SANDIA REPORT

SAND2023-14455
Printed September, 2023

Extracting Ground Truth from Surveillance Video in the Dry Alluvium Geology (DAG) Experiment

Craig D. Ulmer and Nicole D. McMahon

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

ABSTRACT

The Advanced Data Analysis for Proliferation Detection (ADAPD) project is a NNSA NA-22-sponsored Venture that is developing novel data analysis capabilities to detect low-profile nuclear proliferation activities. A key step in the information refinement process for this work is to inspect input sensor datasets and data products produced by our analytics to generate as much "ground truth" as possible about the events that took place during the period of observation. This information helps the team's data scientists improve and validate their algorithms and yields data products that are valuable to analysts and decision makers. In this report we provide information about how we inspected multimodal sensor data from the Source Physics Experiment's Dry Alluvium Geology (DAG) tests and generated ground truth for ADAPD's analysis teams. This work illustrates the front-end data engineering tasks that frequently arise in new studies and documents our efforts to gain greater confidence in the assessments of the data.

ACKNOWLEDGMENT

This work was funded by the U.S. Department of Energy National Nuclear Security Administration's Office of Defense Nuclear Nonproliferation Research & Development.

We are thankful for the support and assistance of several colleagues. Brian Key from LANL provided us with access to the surveillance video dataset and worked through the logistical challenges of shipping a large amount of data during COVID19 shelter-in-place conditions. Mike Rivera shared his insight about activities in the video data and pointed us to interesting time periods to explore. Jonathan MacCarthy from LANL supplied the seismic data and emphasized the complexity associated with inferring activities from the signals. Dan Champion from NNSS provided data for the site and patiently explained how operations normally take place during experiments. Margaret Townsend from NNSS reviewed this report after DAG data was made available to a larger audience. Finally, Dylan Anderson, Erick Draayer, Renee Gooding, Ali Pinar, David Stracuzzi, and Fulton Wang at Sandia provided feedback about what information products would be most useful for their work in ADAPD.

CONTENTS

Executive Summary 7

1. Introduction 9
1.1. Dry Alluvium Geology (DAG) Tests 10

2. Approach 12
2.1. Survey of Input Datasets 12
2.2. Prioritizing Data by Analysis Needs 13
2.3. Initial Output Targets 13

3. Experiments and Results 14
3.1. Timing Challenges 14
3.1.1. Resolving Timezone Uncertainty 14
3.1.2. Estimating Video Camera Time Skew 15
3.1.3. Extracting Video Timestamps 16
3.2. Large Vehicle Movements 17
3.2.1. Manually Creating a Timeline of Large Vehicle Movements 17
3.2.2. Hand-Labeling Large Vehicles 18
3.2.3. Vehicle Locations 21
3.2.4. Hourly Activity 22
3.3. Aggregating Seismic and Video Data 23
3.4. Vehicle Ingress and Egress 25
3.5. Measuring Human Annotation Rates 26

4. Mission Purpose and Impact 28

5. Discussion and Future Work 29
5.1. Challenges 29
5.2. Opportunities for Future Work 29

References 31

LIST OF FIGURES

Figure 1-1. The DAG project was conducted at the Nevada National Security Site (NNSS) in Yucca Flat, Nevada. The instruments and locations involved in this study are highlighted in the bottom photo. 11

Figure 3-1. Comparing the observed sunrise to the expected positions provided in Google Earth provides a means of confirming the timezone. 15
Figure 3-2. Comparing the 06:00 view on consecutive days reveals that the local clock was adjusted for daylight saving time, though two days later than expected. 15
Figure 3-3. Large vehicle movements can be used to estimate how much skew there was between the videos and seismic sensors. 16
Figure 3-4. A summary of different activities observed in the videos was recorded to create a timeline of activities. 18
Figure 3-5. The labelImg tool was used to create bounding boxes for different vehicles observed in the videos. 19
Figure 3-6. The hand labeling effort produced a large number of example images for different vehicles at the site. 20
Figure 3-7. The positions of all the labeled vehicles during the two week period. 21
Figure 3-8. Hourly counts of vehicle movements provided a timeline of activities for a day. 22
Figure 3-9. The hourly vehicle movement counts for week 41 were more subdued and focused primarily on moving fill material. 23
Figure 3-10. The surveillance videos and summary plots were combined in the aggregate video to provide an easy way for users to scan through large amounts of time series data. 24
Figure 3-11. Aggregate video screen capture displaying the utility of joint seismic/video analysis for identifying movement time-stamps and associated vehicles. The yellow box around the seismic data in the lower portion of the figure is associated with the entry of the gravel trailer circled in the upper portion of the figure. 25
Figure 3-12. The DAG video label dataset was examined to determine how many annotations per minute a human made over the course of (a) one day and (b) one month of work. Each red circle represents the number of frames the human annotated at a particular minute in their work schedule. 26
Figure 3-13. The histogram of annotation rates and cumulative number of annotations for this project reveal that labeling work was slow and bursty. 27

EXECUTIVE SUMMARY

The *Advanced Data Analysis for Proliferation Detection* (ADAPD) project is a U.S. Department of Energy, National Nuclear Security Administration-sponsored Venture that is developing novel data analysis capabilities to detect low-profile nuclear proliferation activities. This project is divided into three hard problem areas, with each problem targeting a specific part of the nuclear proliferation process that experts have found difficult to assess. Hard problem three (HP3) focuses on the activity build up corresponding to a nuclear related test. ADAPD's HP3 work is focused on analyzing multimodal surveillance data about a site to not only confirm that test build-up is related to proliferation, but also discover how advanced the tests are and make predictions about when the next test will take place.

As a means of studying how new analytics can be developed to provide better characterization of build-up-to-test activity, the ADAPD HP3 team in FY20 explored datasets that were generated previously for the NNSA in the Source Physics Experiment (SPE). SPE is a long-running program that has conducted a number of tests that detonate conventional explosives at different depths underground. The recent Dry Alluvium Geology (DAG) tests collected a large amount of continuous, seismic data before, during, and after the tests, as well as a significant amount of operations data about the work required to complete the tests. This data is extremely relevant to ADAPD because it documents the process of running a real-life, large-scale, physical experiment in the presence of weather mishaps, construction delays, and various human factors.

The first step in analyzing this worksite was to collect data from available sources, normalize it as best as possible, and then establish a "ground truth" record of what activities took place at different points in time for the experiment. This information helps data scientists locate regions of interest in the data and can help validate whether new analytics are producing meaningful results. This report provides information about how we inspected data from the DAG experiment and extracted ground truth. The majority of this work involved human efforts to manually extract information from the data. We primarily focused on the surveillance video to make and confirm fundamental assumptions about timing information, extract fine-grained timelines for different activities, identify common vehicles and their normal locations, and build ingress/egress logs to estimate how busy the site was at a particular time. The results of this work were used by other ADAPD researchers to develop new analytics that are documented in companion reports. We conclude this report by discussing the key challenges we faced in digesting this data and listing opportunities for improving this process when transferring to new locations.

Editor's Note: This report was originally written as an internal document in 2020 due to release restrictions of the DAG data. In 2021 the DAG data was made available to analysts and organizations that were not participants in the original experiments. NNSS's Data Release Report [4] provides information about this data.

1. INTRODUCTION

The *Advanced Data Analytics for Proliferation Detection* (ADAPD) project is a NA-22-sponsored Venture that is developing novel data analysis capabilities for the detection of low-profile nuclear proliferation activities. ADAPD's hard problem three (HP3) thrust deals with predicting and characterizing build-up-to-test activity and aims to make predictions about when the next test will take place. This research involves developing a number of information-refinement operations that convert raw multimodal data collected from a variety of sensors into information products that yield a greater understanding of an adversary's operations. Detecting activity earlier and improving prediction accuracy improve proliferation detection capabilities.

A significant challenge in this proliferation detection work is that every mission target is effectively a unique scenario with its own operating conditions. Each facility that is monitored has its own unique physical constraints (e.g., in a rural area, mountain, industrial complex, etc.) as well as human factors to consider (e.g., staffed by well-paid civilians vs. unmotivated military, risk averse vs. risk tolerant work cultures, etc.). Similarly, sensor data quality and quantity about a region of interest varies depending on where the site is located. It should be expect that historical, remote-sensing data will be sparse and that in-situ sensors will likely be placed in locations that are not ideal or have sampling rates that miss key events.

