

Disasters seen through Flickr cameras

Romain Fontugne

The Graduate University for Advanced Studies

Kenjiro Cho

IJ Research Lab

Youngjoon Won

IJ Research Lab

Kensuke Fukuda

National Institute of Informatics / PRESTO JST

ABSTRACT

Collecting aftermath information after a wide-area disaster is a crucial task in the disaster response that requires important human resources. We propose to assist reconnaissance teams by extracting useful data sent by the users of social networks that experienced the disaster. In particular we consider the photo sharing website Flickr as a source of information that allows one to evaluate the disaster aftermath. We propose a methodology to detect major event occurrences from the behavior of Flickr users and describe the nature of these events from the tags they post on the Flickr website. Our experiments using two study cases, namely, the Tohoku earthquake and tsunami and the Tuscaloosa tornado, reveals the value of the data published by Flickr users and highlight the value of social networks in disaster response.

1. INTRODUCTION

Reconnaissance missions are crucial tasks enabling efficient responses to severe disasters. Indeed, shortly after a disaster outbreak the rapid identification of the most damaged locations is essential to conduct relief efforts to the appropriate sites. Reconnaissance teams are groups of rescuers exploring the damaged areas and reporting the disaster severity at different places. The collected data are geo-localized photos, videos or text that report the buildings' damage and allow the development of emergency plans.

Recent technologies assist reconnaissance teams in acquiring, sharing and analyzing more data in a systematic manner. For example high-resolution optical satellites provide a coarse view of damage areas and help in guiding responders to hardest hit areas [11]. Although these pictures provide less details than ground-

truth observations, the wide surface covered by satellite pictures allows all affected sites to be rapidly observed. In order to collect and organize the data recorded by reconnaissance teams, Kayen et al. proposed a tool based on Google Earth [4]. This tool allows a greater collaboration among responders and helps to plan further investigation by centralizing data on a map. Similarly, during the response to the 2010 Haiti Earthquake three major US agencies (the U.S. Agency for International Development, the U.S. State Department and the U.S. armed forces) relied on social media to coordinate their knowledge and action [12]. Thereby, using wikis and a collaborative workspaces, the contribution of each responder is available to others and all information is rigorously organized. These social medias help decision makers to systematically acquire knowledge contributions from several sources.

While social medias are valuable tools helping responders to collaborate, public social medias (e.g. Facebook or Twitter) are also important communication mediums for victims. Beyond the technological aspect of social medias, the human interactions flowing through these networks after disaster outbreaks also contains valuable information for responders. Although a social network user may not intend to contribute to a reconnaissance mission, the information he/she sends through the network can convey important information regarding damage caused in his/her neighborhood. Therefore social networks contain numerous chunks of beneficial information from numerous locations that are quickly available to responders. However, the human interactions on social networks during disaster outbreaks have been rarely addressed in the research literature. Sakaki et al. proposed a system that analyzes Twitter users' posts to detect earthquake or typhoon occurrences [9]. Their work emphasizes the value of the information transiting through social medias and their system successfully detects and estimates in real time the position of earthquakes from only short text messages and the geo-localization of users.

Similar to the work of Sakaki and al. [9] we propose to analyze the user posts on the Flickr website after

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM SWID 2011, December 6, 2011, Tokyo, Japan.

Copyright 2011 ACM 978-1-4503-1044-4/11/0012 ...\$10.00.

disaster outbreaks. Flickr is a photo and video hosting service that allows enthusiastic photographers to share their creations and comment on those from other users. Photo uploads on Flickr have been increasing steadily over the past few years and the website was hosting 5 billions pictures taken from all over the world in September 2010 [10]. Each photo is posted along with information provided by the camera or the user. For example the camera provides the date, time and settings used to take the photo whereas the user inputs tags and a description related to the photo.

In this work we consider Flickr users as mobile sensors capturing snapshots of the nearby events. Also we assume that these sensors are particularly active during and after disaster outbreaks because of the uniqueness and impressiveness of these events. While photographers capture and share disaster photos for different reasons, these snapshots have also a potential for contributing in disaster responses.

The goal of this work is to assess the above assertions. We aim at highlighting the particular activity of Flickr users during disaster outbreaks and reveal the benefits of Flickr for investigating damages caused by disasters. This paper presents a methodology to identify the bursts of photos corresponding to major event occurrences, and describe the events using the tags provided by users. The method also precisely localizes the impacted sites and plots the disaster related pictures in Google Maps. The comparison with past pictures from Google Street View allows us to evaluate the damage caused by a disaster.

We stress the potential of Flickr users to help responders for reconnaissance missions, and we present two study cases, the 2011 Tohoku earthquake and 2011 U.S. tornadoes outbreak.

