Multisensor 3D tracking for counter small unmanned air vehicles (CSUAV)

Juan R. Vasquez[a], Kyle M. Tarplee[a], Ellen E. Case[b], Anne M. Zelnio[b], Brian D. Rigling[b]

[a]Numerica Corporation, 2661 Commons Blvd Suite 210, Beavercreek, OH 45431;
[b]Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH 45435

ABSTRACT

A variety of unmanned air vehicles (UAVs) have been developed for both military and civilian use. The typical large UAV is typically state owned, whereas small UAVs (SUAVs) may be in the form of remote controlled aircraft that are widely available. The potential threat of these SUAVs to both the military and civilian populace has led to research efforts to counter these assets via track, ID, and attack. Difficulties arise from the small size and low radar cross section when attempting to detect and track these targets with a single sensor such as radar or video cameras. In addition, clutter objects make accurate ID difficult without very high resolution data, leading to the use of an acoustic array to support this function. This paper presents a multi-sensor architecture that exploits sensor modes including EO/IR cameras, an acoustic array, and future inclusion of a radar. A sensor resource management concept is presented along with preliminary results from three of the sensors.

Keywords: Unmanned air vehicles, tracking, acoustic array, beam form

1. INTRODUCTION

1.1 Problem Description

.Recent concerns regarding the use of small unmanned air vehicles to bring harm to civilian and military personnel by adversarial groups has prompted the need to counter this potential threat. The focus of this research effort is on multi-sensor detection, track, and identification (ID) of SUAVs. The primary objective is to build a point defense system that can be fielded with a focus on countering a surveillance threat with some capability of countering other threats. The small size of these UAVs makes it difficult for any single sensor mode to robustly track and ID these targets. This paper will describe the system design to include multi-sensor multi-target tracking with cueing between sensors and a sensor resource management concept. As an example, we intend to use a radar system to cue EO/IR sensors and incorporate acoustic sensors to validate the target as an SUAV. Fusion of the EO/IR camera data with the radar range data will lead to a 3D track and ID of the target via human target recognition. The acoustic array can increase the level of confidence in the ID of the target as a threat based on classification of the acoustic signatures of potential targets. The ID state of a given track will be Threat / No-Threat / Unknown. It is important to note that other cueing approaches beyond the one example described here will be available, leading to a robust set of options such that reliance on any one sensor will be avoided.

1.2 System Concept

The fundamental approach to this multi-sensor system is to divide the problem into the areas of detect, track, and ID. In order to provide protection for point defense, the sensor suite needs to conduct wide area detection at long range. The current goal is to detect threats at a long range, build tracks from these detections, and recognize that this will occur with some acceptable false alarm rate. The false alarms resulting from the wide area detection will be mitigated by narrow area detection and ID components. The sensors are therefore divided into three categories: Wide Field-Of-View (WFOV), Medium Field-Of-View (MFOV) and Narrow Field-Of-View (NFOV). Some level of processing will be conducted at the sensor level to convert raw sensor data to detection information that is provided to a centralized tracking system. Additional external data sources, such as a cooperating air traffic control center, can provide information on known air targets. While these air targets will be tracked by the CSUAV system, knowledge of their presence and 3D position will be used to mitigate false alarms.

The centralized tracker is a Multiple Frame Association (MFA) tracking system – an optimization-based multi-hypothesis tracker (MHT) capable of initiating, maintaining, and deleting tracks for kinematic and feature-based measurement information [1]. The MFA tracker outputs 3D track and ID information to sensor resource management (SRM) functions which determine the focus areas and dwell time for the NFOV sensors. The track and ID information is also presented to the operator who can interactively control the system's behavior. Note that an accurate Automatic Target Recognition (ATR) system would be extremely challenging, so we are assuming that a single operator with minimal training could provide the target recognition function. The acoustic sensor array can support the human threat evaluation by classifying the target as threat / no-threat by using the acoustic signature of targets to classify tracks as UAV / not-UAV. Other functions could include false alarm mitigation in conjunction with the automated system (i.e., tell the system that a detection is not a target and thus relieve the SRM function from cueing NFOV sensors to this track).

Figure 1 illustrates the primary components of the system, and detailed descriptions of these components are presented in the sections that follow. The radar system is a WFOV sensor that provides long range detection along with range/range-rate data. It is a rotating radar providing 360 x 25 degrees FOV. The EO/IR sensor suite is a gimbaled combination of two EO cameras and a Mid-Wave InfraRed (MWIR) camera. The EO cameras are designated as NFOV (1.7 degrees FOV), which is a color camera used for target ID in conjunction with the NFOV MWIR (4 degrees FOV). The MFOV EO (20 x 13 degrees FOV) camera provides monochrome video for detection and tracking. This MFOV camera has two functions. Note that the radar will cue the video sensors, but given their narrow FOV's required for ID, it will be difficult to provide a cue accurate enough (due to residual biases) to ensure the NFOV cameras acquire the targets. The MFOV camera supports acquisition of the targets for use by the NFOV sensors. The MFOV camera also provides close-in tracking for cases where the SUAV falls below the minimum detection range of the radar. The gimbaled EO/IR platform will be cued via the SRM function based on system level detections/tracks. The acoustic array will employ beam forming to establish the angle of arrival of acoustic data and to increase the signal-to-noise ratio for this data. The MFA tracker will provide track locations to the acoustic array, such that the array can beam form in the direction of potential targets, identify these targets, and provide these IDs to a user for evaluation in conjunction with visual data.