The first step in developing analytics to help monitor a new region of interest is to evaluate the different, multimodal data sources that are available and establish "ground truth" about what events happened at different points in time during the surveillance period. This work is often a manual process that involves human operators examining the available data sources and constructing temporal and positional alignments that make it easier for downstream analytics to extract information. The ground truth process may generate a variety of useful information products, such as a global timeline of activities, a catalog of all actors and actions in the timeline, and synchronized data products that help analytic developers debug, tune, and verify their work. As analytics mature, their results can be used to refine ground truth information and improve confidence in their accuracy.

In order to better understand the challenges of establishing ground truth for a mission, it is useful to work through a practical example and document how different pieces of information can be extracted. For FY20 the ADAPD HP3 team explored surveillance data collected from the NA-22 DAG experiment to develop new analytics for characterizing activities leading up to explosive testing. This report contains an overview of the different surveillance datasets that were collected and focuses on the process by which we established ground truth and initial information products that other HP3 team members could use to improve their algorithms. In addition to providing evidence for our assessments, we document hardships faced in this work and discuss opportunities for improving the ground-truth process.

1.1. Dry Alluvium Geology (DAG) Tests

The U.S. Department of Energy, National Nuclear Security Administration conducted a series of experiments aimed at improving arms control and treaty verification. These experiments represented a nine-year effort known as the Source Physics Experiments (SPE) [7], and were conducted at the Nevada National Security Site (NNSS). The SPE series of underground chemical detonations (high-explosive and nitromethane) at various yields and depths were designed to inform an explosive source prediction capability and improve the ability to detect and verify low-yield nuclear explosions in a noisy geophysical environment. SPE was sponsored by the Office of Defense Nuclear Nonproliferation Research and Development (NA-22) [1, 9].

SPE Phase I consisted of six underground explosions in granite (hard rock). SPE Phase II, also known as the Dry Alluvium Geology (DAG) project, consisted of four underground explosions in the dry alluvium (soft rock) geologic environment of Yucca Flat, Nevada. These DAG experiments consisted of the following detonations [1]:

- DAG-1: 1 metric ton at 385 m depth (July 2018)
- DAG-2: 50 metric tons at 300 m depth (December 2018)
- DAG-3: 1 metric ton at 150 m depth (April 2019)
- DAG-4: 10 metric tons at 100 m depth (June 2019)

The experiments were highly instrumented with up to 1,500 sensors taking measurements. Diagnostics included infrasound, seismic, various borehole instruments, high-speed video, geologic mapping, drone-mounted photography, distributed fiber-optic sensing, electromagnetic signatures, temperature fluctuations, gas-displacement recordings, ground-surface changes from synthetic-aperture radar and lidar [3].

In this study, we focus on continuous, time-lapse video oriented to see emplacement hole and the seismic data from 5 intermediate-field geophones (Figure 1-1).

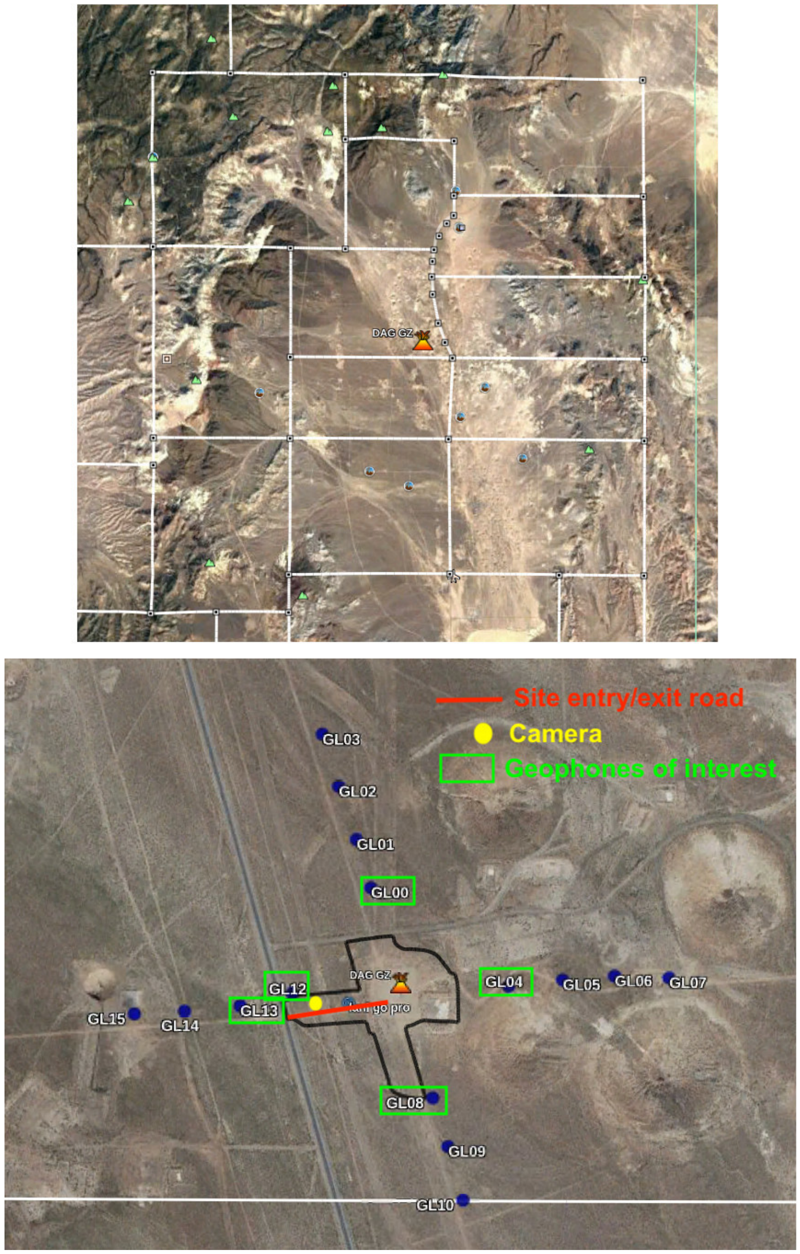


Figure 1-1. The DAG project was conducted at the Nevada National Security Site (NNSS) in Yucca Flat, Nevada. The instruments and locations involved in this study are highlighted in the bottom photo.

2. APPROACH

Our goal in this work was to extract as much relevant and useful information from the available data sources as possible and generate information products that other teams could leverage. We approached this work in three steps. First, we surveyed the available input datasets and assessed their potential value to other team members. Second, we prioritized analyzing the datasets based on needs. Finally, we performed multiple analytics on the data to extract different information products from the data.

2.1. Survey of Input Datasets

The ADAPD venture has collected multiple datasets relating to the DAG experiments, summarized as follows.

Continuous Seismic Data: The DAG experiment collected continuous seismic data from hundreds of sensors at local and regional distances. Our team obtained data from five 3-component 500 Hz intermediate-field geophones (GL00, GL04, GL08 GL12, and GL13) from January 1, 2018 to May 30, 2019.

Time-Lapse Video Surveillance: The DAG experiment placed a Brinno TLC200 PRO time-lapse video camera at the worksite and captured an image every 10 seconds. ADAPD initially acquired two weeks of videos from November 19, 2018 to December 3, 2018. Later in the year we received additional videos that covered October 1, 2018 to January 2, 2019.

DAG Activity Logs: ADAPD team members at NNSS and LANL collected a variety of project planning documents for the DAG work and summarized the information in activity logs. These logs included the number of people and types of equipment scheduled for work on a particular day, but did not provide a timeline of activities within the day.

NNSS Operations Data NNSS also supplied multiple datasets relating to site operations, including power utilization for different locations and network activity information. Unfortunately, this data was only collected at a coarse granularity during the time of the DAG work and lacked coverage of the worksite.

Sentinel Overhead Imagery: The European Space Agency (ESA) provides free access to historical data collected by its Sentinel satellites. The Sentinel-1 and -2 missions were active during the DAG experiments and produced a variety of imagery products at up to 10m resolution. NNSS collected a large amount of imagery for the Nevada site, though the interval between captures varied from 7 to 14 days.

2.2. Prioritizing Data by Analysis Needs

The ADAPD team inspected each input dataset to estimate how valuable the data might be to the analytic teams so that we could prioritize our initial information extraction operations. While the overhead imagery and operations data would normally be extremely valuable for assessing a work site, the sparsity of this data caused us to give the data a low priority. The activity logs and operations data were deemed to be as refined as they needed to be for the time period of the DAG experiment. Given that the raw seismic data was already the primary input for multiple analysis efforts, we chose to focus on inspecting the surveillance videos for our work.

2.3. Initial Output Targets

Based on the activities of the other teams in ADAPD, we determined there were multiple data products that could be extracted from the surveillance video that would be useful in establishing ground truth:

Vehicle Activity Logs: Multiple ADAPD researchers indicated that it would be useful to have accurate information about when different vehicles performed actions in the videos. We monitored the videos and created a log of vehicle activities to document when and where large vehicles moved in the videos. Each activity in these logs includes a start and stop time, the name of the vehicle, and a short description of the activity. These logs give a timeline of activities that helped the broader team navigate the videos.

