Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering
Humanitarian Technology: Science, Systems and Global Impact 2016 00 (2015) 000–000
www.elsevier.com/locate/procedia

Humanitarian Technology: Science, Systems and Global Impact 2016, HumTech2016

CrowdAR: a live video annotation tool for rapid mapping

Elliot Salisbury[a], Sebastian Stein[a], Sarvapali Ramchurn[a]

[a]*Agents Interaction and Complexity Research Group, Department of Electronics and Computer Science, University of Southampton*

Abstract

Digital Humanitarians are a powerful and effective resource to analyse the vast amounts of data that disasters generate. Aerial vehicles are increasingly being used for gathering high resolution imagery of affected areas, but require a lot of effort to effectively analyse, typically taking days to complete. We introduce CrowdAR, a real-time crowdsourcing platform that tags live footage from aerial vehicles flown during disasters. CrowdAR enables the analysis of footage within minutes, can rapidly plot snippets of the video onto a map, and can reduce the cognitive load of pilots by augmenting their live video feed with crowd annotations.

Keywords: real-time, crowdsourcing, search and rescue, UAV

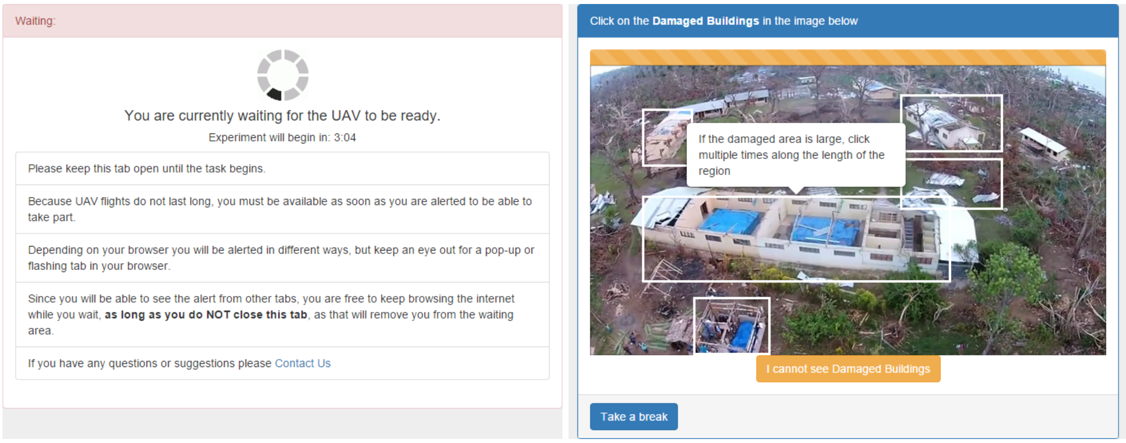
1. Introduction

Major man-made and natural disasters have a significant and long-lasting economic and social impact on countries round the world. The response effort in the first few hours of the aftermath of the disaster is crucial to saving lives and minimising damage to infrastructure. In these conditions, emergency responders on the ground face a major challenge in trying to understand what is happening, where the casualties are, and how to get to them safely. In order to help support this response effort, satellite imagery and aerial footage are increasingly being used by emergency responders and charitable organisations (e.g., Rescue Global, Red Cross, etc..), to help rapidly map disasters. However, the vast amount of imagery coming from such sources takes significant time to analyse in order to determine rescue targets.

Given such challenges, in recent years, a new community of volunteers, called *digital humanitarians* [1] has emerged as a powerful and effective resource to help analyse and organise vast amounts of data coming from disasters. For example, Tomnod, the Standby Task Force, and MicroMappers are but a few organisations that exist to rally volunteers and to carry out rapid data analysis, using typical web-based tools, in the aftermath of major disasters.

To date, these efforts have been focused on annotating satellite and aerial imagery, tracing maps, or classifying tweets [1–4]. In particular, these tasks tended to be ‘offline’ tasks that could take days to complete. However, analysing aerial imagery (e.g., from planes or UAVs) ‘offline’ has a number of drawbacks. First, if the aerial vehicle is flown manually, then the pilot must maintain *continuous focused attention* so that they can spot search targets (e.g., damaged buildings, floating wreckage in the ocean after the missing flight MH370) and then fly the vehicle to get a closer look, and to record footage of it for later crowdsourced analysis. Inevitably, the pilot will eventually become fatigued,

E-mail address: e.salisbury@ecs.soton.ac.uk



(a) The Waiting Room (b) The Tutorial Page

Fig. 1: The interfaces before the live broadcast begins

as these tasks are typically tedious (i.e., flying over an ocean) or visually taxing (i.e., flying over a city), and thus crucial information may be missed [5]. Second, if the aircraft is flown on a fixed flight path, as is typical with UAVs, then the benefit of diverting off course when a target is identified (post-flight) is lost, which can enable responders to gather imagery of potentially important regions that lie at the boundary of the vehicles field of view. Third, in the past, typical crowdsourcing tools have been used to review aerial footage post-flight [6]. Responders who launched the UAV cannot get quantifiable results immediately. Thus, despite being located nearby, cannot know where best to focus their efforts straightaway. Using aerial footage for situational awareness can still be a long and costly process.

