Does Flight Path Context Matter? Impact on Worker Performance in Crowdsourced Aerial Imagery Analysis

Sofia Eleni Spatharioti
Northeastern University
spatharioti.s@husky.neu.edu

Sara Wylie
Northeastern University
s.wylie@northeastern.edu

Seth Cooper
Northeastern University
scooper@ccs.neu.edu

ABSTRACT

Natural disasters result in billions of dollars in damages annually and communities left struggling with the difficult task of response and recovery. To this end, small private aircraft and drones have been deployed to gather images along flight paths over the affected areas, for analyzing aerial photography through crowdsourcing. However, due to the volume of raw data, the context and order of these images is often lost when reaching workers. In this work, we explored the effect of contextualizing a labeling task on Amazon Mechanical Turk, by serving workers images in the order they were collected on the flight and showing them the location of the current image on a map. We did not find a negative impact from the loss of contextual information, and found map context had a negative impact on worker performance. This may indicate that ordering images based on other criteria may be more effective.

Keywords

crowdsourcing; Amazon Mechanical Turk; context

INTRODUCTION

In the aftermath of a natural disaster, swift and efficient coverage of the area for accurate status reports is critical, given the small time window that response and recovery efforts must organize and effectively operate. One popular and cost effective approach which is often deployed is using aerial photography (e.g. drones or small piloted aircraft) for post-disaster image collection. These images must then be analyzed in order to identify locations in need of immediate assistance, using visualization techniques such as cartography and heatmaps. Crowdsourcing has proven an efficient approach to aerial imagery analysis and map generation in scenarios where automated image analysis can be difficult, with systems targeted towards the crowd deployed in numerous disaster events around the world in recent years (Liu 2014; Munro et al. 2013; Barrington et al. 2012).

When dealing with aerial images taken using unmanned aerial vehicles, such as drones, control over captured content may be minimal. Meaningful content can be spread among numerous images, creating a potentially useful context for image analysis. However, in crowdsourcing settings, datasets are often presented randomly to participants, to ensure adequate coverage of the area and an even distribution of labels. This results in a loss of context, as participants enter the process through a series of unrelated images, having no way of situating themselves on the area in view, causing useful information to dissipate. This may lead to user frustration and loss of label quality, in situations where context is necessary to accurately categorize images.

In this work we explored providing two specific types of context to participants: first, varying the *order* in which they encountered images, and second, varying the presence of *map context* during the task. Showing images in their original sequence offered visual overlap to participants, creating a context continuity. We also considered map context, by visualizing the path of the images on a map, along with the current image location, with previously

