TWITTER VISUAL EVENT MINING SYSTEM

The University of Electro-Communications, Tokyo, Japan   Takamu Kaneko, Hiroyoshi Harada and Keiji Yanai

Background

- Spread of smart phones
  - Geotagged photos
- Spread of Twitter
  - Real-time posting

Increase of "geo-tweet photos"

Objective

- Event detection from Twitter
  - Weather, natural events
  - Festivals, sport games
- Understanding event visually
  - Selection representative photos
  - Mapping in a map

Overview
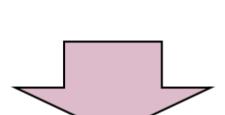
1. Event keyword detection
2. Unification and complement
3. Collection with keywords
4. Extraction image features
5. Event photo clustering
6. Event photo selection
7. Mapping with photos
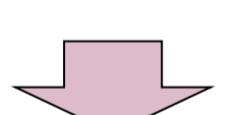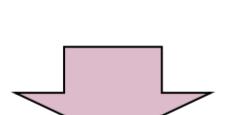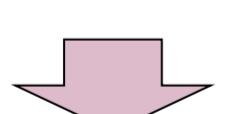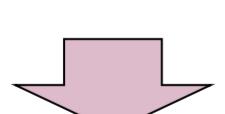
Method

Event Keyword Detection

- Part-of-Speech tagging
- Dividing area
  - Grids by 1 degree lat & lng
- Evaluation area weight

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

- Evaluation keyword score

$$S_{k,d,area} = (N_{k,d,area} - N_{k,d-1,area})W_{area}$$

Keyword "fireworks"

| area weight 10.952 | 3/7/2012 12 users | 4/7/2012 78 users |

In this case,
Score(fireworks, 4/7/2012, this area)
=(78-12) × 10.952
=722.832

Unification & Complement

- Keyword unification
  - over 50% same tweets

  "shuttle", "Endeavour" ➡ "shuttle"

- Keyword complement
  - over 50% same letter

  "Festival" ➡ "Music Festival"

Event Photo Clustering

- Image features
  - Bag-of-Features with SURF
  - RGB color histograms
- A hierarchical clustering algorithm
  - Ward method

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

Event Photo Selection

- Evaluation cluster score

$$V_C = \frac{n_C^2}{E(C)} W_{i,j}$$

- Selection a representative photo in the maximum score cluster
- Elimination lower score cluster

Experiment & Results

Datasets

- in USA
- 1/1/2012 – 31/12/2012
- about 17,000,000 tweets

Keyword Detection

Keyword	Date	Area (lat, lng)
snow	9/1/2012	38~39, -78~-77
sunset	13/1/2012	47~48,-123~-122
Grammy	12/2/2012	34~35,-119~-118
Valentines	14/2/2012	37~38,-123~-122
SXSW	9/3/2012	30~31, -98~-97
Easter	8/4/2012	34~35, -85~-84
shuttle	17/4/2012	38~39, -78~-77
WWDC	10/6/2012	37~38,-123~-122
America	4/7/2012	33~34,-119~-118
hurricane	26/8/2012	25~26, -81~-80
rainbow	5/9/2012	37~38, -123~122
49ers	18/10/2012	37~38, -123~-122
turkey	22/11/2012	25~26, -81~-80
NYE	31/12/2012	41~42, -88~-87

Unification

Before	After
Rangers, Ballpark	Rangers
eclipse, Eclipse	eclipse
West, WWDC, Apple	WWDC
July, Fourth	July
Island, Treasure	Treasure
Bowl, Rose	Rose
House, White	House
Field, LP	Field
Square, Times	Times

Complement

Before	After
Golden	Golden Gate Bridge
Fenway	Fenway Park
Fox	Fox Theatre
Ocean	Ocean Beach
Electric	Electric Daisy Carnival
Festival	Eat Real Festival
Casino	Hotel & Casino
Field	LP Field
Times	Times Square

Photo Clustering

Cluster No.1 num="40" bof="154.64" color="65.45" weight="10.48" score="76.18"

Cluster No.2 num="31" bof="115.88" color="130.91" weight="10.48" score="40.81"

