Performance Measurements for Evaluating Static and Dynamic Multiple Human Detection and Tracking Systems in Unstructured Environments

Barry Bodt,
Richard Camden
Army Research Laboratory
{babodt,rcamden}@arl.army.mil

Harry Scott, Adam Jacoff, Tsai Hong, Tommy Chang, Rick Norcross, Tony Downs, and Ann Virts
National Institute of Standards and Technology
{harry.scott, adam.jacoff, tsai.hong, tommy.chang, rick.norcross, tony.downs, ann.virts}@nist.gov

ABSTRACT
The Army Research Laboratory (ARL) Robotics Collaborative Technology Alliance (CTA) conducted an assessment and evaluation of multiple algorithms for real-time detection of pedestrians in Laser Detection and Ranging (LADAR) and video sensor data taken from a moving platform. The algorithms were developed by Robotics CTA members and then assessed in field experiments jointly conducted by the National Institute of Standards and Technology (NIST) and ARL. A robust, accurate and independent pedestrian tracking system was developed to provide ground truth. The ground truth was used to evaluate the CTA member algorithms for uncertainty and error in their results. A real-time display system was used to provide early detection of errors in data collection.

Categories and Subject Descriptors
B8.2 [**Performance and Reliability**]: Performance Analysis and Design Aids; C.4 [**Performance of Systems**]: Performance attributes.

General Terms
Tracking, Algorithms, Performance, Measurement, Experimentation

PerMIS'09, September 21-23, 2009, Gaithersburg, MD, USA. ACM 978-1-60558-747-9/09/09

Keywords
Unmanned ground vehicle, experimental design, ground truth, pedestrian tracking, metrics, perception, performance evaluation

1. INTRODUCTION
The ARL Robotics Collaborative Technology Alliance (CTA) conducted an assessment and evaluation of multiple algorithms for real-time detection of pedestrians in Laser Detection and Ranging (LADAR) and video sensor data taken from a moving platform in January 2009. In the assessment, the robot vehicle equipped with two pairs of stereo cameras, two sets of General Dynamics Robotic Systems (GDRS) LADAR and two sets of SICK[1] lasers was driven by an operator through a straight route of approximately 240 m containing various configurations of eight moving pedestrians, four mannequins, four barrels, four cones, two trucks, two crates, seven tripods and trees. In addition to the complexity of the environments, the variables included multiple robot vehicle speeds (30 km/h or 15 km/h) and pedestrian speeds (1.5 m/s or 3.0 m/s). The environment was intended to provide some Military Operations in Urban Terrain, or MOUT, characteristics.

The objective of the experiment was to capture the data necessary to evaluate the performance of each CTA team's algorithm, to provide data to support further development of algorithms, and to produce performance analyses based on the captured data to support obstacle avoidance planning. An Ultra

[1] Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST nor does it imply that the materials or equipment identified are necessarily the best for the purpose.

WideBand (UWB) system [1] employed by the National Institute of Standards and Technology (NIST) provided position tracking (≈20 cm uncertainty) of the moving and stationary humans, the robot vehicle, and other objects. Improved performance of the CTA tracking and recognition algorithms has called for improvements in the ground truth solution. Processing techniques were developed and implemented to produce higher quality tracking solutions than those provided by the raw data captured by the ultra wideband system. To address this, we developed a robust filter algorithm. To improve analysis of the performance of the CTA tracking systems, we also developed a temporally consistent algorithm for finding the correspondence between the ground truth data and the CTA tracking data. In addition, a display system was implemented to provide early detection of errors in data collection and to assist in data analysis.

The paper is organized as follows: Section 2 presents a detailed description of the solution based on the UWB system for capturing the ground truth data. Section 3 introduces the filter and interpolation algorithms for improving the quality of the UWB data. Section 4 describes the visualization system for providing early detection of errors. Section 5 presents the correspondence algorithm for data analysis. Sections 6 and 7 present the performance metrics and analysis for evaluating the CTA algorithms. Finally, Section 8 provides a summary and conclusion.

2. GROUND-TRUTH REFERENCE SYSTEM

2.1 Ground Truth Setup

NIST researchers have been working with an asset tracking system employing ultra wideband technology to capture 2-D location and path data for robots, vehicles, and personnel operating within scenarios set up to evaluate robotic perception systems. The goal is to capture quantitative performance data referenced to ground truth positions and time to help compare and improve sensors and algorithms in both indoor and outdoor scenarios.

