Transactions in GIS

Research Article

Transactions in GIS, 2015, 19(5): 694–715

Triangulating Social Multimedia Content for Event Localization using Flickr and Twitter

George Panteras, Sarah Wise, Xu Lu, Arie Croitoru, Andrew Crooks and Anthony Stefanidis

George Mason University

Abstract

The analysis of social media content for the extraction of geospatial information and event-related knowledge has recently received substantial attention. In this article we present an approach that leverages the complementary nature of social multimedia content by utilizing heterogeneous sources of social media feeds to assess the impact area of a natural disaster. More specifically, we introduce a novel social multimedia triangulation process that uses both Twitter and Flickr content in an integrated two-step process: Twitter content is used to identify toponym references associated with a disaster; this information is then used to provide approximate orientation for the associated Flickr imagery, allowing us to delineate the impact area as the overlap of multiple view footprints. In this approach, we practically crowdsource approximate orientations from Twitter content and use this information to orient Flickr imagery accordingly and identify the impact area through viewshed analysis and viewpoint integration. This approach enables us to avoid computationally intensive image analysis tasks associated with traditional image orientation, while allowing us to triangulate numerous images by having them pointed towards the crowdsourced toponym location. The article presents our approach and demonstrates its performance using a real-world wildfire event as a representative application case study.

1 Introduction

Fostered by Web 2.0, ubiquitous computing, and corresponding technological advancements, social media have become massively popular during the last decade. The term social media refers to a wide spectrum of digital interaction and information exchange platforms, ranging from blogs and micro-blogs (e.g. Twitter, Tumblr, and Weibo), to social networking services (e.g. Facebook), and multimedia content sharing services (e.g. Flickr and YouTube). Regardless of the particularities of each platform, these social media services share the common goal of enabling the general public to contribute, disseminate, and exchange information (Kaplan and Haenlein 2010). Traditional web-accessible information has always been rich in geographic content (Silva et al. 2006), and this of course remains true for social media content. But in addition to geographical references within the data, social media is also becoming increasingly geotagged as a result of the proliferation of location-aware devices (Hurst et al. 2007, Valli and Hannay 2010, MacEachren et al. 2011; Stefanidis et al. 2013b). Accordingly, social media content is emerging as a rich source of geospatial information, presenting our community with many opportunities and challenges (Sui and Goodchild 2011). The opportunities are primarily associated with the potential of these crowdsourced data to complement authoritative datasets by contributing timely information (e.g. Gao et al. 2011). The challenges are reflections of the very nature of these datasets: diverse data structures and formats, and variations in quality and accuracy (Agichtein et al. 2008).

Address for correspondence: Dr. Arie Croitor, George Mason University Center for Geospatial Intelligence, 4400 University Drive, MS 6C3 Fairfax, VA 22030, USA. E-mail: acroitor@gmu.edu

doi: 10.1111/tgis.12122

Driven by the allure of opportunity, the geographical community has been experimenting over the past few years with harvesting geospatial information from social media content. For example, studies addressed the use of Twitter reports to gain knowledge regarding the breaking and progression of natural disasters such as wildfires (De Longueville et al. 2009), earthquakes (Crooks et al. 2013) and flooding (Fuchs et al. 2013). The spatiotemporal analysis of Twitter content has also been used to track disease outbreaks (Signorini et al. 2011; Sugumaran and Voss 2012), or to identify the formation of international communities and the communication of information during political crises (Stefanidis et al. 2013a). While these studies are advancing our ability to understand the geospatial content of social media and the manner in which they are used to communicate various forms of information, they were primarily focused on just a portion of social media content: text. However, social media content information is not just textual. Flickr and Instagram offer massive records of imagery, and YouTube videos are rich in visual content, providing an additional dimension through which information is communicated. Some early attempts to exploit the content of these additional services have primarily focused on the analysis of point patterns. For instance, Li and Goodchild (2012) studied point patterns of georeferenced Flickr imagery in conjunction with toponyms in their metadata to identify places through user references to them. Other efforts attempt to recognize activity and behavioral patterns by analyzing these spatiotemporal points of geotagged entries, such as identifying attractive destinations (Kisilevich et al. 2010) or constructing travel itineraries (De Choudhury et al. 2010).

Despite these efforts, the multimedia content of social media remains underexplored. In this article we contribute towards bridging this research gap by examining the benefits of the complementary use of heterogeneous sources of social multimedia feeds to assess the impact of a natural disaster. More specifically, we are introducing a novel social multimedia triangulation process that uses collaboratively Twitter and Flickr content in a two-step integrated process: Twitter content is used to identify toponym references associated with a disaster; this information is then used to provide approximate orientation for the associated Flickr imagery, allowing us to delineate the impact area as the overlap of multiple view footprints. In this approach, we practically crowdsource approximate orientations from Twitter content and use this information to orient Flickr imagery and identify the impact area through viewshed analysis and viewpoint integration. This approach allows us to triangulate numerous images by having them pointed towards the crowdsourced toponym location while avoiding computationally intensive image analysis tasks associated with image orientation (e.g. the identification of conjugate features). In this article we present our approach and demonstrate its performance using a wildfire event as a representative application. The remainder of the article is structured as follows. In Section 2 we discuss the use of social media content in crises. In Section 3 we describe the proposed integrated methodology. In Section 4 we present the results of the proposed methodology using as a test case a wildfire in the central US, and in Section 5 we conclude with an outlook.

2 Social Media and Crowdsourced Crisis Information

With the general public nowadays having at its fingertips technology that a few years ago was available only to advanced computing laboratories, it is only natural that the amount of crowdsourced information of computational merit is rapidly growing, with volunteered geographical information (VGI) being a large portion of this content (Goodchild 2007). Both domain experts and amateurs alike can now generate and disseminate geospatial content through collaborative web mapping services such as Google Map Maker, OpenStreetMap, or WikiMapia (Rouse et al. 2007, Haklay et al. 2008), or pursue novel visual exploration practices through map mashups (Wood et al. 2007). Furthermore, enhanced open-source solutions (e.g. QGIS and R) support more complicated data analysis tasks, with capabilities often comparable to dedicated geographical information software. These capabilities have been put to use in crisis situations, with ad-hoc communities of neocartographers emerging to provide timely updates of geospatial datasets (Liu and Palen 2010). The post-earthquake mapping of Haiti in 2010 is a representative example of a very successful use of the crowd to capture and disseminate information that outperformed authoritative alternatives (e.g. Norheim-Hagtun and Meier 2010, Zook et al. 2010).

While such efforts represent explicit contributions of geospatial content, the same Web 2.0 technological advances have also led to the substantial growth of implicit geospatial content that is contributed by the crowd, especially through social media outlets (Stefanidis et al. 2013b). As public participation in social media is rapidly increasing, the information published through such sites is becoming a new type of big geospatial data (Croitoru et al. 2014). For example, in 2012 Twitter users were posting nearly 400 million tweets daily, or over 275,000 tweets per minute (Forbes 2012), doubling the corresponding rates of 2011 (Twitter 2011). At the same time, 100 million active users are uploading daily an estimated 40 million images in Instagram (2014). Furthermore, every minute, Flickr users upload in excess of 3,000 images (Sapiro 2011), and YouTube (2013) users upload approximately 72 hours of video. Due to their nature, social media are well-suited to communicate information about rapidly evolving situations, ranging from civil unrest in the streets of Cairo during the Arab Spring events (Christensen 2011) or New York during Occupy Wall Street (Wayant et al. 2012), to reporting natural or anthropogenic disasters like wildfires (De Longueville et al. 2009), earthquakes (Crooks et al. 2013), flooding (Vieweg et al. 2010, Triglav-Čekada and Radovan 2013, Fuchs et al. 2013), or nuclear accidents (Fontugne et al. 2011, Utz et al. 2013).

Social media content is often geotagged, either in the form of precise coordinates of the location from where these feeds were contributed, or as toponyms of these locations. Studies have highlighted how the percentage of precisely geolocated (i.e. GPS-derived coordinates) tweets may vary depending on the event and location, ranging approximately from 0.5% to 5.0% of the total data corpus (Mahmud et al. 2012, Stefanidis et al. 2013a). This rate may be higher depending on the area of study, the thematic content, and the underlying conditions. For example, a dataset collected from Japan following the Fukushima disaster reflected a data corpus where 16% of the tweets were precisely geolocated (Stefanidis et al. 2013b). This spike is attributed to the fact that the dataset from Japan reflected a technologically-advanced community that was on the move (following the tsunami and subsequent nuclear accident), so that users were tweeting using primarily their mobile devices. Both of these situations, namely the proliferation of technology in a society and an increased use of mobile (and other location-aware) devices to post tweets, are conditions that tend to produce higher rates of geotagged content in social media. In addition to precisely geotagged tweets, we have observed that approximately 40% to 70% of tweets come with a descriptive toponym related to the location of the user. Regarding imagery and video contributed as part of social media, a recent study has indicated that approximately 4.5% of Flickr and 3% of YouTube content is geotagged (Friedland and Sommer 2010).

The geographic content of social media feeds represents a new type of geographic information. It does not fall under the established geospatial community definitions of crowdsourcing (Fritz et al. 2009) or VGI, as it is not the product of a process through which citizens explicitly and purposefully contribute geographic information to update or expand geographic databases. Instead, the type of geographic information that can be harvested from social media feeds can be referred to as Ambient Geographic Information (AGI) (Stefanidis et al. 2013b); it is embedded in the content of these feeds, often across the content of numerous entries rather than within a single one, and has to be somehow extracted. Nevertheless, it is of great importance as it communicates instantaneously information about emerging issues.