Automating these tasks would require artificial intelligence and computer vision algorithms. These algorithms need to be capable of reliably detecting any number of search targets, despite ever changing conditions (e.g., lighting conditions, orientation, partial occlusion). Work has been done to train computer vision classifiers for this purpose. For example, algorithms have been developed to detect damaged buildings [7], or injured people [8]. However, disasters are dynamic and often have different search requirements, it is infeasible to have an algorithm for each eventuality. Instead, a more general purpose approach would describe the target in natural language and use human recognition to understand and detect them.

Against this background, we propose Crowd Augmented Reality, or CrowdAR,[1] a real-time crowdsourcing platform for annotating live video feeds, including, but not limited to, those from aerial vehicles [9]. CrowdAR uses scalable server-side processing for computer vision and real-time aggregation of crowd sourced data, and a specifically designed UI optimised for real-time crowdsourcing of live video. In the future, we intend to open-source CrowdAR, enabling everyone to use it.

Real-time crowdsourcing is the process of outsourcing work to multiple, simultaneously connected online workers [10]. When applied to disaster response, we can find and tag search targets in a live video feed. These annotations can then be aggregated and used to augment the live feed with the crowdsourced data, thereby reducing the cognitive load of the pilot. For example, informing a fatigued search and rescue pilot to floating wreckage or damaged buildings identified by the crowd, enabling the pilot to make more effective in-flight decisions. Furthermore, real-time crowdsourcing enables many more people to observe the live camera feed, and in so doing, they are more likely to observe and accurately identify important details than a single expert could. Additionally, while a single observer may become fatigued, a real-time crowdsourced team is a tireless workforce. More specifically, individuals of the crowd that tire and stop contributing to the task, can be replaced by new crowd members. As such, the pool of workers is always being replenished with fresh individuals.

Furthermore, through the use of in-flight analysis, CrowdAR reduces the delay between data gathering and the reviewing of footage. Moreover, in-flight analysis allows us to potentially react to events or features the crowd has spotted (e.g., damaged buildings, floating debris) and move the vehicle off its original search path to investigate.

[1] `https://crowdrobotics.org/projects/crowdar`

Fig. 2: The Live Feed Interface

To date, CrowdAR has been evaluated with both paid workers on crowdsourcing platforms, such as Amazon Mechanical Turk (AMT), and with digital humanitarian volunteers, such as MicroMappers, using videofeeds sourced from UAV flights after the Cyclone Pam disaster in Vanuatu, 2015.

2. CrowdAR

CrowdAR is a web-based tool used to crowdsource the annotation of a live video feed for supporting digital humanitarian efforts. CrowdAR can be used, for example, to rapidly map damaged buildings during a UAV flyby, or to identify search and rescue targets in both visually challenging or tedious environments.

CrowdAR consists of four main components, three from the perspective of the volunteers, and one final component that processes the volunteers data. The first stage that the digital humanitarians will encounter is the **Waiting Room**, here users wait for the live video feed to begin broadcasting. The second, and optional, stage is the **Tutorial** page, volunteers can acquaint themselves with the upcoming task. Third, once the live feed is ready users will connect to the **Live Tagging Interface**. Finally, the fourth component involves **Processing Crowd Tags** to the benefit of the emergency responders. In what follows, we cover each of these in turn.

2.1. Waiting Room

Due to the nature of live video, and especially live aerial footage, the video feed will not always be broadcasting. Furthermore, the broadcast will likely only be active for a short interval, given that flight time, especially with UAVs, is limited. To ensure that we have a crowd of workers ready to begin tagging footage the moment the vehicle is airborne and the footage is broadcasting, we need to pre-recruit a crowd. Depending on the situation, there may be a large number of volunteers ready to participate, these could be alerted in a number of ways, for example via emails, push notifications, or Skype calls. However, if the volunteer efforts are low, crowds can be hired using typical crowdsourcing platforms such as Amazon Mechanical Turk. The participants are given a link to the waiting room, where they reside until the live feed begins broadcasting, see Figure 1a.

Participants are informed that they will be alerted when the feed is ready. Provided the volunteers keep the waiting room tab open in their browser, they may continue to use their computer or mobile device as usual. Once alerted, the majority of workers typically return to the task within 2 seconds [10]. Given the freedom to browse other pages, workers may then review the tutorial page.

2.2. Tutorial

The tutorial[2] was designed in partnership with the Standby Task Force to give brief but informative instructions on how to perform the upcoming tasks. A replica of the live interface is shown, with an example still frame from a similar UAV flight, see Figure 1b. The tutorial can be changed and refined depending on the task at hand, and the current disaster scenario. Once a live video feed is ready to broadcast, participants are alerted and can begin tagging.