Our three main contributions are, first, in Section 3 we propose a systematic approach to identify and extract from Flickr relevant information related to disasters. The proposed methodology detects bursts of photos that exhibit the sudden interest of Flickr users in major events and takes advantage of an information retrieval technique to describe the identified events. Second, in Section 4 we evaluate the proposed method by analyzing two disasters that occurred in 2011, the Tohoku earthquake and tsunami and the Tuscaloosa tornado. Using these two examples we demonstrate that social networks provide valuable data to responders although these data were not initially intended to contribute to the response mission. Finally, we make our dataset publicly available which is a snapshot of 10 years of Flickr photos (Section 2).

2. DATA SOURCE: FLICKR

Flickr provides an API¹ to access to the Flickr photo

¹<http://www.flickr.com/services/api/>

database. Each request to the database however will return at most 4000 results. Due to this limitation we crawled the Flickr database to collect the information related to all the photos, videos, and illustrations that have geo-tagging data and that have been taken between January 2000 and July 2011. Our collection provides metadata for more than 130 million geo-tagged pictures taken all over the world. We make this data set available to everyone for non-commercial use only. The key characteristics of this dataset are;

Geotagging The geographical location associated to a picture that is either automatically assigned by the camera built in GPS or manually by the user on the Flickr website.

Tags A list of keywords provided by the user that describes the content of the picture.

Taken date The date and time of the snapshot. In the case of disasters, this date allows us to precisely locate in time the appearance of the disaster.

Upload date The date and time of the picture upload on Flickr servers. For our purpose, this date indicates the delay between the disaster outbreak and the availability of the picture.

Views The number of people that have seen the picture on the Flickr website. This measure conveys the interest of people in the captured event.

3. METHODOLOGY

This article proposes a methodology to identify and investigate disaster through the Flickr database. The proposed method monitors a specific location (e.g. a town or prefecture) and detects major events by looking at the sudden increase in numerous Flickr sensors activity. The event is reported in the form of a collection of photos plotted on a map and a set of tags describing the nature of the event. The proposed method consists of four main steps (Figure 1);

Localization filtering Selecting all photos taken at a specific location for a given time period.

Burst detection Finding bursty sequences of photos corresponding to the major events and evaluating the time interval of the events occurrence.

Burst characterization Identifying meaningful tags that describe the major events previously identified.

Photo extraction and visualization Retrieving all photos corresponding to the identified events and precisely mapping them on a map.

These four steps are detailed in the following sections.

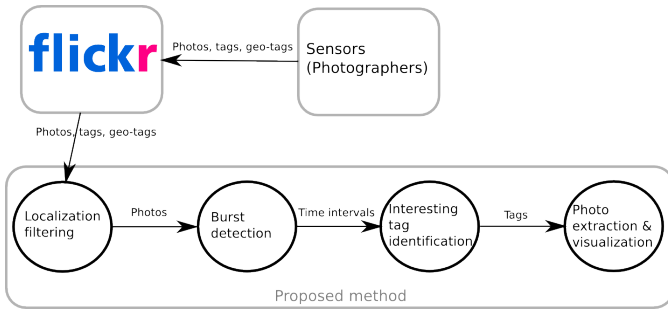


Figure 1: Overview of the proposed method

3.1 Localization filtering

The first step of the proposed methodology is to extract all photos corresponding to the location given by the user. Since our dataset contains only geo-tagged photos we simply select pictures based on their geo-tags (i.e. latitude and longitude).

3.2 Burst detection

The main assumption in the present work is that Flickr sensors suddenly capture numerous photos during disaster outbreaks. Thereby, by looking at the time evolution of sensing data, an abrupt and intensive activity of sensors reveals the occurrence of a disaster or other unusual event. We investigate the activity of sensors over time and look for a bursty sequence of photos taken by several sensors in a short period of time. A straightforward approach is to analyze the total number of photos over time, however, our experiments have shown that this information is biased by a few sensors taking numerous photos of minor events. For example, during a weeding, a photographer takes a hundred photos and posts them on Flickr to share them with family. This event is certainly important for the family but not necessarily for the other photographers nearby. Thereby we propose to monitor the number of sensors that start to take a sequence of photos. Hereafter we refer to all sensors that took at least one photo during a day as *activated sensors*. A sensor is activated only once per day and the timestamp of its activation corresponds to its first daily picture. By looking at the sudden changes in the number of activated sensors we aim to identify major events captured by sensors.

We emphasize that the time evolution of sensor activation is crucial to identify disaster outbreaks. Whereas numerous sensors might be activated during a particular time period such as national holidays or weekends, disaster outbreaks are characterized by a sudden activation of numerous sensors during a short time frame. Therefore we aim at analyzing the number of activated sensors and their inter-activation time. The method we employ is based on the previous work of Kleinberg that identifies important topics in document streams (e.g. e-

mails or news articles) signaled by a burst of activity [5, 6].