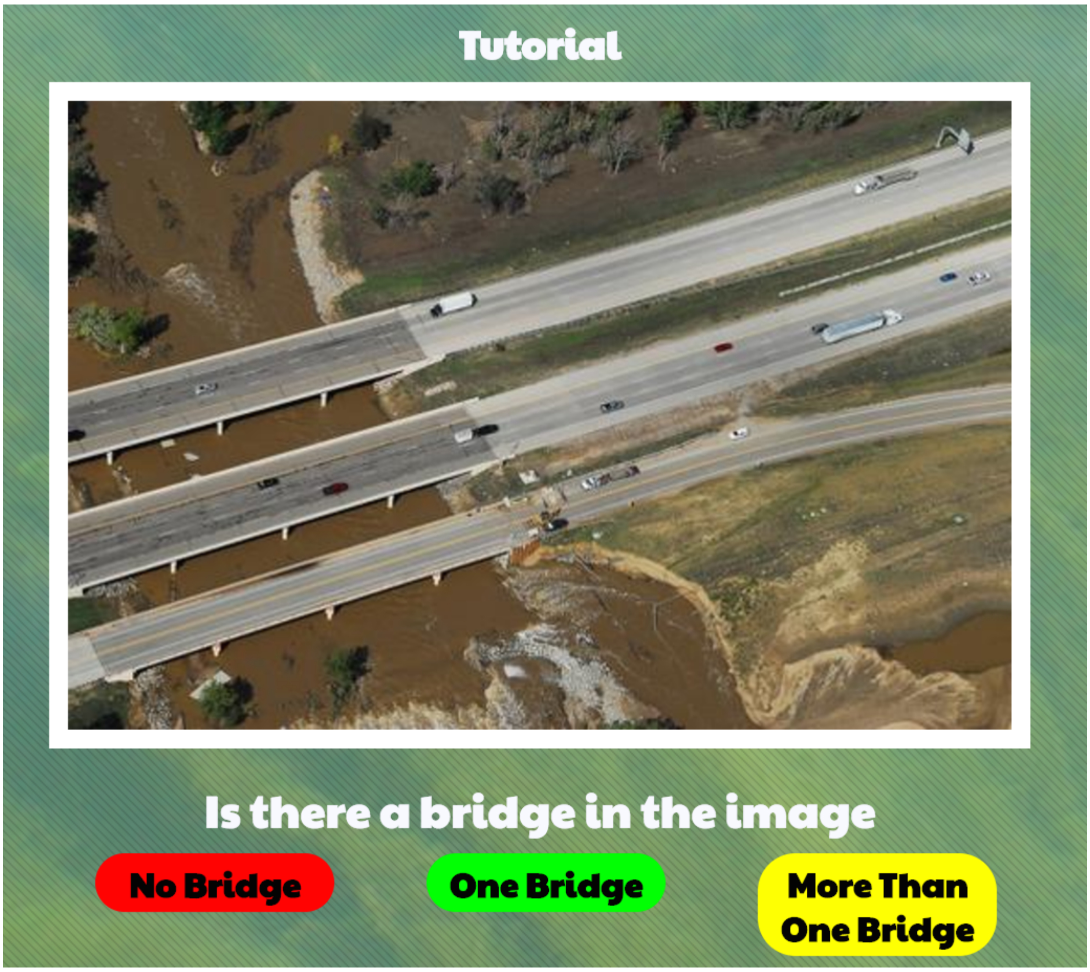

Figure 1. The training page of the task, showing the core interface of the labeling task, with an image, a question and a set of possible labels for workers to choose. The image shows an example for "More Than One Bridge".

submitted labels annotated accordingly, creating a progress context. We were interested in whether the loss of context would have a negative impact on performance in a crowdsourcing setting. To this end, we posted a Human Intelligence Task (HIT) on Amazon Mechanical Turk, asking workers to label a series of images taken from a disaster scenario involving floods in the State of Colorado.

We found that removing context from the task interface did not negatively impact worker performance, including output rate, number of labels provided, and label quality (in terms of accuracy with ground truth). Showing images at random was as effective as preserving order and context, which indicates that other types of ordering, such as decision-theoretic means (Dai et al. 2010) may be better suited when designing a task for disaster response. Further, we found that including a map showing image locations and progress had a negative impact on worker completion of the task. This work contributes an empirical study of how designing a task to include context may impact behavior and performance in crowdsourced aerial imagery analysis, with a focus on disaster response applications.

RELATED WORK

Crowdsourcing has long been used as a means of analyzing significantly sized image datasets for label and characteristics extraction, across multiple domains. In the domain of image labeling for web search, the ESP Game was a multiplayer game designed to allow participants to succeed in labeling an image by submitting the same label at the same time, using required agreement to improve label quality (von Ahn and Dabbish 2004). Cropland Capture was also developed as a crowdsourcing game to monitor cropland from satellite imagery (Sturn et al. 2015). However, Sturn et al. observed a significant drop in classifications in the weeks following press releases about the game.

Galaxy Zoo is viewed as a successful example of crowdsourcing classifications, in the domain of astronomy, attracting more than 200,000 volunteers for classifying images of galaxies from the Sloan Digital Sky Survey

(Raddick et al. 2010). Crowdsourcing image analysis is also used in "space archaeology", with projects such as Global Xplorer, a platform for analyzing satellite images (Global Xplorer 2016), as well as retinal photography (Mitry et al. 2013).

Crowds have often been called upon to aid during severe weather and natural disaster events, by providing information or assisting in analyzing existing data collected through various means. During the Santa Barbara wildfires that took place between 2007 and 2009, volunteers assisted in analyzing and submitting geographic information with the goal of creating a map of the area, using tools like Google Maps, to maintain up to date status reports of the event (Goodchild and Glennon 2010).

After the earthquake that struck Haiti in 2010, a short number code was set up where people could communicate their needs via SMS text messaging. These text messages where in turn translated, geo-coded and categorized by volunteers using a tool developed by Ushahidi, to allow better organization for humanitarian response (Liu 2014).

