FUSION OF MULTIPLE SENSORS FOR SAFE OPERATIONS

Paul H. Haley, PhD
General Dynamics Robotic Systems
Pittsburgh, PA

Susan M. Thornton, PhD
General Dynamics Robotic Systems
Pittsburgh, PA

Robert R. Mitchell, PhD
William P. Zachar
Mike Hoffelder
Steven McLean
General Dynamics Robotic Systems

ABSTRACT

Operating safely in cluttered environments is critical to future autonomous robotic operations as exemplified by FCS Risk 213. In support of this requirement, the Robotics Collaborative Technology Alliance (RCTA) program, sponsored by the Army Research Lab (ARL), has supported research tasks and corresponding integration and test events from 2006 through 2009. Multiple sensor systems, including scanning LADARs and stereo camera pairs, have been used to detect, track, and predict the future motion of obstacles in the close proximity of unmanned ground vehicles. These sensors produce frames of data at rates ranging from 6 to 30 Hertz. Resulting algorithm outputs are correlated to the local world and detection results both above and below the thresholds of the individual algorithms are recorded in a common format. This paper describes two methods for fusing the detection data. The first is a simplistic approach which implements a majority voting scheme amongst the algorithm results. The second, more rigorous, approach uses a "Strength-of-Detection" (SoD) method that utilizes an association step incorporating an error covariance model for each sensor, and also allows for cases where only a subset of the sensors report a detection. Results show that fused detection performance is far better than any single output due to the uncorrelated nature of single-sensor false alarms. We present representative results for both individual sensors and fused outputs.

INTRODUCTION

Safe Operations represents a critical goal for autonomous ground vehicles. The Robotics Collaborative Technology Alliance (RCTA) has been supporting research in support of this goal since 2006. The research includes tasks for detecting and recognizing humans in the vicinity of a ground vehicle using LADARS, EO/IR stereo cameras, and radars. The primary emphasis for each task has been on the use of an algorithmic approach operating on the output of a single sensor to achieve the best possible detection performance while suppressing false alarms.

However, it is well known that the fusion of multiple sensors offers great potential for improvement over the performance of any single sensor modality because the various sensors tend to have complementary strengths and weaknesses. For example, LADARs provide very good range information, but do not discriminate on appearance as well as high resolution imaging sensors. Conversely, cameras provide much better appearance data but, even using stereo pairs, provide comparatively coarse range accuracy.

Consequently, we have undertaken an effort to fuse the results of multiple sensor-algorithm outputs. Here we describe two efforts to implement detection-level fusion. One is a simple "majority vote" approach; which we show has achieved results better than any single-source approach. The second more rigorous approach is currently under development. It is a non-parametric Bayesian approach that makes more complete use of the single-sensor results and leads to an unbiased ROC curve for which performance can be optimized.

In the balance of this paper, we first describe the Safe Operations experiments we have conducted this year to evaluate both single-sensor and fused detection performance. We then discuss human detection results obtained for individual algorithms operating on single-sensor data. Next we describe various possible approaches to fusion and address the issue of associating individual detection results. We describe the technical approach and results to date for our majority vote fusion approach. We also describe our non-parametric Bayesian fusion approach and conclude with discussion of plans for future research.

SAFE OPERATIONS EXPERIMENT DESCRIPTION

The RCTA program has conducted a series of Safe Operations experiments beginning in 2006. The experiment described here occurred on January 21 – 23, 2009 at the National Institute of Standards and Technology (NIST) facility in Gaithersburg, MD. It was conducted by personnel from the Army Research Laboratory (ARL), NIST, and General Dynamics Robotic Systems (GDRS), which is the Lead Industrial Organization for RCTA. The sensor test vehicle was a Chevrolet Suburban equipped with the following sensors (Fig. 1):

- 2 GDRS Fourth Generation scanning LADARs (Gen IV) which each produce approximately 10 frames per second (fps) over a 90 degree horizontal field of view (FOV) and together provide nearly 180 degree FOV in front of the vehicle.
- A Sick LADAR which produces approximately 18.5 frames per second over a 100 degree horizontal FOV and 1 degree vertical FOV.
- A stereo pair of Hitachi HV-F31 Progressive Scan Color 3-CCD Cameras that produce raw color imagery (1024 x 768) at 15 fps with a horizontal FOV of approximately 60 degrees.

