Visual Event Mining from Geo-tweet Photos

Takamu Kaneko and Keiji Yanai
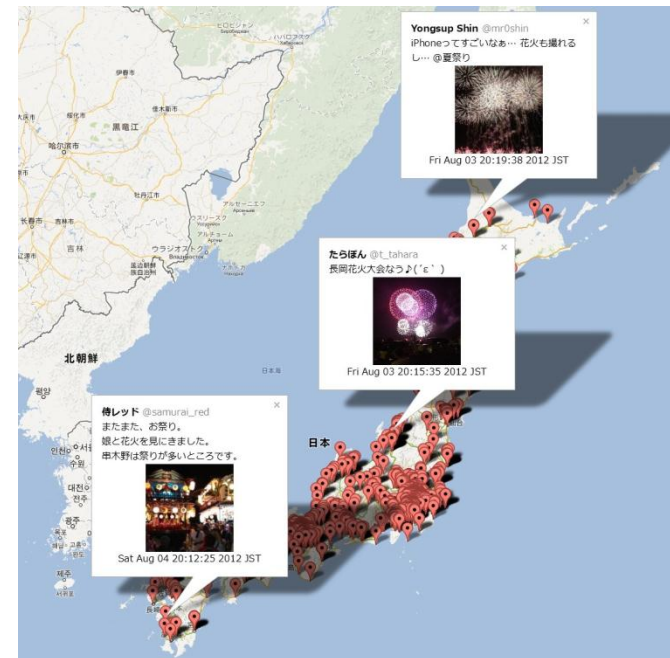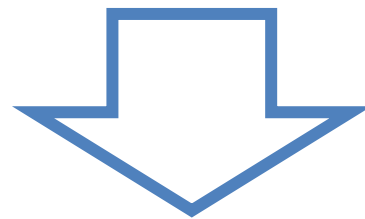
Department of Informatics,

The University of Electo-Communications,

Tokyo, Japan

The University of Electro-Communications

Background

- Spread of smart phones

  -Geotagged photos

- Spread of Twitter

  -Real-time posting

Increase of "geo-tweet photos"

Objective

- Detect events from Twitter stream

  -Weather, natural events

  -Festivals, sport games

- Understand events visually

  -Select representative photos

  -Map in a map

Mapping events with the photo

Related Work: Twitter Event Mining

- Use only textual information

- Sakaki et al. [WWW 2010]
  -Regard Twitter users as social sensors
  -Estimate the location of natural events

- Lee et al. [ACM SIGSPATIAL WS 2010]
  -Divide target area into small sub-regions
  -Monitor the number of tweets
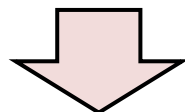
Related Work: Twitter Photo Mining

- Use geo-tweet photos

- Yanai [ACM ICMR 2012]

  -Monitor the Twtiiter stream in real-time

  -Map geo-tweet photos on the map

- Nakaji et al. [ICME WS 2012]