Satellite imagery, among other types of inputs, was used in analyzing the effects of another earthquake event that affected New Zealand in 2011, using a crowdsourcing tool developed by Tomnod (Barrington et al. 2012). Volunteers were asked to identify affected buildings and rate the level of damage. In a project closely related to our work, MapMill was originally developed in response to the Deepwater Horizon Oil Spill (Warren 2010), but was then adapted and used for assessing damage from aerial photographs after Hurricane Sandy affected over eight countries in the Atlantic region in 2012 (Munro et al. 2013). AIDR (Artificial Intelligence for Disaster Response) was initially designed to classify Twitter posts created during disasters using crowdsourcing (Imran et al. 2014) and was later extended to support aerial data captured via unmanned aerial vehicles (UAVs) (Ofli et al. 2016).

A growing body of work has pointed out the negative effects of context loss when performing tasks on crowdsourcing platforms. Contextual interruptions, which are caused by switching between tasks of different types, were shown to cause a significant delay in performance (Lasecki, Marcus, et al. 2014), while the negative impact of long delays between tasks has also been highlighted (Lasecki, Rzeszotarski, et al. 2015). Another example of context enhancing performance can be found in conversations with the crowd in the Chorus and Chorus:View projects, as well as VizWiz, a mobile application developed to assist blind users in providing labels and answering questions using the crowd (Bigham et al. 2010; Lasecki, Thiha, et al. 2013; Lasecki, Wesley, et al. 2013). Allowing users to engage in conversations with the crowd offered more context on the task, which improved crowd response.

Context loss was also explored in Dai et al.'s work (2015), where inserting micro-diversions in long running tasks on Amazon Mechanical Turk was found to lead to increased worker retention without compromising quality. However, none of the types of micro-diversions were relevant to the main task and the main purpose was to offer relief to workers engaging in monotonous tasks.

Other examples of ordering crowdsourcing tasks to improve worker performance include Cai et al. (2016), where writing tasks were presented in different sets of orders to optimize worker behavior. However, the criteria for order were based on complexity and not content overlap, as is the case for our work. Moreover, none of the previously mentioned work on context involved disaster response applications. Specifically in the domain of crowdsourcing for disaster response, recent work has shown that switching between tasks of similar context but different types may also lead to increased worker performance (Spatharioti and Cooper 2017).

STUDY SETUP

The type of natural disaster that was investigated in this study concerned floods, as they occur due to various reasons in different geographical locations around the world all year round. We focused on the 2013 Colorado Floods that were caused due to heavy rain that affected areas from Colorado Springs, to Fort Collins. Participants were asked to assist in labeling aerial imagery captured by Civil Air Patrol, which is publicly available through the Hazards Data Distribution System (HDDS), provided by the U.S. Geological Survey (Hazards Data Distribution System Explorer 2016).

After the event, a sizable number of bridges were affected, with local news stating that around 120 structures in the state of Colorado were in need of repairs (Hendrick 2013). We were interested in assessing whether the crowd would be able to identify these structures through analyzing aerial photography. A first step towards this process was to sort through the available images to identify ones that contained bridges. To this end, we designed a simple labeling task where workers would view an image and answer the question "Is there a bridge in the image?", with possible options being "No Bridge", "One Bridge", and "More Than One Bridge". A screenshot of the training page with the core interface of the task can be seen in Figure 1.

The flow of the experiment was straightforward; participants would first have to consent to the purpose of the experiment, based on IRB guidelines. They would then go through a small training session of example images, to

Flight Path   Bar Ordered / Bar Random

Figure 2. Comparison of interfaces among conditions, showcasing the differences in progress feedback. The interface on the left was used for the *Flight Path* condition and the interface on the right for the *Bar Random* and *Bar Ordered* conditions. Workers in the *Bar Ordered* condition would be presented with the second interface, but the images would be in the order of the flight path, similar to the *Flight Path* order of images.

familiarize themselves with the question and the answers, after which they would proceed to the task. At the end of the task, a small survey was presented.

To generate context, we selected a subset of our image dataset than ran along a clear path geographically on the state of Colorado. These images were then ordered according to the path and selected in a way that created some visual overlap among images, similar to how a dataset of images captured from the air, via drones or small piloted aircraft, would appear. This resulted in an image set of 230 images. The density of images was a bit higher in the beginning, but contextual overlaps occurred consistently across the entire path, as we focused on eliminating gaps between points. We then considered the following three conditions, which can be viewed in Figure 2:

- **Flight Path**: Images were shown in the order of the path of the flight. Progress was shown as a map, with dots indicating the available images in the set and an airplane icon denoting the current location on the task. Dots of previously labeled images were colored according to the participant's vote, mirroring the answer color schema.
- **Bar Ordered**: Images were shown in the order of the path of the flight, but progress was shown in the form of a progress bar.
- **Bar Random**: Images were shown in random order. Progress was shown in the form of a progress bar indicating the percentage of available images completed.

