Three-dimensional Object Tracking in Panoramic Video and LiDAR for Radiological Source-Object Attribution and Improved Source Detection

M. R. Marshall, D. Hellfeld, T. H. Y. Joshi, M. Salathe, M. S. Bandstra, K. J. Bilton, R. J. Cooper, J. C. Curtis, V. Negut, A. J. Shurley, K. Vetter

***Abstract*—Networked detector systems can be deployed in urban environments to aid in the detection and localization of radiological and/or nuclear material. However, effectively responding to and interpreting a radiological alarm using spectroscopic data alone may be hampered by a lack of situational awareness, particularly in complex environments. This study investigates the use of LiDAR and streaming video to enable real-time object detection and tracking, and the fusion of this tracking information with radiological data for the purposes of enhanced situational awareness and increased detection sensitivity. This work presents a novel object detection, tracking, and source-object attribution analysis that is capable of operating in real-time. By implementing this analysis pipeline on a custom developed system that comprises a static $2 \times 4 \times 16$ inch NaI(Tl) detector co-located with a 64-beam LiDAR and 4 monocular cameras, we demonstrate the ability to accurately correlate trajectories from tracked objects to spectroscopic gamma-ray data in real time, and use physics-based models to reliably discriminate between source-carrying and non-source-carrying objects. In this work, we describe our approach in detail and present a quantitative performance assessment that characterizes the source-object attribution capabilities of both video and LiDAR. Additionally, we demonstrate the ability to simultaneously track pedestrians and vehicles in a mock urban environment, and use this tracking information to improve both detection sensitivity and situational awareness using our contextual-radiological data fusion methodology.**

***Index Terms*—Source attribution, object detection, radiological search, object tracking**

I. Introduction

THE detection and localization of radiological and/or nuclear material remains a key challenge in homeland security, particularly in urban environments [1], [2]. These environments are dynamic and highly cluttered from both a physical and radiological perspective, making the detection and localization of radiological threats extremely challenging. The presence of static and dynamic objects, and the constantly evolving nature of the environment, makes the attribution of a radiological signal to a particular physical object (e.g. a given vehicle or person) particularly difficult.

Augmenting spectroscopic radiation data with contextual information, such as streaming video, can provide important situational awareness in these scenarios and help inform the interpretation and adjudication of radiological alarms. While contextual data can undoubtedly aid the analysis of radiological data by a human operator, recent advances in computer-vision based object detection now make it possible to perform radiological-contextual data fusion in real-time, and provide automatic associations between radiological signals and the physical objects in a scene. This has the potential to improve the speed and efficiency with which radiological/nuclear sources can be detected and localized in complex environments while providing significantly enhanced situational awareness to a human operator.

In this paper, we explore the application of this concept by using video and Light Detection and Ranging (LiDAR) to augment a static radiation detector. We demonstrate the use of real-time object tracking information to attribute measured radiological signals to source-carrying objects in the scene.

Previous work has demonstrated the use of video and rudimentary vehicle tracking to enhance the performance of a large gamma-ray imaging system for portal monitoring [3], [4]. More recent advances in contextual sensing, and the processing of contextual data, have enabled the exploration of object tracking and the correlation of the trajectories of objects in a scene with data from isotropic radiation detectors. Such methods have the potential to provide source localization in poorly constrained measurement environments, and without the need for large, complex gamma-ray imagers.

More recently, object tracking and radiological attribution was demonstrated using a LiDAR point cloud projected onto the X-Y plane [5]. This top-down LiDAR projection was used to identify objects in the scene, the 2D trajectories associated with these 2D objects were then correlated with the data recorded by co-located radiation detectors. This correlation was performed using a non-negative least squares method and was used to attribute the radiological data to the person carrying the source. This approach neglects static sources when performing attribution, performs object tracking only in 2D, and is unable to handle dynamic occlusions. Therefore, while they may be adequate in certain situations, these specific methods are likely to offer limited utility in complex, real-world environments.

This material is based on work supported by the US Department of Homeland Security under competitively awarded contract 16DNARI00026. This support does not constitute an expressed or implied endorsement on the part of the United States Government.

M. R. Marshall, K. J. Bilton, A. J. Shurley, and K. Vetter are with the Nuclear Engineering Department at the University of California, Berkeley, Berkeley, CA 94720 USA (email: mattmar2410@berkeley.edu).

D. Hellfeld, M. Salathe, M. S. Bandstra, J. C. Curtis, V. Negut, T. H. Y. Joshi, R. J. Cooper and K. Vetter are with the Applied Nuclear Physics (ANP) Program at Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA 94720 USA.

In this work, we demonstrate 3D object tracking and radiological data fusion using multi-camera video and LiDAR. LiDAR and video are analyzed independently and the object tracking performance from these sensors are combined with radiological signals and their performance compared. Object detection is performed using lightweight neural networks that can be run in real-time on board the sensor system, and object tracking is performed in three dimensions using a class-specific (e.g. pedestrians, vehicles) Kalman filter. Tracking in a 360-degree field of view (FOV) is achieved by using four cameras and a 64-beam LiDAR. These tools are then used to evaluate the performance of object tracking on radiological alarm attribution and optimized spectroscopic search, with the objective of demonstrating these concepts and their performance in a challenging mock urban environment. It should be noted that given the scenario we consider in this paper, a mock urban environment with minimal constraints in the 3D motion of people and vehicles and with various occluding objects, a system without video or LiDAR would be limited so we do not draw comparisons between our system and one without video or LiDAR. The attribution analysis is triggered by spectroscopic alarming using a method based on non-negative matrix factorization [6]. During the attribution analysis, models that describe the time-dependent photopeak count-rate expected in a detector from each trajectory are calculated and compared with the time-series photopeak data. A model for each trajectory includes the $1/r^2$ profile of the trajectory (r is the distance from the detector to the source), the angular detector response, and gamma-ray attenuation in air and in intervening tracked objects. An estimate of goodness-of-fit is used as a discrimination metric to perform attribution (or exclusion) between the radiological data and the model from each trajectory. Optimized spectroscopic search is performed using the same models from trajectories to identify the temporal segments for each trajectory that provide and optimize the signal-to-noise ratio (SNR), demonstrating enhanced detection sensitivity via data fusion. All of these methods are implemented in a framework that is capable of running in real-time ($\sim$ 15 Hz) on a compact multi-sensor system.

The paper is outlined as follows: in Section II, we provide a description of the object detection, tracking, and attribution analysis pipeline as well as an optimized spectroscopic search. In Section III, we detail the results of a study in which we performed a quantitative assessment of the tracking and attribution capabilities as applied to a static system using multi-camera video or LiDAR. In addition, we present object tracking and source-object attribution in a mock urban environment with both pedestrians and vehicles and show the ability to improve detection performance using tracking information. The performance of the methods and their current limitations are discussed in Section IV, and we summarize the work and consider future research opportunities in Section V.

II. Methods

Our complete analysis pipeline was implemented on a custom developed platform which is shown in Fig. 1. This system includes an Ouster 64-beam LiDAR unit [7] and four AR1335 monocular cameras [8] arranged to provide a 360-degree panoramic view around the system. A NVIDIA Jetson AGX Xavier computing platform with 512-core Volta GPU [9] is used to readout the camera and LiDAR sensor and perform on-board data processing. The system is designed as a standalone contextual sensor package (CSP) and can be used to augment a range of radiation detectors. In this work, a $2 \times 4 \times 16$ inch (in) NaI(Tl) gamma-ray detector (hereby referred to as simply the "detector") was co-located with the system. The detector was instrumented with an Ortec DigiBASE [10] multi-channel analyzer and configured to publish gamma-ray spectra at 20 Hz. The system was powered by a 296 Watt-hour (24 V) battery. Sensor readout and synchronization was handled with the Robot Operating System (ROS) [11].

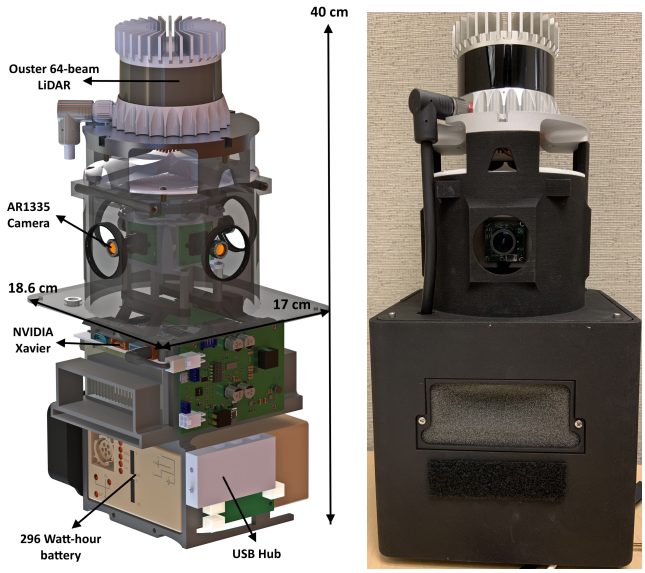


Fig. 1: Object detection, tracking, and source-object attribution system. The system consists of an Ouster 64-beam LiDAR that is mounted above 4 AR1335 monocular cameras to provide a full panoramic field of view. A NVIDIA Xavier is located in the base of the setup for on-board processing. The system is powered by a 296 Watt-hour (24V) battery.