  -Mine representative photos related keywords

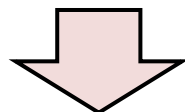 -Compare difference on places and time

Overview

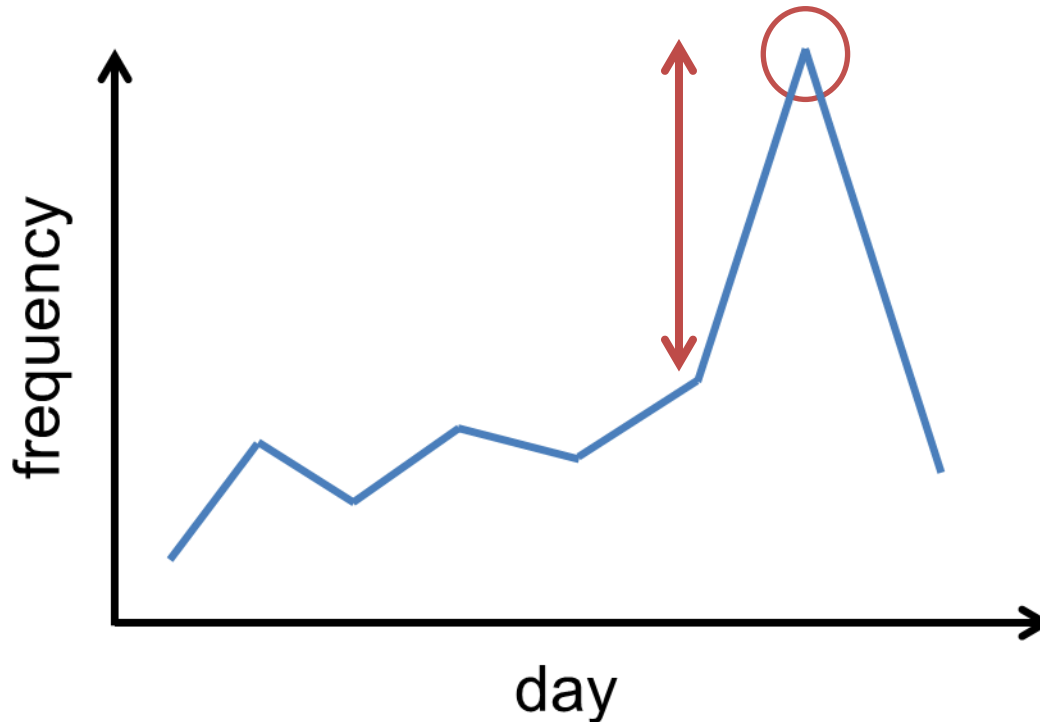| 1. Event keyword detection |
| :-: |

⬇
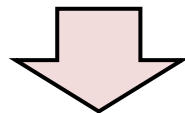
2. Unification and complement

⬇

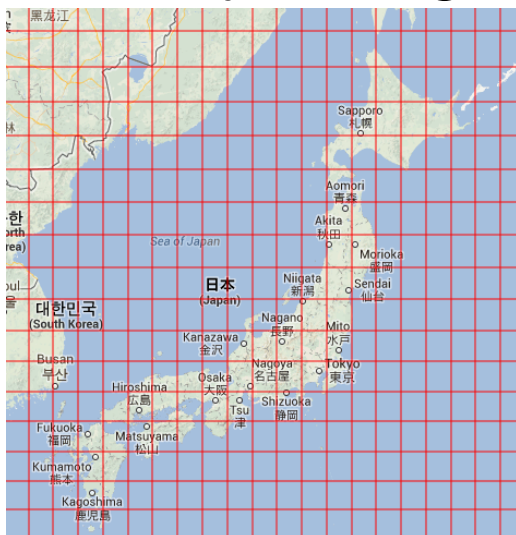| 3. Event photo clustering |
| :-: |

⬇

4. Mapping event with photos

Event Keyword Detection

- Examine change of daily frequency

Event Keyword Detection

- ## Part-of-Speech tagger

  -MeCab(http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html)

  -TweetNLP(http://www.ark.cs.cmu.edu/TweetNLP/)

