

Conf-950793--2
SAND94-3151C

Reliable Motion Detection of Small Targets in Video with Low Signal-to-Clutter Ratios

Scott A. Nichols and R. Brian Naylor

Sandia National Laboratories, Security Technology Department
P. O. Box 5800, Albuquerque, New Mexico 87185-0780
sanicho@sandia.gov (505) 844-1607
rbnaylo@sandia.gov (505) 844-8523

RECEIVED

JUN 19 1995

OSTI

Abstract

Studies show that vigilance decreases rapidly after several minutes when human operators are required to search live video for infrequent intrusion detections. Therefore, there is a need for systems which can automatically detect targets in live video and reserve the operator's attention for assessment only. Thus far, automated systems have not simultaneously provided adequate detection sensitivity, false alarm suppression, and ease of setup when used in external, unconstrained environments. This unsatisfactory performance can be exacerbated by poor video imagery with low contrast, high noise, dynamic clutter, image misregistration, and/or the presence of small, slow, or erratically moving targets.

This paper describes a highly adaptive video motion detection and tracking algorithm which has been developed as part of Sandia's Advanced Exterior Sensor (AES) program. The AES is a wide-area detection and assessment system for use in unconstrained exterior security applications. The AES detection and tracking algorithm provides good performance under stressing data and environmental conditions. Features of the algorithm include: reliable detection with negligible false alarm rate of variable velocity targets having low signal-to-clutter ratios; reliable tracking of targets that exhibit motion that is non-inertial, i.e., varies in direction and velocity; automatic adaptation to both infrared and visible imagery with variable quality; and suppression of false alarms caused by sensor flaws and/or cutouts.

NOTE: This work was supported by the Defense Nuclear Agency and the Department of Energy under contract DE-AC04-94AL85000.

Keywords: video motion detection, clutter suppression, adaptation, false alarm suppression, detection and tracking, human and vehicle tracking

1. Introduction

Passive sensors capable of wide-area, stand-off intrusion detection are gaining increased importance in applications ranging from upgraded fixed perimeter security to rapid-deployment force protection on peace-keeping missions. Using video motion detection (VMD) technologies increases the system's capabilities by providing not only intrusion detection, but also providing the operator immediate visual alarm assessment.

VMD systems have been applied to visible imaging and thermal imaging video sensors around perimeters and sensor fields. However, recent evaluations show nuisance alarms are still excessively high (Vigil 1992), and the sensors and processors are not designed into an integrated unit. Previous attempts to integrate visible and thermal imaging devices with advanced processing have produced a significant technology base (Pritchard 1990; Arlowe 1990) but a challenging problem remains: how to provide robust detection using sophisticated processing in an affordable, reliable package.

Sandia National Laboratories has been tasked by the Defense Nuclear Agency (DNA) to research requirements and technology for, and develop a prototype of, an Advanced Exterior Sensor (AES) for ground-based, stand-off

4

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

intrusion detection and assessment. The AES is a 360-degree scanning, multispectral intrusion detection sensor that simultaneously acquires and processes panoramic images from the infrared, visible, and radar spectrum (Pritchard 1995).

The AES is designed to detect humans and vehicles at ranges from 50–1500 meters and moving as slowly as 0.25 meters/second (0.1 meters/second desired). The current human and vehicle detection requirements for the AES are summarized in Table 1.

Table 1. AES Detection Range Requirements

Target	Conditions	Range (required)	Range (desired)
Upright human walk/run 0.6x1.65 m 1.0 m ²	Clear, good visibility Light rain, humid	500 m 350 m	750 m 525 m
Crawling human head-on 0.5x0.3 m 0.15 m ²	Clear, good visibility Light rain, humid	250 m 200 m	375 m 300 m
Truck/van 1.5x1.5 m 2.3 m ²	Clear, good visibility	1000 m	1500 m

Other AES system goals include: reduction of nuisance alarms by distinguishing humans and vehicles from other moving objects (birds, tree limbs, debris), classification of at least 90 percent of potential threats versus non-threats, and continuous detection while assessing alarms from other locations.

