Standoff Technology Integration and Demonstration Program

STIDP-T1-012, Rev. 0
PNNL-19947

Nano-Corpus Test Report

RL Erikson

March 2010

Homeland Security

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/ordering.htm

STIDP-T1-012, Rev. 0
PNNL-19947

Nano-Corpus Test Report

RL Erikson

March 2010

Prepared for the U.S. Department of Homeland Security,
Science and Technology Directorate/Explosives Division,
and the U.S. Department of Defense
under U.S. Department of Energy Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Revision Log and Approvals

Document Title

Rev. No.	Date	Describe Changes	Pages Changed
0			

Name and Title	Approvals	Date
Todd A. Jokerst, Object Tracking Subtask Lead		
Paul A. Scott Technology Task Lead		
J. Mark Henderson, Deputy Program Manager		
Nick Lombardo, Program Manager		

Acronyms and Abbreviations

GTF	Ground Truth File
OT	Object Tracking
STIDP	Standoff Technology Integration and Demonstration Program
NIST	National Institute of Standards and Technology
ViPER	Video Performance Evaluation Resource
F4DE	Framework for Detection Evaluations
XML	Extensible Markup Language
SFDA	Sequence Frame Detection Accuracy
ATA	Average Tracking Accuracy
MODA	Multiple Object Detection Accuracy
MODP	Multiple Object Detection Precision
MOTA	Multiple Object Tracking Accuracy
MOTP	Multiple Object Tracking Precision
FDA	Frame Detection Accuracy
STDA	Sequence-based Tracking Detection Accuracy

Table of Contents

Figures

Tables

Equations

1.0 Purpose

The purpose of this report is to document the procedure and the results of the Nano-Corpus Object Tracking Challenge. This challenge was used to evaluate the existing object tracking capabilities of commercial vendors. The challenge consisted of 3 videos, approximately 2 minutes each in length, viewing people as they walked about a parking lot. Each video provided a varying degree of difficulty as determined by number of walkers, occlusions, and vehicular activity (see Table 1). Vendors were tasked with tracking all persons in each scene and providing a unique ID and XY coordinate for each person in every frame of the video.

Table 1. Evaluation Data Set Summary

Scene	Pathway	Minutes	# of walkers	Moving Vehicles	Physical Contact	Merging	Splitting	Occlusions (man-frames)
Basic	South	2:26	5	X				288
Challenging	Z Path	1:42	5	X	X	X		1356
Complex	Serpentine	1:49	6	X	X	X	X	1486

The data sets used for this task were acquired at the Toyota Center in Kennewick, WA. Figure 1 below illustrates the approximate field of view for each scene. A Canon Vixia HFS10 was used for recording, and was set at 1920x1080 pixels and NTSC frame rates. The camera was located on the roof of the Toyota Center and is marked by an X on the figure below.

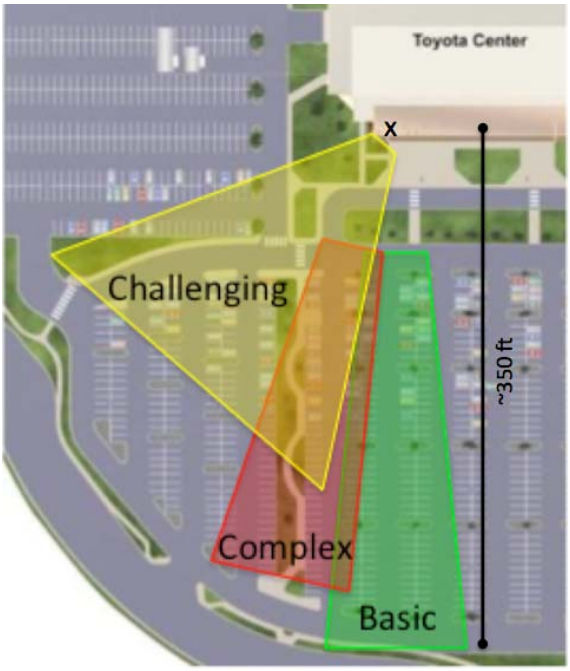

Figure 1. Toyota Center Test Bed

The basic scene depicted a head on shot of 5 people walking towards the building. Limited occlusions were present, and the average distance between people was, with few exceptions, greater than 1 meter. The challenging scene showed 2 groups of people (5 total) approaching the building from different angles, contained more occlusions, and showed side views as well as head on views of the walkers. The complex scene included 6 people merging from different angles, including one getting out of a car to join the group. Many occlusions were present and the traversed path was complex. The lighting conditions during filming varied between 145-195 foot-candles and the sky was overcast producing a diffuse illumination.