An example illustrating overlap between images in the *Flight Path* and *Bar Ordered* conditions, in contrast to the lack of overlap in the *Bar Random* condition, is shown in Figure 3.

Participants were recruited using Amazon Mechanical Turk. We set up a Human Intelligence Task (HIT) on the crowdsourcing platform titled "Disaster Area Map Image Tasks". The description of the HIT informed workers that they would be performing image related tasks such as image labeling and answering a short survey. The HIT region was the United States. Although the HIT title suggested that workers would be working on images from a disaster event, the actual event was not revealed to workers. We chose a fixed payment scheme of 50 cents, regardless of the amount of images that were submitted and offered no further incentive through bonuses for accuracy or volume of contributions. Previous crowdsourcing work that has used fixed payments include the domains of games (Khajah et al. 2016) and writing tasks (Cai et al. 2016), as well as the domain of disaster response (Spathariotí, Govoni, et al. 2017). In addition, no lower bound on the number of images provided was present; workers could proceed at any time to the post-task survey. As evidenced in the work of Ho et al. (2015), workers carry different beliefs in what is considered acceptable performance in crowdsourcing tasks. In order to counteract fears of rejection, we explicitly mention in the instructions of the task that any submission will be accepted as long as the post-task survey

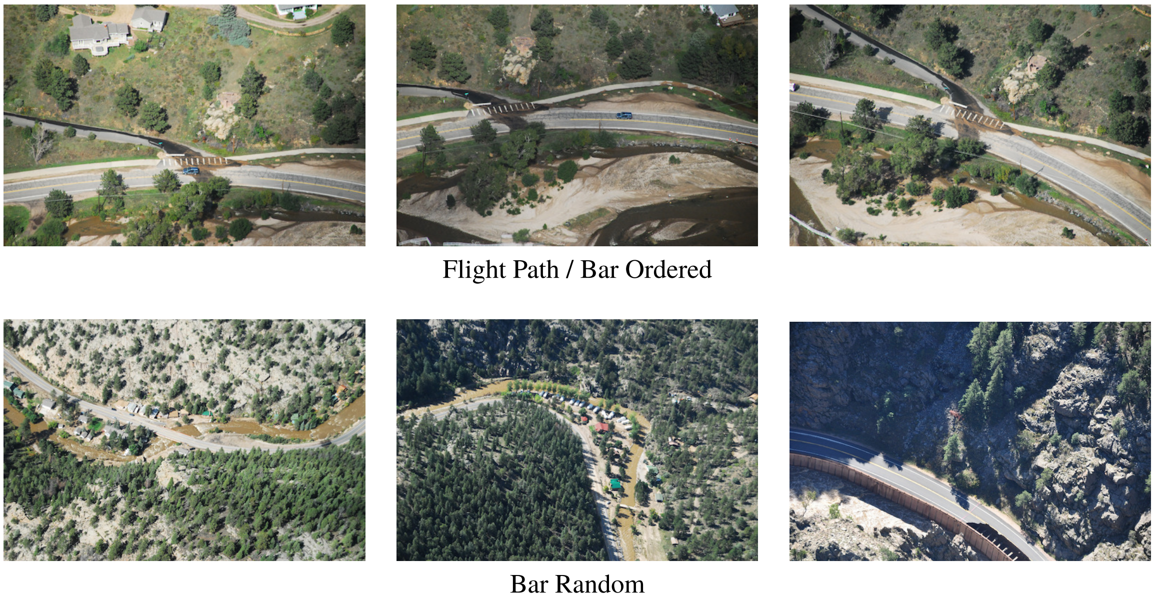

Figure 3. Example of a sequence of images in the task for the different conditions. In the conditions where aerial imagery is shown in order, a visual overlap of content is present.

is completed. The instructions were always available to workers for the entire duration of the image labeling part of the task.

Findings

We collected and analyzed responses from 909 Amazon Mechanical Turk workers, randomly assigned to one of the aforementioned conditions. The setup of the HIT guaranteed that no duplicate submissions were allowed. We also chose to exclude workers who had taken previous iterations of the task that were deemed too similar.