Recognizing the potential of this emerging trend, the term crisis informatics has been introduced to describe the analysis of the responses of the crowd during disasters as they are captured through Web 2.0 enabled applications (Hagar and Haythornthwaite 2005, Palen et al. 2007). This has empowered our community to advance from early endeavours, which were focusing mainly on visualizing diverse datasets through map mashups (e.g. Hudson-Smith et al. 2009), to more advanced analytical approaches. Several studies have addressed particular applications that relate to crisis informatics using social media content. For example, Sakaki et al. (2010) presented a probabilistic spatiotemporal model that uses Twitter responses to detect the epicenter and trajectory of earthquake waves. Crooks et al. (2013) extended this line of work, arguing that Twitter feeds resemble a hybrid form of a sensor system that enables the identification and localization of the impact area of the event. They showcased the use of this approach to quickly locate the epicentre of a large earthquake at an accuracy that is comparable to authoritative systems (such as the US Geological Survey "Did You Feel It" website). Moving from earthquakes to forest fires, De Longueville et al. (2009) examined the use of location-based social networks as a source of information during crises. Starbird and Palen (2011) highlighted the self-organizing nature of the emergency response from Twitter users during the Haiti earthquake. They showed that digital volunteers are inclined to contribute valuable information during the occurrence of sudden and tragic events. These studies have primarily addressed text and accompanying geospatial information, using primarily geotagged tweets as the information source. These studies are prototypical highlights of the value of harvesting social media content to gain situational awareness and understand how such events unfold and impact the population.

Other efforts have also explored patterns of contributions of imagery in social media, to extract meaningful geospatial information from it. For example, Liu et al. (2008) conducted a correlational qualitative study to examine if and how disaster-related Flickr activity evolved for six major disasters between December 2004 and October 2007. This study provided an extensive discussion about the formulation of norms and practices around the contribution of photographic content during disaster response and recovery efforts. The behaviour of Flickr users has also been examined by Fontugne et al. (2011) as an indicator of major event occurrences (e.g. in the form of contribution bursts at specific locations), using as case studies the Tuscaloosa tornado and Tohoku earthquake. Pohl et al. (2012) also utilized the visual content of both Flickr and YouTube to extract crisis sub-events based on images, metadata and videos. However, these efforts have focused primarily on the location of the contributions, rather than attempting to delineate the spatial footprint of the affected area.

While some efforts have addressed the geographical analysis of Twitter content, and other efforts have addressed the analysis of contribution patterns in Flickr (e.g. Senaratne et al. 2013), the challenge of combining these diverse sources of social multimedia information in order to derive event knowledge still remains. This is the challenge that this article addresses. More specifically, we introduce in Section 3 a novel approach that mines Twitter content for extracting toponym references associated with crisis events, and subsequently uses this information to guide a hybrid triangulation of Flickr imagery in order to delineate the event footprint. The aggregate use of these two different social media sources outperforms the potential provided by each one individually, thus rendering such cross-source analysis particularly valuable for the better exploitation of the wealth of information conveyed through various social media platforms.

3 A Cross-Source Triangulation Framework

As discussed above, our main objective is to integrate social media content referring to an event (e.g. a natural or anthropogenic disaster) across sources in order to advance our capability to geolocate this event and delineate its footprint. In order to meet this objective we introduce a novel multimedia triangulation framework. Through this framework, contribution patterns are extended from simple point clouds (indicating the location of the contributors) to become the equivalent of views of a particular event (which involve an understanding of the relationship between the contributor and the event). These views can then be synthesized to delineate the event footprint via viewshed analysis. We accomplish this goal through the two-step process that is summarized in Figure 1. The first component of our approach entails Twitter content analysis for the identification of toponym references associated with the event of interest (presented in Section 3.1). Using this information we then harvest Flickr imagery using geolocation and tag constraints: we query the Flickr Application Programing Interface (API) to retrieve images from the broader vicinity of the toponym, and with tags that are related to it as well. These images are then oriented using the toponym information as a reference point, and their viewable area footprints are integrated via viewshed analysis in order to derive an estimate of the event footprint (as a probability map), as presented in Section 3.2.

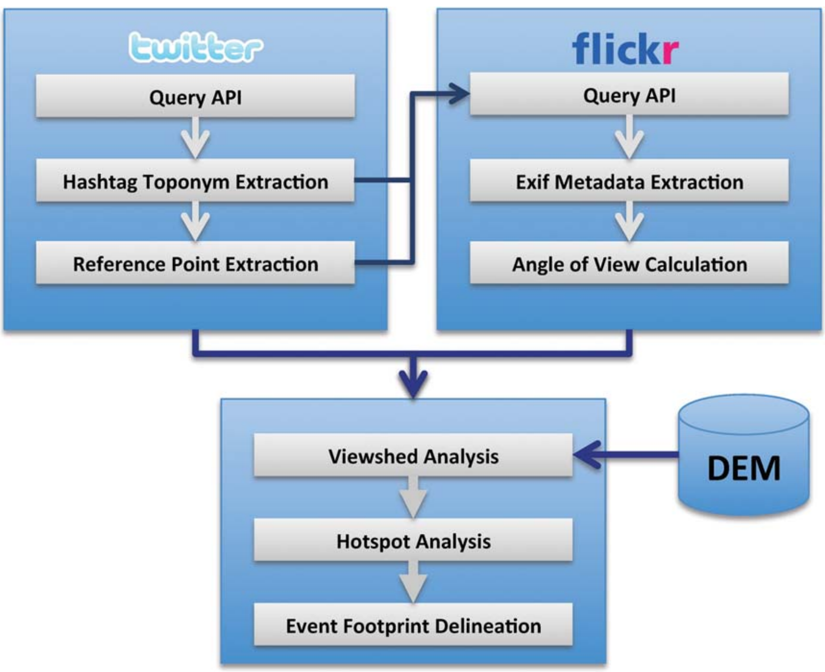

Figure 1 The cross-source triangulation framework

The underlying assumption in our approach is that tweets contain references to the location of the event, whereas Flickr contributions provide views of it. This methodology represents a cross-platform social multimedia analysis approach for event triangulation.

While Twitter is utilized in this framework to derive the approximate location of a given event (which can then be further refined using Flickr), it should be noted that other sources may also be exploited for the extraction of such information. For example, a toponym reference can be extracted from other communication avenues, such as news media feeds or blogs, which could substitute the Twitter reference point extraction process in Figure 1. Another source of such information may very well be Flickr itself, as image annotations may contain toponym references. However, such annotations in Flickr tend to vary in terms of their frequency (Ames and Naaman 2007, Nitta et al. 2014), thus potentially limiting the suitability of Flickr annotations alone for this purpose. This is further attenuated if we consider the data volume differences between Twitter and Flickr. For example in this particular study, the number of Flickr contributions is roughly 0.5% of the number of tweets reporting the same event. This is consistent with the reports of overall data traffic associated with these two social media services (Croitoru et al. 2013).

3.1 Event Localization using Toponym References in Twitter

In order to best communicate how the various components of our framework are operating and integrated, we use the 2012 wildfire of Waldo Canyon in Colorado Springs (Colorado, USA) as a case study. The wildfire started in June 23, 2012 and was not fully contained until July 12, 2012 (the study period), which is used as the study period in this article. During that time, the wildfire consumed a total area of 74 km^2, and was considered the most destructive wildfire in Colorado's history at the time, based on the extent of damage to property (McGhee 2013). Figure 2 provides an overview of study area, showing Waldo Canyon to the northwest of Colorado Springs, with the actual wildfire area overlaid along with the location of geolocated Flickr images during the event.

We collected relevant Twitter data from the Twitter API using the keyword "Fire" over the study period, resulting in a corpus of 97,866 tweets among which 41.4% are retweets. It is worth noting that as we analyze the content of tweets rather than their spatial distribution, the presence of relatively high retweet levels is likely to contribute to the emergence of toponyms in our data corpus, thus further facilitating the detection of the relevant toponyms. We therefore view retweeting as a crowd-sourced curation process, whereby the general public weighs upon twitter content and assigns gravity to it in a variety of ways, with retweeting being the most prominent (e.g. Boyd et al. 2010).

