Real-time Photo Mining from the Twitter Stream: Event Photo Discovery and Food Photo Detection

(Invited Paper)

Keiji Yanai Takamu Takamu Yoshiyuki Kawano
Department of Informatics, The University of Electro-Communications, Tokyo
Chofu-shi, Tokyo, 182-8585, JAPAN
{*yanai,kaneko-t,kawano-y*}*@mm.inf.uec.ac.jp*

***Abstract*—So many people are posting photos as well as short messages to Twitter every minutes from everywhere on the earth. By monitoring the Twitter stream, we can obtain various kinds of photos with texts. In this paper, as case studies of real-time Twitter photo mining, we introduce our current on-going projects on event photo discovery and food photo mining from the Twitter stream.**

***Keywords*-real-time photo mining, Twitter stream, event detection, food photo recognition**

I. Introduction

So many people are posting photos as well as short messages to Twitter every minutes from everywhere on the earth. People send photos with short messages to Twitter soon after taking photos on the spot. Therefore, by monitoring the Twitter stream and picking up Tweet photos, we get to know the current state of the world visually. This is a biggest difference of Twitter to other social media. By taking account of this unique characteristic of Twitter, we are working on mining photos from the Twitter stream [1], [2], [3], [4].

In [1], we proposed a real-time geotagged tweet photo mapping system, "World Seer", which visualizes photo tweets with geotags on the Google Maps in real-time way as well as stores information on geo-photo tweets to our database continuously by monitoring the Twitter streaming via the Twitter Streaming API. We have been collecting both geo-photo tweets and photo tweets without geotags continuously since February 2011. Currently, we are collecting more than five hundred thousand geo-photo tweets a day and more than two million photo tweets a day. In fact the number of photo tweets varies depending on the week of the day. In weekends, the number increases by about 20% compared to weekdays. At present, our geo-photo tweet database has more than 500 million geo-photo tweets and more than two billion photo tweets without geotags, which can be regarded as a huge photo database.

Although Twitter provides Twitter API which enables us to search the Twitter database by keywords, time and date or geotags in addition to Twitter Streaming API which just transmits current tweets continuously with keyword filtering or geo-location filtering, the search function of Twitter API is limited in terms of the number of retrieved tweets (100) and the number of requests per unit time (450/3hours), which is not suitable for research purpose on a large-scale analysis on Twitter data. That is why we are collecting photo tweets from the Twitter stream continuously and storing them to our database.

We think that Twitter is a promising data source of geotagged of photos, while Flickr has been the most popular data source of geotagged photos in the research community of multimedia so far. Since the characteristic of Twitter is quickness and on-the-spot-ness, the photos on Twitter are different from the photos on Flickr. Flickr has many travel photos, while Twitter has many photos related to everyday life such as food, weather, street scene and some events. Therefore, we think that geo-tweet photos are more useful to understand what happens currently over the world than Flickr geo-photos.

By using this database, we are working on even photo mining [2], [3] from Twitter geo-photo tweets and food photo mining [4] from Twitter photo tweets. In this paper, we introduce our two kinds of Twitter photo mining works on events and foods.

II. Twitter Event Photo Mining

In this section, we introduce a system to mine events from the Twitter stream. To do that, we pay attention to the tweets having both geotags and photos. We call such tweets as "geo-photo tweets". So far some works on event mining which utilize geotagged tweets have been proposed such as Sakaki et al.'s typhoon and earthquake detection [5]. However, they used no images but only textual analysis of tweet texts. On the other hand, in this work, we detect events using visual information as well as textual information. In the experiments, we show some examples of detected events and their photos such as "rainbow", "fireworks" and "Tokyo firefly festival".

We propose a Twitter visual event mining system which consists of event keyword detection, event photo clustering and representative photo selection. The processing steps of the proposed system are as follows:

(1) Detect event keyword candidates which frequently appear in the tweets posted from specific areas in specific days.

(2) Unify and complement detected event keywords

(3) Select geo-tweet photos corresponding to the event keywords by image clustering

(4) Select a representative photo to each event

(5) Show the detected events with their representative photos on the map

Note that the current system assumes the tweet messages written by either English or Japanese language, since keyword extraction needs to be taken into account of the characteristics of target language. However, it is not so difficult to extend the proposed system to other languages.

A. The proposed method

In this subsection, we explain each step of the proposed system briefly.

1) Textual Analysis: Tweet messages are written in sentences or sets of words in general. To detect events easier, at first we extract noun words from each tweet message. To do this, for tweets written in English, we apply the English morphological analyzer which is specialized for tweet messages, TweetNLP, while for tweets written in Japanese language, we apply the Japanese morphological analyzer, MeCab. According the output of the morphological analyzer, we extract only noun words as keywords from each tweet after stop-word removal.

To detect events, we search for bursting keywords by examining change of the daily frequency of each keywords within each unit area. The area which is a location unit to detect events are defined in the grids by one degree latitude and one degree longitude. In case that the daily frequency of the specific keyword within one grid area increases greatly, we consider that an event related to the specific keyword happened within the area in that day.