Bounding Boxes for Vehicle Images: ADAPD researchers working on object detection algorithms needed example images of the different vehicles observed at the site to help train and customize their algorithms. We inspected the videos and generated bounding-box labels for many different vehicles. The labels are also of use to others in the project because the coordinates of the boxes can be used to estimate where different vehicles are located at a particular point in time.

Vehicle Ingress/Egress Logs: A team analyzing vehicle traffic through the main entrance to the DAG worksite expressed interest in having a more detailed log of all vehicles (large or small) entering or leaving by the road. The ground truth effort monitored the surveillance video and created a log of all traffic entering or leaving the site.

Aggregate Video Streams: Multiple teams needed a quick way to visually compare the seismic data with activities in the videos. As a means of distilling this information into one stream that consumers could reference while evaluating their analytics, the ground truth effort constructed aggregate videos that combined both the surveillance video and different plots related to the seismic data. These streams were useful because they enabled researchers to rapidly scan through the timeline and visually verify events that their algorithms detected.

3. EXPERIMENTS AND RESULTS

A number of experiments were conducted with the DAG datasets to provide better ground truth about what activities took place at different times during the experiment. This section provides information about how we examined the surveillance video to develop new information products for other users in ADAPD. This work largely focused on resolving timing issues with the videos, characterizing large vehicles and their movements, and monitoring vehicle traffic as it entered and left the work site. In addition to providing evidence to support our ground truth assumptions, this section provides an example of the kinds of problems that must be resolved in the early phases of the analysis pipeline.

3.1. Timing Challenges

One of the challenges of generating ground truth from multiple data sources is resolving time synchronization across all the sources. While the geophones utilize GPS clock synchronization to ensure accurate timing, there was a great deal of uncertainty about the timestamp values that are displayed in the surveillance videos. First, at a coarse granularity it is unclear what timezone is being reported by the local time, and whether the clocks are properly adjusted for daylight saving time. Second, assuming the videos are not tightly synchronized with other time sources, what is the time skew between the videos and the seismic data?

3.1.1. Resolving Timezone Uncertainty

The timestamps displayed in the videos are reported in a local time but lack a timezone identifier (e.g., PDT or PST) for properly converting the value to a Coordinated Universal Time (UTC) representation. The experiment site is in the Pacific timezone and therefore our expectation is that the timestamps are either UTC-7 or UTC-8 depending on daylight saving time. However, it is important to confirm this assumption given that selecting the wrong timezone would result in all timings being off by an hour.

The Sun and Moon are visible in some of the videos and provide a well-known reference point for resolving the timezone question. The camera faces east and records imagery from roughly 06:00 to 19:30. We examined video from November 1, 2018 and extracted images at 06:00, 07:00, and 08:00 local time. We then used Google Earth to render the view from the camera's perspective with both the Sun and its sunlight enabled in the rendering for the UTC times of 13:00, 14:00, and 15:00 on November 1, 2020. As illustrated in Figure 3-1, the Sun's position matches its expected location for UTC-7 times. Given that daylight saving time ended three days later at 02:00

November 4 in 2018, we are confident that the *summer hours* are correctly reported in UTC-7 (PDT) times.

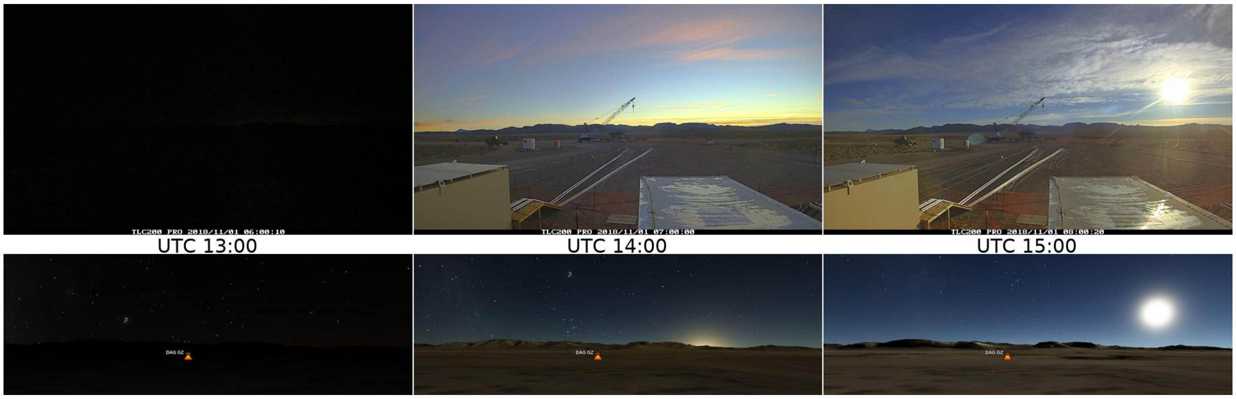


Figure 3-1. Comparing the observed sunrise to the expected positions provided in Google Earth provides a means of confirming the timezone.

We next examined the videos after November 1st to determine how the daylight saving time change was handled. The camera did not capture the actual time change at 02:00 because it does not record at night. As Figure 3-2 illustrates, inspecting the images at 06:00 for the days around the daylight saving time day reveals that the local clock *does get adjusted*. However, this shift happens two days after the expected day (November 6th instead of November 4th). Initial images on the 6th show the camera facing north. We believe the camera did not automatically adjust its time and that a worker manually updated it.

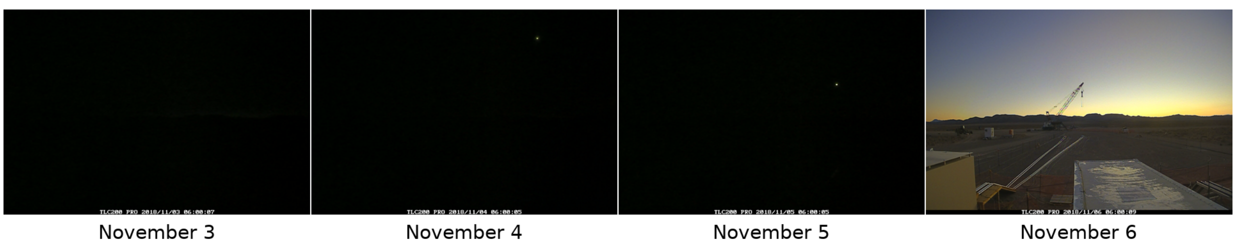


Figure 3-2. Comparing the 06:00 view on consecutive days reveals that the local clock was adjusted for daylight saving time, though two days later than expected.

3.1.2. Estimating Video Camera Time Skew

A second challenge for building ground truth from the surveillance video was estimating how much skew there was between the timestamps in the videos and the seismic data. While the clock for the camera was set to the local time when installed, it is unlikely that the clock was synchronized with an external source and was therefore likely to be out of sync with the geophones. The ground truth team compared the seismic data with large vehicle activity in the video to get a rough estimate of the time skew. This work was subjective and hindered by both the low frame rate of the cameras (1 frame per 10 seconds) and the complexity of how seismic energy propagates through the ground.

We examined multiple types of vehicle movements for this work, including (1) the large crane moving away from the work site, (2) fill trucks driving to the site and unloading material in the distance, and (3) various vehicles entering and leaving the site through the main entrance. While we expected the crane movements to be the easiest to track, we found that the operations exhibited emergent onsets and offsets in the seismic data making identifying definitive starts and stops of activity difficult. The fill trucks generated high-amplitude seismic signals when unloading, but were difficult to observe because they operated in the distant background.