The content of the tweets corpus was analyzed in order to generate the word-cloud shown in Figure 3. This entailed parsing the text to remove all non-hashtag punctuation (e.g. emoticons), removing articles, and converting all text to lowercase. The word-cloud visualizes the frequency of individual words in our Twitter data corpus, with larger words being those encountered more frequently. It is easy to observe that, after the word fire (which was the keyword used to query the Twitter API for this study) the predominant terms that emerge are geographical in nature, with "Waldo" being the dominant among them – either by itself or as a part of a compound hashtag. This heavy use of geographical references in social media narrative when reporting natural disasters has also been noted in other natural disaster studies. Vieweg et al. (2010) stated that in their studies toponym references were present in as many as 40% of tweets reporting wildfires and 18% of tweets reporting flooding. This is also

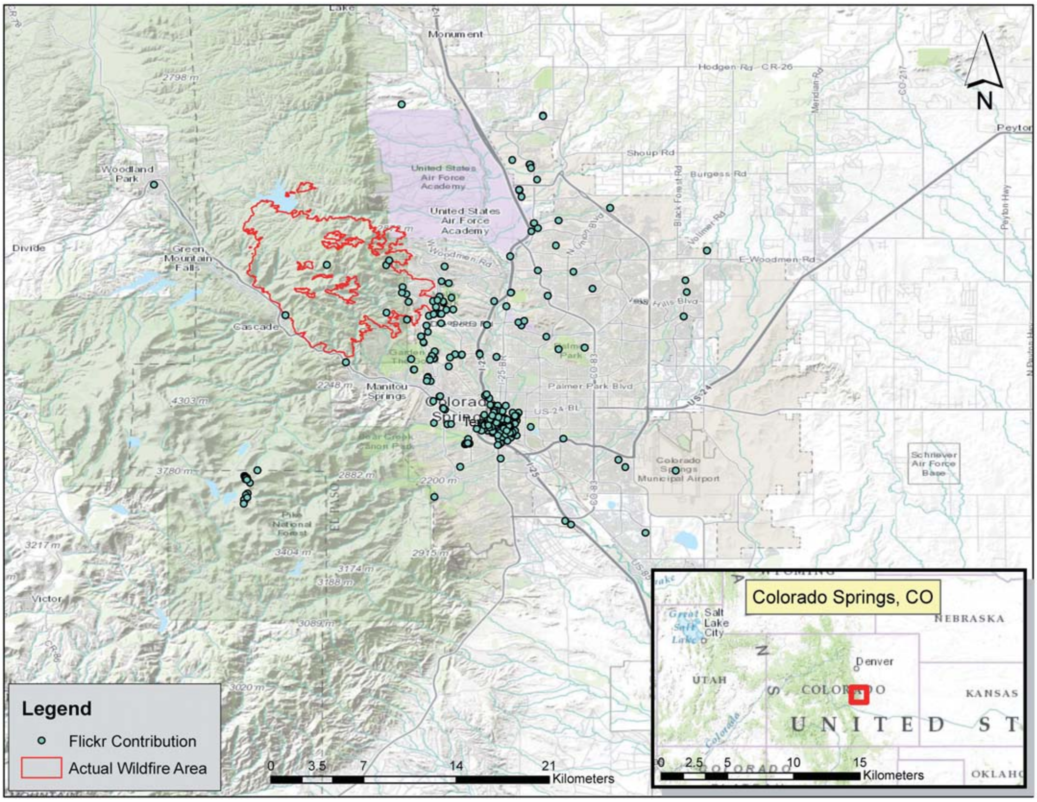

Figure 2 Overview of the study area

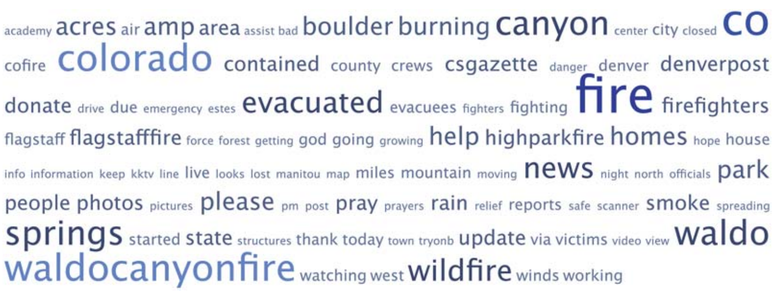

Figure 3 Word-cloud of Tweeter most frequent terms and hashtags during the wildfire

consistent with studies addressing the broad presence of toponyms in reporting various types of breaking news (Lieberman and Samet 2011, Stefanidis et al. 2013a).

The Twitter data corpus was then converted to lowercase and filtered to extract all hashtags. Figure 4 shows the frequency over time of the 10 most popular hashtags for the duration of the wildfire event. As can be seen from it, "#waldocanyonfire" has emerged as the top hashtag associated with this event, a term which encompasses both the nature of the event and the location of it. The emergence of hashtags like this through a bottom-up process, from

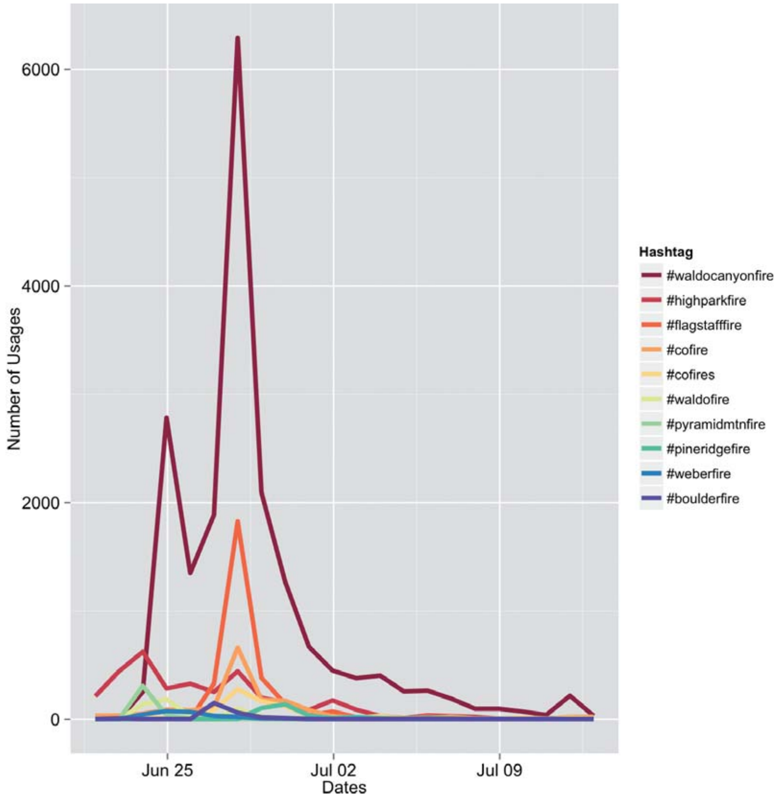

Figure 4 Usage of most frequently adopted hashtags over the wildfire period

the crowd and adopted by the crowd, serves as further indication for the value that the public places on the locational information when referring to major events such as this. In fact, all 10 most popular hashtags were of the form {“#’,<location>,<event>}.

Similar to Figure 4, in Figure 5 we show the frequency of the 10 most popular toponym references in the Twitter narrative associated with the event. The results confirm the popularity of Waldo Canyon, while also suggesting the emergence of a hierarchical structure in the toponym references, with the State (Colorado) leading, and the particular area within it (Waldo) following. The remaining toponym references relate to the areas that were secondarily affected by the wildfire event, e.g. Flagstaff Mountain, and the smaller towns of Manitou and Estes. In our case we selected the toponyms manually for quality control purposes; however, this process can be automated using a gazetteer. Using Waldo Canyon as the prominent location in the Twitter corpus, we retrieved the point location of this toponym from a gazetteer (Google Geocoder), and used it as the reference point of the event in subsequent analysis. Once the approximate geolocation of the event is determined through the analysis of Twitter content (toponyms and hashtags) we proceed with the analysis of Flickr contributions to delineate the impact area of this event, as described in Section 3.2 below.

3.2 Impact Area Delineation through Viewshed Analysis of Flickr Contributions

While Twitter provides textural information of the event, Flickr provides us with visual evidence of the event in the form of images. Such information is often accompanied with

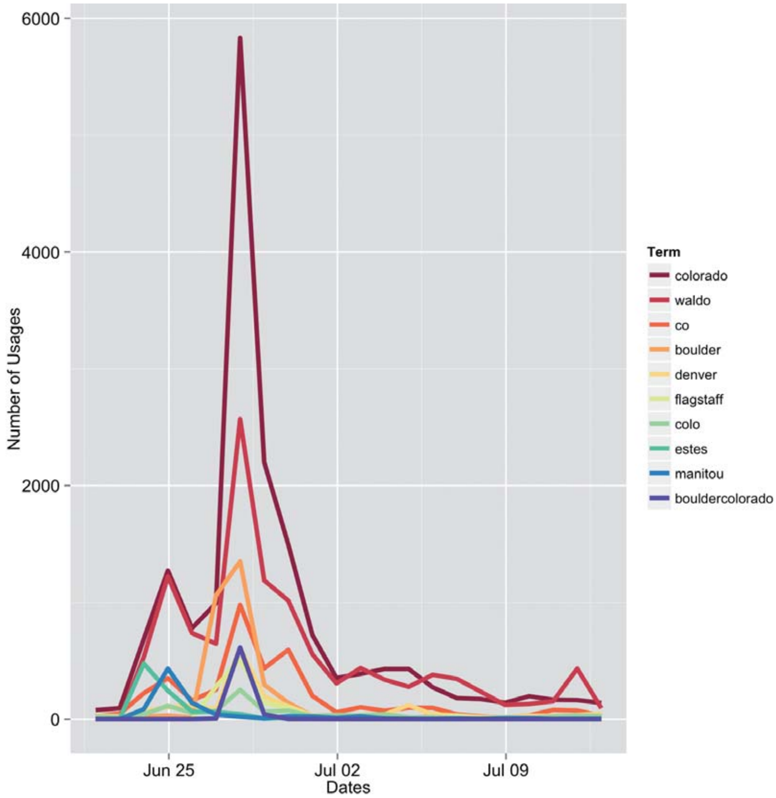

Figure 5 Usage of most frequently adopted toponym terms over the wildfire period

geolocation information either as exact geographical coordinates (via metadata), manual placement of the image on a map (via the Flickr map interface), or as an approximate location (via geographically relevant keywords, i.e. toponyms). In our study we utilize the imagery metadata, which is provided in the Exchangeable Image File (Exif). Exif data provides a range of metadata about the contributed image, including detailed information about the date and time, focal length (f_c), image dimensions (L) and shutter speed. In addition, information about the model and the make of the sensor can be found. Based on such information, all the camera specifications can be retrieved from existing online databases. Finally, in some cases information concerning the direction of view of the image can be found under various Exif fields, for example the "GPS Direction" which is provided when the camera device is equipped with either GPS or an electronic compass. However, such information is often lacking.