Fig. 2 shows an example output from running our analysis pipeline for both video and LiDAR. Each process of the pipeline will be explained in more detail in the subsequent sections.

A. Object Detection and Tracking

Object detection from video frame images was performed using the open-source *You only look once* (YOLO) [12] detector and the ROS implementation, YOLO ROS [13]. In this work, we used YOLOv3-tiny - a lightweight model pre-trained on the COCO dataset [14] capable of inference times of 66 ms on our system. We note that while higher performing YOLO models exist, such as YOLOv3-608, they cannot be run in real-time on our system due to the additional computational burden and limited resources.

YOLO ROS returns identification labels for the detected objects with confidence scores and 2D bounding boxes in image coordinates (see Fig. 2a). The distance (depth) of the detected object to the camera is inferred by comparing the height of the 2D bounding box to the height of a person or a car. For a person, we use a nominal height of 1.75 meter (m) [15] and for a car, we use 1.43 m, which is what we consider to be an average size of a compact car. Heights of 1.80 m, 0.80 m, and 2.5 m are used for labels of truck, motorcycle, and bus, respectively. The 2D bounding boxes are mapped to 3D using the camera intrinsic parameters and the inferred object depth.

A separate object detection process is performed for each of the four cameras on the system. The four camera frames are synchronized, and the object detections from all four cameras are collated. This is done to avoid double-counting detections that take place in regions where the camera images overlap.

The object detection in LiDAR point clouds was performed using the sparsely embedded convolutional neural network (CNN) detection framework SECOND [16] with the PointPillars fast feature encoding [17]. A multi-class detector model was trained on the nuScenes dataset [18] (32-beam LiDAR spinning at 20 Hz) developed for autonomous driving research. In accordance with SECOND, the model was trained on point clouds generated from 10 LiDAR scans to achieve higher point density for inference. It is worth noting explicitly that because the LiDAR provides data which is inherently three-dimensional, the object detection algorithm is able to return 3D bounding boxes, removing the need to infer the distance to the object that is necessary in the case of video.

In the application presented here for our static system, two scans of the Ouster 64-beam LiDAR generated sufficiently dense point clouds without significant motion blurring from dynamic objects. The inference time on the NVIDIA Xavier GPU is sufficiently fast ($\sim$125 ms) to enable the object detection to keep up with the point cloud data (200 ms). The LiDAR object detection returns labelled 3D bounding boxes with associated confidence scores using the same ROS message format as the video results (see Fig. 2b).

To convert sequences of detections into object tracks, we used the Kalman-filter-based Simple Online Realtime Tracking (SORT) algorithm [19]. SORT was extended to process both video and LiDAR detections in three-dimensions. We further modified SORT by replacing the method used to associate detections with the most likely corresponding track. The conventional intersection-over-union (IOU) method was replaced with a multivariate normal (MVN) representation of the 3D bounding boxes and the Hellinger Distance (HD) in order to provide a more statistically founded approach which is capable of accounting for variable uncertainties in object extent. The HD is a measure of similarity between two MVNs that is scaled between 0 and 1, where 0 represents high similarity [20].

To create the MVN representation of a 3D bounding box, the center of the bounding box is converted to the mean of the MVN, and scaled dimensions of the bounding box populate the covariance matrix for position uncertainty. Different scaled dimensions are used for video and LiDAR 3D bounding boxes. In the modified SORT algorithm, we employ a multivariate Kalman filter to track two state variables: position (x, y, z) and velocity (v_x, v_y, v_z). The diagonals of the velocity uncertainty in the covariance matrix are populated according to the detection label. For vehicles, we use 4.44 m^2/s^2 and 0.28 m^2/s^2 for pedestrians; thus, enabling the Kalman filter to simultaneously track both pedestrians and vehicles. Data association between the detected objects and the most likely corresponding track is performed using the HD. Detections with a calculated HD of less than 0.8 are consolidated to a single track. The velocity uncertainties and HD thresholds for pedestrians and vehicles were found by running an optimization over those variables on a scene only containing either pedestrians or vehicles. The objective was to reduce the number of objects with disjoint tracks while also ensuring a low number of incorrect associations by the data association algorithm.

In order to track objects across multiple cameras, the MVNs are transformed from the image frame into a global coordinate frame prior to computation of the HD. Small FOV overlaps exist between the four cameras and thus it is possible for multiple cameras to independently detect the same object. However, once the MVNs are transformed into the global frame, these bounding boxes will overlap. To prevent the same object from being tracked more than once, these overlapping bounding boxes are merged if the calculated HD is less than 0.6. This merging of bounding boxes supplements the nonmax suppression being performed with YOLOv3-tiny in each independent camera frame.

The tracker output is presented to an operator using the object's label (e.g., person or vehicle) and a track ID, where the track ID increases by 1 for each newly generated track. An example is shown in Fig. 2c and Fig. 2d for both video and LiDAR, respectively. The bounding boxes displayed are MVN representations of bounding boxes and will no longer match the bounding boxes output from the object detection CNN. Also, it should be noted that an object's orientation is not considered so all the bounding boxes will have the same orientation.

Given the nature of the object detectors, false detections can occur so the track ID will normally be higher than the total number of objects that have been present in the system's FOV. Additionally, untracked objects and detections not being assigned right away to the correct track contribute to a higher total track ID number. To limit the number of tracked objects, a track is discarded from the Kalman filter if the track is not associated with a detection in 5 consecutive frames, and a new track will be spawned for the object if it is detected again.

B. Modeling and Fitting Radiation Data to Trajectories

The source-object attribution process is invoked when a radiological anomaly is observed. The presence of an anomaly is determined by the Berkeley Anomaly Detection (BAD) algorithm, a spectroscopic detection algorithm based on Non-negative Matrix Factorization (NMF) [6]. BAD provides high sensitivity detection, even in low signal-to-background regimes and serves as a robust trigger for the source-object attribution

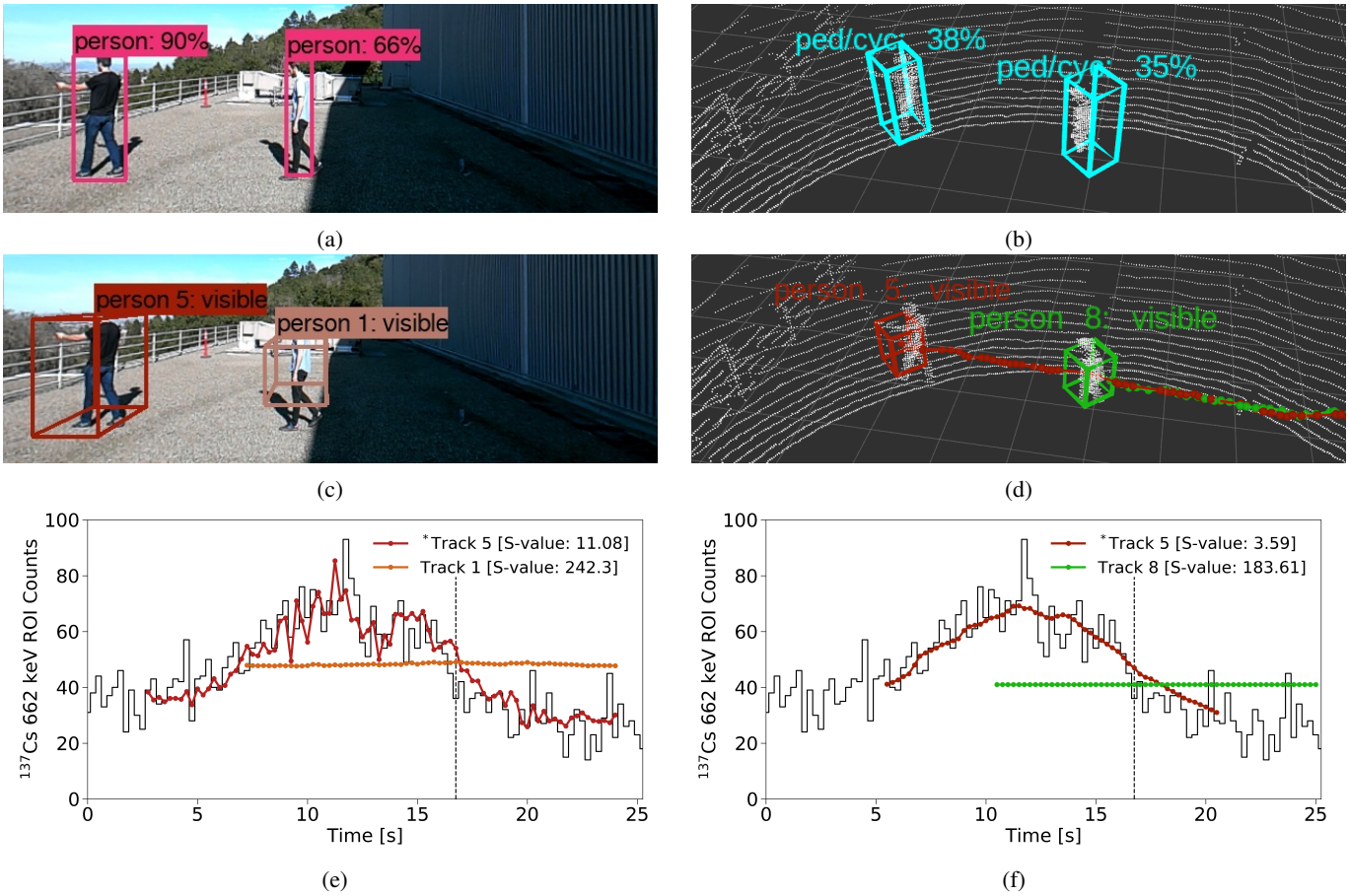


Fig. 2: The object detection, tracking, and attribution pipeline for both video ((a), (c), and (e)) and LiDAR ((b), (d), and (f)). The bounding boxes shown in (a) and (b) are the object detection CNN outputs for video and LiDAR, respectively, with the object label and confidence score. In (b) "ped/cyc" refers to pedestrians or cyclists. The LiDAR CNN groups both labels into the same category. The bounding boxes displayed in (c) and (d) are the object tracking outputs, and the bounding boxes are MVNs converted back to bounding boxes and will not match the bounding boxes in (a) and (b). In (d) the object's trajectory up to that point is indicated by the line proceeding each respective bounding box. In (b) and (d) the white grid squares represent $1\,m^2$, and the area of no points exist from the field of view of the LiDAR. In (e) and (f) the best-fit models for each trajectory from (c) and (d), respectively, to the count-rate data are displayed. The pink bars represent the time interval for a radiological alarm, and the count-rate from the $2 \times 4 \times 16$ in NaI(Tl) detector is shown in black. Additionally, the dashed lines in (e) and (f) depict the moment in time these images were taken from. See supplementary material for the full video.

