and Australia.

Fig.8, Fig.9 and Fig.10 show the results for an adjective concept, "traditional", on Tokyo, Sydney, and Rio de Janeiro. All the results include people wearing traditional clothes.

4. CONCLUSION

In this paper, we proposed a new method to rank geotagged images, GeoVisualRank, which considers both visual similarity and geo-location proximity. The experimental results confirmed that GeoVisualRank has ability to combine both visual similarity and location proximity, which is a natural extension of VisualRank for geotagged images.

5. REFERENCES

[1] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *WWW*, 2009.
[2] T. Haveliwala. Topic-sensitive PageRank: A context-sensitive ranking algorithm for web search. *IEEE trans. on KDE*, 2003.
[3] Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. *IEEE Trans. on PAMI*, 2008.