PNNL employees reviewed all three scenes and ground truth files (GTF) were created for each using ViPER, an annotation tool created by the University of Maryland. These GTFs were XML documents containing an identification (ID) tag, the size (in pixels), and coordinates of each individual for each frame. The GTFs are the standard to which the vendor files are compared and graded against. The montage in Figure 2 below provides a snapshot of the video for each scene with GTF information graphically displayed on top. The green lines represent the path that each walker traversed and the yellow lines indicate when an occlusion was present for that period of time in the walk path.

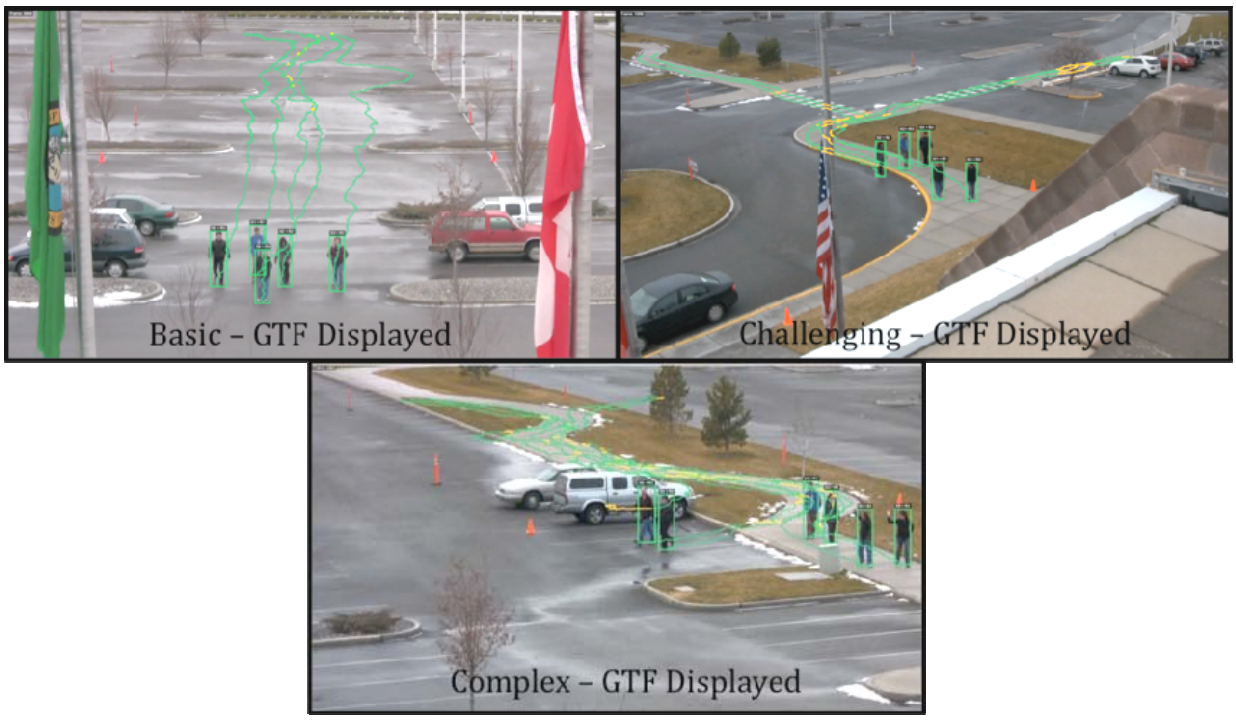

Figure 2. GTF Overlaid on Video

Vendors were expected to deliver an XML file for each scene, in a predetermined format. Results of the comparison between the GTF and a vendor's XML file gauge the ability of each vendor to accurately track an individual through various challenges. The results determine not only if a vendor is capable of meeting current STIDP requirements, but also judges where the general technology stands today and how much work will be required to develop OT technology to the desired level of sophistication.