algorithm. This algorithm also provides an isotope ID which is used to define a spectral region of interest (ROI) for the extraction of time-series count-rate data. The source-object attribution analysis is performed on all trajectories that are within 7 seconds of the start and stop of the radiological alarm.

The first step in the source-object attribution process is the generation of a set of models that describe the time-dependent count-rate that would be expected to be observed were each track to be associated with the detected anomaly. To model the expected counts in the detector, we use the following approach. For a given discrete time step, i, the expected number of detected events, c_i, within a spectral ROI, E, from a radioactive source with gamma-ray flux α in the presence of a constant background b can be described by

$$c_i(E) = \frac{\epsilon(\hat{\Omega}, E)\alpha e^{-\mu(E)\mathbf{r}_i}}{4\pi\mathbf{r}_i^2} \cdot \Delta t_i + b\,, \tag{1}$$

where ϵ is the effective area of the detector, $\mathbf{r}_i$ is the distance from the detector to the source, Δt_i is a given integration time, and μ is an energy and medium dependent linear attenuation coefficient. The effective area is a function of energy and the direction between the tracked object's position and the detector, $\hat{\Omega}$. In our current implementation, ϵ takes the form of a pre-computed response matrix which is used as a look-up table based on the spectral ROI and the direction between the object and the detector at any given time. Details of the process through which this response matrix was generated are provided later in this section.

The best-fit model for each trajectory is found by extracting the time-series ROI count-rate in a time window around the radiological alarm, and maximizing the Poisson likelihood between Eq. 1 and the observed count-rate data with a maximum likelihood estimation algorithm [21], where α and b are free parameters. For the example in Fig. 2, the images in panels

Fig. 2e and Fig. 2f show the best-fit models for the tracks extracted from the video and LiDAR data, respectively, along with the measured count-rate.

The directional response matrix that describes the detector's energy and effective area was generated using a Geant4 simulation [22]. The matrix was populated by modeling the $2 \times 4 \times 16$ in NaI(Tl) detector in air and, for each of several discrete energies, placing a gamma-ray source at 3072 equal-area discretized pixels on a sphere [23] at a constant distance of 10 m (R^2_{sim}) from the detector. At each source position, the total number of counts within the relevant ROI (X_{cnts}) was computed. For a given energy and source position (i.e., direction), ϵ is given by

$$\epsilon = \frac{4\pi \cdot R^2_{\mathrm{sim}} \cdot X_{\mathrm{cnts}}}{N_{\mathrm{SimParticles}}}, \tag{2}$$

where $N_{\mathrm{SimParticles}}$ is the total number of particles emitted into 4π.

These directional response simulations were previously validated against experimental measurements made using a ^{137}Cs source placed at several positions around the detector [24].

The total attenuation coefficient, μ, describes the combination of attenuation in air and in occluding objects in the scene. Before this can be applied, it is necessary to determine whether a given object track is subject to occlusion from another object. To do this, the MVNs are converted back to 2D bounding boxes, and the IOU is calculated. If the IOU is greater than 90%, an object in the scene is considered to be obstructing another object. To determine which object is occluded, the distance from the system to each object is calculated using the pose positions from the Kalman filter and the object farthest from the system is considered to be the one subject to occlusion. However, when objects pass within 1 m or less of each of other, their Kalman filter pose positions can overlap, especially using video trajectories where depth is inferred, making it difficult to accurately determine which object is occluded in the scene. This is a limitation of the current attenuation approach.

The total attenuation imposed by the occluding object is then determined by its size and an estimated linear attenuation coefficient which is based on the label associated with the object. For example, pedestrians are modeled using the average elemental composition of a human (65% Oxygen, 18.5% Carbon, 9.5% Hydrogen, 3.2% Nitrogen, 1.5% Calcium, 1% Phosphorus) [25] while vehicles are modeled as hollow Aluminum boxes with 2 in thick sides. In the current implementation, additional attenuation from other components of the vehicle are not modeled and the orientation of the vehicle is not accounted for. This results in a slight underestimate of the total attenuation in some situations.

Once the attenuation has been calculated, the size of the bounding box is subtracted from the object's total distance from the system, and the remaining distance to the detector is considered to be air. This process is then repeated for each object in the image frame or LiDAR scan. This allows the model to account for cases in which a given object is subject to occlusion by multiple objects. It is important to note that if an object is occluded for more than 5 frames, a new track will be spawned for the object when it is detected again.

Accounting for the effect of occlusion when computing the best-fit models improves the source-object attribution performance by allowing tracks which would otherwise be degenerate, e.g., those which overlap closely in time and space, to be distinguished. A limitation of the current approach is that it accounts only for attenuation in tracked objects. Including attenuation in the environment would likely improve the source-object attenuation power in particularly complex scenes.

C. Attributing Trajectories to Radiological Data

The final step of the attribution analysis identifies the trajectories that are unlikely to be associated with a radiological alarm through the use of an estimated goodness-of-fit metric. The Poisson deviance or log-likelihood ratio statistic [26] was selected as the appropriate statistic for estimating goodness-of-fit. Its distribution is approximated by using the best-fit model to estimate its first three moments and adjusting the moments according to the number of model parameters (e.g., the mean is decreased by 1 for every model parameter). A shifted Gamma distribution that matches these moments was then used to calculate a p-value from the deviance statistic given this distribution. Finally, an S-value [27] was calculated from the p-value and used hereafter, and the S-value enables principled rejection of trajectories that are inconsistent with the data.

The computation assumes that the source is located at the Kalman filter pose estimate. If the source is not located at this position, such as in the case of a source located in the trunk of a vehicle when the pose estimate is at the center of the vehicle, then an offset between the best-fit model and the photopeak count-rate could occur. To account for this potential offset between the assumed and true source location, the best-fit model is calculated multiple times over a two second window while the track is shifted in 0.1 second increments. The Poisson S-value is computed each time and the lowest value is used to determine the best fit. It should be noted that this only accounts for source offsets in the direction of travel, but does not account for potential source offsets in the other two dimensions that could exist.

D. Track-informed Signal-to-Noise Ratio Optimization to Improve Detection Sensitivity

With knowledge of an object's trajectory, it is possible to identify a combination of time-segments that, when combined, maximize the expected signal-to-noise ratio (SNR) for spectral analysis of that object. For this analysis we consider the SNR for a trajectory, present across N discrete time windows, as,

$$\mathrm{SNR} = (\sum_{i}^{N} s_i \Delta t_i)(\sum_{i}^{N} b_i \Delta t_i)^{-1/2} \tag{3}$$

where $s_i = \epsilon(\theta, E)\alpha e^{-\mu(E) r_i}/(4\pi r_i^2)$ is the photopeak count-rate within a fixed integration window Δt at a point in time i, and b_i is the background rate within Δt_i. From this equation it

can be seen that SNR may be maximized by considering only a subset of time-segments ($T \in [1, N]$). Since the radiological sources we want to detect are long-lived (i.e., half-lives on the order of hours or greater) and our analysis system is static, we introduce the reasonable assumptions that source strength and mean background are constant in time, and we describe the sensitivity ($\$_T$) which is proportional to the SNR_T;

$$\mathrm{SNR}_T \propto \$_T = \frac{\sum_{i \in T} \frac{\epsilon(\hat{\Omega}, E)\Delta t_i}{4\pi r_i^2} e^{-\mu(E) r_i}}{\sqrt{\sum_{i \in T} \Delta t_i}}. \quad (4)$$

Eq. 4 may then be used to solve for the subset of measurements T, that maximizes $\$ and, in turn, maximizes SNR for the trajectory in question. We note two observations about this formulation. First, the optimal time windows do not need to be contiguous. Second, under the assumption of constant source strength and background, the optimal integration windows are independent of source strength and background and are entirely determined using tracking information.

