GroundTruth: Bringing Together Experts and Crowds for Image Geolocation

Rachel Kohler, John Purviance, and Kurt Luther
Dept. of Computer Science and Center for Human-Computer Interaction
Virginia Tech, Blacksburg, VA USA
{rkohler1, ffdd4846, kluther}@vt.edu

Abstract

Geolocation, the process of identifying the specific location where a photo or video was taken, is an important task in verifying evidence for investigations in journalism, national security, human rights, and other domains. However, experts typically perform geolocation work as a time-consuming, manual process. This paper introduces GroundTruth, a web-based system that leverages the powerful vision system of crowd workers to support experts in image geolocation tasks. We describe the technical contributions of GroundTruth and present preliminary results from an evaluation with expert geolocators and novice crowds.

Introduction and Related Work

Verification of visual evidence, such as photos or videos, is a fundamental step in investigations across many domains, including journalism, national security, and human rights (Barot, 2014). As social media becomes increasingly popular, investigators increasingly encounter visual evidence online. If these photos or videos can be verified, they may provide key evidence in investigations or lead to important witnesses or sources (Brandtzaeg, Lüders, Spangenberg, Rath-Wiggins, & Følstad, 2016).

For visual material, geolocation is often a key component in the verification process. *Image geolocation* or (*geolocalization*) is the task of identifying the precise geographic location where a photo or video was taken. An independent verification of the image's location helps significantly in evaluating other claims about the image.

For many expert investigators, image geolocation is largely a manual process (Higgins, 2014, 2015; Kohler & Luther, 2017). Experts may begin by inspecting the image's context for hints about its location, such as the person or organization sharing the image, text or other content surrounding the image, and metadata, such as a timestamp or geotag. Next, the expert may examine clues in the image content itself, such as road signs, business names, logos, vehicles, architecture, and clothing, that are suggestive of certain regions. If the above is not sufficient to geolocate the image, the expert may then attempt a systematic search of the location using satellite imagery from Google Earth or other online sources. Experts often begin this brute-force approach by drawing an aerial diagram of the image that abstracts distinctive landmarks to make comparison to the satellite imagery easier.

Given the arduous and time-sensitive nature of these manual efforts, researchers have explored a variety of technological supports for image geolocation. A growing body of research in the field of computer vision seeks to identify photo and video locations using scene recognition, convolutional neural networks, and other automated approaches (e.g. Lin, Cui, Belongie, & Hays, 2015; Weyand, Kostrikov, & Philbin, 2016; Zhai, Bessinger, Workman, & Jacobs, 2016). This work has shown great promise, especially in specific contexts, but high-quality, generalized results are not yet available, and these tools have not been widely adopted by expert investigators.

Another thread of research seeks to leverage the impressive capabilities of the human vision system via crowdsourcing to support geolocation. (Kohler, Purviance, & Luther, 2017) conducted several crowdsourcing experiments and found that novice crowds using an expert geolocator's aerial diagram could reduce a satellite imagery search area by 50% in 10 minutes, whereas showing workers just the ground photo resulted in poor performance.

In this paper, we build on the above approach, but investigate how a software system could allow an expert to collaborate with crowds on a geolocation task. We present GroundTruth, a web-based system that enables experts to designate a search area for a crowd to investigate in parallel, and then review the aggregated results for a potential location match. We also present preliminary results from an evaluation of GroundTruth with expert geolocators and crowd workers recruited from Amazon Mechanical Turk.

System Design and Scenario

Step 1: Expert Launches a Crowd Investigation

When an expert logs into GroundTruth, the system asks her to provide two images. The first is the ground-level photo to be geolocated (videos are not yet supported). The second is a diagram to support their own search efforts as well as the crowd's. The expert can create the diagram with whatever tools or medium she prefers—pen and paper, Photoshop, Powerpoint, etc.—but the result must be an image file or PDF. When providing the diagram, she also estimates the width of the entire drawing in feet or meters.

Next, the expert is presented with an interface with two columns. In the left column, the ground photo appears at the top, and her diagram appears at the bottom, along with rotation controls. The right column is dominated by a large embedded Google Map, with zoom controls and the Map / Satellite mode toggle (set to Satellite mode by default). Based on the context provided with the ground photo, if any, the expert can pan and zoom the map to the appropriate general area (e.g., a city-level view of Los Angeles).

The expert is then presented with a short tutorial that instructs her to identify a promising area of the map for the crowd to help her investigate. She clicks and drags on the map to draw a rectangle that covers the area of interest. As she draws, the interface shows her how many workers and/or how much money would be required for the crowd to search that area, allowing her to adjust to her budget.