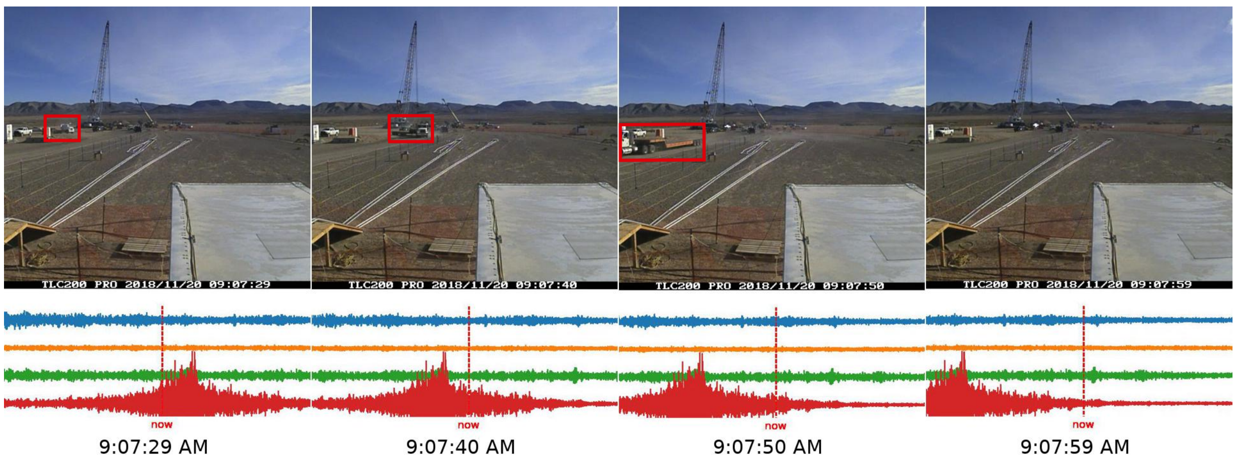


Figure 3-3. Large vehicle movements can be used to estimate how much skew there was between the videos and seismic sensors.

Monitoring vehicle ingress and egress for the site provided a better opportunity for performing time skew estimation. The GL12 sensor was located close to the site entrance and usually recorded strong signals when vehicles drove by it. While the camera did not point at this sensor, we believe there was enough of a view of the road leading to the exit that a human could make a reasonable estimate of when a vehicle passed by the sensor. Based on our observations, we estimated *the video lags the seismic data by approximately 30 seconds.*

3.1.3. Extracting Video Timestamps

Timestamp information for surveillance video was often only recorded graphically in the visible images and required image-to-text processing for extraction. Examining the DAG video revealed that the camera consistently placed the timestamp in a known location using a fixed-width font and a solid black background. Given these ideal conditions, we applied textbook image processing techniques to extract the text with a simple OpenCV script. This script converted the image to grayscale, cropped the image into 19 subimages where the timestamp digits exist, and then compared a simple hash of each digit to a dictionary to determine the digit's value. Through experimentation we found that counting the on bits in the top and bottom halves of the digits provided an easy way to distinguish digits. We created a dictionary by assembling a collection of digit samples and generating hashes for each digit. This approach is not portable to other cameras, but it worked well for this dataset and is fast.

Our datestamp extraction algorithm failed to resolve 69 of the 33,616 frames from the first week of video. Several of these errors can be attributed to image corruption. Given that the camera captured images at fixed intervals, the extraction script was modified to estimate corrupted dates based on a time offset from the previous frame.

For portability, we explored the use of established optical character recognition (OCR) programs that use neural networks to extract text from imagery. These implementations are much more sophisticated and are intended to provide a general solution that works with different fonts, languages, and text warpings. Tesseract OCR[1] is widely referenced as one of the better open-source solutions. We ran Tesseract against multiple camera videos and observed mixed results with the provided data models. Our conclusion is that the tool could be made to extract dates, but a user would first need to train against a number of samples to tune for a particular camera. If an operation is using a particular camera for its video, it may be quicker to simply write an OpenCV script such as ours than to train a general-purpose OCR tool.

3.2. Large Vehicle Movements

Another common task in digesting surveillance video is extracting different pieces of information from the video that are valuable for other analytics. For the DAG experiment there was interest in generating ground truth about seismic signals generated by large vehicles moving about the site. Therefore we focused on two tasks related to the extraction of information about large vehicle movements: (1) building a timeline of activities in which large vehicles moved about the site and (2) extracting bounding boxes for vehicles to help improve object classifiers that could automate finding vehicle movements.

3.2.1. Manually Creating a Timeline of Large Vehicle Movements

A key operation in digesting surveillance video is constructing a timeline of events that took place in the video. This timeline is important because it provides a summary that consumers can use to determine which part of the videos are worth investigating. Timelines also provide a simple text data product that downstream analytics can use as input for statistics (e.g., How many fill trucks visited the site each day?).

We initially focused our analysis on a two-week time period around the Thanksgiving holiday to align with the efforts of our colleagues analyzing the DAG seismic data. We chose to focus primarily on large vehicle movements due to the resolution of the videos and the expectation that the movements of the larger vehicles would generate higher amplitude seismic signals and thus be easily identifiable. No special tools were used in this process: a human simply played the video and recorded information about significant activities. This information included the start/stop times, the vehicle type, and a short summary of what action took place.

[1] `https://github.com/tesseract-ocr`

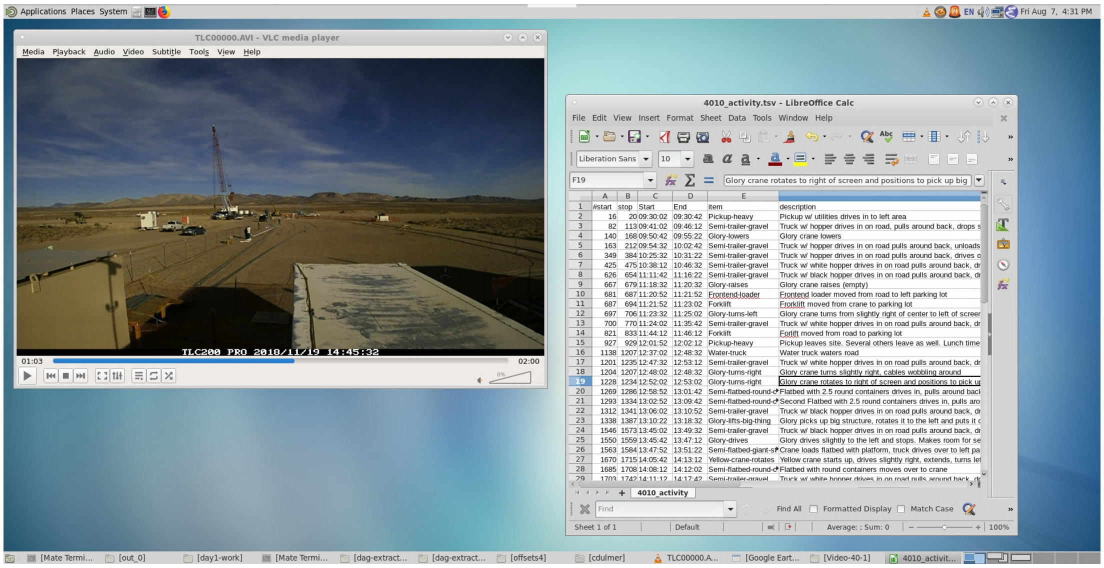


Figure 3-4. A summary of different activities observed in the videos was recorded to create a timeline of activities.

From a labeler's perspective, there were many factors that made this work tedious. First, surveillance videos often have long periods of inactivity interspersed with short windows of activity. It is easy to miss actions where objects return to their starting positions. Second, long videos make it difficult to pan through small intervals effectively. It is easy for humans to lose track of where one scene ends and the next begins when panning. Third, it can be tedious capturing notes about an activity, especially when a similar action happens many times. It is clear that all of these issues could be addressed with better labeling tools that increase the user's efficiency and automate routine tasks.

3.2.2. Hand-Labeling Large Vehicles

The amount of surveillance video from the DAG experiment was large enough that it was impractical for a human to manually summarize all of it. As such, it was important to leverage object detection tools that could automate the information extraction process as much as possible. We initially explored two paths to automating the vehicle extraction process. First, we explored writing traditional change-detection tools in OpenCV to filter out periods of inactivity from the videos to help a human labeler concentrate on sections of the video where there were events of interest. Unfortunately, these approaches were largely ineffective without significant hand tuning due to noise in the videos and the fact that the main region of interest was in the distant background. Second, we explored using open-source implementations of the YOLO [5, 6] algorithm to identify vehicles. However, YOLO's default data models did a poor job of identifying the construction vehicles observed at the site. It was clear that YOLO would need to be trained with more data to be effective.

In support of other efforts in ADAPD that are building better object detection algorithms for surveillance video, we examined two weeks of the DAG video and created bounding boxes for as

Figure 3-5. The labelImg tool was used to create bounding boxes for different vehicles observed in the videos.

many key vehicles as possible. After exploring different image annotation tools, we chose the popular labelImg [10] tool because it received good reviews from other users, was written in Python, provided a simple interface for drawing bounding boxes around different items in a series of images, and generated easy-to-parse output in YOLO-formatted text files. We modified the tool to make it more ergonomic for stepping through a sequence of images in a video. Our process for hand labeling was to watch the video in a player to find the next scene of activity, skip to the corresponding frame in labelimg, draw boxes, advance the frame, and repeat until all frames in a scene were labeled.

