Mining social media to task satellite data collection during emergencies

Guido Cervone

Combining satellite imagery with social media data enables a good match between predictions and observations in emergency situations.

The era of Internet and social media has drastically changed the way individuals all over the world interact and communicate. In particular, the current digital world has changed the dynamics of how we collect and share data across all age groups and socioeconomic statuses. Although social media is not primarily intended for scientific studies, it can be an invaluable source for distributed and timely measurements. This is especially true during disasters and emergencies, where first responders and emergency managers rely on multiple sources of data inputs to make decisions.[1–3]

We have introduced a new methodology that uses social media during disasters or emergencies to prioritize the collection of satellite remote-sensing imagery and to 'fill the gaps' in the collected imagery. We used social media platforms, such as Twitter and Flickr, along with imagery data from satellites, airplanes, and unmanned aerial vehicles (UAVs) to assess the extent of damage on transportation infrastructure during the 2013 Colorado floods, which affected the city of Boulder and the surrounding regions.

Using Twitter's application programming interface, we developed an application that scans tweets and identifies unique messages containing predefined keywords that describe potential catastrophic events and resultant damage to facilities (for example, flood, landfall, and road damage). We geolocated those tweets by using embedded geolocation information (found in roughly 2% of the tweets), mining the text for geographical references of interest that we matched to a gazetteer, or by analyzing data shared through a tweet that contained geolocation information (such as a picture). When a number of tweets clustered in space and time, we analyzed them to determine whether they related to the same event. We usually did this by analyzing
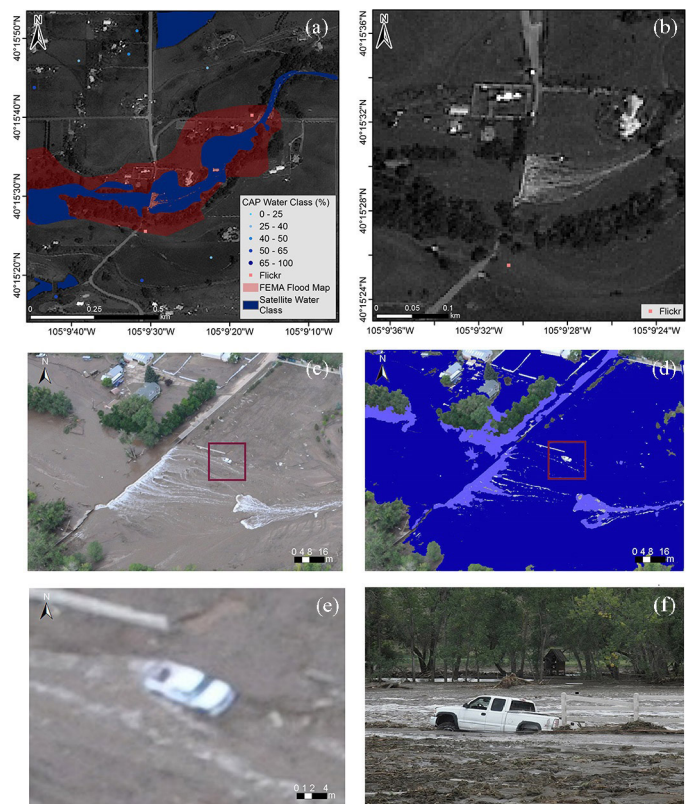

Figure 1. *(a) Map showing WorldView2 satellite band 3 for an area northeast of Longmont, Colorado, which suffered severe flood damage. (b) Close-up view of a submerged road. (c) Civil Air Patrol (CAP) image showing the same submerged road (the square highlights a stranded truck). (d) The same CAP image classified for water pixels. (e) A CAP image close-up view of the stranded truck. (f) The same truck in a Flickr post.*

the hashtags contained in the tweet messages when they were posted. If there was a consensus of common hashtags, we collected additional tweet messages, using both specific geographical locations and common hashtags as query items.