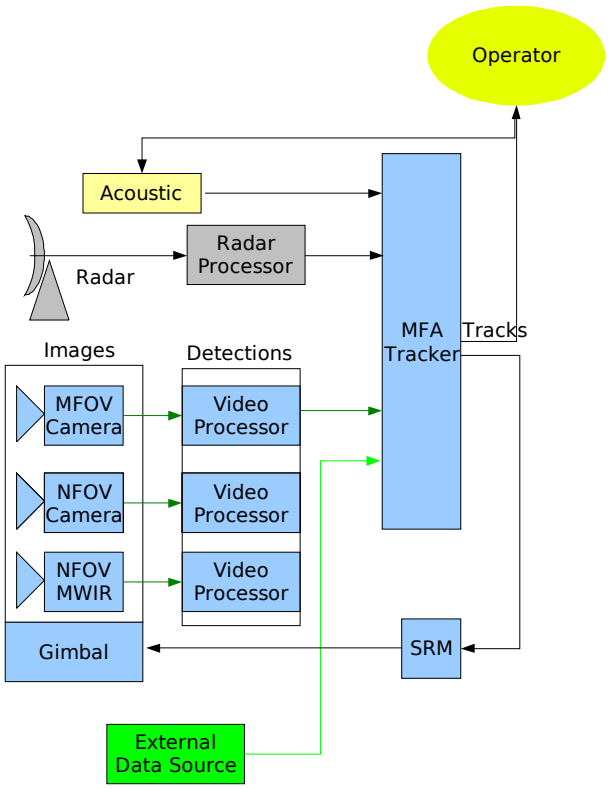

Figure 1: CSUAV system architecture.

Two important aspects of this system design, which drive the need for multiple sensors, are the need for 3D track information and robustness to operating conditions. For instance, the EO/IR sensors, while capable of providing azimuth and elevation estimates, lack the ability to produce accurate range and range rate information provided by the radar. A network of EO/IR sensors in the proper geometry could be used to form 3D tracks, but the radar system is preferred as it also provides a nighttime capability.

Figure 2 shows the sensor coverage (flattened in 2D) for the system giving an indication of the SRM concept that will be employed. The SRM and cueing options are discussed in detail later, but this figure clearly shows the need to provide a control mechanism for the sensors. Notice that the gimbaled MFOV-EO, NFOV-EO, and NFOV-IR sensors must be steered to the moving target to be of any use. The SRM will use track states to keep the gimbal pointed roughly on target. Once the MFOV-EO camera detects the target, the SRM will then fine tune the gimbal so that the NFOV sensors can observe the target.

Figure 3 illustrates the state flow for targets under track. A target list is generated by the MFA tracker, and all tracks are initially identified as Unknown. The SRM function will slew the EO/IR sensor suite to unknown targets, providing the user with high-resolution video for use in identifying the target threat state. These IDs are used to update the target list, which affects future decisions by the SRM. Once a threat is declared, the SRM changes to a single target track mode in which all assets are dedicated to providing a 3D track solution on the Threat.