To test this concept, we apply a spectroscopic analysis to either fixed integration windows (1.0, 2.0, 3.0, 4.0, and 5.0 seconds), a common analysis approach, or to each trajectory's optimized integration window. The spectroscopic analysis computes an anomaly value, the Poisson deviance between the observed data and a mean background spectrum scaled to match the observed counts. We then compare the magnitude of the anomaly value for the different integration windows, a larger anomaly value for the optimized integration window suggests improved detection sensitivity through this track-informed analysis. We present this comparison as a proof-of-concept and note that quantitative evaluation of increased detection sensitivity would require appropriate handling of anomaly thresholds, which must vary with the number of statistical tests performed to maintain a fixed False Alarm Rate (FAR). Such a quantitative analysis would, in turn, require definition of strategies for when to used fixed integration windows (because tracking is not perfect) and when to use track-optimized windows.

III. Results

In the following section we show experimental results using the methods for object detection, tracking, and attribution previously described. First, we evaluate the ability of the attribution analysis to discriminate degenerate tracks by determining the minimum separation between two tracks (in the worst case scenario of identical time of closest approach and velocity) required to allow correct identification of the trajectory associated with the radiological alarm. We then demonstrate the performance of the source-object attribution analysis in a mock urban environment. Finally, using this experimental data, we demonstrate the potential for tracking-enhanced detection sensitivity by showing spectroscopic anomaly values can be increased using track-informed integration windows.

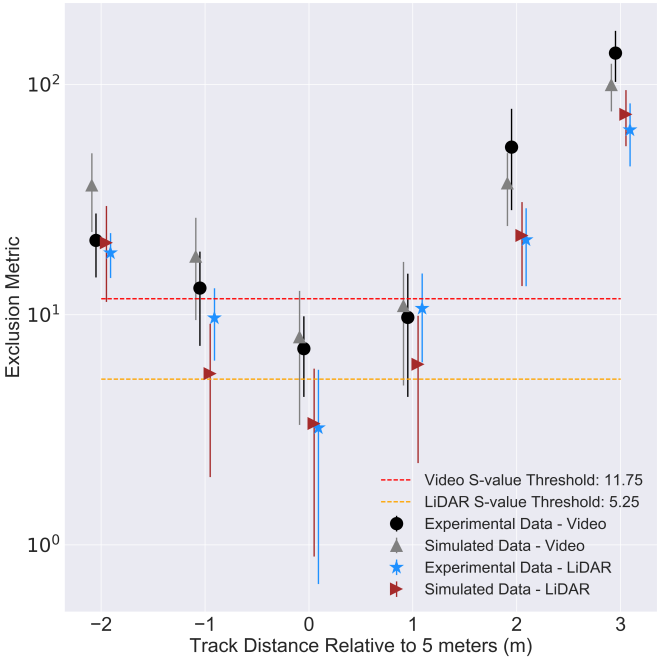


Fig. 3: Discrimination of orthogonal trajectories in the alarm attribution analysis relative to 5 m with both LiDAR and video trajectories. The simulated data represents synthetic trajectories and count-rate data that was randomly sampled for 1000 trials and subjected to the same analysis as the experimental data. The large uncertainties in the synthetic trials reflects the sensitivity of the goodness-of-fit metric. The exclusion metric is the S-value.

A. Track Discrimination in Source-Object Attribution

Most encounters with a radioactive source traveling along a straight trajectory will result in count-rate distributions as a function of time that mimic $1/r^2$, the falloff of detection efficiency with the square of the source distance from the system, and will exhibit the highest count-rate at closest approach. The worst case situation from an attribution perspective are two objects travelling at identical speeds and reaching the closest approach distance simultaneously. However, even slight variations in closest approach distance and the angular sensitivity of the detector prevent full degeneracy. To judge the performance of the attribution pipeline in this type of scenario, a systematic study was performed using a 189 μCi (6.993 MBq) ^{137}Cs source to determine the distance needed between two nearly degenerate tracks to correctly identify the trajectory associated with the radiological alarm. In this analysis occlusion/attenuation from objects was not present/considered because the trajectories for this study were created by a single person walking in front of the system. Radiation data and object trajectories were collected from 3-8 m distance of closest approach in 1 m increments. For each source distance, the ^{137}Cs source was walked past the system at a walking speed of approximately 0.30 m/s, and at each distance, the person carrying the source walked in front of the system 8 times. This analysis was done by using the photopeak ROI (614 keV - 685 keV) for ^{137}Cs.

The radiation data and trajectory from the 5 m source distance at closest approach were chosen as reference. The trajectory at 5 m was replaced with the trajectories from 3, 4, 6, 7, and 8 m to determine how well these trajectories would correlate with the radiological data produced by the 5 m trajectory. The peak count-rate from each trajectory was aligned temporally with the radiation data at 5 m to ensure the trajectories were compared in the most degenerate scenario.

To supplement the limited statistics of these measurements, synthetic count-rate data and trajectories were generated. At each distance from the system, the mean and standard deviation of the position uncertainty for each trajectory was calculated as a function of both distance and angle from the system. These experimentally measured position uncertainties were used to noise simulated trajectories. Count-rate data were sampled according to Poisson statistics from the expected number of detected events calculated through Eq. 1. 1000 trials of this random sampling process were then subjected to the same analysis as the experimental data.

The Poisson S-values from analysis of these experiments, comparing 5 m count-rate data with different trajectories, are shown in Fig. 3. We find that the experimental trajectories from video and LiDAR are both within one standard deviation of each other and their synthetic trials.

S-value thresholds, 11.75 for video and 5.25 for LiDAR, were defined based on a 95% true positive rate for 5 m synthetic trajectories with the 5 m synthetic count-rate data. Under these criteria, degeneracy is successfully broken 45% (58%) of the time for video (LiDAR) trajectories with 1 m difference from the true trajectory. When there is a 2 m separation, we find degeneracy to be successfully broken 99% (99%) of the time for video (LiDAR) trajectories. These results demonstrate the ability of an object tracking and alarm attribution analysis, using either video or LiDAR, to be quite robust against track degeneracy.

An additional phenomena that modulates the radiological signal is attenuation, and in dynamic scenes one would expect objects occluding a particular object to also attenuate a radiological signal associated with that object. This could help to break the degeneracy in tracks, particularly, when two objects cross in front of the detector system. Synthetic data were used to study how occlusion, in conjunction with 3D object tracking, might further improve attribution in scenarios where degenerate tracks are present.

For this analysis, two objects with similar stand-off distances and walking velocities were considered, such that the two objects crossed paths at the distance of closest approach. In each trial, the source-carrier was 5 m from the system at closest approach, and the non-source-carrier was either 6 m or 7 m from the system. To simulate two objects crossing paths, synthetic detections were generated, and random noise based on the derived track uncertainties from LiDAR or videos was applied. The detections were then fed into the MVN SORT algorithm to create tracks. This procedure accounts for the data association algorithm potentially switching track IDs when the two objects cross paths. The count-rate data was simulated based on the object at 5 m using the same approach as outlined in the previous section. 1000 trials were simulated for both when the source-carrier was at 5 m and the non-source-carrier was at 6 m and when the source-carrier was at 5 m and the non-source-carrier was at 7 m.

The attribution analysis was again applied to this synthetic data, with the inclusion of an attenuation factor in the response calculation. The applied attenuation factor accounted for either air or people. The attenuation factor for people was applied for portions of trajectories where objects were being tracked but occlusion was occurring. Using the S-value thresholds of 11.75 (5.25) for video (LiDAR), degeneracy was considered successfully broken if the source-carrier at 5 m had a S-value score below the S-value threshold while the other trajectory at 6 m or 7 m was above the same threshold amount. The results of this analysis show that when the two objects are separated by 1 m, degeneracy is successfully broken 39% (69%) of the time for synthetic data with track uncertainties derived from video (LiDAR). When the distance between the two objects is 2 m, degeneracy is broken 82% (96%) of the time for video (LiDAR).

In theory, by accounting for the effect of occlusion in the attribution analysis, the ability to break degeneracy should increase. However, here we observe that the results are only better for LiDAR at 1 m separation. In the other cases, by including MVN tracking through SORT, the position uncertainties sometimes lead to wrong data associations or disjoint tracks, which negates the gains from occlusion and results in worse performance. While the results, including attenuation, are not better, they provide a more realistic picture of crossing scenarios, where often, it is challenging to track through an occlusion without losing or misidentifying the two tracks, relying on only HD as a metric. Without this problem, occlusion should be an effective tool to deal with track degeneracy, but disentangling the two effects is beyond the scope of this publication.

B. Source-Object Attribution in a Mock Urban Environment

To evaluate the system's source attribution performance for encounters with multiple vehicles and pedestrians simultaneously, data was collected at the University of California's Richmond Field Station (RFS) in a mock urban environment. The contextual system, co-located with the detector, were placed at one corner of an intersection containing traffic lights and crosswalks. The detector was oriented vertically with the 4×16 in face towards the center of the intersection. Long-dwell measurements, spanning $\sim$30 minutes, were performed with dynamic pedestrian and vehicle traffic with vehicle speeds ranging from 0-9 m/s. During data collection a 1.87 mCi (69.19 MBq) ^{137}Cs source was placed in the trunk of a vehicle (white station wagon) and was shielded by 1.5 cm of lead reducing the activity of the ^{137}Cs source to 318 μCi (11.78 MBq). The following analyses are associated with individual source encounters triggered by the spectroscopic alarm.