To accomplish as many of the AES detection goals as possible, a robust detection and tracking algorithm is required that can reduce false alarms caused by background and sensor noise; suppress high contrast edges and boundaries in the imagery; tolerate slight sensor misregistrations; reliably detect small, slow moving, or erratic targets; automatically adapt to various qualities of imagery and environmental conditions; and reliably detect targets with low SCR and high electrical, spatial, and temporal noise conditions.

2. Technical Approach

Development of the detection and tracking algorithm for the AES system began in February 1992. Several literature searches were undertaken to identify the latest published VMD technologies in these subject areas: image registration, clutter suppression, threshold setting and adaptation, target segmentation, data association, target tracking, and sensor fusion (Pritchard and Nichols 1992).

Sandia's goal was not to invent new technologies to increase detection performance. Rather, it was to identify promising technologies which were nearly mature, or technologies which could be integrated into an operational system with a minimum of additional research and development. Technologies that may have promised good performance but were very new, untested, too computationally intensive, or constrained in application scope were not pursued.

The review concluded that, in general, 3-D spatio-temporal matched filters with track-before-detect algorithms or coherent integration algorithms would provide the best detection performance. However, they were deemed unsuitable for AES applications because statistical information about the targets and background was incomplete or unknown, and because they had excessive computational requirements in the temporal image domain. However, the



DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

authors concluded that judicious enhancement of more typical non-integrating clutter suppression, segmentation, clustering, tracking, and sensor data fusion technologies could still satisfy the AES detection requirements. A high-level block diagram of the resultant processing architecture is shown in Figure 1.