Kleinberg proposes to model bursty activities in document streams using an infinite-state automaton [5]. Each state of the automaton stands for a certain rate of document emission distributed according to an exponential density function. Thereby, two states q_i and q_j represent two distinct burst intensities and $i > j$ means that q_i stands for a greater intensity than q_j .

The algorithm proposed by Kleinberg to find the optimal state sequence corresponding to a given sequence of document inter-arrival times is an adaptation of the Viterbi algorithm used for hidden Markov models [8] to the following model and cost function. $C_j(t)$ is the minimum cost of a state sequence for the input x_1, x_2, \dots, x_t that must end with state q_j , and then iteratively builds up the values of $C_j(t)$ using the recurrence relation

$$C_j(t) = -\ln(\alpha_j e^{-\alpha_j x_t}) + \min_l (C_l(t-1) + \tau(l, j)) \quad (1)$$

with initial conditions $C_0(0) = 0$ and $C_j(0) = \infty$ for $j > 0$, and, with α_j the rate of message arrival for the state j and τ a cost function (see [5] for more details).

We implemented this algorithm to identify bursty sequence of activated sensors. Given a sequence of sensor inter-activation times the algorithm provides a sequence of time intervals and the corresponding burst intensity. We consider greater bursts to be the major events captured by the sensors and investigate them using an information retrieval technique.

3.3 Burst description

This section presents a method to describe the major events detected with the algorithm of the previous section. Because the main descriptors of photo content are the set of tags provided by Flickr users, we exhibit the distinctive nature of identified events by uncovering the prominent tags that appear especially during the reported bursts.

Using a time interval t reported by the previous algorithm we classify tags provided by the Flickr users into two categories, those appearing during t and those appearing outside of t . Notice that a tag can belong to both categories. However, tags that significantly describe the singularities of the reported event are those occurring abundantly in t and rarely outside of t . In order to find these significant tags we take advantage of the notion of *interestingness* introduced by Dubinko et al. [2].

Similarly to the *tf-idf* weighting scheme from information retrieval [7], the interestingness as presented by Dubinko et al. [2] aims at exhibiting the most significant object that occurs in a subset even though it is not necessarily the most frequent one. To this end the measure takes into account the number of occurrences of a tag during a particular time interval and during the

total studied period of time.

For our particular case we denote $\gamma(u, t)$ the number of sensors posting the tag u during the period of time t and $\gamma(u)$ the number of sensors posting the tag u for the complete analyzed set of pictures, and, the *interestingness* measure Int is defined as:

$$Int(u, t) = \gamma(u, t) / (C + \gamma(u)) \quad (2)$$

High values of $Int(u, t)$ means that the tag u appears frequently and particularly during the period of time t . Thus, u is a relevant description of the photos taken during the period of time t . Notice that C is a constant allowing to omit the most infrequent tags that occurs accidentally in t , thus, that are not good descriptions of the reported event. This constant is set to 100 for all our experiments.

For each time interval t reported by the burst detector (Section 3.2) we rank all tags that occurs in t depending on their *interestingness* score. From this ranking we extract the most prominent tags by selecting the ones that have statistically higher scores. Namely, we select all tags that are 3 standard deviations higher than the median value and hereafter refer to them as the *descriptor tags*. We employ descriptor tags to describe the distinctive nature of the identified bursts.

3.4 Photo extraction and visualization

The last step of the proposed methodology is to retrieve all the photos corresponding to an identified event. Since the time intervals reported by the burst detector (Section 3.2) contain numerous photos that are not related to the identified event and the event related photos might be taken after the burst of activated sensors (e.g. aftermath photos), we extract photos from inside and outside the identified time interval. The filter we employ consists of extracting all photos that have at least D descriptor tags (Section 3.3). In our experiments we set $D = 2$ as we empirically found that disasters are usually described by at least two tags designating the location and the nature of the disaster.

In order to exhibit the spatial distribution of the disasters the extracted photos are organized on a world map. The location of each photo is determined by the latitude and longitude provided in the geo-tags. Furthermore, our implementation takes advantage of Google Maps and its Street View feature to allow the user to compare the scene captured by Flickr sensors during the disaster with the one captured by the Google cameras in normal conditions.

4. RESULTS

In this section we evaluate our implementation of the above methodology by analyzing two distinct disasters. First, we conduct detailed experiments with data related to a large scale disaster; the 2011 Tohoku

