# Demo: Event Localization using Instagram

Prasanna Giridhar
Department of Computer Science
University of Illinois Urbana Champaign
giridha2@illinois.edu

Tarek Abdelzaher
Department of Computer Science
University of Illinois Urbana Champaign
zaher@illinois.edu

## Abstract

This demonstration exploits the popularity of social networks as participatory sensing systems, where users share events via pictures and videos. We are specifically interested in networks that carry `pictorial` information because taking a picture (of an event) requires physical proximity, thereby revealing the location of the photographed event. We show that such networks may help `localize` important events in space and in time. We consider Instagram as the social network of choice and limit ourselves to `urban` events, since social networks are not dense enough in rural spaces. We demonstrate the feasibility of using Instagram for urban event localization thanks to a simple service for mapping current and past events matching a user-specified query. Users are offered a real-time API that allows them to search for locations of events of choice, and experience them remotely in near real-time through participant pictures.

*Keywords*

Localization, Social Sensing

## 1 Introduction

The number of smartphone users is over 2 billion at present [1]. These devices have a wide range of built-in sensors, such as GPS, cameras, and accelerometers, to name a few. From a sensor network point of view they can be considered as nodes with location-aware data. In this demo, we focus on geotagged pictures, shared on social networks, as a way to determine exact locations of large-scale physical phenomena and other events observed by humans in near real time.

At present, several photo sharing services allow users to share pictures with location information in real time. Instagram [2] is one of the most popular applications in this domain, with over 400 million users including 75 million daily active users across the world [3]. With the advent of such services that can be regarded as participatory sensing systems, it becomes possible to design robust and efficient algorithms capable of utilising the social network data to better understand urban events.

We demonstrate a generic platform for detecting and localizing events of different types such as music concerts, sport events, civil unrest, and natural calamities, using Instagram. To appreciate the feasibility of detectign these events using Instagram, we show the distribution of pictures tagged with event-relevant keywords over time in table 1. The column labeled, `D`, refers to the exact date when the event occurred. The other columns denote the previous and subsequent days. It is clear that a distinguishable peak occurs at the date of the event, if one searches using representative keywords.

We take the advantage of this observation to design a solution capable of detecting events of user-specified types. Further, by inspecting the metadata of matching pictures, we show that we can identify event location. The solution we propose is based on an adaptive mechanism to find clusters of data points related to an event. No event-specific training is required. A distance threshold is determined automatically and used by the algorithm to maximize a cluster-quality metric. We also determine the best candidate clusters for event localization by comparing cluster sizes.

We present the design of the demonstrated tool in Section 2. A description of the proposed demonstration is given in Section 3. Finally, conclusions are presented in Section 4.

## 2 System Design

The goal of our work is to identify the locations of real world events using Instagram photos. In the following subsections we describe the functionality of respective components in our architecture.

### 2.1 Query Search

The first step in our pipeline is to get results from the Instagram API based on a query provided by the user. This query can be as simple as a hashtag associated with the event (for example, `#ParisAttack`). The obtained results are then pre-processed to remove objects without any location information. The filtered dataset is then forwarded to the next stage in the pipeline. It is composed of pictures that refer to the event via tags or keywords. We shall henceforth call them pictures that tag the event.

International Conference on Embedded Wireless
Systems and Networks (EWSN) 2016
15–17 February, Graz, Austria

ISBN: 978-0-9949886-0-7

**Table 1. Distribution of images for events**

| Event | Date | City | Number of Pictures | | | |
|---|---|---|---|---|---|---|
| | | | D-2 | D-1 | D | D+1 |
| Nepal Earthquake | 04/25/15 | Lamjung, NEP | 0 | 0 | 133 | 92 |
| Chile Earthquake | 09/16/15 | Illapel, CHILE | 0 | 0 | 171 | 156 |
| Maroon V | 09/26/15 | Melbourne, AUS | 0 | 0 | 133 | 28 |
| Taylor Swift | 10/3/15 | Toronto, CAN | 1 | 521 | 973 | 586 |
| North American storm complex | 10/4/15 | Charleston, SC, USA | 30 | 128 | 222 | 94 |
| Chicago Marathon | 10/11/15 | Chicago, IL, USA | 33 | 38 | 744 | 147 |
| Paris Attack | 11/14/15 | Paris, FR | 0 | 222 | 349 | 114 |