2.0 Procedure

Evaluation of vendor XML output was standardized to ensure reproducible, defendable results that guarantee fairness across the board. The following details the procedure that was used for this evaluation task, and outlines what documentation was created to allow the data sets to be independently evaluated and confirmed if necessary.

All evaluation was done using the F4DE package from NIST. Specifically, we used the CLEAR Detection and Tracking Scorer algorithms, Version 0.1b (CSV:1.16). Free parameters were set as follows:

- Domain: SV (surveillance)
- Eval: Area
- frameTol: Default (0)
- detthres: Default (1)
- trkthres: Default (1)
- bin: Default (regular)
- MissCost: Default (1)
- FACost: Default (1)
- ISCost: Default (1)

The above parameters produced a command line instruction identical to the following:

./CLEARDTScorer --Domain SV --Eval Area --writeResult result.txt --csv result.csv VendorFile.xml --gtf GroundTruthFile.xgtf --motaLogDir /logDirectory

where the underlined items were changed to call specific files for each run, but all other commands were held constant for all files from all vendors. The Ground Truth Files that were called are specific to a certain video (camera view), but the same file was run for all vendors on each camera view. All result files were labeled with the vendor name, and held in vendor specific folders to ensure no mix-up was made.

A log was created for each vendor file that includes the exact command line text used to run the grading sequence, including specific file/folder locations for each dataset.

All files were edited with a generic header and footer to ensure formatting issues in the front and back matter did not prevent scoring. In the event that a vendor's file did not run through F4DE due to additional formatting issues, reconciliation of the formatting discrepancy was attempted. A log of every character added, deleted or changed was made and saved to ensure traceability and data fidelity. If reformatting the structure did not fix the errors while running the evaluations through F4DE, the file was labeled as "Does Not Run" and received a null score for that scene.

Once a file had been evaluated, the tracking log that was generated by F4DE was used to create an annotated video for human visualization of the data set. This was created using the VidAT tool by NIST. These videos are provided alongside the quantitative grades produced by F4DE in section 3 of this report.

The quantitative grades produced by F4DE were comprised of the following sub-grades:

- SFDA - Sequence Frame Detection Accuracy
- ATA - Average Tracking Accuracy
- MODA - Multiple Object Detection Accuracy
- MODP - Multiple Object Detection Precision
- MOTA - Multiple Object Tracking Accuracy
- MOTP - Multiple Object Tracking Precision

Each sub-grade had a possible score of 1. The data sets were evaluated on 6 criteria; therefore, the highest possible score for each scene was a 6.

The scores from these categories were summed to produce a single score, out of 6 possible points. All three evaluation videos were scored using the above criteria, and the three final scores were used to give a cumulative grade for the evaluation, out of a possible 18 points.

Although, it was stipulated that the nano-corpus would be scored based upon the six metrics above, it was also instrumental to look at the number of ID Switches, False Alarms and Missed Detections in each scene. These faults are components to the various metrics; however, isolation of the values allows one to gain a better appreciation of the detection and tracking capabilities. For this reason, these values were recorded and displayed for all the scenes.

Appendix A provides the mathematical equations behind calculating each sub-grade, as provided by the University of South Florida[1].

3.0 Data

3.1 Participants

Evaluations were performed on five vendors. All data tables and discussion of the results in this document use aliases for each vendor. The names of each vendor are not available at this time due to ongoing contract negotiations.

3.2 Basic Scene Evaluation

The basic scene was a "face-on" film of five people walking directly toward the camera with limited occlusions and large person-to-person distances. See Figure 3 below for an example frame of this video.

Figure 3. Basic Scene

Header and footer file information was generalized for all vendor files to eliminate formatting errors associated with the header structure. In the basic scene, several additional modifications were made to vendor files. Specifically, Company D converted the video file into a separate format, which added 1 frame (out of 3502 original frames). Front and back matter was appended to reflect 3503 frames. Additionally, Company E's file only included data starting at frame 50, and 2 people were tracked for approximately 100 frames (out of 3502 frames). F4DE was unable to recognize the vendor's file as a match to the ground truth file, and subsequently refused to produce a numeric grade for the sequence.