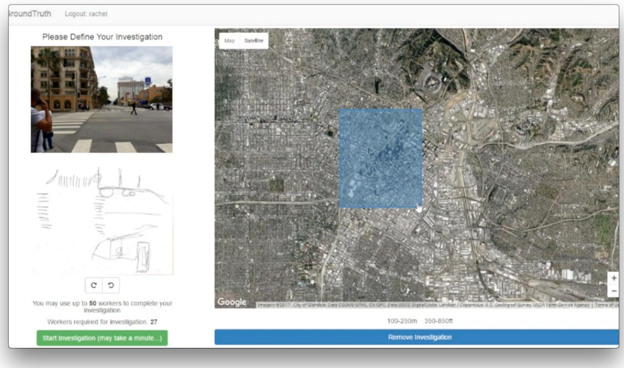
Once satisfied, the expert submits and starts the investigation. The system automatically divides up the search area into a grid of *regions* overlaid on the map. A region is a 4×4 grid of 16 equal-sized subregions. The subregion width is equal to the diagram width provided by the expert, to enable easier visual comparison. Three unique crowd workers analyze each region.

Step 2: Crowd Workers Analyze Imagery

Crowd workers on Amazon Mechanical Turk accept a HIT asking them to sign an IRB consent form, and are then directed to GroundTruth's crowd interface. The left side of the interface shows the expert's diagram. The diagram was randomly rotated, and the worker could rotate it clockwise or counterclockwise by clicking arrow buttons underneath it. It also shows a small Google Map (in Map mode) of the region with the 16-subregion grid overlaid in black lines.

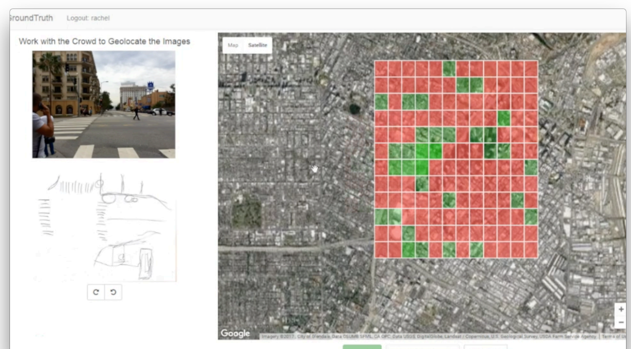
The right side of the interface shows a Google Map (in Satellite mode) of the region, divided by a translucent grid of white lines into 16 equal-sized subregions. The worker can zoom in and out, and toggle Map/Satellite mode, but was confined to that subregion. The worker clicks a green Yes / Maybe button if it looks like a potential match, or a red No button if it does not, and then clicks Next. This advances the worker to the next subregion, and marks in either red or green the corresponding subregion in the small map. The system advances through the subregions in a Creeping Line search pattern, following best practices used in search and rescue (Wollan, 2004).

The decision to have workers evaluate more than one subregion per micro-task may seem surprising. However, we found in pilot studies that workers who saw only one subregion tended to perform poorly due to lack of context. The problem was exacerbated when distinctive features were cropped or located in corners. After experimenting with different-sized regions, we ultimately settled on a 4×4 grid as striking an effective balance of context and effort.



Step 1. Expert launches investigation.

Step 2. Crowd workers analyze imagery.



Step 3. Expert verifies aggregated crowd results.

The top of the interface shows remaining task time and a button to launch the tutorial. The bottom provided a text-box for participants to provide feedback on the task, and a Finish button.

Step 3: Expert Verifies Aggregated Crowd Results

As workers submit judgements, the system aggregates the results and displays them to the expert in real time. Each region is colored red (0 yes judgements) or a shade of green (1–3 yes judgements), indicating priority. The expert can inspect these or other regions of the grid. If the given subregion is not a match, she can click the "Exclude" button, changing its color to black and helping her track where she has already been. The expert can also toggle the visibility of the grid and colors. If she finds a match, she can click "Found It!" to end the investigation.

Implementation Details

We built GroundTruth as a web-based system using a Django/Python framework, a PostgreSQL database, and the Google Maps API for satellite imagery and GIS functions. We hosted the site with Heroku cloud services.

Preliminary Evaluation

Study Design

We recruited four expert image geolocators to gather initial feedback about GroundTruth and how they would interact with the crowd. Two participants were journalists, one was a satellite image analyst, and one was a private investigator. All had at least two years of experience in image geolocation. We compensated them $50 each for their time.

Experts performed the study online while sharing their screen with us via Skype. We asked them to geolocate three ground photos during a single session lasting 30-40 minutes. We provided the photos (but not locations) ahead of time. Experts drew diagrams for each image, and we used those diagrams to gather crowd results. We gave experts a budget of 50 workers and they searched an average area of 10 mi^2 per task. Participants used a think-aloud protocol to externalize their thought process during the task, and we recorded their voice and screen activities. We also interviewed them briefly about their reflections at the end of the study.