Figure 1: Sensor test vehicle used in Safe Operations experiments at NIST in January, 2009

The NIST test site included approximately 300 meters of two-lane paved streets, including an intersection where the test vehicle turned. The test consisted of 40 runs, each with eight human test subjects who were moving on the paths indicated in Figure 2. The test vehicle moved at either 15 kph (20 Runs) or 30 kph (20 Runs). The tests included two levels of clutter, which we refer to as "open" and "cluttered" and are shown in Figures 3 and 5.

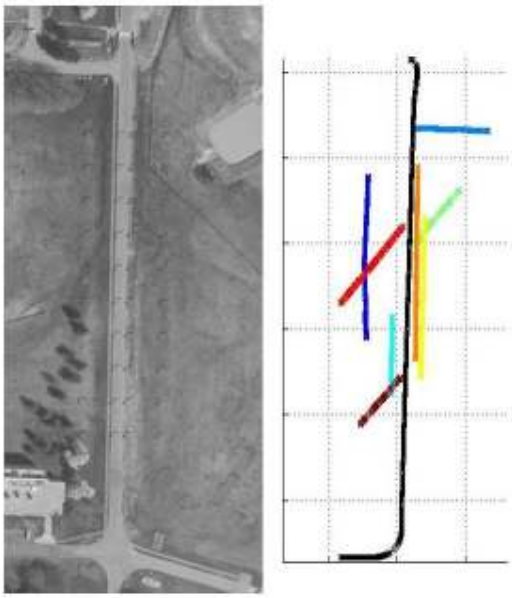

Figure 2: (Left) Overhead imagery of Safe Operations experiment location at NIST in January, 2009. (Right) Test vehicle path (in black) and paths of eight moving humans (in color).

In order to generate ground truth data, we used an automated position recording system as described in [1]. Throughout the testing, all human subjects wore helmets equipped with transceivers for the ground truthing system as shown in Figure 3.

SINGLE SENSOR RESULTS

Figure 4 shows human detection results from four sensor/algorithm sources corresponding to the "open" conditions of Figure 3. It is important to note that the individual results depicted here are on a frame by frame basis. Better individual results can be, and have been, obtained using temporal information.

The results are from two algorithms processing GDRS Gen IV LADAR data, an algorithm processing Sick data, and another algorithm processing data from a stereo camera pair. The LADAR results are all quite good for the Open configuration. Descriptions of the stereo vision algorithms and results on prior data are given in [2]. The LADAR processing algorithms and results on previous data collections have been reported [3.4].

Figure 3: "Open" configuration for Safe Operations experiment at NIST in January, 2009

Figure 5: "Cluttered" configuration for Safe Operations experiment at NIST in January, 2009. Note vehicles, barrels, crates, and other objects not present in the "open" configuration.

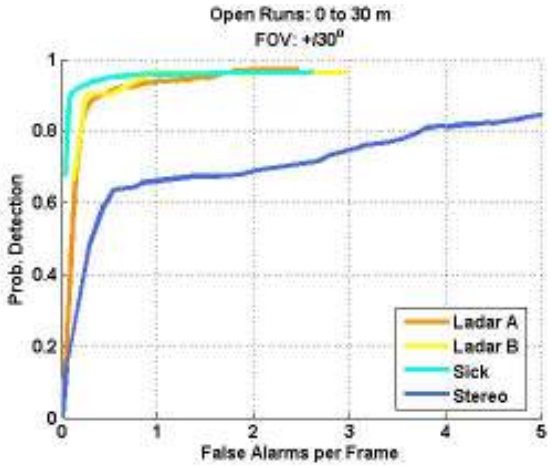

Figure 4: ROC curves for four sensor/algorithm outputs using the "Open" configuration for the Safe Operations experiment at NIST in January, 2009

The stereo vision detection results reported here were degraded by sensor misalignment as well as occlusion issues. We did not attempt to correct these problems because the focus here is on *fusing* the results with those from other algorithms.

Figure 6 shows human detection results from the same four sensor/algorithm sources, but now corresponding to the "cluttered" conditions of Figure 5. Again we note that the individual results depicted here are on a frame by frame basis and that better individual results can be obtained using temporal information.

The LADAR results are all degraded compared to the "open" results. This is expected since the LADAR data do not support appearance-based classification as well as video data do. The Sick results are affected most strongly, probably because of the sensor's limited vertical FOV.