Worker actions on the interface were logged for the entire duration of the task. In order to examine worker behavior, we recorded the following variables for each worker:

- *N_Labels*: The total number of labels provided. Workers could provide one label for each image encountered, and no edits were allowed.
- *Time per Image*: The average amount of time spent labeling each image, in seconds.
- *Total Time*: The total time spent labeling images, in seconds. When analyzing Total Time and Time per Image, we identified and capped breaks over five minutes from our analysis.
- *Accuracy*: The percentage of labels provided agreeing with the ground truth. To determine ground truth, all images were cross-referenced by one of the project team members based on geo-location, extracted from EXIF tags, and content, with a list containing bridge structures monitored by the state of Colorado, provided by the Colorado Department of Transportation.
- *Set Completion*: A boolean indicating whether or not labels were provided for all available images in the set.

Workers were required to complete a survey at the end of the labeling task, in order to complete the HIT and get compensated. To gauge how the task was perceived by workers, we used the NASA Task Load Index (NASA-TLX) which is widely used to gain subjective system assessments in a variety of domains (Hart and Staveland 1988). In particular, we used the raw NASA-TLX, which consists of the six sub-scales of total workload, namely *Mental Demand*, *Physical Demand*, *Temporal Demand*, *Performance*, *Effort* and *Frustration*, separated in 5-point steps from 0 to 100. Workers were required to complete all six sub-scales, which were afterwards combined for the Task Load Index (TLX). We also provided small descriptions of each sub-scale, to facilitate comprehension of what was required. Workers could also communicate their thoughts through the optional *Additional Feedback* field provided.

Our statistical analysis consisted of two types of tests. We initially determined that numerical variables (such as *N_Labels*, *Total Time*, etc.), were not normally distributed, using a Shapiro-Wilk test. Thus we used the Kruskal-Wallis test to determine any differences in numerical variables. We used Pearson's chi-squared test to examine differences in the boolean variable *Set Completion*. We report median values for all numerical variables, unless otherwise noted, in Table 1. Accuracy, although presented as percentage, should be considered a numerical variable.

	Conditions		
Metric	*Flight Path*	*Bar Ordered*	*Bar Random*
N	327	285	297
N_Labels	47	70	55
Time per Image	3.742	3.762	3.959
Total Time	163	249	265
Accuracy	67.5%	66.7%	60.0%
Set Completion (*)	25.7%	34.7%	36.0%
TLX	130	134	127

Table 1. Summary of performance and survey metrics for workers per condition. Results represent median values, unless otherwise noted. (*) denotes $p < 0.05$.

Our statistical analysis did not reveal any significant differences among the three conditions, when viewing numerical variables. Although workers submitted fewer labels and spent less time overall on the task in the **Flight Path** condition, this performance was not deemed significantly different from the other conditions. We did however notice a significant difference in Set Completion, that is the percentage of workers who ended up submitting labels for all available images in the dataset, with 25.7% of workers in the **Flight Path** condition, compared to 34.7% and 36% in the **Bar Ordered** an **Bar Random** conditions. Our post-hoc pairwise comparison analysis, using Pearson's chi-squared test with the Bonferroni correction, revealed a significant difference between the *Flight Path* and *Bar Random* conditions ($p < 0.05$).

DISCUSSION & FUTURE WORK

Our results indicate that stripping workers of provided order and map context when performing the task did not negatively impact performance by an observable amount. It is possible that there were effects, but the sizes were smaller that could be determined by our study; though we did include nearly a thousand subjects.

We observed similar levels of worker accuracy, when measured as agreement with the ground truth acquired for the dataset. Workers in the condition where no type of context was available (*Bar Random* condition) were able to perform similarly to both conditions with context, whether that was in the form of order (*Bar Ordered*), or in the form of map progress context and order (*Flight Path*). The highest median number of labels provided was recorded in the *Bar Ordered* condition, though not significantly.

When focusing on time spent labeling per image, workers recorded little variance among conditions. As the *Flight Path* condition contained a new visual element not present in the no context condition (*Bar Random*), this result suggests that workers chose to pay limited attention to the flight path progress element, or to ignore it completely. However, this context path element was not explicitly required for the completion of the task; workers could still choose to label the image based on the question shown without having to perform any action on the map at all, whose purpose was more of an optional aid for the task. We believe that lack of map utilization suggests that workers chose to complete the task by optimizing time and minimizing the amount of steps taken to reach payment. This result is supported by previous work on task design and on worker motivation on Amazon Mechanical Turk (Paolacci et al. 2010; Ross et al. 2010).