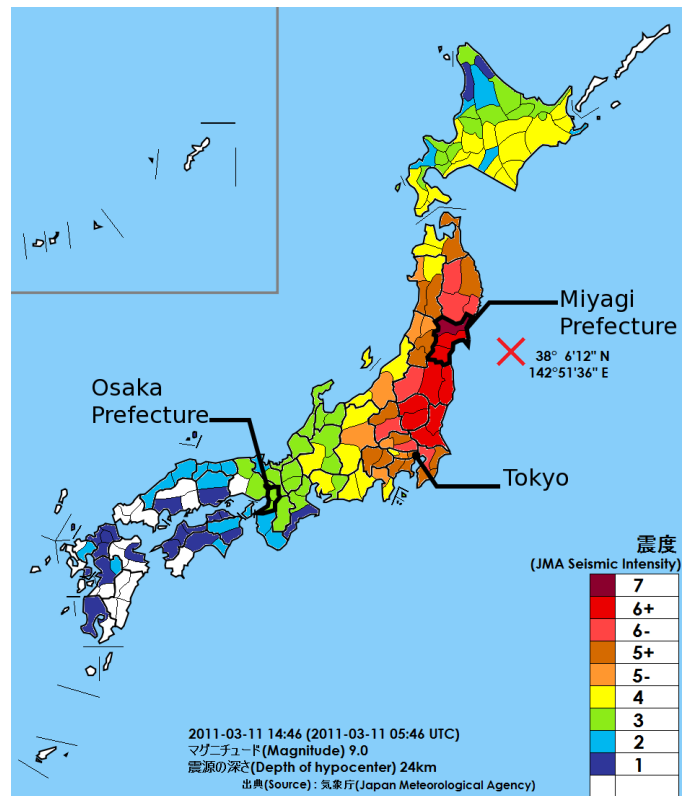


Figure 2: Seismic intensity observations resulting from the Tohoku earthquake (Source: Japan Meteorological Agency)

earthquake and tsunami that struck Japan on March 11, 2011. Second, we investigate the ability of the proposed methodology to localize in space a disaster that caused important damages in a restricted area; the Tuscaloosa tornado which is a destructive tornado of the 2011 US tornado outbreak.

4.1 Tohoku earthquake and tsunami

On March 11 2011 at 14:46 JST a magnitude 9.0 M_w earthquake occurred in the Pacific Ocean at 70km east of Japan coast (Figure 2). This is the fifth most powerful earthquake recorded by world seismometers. It shook the entire northern part of Japan and triggered a devastating tsunami.

In order to show the behavior of Flickr sensors at different locations this section presents the results of our methodology for three sites in Japan; (1) the Miyagi prefecture is the closest coast to the earthquake epicenter and the most damaged by the tsunami. There the earthquake has been measured at 7 on the JMA seismic intensity scale (i.e. the highest score for this scale). (2) Tokyo, the Japanese capital has also experienced the earthquake but damage was minor mainly because the tsunami did not reach Tokyo bay. In Tokyo the earthquake has been measured at 5 on the JMA scale.

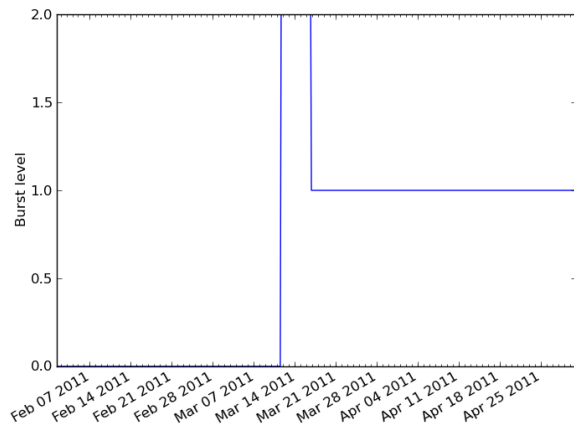


Figure 4: Bursts reported by the burst detector (Section 3.2) for the pictures taken in Miyagi prefecture

(3) Osaka prefecture, a major prefecture located at the south of the Japan’s main island, several people felt the earthquake as a normal one that strikes frequently in Japan. In Osaka the earthquake has been measured around 3 on the JMA scale. In order to exhibit the behavior of Flickr sensors before and after the Tohoku disaster our experiments span on three months, February, March and April 2011.

4.1.1 Miyagi prefecture

Miyagi is the most populated prefecture in the northern part of Japan. It is also the prefecture most devastated by the Tohoku earthquake and tsunami. Figure 3 summarizes the time evolution of the number of photos, the number of activated sensors, and the number of views per day for the 4208 photos taken in the Miyagi prefecture from 2011/02/01 to 2011/03/31. Otherwise stated the dates in these figures refer to the day the photos were taken (only the Fig.3(a) provides also the photos uploading date).

All the monitored quantities in these figures exhibit a sudden increase during the Tohoku disaster outbreak. The high peak in the number of views for the pictures taken during the day of the Tohoku earthquake and tsunami highlights the importance of this disaster that caught the attention of many people around the globe (Figure 3(c)).

Analyzing these photos with the proposed methodology confirms this observation as a significant burst is detected while the previous quantities suddenly increase (Figure 4). The identified burst starts at 14:35:03 JST, highlighting the inaccuracy of the timestamp provided by cameras users (Table 1). The intense activity of the sensors is maintained during the aftermath evaluation

as the reported burst lasts for 6 days at the level 2. By lasting until the end of the analyzed time period at the level 1 the proposed method highlights that sensors are still focused on this disaster one month and half after its outbreak (Figure 4).