Scores for the basic scene are shown below in Table 2. Accuracy was superior for several vendors, and the precision scores were reasonable. Company E was given a null score.

Table 2. Basic Scene Score

Vendor	SFDA	ATA	MODA	MODP	MOTA	MOTP	Total
A	.33	.29	.75	.37	.74	.38	2.87
B	.61	.43	.87	.67	.87	.67	4.11
C	.53	.48	.94	.56	.94	.56	4.02
D	.59	.49	.99	.60	.99	.60	4.26
E	0	0	0	0	0	0	0*

*File was not gradable.

Company D performed particularity well in both detection and tracking accuracy, with only 3 ID splits, 2 ID merges, and 0 false alarms. Figure 4 below provides an overview of the MOTA components for each vendor.

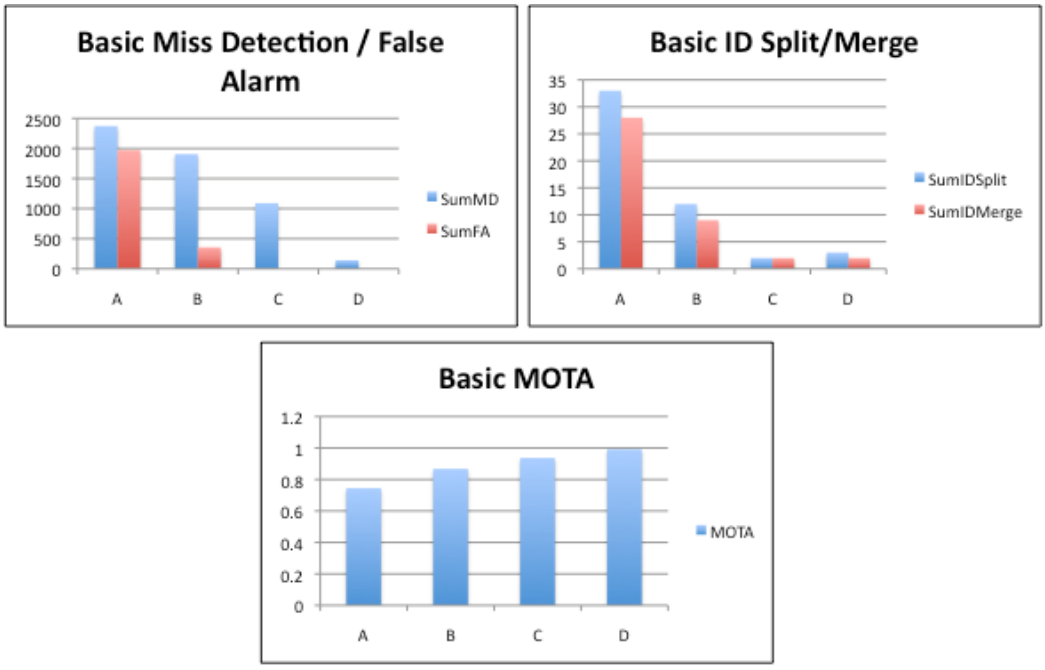

Figure 4. Basic MOTA Summary

3.3 Challenging Scene Evaluation

The challenging scene was a film of two groups walking from different directions before merging and winding their way towards the camera. More occlusions were present in this film than in the basic, with individuals walking behind trees and flagpoles. See Figure 5 below for an example frame of this video.

Figure 5. Challenging Scene

Header and footer file information was generalized for all vendor files to eliminate formatting errors associated with the header structure. In this scene, Company E's file did not include any tracking information for more than the first 100 frames of the video. F4DE was unable to recognize the vendor's file as a match to the ground truth file, and subsequently refused to produce a numeric grade for the sequence. All other vendor files were run without additional format changes for this scene.

Scores for the challenging scene are shown below in Table 3. Performance was lower than the basic scene, which was expected. Company E was given a null score.