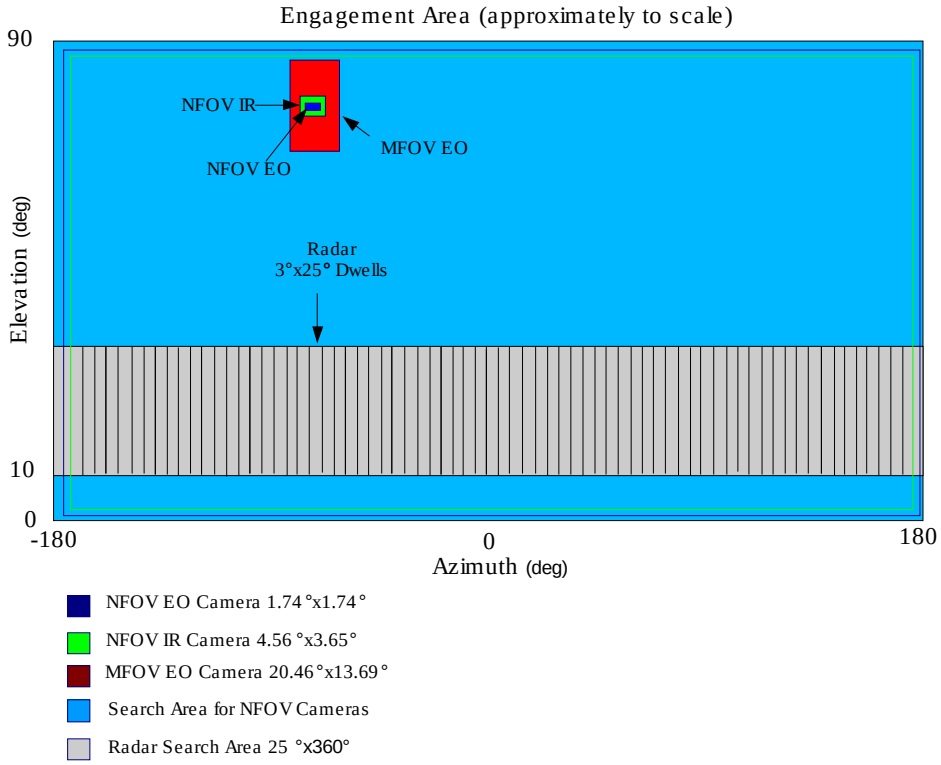

Figure 2: Sensor FOVs.

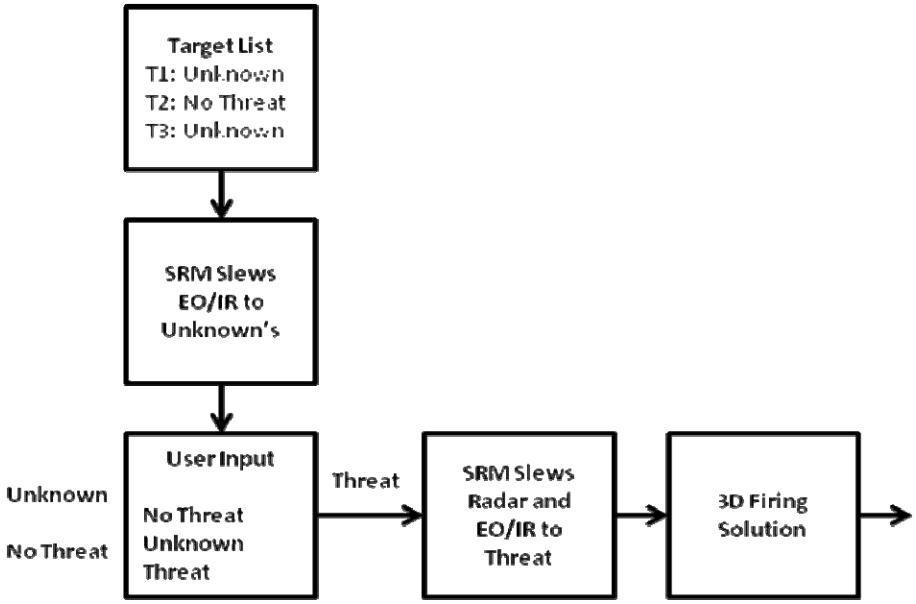

Figure 3: State flow diagram.

2. SYSTEM DESIGN

2.1 Sensor Resource Manager

The SRM is used to assign the region of focus for the EO/IR sensor suite based on system level tracks. The system level tracks are initially generated by the long range sensors such as the radar, and later updated by fusion of all of the sensor data. The SRM function provides the EO/IR sensors with a required pointing location (azimuth and elevation) based on cues from the WFOV sensors. Detections from any WFOV sensor will be used to initiate a target track via the MFA. These target tracks are initially identified as Unknown and added to the SRM's target list. The SRM uses this target list to determine when and where to point the NFOV sensors. The basic idea is that, given a list of tracks with ID of Unknown, the SRM will assign the NFOV sensors to these tracks thus providing the operator with the high resolution information necessary to declare them as either Threat or No Threat. The SRM will maintain a list of tracks that remain in the Unknown and Threat state, and continually cycle through this list. The SRM will provide the user with a queued set of target images for threat determination. Targets that are deemed No Threat will be removed from the target list presented to the user, and targets deemed as Unknown will remain on this target list. Once a target is declared a threat, the SRM will focus all sensors to this track. Note that other external data sources may provide knowledge that a given track is not a threat, and the SRM function will automatically take this into account since this information will be associated with the track via the MFA tracker. The SRM takes the sensor limitations into account. Specifically, the gimbal slew rates and settling time required to provide accurate data to the operator. The hardware we are sourcing is designed to mitigate these issues along with the obvious tradeoff of cost.

2.2 Operator Interface

The operator interface is being developed as a Graphical User Interface (GUI) with functions that allow the user to monitor target tracks, provide input to the MFA tracker for threat declaration and provide input to the SRM for the gimbaled sensors.