Furthermore, the proposed method successfully identified the nature of this burst as the reported descriptor tags identify the two main characteristics of the Tohoku disaster, namely earthquake and tsunami (Table 1).

To evaluate the damage caused by the disaster our implementation of the proposed methodology displays photos with descriptor tags in Google Street View. Therefore one can easily compare the views taken by the Google teams with those of Flickr sensors that captured damaged areas. Figure 5 is an example of 3 views from Street View and 3 photos of Flickr sensors from a city in Miyagi called Ishinomaki. In particular Fig. 5(a) shows the coast area of Ishinomaki and reveals that this part of the city has been entirely demolished. Figure 5(b) represents a house that is located more in the land and exhibits that the tsunami was still strong enough to move cars but houses didn’t collapse. Thereby, we deduce from these observations that the major damage in Ishinomaki was mainly caused by the tsunami rather than the earthquake, and the elevation in the city could protect houses.

Since the tsunami has been particularly devastating in Miyagi the Internet infrastructure was altered and prevented Flickr sensors from rapidly uploading their photos. Figure 6 depicts the delay of the photos that correspond to the Tohoku disaster and that are taken the day of the earthquake (i.e. 2569 photos out of the 4208 analyzed photos). Only 7% of these photos have been uploaded the day of the earthquake, however, 65% were accessible on the website 6 days after the earthquake.

4.1.2 Tokyo

Although the Tohoku earthquake shook Tokyo skyscrapers and altered its train system, the tsunami did not reach the Japanese capital and the damage there was limited. From our dataset 41858 photos were taken in Tokyo from 2011/02/01 to 2011/04/30, the number of photos, activated sensors and number of views over time are depicted in Figure 7. Contrary to Miyagi, these figures do not highlight significant increase in the number of photos or activated sensors the day of the earthquake. The day of the disaster, the number of photos is lower than the two preceding weekends, whereas, the number of activated sensors is similar to those of the two preceding weekends (Figure 7(a) and 7(b)).

However, a significant increase of these two values is observed on Sunday April 10. Our investigation revealed that the major event in Tokyo on April 10 is the cherry blossom viewing which is one of the main

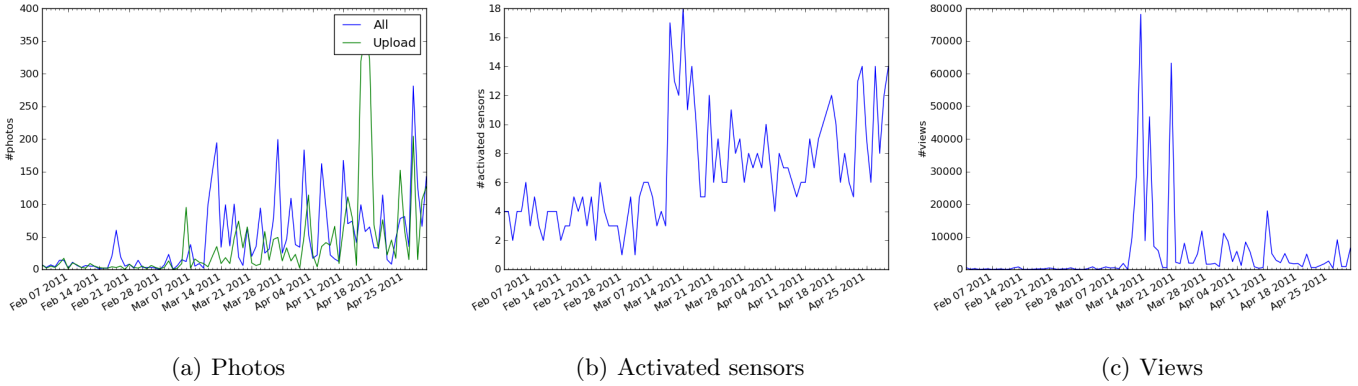


Figure 3: Photos taken in Miyagi prefecture from 2011/02/01 to 2011/03/31

Location	Start	End	Descriptor tags
Miyagi	03/11 14:35:03	03/17 02:17:56	earthquake, tsunami, japan, quake, 2011, sendai, <i>jishin</i> , <i>tsunami</i> , relief
Tokyo	03/11 14:50:00	03/11 16:43:56	earthquake, <i>jishin</i> , quake
Tokyo	04/10 09:43:29	04/10 17:36:12	cherry, sakura, hanami, <i>sakura</i> , blossom, spring, cherryblossom, ...
Osaka	04/09 11:12:21	04/10 20:13:21	<i>sakura</i> , cherry, <i>Osaka</i> , sakura, eyefi, osaka, blossom, park, <i>nihon</i> , ...