Figure 3-6 provides a sample of some of the images the video labeling effort provided as training data for other ADAPD efforts. Over 2,200 image samples were extracted for eighteen types of vehicles. Key vehicles such as the crane were further decomposed into additional categories to capture orientation and types of operation. The label boxes are defined in plain text files that are stored in a git repository. This repository provides consumers an easy way to verify they are up to date with the data and allows us to keep a fine-grained log of changes made by different people to the data.

Figure 3-6. The hand labeling effort produced a large number of example images for different vehicles at the site.

3.2.3. *Vehicle Locations*

A benefit of placing boxes around the vehicles during the labeling process was that it became easier to extract information about where different vehicles normally resided while at the experiment site. These locations may help downstream analytics establish relationships between different vehicles and make better predictions about sequences that take place during different operations. Figure 3-7 plots the lower-center box location of every label generated during the two week period. As expected, vehicles typically follow the roads and gravitate to areas related to their function.

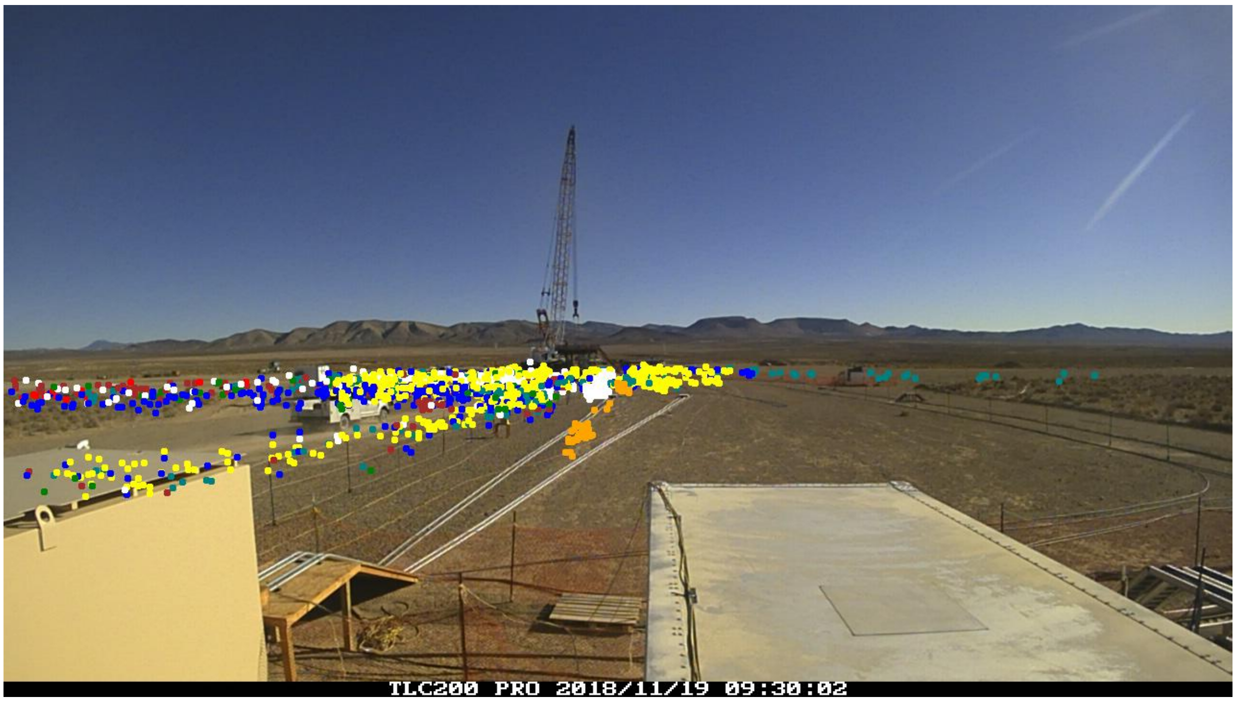


Figure 3-7. The positions of all the labeled vehicles during the two week period.

3.2.4. *Hourly Activity*

Another benefit of vehicle labels was that they could provide a simple metric for tracking how active each vehicle was at different points in time. For example, we counted the number of labels generated in an hour for each major vehicle category during the monitoring period. As illustrated in Figure 3-8, a great deal of work took place during the first two days of week 40. The Old Glory crane (green) operated for a day and a half before being driven off site. Gravel (red) and sand (yellow) trucks worked continuously and deposited a significant amount of material at the site. In terms of site services, the water truck (blue) visited both days while the septic truck (pink) visited on the second day after most activities finished.

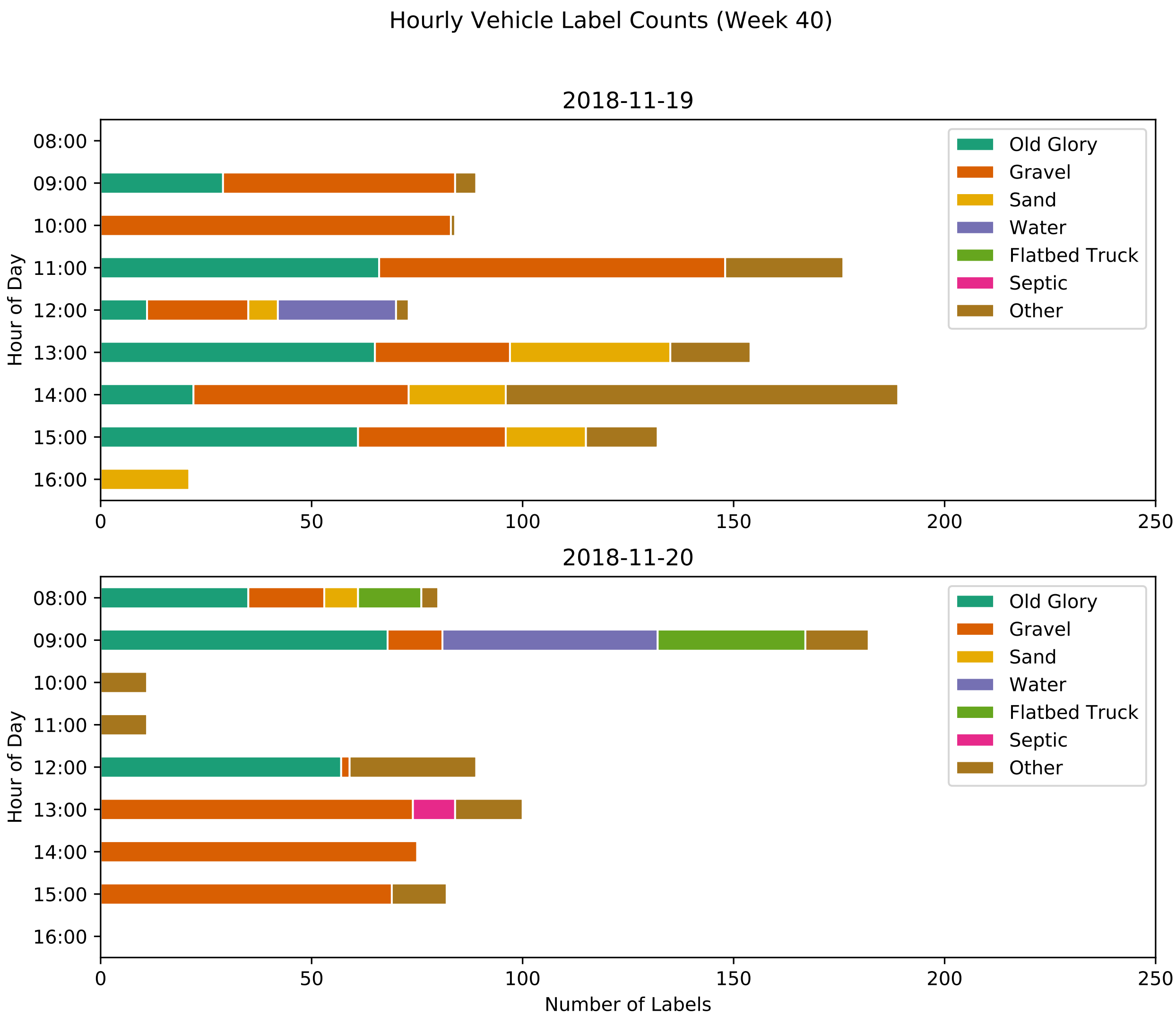


Figure 3-8. Hourly counts of vehicle movements provided a timeline of activities for a day.

As depicted in Figure 3-9, the following week was much more subdued. Multiple loads of gravel were deposited at the site while workers continued their stemming operations. The water truck sprayed the roads (even on a rainy day) and the septic truck serviced the restrooms.