In the previous step, we limited an event keyword to a single noun keyword. However, since some events are represented by compound keywords, the same event are sometimes detected by several keywords independently. In such case, we unify them into a compound keyword related to the same event according to the following heuristics:

(1) In case that more than half of the tweets related to a specific event keyword overlaps the tweets related to another event keyword, the former keywords are integrated and replaced with the latter keywords.

    - E.g. “rain” and “typhoon” ⇒ “typhoon”

(2) In case that words just after or before the detected event keyword are the same in more than 80% tweets including the keyword, such words are regarded as being part of a compound event keyword.

    - E.g. “Tokyo”, “sky” and “tree” ⇒ “Tokyo Skytree”

B. Visual Analysis

Until the previous step, event keywords and their corresponding tweets have been selected. In this step, we carry out clustering of the photos embedded in the selected event tweets and selecting representative ones from them.

As image features, we use bag-of-features (BoF) with densely-sampled SURF local features and 64-dim RGB color histograms. SURF keypoints are sampled with every 10 pixels in the scale 5, 10 and 15. The size of the codebook for BoF was set as 1000. Both feature vectors are L1-normalized.

For clustering photos, we use the Ward method which is one of agglomerative hierarchical clustering methods. It creates clusters so to minimize the total distance between the center of each cluster and the cluster members. It merges the cluster pairs which bring the minimum total error calculated in the following equation one by one.

We evaluate each of the obtained clusters in terms of visual coherence. We calculate visual coherence score V_C. When V_C is high, the corresponding cluster is likely to strongly related to the event. On the other hand, in case that V_C is lower, the cluster is expected to be a noise one which is less related to the event.

In addition, the cluster having the maximum value of V_C is regarded as a representative cluster, and the photo the visual feature vector of which is the closest to the cluster center is selected as a representative photo for the corresponding event. Please see the detail in [3].

C. Experimental results

In the experiment, we prepared two large-scale geo-photo tweet databases which are extracted from our geo-photo tweet database: The first one is a Japan geo-photo tweet database which consists of about three million geo-photo tweets posted from Japan from February 10th, 2011 to September 30th, 2012. The second one is a United States geo-photo tweet database, which consists of seventeen million geo-photo tweets posted from United States from January 1st, 2012 to December 31st, 2012.

As results of event keyword detection for the given datasets, we obtained 258 and 1676 event keywords from the Japan and US dataset related to natural phenomena such as “rainbow” and “typhoon” and local events related to “fireworks” and“festival”, and the accuracy of the event keyword detected finally were 86.4% and 88.9%, respectively.

Some detected events are shown on the map with their representative photos in Figure 1 and Figure 2. These map are interactive maps based on Google Maps API, and a user can see any event photos by clicking markers on the maps. Figure 3 shows detected beautiful sunset photos after after clicking the representative photo shown in the pop-up maker. This map-based interactive event viewing system is available via Web at http://mm.cs.uec.ac.jp/event/ for the US dataset and at http://mm.cs.uec.ac.jp/event_jp/ for the Japan dataset.

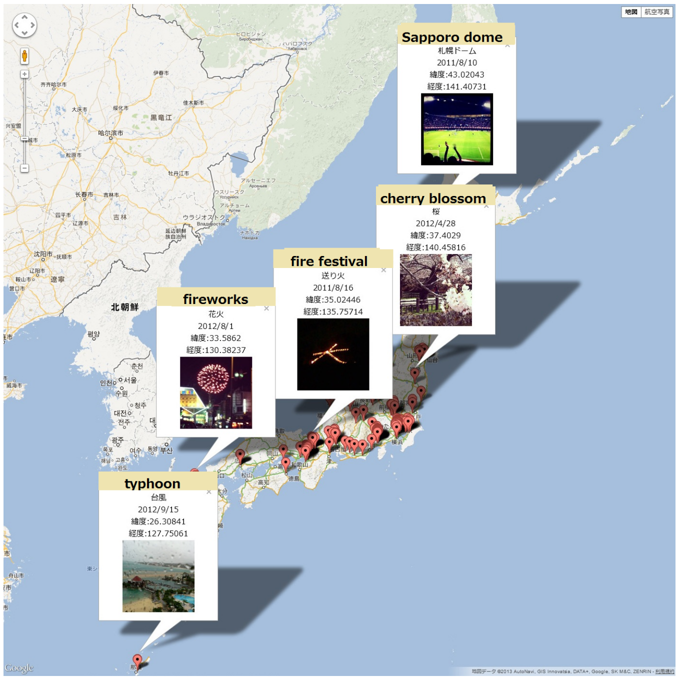


Figure 1. Some detected events in Japan are shown on the map with their representative photos.

With the proposed method, we implemented a real-time system. Because our method requires relatively light computation, the proposed method can be used as a method of the real-time event detection with multi-thread processing. While in the previous experiments we detected event photos from our tweet photo database, we show a result we detected by the real-time event photo detection system from the Twitter stream in Figure 4 which corresponds to "fireworks" in Tokyo area at July 26th, 2014.