The tracking system uses state of the art ultra wideband radio receivers posted around the perimeter of the scenario to track multiple static and dynamic targets with badge-size transmitters (see Figure 1). It works in open outdoor areas and indoor areas through certain types of walls, though overall accuracy can vary. NIST has performed system characterization tests in ideal conditions to determine the best possible 2D accuracy of the system, which is approximately 15 cm. We have used it to track vehicles and personnel throughout an area over 80 000 square meters with an average accuracy of approximately 20 cm and an update rate of approximately 50 Hz, which is sufficient for tracking vehicles at highway speeds. We have also used it to track robots through random mazes with plywood walls (non-line-of-sight) achieving similar accuracies. We have not been successful tracking through concrete walls, but have used additional receivers in hallways to compensate during indoor building deployments. The total number of dynamic and static transmitter tags used simultaneously thus far has been approximately 15 and 30 respectively for marking obstacles and known fiducial points to check accuracy. Setup time for a new site takes about 5 days. Returning to a previously setup site takes approximately two days for calibration prior to testing.

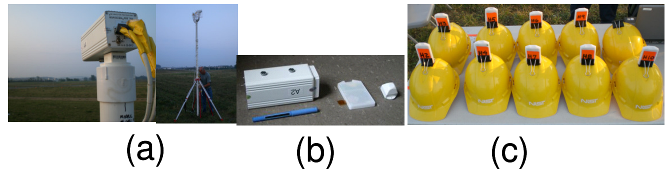

Figure 1. (a) shows a receiver deployed in the field atop a mast centered over a known fiducial marker. (b) shows the asset tracking system components, ultra-wideband radio frequency receiver (shown with integrated high-gain antenna), 1 W transmitter tag, and 30 mW transmitter tag. c) shows Several badge tags attached to helmets to track personnel in the scenario. Typically two tags are placed on moving vehicles to identify orientation.

Figure 2 shows a plot of the tracking results for a ground truth system coverage and accuracy test on the January course configured at NIST. Green and orange plots show the vehicle path and the other plots show pedestrian tracks.

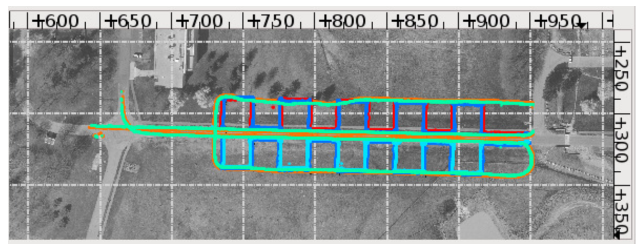

Figure 2. A calibration run with two transmitter tags mounted on a vehicle and two tags on each of two pedestrians to check coverage.

2.2 Filter and Interpolation Algorithms for the Ground-Truth Data.

The goal of the filter process is to remove outlier and error measurements from the ground-truth data. We identify outliers based on the maximum plausible speed of the tag. A polynomial least-squares algorithm filters the remaining data points. We then fit a spline through the filtered points to identify the tag's position as a function of time. We interpolate the trimmed, filtered, and splined data at timestamps obtained from the CTA performers' data. This interpolated ground-truth is later used to establish temporal constraints for correspondence.

The UWB position data contain anomalies that, while generally minor, diminish the usefulness of the data in subsequent evaluations and displays. The filter combines previous and subsequent readings to remove anomalies and identify a more accurate and timely position. For example, Figure 3 below is the UWB position data for tag A01D of Run3. The data show two significant anomalies: a gap in the center area and outlier points away from and along the track.

The filter has three components: trim, filter, and spline. The green dots represent the raw data that were trimmed as outliers. The red dots are the remaining raw data points. The white line is the result of the filter. And the blue points are the positions at the CTA timestamps based on a spline fit of the filtered points.

The filter's trim component removes outlier data points. Physical constraints limit the distance that an UWB tag can move between readings. The trim component computes the velocity between the current point and the last good point. The filter trims the current point when the velocity between these points is excessive. The filter checks subsequent points for a point that represents a reasonable velocity. The filter then uses that point as the last good point for subsequent evaluations. The filter passes the trimmed position list to the window component (see red in Figure 3).

