Crowdsourced Image Geolocation as Collective Intelligence

RACHEL KOHLER, KURT LUTHER, Dept. of Computer Science, Virginia Tech[1]

1. INTRODUCTION

As internet access grows more commonplace around the world, an increasing share of the population is exposed to a massive and exponentially growing quantity of online information and news. Journalists are among the most critical professional gatekeepers in helping societies sort through this barrage of data to separate fact from fiction and truth from "fake news."

This paper focuses on a subset of this gatekeeping work, the practice of *image verification*, in which journalists and other investigators seek to research and confirm or refute the claims of image or video data shared with the public, often online via social media, with little or no context (Barot, 2014). For example, witnesses may upload photos of a terror attack or natural disaster, or governments may release videos of military actions. Image verification experts investigate this evidence to identify sources or witnesses to interview, to raise awareness of influential events, to defuse controversy around false information, or to hold governments accountable, among many reasons.

We focus on one type of image verification, *image geolocation*, in which investigators work to identify the geographic location where a photo or video was made, as accurately and specifically as possible (Higgins, 2014). Image geolocation has been used to find terrorist training camps and to dispute official stories of government military intervention, among other purposes. A growing body of research seeks to use computer vision techniques to automatically geolocate images (Weyand, Kostrikov, & Philbin, 2016), but little attention has been paid to the online communities of professional and amateur journalists who perform manual image geolocation using entirely public or "open source" information, often with superior results.

We report on in-depth interviews with experienced image geolocators, focusing on their motivations, practices, and the role of online collaboration, social media, and technology in their work. This work has implications for the theory and design of collective intelligence systems that support sensemaking and image analysis (Bigham et al., 2010; Hahn, Chang, Kim, & Kittur, 2016; Hara et al., 2015).

2. METHOD

For this ongoing study, we conducted in-depth, semi-structured interviews with five journalists who had expertise in image verification, especially photos and videos shared on social media. Participants represented a breadth of experiences in journalism. Some worked at major news bureaus in the US and Europe; others worked as freelancers while completing graduate degrees or working as corporate security experts or private investigators. The interviews were audio-recorded and fully transcribed. Our questions focused on topics such as participants' motivations, process, use of tools, and collaboration. We analyzed the data by iteratively coding participant responses from the transcripts within each of the major topics we asked about. Participant names and some occupational details are anonymized.

[1] This research was funded in part by NSF awards 1527453 and 1651969.

3. RESULTS

3.1 Motivation

Participants mentioned a variety of motivations for geolocating images, and many expressed multiple overlapping motivations. P5 emphasized the societal impact his work could have on ongoing conflicts, and a moral obligation to uncover the truth, in particular if civilians are claimed to be involved. He said, "I think [it] is very important especially when it's about civilian deaths to check claims, to dig deep into such claims."

Other participants mentioned the fun or thrill of trying to solve a mystery or puzzle. P2 said, "It's a bit of an investigative chase, you know, and there's a little bit of an adrenaline rush when you geolocate something." P5 agreed, "I'm very passionate about it. I think it's a lot of fun, actually, to try to figure out the exact location." P3 pointed out that these elements can attract participants who, in contrast to participants like P5, aren't compelled by the subject matter per se. "People always have fun even when it's something they're totally not interested in, you know, some war or environmental issue or crime or whatever. It's fun to gather all this evidence and piece it together and try to come to a conclusion, especially if you're the first one to do it."

3.2 Speed

Participants agreed that the speed or timeframe for investigations can vary widely. P5 described examples of in-depth investigations that were less time-sensitive, saying, "When it's an investigation, like a big one, it can really take up days and days and days, like many, many hours." P3 explained, "You do more obscure kind[s] of geolocations where you would go back and look for changes over time, because maybe the newest photo you have is from 2010. So you look [at] what are the changes to the landscape over six years, new construction, check the elevation in the background, are there any mountains, and so on."