An example of a single alarm encounter, showing multi-camera object tracking and attribution analysis results, is shown in Fig. 4. The colors used for each trajectory are consistent across the sub-figures. Fig. 4a shows images from

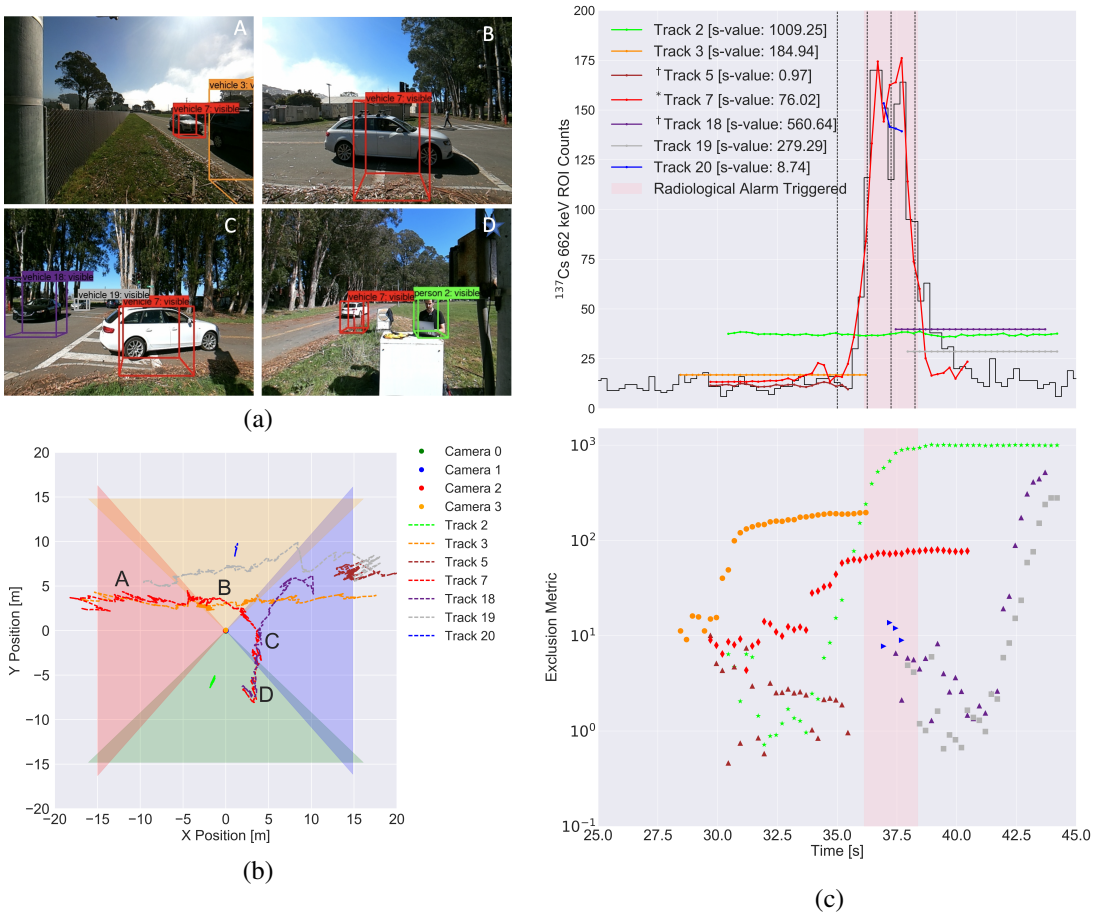


Fig. 4: Source-object attribution in a mock urban scene with both pedestrians and vehicles using camera data. The progression of the scene as the vehicle - Track 7 (the white station wagon) - carrying the radiological source moves past the system is shown in (a). The associated trajectories for each object in the scene with each camera's FOV overlaid on top of it is displayed in (b), and the different letters correspond with (a). In (b), Track 20 corresponds with the pedestrian in the background of (a) image B. Track 5 and Track 18 are examples of disjoint tracks, and both tracks correspond with the black car in (a) image C. The black car is not tracked continuously through the occlusion temporarily caused by Track 3. The top plot in (c) is the associated best-fit models of each object in the scene to the ROI count-rate data from the $2 \times 4 \times 16$ in NaI(Tl) detector, which is shown in black, and the dashed lines correspond with (a) starting with the upper left image, then upper right image, then lower left image, and finally the lower right image. The * indicates the source-carrier, and the † indicates the tracks were generated by the same object. The bottom plot in (c) displays the exclusion metric as a function of time for each trajectory. The full video is available in the supplementary material.

the four cameras (at different times throughout the alarm encounter), overlaid with the 3D bounding boxes of the labeled object trajectories, illustrating the consistent tracking of the white vehicle through the encounter. A top-down view of the object trajectories during this encounter are shown in Fig. 4b, with the FOV of the cameras also shown. The noise in the trajectories is expected given the heuristic approach to depth estimation used for visual tracking.

The best-fit results from the attribution analysis and the conclusion of the encounter are shown in the top plot of Fig. 4c. The majority of tracks are clearly excluded as having an association with the alarm, but three tracks (Track 7 (white vehicle), Track 20 (pedestrian in the background of Fig. 4a image B), and Track 5 (black car)) show consistency with the radiological signals. Of these, the source-carrier (Track 7) is tracked for an extended period of time, while the other trajectories were not continuously tracked throughout the alarm encounter and cannot be excluded.

The temporal evolution of the attribution analysis scores (S-values) are shown in the bottom plot of Fig. 4c. Short trajectories often have little modulation of $1/r^2$ and are thus likely to result in a model that fits the radiological data. Furthermore, before and after the alarm window, tracks are associated mostly with flat background, which, with α set to zero in Eq. 1, can always be well described by the model. However, continuous and consistent tracking of objects throughout an alarm encounter allows for effective attribution (or exclusion).

In this encounter, it is seen that the total number of objects present (6) does not correspond with the total track ID number,

which is 20 by the end of the alarm encounter. This is caused from both a high number of untracked objects and detections not being assigned right away to the correct track. Short trajectories, which we define as 5 or less Kalman filter poses, are more easily fit to the count-rate data and therefore, cannot be excluded from the analysis. In an effort to reduce false positives in the attribution analysis, the attribution analysis is only performed on tracks that have 3 or more Kalman filter poses.

In Fig. 5a we show the results from the same source-encounter, but now using LiDAR tracking. The color-coding of the tracked objects (but not their labels) has been kept consistent with Fig. 4. The accurate depth estimates from the LiDAR result in substantially smoother object trajectories. In addition, it is seen again that the track ID's are higher than the total number of objects present in the scene, and this is due to false detections from the object detection CNN and detections not being assigned right away to the correct track.

Furthermore, the LiDAR object detector is more robust at detecting objects further away from the system and tracking through temporary occlusions compared to the visual object detector. In particular, Track 8 was continuously detected and tracked throughout the scene with the LiDAR; whereas, the same pedestrian using video data (Track 20, the pedestrian from the background of Fig. 4a image B) was not continuously detected or tracked. Additionally, Track 44 was tracked through a temporary occlusion from Track 3 driving straight, but with the video-based trajectories, this same object (Track 5 and Track 18 which are both associated with the black car from Fig. 4a image C) has disjoint tracks from the temporary occlusion. The improved tracking from LiDAR allowed both this pedestrian and vehicle to be rejected as source-carriers, which was not the case in video. As seen in Fig. 5b, by continuously tracking objects, the source-object attribution analysis is able to take advantage of the full dynamics in the scene to exclude trajectories that do not correlate with the count-rate data.

In this scenario, the best-fit model (Track 50 - white station wagon from Fig. 4a) clearly follows the count-rate data observed in the detector. This attribution is correct as the ^{137}Cs source was located in the trunk of Track 50. The time-dependence of the source-object attribution metric is shown in Fig. 5c. The continuous tracking of object through LiDAR allows rapid exclusion of all tracks except Track 50 by the time of closest approach (38 sec).

Though there is a clear correlation for both Track 7 and Track 50 in video and LiDAR, respectively, the S-values are larger than the statistically motivated thresholds for association defined in Sec. III-A. This result is most likely driven by using a simple physics model within a scenario that contains anisotropic shielding/attenuation, which is not included in the model.

Fig. 6 shows the results from all of the alarm encounters at RFS for both video and LiDAR. The exclusion metric for both video and LiDAR show large variability between trials mainly due to the use of the simple physics model. Nonetheless, with LiDAR trajectories, the source-carrier was still assigned the lowest S-value score in 21 out of 26 alarm encounters. In the 5 encounters where the source-carrier could potentially be excluded as the object responsible for the radiological alarm, there is a clear correlation between the count-rate data and the best-fit model for the source-carrier, and an operator monitoring the system in real-time would be able to sift through the trajectories present to identify the most likely trajectory for alarm adjudication.

Using video trajectories, the results show a large number of tracks that have 5 or fewer poses in a majority of the alarm encounters. These short tracks are caused by YOLOv3-tiny's inability to consistently detect objects, especially pedestrians, more than about 10 m from the system. This results in short trajectories that are more easily fit to the count-rate data and therefore, cannot be excluded from the analysis. In addition, inconsistent tracking allows tracks to only associate mostly with flat background, which will be well described by the model since α will be set to zero in Eq. 1; thus, preventing these tracks from being excluded from the analysis. However, in 16 out of the 26 alarm encounters an operator monitoring the system would be able to correctly identify the trajectory responsible for the radiological alarm. In these 16 cases, there was a clear correlation between the source-carrier and the radiological data. Of these 16, there were 5 alarm encounters where the source-carrier had both a clear correlation and the lowest S-value score. There were 10 alarm encounters out of the 26 where a clear correlation did not exist because the source-carrier had disjoint tracks from either inconsistent tracking or the data association incorrectly assigning the wrong track id to the source-carrier when it passed closely by another vehicle.