- ## Divide target area into sub-regions

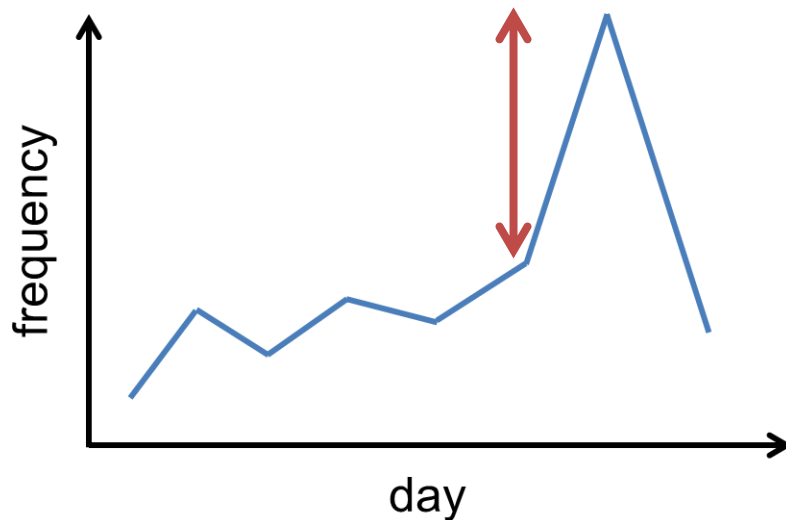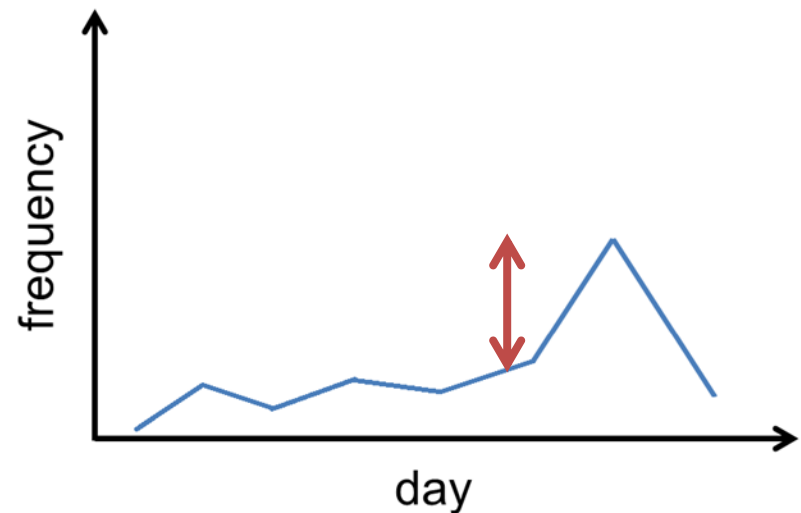 -Grids by 1 degree latitude and longitude

US

Japan

Event Keyword Detection

- ## Difference of user population
  -Difference of increase scale

high population area   low population area

Event Keyword Detection

- Evaluate area weight

$$W_{area} = \frac{\#users_{max} + sd}{\#users_{area} + sd}$$

- Evaluate keyword score

 -over 50(JPN), 200(USA)

$$S_{kw,date,area} = (N_{kw,date,area} - N_{kw,date-1,area})W_{area}$$

Unification and Complement

- Unification keyword

  -more than half of the same tweets

  -Use the highest score keyword

  > "shuttle", "Endeavor" ⟹ "shuttle"

- Complement keyword

  -more than 80% of the same word

  -before and after the keyword

  > "Festival" ⟹ "Music Festival"

Event Photo Clustering

- Image features

  -Bag-of-Features with SURF

  -Color histograms

- Ward method

  -a hierarchical clustering method

  -threshold is 300(both)

$$E(C) = \sum_{x \in C}((x_{BoF} - \overline{x_{BoF}})^2 w_{BoF} + (x_{RGB} - \overline{x_{RGB}})^2 w_{RGB})$$

Event Photo Selection

- ## Select a representative cluster

-evaluate cluster score on representativeness

$$V_C = \frac{\#photos_C{}^2}{E(C)} W_{area}$$

- ## Select a representative photo

-from the maximum score cluster

- ## Eliminate lower score cluster

-less than 5(JPN) , 20(USA)

Experiments

- Dataset 1

  -in Japan

  -from Feb 10th, 2011 to Sep 30th, 2012

  -about 3 million geo-tweet photos

- Dataset 2

  -in US

  -from Jan 1st, 2012 to Dec 31st, 2012

  -about 17 million geo-tweet photos

Results of Keyword Detection

Keyword	Date
snow	11/02/2011
earthquake	11/03/2011
fireworks	06/08/2011
typhoon	21/09/2011
Mt. Fuji	24/09/2011
Apple	06/10/2011
eclipse	10/12/2011
illumination	10/12/2011
Christmas	24/12/2011
New years eve	31/12/2011
sunrise	01/01/2012
firefly	06/05/2012

Keyword	Date
snow	09/01/2012
sunset	13/01/2012
Grammy	12/02/2012
Valentines	14/02/2012
SXSW	09/03/2012
Easter	08/04/2012
shuttle	17/04/2012
WWDC	10/06/2012
hurricane	26/08/2012
rainbow	05/09/2012
49ers	18/10/2012
NYE	31/12/2012