Table 3. Challenging Scene Score

Vendor	SFDA	ATA	MODA	MODP	MOTA	MOTP	Total
A	.20	.03	.23	.31	.20	.32	1.29
B	.54	.41	.70	.71	.69	.70	3.74
C	.43	.16	.89	.48	.88	.48	3.31
D	.58	.17	.96	.60	.95	.60	3.86
E	0	0	0	0	0	0	0*

*File was not gradable.

Company D again did particularity well in both detection and tracking accuracy, but Company B has a high overall score with good precision as well as accuracy. Figure 6 below provides an overview of the MOTA components for each vendor.

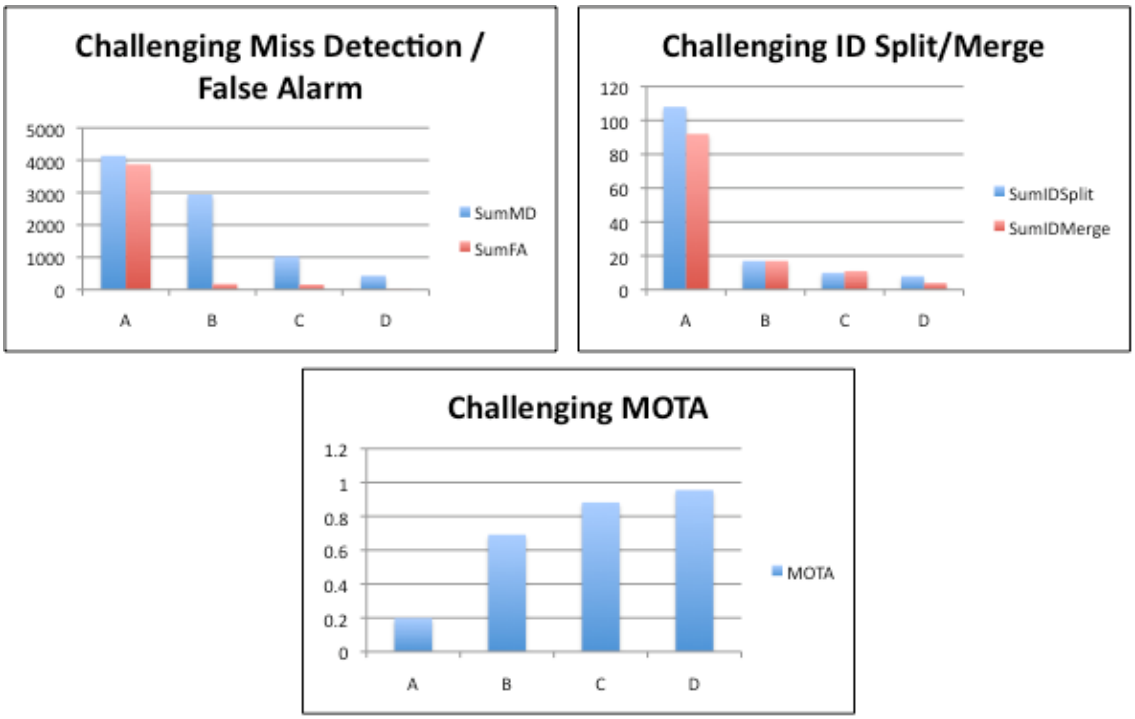

Figure 6. Challenging MOTA Summary

3.4 Complex Scene Evaluation

The complex scene was a film of two groups walking from different directions before merging and making their way down a snaking sidewalk. Another individual drove up in a car, exited and joined the group. Towards the end the group split in two, and the individuals exited from different directions. Significant occlusions were present. See Figure 7 below for an example frame of this video.

Figure 7. Complex Scene

Header and footer file information was generalized for all vendor files to eliminate formatting errors associated with the header structure. In this scene, several additional modifications were made to vendor files. Specifically, Company A had several values in the "Y" pixel location that were -1 and -2. Since a negative pixel value can not exist, these few numbers were changed to 0, with a log created to show the exact locations. This did not have a significant impact on overall score. Additionally, Company E's file only included data starting at frame 50 and ending at frame 528 for all walkers (out of 2629 frames). F4DE was unable to recognize the vendor's file as a match to the ground truth file, and subsequently refused to produce a numeric grade for the sequence.