When the initial flooding occurred in Colorado, the volume of geolocated and non-geolocated tweets mentioning floods in the Boulder area increased significantly, creating hotspots. Specifically, we concentrated on tweets that mentioned both flood and transportation damage. We used these areas to task the WorldView2 commercial satellite to collect high-resolution images of areas in Colorado that had been identified by social media, and which potentially contained damage to transportation infrastructure.

We integrated the collected satellite imagery and social media data with additional non-traditional data sources to assess damage to transportation infrastructure.[4, 5] We applied a spatial kernel interpolation to help identify road conditions in downtown Boulder and surrounding regions. The goal was to use multi-sensor data fusion analysis to generate a map that highlighted areas most likely to have been affected by the flood. Using geographic information system techniques, we overlaid areas identified as being more likely to have suffered damage, with available road network layers. The intersection between the roads and damage areas gave an early assessment of where roads were likely to be affected and potentially impassable.

Results show that this methodology is sound both in cases when high-resolution remote sensing images are available, and when data is incomplete due to cloud cover or satellite revisiting time. For example, in the city of Boulder, Twitter messages containing images helped augment WorldView2 and Landsat observations and increased the prediction of road damage. We compared the outcomes using our methodology with a list of officially closed roads, and found a good match between predictions and observations. We would not have been able to establish which roads were closed with such accuracy if we had used satellite data alone.

In the city of Longmont, Colorado, WoldView2 data contained large amounts of cloud cover in the areas where most damage was recorded. Once the satellite image was combined with tweets and photographs, we were able to identify multiple areas as flooded and several roads as impassible. Figure 1 shows such a case, where road closure was predicted based on a combination of satellite and aerial imagery, ground photos, and Twitter messages. Figure 1(a) shows the WorldView2 band 3 as background, along with the Federal Emergency Management Agency flood map (available after the fact for verification purposes). Figure 1(b) is a close-up view of a road that is totally submerged. The cluster of white pixels to the right of the water

flowing over and past the road is likely to be a stranded truck. Figure 1(c) shows an aerial image where it is possible to discern the submerged road and a stalled truck, highlighted in a square. Figure 1(d) shows the same image automatically classified by machine learning classifiers. Figure 1(e) shows a close-up image of the area with the stranded truck, and Figure 1(f) shows a picture taken on the ground and posted on Flickr of the same partially submerged truck.

To summarize, our combined social media/satellite data methodology is particularly well suited for use in unplanned and unpredicted emergencies, where the location, time, and nature of an event are unknown. We have designed this method for satellite data collection, but in future we will seek to develop it further for the effective collection of aerial imagery from airplanes, unmanned aerial vehicles, and/or drones.

Author Information

Guido Cervone
Pennsylvania State University
University Park, PA

Guido Cervone is associate professor of geoinformatics in the Department of Geography and Institute for CyberScience. His research focuses on the development and application of computational algorithms for the analysis of spatiotemporal remote sensing, numerical modeling, and social media big data related to environmental hazards and renewable energy.

References
1. A. Acar and Y. Muraki, *Twitter for crisis communication: lessons learned from Japan's tsunami disaster*, **Int'l J. Web Based Communit. 7**, pp. 392–402, 2011.
2. P. Earle, D. Bowden, and M. Guy, *Twitter earthquake detection: earthquake monitoring in a social world*, **Ann. Geophys. 54**, pp. 708–715, 2012.
3. Q. Huang and Y. Xiao, *Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery*, **ISPRS Int'l J. GeoInf. 4**, pp. 1549–1568, 2015.
4. J. Fohringer, D. Dransch, H. Kreibich, and K. Schröter, *Social media as an information source for rapid flood inundation mapping*, **Nat. Haz. Earth. Syst. Sci. 15**, pp. 2725–2738, 2015.
5. E. Schnebele and G. Cervone, *Improving remote sensing flood assessment using volunteered geographical data*, **Nat. Haz. Earth Syst. Sci. 13**, pp. 669–677, 2013.