Japan   USA

Results of Unification & complement

Japan

Before	After unification	After complement
Sky, Tree	Sky	Sky Tree
fireworks, Yodo-river	fireworks	fireworks
eclipse, solar	eclipse	solar eclipse
marathon	marathon	Tokyo marathon
Dome	Dome	Sapporo Dome

US

Before	After unification	After complement
Rangers, Ballpark	Rangers	Rangers Ballpark
West, WWDC, Apple	WWDC	WWDC
Golden, Gate, Bridge	Golden	Golden Gate Bridge
Square, Times	Times	Times Square
Carnival, Electric, Daisy	Electric	Electric Daisy Carnival

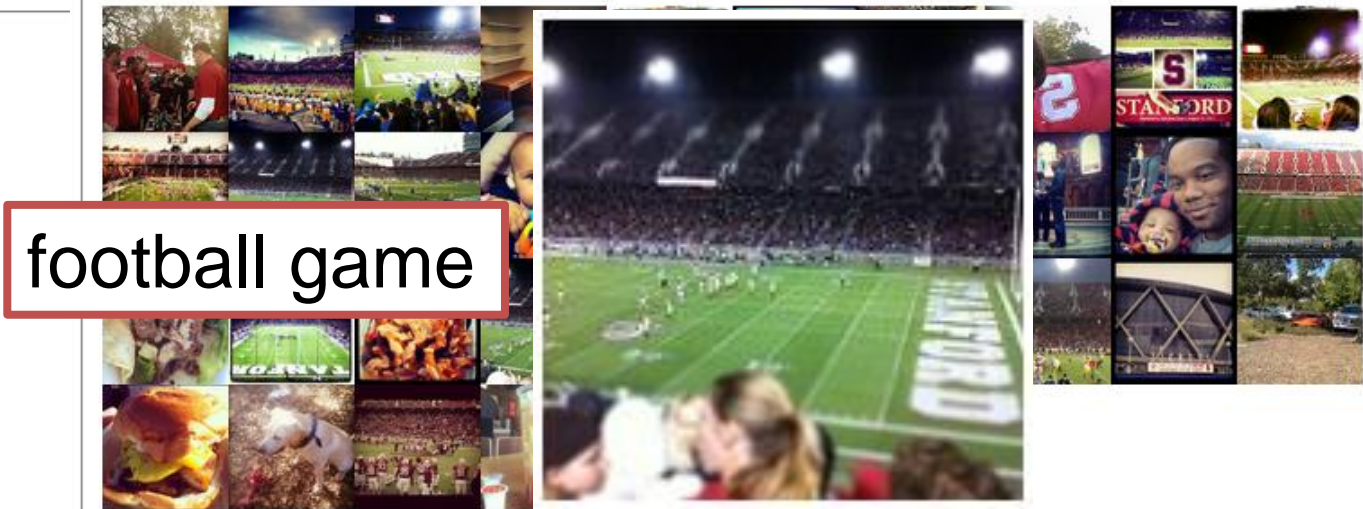# "fireworks" photo clusters

"cherry blossoms" photo clusters

"Stanford Stadium" photo clusters

Cluster No.1 num="33" bof="142.56" color="167.23" weight="10.95" score="38.5"