The filter component is based on the Savitsky-Golay algorithm [2]. Savitsky applies a polynomial least-squares fit to a set of points before and after the current point. The filtered value is the sum of the products of the points with an array of coefficients determined by the order of the algorithm (generally 3) and the size of the point set. The Savitsky algorithm relies on evenly spaced data. The gaps in the trimmed data would cause the Savitsky algorithm to inappropriately shift the data points. To compensate, our algorithm fills gaps in the trimmed data with linearly interpolated data points before applying the Savitsky-Golay algorithm. Our algorithm discards the fill points prior to passing the data onto the spline component (see the white in Figure 3).

The filter's spline component is based on a cubic Hermite spline. The spline component identifies positions at a time of interest rather than at the time of data collection and allows researchers to determine the position of the UWB tag at times provided by the CTA systems (see the blue in Figure 3).

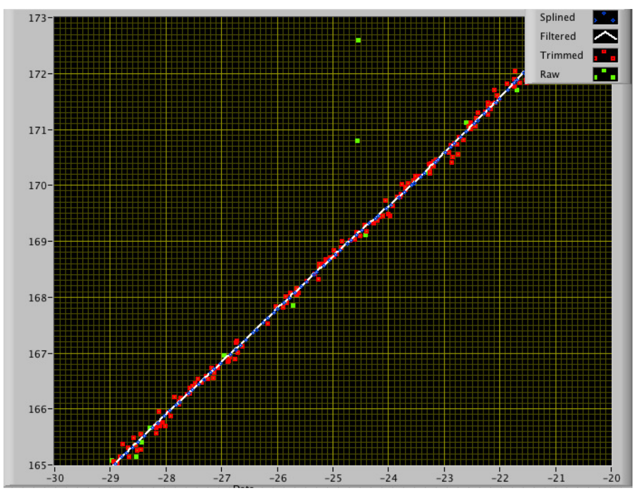

Figure 3. Green is the raw data. Red is the trimmed data. White is filtered data. Blue is the interpolation data

3. CTA REAL-TIME HUMAN DETECTION AND TRACKING ALGORITHMS

In this experiment, six algorithms were included from the CTA. Five use LADAR sensing and one uses a vision system to provide data for the algorithms. During each algorithm cycle, the algorithm reports information about the detected humans. The report includes the number of detections, their locations, strength of detections, the time of detection, as well as vehicle status such as its location, speed, orientation, etc. All detections from the same algorithm cycle have the same detection time. The reports are collected and saved into files, one per algorithm.

Reporting rate varies from algorithm to algorithm and may not be fixed within the algorithm itself. For example, an algorithm's cycle time may increase when the number of detections increases.

In general, there are two data sets: ground-truth data and detection data. Detection data are the locations and detection times of all humans reported by a CTA algorithm, whereas ground-truth data are the corresponding UWB locations for the same time.

Both the detection data and the ground-truth data are independently grouped with unique identifications (ID). For the ground-truth data, the group IDs are also referred to as the "tag ID". Different tag IDs always refer to different humans or physical objects.

For the detection data, the group IDs are referred to as "tracking ID". With perfect CTA system detection and tracking performance, the number of tracking IDs would be the same as the number of tag IDs. In reality, the detection data can have more than one tracking ID for the same human due to occlusion or imperfect tracking capability of the algorithms.

4. DATA VISUALIZATION FOR EARLY DETECTION OF ERRORS IN DATA COLLECTION.

Data visualization is important for verifying the integrity of both the ground-truth data and the outputs of the CTA algorithms prior to, and during, the data collection. Bad data could arise due to sensor malfunction or unforeseen circumstances prior to or during the data collection. Since data collection is expensive, time consuming, and labor intensive, it is advantageous to detect bad data as soon as possible and prevent waste of resources.

A software-based interactive viewer was developed for this purpose. The viewer uses various open source libraries and runs natively on Linux, Windows and Mac OS X. Figure 4 shows a typical screen-shot of the viewer displaying both the detection data from a CTA algorithm and the corresponding ground-truth data. An individual entity can be toggled on or off by clicking on its tag ID or tracking ID.

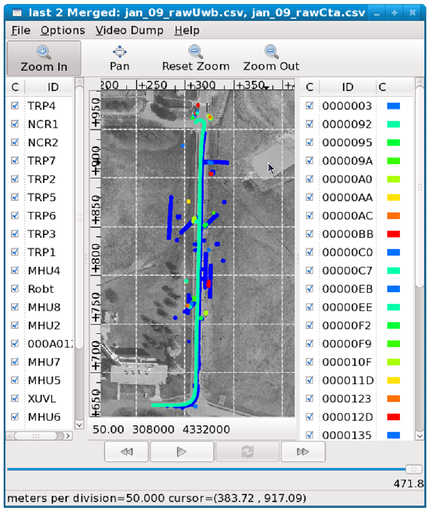

Figure 4. CTAviewer screenshot showing both detection data and ground-truth. The left panel lists the tag ID and the right panel lists the tracking ID. An individual entity can be toggled on/off by its ID.