It is noteworthy that the stereo vision results are actually better for the cluttered configuration. This is probably due to both the more robust appearance-based classification with more pixels on target, and to a richer set of features for stereo processing.

This variation in performance across sensors and algorithms provides motivation to fuse the results together. The four sources considered here tend to be correlated for targets, while being uncorrelated for false alarms. The future addition of other sensors such as IR and radar should provide further complementary capabilities that enhance the value of fusion processing.

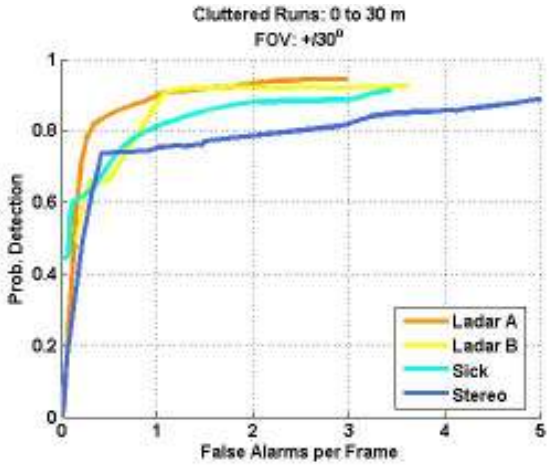

Figure 6: ROC curves for four sensor/algorithm outputs using the "Cluttered" configuration for the Safe Operations experiment at NIST in January, 2009

PROMISE AND CHALLENGE OF SENSOR FUSION

Despite the obvious potential advantages of sensor fusion, there are significant challenges as well. Attempting fusion at the sensor/data level is theoretically best but is computationally daunting and is not addressed here. Other options are feature-level and detection-level fusion.

Thus far we have addressed detection-level fusion because it is computationally tractable and has produced the promising results shown below. A key challenge for detection-level fusion is the correct spatial and temporal association of the detections from the various sensor/algorithm sources, which we discuss next.

Association of Multi-Sensor Detections

Incorrect associations of detections can produce results that are worse than single-source results. Failure to associate two detections of the same object can produce a false alarm. Incorrect association of two detections that actually correspond to distinct objects can produce missed target detection.

Consequently, we have devoted a great deal of effort to intrinsic and extrinsic calibration of sensors to minimize sensor pointing errors. We also model and account for remaining errors during the data association process.

Shown in Figure 7 is the association process for one frame of results from the January '09 RCTA Safe Operations experiments. The inset at the lower left shows detections from each source using the color key above the inset. The arrows represent velocities of movers. The blue circled set of detections is expanded at the upper right to show all five responses in this case.

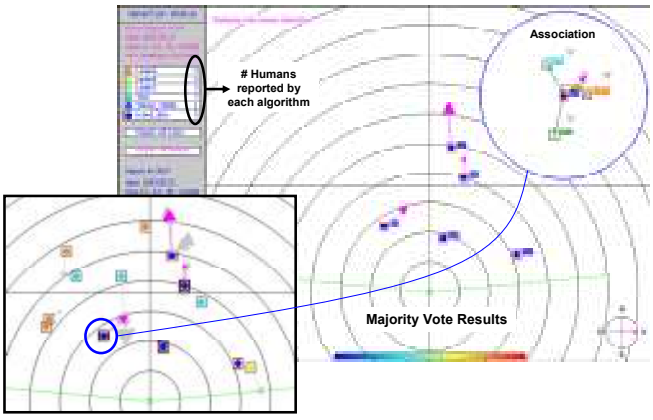

Figure 7: Association of multi-source detections are shown in the inset, with one example association expanded to illustrate all responses to the same ground truth object. Also shown are the fusion results using the Majority Vote approach.

After the association process, we have a list of detections where, for each detection, there is one of 3 outcomes for each of N sensor/algorithm sources:

1. Response above threshold (target), with SoD – this increments r_h, the counter for human detections
2. Response below threshold (clutter), with SoD – this increment r_c, the counter for clutter detections
3. No response – this increments r_x

The response set must satisfy

$$r_h + r_c + r_x = N \tag{1}$$

MAJORITY VOTE FUSION

Our Majority Vote approach for fusion is quite simple, but is proving to be effective. It does not consider the strength of detection (SoD), and does not explicitly consider the number of responses or which sensor/algorithm responded. A fused detection is declared if the response set satisfies:

$$\frac{r_h}{r_h + r_c} > 0.5 \tag{2}$$

Results using the Majority Vote approach are shown in Figures 8 and 9. The ROC curves represent the fused result while the points are the operating points for each individual algorithm. The fused curves allow us to select an operating point that is superior to each of the individual ones.