However, most participants, P5 included, emphasized the importance of speedy results in the majority of cases. For breaking news in particular, time is of the essence. P3 said that journalists and officials "want to very quickly verify eyewitness accounts and photos and videos that emerge and make sure they're real and not fake… So you have to very quickly figure out the location, any important landmarks, and then cross-reference them." Furthermore, rapidly changing situations mean that locations may only be temporarily relevant, e.g., if subjects are moving quickly. "If you can't figure it out in fifteen, twenty minutes," P3 said, "then you're probably already too late and more information's already come out by then."

Besides the dynamic nature of the situation itself, competition creates other incentives to geolocate quickly. P1 emphasized his conflicting goals of speed and accuracy with respect to competition. "I'm a journalist first and foremost in this corporation. I'm not a forensic scientist. So yes, I'm interested in getting the story right, but I'm particularly interested in getting it right for [my news agency] and not [competing agencies]."

Several participants pointed to the value of social media, especially Twitter, for soliciting very fast responses to geolocation-related queries. P4 said, "If you post something on Twitter asking for Twitter communities to help, it's crazy how fast that sort of crowdsourcing can get a result." P3 shared a technique that brought a new perspective to P1's tension between speed and accuracy: "The best way to get any response on Twitter is not to post something that's cool or useful, but to post something and misidentify a type of weapon or a type of vehicle. You'll immediately get fifty replies trying to correct you because people like to be right and show off their knowledge about their obscure things they know about."

3.3 Process

We saw strong consistency in the process participants described using to perform image geolocation. Typical early steps included an initial scan to identify potential clues and an iterative process of narrowing down geographic regions into smaller, more tractable areas. P5 detailed his approach:

> I start looking for clues in the footage and I see a landmark, you know, like a minaret or a mosque or a very interesting or specific building, the shape of a building. I start looking for clues in the media that could give away something about the location. When it's an airstrike video, you see them from the top so if it's mountainous, I already know. When I'm looking at a certain country, let's say Syria, then I know in what kind of region I should start looking. If it's desert, I also know, okay, it's at least not at the coast, you know. So you start to narrow down the field of search, the area which I search, smaller and smaller.

From this point, participants described a systematic effort to compare the target image to reference image sources, including aerial and satellite sources like Google Earth and Digital Globe, as well as human-scale photographs from Google Street View and Panaramio. In P4's words, "If you do have somewhere that you think matches up, you just refer back and forth from the satellite imagery to the picture and try to make sure everything lines up, or you can make sense of everything."

Participants talked at length about the important cognitive skill of mapping ground-level photos to aerial or satellite reference imagery, to enable comparisons. Some would perform the re-orientation entirely in their heads. Other, like P5, created physical representations with pen and paper: "Let's say if it's a video that was taken from the ground, and somebody's walking and filming the surrounding[s], I just pause the video—sometimes [each] second—and just draw a bird's eye perspective, or a satellite image perspective, of how I think it may look like from the air. So I can then compare it with satellite imagery just to get a better impression, to basically transform the 3-D surrounding somebody's walking in into a more 2-D overview of the area."

Participants emphasized that this skill was difficult at first, but learnable with practice. P3 explained, "The real thing people have trouble with is orienting themselves in their minds to be able to match things to satellite imagery that they see on the ground. On the ground, maybe something seems like it's really, really far away, or really, really close, but on a satellite map your mind doesn't perceive things the same way zoomed out rather than zoomed in." Consequently, he teaches workshops to help journalists improve this skill. P4 agreed that practice makes a big difference. "I definitely do think it's like a muscle that you work out so you can just better associate things that you see in pictures with satellite imagery. Perspective distortion can throw off a novice or a beginner really easily because things that you see from the air tend not to look how you would think they would from the ground. But with enough experience, you start to figure that out and know what to look for."