Example of "fireworks"

Cluster No.1 num="48" bof="181.75" color="212.1" weight="10.95" score="64.07"

Cluster No.2 num="24" bof="109.96" color="149.59" weight="10.95" score="24.3"
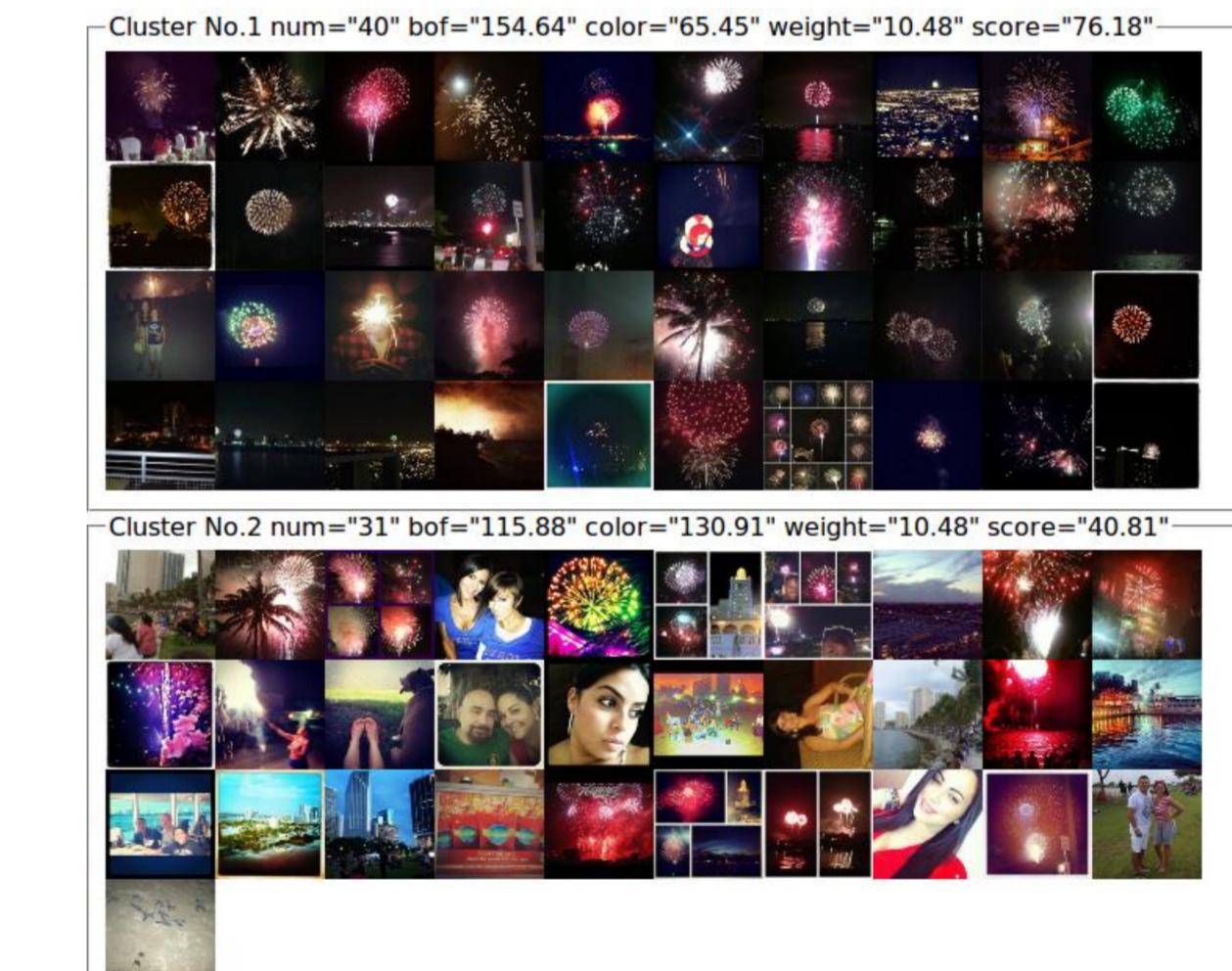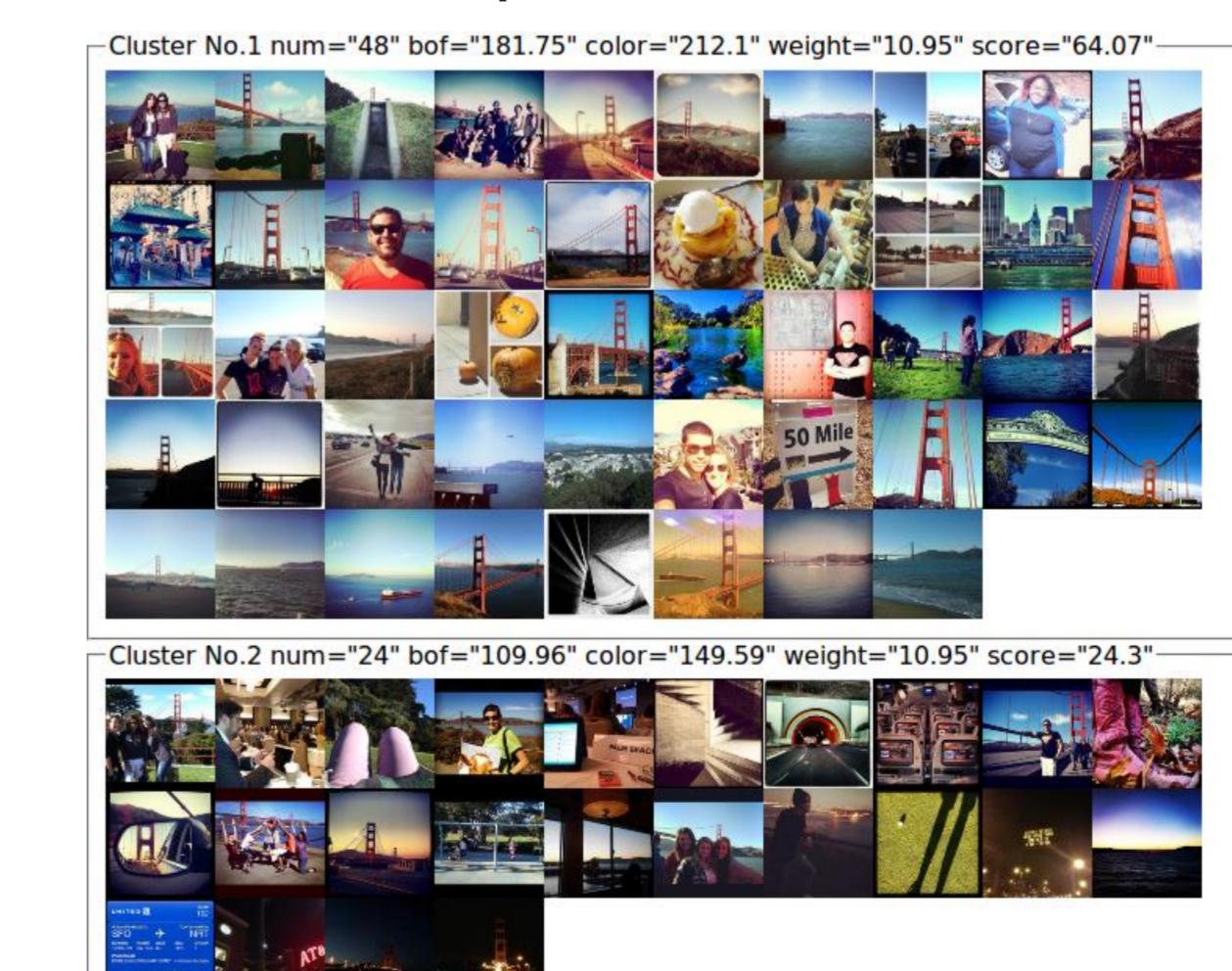
Example of "Golden Gate"

Mapping Results

- 1676 events detected

Example of mapped events

Conclusion

- Event Detection with geo-tweet photos
- Understanding event visually

Future Work

- Variable grid size & Variable term for detection
- Real-time detection