## 2.2 Adaptive Clustering

In this work, we use adaptive distance-based clustering to group together pictures that tag the event based on their locations. It is a non-paramteric technique that identifies the best distance threshold to use for clustering a given dataset. At every stage, we determine the quality of clusters formed using a silhouette measure for the formed cluster objects. This scoring mechanism comprizes a three-step process:

- Cohesion Factor ($a_i$): For the $i^{th}$ data point, we find the average distance to all other data points within the same cluster.
- Separation Factor ($b_i$): For the $i^{th}$ data point, we find the average distance from all the data points of another cluster to which it does not belong. Then, we take the minimum of the average distances from all the clusters.
- Silhouette Coefficient: Finally, we assign a score to the $i^{th}$ data point using the equation $s_i = \dfrac{(b_i - a_i)}{max(a_i, b_i)}$.

The silhouette coefficient for any data point is in the range $[-1, 1]$. The ideal best case is when $a_i = 0$ for which the maximum value of 1 is attained. Using this scoring technique we try to determine the distance threshold for which the best average score is acheieved using the given data points.

## 2.3 Candidate Selection

The next step in the pipeline is to identify the candidate clusters of pictures that correspond to the real location of the event. An important challenge is to automatically determine whether a single event or multiple events of the same type have occurred at a given time frame. For example, when looking for concerts by a popular artist, say, Taylor Swift, we expect to see only one event at a time. The artist, being a single entity, can perform only at one valid location at a given time. Selecting multiple clusters for single-entity events will result in false positives. On the other hand, when looking for a generic event, such as a marathon, we accept the possibility of finding multiple concurrent events at different locations.

In order to differentiate between single-entity and multiple-entity events, we performed cluster-size ratio analysis for the largest clusters found that match specific keywords in all the datasets we collected. The average ratio of the largest cluster size to the next largest cluster size in case of single entity and multiple entity events were 7.1 and 1.25, respectively. Hence an empirical threshold was set to distinguish single-entity and multi-entity events based on the drop-off in cluster sizes matching the event. We return only the top most cluster if the ratio exceeds the threshold. Otherwise, we return all clusters above a given size.

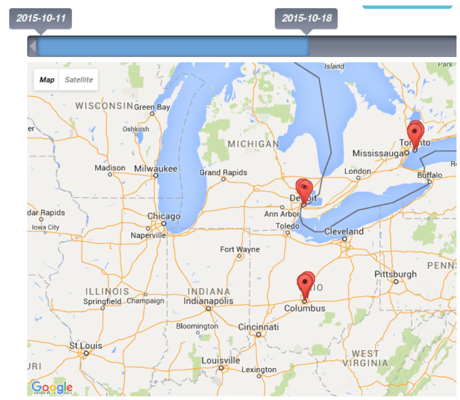


**Figure 1. Timeline map based interface**

## 2.4 Timeline Interface

The final step in the pipeline is to output the localized events on a timeline based visual map interface. The timeline slider spans the date range provided by the user during query search and can be moved in steps of 24 hours to see all the localized events on a particular date. The map interface shows pins corresponding to each event to the exact geo-coordinate location. Figure 1 is a sample map interface for localized `marathon` events on `10/18/15`.

# 3 Demonstration

The demo will allow users to interact with our search tool, allowing visualization of locations of past and current events (data collected in real time at the time of the demo). The tool will be hosted on a public server so that users can interact with the system to visualize and localize events. Users will be able to view the unfolding events at the different identified locations.

# 4 Conclusions

In this demonstration, we illustrate a visualization tool capable of automatically identifying the locations of physical events of interest from social network data. Preliminary evaluation shows that the tool correctly identifies event locations most of the time. The accuracy of our tool in localizing events is expected to increase with more Instagram use. A full manuscript describing the work is currently in preparation.

# 5 References

[1] http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.

[2] http://www.instagram.com/.

[3] http://expandedramblings.com/index.php/important-instagram-stats/.