3.4 Collaboration

Participants described collaboration as critical to their success, and gave examples at multiple scales, ranging from small groups to large online communities. Some participants formed small collectives of investigators to perform collaborative image geolocation. P2, P3, and P4 mentioned Slack as an invaluable communication tool for this purpose. P4 elaborated on some of the benefits of collaboration in geolocation tasks:

> Anyone can try to do an investigation solo and plenty of people have. I have. But it's always great to have another set of eyes, another perspective, someone that might point something out that you missed. Just bouncing ideas off of people and being able to—I mean, not only just making sure you get it right, but it's also a lot quicker if multiple people are scouring the web with the same goal in mind.

P3, who also works in a collective, emphasized the benefits of specialization and complementary expertise. He speaks multiple languages, while another of his group is a graphic designer with strong knowledge of photo manipulation, and a third member is an engineer with expertise in manipulating datasets.

Beyond their collectives, participants also mentioned the benefits of seeking help from broader online communities, such as social media. For example, P4 said that Twitter is host to both a niche community of geolocation experts, as well as a much larger community focused on military verification, such as identifying weapons and aircraft.

Building on the idea of specialization, many participants spoke to the benefits of social media and crowdsourcing for accessing pockets of deep local knowledge. P5 acknowledged that, despite being very well-traveled, many places were still unfamiliar to him. However, he noted, "If you have a large outreach on Twitter, for example, then you may get very good hints from people that actually live there and recognize places." P4 provided one such example, when attempting to identify an Islamic State sympathizer simply by a photo of his backyard in Germany. "It would have been like a brute force, just going through everyone's backyard looking for certain shapes. But just normal people who knew the neighborhood were able to chime in and be like, 'Oh yeah, that's down the street.'" P3 pointed out that many world locations lack Google Street View photos, or the latest satellite imagery is outdated, so crowdsourced local knowledge is the only viable option in these cases.

4. CONCLUSIONS AND FUTURE WORK

Our ongoing interview study revealed rich details about the attitudes, practices, and challenges surrounding manual image geolocation. Participants demonstrated remarkable creativity, resourcefulness, and persistence in identifying critical information to confirm or deny claims associated with image and video data with little or no context. They also developed unique skills, such as spatial and cognitive abilities to change the perspective of a photographed region from 2D to 3D and vice-versa. Participants also engaged heavily in online collaboration, both in small collectives to share information, trade specialized expertise, and give feedback; and through engagement with larger online communities and social media to solicit targeted local knowledge. We also found that participants succeeded even with off-the-shelf, general-purpose tools such as Twitter, Slack, Google Earth and Google Street View, and even pen and paper. These results shed light on an important form of collective intelligence and crowdsourced sensemaking that has seen little research attention.

In our future work, we are conducting more interviews, and exploring how specialized tools could further support the image geolocation practices identified in this paper. We are developing GroundTruth, an online platform that allows investigators to delegate some of their image geolocation tasks to crowd workers, who are responsible for searching a small section of the grid and relaying their findings back to the investigator. We are experimenting with giving crowds in GroundTruth different representations, e.g. ground-level photos versus bird's-eye diagrams of varying detail, to overcome the difficulties of perspective shifting described above.

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

Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., … Yeh, T. (2010). VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology* (pp. 333–342). New York, NY, USA: ACM. https://doi.org/10.1145/1866029.1866080

Hahn, N., Chang, J., Kim, J. E., & Kittur, A. (2016). The Knowledge Accelerator: Big Picture Thinking in Small Pieces. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2258–2270). New York, NY, USA: ACM. https://doi.org/10.1145/2858036.2858364

Hara, K., Azenkot, S., Campbell, M., Bennett, C. L., Le, V., Pannella, S., … Froehlich, J. E. (2015). Improving Public Transit Accessibility for Blind Riders by Crowdsourcing Bus Stop Landmark Locations with Google Street View: An Extended Analysis. *ACM Trans. Access. Comput.*, *6*(2), 5:1–5:23. https://doi.org/10.1145/2717513

Higgins, E. (2014, December 19). Geolocating Tunisian Jihadists in Raqqa. Retrieved August 3, 2016, from https://www.bellingcat.com/resources/case-studies/2014/12/19/geolocating-tunisian-jihadists-in-raqqa/

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