VISUALIZATION OF REAL-WORLD EVENTS WITH GEOTAGGED TWEET PHOTOS

Yusuke Nakaji and Keiji Yanai

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

ABSTRACT

Recently, microblogs such as Twitter have become very common, which enable people to post and read short messages from anywhere. Since microblogs are different from traditional blogs in terms of being instant and on the spot, they include much more information on various events happened over the world. In addition, some of the messages posted to Twitter include photos and geotags as well as texts. From them, we can get to know what and where happens intuitively.

Then, we propose a method to select photos related to the given real-world events from geotagged Twitter messages (tweets) taking advantage of geotags and visual features of photos. We implemented a system which can visualize real-world events on the online map.

Index Terms— Twitter, Tweet photo, event visualization, geotag, mean-shift

1. INTRODUCTION

Recently, smart phones such as iPhone and Android phones have become very common. At the same time, microblogs such as Twitter have become to gather many users, since microblogs enable people to post and read short messages from anywhere using smart phones. Since microblogs are different from traditional blogs in terms of being instant and on the spot, they includes much more information on events happened over the world. In addition, some of the messages posted to Twitter, which are commonly called as "tweets", include photos and geo-tags as well as texts. By using photos and geotags embedded in tweets, we can get to know what and where happened over the world intuitively.

So far, there has been many works on Twitter. Some of them handle geotags. By analyzing geotagged tweets, we can detect events with their locations. For example, Sakaki et al. proposed a method to track typhoon move and monitor earthquake using geotagged tweets [1]. They regarded Twitter users as "social sensors" which react real-world events including natural phenomenon such as earthquake and typhoon and human events such as festivals and sport events.

On the other hand, Lee et al. proposed a method to detect local events automatically by monitoring the number of tweets, the number of tweet users, and movement of tweet users within each local area [2]. If a sudden rise in these numbers is detected, some kinds of "events" are regarded as being taking place. Their methods succeeded in detecting special local festivals and famous seasonal festivals.

In this way, most of the works on Twitter utilize texts as well as geotags embedded in tweets. On the other hand, none of the works on Twitter focus on photos attached to tweets, as long as we know. Then, in this paper, we propose a method to detect representative photos related to the given events in the real world automatically from tweets containing geotags and photos using visual analysis as well as geotag analysis. To provide events to the system, we can use time periods and locations as well as keywords. By comparing representative tweet photos over the time periods or over some locations, we can get to know transition of events regarding locations and time.

Our final objective is to implement a system which can let us know what happens currently and in the past over the world visually, which includes detection of events and selection of representative photos corresponding to the detected events.

In this paper, we focus on only selection of representative photos related to events which are assumed to be given by hand. We propose a method to do that and show some preliminary results.

2. OVERVIEW

At first, we have to decide areas and time periods where and when we like to detect events. In addition, we can use optional keywords to limit tweets as well, which helps examine specific events such as typhoon. Since this paper focuses on only selecting representative photos corresponding to the given event, detecting salient events will be one of our future works.

Firstly, we gather geo-photo-tweets from Twitter and store them into the in-house database. Secondly, we download photos contained in the gathered tweets. Thirdly, we extract visual features from the downloaded images. Fourthly, we cluster images and geotags embedded in the tweets related to the given events, and finally select representative photos regarding the given locations and time periods by using a method based on the GeoVisualRank method [3].

To summarize it, the method to detect representative event photos from geo-photo-tweets is as follows:

1. Gather tweets with geotags and photos using Twitter Streaming API.

2. Extract visual features from tweet photos using color histograms and SURF-based bag-of-feature (BoF) representation.
3. Perform k-means clustering over visual features and mean-shift clustering over geotags.
4. Select representative tweet photos with time-extended GeoVisualRank regarding each place and each time period.

2.1. Collecting Geo-tweets with Photos

We use Twitter Streaming API to collect geo-tweets with photos. We have been collecting tweets for about one years so far. Since February 2011 for one year, we collected about 18 million geo-photo-tweets and stored all of them into our in-house geo-photo-tweet database.

Since each tweet itself includes not photos, but URLs of photos, we extract URLs and download the corresponding photos from photo storage sites for twitter such as Twitpic, yfrog and instagr.am.

In fact, we limit geo-photo-tweets to download corresponding photos with only the tweets posted within Japan currently, since we can catch a geo-photo-tweet every two seconds and the total amount is too huge to download all the photos.

2.2. Extraction of Visual Features

As visual features to represent tweet photos, we use color histogram and Bag-of-Features representation of SURF features [4] with the codebook the size of which is 2000.