Flickr data can be retrieved through a dedicated API (http://www.flickr.com/services/api/flickr.photos.search.html), similarly to Twitter, which supports the user-defined queries. For our study we retrieved data based on a number of query parameters: (1) photos must be geotagged (i.e. *has_geo=1)*; (2) photos must have the tags wildfire and Colorado (i.e. *tags="wildfire,colorado", tag_mode="all")*; (3) photos must have the title or description that contains Waldo Canyon Fire (i.e. *text="Waldo Canyon Fire"*); (4) photos must be within a bounding box (bbox) defined by the study area (i.e. *bbox="-105.316,38.523, -104.291,39.224"*); and (5) the time stamp of the photo must be in the time period of the study (i.e. *min_taken_date="2012-06-24", max_taken_date="2012-07-04"*). Using these parameters a total of 427 images were retrieved of which only 191 (less than 50%) had Exif information. However, while for some of these images the angle of view (AOV) can be derived from Exif information, none of these images included the observer's orientation (i.e. azimuth). This fact, which appears to be frequent in Flickr data (Wueller and Fageth 2008), serves as one of the primary motivations for developing our viewshed analysis methodology. As a result, we use the coordinates of the toponyms and the Exif information to derive both the direction of view (as estimated by the azimuth) and the AOV (as estimated from the focal length and the image size), which we turn to next.

As expected, the contributions in this case are consistent with observed social media and blogosphere patterns (e.g. Stefanidis et al. (2013b) and Shi et al. (2007) respectively): approaching a power law distribution, with few users contributing large portions of the data, and a majority of users making minimal contributions. In our case study the 427 Flickr images that were retrieved were contributed by 38 distinct users, with the median contribution per user being one photo (compared to the average of 11). This deviation between the median and the average values is indicative of the degree of skewness of the contributions among users.

3.2.1 Azimuth and angle of view calculation

The purpose of estimating the azimuth and the AOV is to orient and constrain the extent of the view from each image location as shown in Figure 6. For this purpose, we first establish the AOV using the sensor parameters (i.e. focal length and image dimensions as provided by the image Exif file), and then orient the AOV by calculating the azimuth between the observer location and the event reference point. Generally, three AOVs can be calculated for a given image: the horizontal, the vertical, and the diagonal. As our objective is to establish the extent of the footprint of the event (i.e. wildfire), we utilize the horizontal AOV, which is calculated as:

$$\varphi_{AOV} = 2tan^{-1}\left(\frac{L}{2f_c}\right) \tag{1}$$

where L is the image width and f_c is the sensor focal length. Using Equation (1), the AOV has been calculated for the 191 images for which an Exif file was available. For the remaining 236 images that did not include Exif metadata, the average of the 191 AOVs that were calculated using the Exif data was used as an approximation. Considering that Flickr imagery is increasingly contributed by mobile devices with relatively similar camera characteristics (https://www.flickr.com/cameras), the use of an average value for imagery lacking AOV information is a reasonable approximation (Singla and Weber 2011).

In order to orient the AOVs, we calculated the azimuth between each image location and the event reference point (as described in Section 3.1). More specifically, the calculation of the azimuth for every image was based on the geodetic azimuth using the following formula (Yang et al. 1999):

$$\theta = tan^{-1}\left(\frac{\sin(\lambda_2 - \lambda_1)\cos(\varphi_2)}{\cos(\varphi_1)\sin(\varphi_2) - \sin(\varphi_1)\cos(\varphi_2)\cos(\lambda_2 - \lambda_1)}\right) \tag{2}$$

where φ_1, λ_1 and φ_2, λ_2 are the geographical coordinates of the Flickr image location (or the observer) and the event reference point, respectively. As a result of this calculation, each Flickr image is now associated with a geographic location and an oriented AOV from which a viewshed analysis can be carried out in order to delineate the footprint of the event.

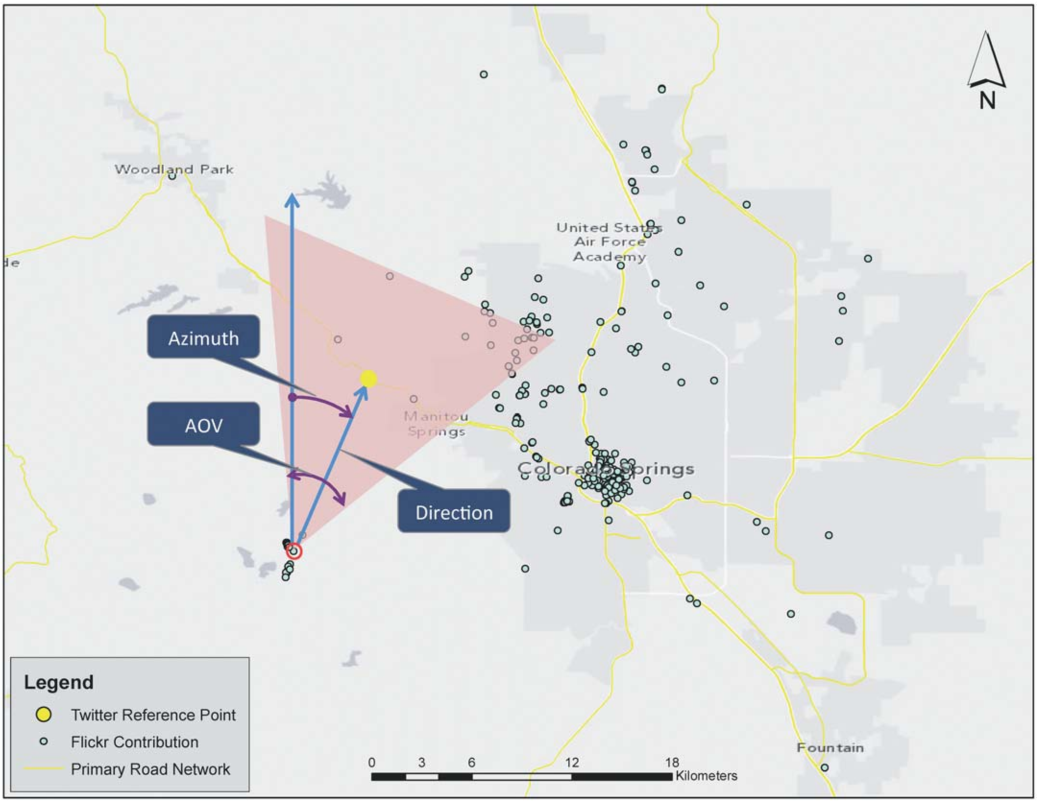

Figure 6 The AOV and azimuth of a given Flickr image

3.2.2 Viewshed analysis

The information extracted in the previous section (i.e. the event reference point, and the azimuth and AOV of each image) can be utilized for estimating the *footprint* of the event we analyze. The underlying principle of this estimation process is that observers who contribute images related to the event are doing so from locations at which the event is visible. It should be noted that here we do not assume that the viewable area of all images is identical, but that these viewable areas share one or more *common* areas that are of interest. Based on this, we apply a crowdsourcing approach for estimating the footprint of the event: while each observation may cover a different viewable area and a corresponding footprint on the ground, by *superimposing* all footprints we can derive an estimation of the event footprint. This process can also be seen as a spatial voting process, where each observer – through the contributed Flickr image – casts a vote on the location of the event in the form of a viewable area. The accumulation of these votes, as measured per unit area in the form of a heat map, can then lead to "hotspots" in which the event is most likely to be found.

In order to estimate the footprint of the event through the superimposition of the footprints of individual views, we must first calculate the footprint of each view separately. Given a viewer location, an AOV and a view direction, the problem of estimating the footprint can be transformed into a viewshed analysis problem. In this problem setting, the viewer parameters are used together with a Digital Elevation Model (DEM) of the area for finding the visible areas of a surface from a given observer location. Viewshed analysis is a well-established technique, which spans across various application areas, from navigation and site selection to landscape planning and telecommunication systems (e.g. Nagy 1994, Fisher 1995, De Floriani and Magillo 2003, Sander and Manson 2007). In our framework, viewshed analysis is utilized to calculate the viewable areas (or cells in the case of a raster grid) between observer and points in the study area, given the reference point of interest (i.e. the event reference point) based on the elevation difference between these points. By systematically applying this calculation to all cells in the study area, we generate a binary map showing the viewable area for each observer. The superimposition of all binary maps for all observers then results in a heat map, where each cell in the map accumulates the number of times the cell was flagged as viewable. It should be noted that while here we assign the same weight to each binary viewshed map during the superimposition process, other weighting schemes could be applied in order to enhance the heat map fidelity for a specific purpose. For example, given a time interval, viewshed maps may be weighted according to their timestamp in order to generate a heat map that highlights the extent of the wildfire during that time interval. However, as in this case study we aim to explore the full extent of the fire, this option was not pursued.