These results, collected in the mock urban environment, demonstrate that situational awareness can be improved for both LiDAR and video using our source-object attribution analysis. In the majority of alarm encounters for both LiDAR and video, a connection between the object carrying the radiological source and the radiological data could be made.

C. Improved Detection Sensitivity with Track-informed Optimized Integration Windows

In addition to enabling automated identification of pedestrians and/or vehicles correlated with radiological alarms, object tracking has the potential to enable increased detection sensitivity through track-specific time integration windows. As discussed in Section II-D, using a model encapsulating the geometric and detector response associated with a trajectory, one may identify the selection of time-segments that should optimize SNR under the assumption of isotropic source emission and constant background. This optimum integration window analysis was performed on the same alarm encounters in the mock urban environment as the previous analyses and compared with fixed integration times.

In Fig. 7 we show the results for a single alarm encounter, showing the maximum relative anomaly value, normalized against the maximum anomaly value with a 1.0 second integration time, for all tracks in the scene as well as different fixed integration times. The trajectories used for this analysis correspond with the trajectories from Fig. 4 and Fig. 5,

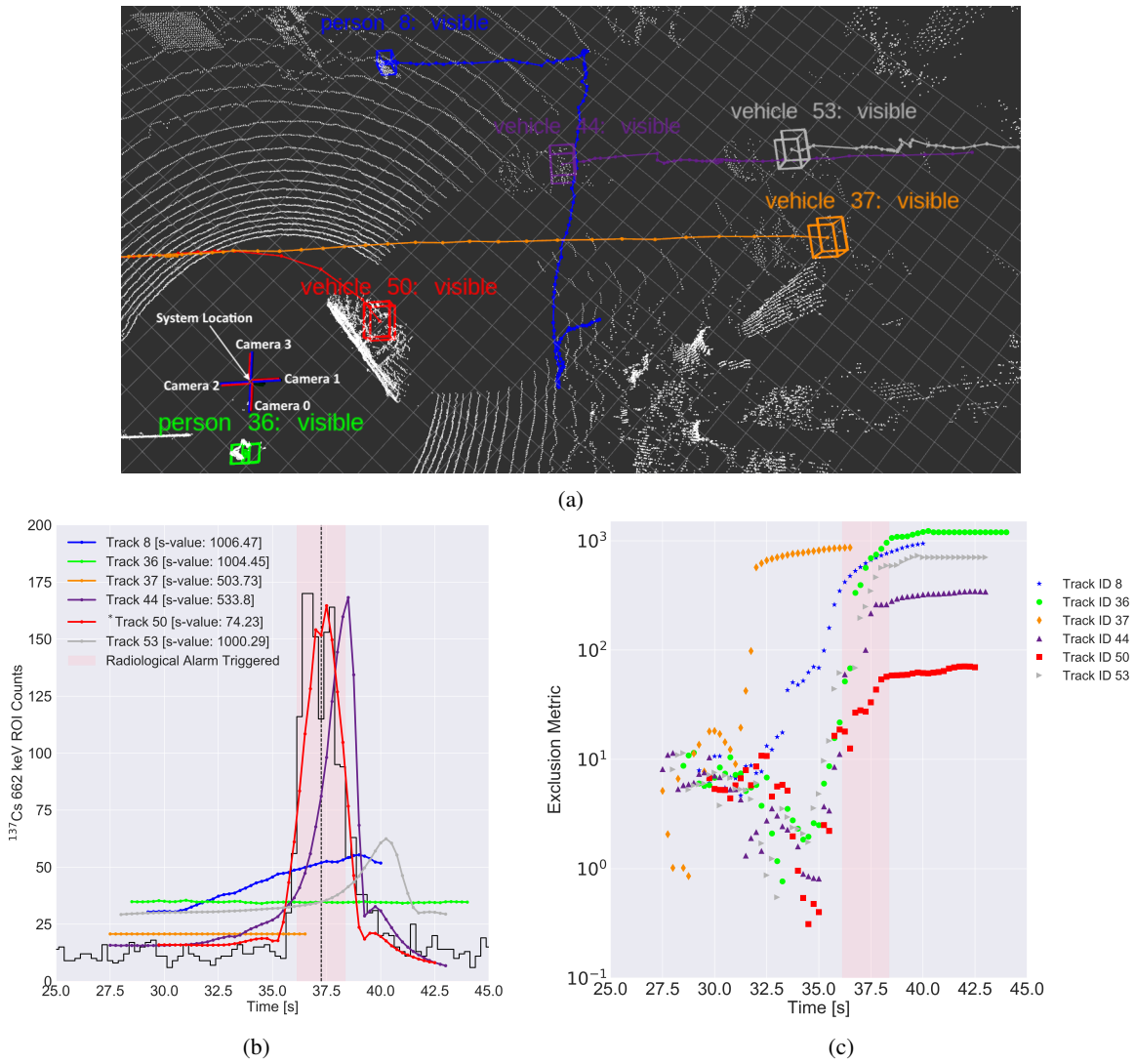


Fig. 5: Source-object attribution in a mock urban scene with both pedestrians and vehicles using LiDAR data. The same source-encounter from Fig. 4a, where a vehicle - Track 50 (white station wagon from Fig. 4a) - carrying a ^{137}Cs source drove past the system, is shown in (a), and the * in (a) indicates the source-carrier. (a) corresponds with camera image C in Fig. 4a. The bounding box colors (but not the labels) for each object are consistent with Fig. 4a. In (a), the trajectory of each object in the scene up to that point in time is indicated by the lines proceeding each respective bounding box. The white grid lines represent $1\,\mathrm{m}^2$, and the area in the middle without points is caused by the field of view of the LiDAR. The count-rate data from the $2 \times 4 \times 16$ in NaI(Tl) detector (black line) and the best-fit models to the ROI count-rate data is displayed in (b). The dashed line corresponds with the image from Fig. 5a, and (c) displays the evolution of exclusion metric as a function of time for each trajectory, and he pink bars indicate the time interval for the radiological alarm. The full video is available in the supplementary material.

and the color-coding for the tracked objects is consistent with the color-coding from each respective figure. In this encounter we find results consistent with the hypothesis that the optimized time-window for the source-carrying trajectory should yield a larger anomaly value. However, in both video and LiDAR, the source-carrying trajectory did not have the maximum anomaly value amongst the different integration times which was produced using a 4 second integration time. In this case, the optimal window does not produce the highest anomaly value due to anisotropic shielding from the vehicle, which is significant enough such that the assumptions of the analysis fail. We also observe an elevated anomaly value for another track which corresponds to the researcher who was operating the system, who was nearby and stationary during this particular alarm encounter.

The results from applying this analysis to the 26 different alarm encounters in a mock urban environment are shown in Fig. 8 for both video and LiDAR. The results show that using a tracking-informed integration window can improve the anomaly value, a proxy for detection sensitivity, compared to various fixed integration windows. Two alarm encounters (22** and 24**), in particular, produced large relative anomaly

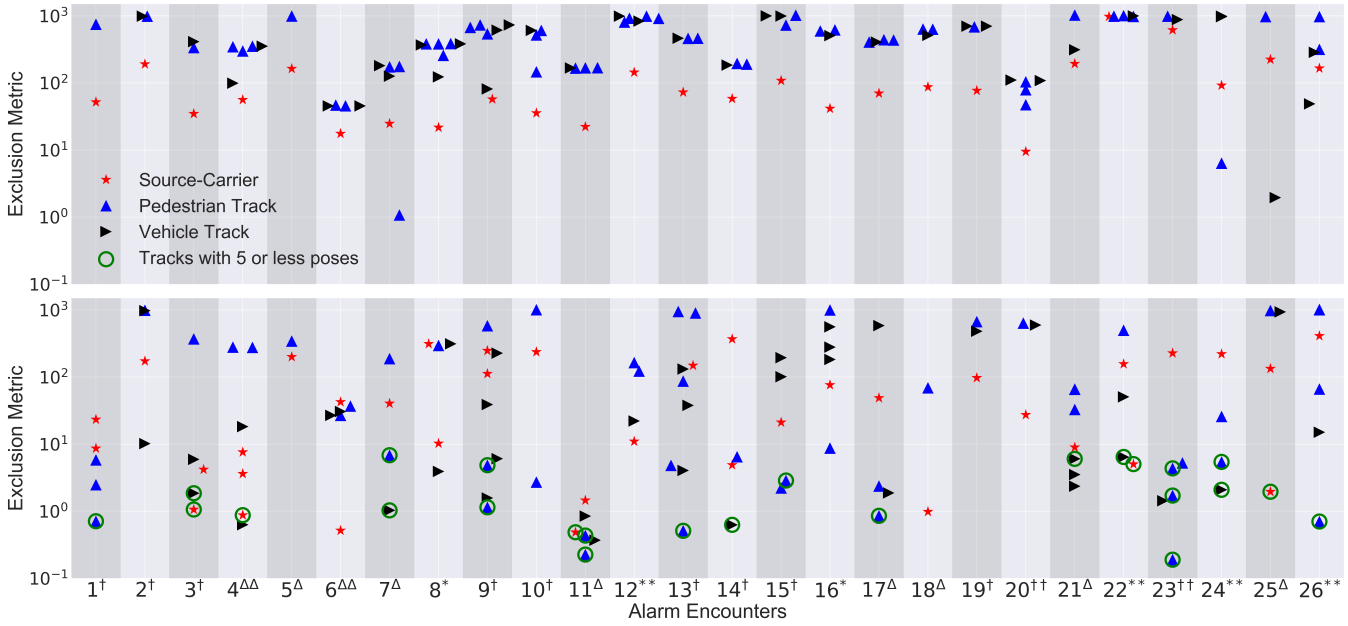


Fig. 6: Source-object attribution analysis results from 26 alarm encounters in a mock urban environment. The total number of tracks generated during each alarm encounter for pedestrians, vehicles, and the source-carrier are presented. The top (bottom) plot was performed with LiDAR (video) trajectories. † indicates that the source carrier went straight past the operator during the trial. ‡ indicates that the source carrier was first stopped at a red light then proceeded straight past the operator during the trial. $*$ indicates that the source carrier turned right during the trial, and $**$ indicates that the source carrier was stopped at a red light and then turned right. Lastly, Δ indicates a left turn by the source carrier during the trial, and $\Delta\Delta$ indicates the source carrier was stopped at a red light and then performed a left-hand turn.