III. Twitter Food Photo Mining

In this section, we focus on food photos embedded in Tweets as the second case study on a large-scale Twitter photo analysis. Food is one of frequent topics in Tweets with photos. In fact, we can see many food photos in lunch and dinner time in the Twitter stream.

In this section, by combining keyword-based search and food image recognition, we mine food photos from the Twitter stream. To collect food photos from Twitter, we monitor the Twitter stream to find the tweets containing both food-related keywords and photos, and apply a "foodness" classifier and 100-class food classifiers to them to verify whether they shows foods or not after downloading the corresponding photos. We used the state-of-the-art Fisher Vector

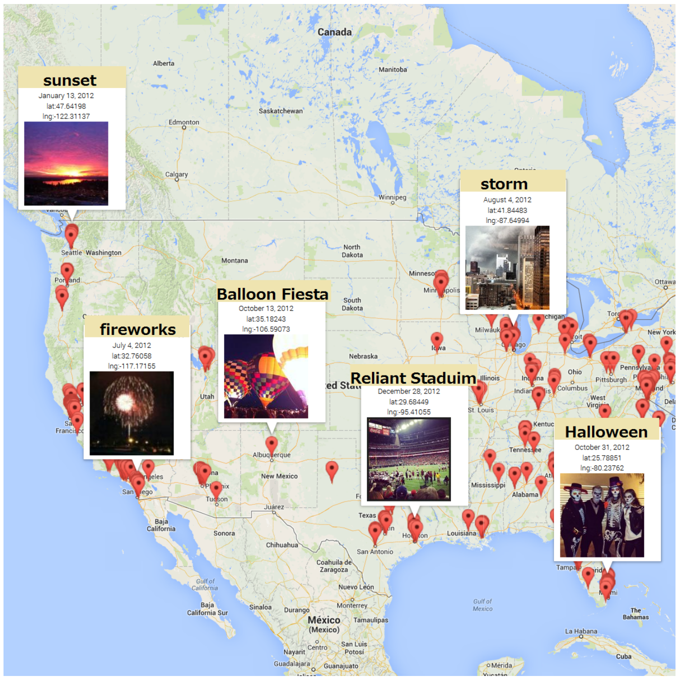
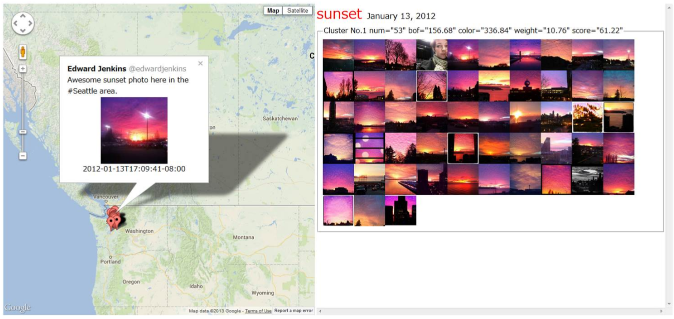


Figure 2. Some detected events in US are shown on the map with their representative photos.



Edward Jenkins @edwardjenkins
Awesome sunset photo here in the #Seattle area.
2012-01-13T17:09:41-08:00

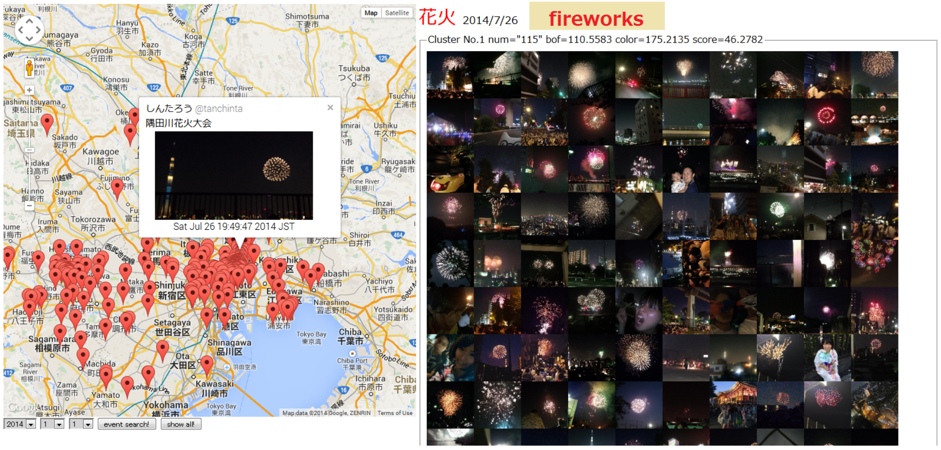
sunset January 13, 2012
Cluster No.1 num="53" bof="156.68" color="336.84" weight="10.76" score="61.22"

Figure 3. "Sunset" photos after clicking the representative photo shown in the pop-up maker.

coding with HoG and Color patches for food classifiers which is slightly modified with the rapid food recognition method for mobile environments proposed in [6], and trained them with the UEC-FOOD100 dataset [7][1] which consists of 100 kinds of foods commonly eaten in Japan. Since we employ the improved method of the real-time mobile recognition, it takes only 0.024 seconds to recognize one image and it achieved about 83% classification rate within the top five candidates.