Figure 5 and Figure 6 are examples of bad data. The plots show the locations of two ground-truth tags mounted on a moving vehicle. The two tags were separated by about 1 meter and the vehicle drove at a constant speed.

The viewer software allows us to quickly view and evaluate the data immediately after the each run and investigate the cause of any anomalies. Problems in ground-truth can often be eliminated by placing more UWB receivers in the problematic areas.

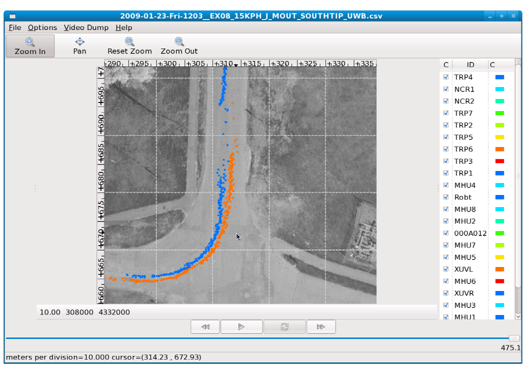

Figure 5. Correctable gaps in the ground-truth data.

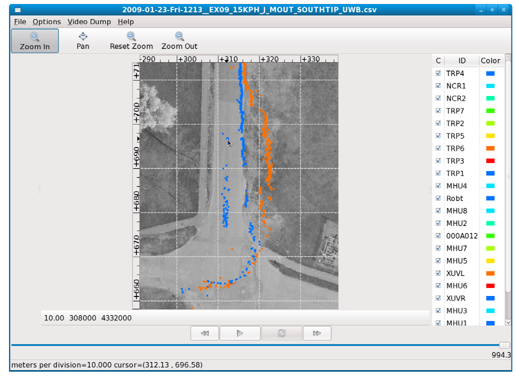

Figure 6. Uncorrectable severe distortion and gaps in the ground-truth data.

5. MAP BETWEEN GROUND-TRUTH AND DETECTION

All CTA teams output their results with time stamps which are used to synchronize to the UWB ground-truth data. Since each team has different output rates, the linear interpolation described in Section 2.2 is used to handle the different rates. All timestamps in the data collection systems come indirectly from a common clock source via a Network Time Protocol (NTP) server. All systems synchronize to the NTP server at the beginning of each run. A run lasts about 1 minute. This allows our data collection computers to stay synchronized to each other within 20 milliseconds.

All CTA algorithms output their results in a standard format, and are stored in Comma Separated Values (CSV) for viewing using the viewer described in section 3.

The requirements for the performance evaluation of the CTA systems include:

1. Timestamp correspondence between ground-truth and detection.

2. Object/human correspondences between ground-truth and detection.

3. Definition and computation of metrics and measurements for performance evaluation.

Establishing time correspondence between the ground truth and the detection system is important for resolving ambiguity that involves time. Finding a mapping between the ground truth objects and the objects detected by the CTA algorithms is crucial for evaluating the algorithms' performance. In Figure 7 using a nearest neighbor criterion, inside the red circle, the blue star T_3 will correspond to the yellow T_2 circle since the distance is less than the distance to the blue T_3 circle. When time correspondence is established, the star T_3 ground-truth will correspond to the ground truth represented by the blue T_3 circle. The time correspondence algorithm will be described in section 5.1. Before defining the metrics, it is necessary to have a good way of assigning detected objects to ground truth. The detail of the correspondence algorithm will be described in Section 5.2. The assessment of the performance of the CTA tracking systems required several measurements. These measurements will be described in Section 5.3.