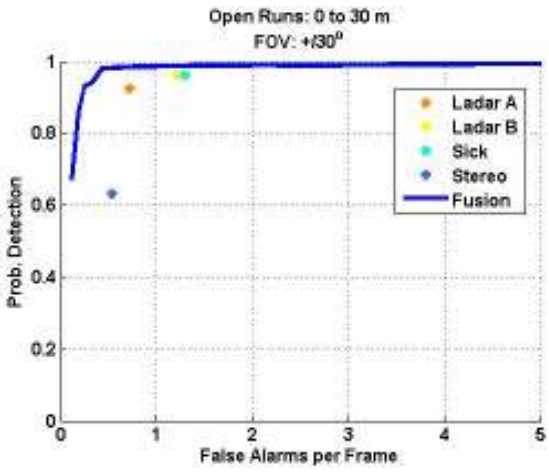

Figure 8: Fusion results using the Majority Vote approach for the "open" configuration of the RCTA '09 Safe Operations experiment.

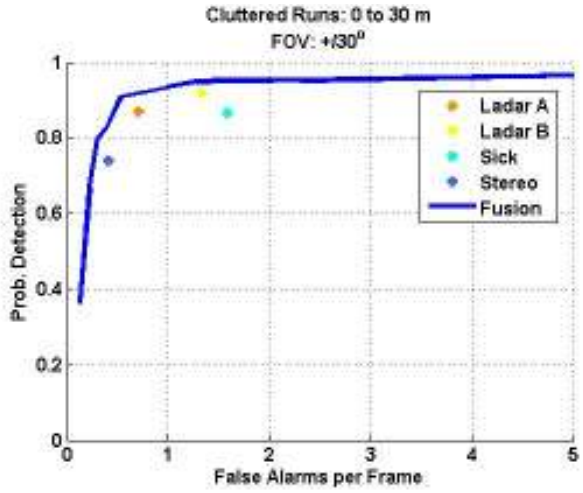

Figure 9: Fusion results using the Majority Vote approach for the "cluttered" configuration of the RCTA '09 Safe Operations experiment.

NON-PARAMETRIC BAYESIAN FUSION

Although the previously described results using simple majority vote fusion are encouraging, this approach suffers from a number of weaknesses:

- It does not explicitly use the number of responses. Thus if we have 4 cuers, for example, the case of two target responses with the other two cuers not responding (r_h=2, r_x =2) is treated the same as four target responses (r_h=4).
- No direct use is made of the SoD values from the cuers so that a detection barely above threshold is treated the same as a very strong detection.
- Conversely, a response whose strength is just below target threshold for that cuer is treated the same as one with a very low SoD value.
- All sources are treated equally so there is no weighting of results even though analysis of individual results may indicate good reason to do so. A certain response set (with SoDs) might be strongly indicative of a true target even though majority voting would tend to reject it.

Consequently, we are also pursuing a more rigorous fusion approach, which is described next. In this approach, we consider an isolated object in the field of view of N different collinearly mounted (or nearly collinearly mounted) sensors with associated detection processing algorithms. Depending on various factors there are conceptually 2^N responses of these sensor/algorithm pairs as to whether they declare a detection. All N may respond, none may respond or any r out of N may respond. We assume that each algorithm puts out an SoD, q_i, giving its strength of detection. Let S_k , k = 1 to 2^N-1, be sets giving the indices of q_i for all possible responses of the N algorithms. We need not model the null response. For example if N = 3 then the S_k would be defined as

$$\begin{aligned} S_1 &\equiv [123] \\ S_2 &\equiv [12] \\ S_3 &\equiv [13] \\ S_4 &\equiv [23] \\ S_5 &\equiv [1] \\ S_6 &\equiv [2] \\ S_7 &\equiv [3] \end{aligned} \qquad (3)$$

In this fusion approach, we construct joint densities modeling the SoD response of the algorithms to targets and False Alarms. We define

$$f_T(q)_k = \text{density of q for index set k over targets} \qquad (4)$$

$$f_{FA}(q)_k = \text{density of q for index set k over False Alarms} \qquad (5)$$

Then a fusion algorithm can be implemented by first determining k and then comparing

$$r_k(q) = f_T(q)_k \,/\, f_{FA}(q)_k \qquad (6)$$

to a threshold and declaring a target if the threshold is exceeded.