soccer game
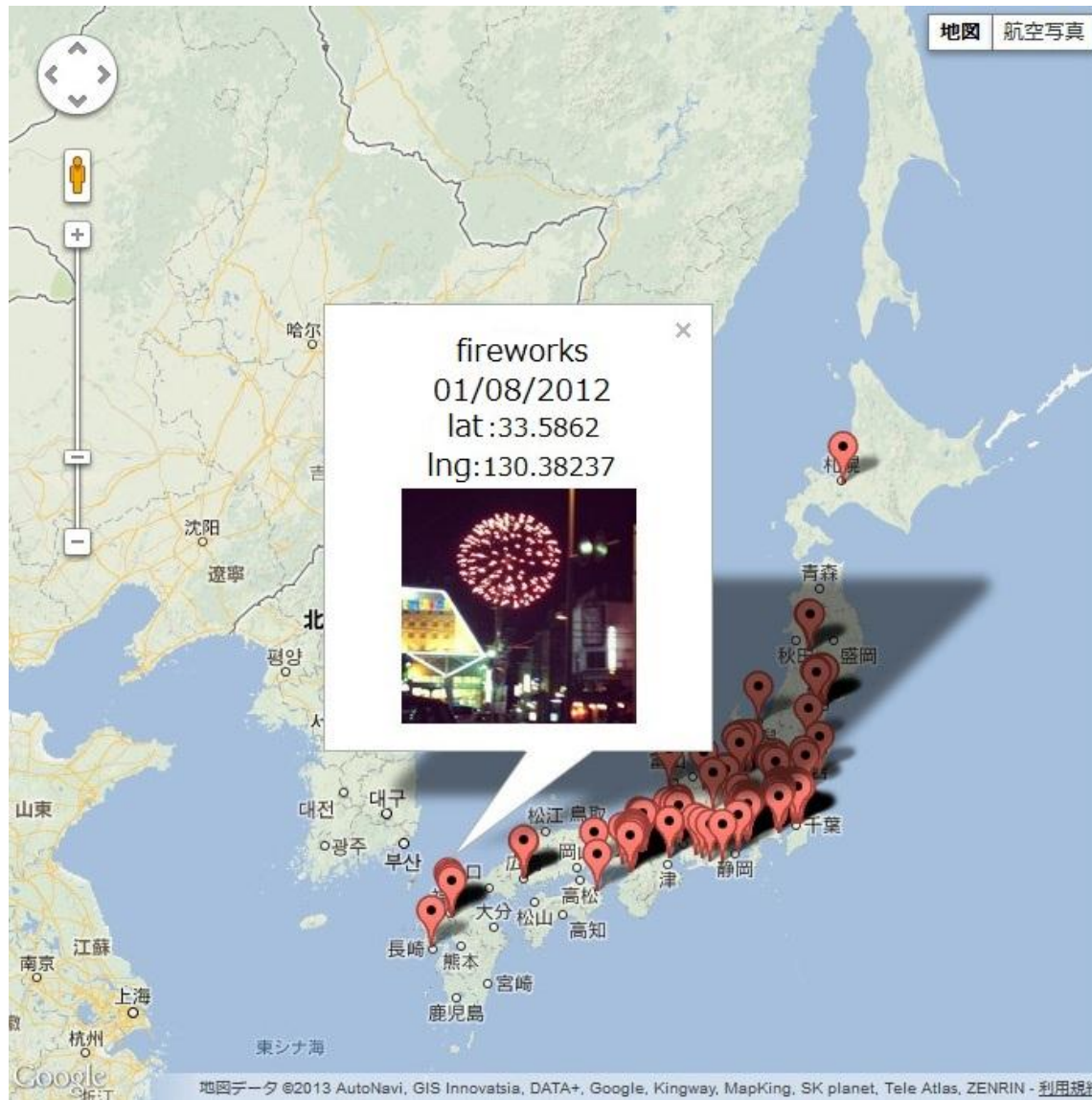
109.82" color="119.27" weight