In the experiments, we report the results of our food photo mining on 100 kinds of foods in the UEC-FOOD100 dataset from the photo tweet log data we have collected for two years and four months. As results, we detected about

[1] http://foodcam.mobi/dataset/


花火 2014/7/26 fireworks

Cluster No.1 num="115" bof=110.5583 color=175.2135 score=46.2782

しんたろう @tanchinta
隅田川花火大会
Sat Jul 26 19:49:47 2014 JST

Figure 4. The event keyword, "fireworks", detected by real-time event photo detection.

470,000 food photos from Twitter with about 99% accuracy. With this data, we have made spatio-temporal analysis on food photos. In addition, we have implemented the real-time food photo detection system from the Twitter stream.

A. Overview

In this section, we describe an overview of the proposed method to mine food photos from the stored Twitter logs as well as the Twitter stream. We employ the following three-step processing.

(1) We perform keyword-based search with the names of target foods over a set of photo Tweets.

(2) We apply a newly-proposed "foodness" classifier to the tweet photos selected by the keyword-based search for classifying then into either of "food" or "non-food" photo.

(3) We apply individual food classifiers corresponding to the food names. In the experiments, we prepared multi-class discriminative classifiers trained by SVM with the UEC-FOOD100 dataset in the one-vs-rest manner.

The food classifiers employed in the third step is a slight modification of the method for mobile food recognition proposed in [6], while the foodness classifier is newly proposed for removing non-food photos.

B. Detail of the Proposed Method

1) Keyword-based Photo Tweet Selection: In the first step, we select photo tweets by keyword-based search with the names of the target foods. We search tweet message texts for the words of the target food names.

As the target foods, we used 100 kinds of foods in the UEC-FOOD100 dataset in the experiments. Because the UEC-FOOD100 dataset includes common foods in Japan such as ramen noodle, curry, and sushi, we searched only photo tweets the message texts of which are written in Japanese language. We can easily select them by checking the language attribute of each tweet obtained from the Twitter Streaming API.

2) Foodness Classifier: We construct a "Foodness" Classifier (FC) for discriminating food images from non-food images. FC evaluates if the given image is a food photo or not. We use FC to remove noise images from the images gathered from the tweet photos selected by the food names.

We construct a FC from the existing multi-class food image dataset. Firstly, we train linear SVMs [8] in the one-vs-rest strategy for each category of the existing multi-class food image dataset. As image features, we adopt HOG patches [9] and color patches in the same way as [6]. Although HOG patches are similar local features to SIFT [10], HoG can be extracted much faster than SIFT. Both descriptors are coded by Improved Fisher Vector (IFV) [11], and they are integrated in the late fusion manner. We perform multi-class image classification in the cross-validation using the trained liner SVMs, and we build a confusion matrix according to the classification results. In the experiments, we used 64 GMMs for IFV coding and two-level spatial pyramid [12], which is much improved from mobile food recognition [6] in terms of the feature dimension.

Secondly, we make some category groups based on confusion matrix of multi-class classification results. This is inspired by Bergamo et al.'s work [13]. They grouped a large number of categories into superordinate groups the member categories of which are confusing to each other recursively. In the same way, we perform confusion-matrix-based clustering for all the food categories. We intend to obtain superordinate categories such as meat, sandwiches, noodle and salad automatically. As results, in the experiments, we obtained 13 food groups as shown in Table I.

Table I
13 FOOD GROUPS AND THEIR MEMBER FOODS FOR THE "FOODNESS" CLASSIFIER.

food group	food categories
noodles	udon nooles, dipping noodles, ramen
yellow color	omlet, potage, steamed egg hotchpotch
soup	miso soup, pork miso soup, Japaneses tofu and vegetable chowder
fried	takoyaki, Japaneses-style pancake, fried noodle
deep fried	croquette, sirloin cutlet, fried chicken
salad	green salad, macaroni salad, macaroni salad
bread	sandwiches, raisin bread, roll bread
seafood	sashimi, sashimi bowl, sushi
rice	rice, pilaf, fried rice
fish	grilled salmon, grilled pacific saury, dried fish
boiled and seasoned	seasoned beef with potatoes simmered ganmodoki seasoned beef with potatoes
sauteed	sauteed vegetables, go-ya chanpuru, kinpira-style sauteed burdock
sauce	stew, curry, stir-fried shrimp in chili sauce

To build a "foodness" classifier (FC), we train a linear SVM of each of the superordinate categories. The objective of FC is discriminating a food photo from a non-food photo, which is different from the objective of the third step for discriminating a specific food photo from other kinds of food photos. Therefore, abstracted superordinate categories are desirable to be trained, rather than training of all the food categories directly. The output value of FC is the maximum value of SVM output of all the superordinate food groups.