Results and Discussion

All four participants were enthusiastic about GroundTruth and having geolocation support from the crowd. One journalist was excited to have crowd support for geolocation, given the hectic nature of his news agency, when he is frequently multi-tasking:

> *In looking at a video and verifying it, I'm looking at what media reports are saying about it, I'm talking to people on the ground, and I'm also sort of typing up a report, and looking at updates as I'm sort of writing my own quick article or video verification about the actual incident. If, while I was doing that I was able to say, these are the coordinates for the town at a search radius of 1500 meters and these are the things were looking for, I know probably I would have been able to do it in maybe 60% of the time that it would normally take.*

The experts also made use of the crowd results to direct their search strategy. The satellite image analyst talked about his process of using the crowd feedback as an overview, and then following up: "*Basically, what I'm doing is obviously I look for where are the darkest greens, zoom in there, and then turn [the color] off.*" The private investigator had a similar strategy:

> *I definitely paid more attention to the spots that were green, and once I would hit a patch of red, I would see, yeah, [the crowd is] probably right, and then I would zoom out to see where another patch of green was and move over there and start my search again.*

The second journalist put the most trust in the crowd results: "Places not to look—the red squares—have definitely been the most useful because it significantly narrows down the places that I am looking… I didn't spend any time looking in the areas that were red." Notably, she performed the best of all the experts, geolocating all three photos within the time limit.

Conclusion and Next Steps

Image geolocation is an arduous, yet important task in many types of investigations. This paper presents the GroundTruth system, which enables novice crowds to support expert investigators in the often manual and tedious geolocation process. Our preliminary evaluation suggests that investigators are enthusiastic about this type of crowd support, but further studies are needed to understand the complex dynamics between experts and crowds.

In future work, we plan to quantitatively and qualitatively evaluate the performance of expert geolocators with and without GroundTruth's crowdsourced support. We are also exploring how expert tasks like specifying the search area and drawing the diagram might be crowdsourced. Future work may also consider how experts and crowds could augment computer vision-based geolocation systems.

Acknowledgements

We thank our experts and crowds, anonymous reviewers, and members of the Crowd Intelligence Lab. This research was funded in part by NSF awards 1527453 and 1651969.

References

Barot, T. (2014). Verifying Images. In *Verification Handbook: A Definitive Guide to Verifying Digital Content for Emergency Coverage*. Retrieved from http://verificationhandbook.com/book/chapter4.php

Brandtzaeg, P. B., Lüders, M., Spangenberg, J., Rath-Wiggins, L., & Følstad, A. (2016). Emerging Journalistic Verification Practices Concerning Social Media. *Journalism Practice*, *10*(3), 323–342. https://doi.org/10.1080/17512786.2015.1020331

Higgins, E. (2014, July 9). A Beginner's Guide to Geolocating Videos. Retrieved May 4, 2017, from https://www.bellingcat.com/resources/how-tos/2014/07/09/a-beginners-guide-to-geolocation/

Higgins, E. (2015, July 25). Searching the Earth: Essential Geolocation Tools for Verification. Retrieved May 4, 2017, from https://www.bellingcat.com/resources/how-tos/2015/07/25/searching-the-earth-essential-geolocation-tools-for-verification/

Kohler, R., & Luther, K. (2017). Crowdsourced Image Geolocation as Collective Intelligence. In *Collective Intelligence 2017*. New York, NY, USA.

Kohler, R., Purviance, J., & Luther, K. (2017). Supporting Image Geolocation with Diagramming and Crowdsourcing. In *Proceedings of AAAI HCOMP 2017*. Quebec City, Canada.

Lin, T.-Y., Cui, Y., Belongie, S., & Hays, J. (2015). Learning Deep Representations for Ground-to-Aerial Geolocalization. In *Computer Vision and Pattern Recognition (CVPR)*. Boston, MA. Retrieved from http://vision.cornell.edu/se3/wp-content/uploads/2015/04/geocvpr15.pdf

Weyand, T., Kostrikov, I., & Philbin, J. (2016). PlaNet - Photo Geolocation with Convolutional Neural Networks. *ArXiv:1602.05314 [Cs]*. Retrieved from http://arxiv.org/abs/1602.05314

Wollan, H. (2004). *Incorporating Heuristically Generated Search Patterns in Search and Rescue* (Master's Thesis). University of Edinburgh. Retrieved from https://pdfs.semanticscholar.org/a20f/5518053ab18c78dfcae58f53a8702c84b380.pdf

Zhai, M., Bessinger, Z., Workman, S., & Jacobs, N. (2016). Predicting Ground-Level Scene Layout from Aerial Imagery. *ArXiv:1612.02709 [Cs]*. Retrieved from http://arxiv.org/abs/1612.02709