="10.95" score="65.45"
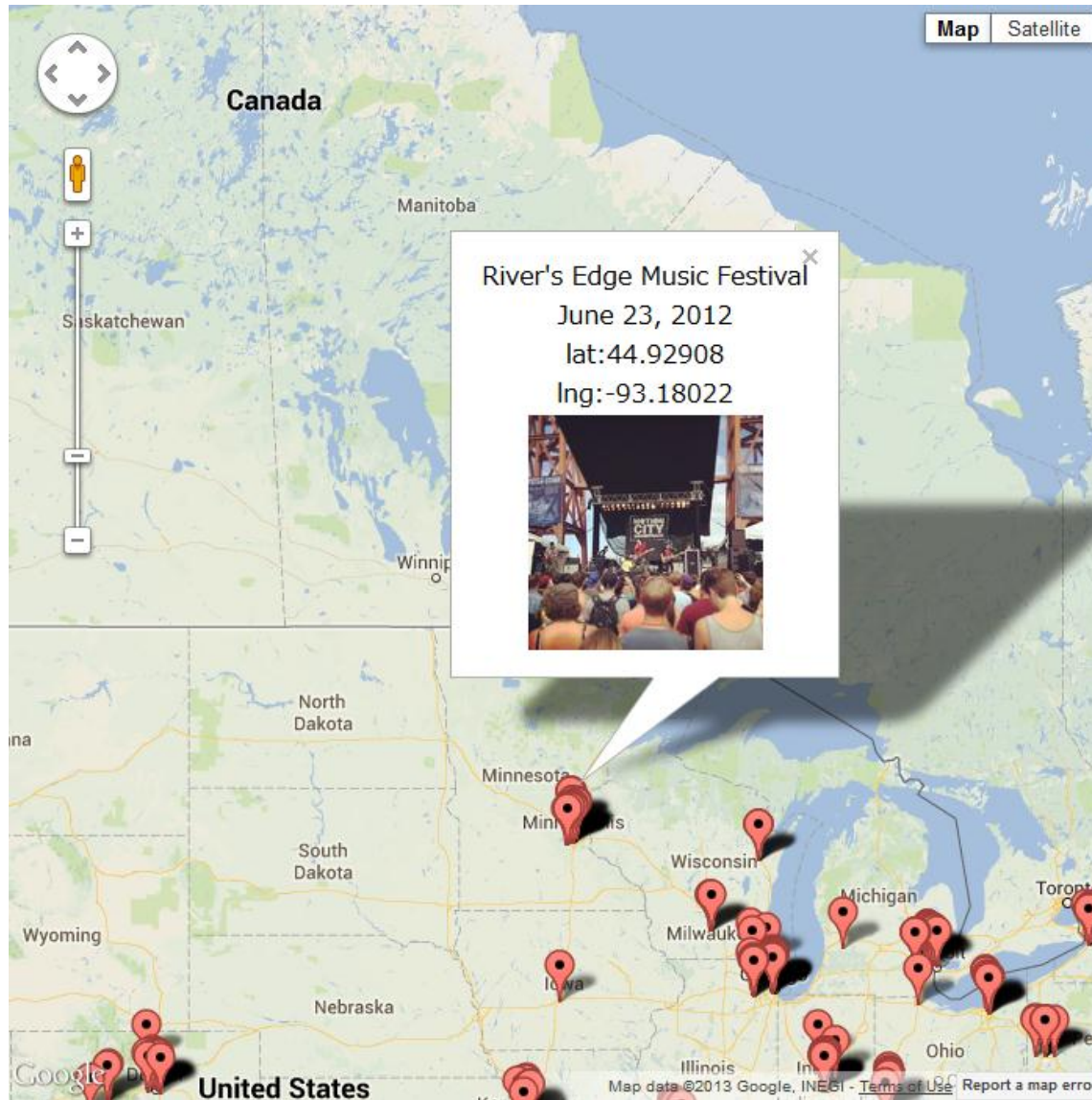
football game