When training SVMs, we used all the images of the categories under the superordinate category as positive samples. For negative samples, we built a negative food image set in advance by gathering images using the Web image search engines with query keywords which are expected to related to noise images such as "street stall", "kitchen", "dinner party" and "restaurant" and excluding food photos by hand. All the images are represented by Fisher Vector of HoG patches and color patches. SVMs are trained in the late fusion manner with uniform weights.

In the second step, we apply FC for the selected tweet photos and remove non-food photos from the food photo candidates.

3) Specific Food Classifiers: In this step, we classify a given photo into one of the prepared food classes.

First, we extract HOG patches and Color patches in a dense grid sampling manner in the same way as the previous step. Then, we apply PCA to all the extracted local features, and encode them into Improved Fisher Vectors. The method to extract features is the same as the previous step including the parameter settings. Next, we evaluate linear classifiers in the one-vs-rest way by calculating dot-product FVs. Finally we output the top-N categories in terms of the descending order of evaluation scores of all the linear classifiers.

In the experiments, we regarded the given tweet photo as a photo of the corresponding food if the food names contained in the tweet messages are ranked in the top five categories by evaluation of 100-kind food classifiers. This is because the top-5 classification rate exceeds 83%, while the top-1 rate is still around 60%.

C. 100 Food Categories and Their Classifiers

In this subsection, we describe the detail of the 100-class food dataset we used in the experiments. In the experiments, as target foods, we used 100 foods in the UEC-FOOD100 [7][2], because we employ supervised food photo classification which requires training data to select the target foods in the third step. It contains more than 100 images per category, and all the food item in which are marked with bounding boxes. For training and evaluation, we used only the regions inside the given bounding boxes. The total number of food images in the dataset is 12,905. UEC-FOOD100 dataset consists of common foods in Japan. Then, we restricted tweets from which we mine food photo tweets to only the tweets with Japanese messages.

In [6], they implemented a mobile food recognition system using the same dataset. Although basically we followed their method for individual food classification in the third step, we extended the parameter setting to improved accuracy. To say it concretely, we doubled the size of GMM for FV encoding from 32 to 64, and added two-level spatial pyramid. As a result, the total feature dimension are raised from 3072 to 35840, which boosted the classification performance evaluated by 5-fold cross-validation as shown in Figure 5. Regarding the processing time, it takes only 0.024 seconds per image to recognize on Core i7-3770K 3.50GHz with multi-threaded implementation optimized for a quad-core CPU.

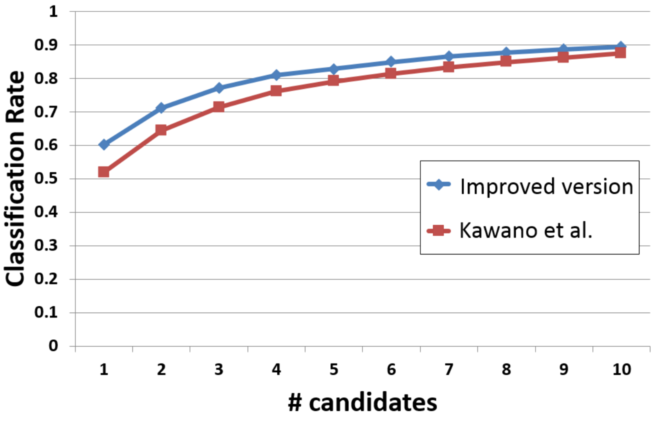


Figure 5. Comparison on the top-k classification rates with the UEC-FOOD100 dataset evaluated by 5-fold cross validation between [6] and this paper.

[2] http://foodcam.mobi/dataset/

D. Experimental Results

In this subsection, we describe the experimental results on twitter food photo mining. We have been collecting photo tweet logs by monitoring the Twitter stream by using Twitter Streaming API. Here, we used 122,328,337 photo tweets with Japanese messages out of 988,884,946 photo tweets over all the world collected from May 2011 to August 2013 for two years and four months.

From these photo tweets, we selected 1,730,441 photo tweets the messages of which include any of the name words of the 100 target foods in the first step of the proposed processing flow. Then, in the second step, we applied a "foodness" classifier (FC) to all the selected images. After applying FC, we applied 100-class one-vs-rest individual food classifiers. As a result, we obtained 470,335 photos which are judged as food photos corresponding to any of the 100 target food categories by our proposed processing pipeline.