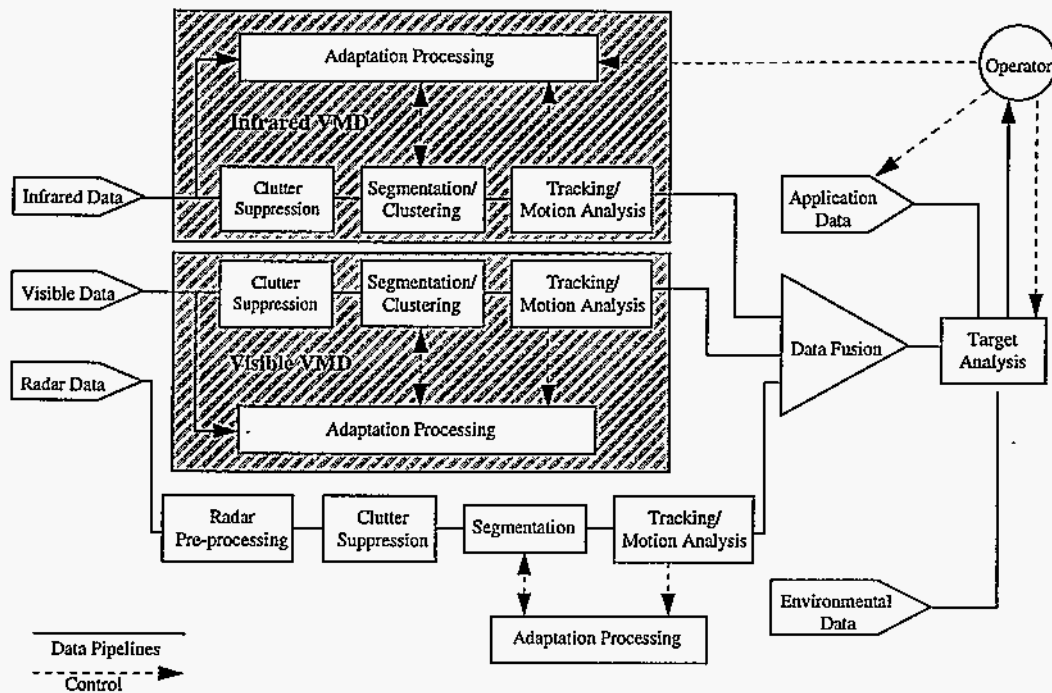


Figure 1. AES Detection and Tracking System Block Diagram

The two shaded blocks in the figure illustrate the principal modules of the infrared and visible AES detection and tracking algorithm pipelines—the topics of this paper. The radar processing pipeline looks much like the infrared and visible pipelines and follows similar processing steps; however, the resolution and nature of the AES radar data calls for a somewhat different processing approach. For this reason its details are outside the scope of this paper. The other blocks, such as data fusion and target analysis, have not been implemented at the time of this report.

Over the course of the advanced algorithm development, four areas have been identified as the keys that provided the best tradeoff between performance and increased memory and computation requirements. These areas are:

1. Advanced clutter suppression
2. Multilayered thresholding and feedback adaptation
3. Purposeful motion analysis
4. Target prediction and expectation

Advanced Clutter Suppression

The AES detection and tracking algorithm uses a unique approach to remove the background clutter and noise which can result in spurious detections not caused by coherent target motion. This method is based on a computationally efficient approximate solution to the spatio-temporal constraint equation (STCE) commonly used to compute image flow fields. A detailed derivation and discussion of this equation is provided in previous literature, hence, only the final equation is presented here (Horn and Schunk 1980).

$$\left(\frac{\partial I}{\partial t} + V_x \frac{\partial I}{\partial x} + V_y \frac{\partial I}{\partial y} \right)^2 = (Error)^2 \quad (1)$$

In its above form, the STCE equation is ill-posed and cannot be solved without additional constraints. It is also difficult to solve with precision. Instead, Sandia's approach is to compute one iteration of a gradient descent to obtain a very coarse solution for the velocities of the image flow field. But rather than using these velocity values (V_x , V_y) directly to indicate moving targets, Sandia's approach is to use the residual squared error energy of the equation in a small, localized region to indicate motion (Wayman 1993a, 1993b). This is much more computationally efficient because the velocity values are not solved for with precision. With this method, motion is indicated when the solution to the equation cannot be found without yielding a large residual error (non-zero) value. This typically occurs when the equation is solved across the boundary of a moving object. This moving boundary violates the linear constant velocity constraint of the equation, hence allowing the STCE to be used as a moving target indicator (MTI).

While equation (1) may seem complicated for use as an MTI, closer scrutiny will reveal that if the velocity terms are constrained to always be zero, the equation degenerates to simple differencing. The velocity terms in the STCE are beneficial because they add some degree of robustness against small (1-2) pixel image misregistrations. This is true because a global image shift or translation can occur without violating the equation's constraint of constant linear velocity.

Multilayered Local Thresholding and Feedback Adaptation

Localized adaptive thresholding has played a key role in the detection performance obtained with the algorithm on test imagery. Experience has shown that a single fixed global threshold for the entire image is ineffectual; a single adaptive threshold for the entire image is better, but it still suffers from severe limitations. This lack of effectiveness is because clutter characteristics can vary widely over the scene, and the statistic for updating this threshold often cannot not be found or does not apply uniformly. If one pays the performance penalty and resorts to localized thresholds, i.e., thresholds which represent areas adapted to much smaller regions of the image, then performance improves dramatically. However, the problem still exists of how to adjust these thresholds so that maximum detection sensitivity is achieved while still maintaining a constant false alarm rate (CFAR). In theory, the CFAR processor can achieve good performance using a threshold proportional to the standard deviation of the clutter, i.e., $T = K\sigma_{clutter}$ for each local region. However, in almost all cases, empirical data suggests that the proportionality constant K varies in both space and time over the region, and thresholds set with this rule do not provide maximum detection sensitivity on the average. This is especially true in an exterior, noisy, unconstrained environment, where it is impossible to accurately characterize the clutter and targets in the sensor data. Sandia's detection and tracking technique mitigates this deficiency by using a combination of localized thresholds applied at two stages of the processing pipeline. In effect, the subsequent thresholds in the pipeline compensate for the errors induced by imprecise settings of the initial thresholds. This is akin to admitting imprecise knowledge of the true clutter characteristics in the imagery; therefore we strive to obtain the best model by successively approximating them throughout the pipeline.