While not significant, the Task Load Index measure was lowest in the no-context condition *Bar Random*. Although special care was given to the design of the core interface to minimize distractions and unnecessary elements from the task, the increased TLX measures may hint to cognitive load due to the addition of the flight path progress map. Cognitive load on task design has been shown to negatively impact quality of work on crowdsourcing platforms (Finnerty et al. 2013; Kittur et al. 2013; Rahmanian and Davis 2013).

We did observe a significant difference in the percentage of workers completing the entire set of available images between the condition with both types of context (*Flight Path*) and the no context condition (*Bar Random*), which may also suggest some level of cognitive load presence in the first condition. Nevertheless, none of the workers commented on the presence of the map in the survey.

Our work was primarily based on recruiting workers from Amazon Mechanical Turk. We would like to explore the effects of context in other paid crowdsourcing platforms, as well as in pure volunteer work systems. We were also focused on one specific type of task, image analysis. Other types of tasks relevant to disaster response include structure identification on maps as well as emergency text transcription and analysis. It also remains to be seen if similar observations can be made in non-disaster response settings on crowdsourcing platforms such as Amazon Mechanical Turk.

The selected task of this work was characterized by an increased level of complexity, as it requires meticulous image analysis for object extraction, in our case bridge structure identification and quantification. The level of complexity is validated both by the levels of accuracy posted by workers, as well as direct comments on the Additional Feedback optional field in the post-task survey. This also offers a new direction towards exploring better training concepts as well as breaking down the task into smaller, less complex subtasks to achieve our goal.

Other approaches to integrating flight path information may offer more promising directions for future work in motivating engagement in crowdsourcing aerial imagery analysis. We are particularly interested in extracting diverging flight paths from image datasets and allowing participants to branch out from the main path and explore alternative directions, as part of a larger research project focused on exploring task variety. We believe that worker flexibility during the task can contribute to the growing work on how task variety can positively impact worker performance on crowdsourcing platforms.

ACKNOWLEDGMENTS

This work was supported by a Northeastern University TIER 1 grant and Google. We would like to thank the Civil Air Patrol and the U.S. Geological Survey for making the images available and the Mechanical Turk workers for their participation.

REFERENCES

von Ahn, L. and Dabbish, L. (2004). “Labeling images with a computer game”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 319–326.

Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., Lin, A. Y.-M., and Huyck, C. (2012). “Crowdsourcing earthquake damage assessment using remote sensing imagery”. In: *Annals of Geophysics* 54.6.

Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S., et al. (2010). “VizWiz: Nearly Real-time Answers to Visual Questions”. In: *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*. UIST ’10. New York, NY, USA: ACM, pp. 333–342.

Cai, C. J., Iqbal, S. T., and Teevan, J. (2016). “Chain reactions: the impact of order on microtask chains”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3143–3154.

Dai, P., Mausam, and Weld, D. S. (2010). “Decision-theoretic control of crowd-sourced workflows”. In: *Proceedings of the 24th AAAI Conference on Artificial Intelligence*.

Dai, P., Rzeszotarski, J. M., Paritosh, P., and Chi, E. H. (2015). “And now for something completely different: improving crowdsourcing workflows with micro-diversions”. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 628–638.

Finnerty, A., Kucherbaev, P., Tranquillini, S., and Convertino, G. (2013). “Keep It Simple: Reward and Task Design in Crowdsourcing”. In: *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*. CHItaly ’13. New York, NY, USA: ACM, 14:1–14:4.

Global Xplorer (2016). http://www.globalxplorer.org/.

Goodchild, M. F. and Glennon, J. A. (2010). “Crowdsourcing geographic information for disaster response: a research frontier”. In: *International Journal of Digital Earth* 3.3, pp. 231–241.

Hart, S. G. and Staveland, L. E. (1988). “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”. In: *Advances in Psychology*. Ed. by P. A. Hancock and N. Meshkati. Vol. 52. Human Mental Workload. North-Holland, pp. 139–183.

Hazards Data Distribution System Explorer (2016). http://hddsexplorer.usgs.gov/.

Hendrick, T. (2013). *CDOT: 120 bridges damaged by Colorado flooding*.