For the 470,335 selected photos as food photos, we evaluate the number of selected photos for each category. Table II shows the ranking of 100 food categories in terms of the number of mined tweet food photos. The number of "Ramen noodle" and "curry" photos are the most and the second most with the large margin to the third or less ranked food categories, respectively. In fact, "ramen" and "curry" are regarded as the most popular foods in Japan. "Sushi", "dipping noodle (called as Tsukemen in Japanese)" and "omelet with fried rice (called as Ome-rice in Japanese)" are also popular foods in Japan. The results of twitter food image mining reflects food preference of Japanese people. In addition, we found that many of "ome-rice" had drawings or letters drawn with ketchup, as shown in Figure 6. These are estimated to be made at home, while most of "ramen" and "sushi" photos are taken at food restaurants, because there are many ramen noodle and sushi restaurant in Japan. Although "hamburger" and "beef bowl" are also popular in Japan as fast food served at fast-food restaurants such as McDonald and Yoshino-ya, they are ranked at more than twentieth. This is because the foods provided by nation-wide fast-food chain restaurants such as McDonaldo are the same everywhere in the same chain restaurants, and they are not worth posting their photos to Twitter. On the other hand, since there are no monopolistic restaurant chains on ramen noodle and curry in Japan, the foods served at every ramen or curry restaurants have originality and are different from each other.

Next, we evaluated the precision rate of the selected food photos in the each steps regarding the top five foods and two sub-categories of "ramen noodle" and "curry". Table III shows the results in case of four types of the combinations of the three kinds of the selection methods, (1) only keywords, (1)+(2) keywords and foodness classifier (FC), (1)+(3) keywords and individual food classifier(IFC), and (1)+(2)+(3) keywords, FC and IFC. Note that this evaluation was done for the 300 random-sampled photos for each cell in the table.

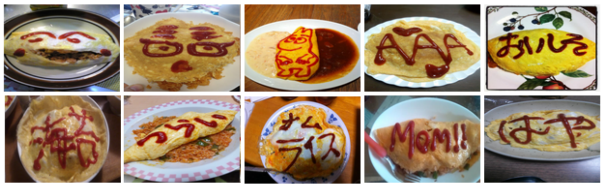


Figure 6. Examples of "omelet" photos. Most of them have drawings drawn by ketchup.

Regarding (1), the precision of two sub-categories, "beef ramen noodle" and "cutlet curry", are relatively higher, 94.3% and 92.7%, than "ramen noodle" and "curry". From this results, we can assume that when tweeting detailed food names with photos, the photos probably represent the corresponding foods. After applying both FC and IFC, (1)+(2)+(3), the precision of all the seven foods achieved the best compared to the cases of applying only single methods or only keyword-based search, (1), (1)+(2) and (1)+(3). Except for "sushi", the precision reached 99.0%, which means nearly perfect. This shows the effectiveness of introducing both FC and IFC after keyword-based search. Exceptionally, "sushi" is a difficult food to recognize by object recognition methods, because the appearances of "sushi" varies greatly depending on the kinds of the ingredients on the pieces of hand-rolled rice.

Finally, we describe simple spatio-temporal analysis on Twitter food photos. Figure 7 shows the prevailing-food map where the red marks, the yellow marks and the blue marks represent the areas where "ramen noodle", "curry" and "okonomiyaki" are most popular in terms of the number of food photo tweets, respectively. The left map, the center map, and the right map show the prevailing-food map on all the term (May 2011-Aug. 2013), Dec. 2012 (in winter), and Aug. 2013 (in summer), respectively. From the leftmost map, "ramen noodle" is the most popular over Japan on average through a year. However, compared between the center map and the rightmost map, popularity of "curry" increases in summer, while "ramen noodle" becomes the most popular in winter. Exceptionally, in the area around Hiroshima where the blue marks appear, "okonomiyaki" is always the prevailing food in Twitter food photos, this is partly because Hiroshima has a very popular regional food, "Hiroshima-yaki", which is a variant of "okonomiyaki".

As another temporal analysis on the mined food photos, we examined the time when each food are eaten the most frequently over a day. As results, the most frequent time when "ramen noodle" and "curry" are eaten is between 12pm and 2pm, while the most frequent time of "sushi" and "okonomiyaki" is between 7pm and 9pm. This reflects the difference of the characteristic of the foods. As shown in

Table II
THE RANKING OF 100 FOODS IN TERMS OF THE NUMBER OF MINED TWEET FOOD PHOTOS.