The SoD is a measure of how strongly a particular algorithm rates an object as being a target. Targets should have large SoDs while false alarms should have small SoDs. If one can compute a valid ROC curve that is everywhere concave, then one can use the parameter implementing the ROC curve as an SoD value.

Because not all of the individual algorithms whose results we are fusing can readily produce SoDs with the desired properties, we have developed and implemented procedures to remap each dimension of the sample features (SoDs) to the uniform density for both target and false alarm features. We also detect point mass components of the features and insert in the cumulative map an interval proportional to their number at appropriate points. An example of this remapping is shown in Figures 10 and 11.

We next combine linearly the Target cumulative map and the False Alarm cumulative map as a convex combination for each dimension of the feature space. We apply this map to the feature data of Targets and False Alarms and replace the point mass components by uniformly distributed pseudo data in proportion to the point mass values at the appropriate

intervals. This produces two new remapped sets of feature data across Targets and False Alarms. An example of the remapping corresponding to Figures 10 and 11 is shown in Figure 12.

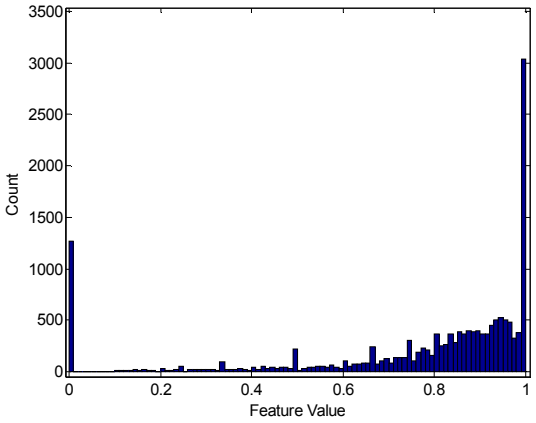

Figure 10: Example histogram of target feature.

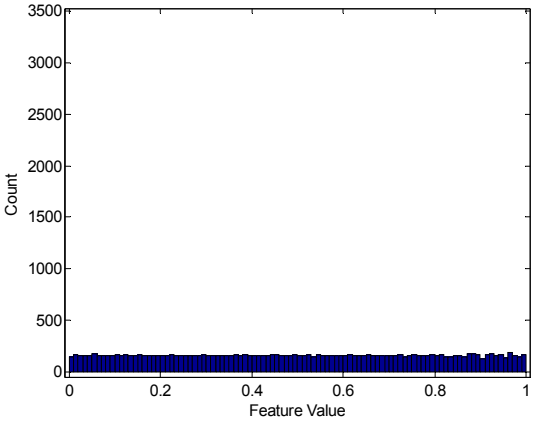

Figure 11: Example histogram of target feature mapped to uniform density.

Finally we construct a filter which is a hypercube in the n-dimensional space whose dimension is the number of responses to be fused. The size of the filter in each dimension constitutes the smoothing for that input result to fusion processing. The final fusion result is the convolution of the filter with the previously described joint density function.

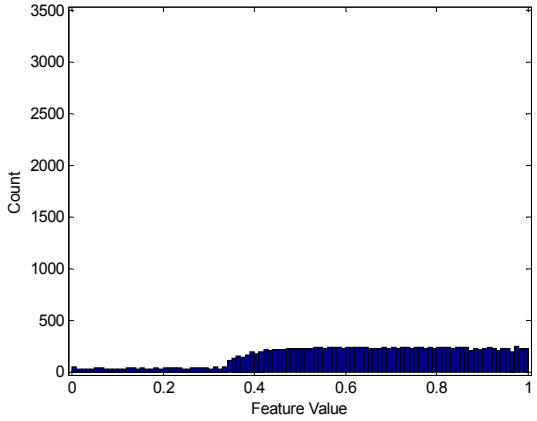

Figure 12: Example histogram of normalized target feature after weighting remapping.