Track Monitoring: The GUI will have a real-time display of each sensor's output, which is switchable to allow monitoring of a single sensor feed. A split screen option will permit viewing of multiple sensors such that a WFOV sensor could be viewed alongside a NFOV sensor. The EO/IR data will be presented as a video feed from the video processor (discussed below). Each of these displays will be overlaid with track data, color coded to represent the current track ID. The track currently being viewed by the NFOV-EO/IR cameras as dictated by the SRM will flash on the

display. Note that the operator interface will be completely developed under this effort, so the discussion here is just one alternative.

User Input: The GUI will allow the user to designate the threat status (ID) of each track as the SRM selects the track for NFOV sensing. The SRM will actively cycle through its target list and generate a queue of target images for the user to evaluate. The user can control the queue by determining when to transition to the next target image. An additional capability will be for the user to manually supersede the SRM function entirely and cycle through tracks in the WFOV display.

2.3 Radar

We have two options for an RF long range detection solution. The first radar is an L-band system from Syracuse Research Corporation (SRCTec) [2]. SRCTec offers a Light Weight Search and Track Acquisition Radar (LSTAR) which provides 360 degrees of coverage. The LSTAR is electronically steerable and has scans rates of 360 degrees per second.

The second radar is an X-band system from Detection Monitoring Technologies (DMT) which offers a family of Area Intrusion Monitoring Systems (AIMS) which also provides 360 degree azimuth coverage [3]. The AIMS radar is pointed mechanically and has scan rates of up to 48 degrees per second.

Figure 4: SRCTec LSTAR system.

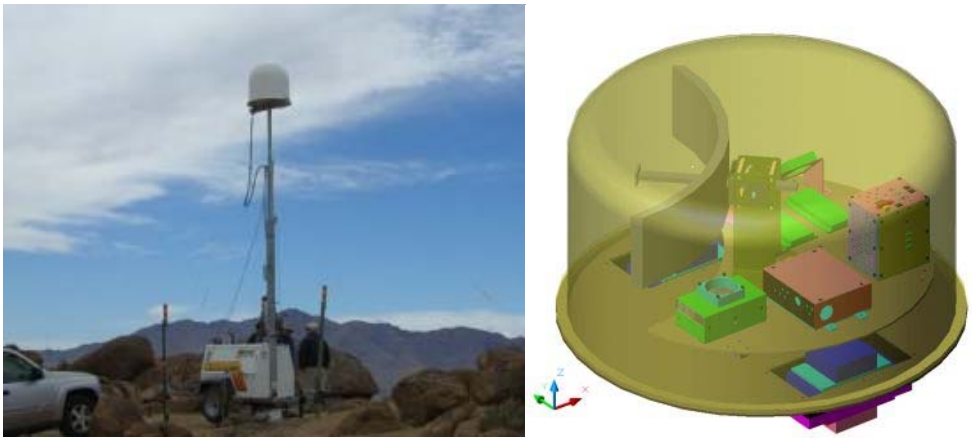

Figure 5: DMT AIMS radar.

We plan to tradeoff the sensors and determine which one offers the optimum solution. Figure 4 shows the LSTAR candidate radar and Figure 5 shows the AIMS radar.

2.4 EO/IR Sensors

The EO/IR sensors consist of commercial-off-the-shelf (COTS) video cameras and MWIR sensors. The current research efforts have used a combination of EO and MWIR cameras to characterize the detection capability at long range. Figure 6 shows the functional block diagram of each video processor. The raw video is streamed to an acquisition card which feeds the framed data to a lossless compression function for file storage. The file storage is primarily for data collection and post processing efforts. The data is also processed by an N-frame differencing and clustering algorithm operating in real time to generate measurements (detections) for use by both the MFA tracker and the operator. A compressed version of the video is also provided to the operator and later annotated with the track information. All of the data that leaves the video processing function is published on the local domain via OMG's Data Distribution Service (DDS) interface. As such, the system architecture is built on a real-time publish/subscribe communication system. The physical sensor layout for the EO/IR sensors is shown in Figure 7 to include the supporting components such as a power supply, cooling fan, server switch and computer rack. A final packaging design is still underway, but it is anticipated that some form of protective glass (fish eye) will be used to handle the effects rain, sand, etc.