2.3. Clustering Photos and Geotags

Before performing clustering, we search the geo-photo-tweet database for the geo-photo-tweets which meet the given conditions on locations, time periods and optional keywords. And then, we perform k-means clustering for the selected tweet-photos on visual features with the number of clusters set as $n/10$ where n is the number of the selected photos. This visual clustering result is used for showing similar photos to representative photos on the system.

At the same time, we carry out mean-shift clustering [5] on geotags of the selected geo-photo-tweets in order to select representative areas. Mean-shift clustering is based on kernel density estimation, which consists of iteration of setting circular regions $T_h(x)$ with radius h where its center is a data point x and updating the value of the point x to the mean value weighted by Gaussian kernel function $K(x_i; x, h)$ shown in Equation (1). This iteration is repeated until being convergent.

$$m(x) = \frac{\sum^{x_i \in T_h(x)} K(x_i; x, h) w(x_i) x_i}{\sum_{i=1}^{n} K(x_i; x, h) w(x_i)} \quad (1)$$

$$K(x_i; x, h) = exp\Big(-\frac{||x - x_i||^2}{h^2}\Big) \quad (2)$$

In the experiment, we set $w(x_i)$ as 1 and h so that the radium of each cluster is 50km.

2.4. Select Representative Photos

We use a method based on the GeoVisualRank method [3] to select representative photos regarding the given locations and time periods.

GeoVisualRank [3] is the extension of VisualRank [6] which is a method to select representative images from a set of images. VisualRank is based on the widely-known Web page ranking algorithm, PageRank [7], 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 and GeoVisualRank 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.

VisuakRank is computed with the following equation:

$$\boldsymbol{r} = \alpha S \boldsymbol{r} + (1 - \alpha)\boldsymbol{p}, \quad (0 \le \alpha \le 1) \quad (3)$$

where 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 $\boldsymbol{p}$. In the experiments, α is set as more than 0.85 in the same as [6]. The final value of $\boldsymbol{r}$ is estimated by updating $\boldsymbol{r}$ iteratively with Equation (3). Because S is column-normalized and the sum of elements of $\boldsymbol{p}$ is 1, the norm 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 $\boldsymbol{r}$ [8].

In GeoVisualRank [3], a geo-location-based bias vector is used in calculation of VisualRank instead of a uniform damping vector to take account of geo-location proximity. GeoVisualRank computes ranking of geotagging images considering both visual similarity and geo-based bias. In this paper, we extend it by making it take account of time proximity. We compute a bias vector based on time differences between the time when the corresponding tweet is posted and a given time as well as distances between geotagged location of a tweet photo and a given reference location.

In the extended GeoVisualRank we proposed here, reference locations and reference times need to be given to calculate a bias vector. It gives higher PageRank values to geotagged images the geotag location of which are closer to the given reference locations and the taken time of which are closer to the given reference time. To do that, we set each element of the bias vector as follows:

$$\boldsymbol{p}^{geo}(i) = \begin{cases} 1/n_{geo} & (g_i \in C^{geo}) \\ 0 & (g_i \notin C^{geo}) \end{cases} \quad (4)$$

$$\boldsymbol{p}^{time}(i) = \begin{cases} 1/n_{time} & (t_i \in C^{time}) \\ 0 & (t_i \notin C^{time}) \end{cases} \quad (5)$$

where C^{geo} represents a location cluster containing the given reference location, C^{time} represents time cluster containing the given time, and n_{geo} and n_{time} represents the number of images belonging to C^{geo} and C^{time}, respectively. Note that time clusters are constructed by dividing times every six hours as shown in Table 1.

Table 1. Time clusters

No.	time period
0	Midnight∼6 A.M. (after midnight)
1	6 A.M. ∼Noon (morning)
2	Noon∼6 P.M. (afternoon)
3	6 P.M. ∼Midnight (night)

Finally the bias vector $\boldsymbol{p}$ is obtained by linear combination of $\boldsymbol{p}^{geo}$ and $\boldsymbol{p}^{time}$ as shown in Equation (6).

$$\boldsymbol{p} = \beta \boldsymbol{p}^{geo} + (1-\beta)\boldsymbol{p}^{time} \quad (0 \leq \beta \leq 1) \qquad (6)$$

In the experiments, we set β as 0.5.