Table 1: Prominent bursts and descriptor tags reported by the proposed method. Tags are ranked by their *interestingness* score (Section 3.3) and the most left hand side tag has the highest score. Japanese characters are written in Latin alphabet and italicized; *jishin*, *sakura*, and *nihon* refer to earthquake, cherry blossom, and Japan in Japanese, respectively.

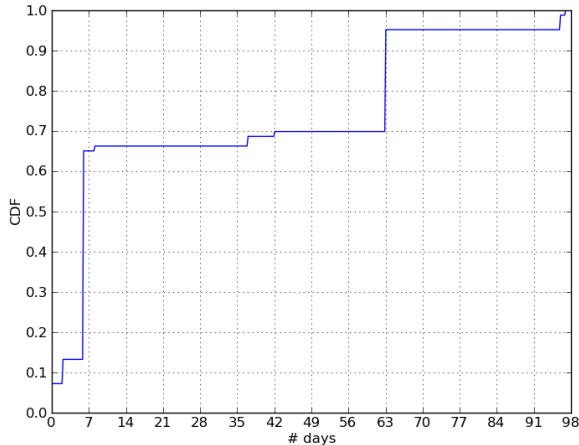


Figure 6: CDF of the delay in day between the shooting of the photo and the upload on Flickr. These are the only photos taken in Miyagi at the day of the earthquake and that have at least two descriptor tags (i.e. 2569 photos).

tourist attractions in Japan. The Japan Meteorological Agency also reported that cherry trees reached full bloom around April 6 in Tokyo [3] and Japanese had the chance to celebrate this tradition on the following weekend (i.e. April 09 and 10).

The proposed method reports two prominent bursts corresponding to the Tohoku earthquake and the cherry blossom view (Table 1). Although the number of photos and the number of activated sensors per day do not highlight singularities the day of the disaster (Figure 7), numerous sensors have been suddenly activated at the same time when the earthquake occurred. This particular distribution in time of the sensor activation is detected by the burst detector of Section 3.2 at 14:50:00 (i.e. 6 minutes after the earthquake occurred). The time length of the burst is significantly shorter than the one reported in Miyagi mainly because the aftermath has been substantially less important in Tokyo.

Also, the proposed method successfully identified the nature of the event in Tokyo as it reports descriptor tags that are only related to the earthquake and none are referring to the tsunami that did not reach Tokyo (Table 1).

The second reported burst corresponds to the cherry blossom viewing that mainly took place during the second weekend of April. All descriptor tags that stand for this event are related to cherry blossoms and cherry



(a) Two views from StreetView and two from Flickr of the Ishinomaki coast, Miyagi.



(b) A view from StreetView and one from Flickr taken in Ishinomaki city, Miyagi.