Training our non-parametric Bayesian Fusion Classifier requires statistics for the 2^{N-1} response sets resulting from N (for the current usable data set N = 4) different cuers detecting commonly sensed objects. Hence algorithms were developed to match ground truth data with tracks to mark the track as a target or non-target and then to accumulate statistics on the value and source of each SoD value attached to the track to generate joint statistics over 15 different response sets. Below in Table 1 is a listing of the number of detections that fell into each set for a particular run of tests conducted in September, 2007. For these 28 test runs there were over 68,000 target instances in the data as well as nearly 500,000 non-target instances.

Response Sets	Cuer 1	Cuer 2	Cuer 3	Cuer 4	Targets	Non-targets
1	X				600	31570
2		X			1322	28576
3	X	X			8684	78229
4			X		11031	285142
5	X		X		344	6442
6		X	X		13148	26207
7	X	X	X		20148	34294
8				X	23	388
9	X			X	3	111
10		X		X	5	409
11	X	X		X	41	1342
12			X	X	60	411
13	X		X	X	31	373
14		X	X	X	306	326
15	X	X	X	X	12533	2202
Totals					68279	496022

Table 1: Response set data for RCTA Safe Operations testing conducted in September, 2007.

For each response set the number of variates is equal to the number of columns marked with 'X' in a row. For each response set a density for the corresponding SoD values over

the target and non-target sets is estimated and the Bayesian classifier is implemented using those densities

Our non-parametric Bayesian fusion algorithm requires very little computation at runtime. The computationally intensive portion is the off-line training of the classifier. The algorithm consists of the following steps:

1. Obtain SoDs at time t_k
2. Determine response set k on non-empty SoD pattern
3. Pick up all tables associated with the response set k
4. Make decision as follows:
   - If zero density target set → declare False Alarm
   - If zero density false alarm set → declare Target
   - Else map q to q_{map} using previously defined mapping and compute from density tables

$$r_k(q_{map}) = f_T(q_{map})_k \,/\, f_{FA}(q_{map})_k \qquad (7)$$

and declare Target if

$$r_k(q_{map}) \geq r_{threshold} \qquad (8)$$

The approach of using a mutually exclusive response set Bayesian classifier offers a number of gains. First, the response set parsing of detection outcomes even without SoD values yields 15 points on the system ROC curve using ratios of true detections and false alarms for each response set. Adding SoD values spreads these 15 points via the associated likelihood ratio. By the Neyman-Pearson lemma, this will be as good, or better, than the original 15-point curve.

Second, the mutually exclusive response set parsing makes optimal use of detection outcomes versus simplistic "OR"ing which improves P_d at the expense of increased false alarms or versus "AND"ing which reduces false alarms at the expense of p_d.

Third, the common use of likelihood ratio across ROC curves specific to each set permits calculation of a system ROC curve via addition of target cumulative sums across each set and addition of false alarm cumulative sums across each set.

CONCLUSION

Our work is based on Safe Operations experiments being conducted by the Robotics Collaborative Technology Alliance program. To date the focus of those experiments has been on individual sensor modalities and algorithms. As a basis for our fusion processing, we first present individual results from an experiment in January 2009 using the GDRS Gen IV scanning LADAR, a Sick LADAR, and a stereo vision system. Then we describe a simple majority vote fusion approach and present ROC curves demonstrating its use. Finally we describe a more rigorous non-parametric Bayesian fusion approach that use the strength of detection (SoD) reported by each of N contributing sensor/algorithm inputs. This approach models the resulting $2^N - 1$ response sets as joint densities based on prior ground truthed results. We report the application of this approach to an experimental data set from 2007. We will report final results on that and other data in future work.

REFERENCES

[1] B. Bodt, "A Formal Experiment to Assess Pedestrian Detection and Tracking Technology for Unmanned Ground Systems," 26th Army Science Conference, Orlando FL, December, 2008.

[2] M. Bajracharya, B. Moghaddam, A. Howard, L. Matthies, "Detecting personnel around UGVs using stereo vision," Proc. SPIE, Vol 6962: Unmanned Systems Technology X, April 2008.

[3] L. Navarro-Serment, C. Mertz, M. Hebert, "Predictive Mover Detection and Tracking in Cluttered Environments," Proc. of the 25th. Army Science Conference, November, 2006.

[4] S. Thornton, M. Hoffelder, D. Morris, "Multi-sensor Detection and Tracking of Humans for Safe Operations with Unmanned Ground Vehicles," Human Detection from Mobile Vehicles Workshop, ICRA, May, 2008.