The mechanics of calculating these thresholds is relatively simple, however, finding the gains and adaptation parameters needed for the adaptive feedback control of the thresholds can be difficult. The diagram shown in Figure 2 illustrates the diversity of feedback paths used to continuously adapt the algorithm's detection thresholds so that processing overload does not occur, false alarms are suppressed, and desired detection outcomes are promoted. Principally, the diagram illustrates the fact that most of the threshold adaptation depends on the detection activity in the processing pipeline. Little information is required from static scene analysis or external inputs.

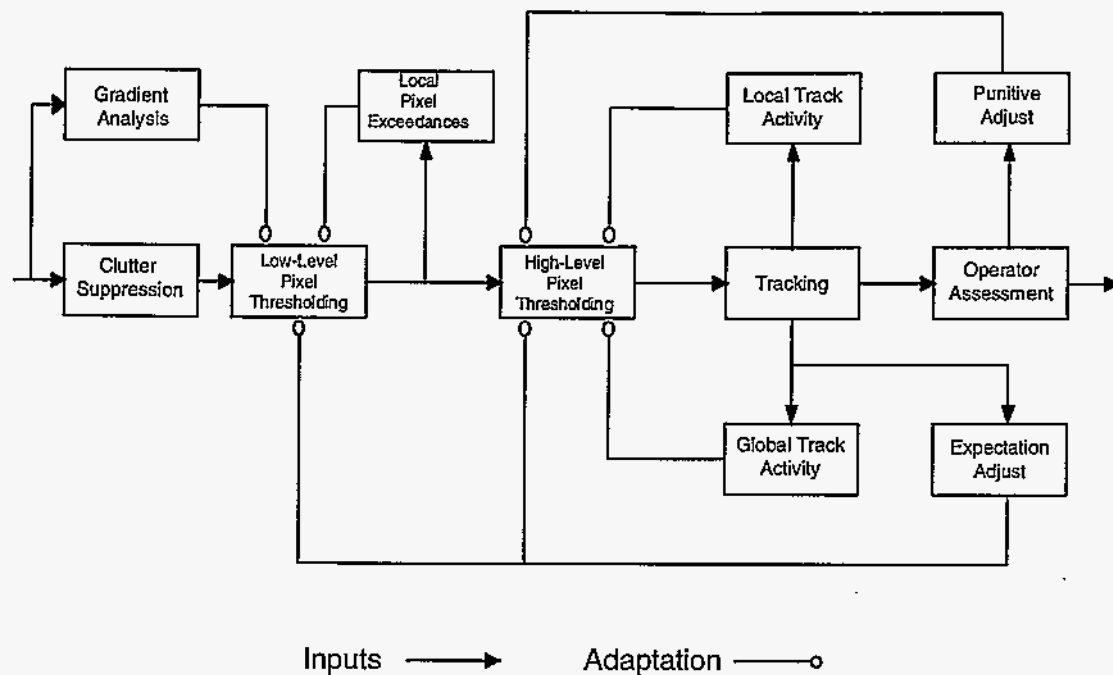


Figure 2. AES Threshold Adaptation Diagram

The thresholding in the low-level segmentation block of Figure 2 is based upon a gradient analysis of the imagery, the level of activity in the localized region, and an expectation adjustment. Any pixel value that does not exceed the threshold is set to zero prior to high-level segmentation processing. The high-level segmentation threshold is based upon global track activity, local track activity, an operator feedback adjustment, and the expectation adjustment. Graphically, the effect of the two thresholds can be seen in Figure 3. Only those pixels whose magnitude exceed both of the individually calculated thresholds will be passed on to the clustering and tracking processes.