Figure 5: Ishinomaki, Miyagi, before and after the tsunami

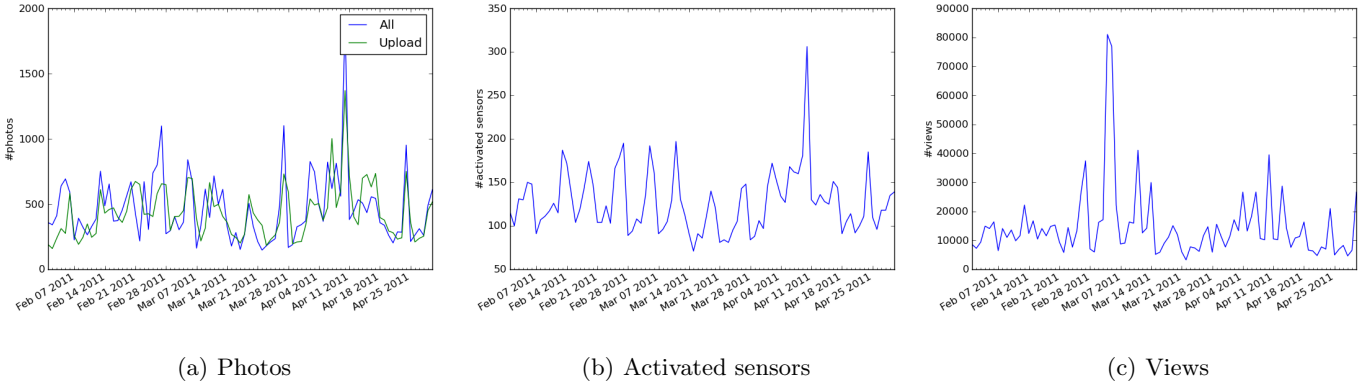


Figure 7: Photos taken in Tokyo from 2011/02/01 to 2011/03/31

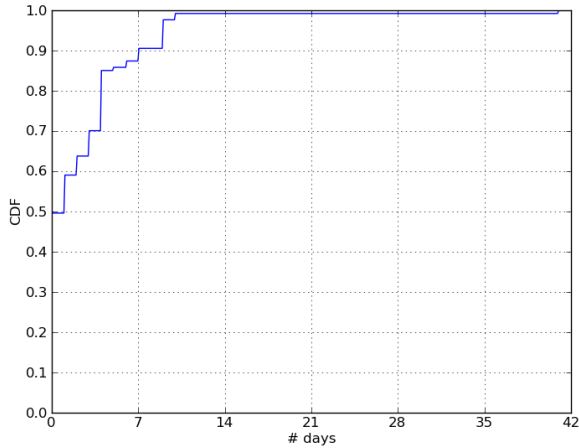


Figure 8: CDF of the delay in day between the shooting of the photo and the upload on Flickr. These are the only photos taken in Tokyo at the day of the earthquake and that have at least one descriptor tag (i.e. 422 photos).

blossom viewing, respectively *sakura* and *hanami* in Japanese.

Figure 8 depicts the delay between the shooting of the photo and the upload on the Flickr website of all photos taken the day of the earthquake and related to the disaster. 49% of these photos were available the day of the earthquake and 90% were available within one week. Compared to Miyagi, the photos from Tokyo were rapidly available mainly because the aftermath was considerably less important than in the northern part of Japan.

4.1.3 Osaka

In Osaka the Tohoku earthquake has been measured at 3 on the JMA scale and no damage was reported. We

notice that the Japan Meteorological Agency reports yearly about 100 earthquakes of seismic intensity 3 or greater [1] (earthquake that are ranked below 3 on the JMA scale are barely discernible by people).

From our dataset 8657 photos are taken in Osaka from February to April 2011, the time evolution of the number of photos, activated sensors and views are depicted in Figure 9. Similarly to the photos taken in Tokyo, these three quantities do not exhibit anomalous values the day of the earthquake, but we also observe a sudden increase of the number of photos and activated sensor on April 10.

The proposed method reports a single burst for the data collected in Osaka (Table 1). This burst corresponds to the cherry blossom viewing and shows the importance of this activity all over Japan. Since no particular pattern of sensor activation is exhibited in Osaka when the earthquake occurred, the proposed method does not reveal any specific event in Osaka at this period of time.

4.1.4 Discussion

Our results highlight the occurrence of two major events during March and April 2011; the earthquake-triggered disasters and a Japanese traditional custom, cherry blossom viewing. Both events reveal the main singularity of the social networks, the human sensing.

Regarding the Tohoku disaster, the reported bursts decrease as the distance from the epicenter increases. Therefore, the reported bursts convey the seismic intensity of the earthquake as it is felt by the population. Since the proposed approach rely on the apprehension and judgment of humans, the output of the proposed method reveals events that are impressive or interesting to the population. In other words, in the case of the earthquake, our approach does not translate the magnitude of the earthquake but how the people felt it. Therefore, because earthquake happens on a daily basis in Japan the low seismic intensity felt in Osaka did not

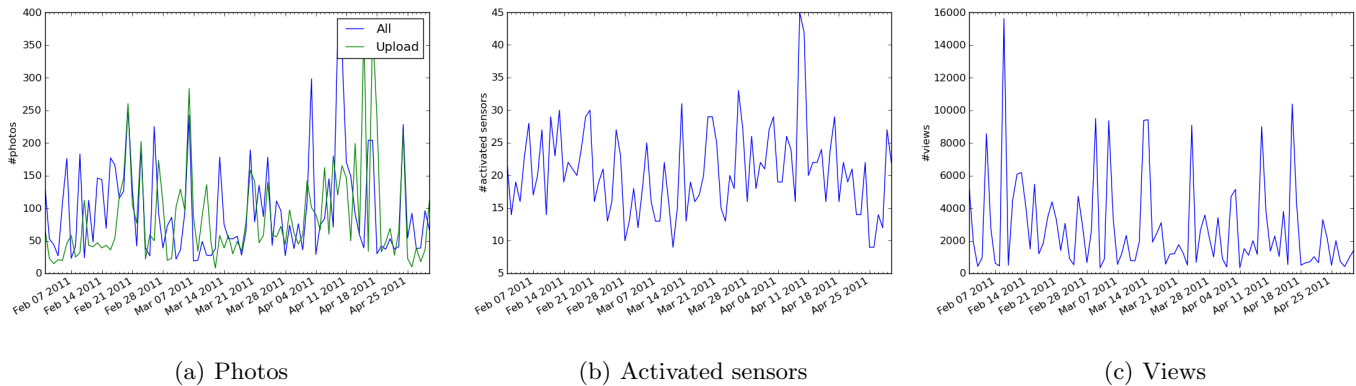


Figure 9: Photos taken in Osaka from 2011/02/01 to 2011/03/31

frightened the population whereas a similar seismic intensity earthquake is apprehended differently by a population that is not used to earthquakes. This human sensing is essential to measuring the impact of disasters on the population and cannot be achieved with traditional sensors that measure physical quantities.