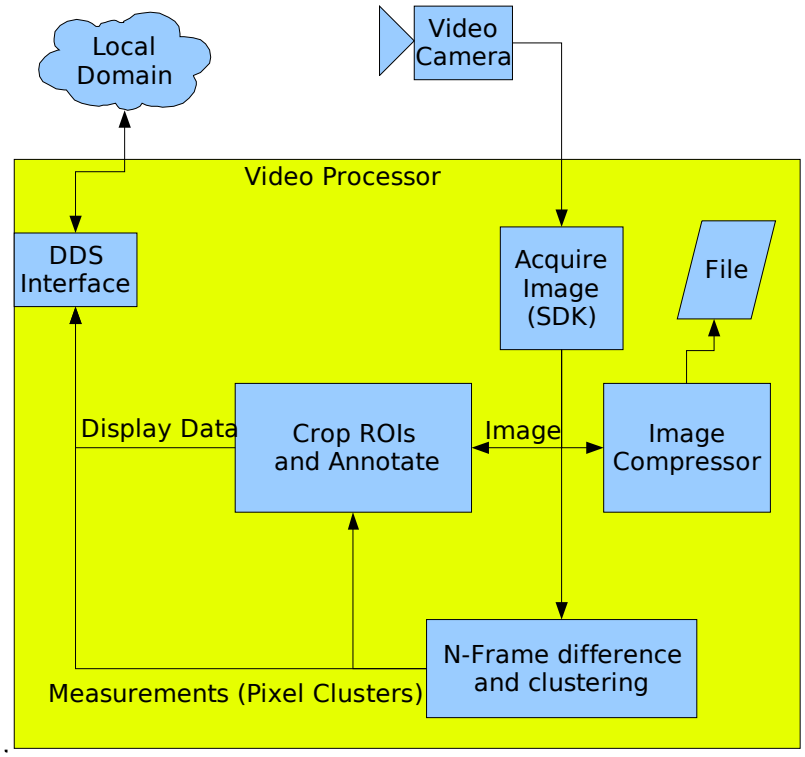

Figure 6: EO/IR video processor.

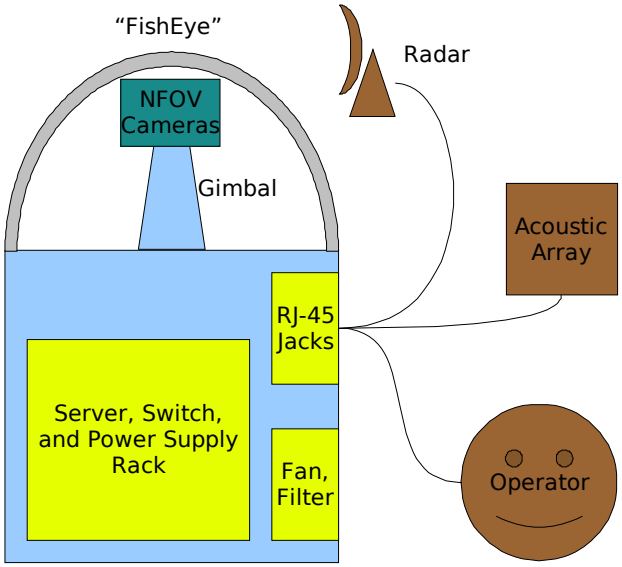

Figure 7: EO/IR physical layout.

2.5 Acoustic Array

The acoustic array (Figure 8) consists of twenty-four microphones used to collect audio data coupled with an EO/IR video camera to act as a reference frame for calibration. Collected data is currently processed using MATLAB software that first calibrates [4-6] the location of the microphones relative to the field-of-view of the camera and then performs beam forming [7].

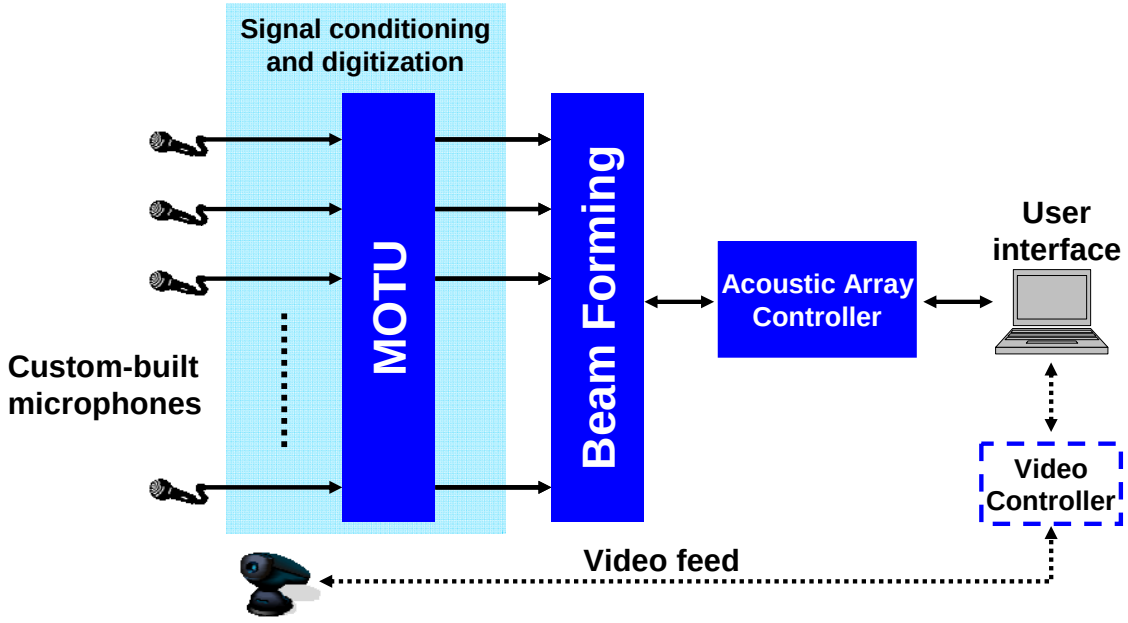

Figure 8: Acoustic array system concept.