Scores for the complex scene are shown below in Table 4. In general, performance was higher than expected, and in one case was better than the same vendor's performance in the basic scene. Accuracy was superior for Company D. Since Company E's video was unable to be graded, they were given a null score for each metric.

Table 4. Complex Scene Score

Vendor	SFDA	ATA	MODA	MODP	MOTA	MOTP	Total
A	.53	.28	.86	.57	.85	.57	3.66
B	.54	.21	.74	.67	.73	.65	3.54
C	.49	.27	.87	.52	.85	.51	3.52
D	.59	.23	.97	.59	.97	.58	3.93
E	0	0	0	0	0	0	0*

*File was not gradable.

Company D again performed particularity well in both detection and tracking accuracy. Company D also had 0 false alarms for the sequence with the next closest vendor Company B, at 184 false alarms.

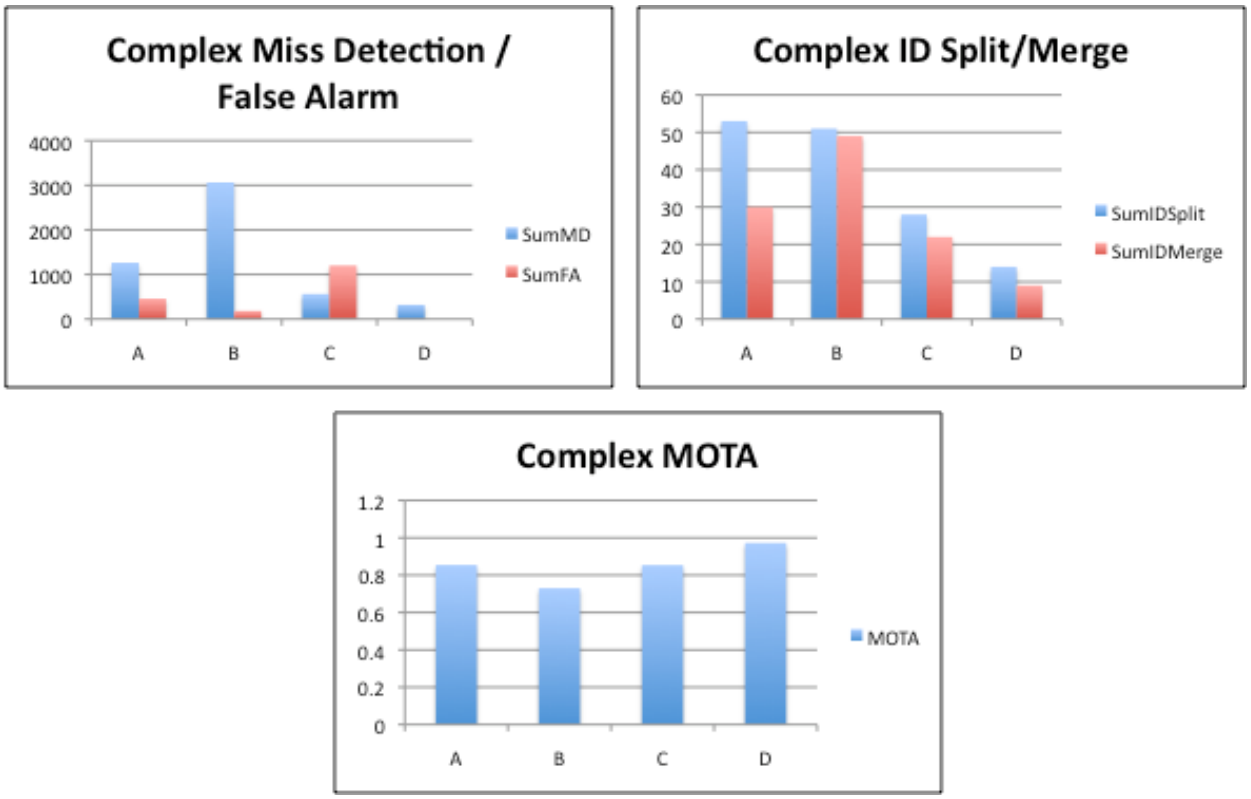

Figure 8. Complex MOTA Summary

4.0 Conclusion

Multiple object tracking scores were reasonably high with several companies standing out as clear leaders in object tracking technology. For the selected grading scheme, the highest score any scene could yield was a 6. Below in Table 5 grand totals of overall vendor performance are shown, out of a possible 18 points.