Since we use color histogram and Bag-of-Features representation of SURF features as visual features, we calculate histogram intersections as visual similarity between image features. As shown in Equation (7), 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 γ as 0.5.

$$S_{combine} = \gamma\ S_{color} + (1-\gamma)\ S_{BoF} \quad (0 \leq \gamma \leq 1) \qquad (7)$$

3. EXPERIMENTS

As the experiments, we tried to select event photos related to typhoon on September 2011 (Event no.1), new year habits in Japan (Event no.2) and the huge earthquake hit on Northern-East Japan in March 11th 2011 (Event no.3) as shown in Table 2.

3.1. Selected Photos for Each Event

For Event No.1, before the processing, we selected 616 geo-photo-tweets with keyword search for tweet messages and time periods limitation on September 2011. And then, we applied clustering on geotags and visual features, and compute the extended GeoVisualRank. Finally we obtained representative photos as shown in the upper left of Figure 1. This figure shows the representation photos on "typhoon" selected in the largest time cluster in the largest geotag cluster which corresponds to the area around Tokyo. Most of the selected photos shows the black sky which is typical way of the sky when typhoon is approaching.

For Event No.2, we selected photo-tweets in the same way as Event No.1 at first. The upper left figure of Figure 1 shows representative photos related to "New Year's life in Japan" (Event No.2), which represents photos taken at shrines or temples. We have a traditional custom to go shrines or temples to pray safe life of the new year in Japan. The selected photos reflected it.

For Event No.3, we did not use keyword search to pick up geo-photo-tweets from the database. We used only the condition on time and area, since we liked to examine the difference of the kinds of representative photos depending on locations in the day when the biggest event for these ten years happened in Japan. The time period was set from the day, March 11th, for two days.

The bottom figure of Figure 1 shows the representative photos taken in Tokyo, which expresses the scene that the trains stopped and people walks to home instead of getting trains.

3.2. Comparison of the Representative Photos Depending on Locations

In this subsection, we examine the differences of the representative photos on the same event depending on locations. Figure 2 shows results of the four location, Fukushima, Tokyo, Okayama and Fukuoka, on Event No.3 (earthquake) in the afternoon of March 11th 2011. Some of the photos in Fukushima and Tokyo which the earthquake affected greatly show the rooms where many books are scattered, broken building, and many people walking to home due to train stop. On the other hand, most of the photo in Okayama and Fukuoka where they are enough far from seismic center of the earthquake consists of everyday life photos such as foods and normal street scene.

From these results, we can get to know areas where the earthquake affected intuitively.

3.3. Comparison of the Representative Photos Depending on Times

In this subsection, we examine the differences of the representative photos on the same event depending on times. Figure 3 shows the representative photos on the four time periods in the New Year's day (Event No.2) in Tokyo. From midnight to 6AM (left top), the photo of the first sunrise of the year were selected, from 6AM to 12PM (right top) the photo of the temples and shrines were selected, and from 12PM to 6PM (left bottom) and from 6PM to 12AM (right bottom) many photos related to new year's day foods were selected.

From these results, we can get to know transition of way of the life depending on times on the New Year's day in Tokyo visually.

4. CONCLUSIONS

In this paper, we proposed a method to select photos related to the given real-world events from geotagged Twitter messages (tweets) taking advantage of geotags and visual features of photos. We implemented a system which can visualize real-world events on the online map. The experiments showed that the method was helpful to visualize the differences of the

Fig. 1. Representative tweet photos in the largest time clusters in Tokyo (upper left: typhoon (Event No.1), upper right: New Year (Event No.2), bottom: earthquake (Event No.3)

representative photos related to the given event depending on locations and times.

In this paper, we focused on only selection of representative photos related to events which are assumed to be given by hand . As future work, we will work on a method to detect event automatically and in the real-time way, and combine it with the proposed method in this paper to implement a system which can let us know what happens currently and in the past over the world visually.

5. REFERENCES

[1] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: real-time event detection by social sensors," in *Proc. of the International World Wide Web Conference*, pp. 851–860, 2010.

[2] R. Lee and K. Sumiya, "Measuring geographical regularities of crowd behaviors for twitter-based geo-social event detection," in *Proc. of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pp. 1–10, 2010.

[3] H. Kawakubo and K. Yanai, "Geovisualrank: A ranking method of geotagged images considering visual similarity and geo-location proximity," in *Proc. of the ACM International World Wide Web Conference*, 2011.

[4] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. of European Conference on Computer Vision*, pp. 404–415, 2006.

[5] D. Comaniciu and P. Meer, "Mean shift analysis and applications," in *Proc. of International Conference on Computer Vision*. Ieee, vol. 2, pp. 1197–1203, 2002.