The hardware used to construct the array consists of twenty-four custom-made microphones, a sound-recording board, and a desktop computer for hardware control. The microphone design includes a condenser microphone cartridge from

Panasonic and a MOSFET for pre-amplification. Each microphone is encased in a small metal box and is powered by a nine-volt battery. Ports are drilled in the metal casing for the microphone cartridge, a power switch, and a 20-foot output cable with a ¼" jack. The microphones interface to a sound recording board by Mark of the Unicorn (MOTU) that supports 24 I/O channels with 96 kHz of bandwidth per channel. The MOTU recording software stores each channel of data in a separate WAV file for off-line processing.

This microphone array is designed to be mobile so that it can be set up in different locations and in different configurations as dictated by the application. As a result, the beam forming software cannot rely on precision placement of the microphone phase centers. For this reason, calibration is required after each setup to determine the locations of the microphones relative to one other. This particular data collection employed only 16 microphones in a roughly linear configuration with 10-inch element spacing.

Calibration is accomplished by generating known reference signals that can be simultaneously observed by the video feed and the acoustic array. In this case, a series of acoustic chirps were produced over a range of positions within the video field-of-view. The recorded chirps are then pulse-compressed on each audio channel. The compressed chirps are detected, and their relative time delays across channels are then calculated. This information allows one to solve for relative microphone locations (*e.g.*, $\{(x_n,y_n)|n=1\ldots24\}$ in 2-D) of such sufficient fidelity for beam forming. Typically, the array centroid is then shifted to the coordinate origin.

After calibration, delay-and-sum beam forming is straightforward. To steer a beam in a given direction θ, the far-field time delay for the nth channel is

$$t_n = x_n \cos\theta + y_n \sin\theta ,$$

and thus the beam output is

$$s(t;\theta) = \sum_{n=1}^{24} s_n(t - t_n)$$

where $s_n(t)$ is the audio signal observed by the nth microphone. This process is repeated for every aspect angle θ within the desired field-of-view. Prior to detection processing or display in a waterfall plot (*i.e.*, direction-of-arrival versus time), one may integrate the signal energy over discrete time intervals for each beam.

3. RESULTS AND ANALYSIS

A series of data collections with Alpha 60 remote controlled aircraft provide preliminary results for the EO/IR cameras and the acoustic array. The goal was to characterize the detection ranges of these sensors and verify that sufficient measurements could be generated to track the target. Truth data on board the aircraft was used to validate performance, but due to the nature of this research, exact detection ranges cannot be disclosed.

3.1 EO/IR Detection and Tracking

The NFOV-MWIR and the MFOV-EO cameras were used to test the detection and tracking ability of the each sensor mode. Frame-to-frame change detection is employed to generate raw measurements at each pixel in the camera FOV. Morphological operations then cluster pixels and form blob detections. The blob centroids provide measurements to the MFA tracker for kinematic tracking. Figure 9 shows snapshots of the MWIR video at both long and short ranges. The short range image on the right clearly shows the profile of the SUAV. The image on the left indicates the ability of the image processing to generate detections of the SUAV at long range.

Similarly, Figure 10 shows the ability of the EO camera to detect the SUAV. The short range image on the right provides a close look at the SUAV, which at first look closely resembles a small manned aircraft. The image on the left depicts the relatively significant challenge associated with detecting the target with the EO camera at long range. While the target is not obvious to the naked eye, the change detection algorithm is capable of segmenting the target in the video. These examples do however raise a key issue associated with reliance on either of these sensor modes to independently provide a high probability of detection with a low rate of false alarms. Other samples collected during the experiments generated false detections from birds and commercial aircraft. Figure 11 highlights this point by showing

that a commercial jet in the distance presents a return very similar to the SUAV that is considerably closer to the sensor. Since the video cameras are only capable of providing azimuth and elevation data, the need for range information is paramount to eliminate false alarms of this type. This has motivated the project to include a radar system, and future results will demonstrate the capability of combining the radar with the current EO/IR suite. Alternatives to the radar include laser range finders, which may be pursued in a future effort.

It should be noted that the detections provided by the EO/IR imagery were readily tracked with the MFA tracker. The truth data validated that successful tracks were generated and maintained. Other clutter objects generated short lived tracks, many of which were pruned by the MFA's track maintenance logic. However, in the cases of persistent detections such as those generated by commercial aircraft, the need to incorporate additional sensor data for the purpose of ID was clear. The next section demonstrates the fundamental capability of the acoustic array to generate measurement information for the SUAVs, with the ultimate goal of providing ID information to mitigate false alarms.