[2] See a live example of the tutorial here: `https://crowdrobotics.org/projects/crowdar/howto`

(a) Augmented Live Video Shown to Pilot

(b) Mapped Snippets

Fig. 3: The Output to First Responders

2.3. *The Live Tagging Interface*

The live tagging interface[3] enables users to identify search targets in the live video feed. Users are shown a single still frame from the live feed for a short duration, a countdown timer bar that displays the duration left of the still frame, and two buttons, indicating that there is no damage visible in the frame, or that the user wishes to take a break, see Figure 2.

A still frame is shown to the workers, rather than the live feed, as this requires less bandwidth (which may be limited in disaster scenarios), allows workers to annotate multiple targets per frame, and because previous experimentation has shown that because the search targets are not moving it is easier for workers to accurately and quickly click on them. Thus the workers identify more targets, more accurately. Previous experimentation suggests that a 2 second duration for the still frame provides the best trade-off between worker accuracy and input-rate for CrowdAR to receive the data soon enough to augment the pilot's feed with the crowdsourced annotations [9]. Participants then click the frame on areas they believe to be damaged. After 2 seconds, a new frame is loaded, and the user's clicks are submitted to CrowdAR for tracking.

The tags generated by the volunteers are sent to the CrowdAR server, processed and tracked through the live video, and if an area of damage is identified by multiple participants, it is considered an identified search target and can then be forwarded for further processing.

2.4. *Processing Crowd Tags*

CrowdAR is capable of analysing the live crowd data in a number of ways. For example, the following two methods are of benefit to humanitarian aid. First, CrowdAR can augment the live video feed to the pilot, showing the pilot the crowdsourced annotations during the flight, see Figure 3a. Thus, allowing the pilot to divert the path of the aircraft to investigate areas of interest indentified by the crowd. This could be useful for search and rescue flights, especially over the ocean looking for missing boats and floating wreckage. During the search for flight MH370, crew would look out a window for hours at a time. CrowdAR would instead notify the pilot the moment the crowd spots something of interest. This enables the pilot to make a better judgement and decide to investigate immediately. In contrast, if the footage was analysed post flight, returning to the same area again would not be possible as ocean currents take the debris away.

Second, rapidly mapping the search targets (e.g., damage) onto a GIS application (e.g., Google Maps). Once an area of damage has been identified by the crowd and tracked through the video by CrowdAR, a short snippet can be generated by cropping the video and highlighting just the search target. These short snippets of video may then be sent off for further verification and classification, by either an expert or a trained crowd. This significantly reduces the workload of the experts, who need only accept or reject the snippets, rather than watch a whole video. These snippets can then be plotted onto a map, see Figure 3b. All of this can be performed simultaneously with the live video feed tagging, and snippets can be verified and plotted within a matter of minutes of the UAV first observing the target.

[3] A video of how to interact with the interface is shown here `https://vimeo.com/140445323`

3. Conclusion

CrowdAR is a system for crowdsourcing the analysis of live video feeds. This analysis can be achieved in a matter of minutes, in contrast to traditional methods which can take days. The crowdsourced annotations can be used in a number of ways. First, augmenting the live video for a pilot, reducing their cognitive load and enabling more effective and timely decision making. Second, extracting snippets from the feed and rapidly plotting their location on to a map.

References

[1] P. Meier, Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response, CRC Press, 2014.

[2] A.-M. Lin, A. Huynh, L. Barrington, G. Lanckriet, Search and discovery through human computation, in: P. Michelucci (Ed.), Handbook of Human Computation, Springer New York, 2013, pp. 171–186.

[3] S. Vieweg, A. L. Hughes, K. Starbird, L. Palen, Microblogging during two natural hazards events, Proceedings of the 28th international conference on Human factors in computing systems - CHI '10 (2010) 1079.

[4] P. Meier, Using MicroMappers to Make Sense of UAV Imagery During Disasters, 2014.

[5] M. L. Cummings, C. Mastracchio, K. M. Thornburg, A. Mkrtchyan, Boredom and distraction in multiple unmanned vehicle supervisory control, Interacting with Computers (2013) iws011.

[6] P. Meier, Debrief: UAV/Drone Search and Rescue Challenge, 2014.

[7] G. Ouzounis, P. Soille, M. Pesaresi, Rubble detection from vhr aerial imagery data using differential morphological profiles, in: Proceedings of the 34th International Symposium for Remote Sensing of the Environment, International Center for Remote Sensing of Environment (ICRSE), 2011.

[8] W. T. L. Teacy, S. Julier, R. D. Nardi, A. Rogers, N. R. Jennings, Observation modelling for vision-based target search by unmanned aerial vehicles, in: 14th International Conference on Autonomous Agents and Multi-Agent Systems, 2015, pp. 1607–1614.

[9] E. Salisbury, S. Stein, S. Ramchurn, CrowdAR: Augmenting Live Video with a Real-Time Crowd, in: Conference on Human Computation & Crowdsourcing (HCOMP 2015), 2015.

[10] M. S. Bernstein, J. Brandt, R. C. Miller, D. R. Karger, Crowds in Two Seconds: Enabling Realtime Crowd-powered Interfaces, in: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11, ACM, New York, NY, USA, 2011, pp. 33–42.