The implementation of the viewshed analysis was carried out in the ArcGIS environment through a workflow consisting of a set of python scripts. This workflow, which systematically applies the viewshed calculation for each image in our data set, provides the ability to control the calculation parameters used. In particular, for each image we set the angular limits of the viewshed calculation as the left and right azimuths of the AOV of the image (which can be derived from the AOV and the azimuth of each image), and set a minimum and a maximum range parameter (measured from the viewer's location) to limit the distances from the viewer for which the viewshed calculation is carried out. The values of these range parameters are set as a function of the average distance between the event reference point and the location of each Flickr image. It is worth noting that in our experiments we utilized the National Elevation Data (NED) data, a 10 m resolution DEM that is available through the US Geological Survey (USGS). The final step of our viewshed analysis includes the superimposition of all viewshed raster grids, resulting in a heat map. Cells having high values in this heat map indicate locations that have been visible more in Flickr imagery in relation to the event, while cells having low or zero values indicate locations that have not been visible in such imagery. Based on these values, we can then analyze hotspots the heat map in order to identify highly visible locations, i.e. locations that were of interest to many viewers on the ground.

3.2.3 Hotspot detection

In the final step of our framework we utilize the heat map that was generated in order to identify hotspots and delineate their extent as an approximation for the footprint of the wildfire event. Here, we refer to a hotspot as a spatial cluster of cells for which high heat map values exist, i.e. clusters that are highly visible to observers in the viewshed analysis. Several well-studied spatial analysis methods exist for the detection of hotspots, among which are Kernel Density Estimation (KDE), Moran's I, and Getis-Ord (Gi*) (Kuo et al. 2012). KDE, which is based on a spatial filtering process, produces a smooth density surface by estimating the surface density (Silverman 1986, Xie and Yan 2008). However, a key difficulty in implementing KDE is the filter bandwidth as well as the ability to test the statistical significance of the results.

Another possible measure is Moran's I, which estimates spatial autocorrelation among similar (low or high) values. While Moran's I could be used for detecting hotspots, its inability to automatically distinguish between high or low hotspots (Griffith 1987) limits its usability for our purpose. In view of these limitations, we utilize the Getis-Ord Gi* statistic (Ord and Getis 1995; De Smith et al. 2007), which enables one to identify statistically significant spatial clusters of both high cell values ("hotspots") and low cell values ("cold spots") in the heat map. A key advantage of the Gi* statistic is that it allows testing the results for statistical significance using easily calculated z-scores (Burt et al. 2009). In order to identify hotspots in the viewshed heat map we applied the Gi* statistic to the heat map, calculated the corresponding z-scores, and used them to generate four classes according to the following p-value thresholds: 90% significant (z-score ≥1.645), 95% significant (z-score ≥1.960), 99% significant (z-score ≥2.576), 99.9% significant (z-score ≥3.291). All non-significant cells were grouped in a fifth class. It should be noted that by overlaying two or more significant level heat maps, it is possible to generate a heat map of significance level ranges. For example, overlaying the 95% significance heat map on top of the 90% significance level would result in three types of pixels, namely pixels below 90%, between 90% and 95%, and above 95%.

4 Case Study: The Waldo Canyon Wildfire

In order to showcase the utility of our approach in a real-world crisis setting, we applied it to the 2012 Waldo Canyon wildfire. For this purpose we collected both Twitter and Flickr data, as discussed in Section 3.1, and applied the proposed analysis framework in order to delineate the impact area of the fire using our cross-sourced triangulation approach. To demonstrate the benefit of using cross-sourced social media in the triangulation process we applied three modes of the analysis:

- **Mode 1**: the impact area was estimated as the overlap of all viewsheds that were generated from all Flickr contribution locations *without* calculating a reference point or evaluating the AOV for each image. Accordingly, in this mode, we use only Flickr data, without constraining the viewshed analysis with any AOV information.
- **Mode 2**: the impact area was estimated by using the centroid of the locations of all Flickr contributions as the reference point for the AOV calculation, followed by a viewshed analysis of each image. Accordingly, in this mode we use only Flickr data, ignoring any toponym information from Twitter.
- **Mode 3**: the impact area was estimated by using the toponym reference, as derived from Twitter, as the reference point for the AOV calculation, followed by a viewshed analysis of each image. Accordingly, in this mode we use Twitter content to orient Flickr data and guide the viewshed analysis.

The results of analysis modes 1, 2, and 3 are shown in Figures 7, 8, and 9, respectively. In each of these figures we present for each significance level range (as discussed in Section 3.2.3) the resulting impact area of the wildfire as an overlaid raster heat-map as well as the known wildfire impact area as provided by the US National Oceanic and Atmospheric Administration (NOAA 2013) following the event. In order to estimate the accuracy of the analysis results with respect to the known wildfire impact area and examine the benefit of using social multimedia for our analysis approach we calculated a confusion matrix for each analysis mode. For this purpose every heat map cell p at location (x,y) was labeled as belonging to one of four

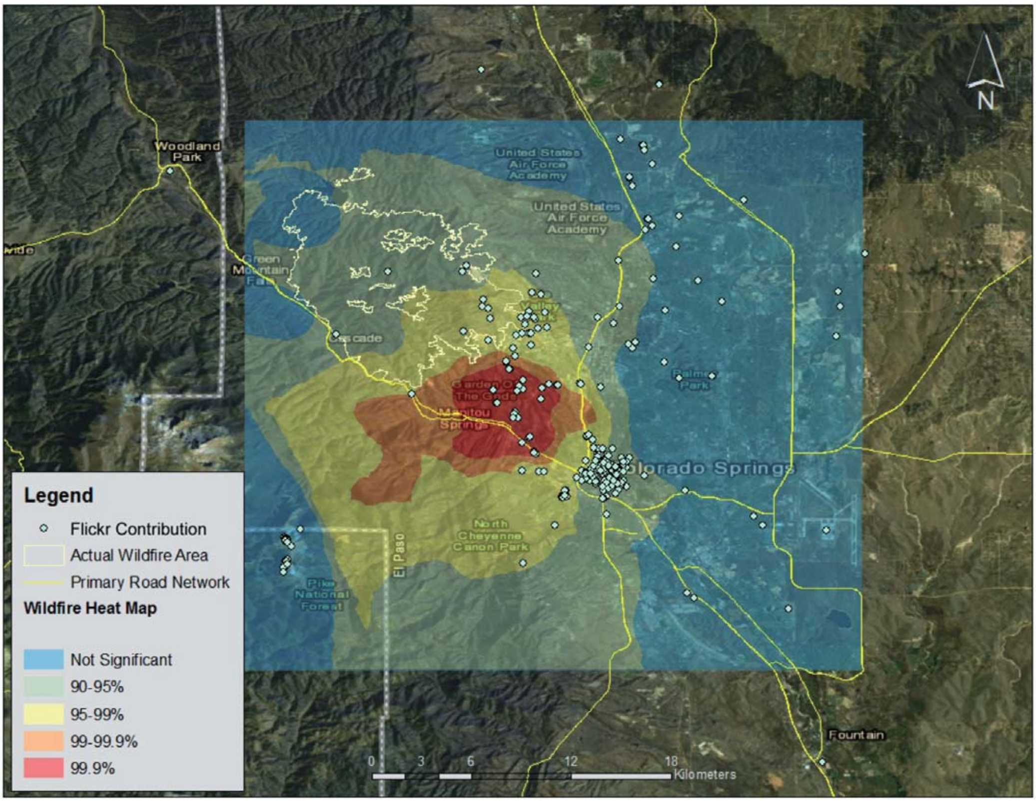

Figure 7 Wildfire location assessment as derived by analysis mode 1

classes according to a statistical significance condition and a spatial condition as shown in Table 1.

In this table, C is the user-defined confidence levels, $S()$ is the significance level operator which assigns a significance level for a given cell p, and A_{fire} is the known wildfire impact area. Based on these results we then calculated the rate of each class (e.g. True Positive rate) by dividing the area covered by each class by the corresponding reference area. For example, the TP rate was calculated as the ratio between the total area that was detected as fire through our analysis method and the known wildfire area. The results of this labeling process for analysis modes 1, 2, and 3 are summarized in Tables 2, 3, and 4, respectively.

Comparing the accuracy analysis results of the three analysis modes highlights the benefit of using our cross-source viewshed analysis: while the TP rate is only 29% at 95% confidence level when no reference point is used (analysis mode 1), this rate increases to 61% when a reference point is derived using only Flickr data (analysis mode 2), and reaches 75% when both Twitter and Flickr data are used (analysis mode 3). It is worth noting that in our accuracy analysis the TN rate for all three analysis modes remained approximately the same (between 79% and 87%), the false detection rates were reduced from 79% for FP and 90% for FN in analysis mode 1 to 68% and 25% for FP and FN, respectively in analysis mode 3. These results demonstrate that by combining crowdsourced data from both Flickr and Twitter in the viewshed analysis we are able to improve not only the detection accuracy but also the false detection accuracy. A three-dimensional rendering of the results obtained from the analysis in