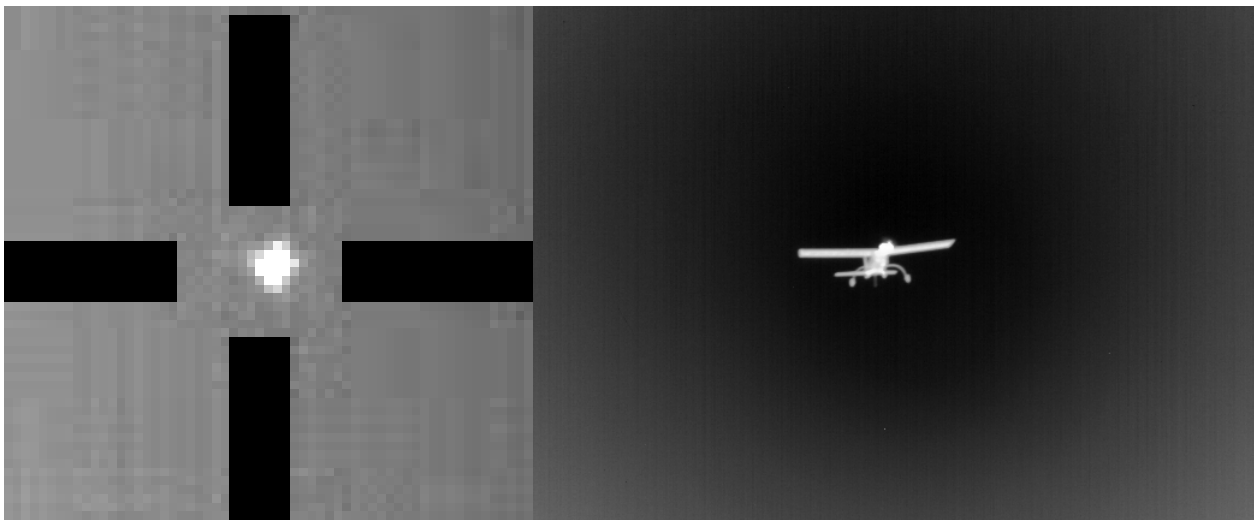

Figure 9: MWIR camera data. (left) target detected at long range with crosshairs to indicate automated detections (right) this image clearly shows the UAVs IR signature

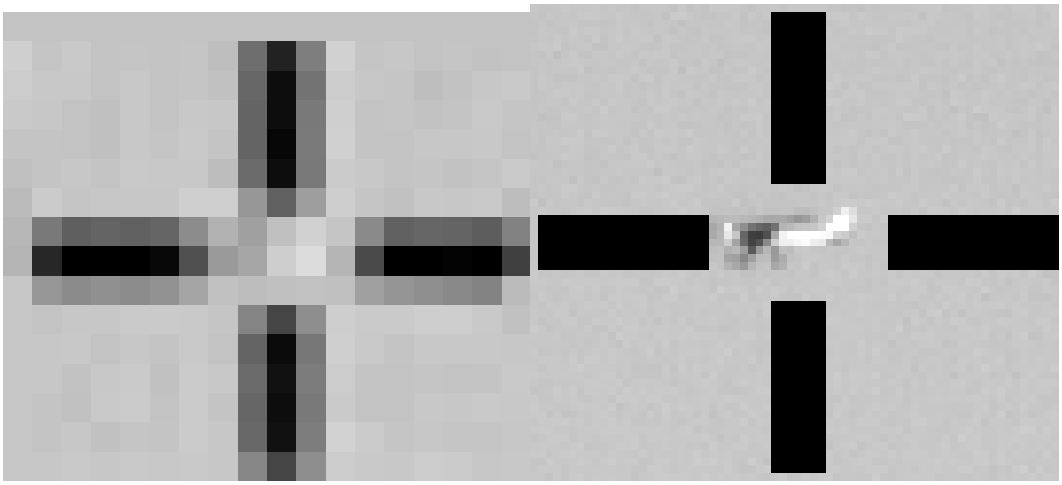

Figure 10: EO camera data. (left) target detected at long range with crosshairs to indicate automated detections (right) this image clearly shows the UAVs EO signature

Figure 11: Illustration of potential target clutter from commercial aircraft

3.2 Acoustic Beam Forming

For the collected data, the acoustic array calibration was performed as described in Section 2.5, and the estimated set of microphone locations was used in all of the following scenarios. Some preconditioning of the recorded signals was required prior to beam forming in order to suppress interference in the near field and ambient noise clutter. Specifically, a band bass filter from 450 Hz to 3 kHz was applied to each channel of data. The 450 Hz cutoff is low enough to remove most of the ambient clutter noise and voices in the near field. A 3 kHz cutoff was selected to allow the recorded data to be down-sampled from 44.1 kHz to roughly 8 kHz, thus greatly increasing the processing throughput without loss of significant signal energy.