1	ramen noodle	80021
2	curry	59264
3	sushi	25898
4	dipping noodle	22158
5	omelet with fried rice	17520
6	pizza	16921
7	jiaozi	16014
8	Japanese-style pancake	15234
9	steamed rice	14264
10	sashimi	13927
11	hambarg steak	11583
12	beef stake	9503
13	takoyaki	9004
14	fried rice	8383
15	fried noodle	7905
16	oden	7453
17	toast	6350
18	cutlet curry	6339
19	tempura	5905
20	rice ball	5462
21	gratin	5223
22	croquette	4837
23	stew	4797
24	sashimi bowl	4730
25	chicken-'n'-egg on rice	4513
26	tempura bowl	4464
27	beef bowl	4285
28	spicy chili-flavored tofu	4081
29	yakitori	3829
30	hamburger	3662
31	chilled noodle	3473
32	sukiyaki	3408
33	miso soup	3295
34	fish-shaped pancake with bean jam	3281
35	pork cutlet on rice	3188
36	omelet with grilled minced meat	2592
37	bibimbap	2368
38	spaghetti	2171
39	lightly roasted fish	2162
40	seasoned beef with potatoes	2129
41	natto	2094
42	spaghetti with meat source	1994
43	steamed egg hotchpotch	1843
44	egg sunny-side up	1635
45	croissant	1579
46	udon noodle	1500
47	simmered pork	1443
48	mixed sushi	1371
49	pork miso soup	1229
50	ginger-fried pork	1158
51	potato salad	1150
52	egg omelet	1146
53	eels on rice	1071
54	egg roll	1058
55	sweet and sour pork	1049
56	fried shrimp	1049
57	sauteed vegetables	1040
58	shrimp with chill source	1003
59	cabbage roll	965
60	mixed rice	901
61	pilaf	891
62	soba noodle	880
63	potage	816
64	hot dog	795
65	chicken rice	736
66	wiener sausage	577
67	dried fish	563
68	steamed meat dumpling	561
69	french fries	561
70	beef ramen noodle	555
71	sandwiches	551
72	cold tofu	517
73	boiled chicken and vegetables	352
74	sirloin cutlet	331
75	nanbanzuke	323
76	fried chicken	314
77	stir-fried beef and peppers	312
78	roll bread	288
79	roast chicken	263
80	macaroni salad	239
81	boiled fish	228
82	kinpira-style sauteed burdock	225
83	tempura udon	213
84	raisins bread	205
85	goya chanpuru	198
86	green salad	145
87	chinese soup	141
88	Japanese tofu and vegetable chowder	137
89	salmon meuniere	96
90	grilled pacific saury	84
91	chip butty	76
92	fried fish	72
93	begitable tempura	71
94	tensin noodle	69
95	ganmodoki	34
96	grilled salmon	25
97	sauteed spinach	12
98	teriyaki grilled fish	3
99	grilled eggplant	2
100	pizza toast	0

Table III
THE NUMBER OF SELECTED PHOTOS AND THEIR PRECISION(%) WITH FOUR DIFFERENT COMBINATIONS.

food category	(1)	(1)+(2)	(1)+(3)	(1)+(2)+(3)
ramen noodle	275652 (72.0%)	200173 (92.7%)	84189 (95.0%)	80021 (99.7%)
beef ramen noodle	861 (94.3%)	811 (99.0%)	558 (99.7%)	555 (99.7%)
curry	224685 (75.0%)	163047 (95.0%)	62824 (97.0%)	59264 (99.3%)
cutlet curry	10443 (92.7%)	9073 (98.0%)	6544 (98.7%)	6339 (99.3%)
sushi	86509 (69.0%)	43536 (86.0%)	48019 (72.3%)	25898 (92.7%)
dipping noodle	33165 (88.7%)	24896 (96.3%)	28846 (93.7%)	22158 (99.0%)
omelet with fried rice	34125 (90.0%)	28887 (96.3%)	18370 (98.0%)	17520 (99.0%)

this subsection, the data we collected through Twitter food photo mining is useful for food habit analysis.

E. Real-time Food Photo Detection System

We implemented a real-time Twitter food photo detection system which continuously detects 100 kinds of food photos from the Twitter stream. We detect the photo tweets including any of 100 kinds of Japanese food names about ten times per minute at most. Because the time to download a thumbnail image is about 2 or 3 seconds and the processing time for food recognition for each image is less than 0.1 seconds, we can process all the pipeline on a single machine in the real-time way. The very fast food recognition method which was originally designed for a mobile application [6] made it possible. With this system, currently we always keep running the real-time food photo detection system and collecting new food photos. For example, we are collecting about 20,000 "ramen noodle" and 15,000 "curry" photos per month.

As shown in Figure 8, the detected food photos are shown on the map if they have geotags or geo-related words such as place names in their Tweet messages, and on the right side the photos are displayed as the results by online k-means clustering. This system can be accessible via http://mm.cs.uec.ac.jp/tw/.

REFERENCES

[1] K. Yanai, "World seer: A realtime geo-tweet photo mapping system," in *Proc. of ACM International Conference on Multimedia Retrieval*, 2012.

[2] Y. Nakaji and K. Yanai, "Visualization of real world events with geotagged tweet photos," in *Proc. of IEEE ICME Workshop on Social Media Computing (SMC)*, 2012.

[3] T. Kaneko and K. Yanai, "Visual event mining from geo-tweet photos," in *Proc. of IEEE ICME Workshop on Social Multimedia Research (SMMR)*, 2013.