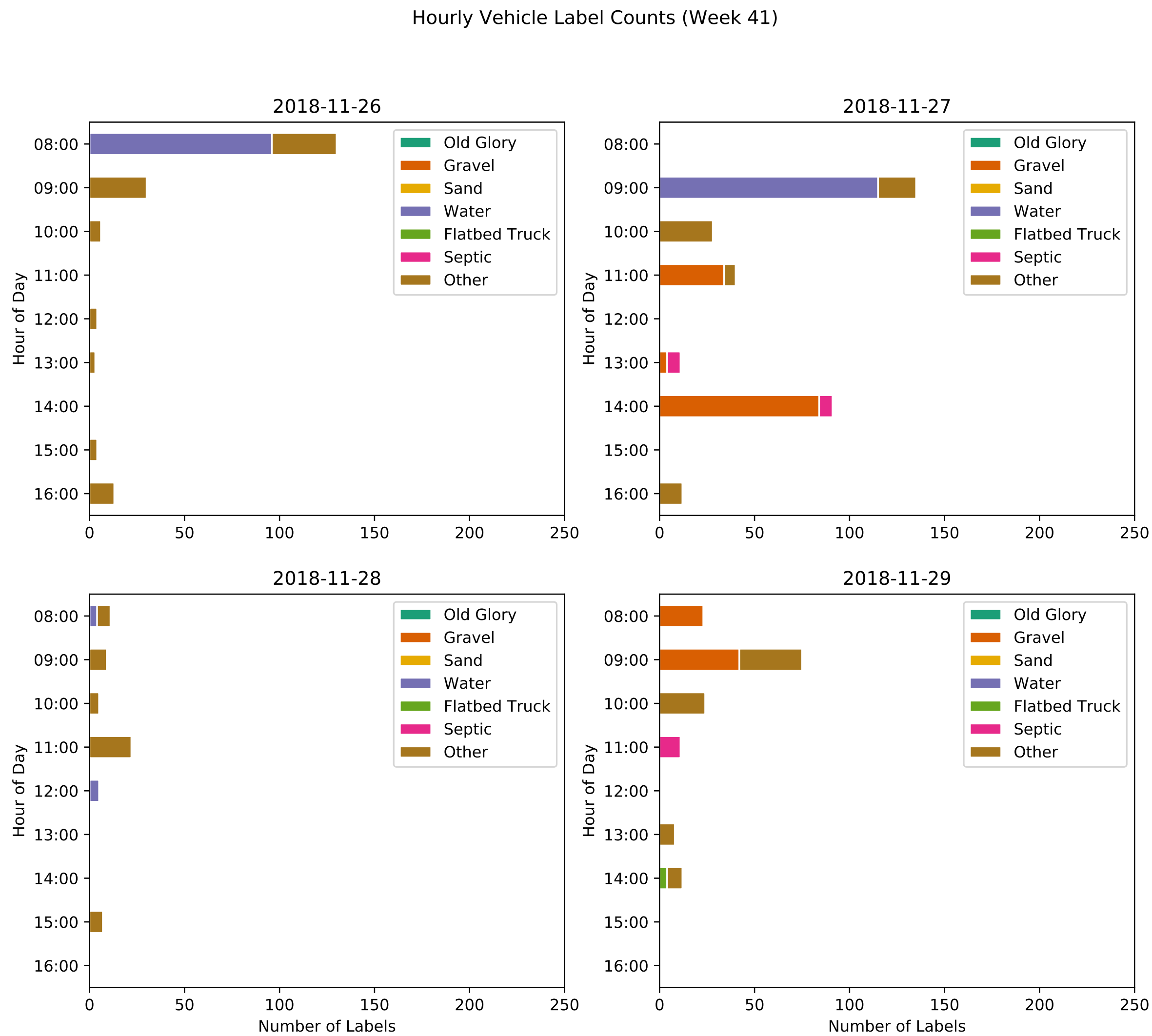


Figure 3-9. The hourly vehicle movement counts for week 41 were more subdued and focused primarily on moving fill material.

3.3. Aggregating Seismic and Video Data

It is difficult for humans to parse and understand large amounts of multimodal, time series data. In addition to the DAG seismic data covering more than a hundred days of surveillance, the seismic data was large enough that there was a noticeable overhead for reading and plotting a desired window of activity. As such, it was useful to provide visualization aids that allow consumers to rapidly scan through the data and get rough estimates of the signal content that could be representative of different activities as well as exact timing of the activities of interest. After considering multiple approaches to summarizing the videos, we decided that one of the most effective ways to make the data more accessible was simply to create an aggregate video that displayed both the surveillance video and plots from several of the seismic sensors. These videos enabled users to pan to a particular point in the timeline and get seismic time- and frequency-domain plots for that point in time.

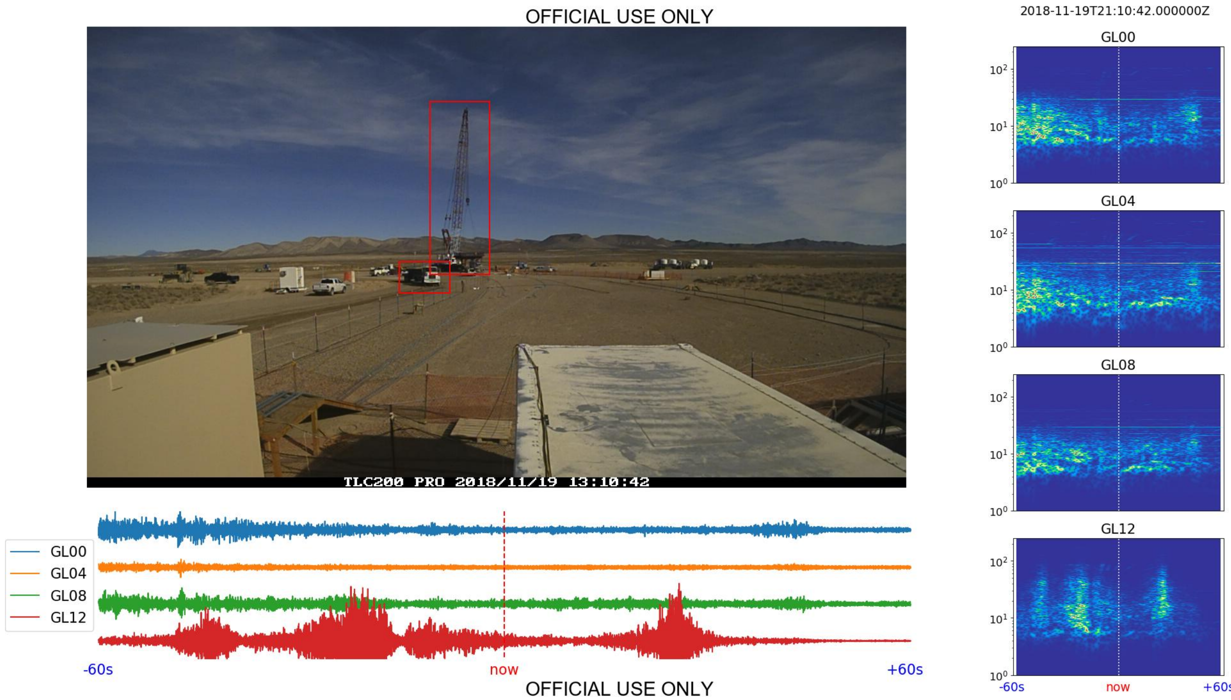


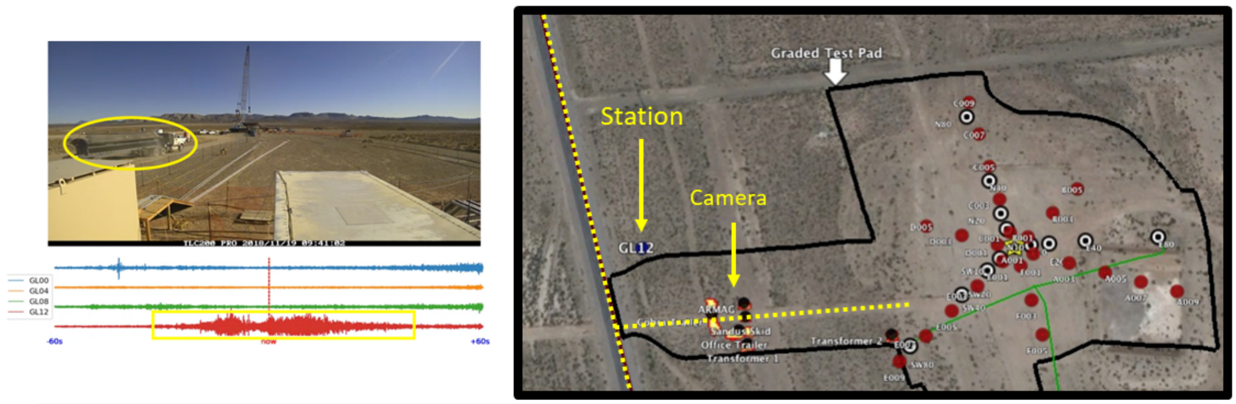
Figure 3-10. The surveillance videos and summary plots were combined in the aggregate video to provide an easy way for users to scan through large amounts of time series data.