Ho, C.-J., Slivkins, A., Suri, S., and Vaughan, J. W. (2015). “Incentivizing high quality crowdwork”. In: *Proceedings of the 24th International Conference on World Wide Web*, pp. 419–429.

Imran, M., Castillo, C., Lucas, J., Meier, P., and Vieweg, S. (2014). “AIDR: Artificial Intelligence for Disaster Response”. In: *Proceedings of the 23rd International Conference on World Wide Web*, pp. 159–162.

Khajah, M. M., Roads, B. D., Lindsey, R. V., Liu, Y.-E., and Mozer, M. C. (2016). "Designing engaging games using Bayesian optimization". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5571–5582.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., and Horton, J. (2013). "The future of crowd work". In: *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pp. 1301–1318.

Lasecki, W. S., Marcus, A., Rzeszotarski, J. M., and Bigham, J. P. (2014). *Using microtask continuity to improve crowdsourcing*. Tech. rep. CMU-HCII-14-100. School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Lasecki, W. S., Rzeszotarski, J. M., Marcus, A., and Bigham, J. P. (2015). "The effects of sequence and delay on crowd work". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1375–1378.

Lasecki, W. S., Thiha, P., Zhong, Y., Brady, E., and Bigham, J. P. (2013). "Answering Visual Questions with Conversational Crowd Assistants". In: *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ASSETS '13. New York, NY, USA: ACM, 18:1–18:8.

Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., and Bigham, J. P. (2013). "Chorus: A Crowd-powered Conversational Assistant". In: *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. UIST '13. New York, NY, USA: ACM, pp. 151–162.

Liu, S. B. (2014). "Crisis crowdsourcing framework: designing strategic configurations of crowdsourcing for the emergency management domain". In: *Computer Supported Cooperative Work* 23.4-6, pp. 389–443.

Mitry, D., Peto, T., Hayat, S., Morgan, J. E., Khaw, K.-T., and Foster, P. J. (2013). "Crowdsourcing as a novel technique for retinal fundus photography classification: analysis of images in the EPIC Norfolk cohort on behalf of the UKBiobank Eye and Vision Consortium". In: *PLoS ONE* 8.8, e71154.

Munro, R., Schnoebelen, T., and Erle, S. (2013). "Quality analysis after action report for the crowdsourced aerial imagery assessment following Hurricane Sandy". In: *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management*.

Ofli, F., Meier, P., Imran, M., Castillo, C., Tuia, D., Rey, N., Briant, J., Millet, P., Reinhard, F., Parkan, M., et al. (2016). "Combining human computing and machine learning to make sense of big (aerial) data for disaster response". In: *Big Data* 4.1, pp. 47–59.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). "Running experiments on Amazon Mechanical Turk". In: *Judgment and Decision Making* 5.5, pp. 411–419.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., Szalay, A. S., and Vandenberg, J. (2010). "Galaxy Zoo: exploring the motivations of citizen science volunteers". In: *Astronomy Education Review* 9.1.

Rahmanian, B. and Davis, J. (2013). "Crowdsourcing, cognitive load, and user interface design". In: RMIT University, pp. 1–12.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. (2010). "Who Are the Crowdworkers?: Shifting Demographics in Mechanical Turk". In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '10. New York, NY, USA: ACM, pp. 2863–2872.

Spatharioti, S. E. and Cooper, S. (2017). "On Variety, Complexity, and Engagement in Crowdsourced Disaster Response Tasks". English. In: *Proceedings of the 14th International Conference on Information Systems for Crisis Response And Management*. Ed. by F. B. Tina Comes and 14th International Conference on Information Systems for Crisis Response And Management. Albi, France, pp. 489–498.

Spatharioti, S. E., Govoni, R., Carrera, J. S., Wylie, S., and Cooper, S. (2017). "A Required Work Payment Scheme for Crowdsourced Disaster Response: Worker Performance and Motivations". English. In: *Proceedings of the 14th International Conference on Information Systems for Crisis Response And Management*. Ed. by F. B. Tina Comes and 14th International Conference on Information Systems for Crisis Response And Management. Albi, France, pp. 475–488.

Sturn, T., Wimmer, M., Salk, C., Perger, C., See, L., and Fritz, S. (2015). "Cropland Capture – a game for improving global cropland maps". In: *Proceedings of the 10th International Conference on the Foundations of Digital Games*.

Warren, J. Y. (2010). "Grassroots mapping: tools for participatory and activist cartography". Thesis. Massachusetts Institute of Technology.