[6] Y. Jing and S. Baluja, "Visualrank: Applying pagerank to large-scale image search," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 11, pp. 1877–1890, 2008.

[7] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proc. of the Seventh International World Wide Web Conference*, 1998.

[8] T.H. Haveliwala, "Topic-sensitive PageRank: A context-sensitive ranking algorithm for web search," *IEEE trans. on Knowledge and Date Engneering*, vol. 15, no. 4, pp. 784–796, 2003.

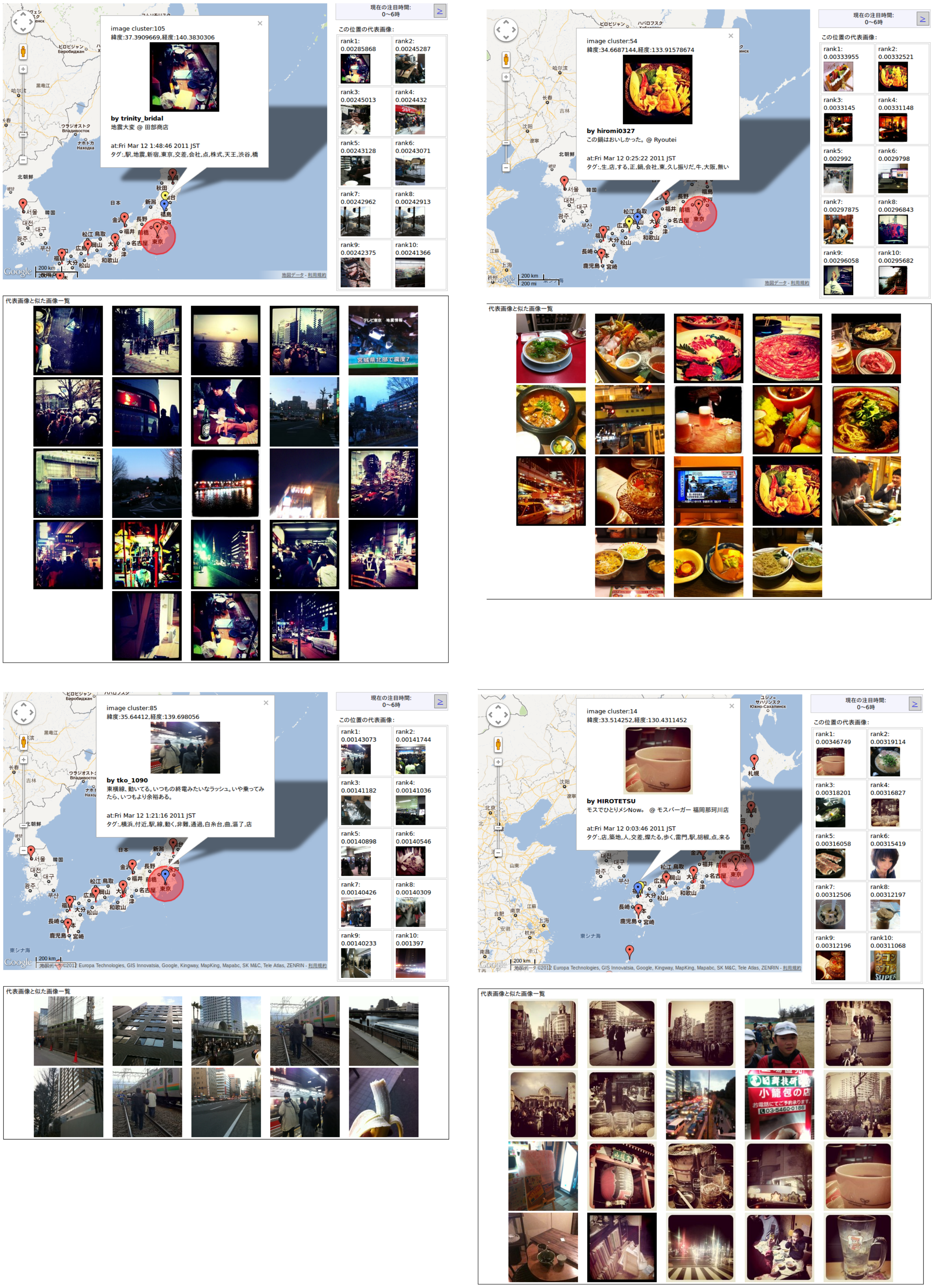

Fig. 2. Representative tweet photos related to "earthquake" (left top: Fukushima, left bottom: Tokyo, right top: Okayama, right bottom: Fukuoka)