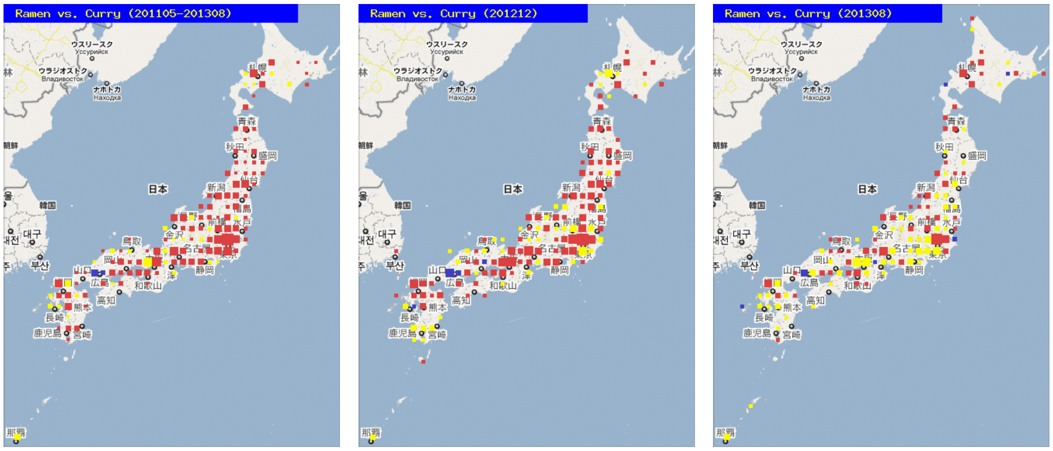


Figure 7. The prevailing-food map of Japan. The red marks, the yellow marks and the blue marks represents the areas where “ramen noodle”, “curry” and “okonomiyaki” are most popular in terms of food photo tweets, respectively. The left map, the center map, and the right map shows the prevailing-food map on all the term (May 2011-Aug. 2013), Dec. 2012 (in winter), and Aug. 2013 (in summer), respectively.

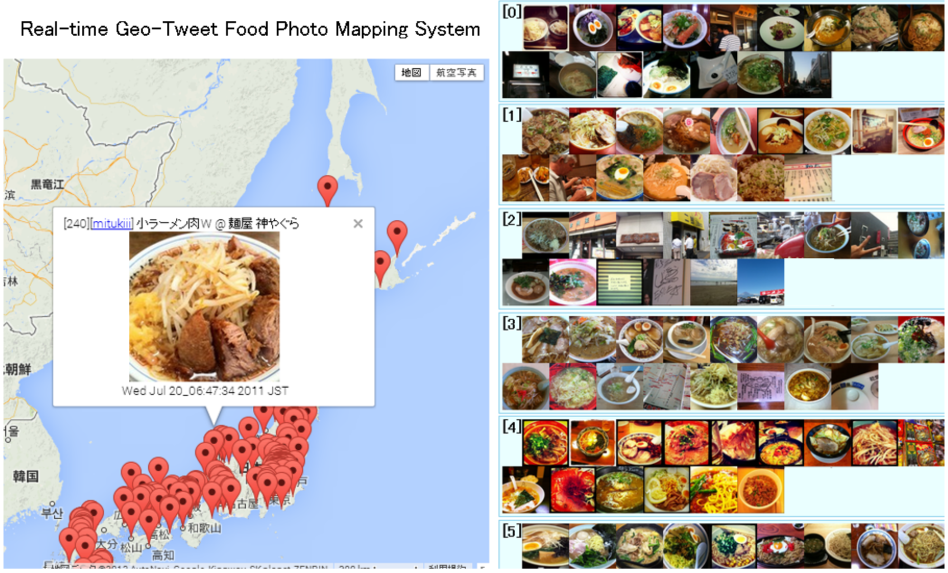

Real-time Geo-Tweet Food Photo Mapping System
地図
航空写真
[240][mitukiii] 小ラーメン肉W @ 麺屋 神やぐら
Wed Jul 20_06:47:34 2011 JST
黒竜江
朝鮮
韓国
釜山
和歌山
高知
松山
大分
[0]
[1]
[2]
[3]
[4]
[5]

Figure 8. Detected food photos are displayed on the map when geo-information are available.

[4] K. Yanai and Y. Kawano, “Real-time food image mining and analysis from the twitter stream,” in *Proc. of Pacific-Rim Conference on Multimedia (PCM)*, 2014.

[5] 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*, 2010, pp. 851–860.

[6] Y. Kawano and K. Yanai, “Foodcam: A real-time food recognition system on a smartphone,” *Multimedia Tools and Applications*.

[7] Y. Matsuda and K. Yanai, “Multiple-food recognition considering co-occurrence employing manifold ranking,” in *Proc. of IAPR International Conference on Pattern Recognition*, 2012.

[8] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, “LIBLINEAR: A library for large linear classification,” *The Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.

[9] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2005.

[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] F. Perronnin, J. Sánchez, and T. Mensink, “Improving the fisher kernel for large-scale image classification,” in *Proc. of European Conference on Computer Vision*, 2010.

[12] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2006.

[13] A. Bergamo and L. Torresani, “Meta-class features for large-scale object categorization on a budget,” in *Proc. of IEEE Computer Vision and Pattern Recognition*, 2012.