Figure 3-10 provides a snapshot of a single frame from one of the aggregate videos. The top-left of the video contains the surveillance video and includes label boxes around large vehicles that were identified during the hand labeling. Time series plots for four geophones are presented in the bottom of the video. These plots flow to the left when the video is watched and show signals a minute before (left) and a minute after (right) the current time. Spectral plots for the four sensors are available on the right of the video with equivalent time windows. While the time series plots are useful for analysis of the seismic amplitude of an activity, the spectral plots help illustrate changes in frequency content of the signal.

Rendering the aggregate videos is time consuming due to the sheer number of frames in the video. Python scripts were constructed to automate the process and farm out the work to the compute nodes of a data-intensive cluster computer. We are currently enhancing these scripts to make them easier to use and portable to other environments.

3.4. Vehicle Ingress and Egress

Geophone GL12 was located near the intersection of a highway and the main road utilized for entering/exiting the worksite (Figure 3-10). The geophone was located behind and northwest of the eastward facing camera. Seismic energy attributed to vehicles entering and leaving the site on the marked road was easily detectable in both the time and frequency seismic plots (Figure 3-11). The aggregate video allowed for the extraction of more detailed vehicle movement time-stamps and subsequent association with the type of vehicle generating the energy (Figure 3-11). Additional details on seismic detection of vehicle ingress and egress are available in other ADAPD FY20 technical reports.



Begin	End	Movement	EquipmentType
2018-11-19T17:39:15.890914Z	2018-11-19T17:40:55.358472Z	enter	semi-trailer-gravel
2018-11-19T17:46:48.628900Z	2018-11-19T17:47:43.642177Z	exit	semi-trailer-gravel
2018-11-19T17:51:24.626386Z	2018-11-19T17:52:03.885087Z	exit	4-door-pickup
2018-11-19T17:53:04.747841Z	2018-11-19T17:54:16.622772Z	enter	semi-trailer-gravel
2018-11-19T17:58:45.733465Z	2018-11-19T17:59:35.381669Z	enter	4-door-pickup

Figure 3-11. Aggregate video screen capture displaying the utility of joint seismic/video analysis for identifying movement time-stamps and associated vehicles. The yellow box around the seismic data in the lower portion of the figure is associated with the entry of the gravel trailer circled in the upper portion of the figure.

3.5. Measuring Human Annotation Rates

A common question about ground truth work is *How long does it take a human to manually inspect surveillance video and annotate objects?* The answer depends on several factors, including the experience of the labeler, the desired amount of detail in the results, the number of objects to label, and the density of activities in the video. Our estimate for annotating a single day of video similar to the DAG surveillance video was either (1) about an hour if only a high-level description of activities is required or (2) about five hours if object position labels are required.

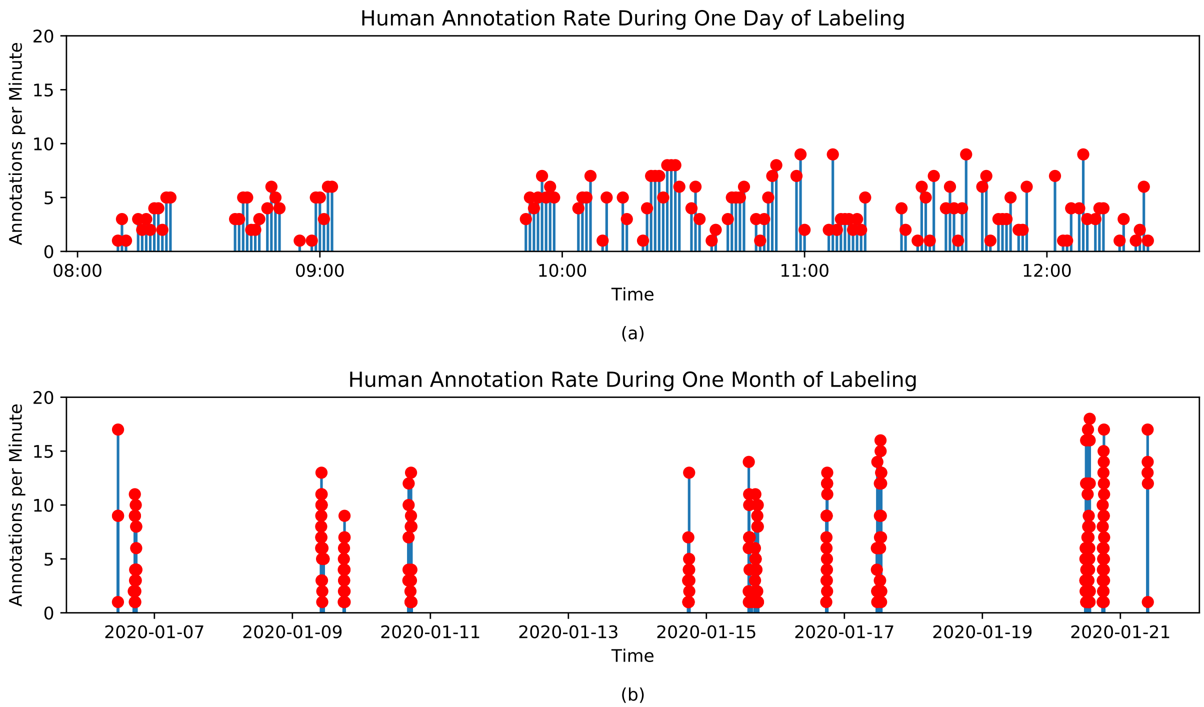


Figure 3-12. The DAG video label dataset was examined to determine how many annotations per minute a human made over the course of (a) one day and (b) one month of work. Each red circle represents the number of frames the human annotated at a particular minute in their work schedule.

As a means of estimating the effort required for labeling data, we extracted the file creation timestamps for the labels that were generated during the DAG video labeling process and measured how many frames the human annotated per minute over the course of this project. As illustrated in Figure 3-12(a-b), the DAG labeling took place in two phases: (a) an initial period in December where a single day of video was annotated quickly to get team members data for their experiments and (b) a period of gradual updates in January where additional days were annotated as time permitted. The December plot shows that a single day of surveillance video could be fully annotated by someone with little experience in about a half day of work. The labeler generated annotations at a maximum rate of 5-10 annotations per minute during this period. As illustrated in the January plot, this rate improved to as much as 17 annotations per minute as the labeler became more experienced with the data and tools.

A histogram for the annotation rates and the cumulative number of annotations for the work are presented in Figure 3-13. As expected, it was common for the labeler to generate less than 8 annotations per minute and rare to generate more than 12 annotations per minute. The cumulative plot shows that work took place in bursts, where 200-500 annotations were made per day. While the gaps in the data made it difficult to estimate the total amount of time the labeler actively worked on this task, there were a total of 388 unique minutes spread over the month of work where at least one label was generated.