The cherry blossom viewing also exhibits the human behavior of the Flickr sensors. While this traditional event is nation wide, the corresponding bursts are identified only in Tokyo and Osaka. Although cherry trees also bloomed in mid-April in Miyagi [3], the population there was uniquely concerned in the tsunami aftermath and disregarded the traditional event.

4.2 Tuscaloosa Tornado

In addition to the Tohoku earthquake and tsunami we briefly present in this section the analysis of another disaster, the 2011 US tornado outbreak. In particular we focus on one tornado that struck Tuscaloosa, Alabama, on April 27 and caused 47 deaths and significant damage. The key difference of this event compared to the Tohoku disaster is the limited size of the damaged area. Thus, using the Tuscaloosa tornado we highlight the efficiency of our method to locate disasters in space.

The proposed method identified a burst corresponding to the tornado and 65 photos having the two descriptor tags: *tuscaloosa*, *tornado*. As shown on Figure 10, the proposed method takes advantage of the photo’ geo-tags to locate them on a map. The position of the 65 photos clearly reveal the path of the tornado that struck Tuscaloosa. We also observe that the density of taken photos follows the distribution of the population as more photos are identified in the downtown than in the suburbs.

5. CONCLUSION

In this paper we analyze the behavior of Flickr users during disaster outbreaks and show that their photos

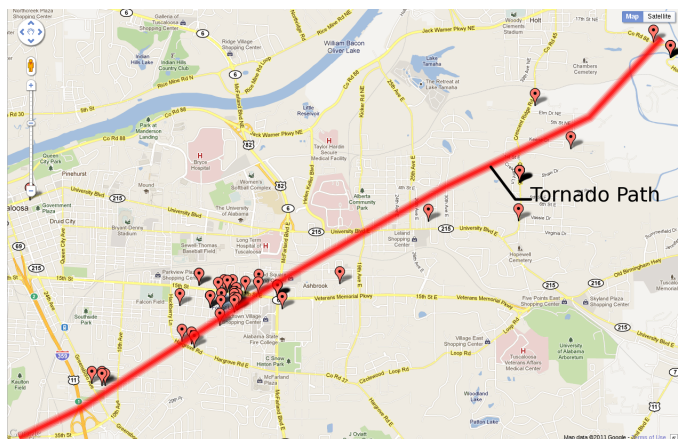


Figure 10: The path of the tornado that struck Tuscaloosa on 2011/04/27 and corresponding Flickr photos retrieved by the proposed method (red markers)

allows one to locate a disaster in time or space and evaluate the disaster’s aftermath. We proposed a methodology to detect disasters through Flickr photos and to extract information concerning the nature of disasters. We carefully evaluated the proposed method using two study cases; the Tohoku earthquake/tsunami and the Tuscaloosa Tornado. Our results highlight the sensitivity of Flickr users to disaster outbreaks and the accuracy of the information they provide.

Acknowledgments

This research was partially supported by the JSPS fellowship program.

6. REFERENCES

- [1] S. B. Bozkurt, R. S. Stein, and S. Toda. Forecasting probabilistic seismic shaking for greater tokyo from 400 years of intensity

- observations. *Earthquake Spectra*, 23(3):525–546, 2007.
- [2] M. Dubinko, R. Kumar, J. Magnani, J. Novak, P. Raghavan, and A. Tomkins. Visualizing tags over time. *WWW '06*, pages 193–202, 2006.
- [3] JMA. Observation: 2011 cherry blossoms. http://www.data.jma.go.jp/sakura/data/sakura2011_mankai.html.
- [4] R. Kayen, W. Steele, B. Collins, and K. Walker. Google earth mapping of damage from the niigata-ken-chuetsu m6.6 earthquake of 16 july 2007. *6th Intl. Conf. Case Hist. Geotech. Engr.*, 2008.
- [5] J. Kleinberg. Bursty and hierarchical structure in streams. *KDD '02*, pages 91–101, 2002.
- [6] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins. On the bursty evolution of blogspace. *World Wide Web*, 8:159–178, 2005. 10.1007/s11280-004-4872-4.
- [7] C. D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.
- [8] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, feb 1989.
- [9] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. *WWW '10*, 2010.
- [10] J. D. Sutter. 5 billionth photo uploaded to flickr. http://articles.cnn.com/2010-09-20/tech/flickr.5.billion_1_photo-sharing-site-flickr-facebook.
- [11] F. Yamazaki, Y. Yano, and M. Matsuoka. Visual damage interpretation of buildings in bam city using quickbird images following the 2003 bam, iran, earthquake. 21(S1):S329–S336, 2005.
- [12] D. Yates and S. Paquette. Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake. *International Journal of Information Management*, 31(1):6 – 13, 2011.