values compared to fixed integration windows for both video and LiDAR. In both of these cases, the vehicle carrying the source was stopped at a red light next to the system for an extended period of time before the vehicle proceeded making a right-hand turn past the system.

From the results in Fig. 8, we find that the LiDAR-generated trajectories are more consistent than the video-based trajectories in yielding this enhanced sensitivity. This is consistent with the previous results showing that the LiDAR detection and tracking is more effective than video at consistently tracking objects through the scene. Even in the case of LiDAR tracking, there are alarm encounters where a fixed integration window of 1, 2, 3, 4, or 5 seconds yields larger anomaly values. These cases were often driven by the object not being detected or fully tracked across the system's entire FOV. In several cases, however, the object was fully tracked and again we hypothesize that the anisotropic shielding from the vehicle was sufficiently significant that the assumptions of the analysis fail. On the whole, this analysis suggests that contextual information can be used to improve the detection sensitivity of a static spectrometer.

IV. Discussion

The results of this study have shown that our source-object attribution analysis has the ability to improve both situational awareness and detection sensitivity. These results reveal that effectively leveraging video or LiDAR data, combined with edge computing, can enhance the localization and detection performance of a single radiation detector. However, the methods employed in this work have several limitations.

The attribution analysis does not account for uncertainties related to trajectories. Tracks are considered to be fixed and assumed to be the source of the radiation signal. The result of this assumption is that S-values are often quite large and don't reflect realistic probabilities, particularly for situations, where, when inspected by eye, a clear association between tracks and radiological data is obvious. Furthermore, large S-values are often caused by an inadequate physics model not taking into account anisotropic attenuation, as would be expected for sources inside of vehicles. While some uncertainty, for example those originating from tracking, could be incorporated into an improved attribution metric, many other uncertainties are of systematic nature and are not known. In these cases, the S-value, providing a statistical assessment of the situation, might not be the optimal metric and another description might provide an answer closer to the system operator's needs.

Considering attribution capabilities based on video, it can be noted that, first, it is limited by the unreliable distance estimate of objects in the scene. Track switching and disjoint tracks are a result of this uncertainty, particularly, for objects in close proximity to each other. There are depth extractor CNNs that produce depth maps for monocular cameras. They should provide better distance information but are currently too GPU intensive to run in parallel with object detection on a low-power system. Given the rate of improvement of these depth-estimating algorithms, we expect that a depth estimating CNN could be run along with an object detection, tracking, and

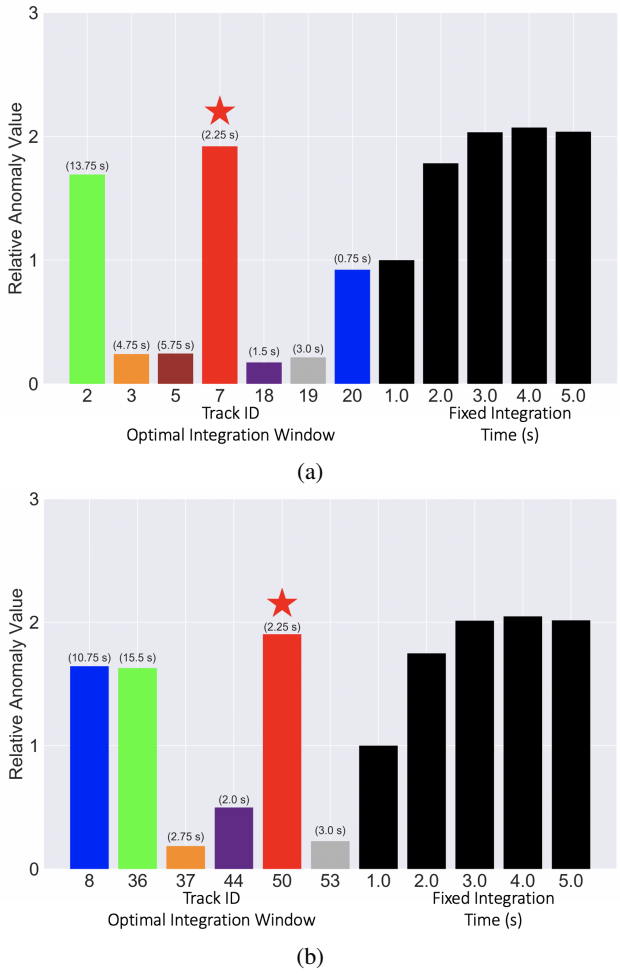


Fig. 7: Relative anomaly values from track-informed optimized integration windows and fixed integration window of 1, 2, 3, 4, and 5 seconds for both video (Fig. 7a) and LiDAR (Fig. 7b) data. The anomaly values are normalized to the 1.0 second fixed integration window. These results are from the alarm encounter shown in Fig. 4 and Fig. 5. The star indicates the track carrying the source in each scenario, and the time duration for each optimal window is provided in the parentheses.

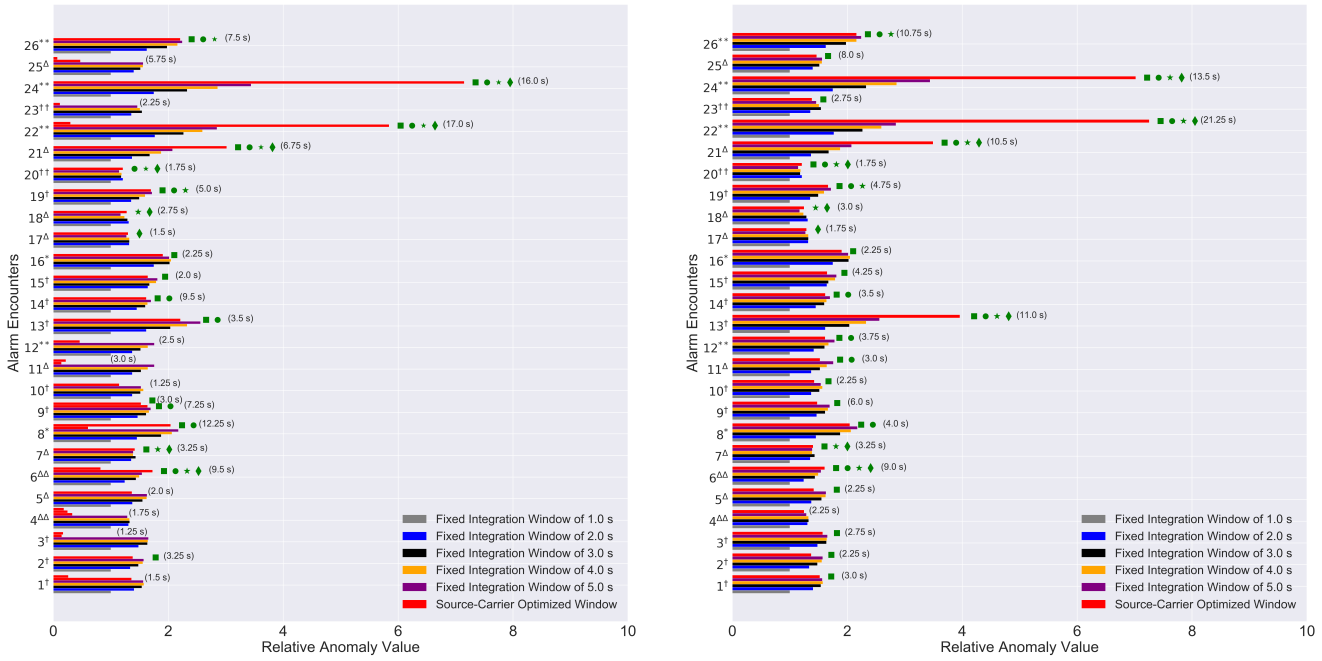
source-object attribution analysis in the near future. Another approach to improving distance estimation for video data is to use a stereoscopic camera, which provides a depth map. We currently have a setup that runs our attribution analysis with a single stereoscopic camera with reduced coverage compared to a 4 camera setup. Second, the current object detection network is not built to deal with small objects in images that are far away from the system. At distances larger than 10 m pedestrians cannot reliably be tracked anymore. This leads to short or split tracks that often score better on the attribution metric than the true source-carrier's trajectory. New developments, particularly feature pyramid networks [28], are trying to address these issues and might result in lightweight networks with better detection performance. Recent tests with yolov4 [29] show very promising results but it is currently too computationally demanding to be run on a low-power system. Finally, tracking of objects within 1 m to 2 m of each other is restricted by the distance uncertainties in the multi-variate normal representation, which can cause overlaps and result in wrong assignments. By using additional object appearance descriptions (features) derived from imagery, more reliable tracking could be achieved, and notably the issue of track switching could be resolved. An implementation of such a method that can already run on a low power system is described in [30], and will be explored in future studies.