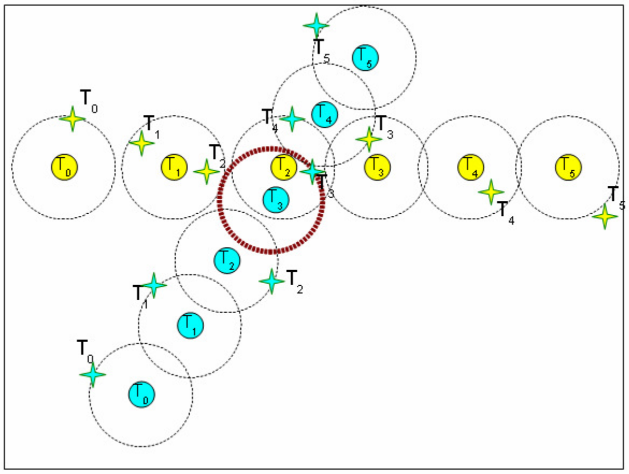

Figure 7. The CTA data are represented by star shapes and the ground-truth data are represented by shaded circles. T_n represents the time n. It is sufficient to use time to match the closest detection ground-truth pair. The outer circle around the ground-truth data indicates the threshold radius used for establishing a spatial constraint.

5.1 Establish Time Correspondences between Ground-Truth and Detection Tracking.

After the filtering and interpolation process, all ground-truth data are interpolated at timestamps compiled from all the detection data. Two matching files are generated for each team - one containing the ground truth and the other containing the detection data. These two matching files have the same number of entries. Each entry contains a timestamp and information about object locations detected at, or interpolated to, that timestamp. Since the timestamps are matched, there is a one-to-one correspondence among the timestamps in the entries between the two files.

5.2 Establish Object/Human Correspondences between Ground Truth and Detection Tracking

Previously [3], correspondence was determined solely based on spatial constraints. One such constraint is the distance between the ground-truth and the detection data. Although spatial constraints are essential, they alone can not resolve ambiguities that arise when close data are taken at different times. Such ambiguity and an unnecessary spatial search can be avoided when we take the temporal consistency into account

Several map correspondence algorithms [4][5][6][7] were investigated. The correspondence algorithm, adopted by Classifications of Events, Activities and Relationships (CLEAR)[8] evaluation workshop group, was implemented. Using matched-pair data from Section 5.1, the algorithm computed the averaged error distances over the cluster with respect to each object. Clusters were associated with ground-truth objects based on a minimum average distance subject to meeting a 3 m proximity threshold based on our experiments. In addition, we used velocity to differentiate stationary from moving objects. Clusters were then labeled as a human (moving), mannequin (stationary human), misclassification objects (moving or stationary), or false positive. The results of this correspondence procedure will be used for computing several measurements for the analysis of the human detection in the following section.

5.3 Post Process the Data Acquisition files for Human Detection Analysis

In order to analyze the CTA tracking algorithms correctly and accurately, post processing of the data is necessary. CTA algorithms differed in cycle time ranging from 7 Hz to 20 Hz. At the end of each algorithm cycle, each algorithm reported detection information such as positions and velocities of the humans. The underlying assumptions for the outputs of the algorithms included the following:

- Only obstacles seen and classified as human were reported.
- Unique identification numbers were assigned to individual algorithm detections within a run.
- Algorithms demonstrated tracking of an individual by maintaining the same ID in successive frames.
- Algorithms also reported velocity of the detected humans.

Since we only instrumented the ground-truth data in the 300 m x 150 m test area, all detections are excluded if they occurred outside the test area. The correspondence algorithm described in Section 5.2 found the correspondence between the detections and the ground-truth based on location and time stamp. Detections were compared with all the ground-truth objects on the course. Absolute error distances were computed, summed, and averaged over the cluster with respect to each course object. The absolute velocity error between detection and ground truth objects was also computed and averaged over the cluster. Clusters were associated with a ground-truth object based on minimum average distance subject to meeting a 3 m proximity threshold. Using velocity to establish stationary/moving objects, clusters were then classified as a human, mannequin, misclassification, or false positive. Other values were reported in the post processing. For example, reports included the distance from the moving vehicle at the time of first detection for individual detections within the common ID cluster, the shortest distances and velocities, and dispersion measures for distance and velocity.

6. PERFORMANCE METRICS

Post processing of the data above results in a spreadsheet for each algorithm with metrics for analysis. A record is formed for each algorithm-reported human. Each algorithm assigns an identifier to an entity on the course classified by the algorithm to be a human. All information related to that algorithm identification is condensed to a single record. This record may hold information from many cycles of the algorithm. Post processing determines whether that entity is, in truth, a human or mannequin (true positive), another known course entity not human or mannequin (misclassification), or an unknown course feature with no associated ground-truth (false positive). Distinctions are also made between moving and stationary entities and various classes of nonhuman entities (e.g., barrels, cones, crates). Field notes describe test conditions under which the data were collected, absolute and relative positioning of the robot platform and detected entities recorded at the time detections first occurred for an identification, time and cycle number indicators of the persistence of detection, or the accuracy of the algorithm classification decision.