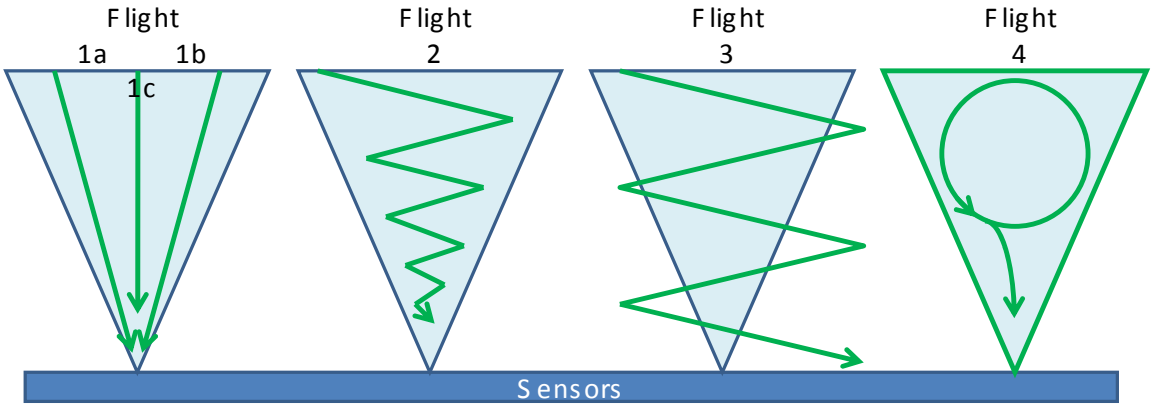

Figure 12: Recorded flight trajectories.

Beam forming was applied to four of the recorded flight trajectories shown in Figure 12. In each case, a waterfall plot was generated by integrating the beam-formed signal energy over 33 ms intervals. The results are shown in Figure 13 for trajectories 1c, 2, 3, and 4. In each, one can observe the presence of grating lobes due to the aliasing of high frequency signal content, particularly at later times in a flight when the aircraft is closer to the array and at higher SNR. With 10-inch element spacing, acoustic signals with wavelengths shorter than 20-inches or frequencies larger than roughly 700 Hz will experience spatial aliasing. Several low-level interference sources appear consistently at aspects of -40°, -30°, -20°, and 10° for example. There are also short-time events in the near field, such as roughly 3 seconds into flight 1c. In general, the beam formed results and waterfall plots are consistent with executed flight trajectories.

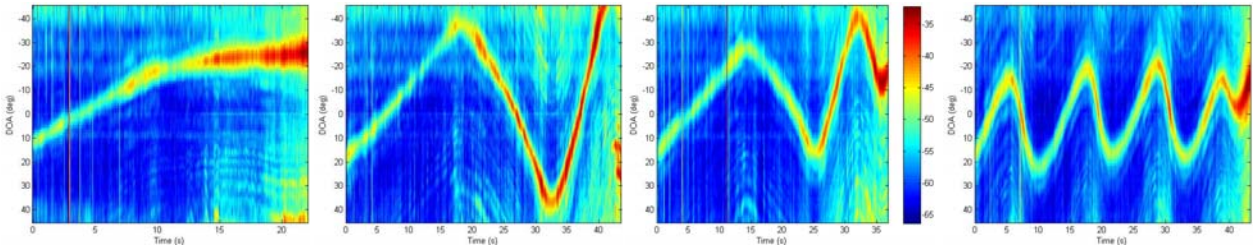

Figure 13: Waterfall plots resulting from beam forming flights 1c, 2, 3, and 4.

3.3 Summary

The preliminary results for EO/IR cameras and acoustic array are very promising. Future efforts include the incorporation of the radar sensor and adding a classification capability based on the acoustic data to mitigate false alarms. Additional development of the sensor resource management function will be completed along with integration of the sensor suite for real-time operation.

REFERENCES

[1] Poore, Aubrey B., Multidimensional Assignment Formulation of Data Association Problems Arising from Multitarget Tracking and Multisensor Data Fusion. *Computational Optimization and Applications.* 1994, Vol. 3.

[2] Lightweight Surveillance and Track Acquistiiont Radar, http://www.srctecinc.com/products.html, Feb 2008.

[3] Area Intrusion Monitoring Systems, http://www.ctech.com.tr/defense2.htm, Feb 2008.

[4] Ash, J. N. & Moses, R. L. Acoustic time delay estimation and sensor network self-localization: Experimental results. Journal of the Acoustical Society of America, 2005, 118, 841-850.

[5] Khong, A. W. H. & Brookes, M. The Effect of Calibration Errors on Source Localization with Microphone Arrays. Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, 2007.

[6] Ianniello, J. Time delay estimation via cross-correlation in the presence of large estimation errors. Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on, 1982, 30, 998-1003.

[7] Chen, J. C.; Yao, K. & Hudson, R. E. Source localization and beamforming. IEEE Signal Processing Magazine, 2002, 19, 30-39.