Table 5. Grand Total Scores

Vendor	Basic Total	Challenging Total	Complex Total	Grand Total
A	2.87	1.29	3.66	7.82
B	4.11	3.74	3.54	11.40
C	4.02	3.31	3.52	10.85
D	4.26	3.86	3.93	12.05
E	0	0	0	0*

*All files were not gradable.

Company D's performance stands out from a numerical standpoint, as well as from review of the amount of ID Switches, False Alarms and Missed Detections in each scene. Company B's overall numerical score was second best; however, this was misleading. Although Company C's overall score was lower, they performed significantly better than Company B in the level of ID Switches, False Alarms and Missed Detections. See Appendix B for a detailed breakdown of the MOTA components.

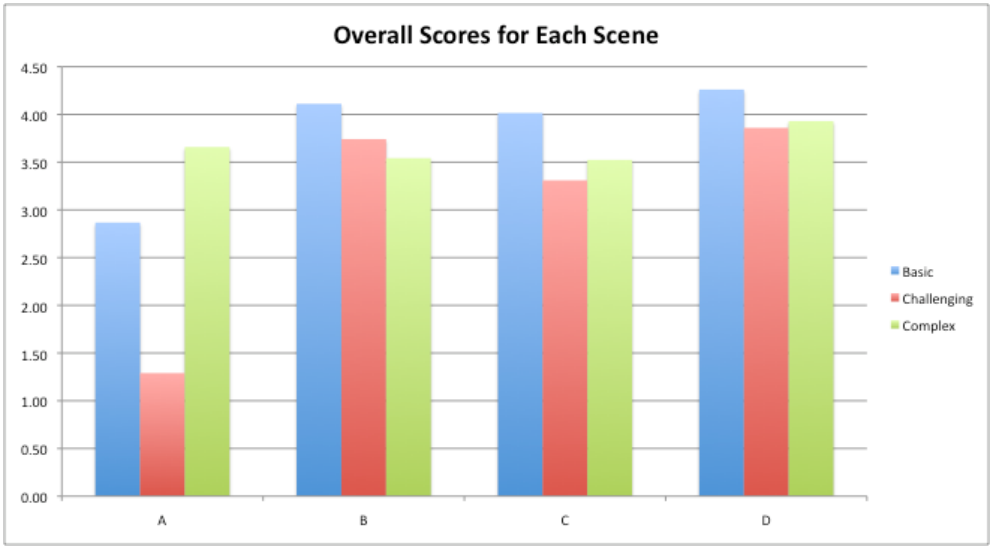

Figure 9. Overall Score Summary

5.0 Bibliography

1. R. Kasturi et al. *Performance Evaluation Protocol for Face, Person, and Vehicle Detection & Tracking in video Analysis and Content Extraction* from Supplemental Materials for *Framework for Performance Evaluation of Face, Text and Vehicle Detection and Tracking in Video: Date, Metrics and Protocol* IEEE TPAMI February 2009 vol. 31 no. 2 pp. 319-336

   http://www.computer.org/portal/web/csdl/doi/10.1109/TPAMI.2008.57

Appendix A: Sub-Grade Equations[1]

Sequence Frame Detection Accuracy or SFDA is specified as:

$$SFDA = \frac{\sum_{t=1}^{t=N_{frames}} FDA(t)}{\sum_{t=1}^{t=N_{frames}} \exists(N_G^{(t)} \; OR \; N_D^{(t)})} \tag{A.1}$$

where

$$FDA(t) = \frac{\text{Overlap Ratio}}{\frac{N_G^{(t)}+N_D^{(t)}}{2}} \tag{A.2}$$

and

$$\text{Overlap Ratio} = \sum_{i=1}^{N_{mapped}} \frac{|G_i^{(t)} \bigcap D_i^{(t)}|}{|G_i^{(t)} \bigcup D_i^{(t)}|} \tag{A.3}$$

N_G is the number of ground truth objects and N_D is the number of detected objects. G_i is a given ground truth object, and D_i is a given detection object matched to the same *i*.