7. EXPERIMENTS AND PERFORMANCE EVALUATION

The purpose of this section is to outline the experiment and to illustrate the importance of the ground truth system in assessment of the algorithms. A complete analysis is not given. The principal experiment consisted of thirty-two runs conducted at the south end of Center Drive on the NIST campus (Figure 9). An autonomous vehicle platform, with sensors and algorithms on board as discussed above, was driven south to north over an approximately 240 m run. Scripted scenes with human motion, mannequins, and course clutter were sensed and interpreted and reported by the algorithms in real time. Eight humans were present in each run, four to either side of the road. Four moved in a manner parallel to the road, three at 45 degrees toward the road, and one perpendicular toward the road. Three parallel runs were receding from the platform and one was approaching. Among the 45 degree runs, all were approaching the road, but only the run to the right was approaching the vehicle.

Movements of the humans were choreographed and timed to ensure that regardless of test conditions the scene sensed remained consistent across runs. See Figure 8. All humans were upright in the principal experiment. Excursion runs not reported here explored other postures and group movement patterns.

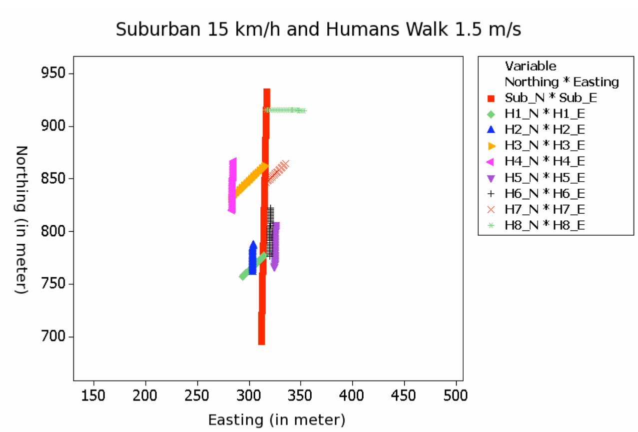

Figure 8. Human paths relative to platform route. The units are in meters.

Test conditions were formed based on three factors: platform speed, human speed and course clutter. The platform was driven at either 15 km/h or 30 km/h, humans moved at 1.5 m/s or 3.0 m/s, and the course was cluttered to approximate MOUT complexity or was open except for the human movers. The test conditions were allotted equally in accordance with a 2^3 factorial design with 4 replications per condition. Under the MOUT conditions only, 8 mannequins, 4 barrels, 4 cones, 2 crates, and 2 trucks were included on the course. Seven NIST tripods were on the course for all 32 runs. Figure 9 shows a view of the course during a MOUT run.

The assessment of algorithms focuses on the questions of what an algorithm saw, when it saw it, and how long the sighting persisted. These questions will be pursued for each algorithm and in the context of the experimental conditions under which the data were collected. The ground-truth system allows definitive answers to these questions. We share some preliminary high-level results to illustrate performance.

Figure 9. Right side of course during a MOUT run.

Table 1 summarizes the performance of the six algorithms in terms of detections, misclassifications, and false positives over the complete set of 32 runs. Entries are percentages except for the false positive entries, which report the number per run. Note that true positives are in bold, and false positives are in italics. All other entries show the algorithm misclassification of other course entities as human. At a high level, this table addresses what was seen.

Table 1. Algorithm performance expressed in terms of the percentage of course entities detected and the number of false positives per run.

	CTA Algorithm					
Object Type	Alg 1	Alg 2	Alg 3	Alg 4	Alg 5	Alg 6
Human(%)	**97.3**	**90.8**	**98.4**	**98.0**	**89.5**	**85.7**
Mann.(%)	**10.2**	**-**	**97.7**	**98.4**	**91.4**	**62.5**
Cones(%)	0.0	-	4.7	0.0	65.6	0.0
Barrels(%)	14.1	-	54.7	70.3	89.1	0.0
Crates(%)	46.9	-	100.0	90.6	100.0	50.0
Trucks(%)	25.0	-	100.0	25.0	100.0	75.0
Tripods(%)	1.3	46.7	53.6	60.7	58.9	29.8
False Positives	*29.8*	*77.9*	*155*	*37.3*	*29.8*	*1.3*

Performance varies widely across algorithms. While some demonstrate a high probability of detection, misclassification of other course entities is clearly a problem. Moreover, the number of false positives recorded, if not addressed, ultimately would provide a greater challenge for dynamic planning in an autonomous mode.