Mapping Results

- Map event in a map

  -Calculate coordinates of event

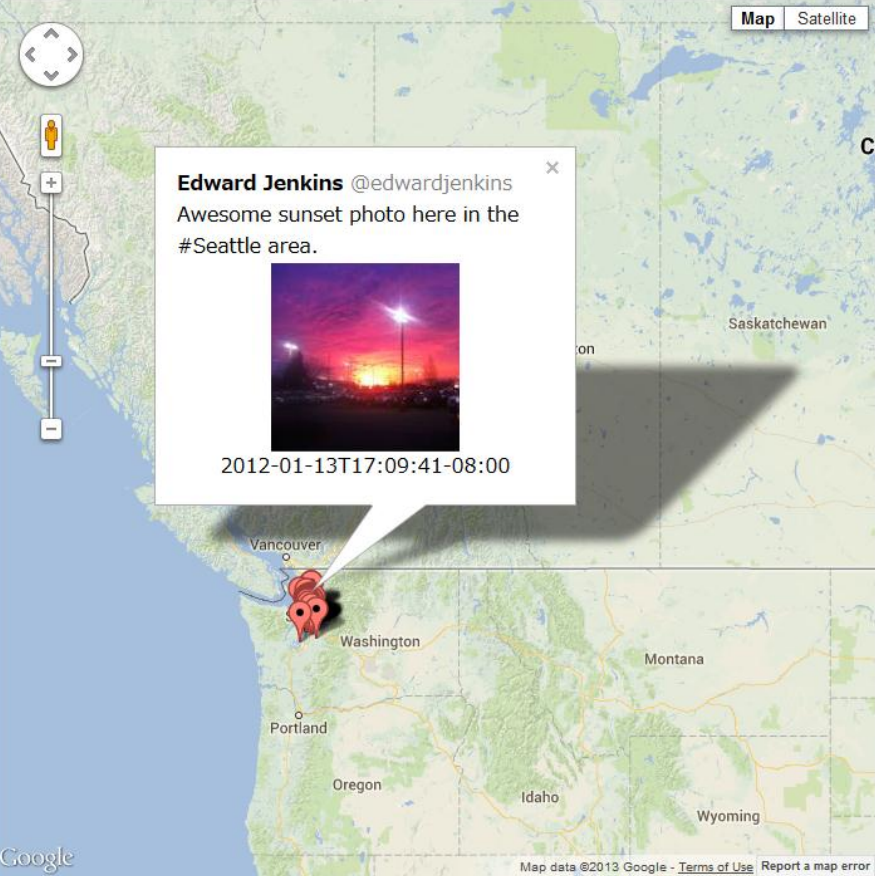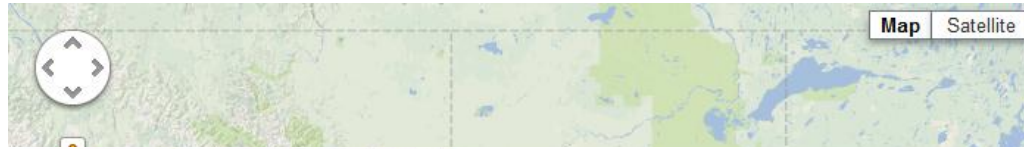 -Correspond information and the photo

- Summary of results

	Japan	USA
# events	258	1676
accuracy	65.5%	72.5%

Rescheduled Expo data set

"River's Edge Music Festival"
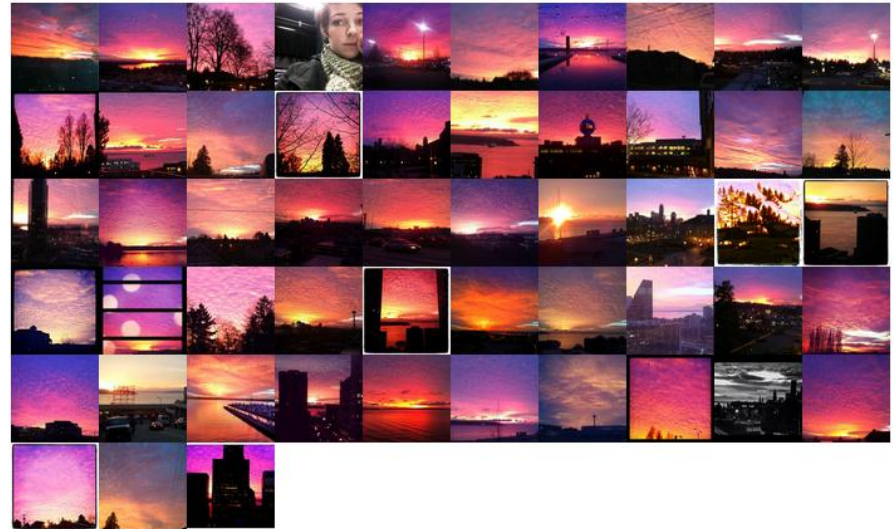
"sunset"

Conclusion

- Event detection with geo-tweet photos

    -Detection of keywords

    -Selection of representative photos

- Detection results

    -258 events in Japan dataset

    -1676 events in US dataset

- Accuracy of representative photos

    -65.5% in Japan dataset

    -72.5% in US dataset

Future Works

- Flexible detection
  - Variable size of grid
  - Variable term for detection

- Real-time detection

- Improvement of methods for selection representative photos

More results…

- ICME demo session on July 18th