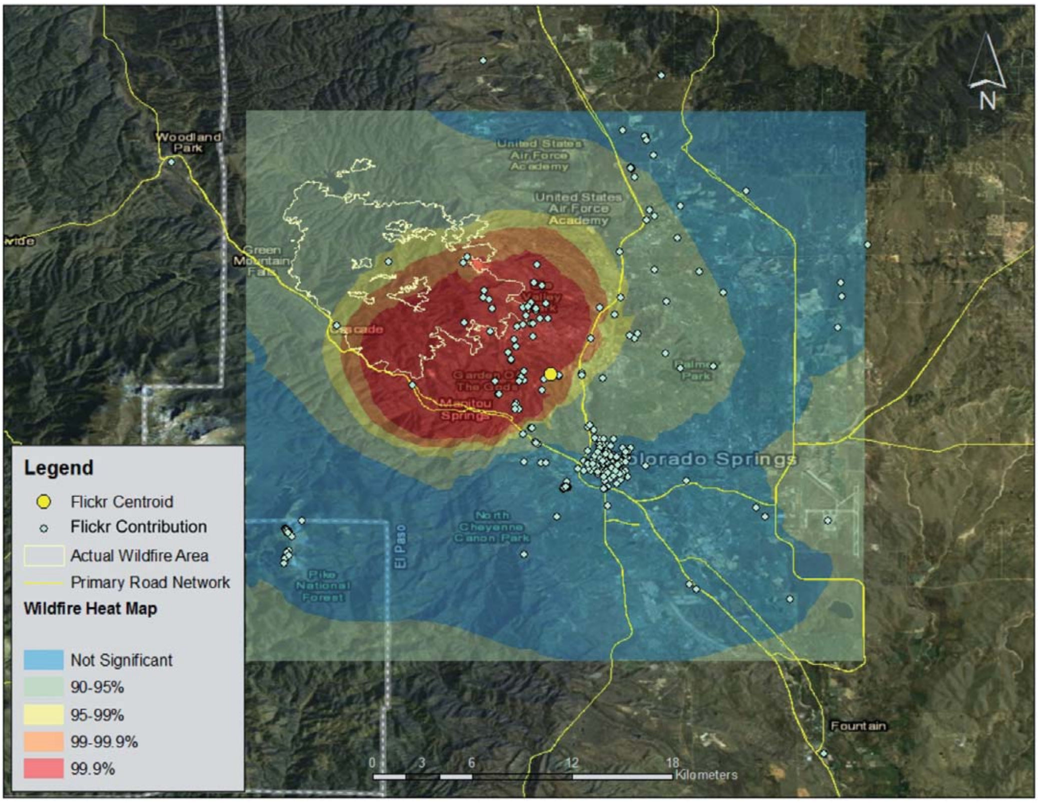

Figure 8 Wildfire location assessment as derived by analysis mode 2

Table 1 Heat map cell labeling according to significance and spatial conditions

Class	Confidence Level Condition	Spatial Condition
True Positive (TP)	$S(p) \geq C$	$p \in A_{fire}$
True Negative (TN)	$S(p) < C$	$p \notin A_{fire}$
False Positive (FP)	$S(p) \geq C$	$p \notin A_{fire}$
False Negative (FP)	$S(p) < C$	$p \in A_{fire}$

mode 3 is given in Figure 10, showing the topography of the case study area as well as the known extent of the wildfire. As can be seen, the pattern of contributions is driven by the terrain conditions and population distribution: while we have numerous contributions from the semi-urban area southeast of the wildfire, we have practically no contributions from the mountainous rural areas to the northwest of the event.

5 Conclusions and Outlook

The analysis of social media content to extract geospatial information and event knowledge from such crowd-contributed data has become the subject of substantial research activities. In

Table 2 Accuracy assessment for the results of analysis mode 1

	Class			
Confidence Level	TP	TN	FP	FN
90.0%	92%	45%	88%	63%
95.0%	29%	79%	90%	71%
99.0%	0%	93%	100%	100%
99.9%	0%	97%	100%	100%

Table 3 Accuracy assessment for the results of analysis mode 2

	Class			
Confidence Level	TP	TN	FP	FN
90.0%	100%	52%	88%	39%
95.0%	61%	87%	71%	39%
99.0%	49%	90%	70%	51%
99.9%	35%	93%	70%	65%

Table 4 Accuracy assessment for the results of analysis mode 3

	Class			
Confidence Level	TP	TN	FP	FN
90.0%	92%	70%	80%	18%
95.0%	75%	87%	68%	25%
99.0%	69%	89%	66%	31%
99.9%	49%	92%	65%	51%

this article we presented an approach that makes use of the multimedia nature of social media content by examining the benefits of the complementary use of heterogeneous sources of social multimedia feeds in order to assess the impact of a natural disaster. More specifically, we introduced a novel social multimedia triangulation process that uses collaboratively Twitter and Flickr content in a two-step integrated process. In this approach, we practically crowdsource approximate orientations from Twitter content and use this information to orient accordingly Flickr imagery and identify the impact area through viewshed analysis and viewpoint integration.

To demonstrate how our approach leverages multimedia content from social media in order to locate events, we used as a representative case study a natural disaster event,

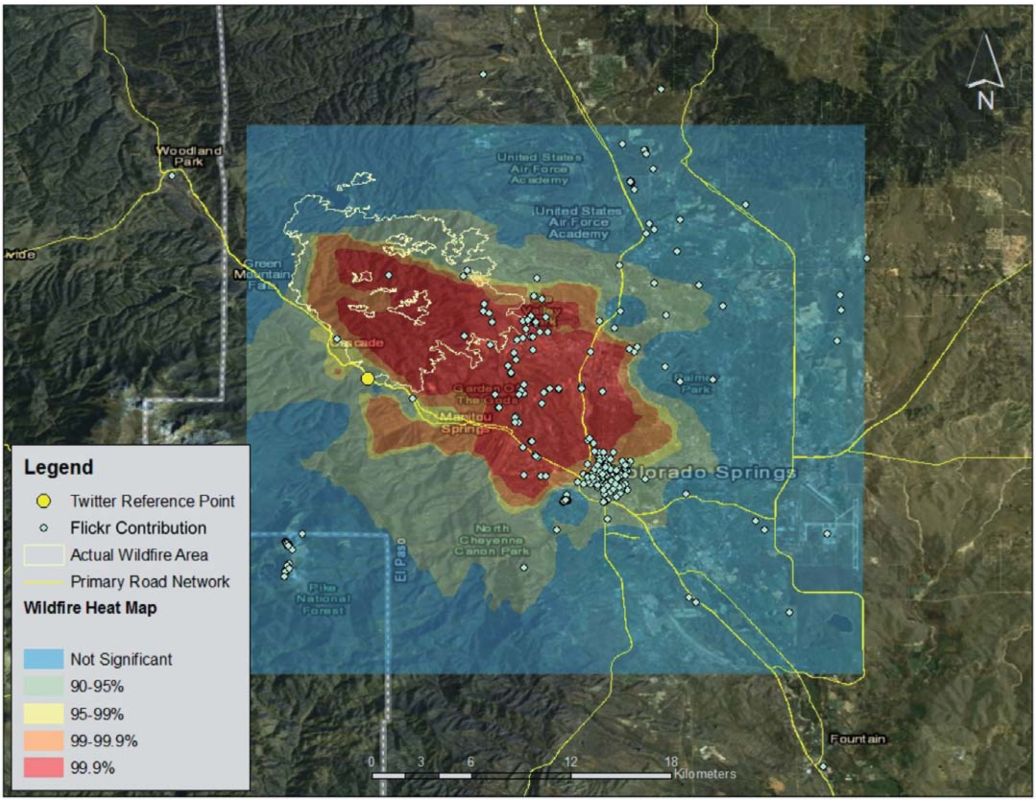

Figure 9 Wildfire location assessment as derived by analysis mode 3

namely the 2012 wildfire of Waldo Canyon in Colorado Springs. For our study we collected relevant Twitter data from the Twitter API using the keyword "fire", resulting in a corpus of 97,866 tweets (of which 5.57% were precisely geolocated) referring to the fire. We also collected 427 geolocated images contributed to Flickr during the event with "wildfire" as a tag. Combined, these datasets comprise multimedia crowd contributions communicating the event, and complement each other with respect to their thematic content.

Our objective was to pursue an innovative solution that harnesses these diverse crowd contributions in order to delineate the impact area of this particular event. The two-step approach that we introduced here proceeded by first using Twitter content to identify toponym references associated with a disaster. This information was then used to provide approximate orientation for the associated Flickr imagery, allowing us to delineate the impact area as the overlap of multiple view footprints. This is a two-step crowdsourcing process that crosses platforms and media in order to delineate an event: we use the text in Twitter to crowdsource a compass, in the form of a reference viewpoint, and then use this information to aggregate the views of another crowdsourced dataset, namely Flickr imagery. In essence, this extends the scope of VGI, in that crowdsourced content is not limited to the datasets, but also extends to harvesting information that is critical for the analysis of these datasets. This approach allows us to bypass computationally intensive image analysis tasks associated with traditional image orientation (e.g. the