In Figure 10, a boxplot (with mean) is shown for the distance between the platform and the target entity at the time of first detection. This distance is as perceived by the algorithm, but generally this does not vary greatly with the actual ground truth. Detections (green), misclassifications (yellow), and false positives (red) are shown. There are differences according to the type of obstacle. Tripods and trucks (Trks), for example, tend to be recognized in the neighborhood of 30 m away; whereas humans are detected on average at more than 50 m away. Confidence in this graph is based on the ground-truth system. A similar graph exists with actual distances based on the ground-truth, but a distance for false positives requires the algorithm-produced values. In this fashion, the question of when entities were seen is addressed.

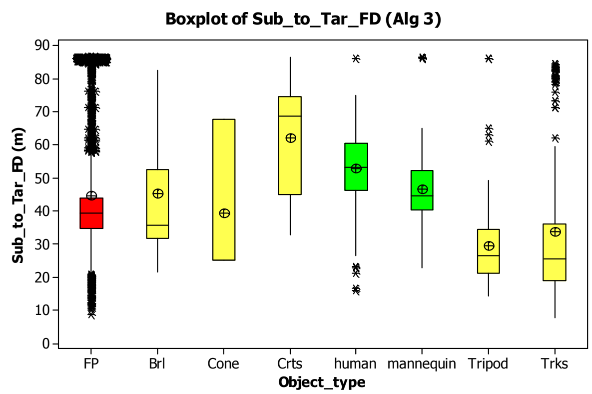

Figure 10. Boxplots of distance from platform to targets detected by the algorithms for different object types.

Figure 11 shows a boxplot for the duration of time all entities of a certain type were tracked. Ideally, humans and mannequins would be tracked persistently; whereas, other entities would not.

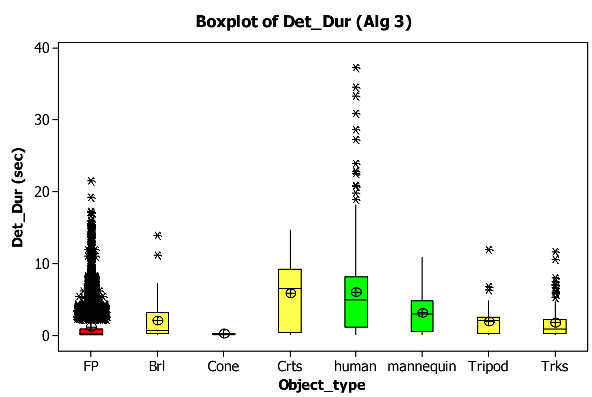

Figure 11. Boxplots of duration of tracking for different object types.

From Table 1 we learn that high detection rates are accompanied by higher than desired misclassification rates and numbers of false positives. The intent of examining data in Figure 11 is to determine if persistent tracking requirements might greatly reduce false alarms and misclassifications while retaining a high level of detection. In this case, at least 75 % of the false positives fall below the 25th percentile for humans and mannequins detected, suggesting persistent tracking may reduce false positives. However, tracking of misclassified entities would not be greatly influenced.

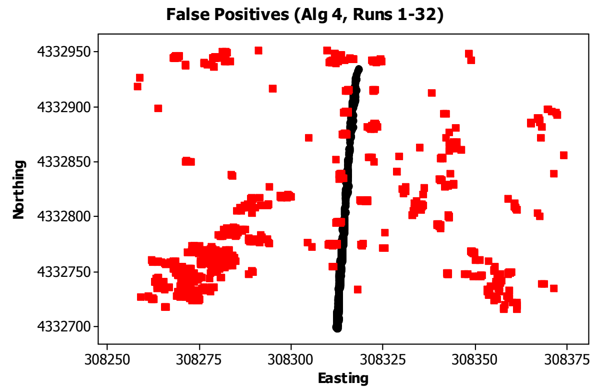

Figure 12. Scatterplot of false positive locations for Alg4.

Ground-truth allows a definitive decision on false positives. Figure 12 shows the false positives recorded for ALG4 over the 32 runs of the experiment. The path of the vehicle (black) and the false positive locations (red) are shown. As part of the analysis, the cause of false positives (e.g., bushes, trees, high grass) will be pursued.