Table 2. The conditions of three kinds of experiments

No.	Search keyword (OR search)	time periods	area	# tweet photos
1	typhoon	Sep. 1 2011 to Sep. 30 2011	posted from Japan	616
2	New year's day first-shrine-visit first-sunrise	Jan. 1 2012 to Jan. 20 2011	posted from Japan	1400
3	(none)	Mar. 11 2011 to Mar. 12 2011	posted from Japan	1080

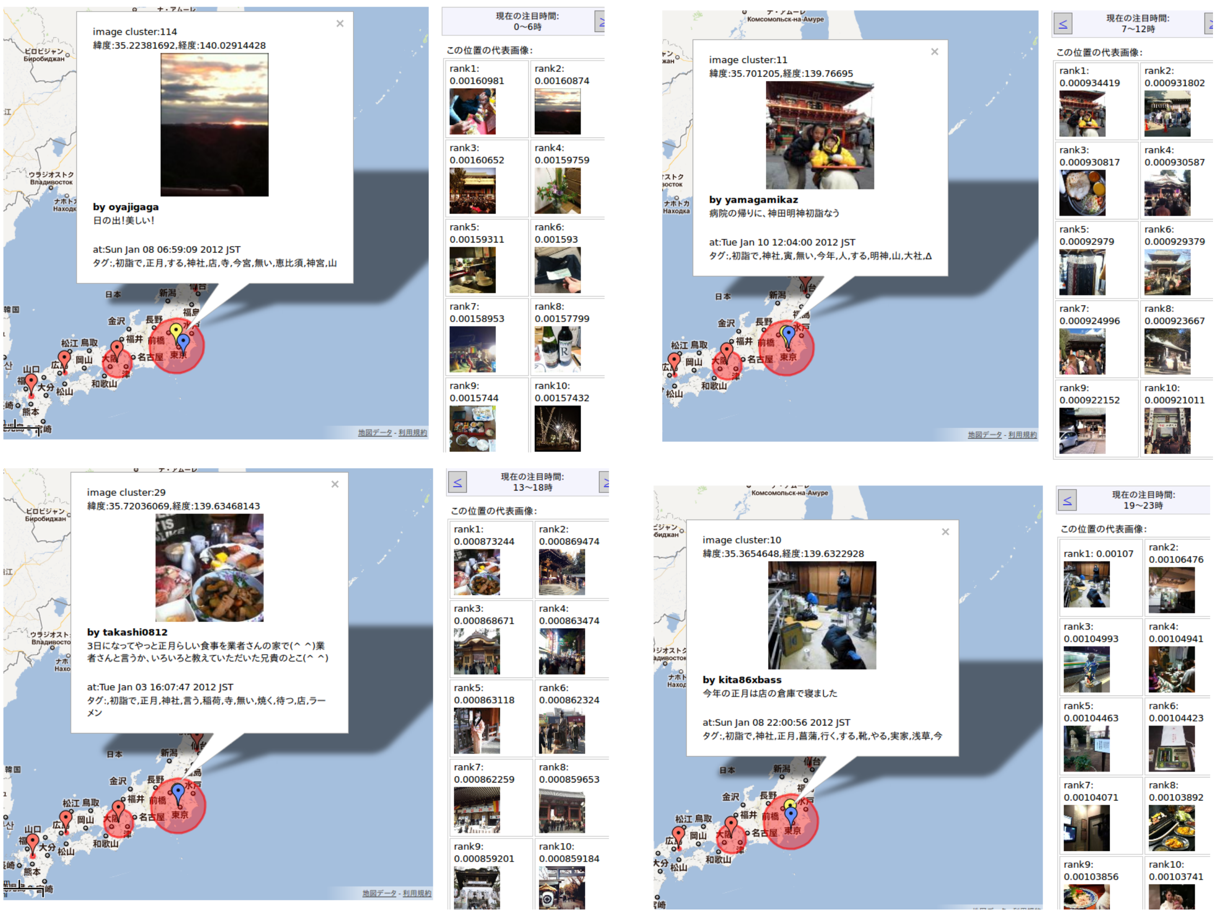

Fig. 3. Representative tweet photos related to "New Year" on the New Year's day (Jan. 1st) (left top: 0AM∼6AM, right top: 6AM∼12PM, left bottom:12PM∼6PM, right bottom:6PM∼0AM)