Average Tracking Accuracy or ATA is defined as:

$$ATA = \frac{STDA}{\frac{N_G+N_D}{2}} \tag{A.4}$$

where

$$STDA = \sum_{i=1}^{N_{mapped}} \frac{\sum_{t=1}^{N_{frames}} \left[\frac{|G_i^{(t)} \cap D_i^{(t)}|}{|G_i^{(t)} \cup D_i^{(t)}|}\right]}{N_{(G_i \cup D_i \neq \emptyset)}} \tag{A.5}$$

Multiple Object Detection Accuracy or MODA is specified as:

$$MODA(t) = 1 - \frac{c_m(m_t) + c_f(fp_t)}{N_G^t} \quad \text{(A.6)}$$

where c_m and c_f are the cost functions for the missed detects and false alarm penalties. Additionally m_t and fp_t are the number of misses and the number of false positives for a given frame *t*, respectively.

Multiple Object Detection Precision or MODP is specified as:

$$MODP(t) = \frac{(Overlap\ Ratio)}{N_{mapped}^t} \quad \text{(A.7)}$$

where N^t_{mapped} is the number of mapped object sets in frame *t*. Overlap ratio is defined above in equation A.3.

Multiple Object Tracking Accuracy or MOTA is defined as:

$$MOTA = 1 - \frac{\sum_{i=1}^{N_{frames}} (c_m(m_i) + c_f(fp_i) + \log_e(id_{switches}))}{\sum_{i=1}^{N_{frames}} N_G^i} \quad \text{(A.8)}$$

Multiple Object Tracking Precision or MOTP is defined as:

$$MOTP = \frac{\sum_{i=1}^{N_{mapped}} \sum_{t=1}^{N_{frames}} \left[\frac{|G_i^{(t)} \cap D_i^{(t)}|}{|G_i^{(t)} \cup D_i^{(t)}|} \right]}{\sum_{j-1}^{N_{frames}} N_{mapped}^j} \quad \text{(A.9)}$$

where N_{mapped} refers to the mapped objects over the entire track as opposed to just the frame and N^j_{mapped} refers to the number of mapped objects in the j^{th} frame.

Appendix B: MOTA Component Details

Details are provided of the MOTA summary for each scene. Terms are defined as follows:

Missed Detection: Ground Truth File labeled this individual on this frame as present and un-occluded, but the vendor file failed to identify an overlapping region for the given walker.

False Alarm: Vendor file contains a region that has no matching Ground Truth region for that frame.

ID Split and Merge: Vendor file contains a region that changes which walker in the Ground Truth it is matched to.

# of Evaluated GT: Sum of the number of un-occluded, present walkers in all frames of the Ground Truth. (5 walkers in 100 frames = 500; 5 walkers in 100 frames with 1 person occluded for 50 frames =450)

Basic MOTA:

Vendor	Missed Detection	False Alarms	ID Splits	ID Merges	# of Evaluated GT	MOTA
A	2371	1970	33	28	17222	0.74
B	1905	357	12	9	17222	0.87
C	1091	0	2	2	17222	0.94
D	142	0	3	2	17222	0.99

Challenging MOTA:

Vendor	Missed Detection	False Alarms	ID Splits	ID Merges	# of Evaluated GT	MOTA
A	4137	3870	108	92	10241	0.20
B	2943	182	17	17	10241	0.69
C	1025	162	10	11	10241	0.88
D	431	19	8	4	10241	0.95

Complex MOTA:

Vendor	Missed Detection	False Alarms	ID Splits	ID Merges	# of Evaluated GT	MOTA
A	1264	460	53	30	12401	0.85
B	3062	184	51	49	12401	0.73
C	563	1203	28	22	12401	0.85
D	320	0	14	9	12401	0.97

Standoff Technology Integration and Demonstration Program

http://stidp.labworks.org