One of the most advantageous features in the measurement technology employed here is the time sequenced display of detection. For individual runs, it is necessary to drill down into the data to investigate anomalies more carefully. A static display in Figure 13 shows a run for ALG4. Labels over the points for the path of the vehicle indicate when a specific moving human was detected during the run. The moving human is also plotted. In this instance, there are replicates for some of the moving humans (e.g., MUH4). This tells us that multiple unique identifications were assigned to this human. For some reason, the algorithm judged the human to be a different entity at different time. One possibility is occlusion, as might be the case here when MUH4 was obscured from view by crates during a MOUT run. Although informative, the static display does not reveal the same detail as movies of the run as it unfolds. The CTA Viewer illustrated in Figure 4 is critical to detailed analysis.

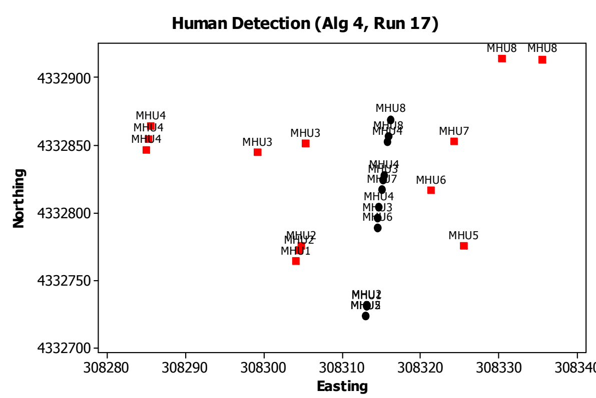

Figure 13. Human detection for run 17 using ALG4.

8. CONCLUSION

We presented details of several components of a system for determining performance of sensors and perception algorithms tasked with detection, tracking, and classification of moving and fixed objects, including pedestrians, around a moving robot vehicle. We presented a filter algorithm developed to improve ground-truth data for analysis. In addition, we developed a group-based correspondence between ground-truth and detection for data analysis and performance evaluation.

From an analysis perspective, the advances in measurement technology of good ground truth data improve the assessment process markedly. The ground truth precision provides an objective evaluation of the results reported by the algorithms. It makes possible the exact tracking of moving entities on the course, essential given the planned assessment of the "detection and tracking" purposes of the algorithms. This was previously not possible. The CTA viewer has proven to not only be a useful tool in visual analytics, but has also provided an instant check during the conduct of the experiment as to whether or not data are being collected and whether systems are in good calibration.

9. ACKNOWLEDGMENTS

The authors would like to thank the CTA partners for their research and willingness to conform to the reporting format, and to thank Jeff Jones and his team at General Dynamic Robotics Systems (GDRS) in Westminster for their help with data collection.

10. REFERENCES

[1] Robert Fontana, Recent System Applications of Short-Pulse Ultra-Wideband (UWB) Technology (Invited Paper), IEEE Microwave Theory & Tech., Vol. 52, No. 9, Sept. 2004.

[2] Abraham Savitzky and Marcel Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures". Analytical Chemistry 36 (8): 1627–1639, 1964.

[3] Barry Bodt and Rickard Camden, "Detecting and Tracking Moving Humans form a Moving Vehicle,", SPIE, Orlando, Florida, 2008.

[4] Keni Bernardin, A. Elbs, Rainer Stiefelhagen, "Multiple Object Tracking Performance Metrics and Evaluation in a Smart Room Environment". The Sixth IEEE International Workshop on Visual Surveillance, VS 2006, Graz, Austria, 2006.

[5] Kevin Smith, Sileye Ba, Jean-Marc Odobez and Daniel Gatica-Perez, "Evaluating Multi-Object Tracking", Workshop on Empirical Evaluation Methods in Computer Vision (EEMCVF) San Diego, CA, 2005.

[6] Sarma Pingali and Jakub Segen, "Performance Evaluation of People Tracking Systems," wacv, pp.33, Third IEEE Workshop on Applications of Computer Vision (WACV '96), 1996

[7] Faisal Bashir and Fatih Porikli., "Performance Evaluation of Object Detection and Tracking Systems", IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), June 2006.

[8] Keni Bernardin, " CLEAR 2007 Evaluation Plan 3D Person Tracking Task,", Classification of Event, Activities and Relationships Evaluation and Workshop (CLEAR) 2007.

[9] Richard Camden and Barry Bodt, "Safe Operations Experiment Report," ARL-TR-3773, U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, April 2006.

[10] Elias Rigas, Barry Bodt and Richard Camden, "Detection, tracking, and avoidance of moving objects from a moving autonomous vehicle," Proc. SPIE 6561 (2007).