The analyses presented in Section III suggest that object detection and tracking with LiDAR is currently more robust than a video-based approach. This is mostly driven by the direct availability of distance information. However, there is a benefit to running both video and LiDAR. SECOND/PointPillar's object detector doesn't work reliably for people standing near walls or large objects so the object may go undetected for LiDAR, and we believe this is because spurious points interfere with its anchoring algorithm. A training set including more data with pedestrians nearby objects could potentially increase the networks sensitivity, but is not readily available.

As discussed in Section III-A, the inclusion of attenuation in the attribution analysis could potentially help to break degeneracies that exist when objects approach the system in a similar fashion. In our current implementation, attenuation through an object in the scene is only calculated when an object is fully occluded. We neglect how the fraction of the solid angle covered by an occluding object between the detector and radiological source changes with time, which influences the expected counts in the detector. While this is a reasonable assumption for the occlusion that occurs between two pedestrians, for objects with larger extents, such as vehicles or trucks, this assumption is not valid. To better account for how the expected time-dependent count-rate in our physics model changes as two objects pass by each other, we need to determine how the fraction of the solid angle between the detector and radiological source changes as a function of time. This will require more complex experimental data and simulations to better understand both the attenuation effect of objects in the scene as well as the attenuation effect of the environment. In addition, to better account for attenuation in the physics model, we need to improve our tracking capabilities through occlusions to limit the number of track switches or disjoint tracks that currently occur. With better handling of attenuation and tracking capabilities through occlusions, the improvements yielded by including attenuation in the attribution analysis may improve compared to the attenuation results presented in this work.

V. Conclusion

We have developed and demonstrated a system that performs multi-class object detection, 3D tracking, and source-object attribution in real-time using either LiDAR or visual cameras. This system, and the methods it implements, can improve the localization of radiological/nuclear materials in urban environments by enhancing situational awareness and allowing non-directional detectors to provide localization information. Furthermore, in urban environments, this system



(a) (b)

Fig. 8: Maximum relative anomaly values from each alarm encounter using either track-optimized time-windows or fixed integration windows from 26 alarm encounters in a mock urban environment. The anomaly values are normalized to the 1.0 second fixed integration window. Results from the video and LiDAR are shown in sub-figure (a) and (b), respectively. The green squares, circles, stars, or diamonds indicate the encounters where the optimal integration window yielded a higher anomaly value than a 2, 3, 4, or 5 second integration window, respectively. The time duration for each optimal window is provided in the parentheses, and in the case for (a) the time duration with the longest window is provided in certain alarm encounters where the source-carrier had disjoint tracks. † indicates that the source carrier went straight past the operator during the trial. ‡ indicates that the source carrier was first stopped at a red light then proceeded straight past the operator during the trial. $*$ indicates that the source carrier turned right during the trial, and $**$ indicates that the source carrier was stopped at a red light and then turned right. Lastly, Δ indicates a left turn by the source carrier during the trial, and $\Delta\Delta$ indicates the source carrier was stopped at a red light and then performed a left-hand turn.

has the potential to improve detection of radiological/nuclear materials by increasing detection sensitivity.

The findings from this study demonstrate that radiological sources can be successfully attributed to objects derived from video and LiDAR in a mock urban environment.

The findings further show that LiDAR offers superior tracking performance as compared to video, and this improves the ability of the source-object attribution analysis to reject trajectories that are inconsistent with the data, while providing a more apparent correlation between the object responsible for the radiological alarm and the count-rate data. Additionally, it is shown that by using track-specific integration windows, object tracking can improve the anomaly value, a proxy for detection sensitivity, compared to fixed integration windows for both video and LiDAR. From these results, we conclude that using video and LiDAR to augment a static radiation detector does enhance situational awareness and strongly suggests that contextual information can be used to improve the detection sensitivity of a static radiation detector.

Future work will explore improving tracking capabilities for video to limit track switches and disjoint tracks by using object features derived from imagery. We will also explore different lightweight neural networks that offer better performing object detection algorithms for video as well as methods to improve the distance estimation of objects for video. Furthermore, we will explore other attribution metrics to identify the trajectories that are most (and least) likely to have been associated with a radiological alarm. Finally, the results presented were limited to a static system. Future work will investigate application of tracking and attribution performance for mobile systems where both sensors and objects are in motion.

REFERENCES

[1] R. C. Runkle, L. E. Smith, and A. J. Peurrung, "The photon haystack and emerging radiation detection technology," *Journal of Applied Physics*, vol. 106, no. 4, p. 7, 2009.

[2] R. C. Byrd, J. M. Moss, W. C. Priedhorsky, C. A. Pura, G. W. Richter, K. J. Saeger, W. R. Scarlett, S. C. Scott, and R. Wagner, "Nuclear detection to prevent or defeat clandestine nuclear attack," *IEEE Sensors Journal*, vol. 5, no. 4, pp. 593–609, 2005.

[3] T. P. Karnowski, M. F. Cunningham, J. S. Goddard, A. M. Cheriyadat, D. E. Hornback, L. Fabris, R. A. Kerekes, K.-P. Ziock, and T. F. Gee, "Motion estimation accuracy for visible-light/gamma-ray imaging fusion for portable portal monitoring," in *Image Processing: Machine Vision Applications III*, vol. 7538. International Society for Optics and Photonics, 2010, p. 75380F.

[4] K. P. Ziock, J. Collins, M. Cunningham, L. Fabris, T. Gee, J. Goddard, F. Habte, and T. Karnowski, "The use of gamma-ray imaging to improve portal monitor performance," *IEEE Transactions on Nuclear Science*, vol. 55, no. 6, pp. 3654–3664, 2008.
[5] K. Henderson, X. Liu, K. Stadnikia, A. Martin, A. Enqvist, and S. J. Koppal, "Proximity-based sensor fusion of depth cameras and isotropic rad-detectors," *IEEE Transactions on Nuclear Science*, 2020.
[6] K. Bilton, T. Joshi, M. Bandstra, J. Curtis, B. Quiter, R. Cooper, and K. Vetter, "Non-negative matrix factorization of gamma-ray spectra for background modeling, detection, and source identification," *IEEE Transactions on Nuclear Science*, 2019.
[7] "High-resolution os1 lidar sensor: robotics, trucking, mapping — ouster," https://ouster.com/products/os1-lidar-sensor/, (Accessed on 06/17/2020).
[8] e-con Systems, https://www.e-consystems.com/nvidia-cameras/jetson-agx-xavier-cameras/four-synchronized-4k-cameras.asp, (Accessed on 02/04/2020).
[9] "Jetson agx xavier developer kit — nvidia developer," https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit, (Accessed on 06/17/2020).
[10] "digibase 14-pin pmt base with integrated bias supply, preamplifier, and mca with digital signal processing," https://www.ortec-online.com/products/electronics/photomultiplier-tube-bases/digibase, (Accessed on 08/12/2020).
[11] Stanford Artificial Intelligence Laboratory et al., "Robotic operating system." [Online]. Available: https://www.ros.org
[12] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv*, 2018.
[13] M. Bjelonic, "YOLO ROS: Real-time object detection for ROS," https://github.com/leggedrobotics/darknet_ros, 2016–2018.
[14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.
[15] C. D. Fryar, Q. Gu, C. L. Ogden, and K. M. Flegal, "Anthropometric reference data for children and adults; united states, 2011-2014," 2016.
[16] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, 2018.
[17] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," *CVPR*, 2019.
[18] H. Caesar *et al.*, "nuscenes: A multimodal dataset for autonomous driving," *arXiv:1903.11027v4*, 2019.
[19] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 3464–3468.
[20] L. Pardo, *Statistical inference based on divergence measures*. Chapman and Hall/CRC, 2018.
[21] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.
[22] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. . Barrand *et al.*, "Geant4—a simulation toolkit," *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003.
[23] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, "Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere," *The Astrophysical Journal*, vol. 622, no. 2, p. 759, 2005.
[24] J. C. Curtis, R. J. Cooper, T. H. Joshi, B. Cosofret, T. Schmit, J. Wright, J. Rameau, D. Konno, D. Brown, F. Otsuka *et al.*, "Simulation and validation of the mobile urban radiation search (murs) gamma-ray detector response," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2018.
[25] L. D. Meyers, J. P. Hellwig, J. J. Otten *et al.*, *Dietary reference intakes: the essential guide to nutrient requirements*. National Academies Press, 2006.
[26] S. Baker and R. D. Cousins, "Clarification of the use of chi-square and likelihood functions in fits to histograms," *Nuclear Instruments and Methods in Physics Research*, vol. 221, no. 2, pp. 437–442, Apr. 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167508784900164
[27] S. Greenland, "Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with s-values," *The American Statistician*, vol. 73, no. sup1, pp. 106–114, 2019.
[28] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.
[29] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv e-prints*, p. arXiv:2004.10934, Apr. 2020.
[30] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649.