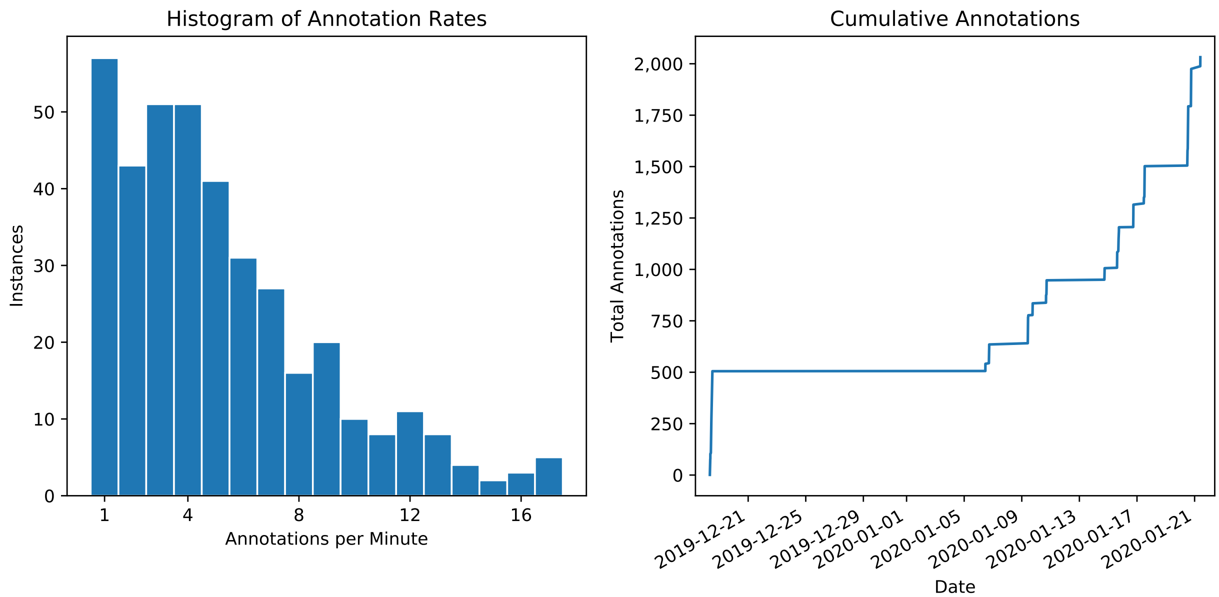


Figure 3-13. The histogram of annotation rates and cumulative number of annotations for this project reveal that labeling work was slow and bursty.

4. MISSION PURPOSE AND IMPACT

Many nonproliferation efforts analyze surveillance data to help analysts make better assessments about observed activities. While ADAPD's primary focus was to develop new analytics, improving ground truth was an essential first step in our work and a task that was worthy of examination on its own. As illustrated in this report, there are many tedious data-engineering tasks that are typically employed to improve the amount of ground truth available for a surveillance application. There are opportunities for automating many of these tasks. By documenting the steps taken in our work, we highlight techniques that improved our confidence in the data as well as provide examples of where to draw the line between using humans and using algorithms to do the work.

There are multiple analysis efforts within ADAPD that benefited from this work:

Object Detection for Timeline Generation: Mike Rivera leveraged the vehicle images extracted from the videos to serve as training data for a vehicle classifier. This data enabled Mike to customize the classifiers to recognize the specific vehicles used at the site, thereby allowing the larger collections of videos to be inspected in an automated manner.

Seismic Clustering: Erick Draayer used the labeled data for validating the seismic clustering results. He paired the labels created from the video data with seismic sequences. These labeled seismic sequences were used for validating his methodology for feature extraction and clustering on seismic data.

Activity Recognition from Seismic Signals: Fulton Wang used the ground truth on the times during which a particular activity was being performed to train classifiers that ingest a seismic signal from the sensors for a time window, and predict the probability that a particular activity was being performed during the time window. Without the ground truth provided by the video label data, it would have been impossible to train these classifiers.

Similarity Measures for Clustering Non-traditional Seismic Waves: Renee Gooding used the timestamps of the video labels to reduce the set of seismic segments that need to be clustered. The algorithm only clustered segments that occurred at least partially during a labeled video activity.

5. DISCUSSION AND FUTURE WORK

5.1. Challenges

In retrospect, there were a number of reasons why building ground truth for the DAG environment was challenging:

Surveillance Video Hindrances: While it was extremely valuable to have the surveillance video for ground truth tasks, there were a number of quality factors that made analysis challenging. First, multiple points of interest were either in the distance or out of view (e.g., all but one geophone). Second, the 10 second time interval between frames made tracking individual objects difficult due to the distance a vehicle can travel between frames. Finally, the limited amount of metadata embedded in the video stream meant we had to be skeptical of the camera's settings (e.g., orientation and time) and verify fundamental assumptions of its operation.

Locating Practical Implementations of Algorithms: A reoccurring problem in the ground truth effort was that while we knew what information we wanted to extract and the theory behind current algorithms that could extract that information optimally, we often lacked a practical tool that implemented the algorithm and worked with our data and hardware. As such, a nontrivial portion of our time was spent searching for new tools, building them, and adjusting them to work in our environment.

Boundary Between Manual and Automated Methods: As data scientists we always try to automate as much of the ground truth work as possible to minimize tedious tasks. However, we often face "one-off" tasks where the fastest path to producing results is through brute-force manual labor. For example, labeling the initial two weeks of video sequences was tedious, but we were able to get reasonable-quality results to others much more quickly than if we had spent time tuning off-the-shelf algorithms. Sequences longer than several days would not have been feasible.

Ad hoc Data Pipelines: While we would like to think of the information refinement process as a structured system where data is run through multiple analytics stages in a structured, repeatable manner, ADAPD implemented its data flow in a manual ad hoc manner. Initial and intermediate datasets were placed in a shared location where different researchers could read and write different data products as needed. This process was sufficient for ADAPD's research objectives. However, it would be useful to leverage an existing data flow framework in future similar efforts for the same reason that source control management software improves code development.

5.2. Opportunities for Future Work

Based on our experiences we see there are a number of opportunities for future work.

Super-Resolution for Resolving Background Objects: Vehicles in the background were difficult to automatically identify correctly because of their low pixel counts. Current super-resolution [8] techniques have made impressive progress [2] in developing neural networks that can scale the resolution of an image by inferring what training data suggests would normally be available in natural, higher-resolution images. It is possible that a neural network could be trained with mission-relevant imagery, thereby allowing an analyst to scale background objects before passing them to an object classifier.

Scene-Specific Training: Most object classifiers are designed to be position independent as they are expected to provide a general solution that works with a large number of variances. However, the object tracking needed in this type of surveillance work is much more specific: there are exactly two fill trucks and they never drive anywhere in the video that is not a road. For high-value surveillance sites it may be useful to build custom neural networks that specifically leverage these factors to provided a more accurate classifier that is tuned for a specific scenario.

Consider Multiple Frames in Object Tracking: Current object tracking approaches typically break the work into two separate tasks: (1) an object classifier algorithm identifies all items in a frame and then (2) a motion estimation algorithm determines where each object moved in later frames. From our experiences in hand-labeling objects in the background, we have low confidence that this approach will work well for distant objects. Often we could only locate a vehicle by looking at several frames to estimate its trajectory and then iterating between neighboring frames to infer position. We expect that an deep neural network that considers a sequence of frames could provide better results for distant objects than current approaches.

Smarter Cameras: There are a variety of ways that surveillance cameras could be improved by using embedded processors at the camera source to supplement the video stream with valuable metadata. Simple metrics such as current location, orientation, and time synchronization would simplify the process of aligning the data with other sources. Variable frame rates would enable an object to be tracked at a higher time fidelity without impeding file sizes. Finally, both change detection and object detection would help consumers locate periods of interest more quickly and could enable the camera to relay tracking information for items when the full frame rate is not recorded. Other ventures in NA22 are beginning to embed more processing at the camera source and are a significant step in the direction of sensors that produce *information* instead of just data.

REFERENCES

[1] Jesse Bonner. Defense Nuclear Nonproliferation Programs Source Physics Experiments (SPE) Dry Alluvium Geology (DAG). Technical Report DOE/NV–1596, Nevada National Security Site, North Las Vegas, NV (United States), 2018.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.

[3] Manette Fisher. Earthquake or underground explosion? Experiments help differentiate between nuclear test and natural events. `https://www.sandia.gov/news/publications/labnews/articles/2019/08-16/source_physics.html`, 2019.

[4] Jennifer Larotonda and Margaret Townsend. Data Release Report for the Source Physics Experiment Phase II: Dry Alluvium Geology Experiments (DAG-1 through DAG-4), Nevada National Security Site. Technical Report DOE/NV/03624–1220, Nevada National Security Site, North Las Vegas, NV (United States), 2021.

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[6] Joseph Redmon and Ali Farhadi. YOLOv3: An Incremental Improvement. *arXiv*, 2018.

[7] Catherine Snelson, Christopher Bradley, William Walter, Tarabay Antoun, Robert Abbott, Kyle Jones, Veraun Chipman, and Lloyd Montoya. The Source Physics Experiment (SPE) Science Plan (Version 2.4c). Technical Report LLNL-TR-654513, Lawrence Livermore National Laboratory, Livermore, CA (United States), 2019.

[8] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-Resolution Image Reconstruction: A Technical Overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003.

[9] Scott Traeger. Source Physics Experiment Phase II, Dry Alluvium Geology (DAG) Experiments Using Nitromethane. Technical Report LA-UR-16-29587, Los Alamos National Laboratory, Los Alamos, NM, USA, 2016.

[10] D. Tzutalin. LabelImg. `https://github.com/tzutalin/labelImg`, 2015.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop

Hardcopy—External

Number of Copies	Name(s)	Company Name and Company Mailing Address