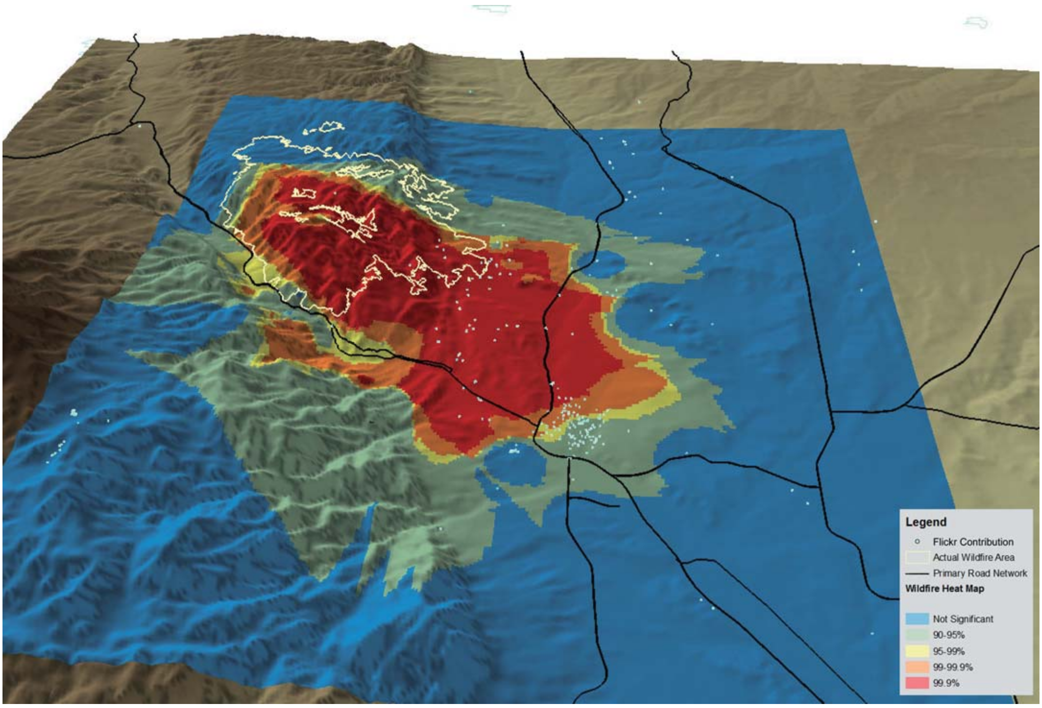

Figure 10 A three-dimensional perspective of wildfire location assessment as derived by analysis mode 3

identification of conjugate features), yet supports the aggregation of multiple image views in order to delineate the impact area as the aggregate of multiple views. The results of our analysis show the improvements in delineating the impact area through the introduction of such information.

As we are moving towards a wider adoption of crowdsourced content, we have to continue being aware that such content is the outcome of a geosocial process: the level of participation, and the patterns of contributions are driven by the particularities of the corresponding physical and social environments. In our particular case, contributions were primarily made from south and southeast areas, not only due to the presence of urban areas in them, but also due to accessibility issues and the nature of the event itself. Having had a more broad distribution of contributions around the impact area would have resulted in further improvements. However, even for such adverse conditions as the ones we encountered in this case study we showed that at a confidence level of 95% we can raise the true positive (TP) rate to 75% when we use our two-step triangulation process, in contrast to a TP rate of only 29% when no reference point was extracted from Twitter. This supports the argument that by harvesting various types of information from diverse crowdsourced content we can better infer event-specific information from these citizen contributions. Attempting to consider this problem from the point of view of a decision maker, it may be considered advantageous to have a high TP level, even at the cost for a relatively high (but manageable) FP level, as this will likely raise awareness and preparedness levels among the potentially affected population. Furthermore, FP levels in these crisis situations carry a particular value of their own, as they can be viewed as capturing the public's perception of the potential threat from their respective locations.

One thing to consider in conjunction with the abovementioned levels of accuracy is that they would be affected by the granularity of the reference point. For example, if people were referring to the "Colorado wildfires", our approach would not be able to generate meaningful results. Generally, one would reasonably expect a link between the granularity level of event references, as they emerge through public discourse, and the type of the corresponding event. While some events have a rather localized footprint (e.g. wildfires), others have a broader impact (e.g. hurricanes). This can be viewed as an extension of the problem of geo-parsing text at global- and local-scales (Leidner and Lieberman 2011). Furthermore, the dynamicity of an event may impact the analysis: a fire is a very dynamic event, but was (in this case) still spatially contained. If it were to be spreading across large areas our analysis would have to be segmented across temporal intervals, within which the event would be mapped at distinct instances, and its evolution tracked accordingly. Presumably, this could also lead to the emergence of sequences of toponyms for the same event.

It is worth noting that rather than focusing on fine-tuning the accuracy of the outcome of the analysis, our main objective in this article was to demonstrate the feasibility of our approach in the context of a rapid assessment of the impact area of an event, given non-curated data corpus such as the one presented here. As we have shown above, even with certain approximations, e.g. using average camera model values for images without Exif information, we are able to assess the impact area quite well. Such approximations could be further improved and refined by using techniques for estimating missing camera parameters, e.g. Bujnak et al. (2010) or De Oliveira Costa et al. (2014). Similarly, the viewshed analysis can be refined to account for the combined effects of the accuracy of the DEM, e.g. Oksanen and Sarjakoski (2006), as well as the accuracy of the technique used to calculate it, e.g. Fisher (1993).

Nevertheless, we need to remain aware that the particular nature of social media contributions may result in biases in their patterns of contribution. For example, Li et al. (2013) focused on social media usage in Twitter and Flickr, finding a relationship between Twitter usage and well-educated high-income people, particularly white and Asian populations. More relevant to this work, Kent and Capello (2013) studied the use of social media during a crisis situation (a wildfire). Their analysis showed that demographic characteristics of the area impacted by the emergency situation could be used to reveal the propensity of its population to contribute information in social media during such a crisis. These works reveal some of the intrinsic nature of social media contributions as they relate to geospatial information, warranting the further study of such activities in order to gain a better understanding of the value and quality of this crowdsourced content.

In order to overcome the demographics-related limitations (and the resulting biases) it is possible to consider active social media approaches, whereby requests for contributions are issued for locations that are underrepresented in the harvested data. This nevertheless would not address the limitations of population gaps, where low population density results in lack of data (thus limiting the accuracy of the analysis). Towards that end, one could consider the integration of social media feeds with traditional geosensor networks, in order to collect from the latter focused information in response to the breaking events that are detected in the former. While such integration still remains largely unexplored, it clearly emerges as a promising future direction due to the substantial advancements in social media harvesting and processing.

References

Agichtein E, Castillo C, Donato D, Gionis A, and Mishne G 2008 Finding high quality content in social media. In *Proceedings of the International Conference on Web Search and Web Data Mining*, Stanford, California: 183–94

Ames M and Naaman M 2007 Why we tag: Motivations for annotation in mobile and online media. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, San Jose, California: 971–80

Boyd D, Golder S, and Lotan G 2010 Tweet, tweet, retweet: conversational aspects of retweeting on Twitter. In *Proceedings of the Forty-third IEEE Hawaii International Conference on System Sciences*, Kauai, Hawaii: 1–10

Bujnak M, Kukelova Z, and Pajdla T 2010 Robust focal length estimation by voting in multi-view scene reconstruction. In Zha H, Taniguchi R, and Maybank S (eds) *Computer Vision: ACCV 2009*. Berlin, Springer: 13–24

Burt J, Barber G, and Rigby R 2009 *Elementary Statistics for Geographers* (Third edition). New York, Guilford

Christensen C 2011 Twitter revolutions? Addressing social media and dissent. *Communication Review* 14: 155–57

Croitoru A, Crooks A T, Radzikowski J, and Stefanidis A 2013 GeoSocial gauge: A system prototype for knowledge discovery from geosocial media. *International Journal of Geographical Information Science* 27: 2483–508

Croitoru A, Crooks A T, Radzikowski J, Stefanidis A, Vatsavai R R, and Wayant N 2014 Geoinformatics and social media: A new big data challenge. In Karimi H A (ed) *Big Data Techniques and Technologies in Geoinformatics*. Boca Raton, FL, CRC Press: 207–32

Crooks A T, Croitoru A, Stefanidis A, and Radzikowski J 2013 #Earthquake: Twitter as a distributed sensor system. *Transactions in GIS* 17: 124–47

De Choudhury M, Feldman M, Amer-Yahia S, Golbandi N, Lempel R, and Yu C 2010 Constructing travel itineraries from tagged geo-temporal breadcrumbs. In *Proceedings of the Nineteenth International Conference on World Wide Web*, Raleigh, North Carolina: 1083–84

De Floriani L and Magillo P 2003 Algorithms for visibility computation on terrains: A survey. *Environment and Planning B* 30: 709–28

De Longueville B, Smith R S, and Luraschi G 2009 OMG, from here, I can see the flames!: A use case of mining location-based social networks to acquire spatiotemporal data on forest fires. In *Proceedings of the International Workshop on Location Based Social Networks*, Seattle, Washington: 73–80

De Oliveira Costa F, Silva E, Eckmann M, Scheirer W J, and Rocha A 2014 Open set source camera attribution and device linking. *Pattern Recognition Letters* 39: 92–101

De Smith M J, Goodchild M F, and Longley P A 2007 *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools* (Second edition), The Winchelsea Press, Winchelsea, UK

Fisher P F 1993 Algorithm and implementation uncertainty in viewshed analysis. *International Journal of Geographical Information Systems* 7: 331–47

Fisher P F 1995 An exploration of probable viewsheds in landscape planning. *Environment and Planning B* 22: 527–46

Fontugne R, Cho K, Won Y, and Fukuda K 2011 Disasters seen through Flickr cameras. In *Proceedings of the Special Workshop on Internet and Disasters*, Tokyo, Japan

Forbes 2012 Twitter's Dick Costolo: Twitter Mobile Ad Revenue Beats Desktop on Some Days. WWW document, http://onforb.es/KgTWYP

Friedland G and Sommer R 2010 Cybercasing the joint: On the privacy implications of geotagging. In *Proceedings of the Fifth USENIX Workshop on Hot Topics in Security (HotSec 10)*, Washington, DC

Fritz S, MacCallum I, Schill C, Perger C, Grillmayer R, Achard F, Kraxner F, and Obersteiner M 2009 Geo-Wiki.Org: The use of crowdsourcing to improve global land cover. *Remote Sensing* 1: 345–54

Fuchs G, Andrienko N, Andrienko G, Bothe S, and Stange H 2013 Tracing the German centennial flood in the stream of tweets: First lessons learned. In *Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, Orlando, Florida: 31–38

Gao H, Barbier G, and Goolsby R 2011 Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems* 26(3): 10–14

Goodchild M F 2007 Citizens as sensors: The world of volunteered geography. *GeoJournal* 69: 211–21

Griffith D A 1987 *Spatial Autocorrelation: A Primer*. Washington, DC, Association of American Geographers

Hagar C and Haythornthwaite C 2005 Crisis, farming and community. *Journal of Community Informatics* 1: 41–52

Haklay M, Singleton A, and Parker C 2008 Web Mapping 2.0: The neogeography of the GeoWeb. *Geography Compass* 2: 2011–39
Hudson-Smith A, Crooks A T, Gibin M, Milton R, and Batty M 2009 Neogeography and Web 2.0: Concepts, tools and applications. *Journal of Location Based Services* 3: 118–45
Hurst M, Siegler M, and Glance N S 2007 On estimating the geographic distribution of social media. In *Proceedings AAAI International Conference on Weblogs and Social Media*, Boulder, Colorado
Instagram 2014, Available at http://instagram.com/press/ [Accessed 3rd April 2014].
Kaplan A M and Haenlein M 2010 Users of the world unite!: The challenges and opportunities of social media. *Business Horizons* 53: 59–68
Kent J D and Capello H T 2013 Spatial patterns and demographic indicators of effective social media content during the Horsethief Canyon fire of 2012. *Cartography and Geographic Information Science* 40(2): 78–89
Kisilevich S, Krstajic M, Keim D, Andrienko N, and Andrienko G 2010 Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. In *Proceedings of the Fourteenth International Conference of Information Visualisation*, London, UK: 289–96
Kuo P F, Zeng X, and Lord D 2012 Guidelines for choosing hot-spot analysis tools based on data characteristics, network restrictions, and time distributions. In *Proceedings of the Ninety-first Annual Meeting of the Transportation Research Board*, Washington, DC: 22–26
Leidner J L and Lieberman M D 2011 Detecting geographical references in the form of place names and associated spatial natural language. *SIGSPATIAL Special* 3(2): 5–11
Li L and Goodchild M F 2012 Constructing places from spatial footprints. In *Proceedings of the First ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, Redondo Beach, California: 15–21
Li L, Goodchild M F, and Xu B 2013 Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science* 40: 61–77
Lieberman M D and Samet H 2011 Multifaceted toponym recognition for streaming news. In *Proceedings of the Thirty-fourth International ACM SIGIR Conference on Research and Development in Information Retrieval*, Beijing, China: 843–52
Liu S B and Palen L 2010 The new cartographers: Crisis map mashups and the emergence of neogeographic practice. *Cartography and Geographic Information Science* 37: 69–90
Liu S B, Palen L, Sutton J, Hughes A L, and Vieweg S 2008 In search of the bigger picture: The emergent role of online photo sharing in times of disaster. In *Proceedings of the Information Systems for Crisis Response and Management Conference*, Washington, DC
MacEachren A M, Jaiswal A, Robinson A C, Pezanowski S, Savelyev A, Mitra P, and Blanford J 2011 Senseplace2: Geotwitter analytics support for situational awareness. In *Proceedings of the IEEE Conference in Visual Analytics Science and Technology (VAST)*, Providence, Rhode Island: 181–90
Mahmud J, Nichols J, and Drews C 2012 Where is this tweet from?: Inferring home locations of Twitter users. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, Dublin, Ireland
McGhee T 2013 4,167 Colorado Wildfires Caused Record Losses of $538 Million in 2012. *Denver Post* (1/17) (available at http://www.denverpost.com/ci_22396611/4–167-colorado-wildfires-caused-record-losses-538)
Nagy G 1994 Terrain visibility. *Computers and Graphics* 18: 763–73
Nitta N, Kumihashi Y, Kato T, and Babaguchi N 2014 Real-world event detection using Flickr images. In Gurrin C, Hopfgartner F, Hurst W, Johansen H, Lee H, and O'Connor N (eds) *Multimedia Modeling*. New York, Springer: 307–14
NOAA 2013 National Weather Service Weather Forecast Office (Waldo Canyon Wildfire dss). WWW document, http://www.crh.noaa.gov/pub/?n=waldocanyonwildfiredss
Norheim-Hagtun I and Meier P 2010 Crowdsourcing for crisis mapping in Haiti. *Innovations: Technology, Governance* 5(4): 81–9
Oksanen J and Sarjakoski T 2006 Uncovering the statistical and spatial characteristics of fine toposcale DEM error. *International Journal of Geographical Information Science* 20: 345–69
Ord J K and Getis A 1995 Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27: 286–306
Palen L, Vieweg S, Sutton J, Liu S B, and Hughes A L 2007 Crisis informatics: Studying crisis in a networked world. In *Proceedings of the Third International Conference on E-Social Science*, Ann Arbor, Michigan
Pohl D, Bouchachia A, and Hellwagner H 2012 Automatic sub-event detection in emergency management using social media. In *Proceedings of the Twenty-first International Conference Companion on World Wide Web*, Lyon, France: 683–86
Rouse L J, Bergeron S J, and Harris T M 2007 Participating in the Geospatial Web: Collaborative mapping, social networks and participatory GIS. In Scharl A and Tochtermann K (eds) *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. London, Springer: 153–58

Sakaki T, Okazaki M, and Matsuo Y 2010 Earthquake shakes Twitter users: Real-time event detection by social sensors. In *Proceedings of the Nineteenth International Conference on World Wide Web*, Raleigh, North Carolina: 851–60

Sander H A and Manson S M 2007 Heights and locations of artificial structures in viewshed calculation: How close is close enough? *Landscape and Urban Planning* 82: 257–70

Sapiro G 2011 Images everywhere: looking for models: technical perspective. *Communications of the ACM* 54(5): 108–18

Senaratne H, Bröring A, and Schreck T 2013 Using reverse viewshed analysis to assess the location correctness of visually generated VGI. *Transactions in GIS* 17: 369–86

Shi X, Tseng B, and Adamic L A 2007 Looking at the blogosphere topology through different lenses. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2007)*, Boulder, Colorado

Signorini A, Segre A M, and Polgreen P M 2011 The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS One* 6(5): e19467

Silva M J, Martins B, Chaves M, Afonso A P, and Cardoso N 2006 Adding geographic scopes to Web resources. *Computers, Environment and Urban Systems* 30: 378–99

Silverman B W 1986 *Density Estimation for Statistics and Data Analysis*. Boca Raton, FL, CRC Press

Singla A and Weber I 2011 Camera brand congruence and camera model propagation in the Flickr social graph. *ACM Transactions on the Web* 5(4): 20

Starbird K and Palen L 2011 Voluntweeters: Self-organizing by digital volunteers in times of crisis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, British Columbia: 1071–80

Stefanidis A, Cotnoir A, Croitoru A, Crooks A T, Radzikowski J, and Rice M 2013a Statehood 2.0: Mapping nation-states and their satellite communities through social media content. *Cartography and Geographic Information Science* 40: 116–29

Stefanidis T, Crooks A T, and Radzikowski J 2013b Harvesting ambient geospatial information from social media feeds. *GeoJournal* 78: 319–38

Sugumaran R and Voss J 2012 Real-time spatio-temporal analysis of West Nile virus using Twitter data. In *Proceedings of the Third International Conference on Computing for Geospatial Research and Applications*, Washington, DC

Sui D and Goodchild M F 2011 The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science* 25: 1737–48

Triglav-Čekada M and Radovan D 2013 Using volunteered geographical information to map the November 2012 floods in Slovenia. *Natural Hazards and Earth System Science* 13(11): 2753–62

Twitter 2011 200 Million Tweets per Day. WWW document, http://bit.ly/laY1Jx

Utz S, Schultz F, and Glocka S 2013 Crisis communication online: How medium, crisis type and emotions affected public reactions in the Fukushima Daiichi nuclear disaster. *Public Relations Review* 39: 40–6

Valli C and Hannay P 2010 Geotagging where cyberspace comes to your place. In *Proceedings of the 2010 International Conference on Security and Management*, Las Vegas, Nevada: 627–32

Vieweg S, Hughes A L, Starbird K, and Palen L 2010 Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. In *Proceedings of the Twenty-eighth International Conference on Human Factors in Computing Systems*, Atlanta, Georgia: 1079–88

Wayant N, Crooks A T, Stefanidis A, Croitoru A, Radzikowski J, Stahl J, and Shine J 2012 Spatiotemporal clustering of social media feeds for activity summarization. In *Proceedings of the Seventh International Conference for Geographical Information Science*, Columbus, Ohios

Wood J, Dykes J, Slingsby A, and Clarke K 2007 Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE Transactions on Visualization and Computer Graphics* 13: 1176–83

Wueller D and Fageth R 2008 Statistic analysis of millions of digital photos. In *Proceedings of the International Society for Optics and Photonics on Electronic Imaging*, San Jose, California

Xie Z and Yan J 2008 Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems* 32: 396–406

Yang Q, Snyder J, and Tobler W 1999 *Map Projection Transformation: Principles and Applications*. Boca Raton, FL, CRC Press

YouTube 2013 YouTube Pressroom Statistics. WWW document, http://bit.ly/gzYBVx

Zook M, Graham M, Shelton T, and Gorman S 2010 Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical and Health Policy* 2: 2