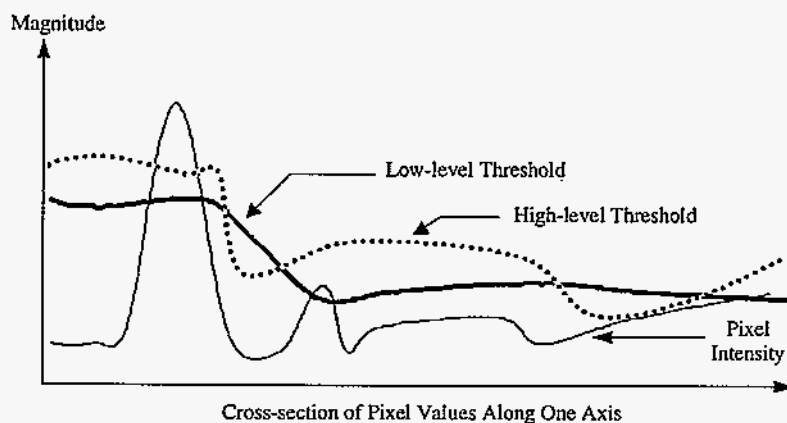


Figure 3. Segmentation Thresholds

Purposeful Motion Analysis

The primary mechanism for purposeful motion analysis is the Motion/Prediction Error Envelope Metric (MPEEM). This performance envelope is a function of total "purposeful" distance displaced since track initialization and an accumulated error measure of the object's predicted motion and feature consistency.

When viewed temporally, an object which is moving "purposefully" away from the original detection point would likely generate an MPEEM envelope which resembles the example in Figure 4.

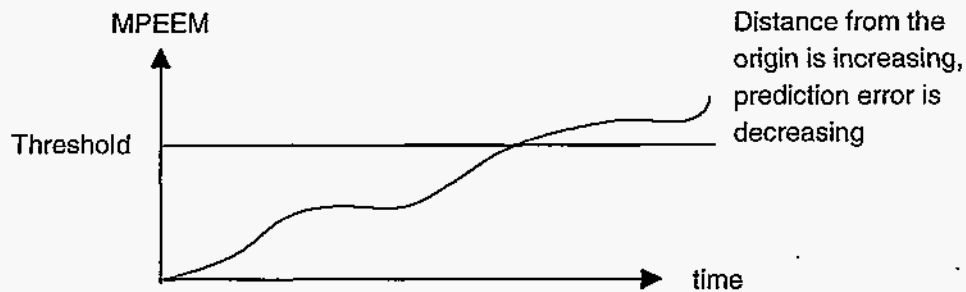


Figure 4. MPEEM as a function of time for a typical target

Operating in a stand-alone mode, the detection and tracking algorithm would signal an alarm when the MPEEM value exceeded some threshold based on the desired level of detection sensitivity or false alarm rejection. This threshold may be set a priori by the user, based upon past system performance, or automatically calculated during run-time. Other factors such as the size and brightness of a target, the length of time a target has been tracked, or the particulars of the target's physical parameters can be used to bias the MPEEM value to further enhance performance

Target Prediction and Expectation

Detection sensitivity for maintaining marginal target tracks can be increased without globally increasing false alarms by adjusting thresholds within some region where the target's next appearance or location is predicted. Figure 5 illustrates this concept. The area indicated by the cross-hatched region would have its detection thresholds lowered according to some bowl-shaped function with the maximum change or sensitivity occurring at the center of the region. This procedure makes intuitive sense and is consistent with human perception insofar as humans are more likely to perceive an object in a region in which they have an expectation of seeing that object based on past experience.

Although simple and effective, this technique must be used with caution. Empirical results have shown that excessive threshold reduction in a region of expectation can cause a rapid degradation in performance due to detections becoming a self-fulfilling prophecy.

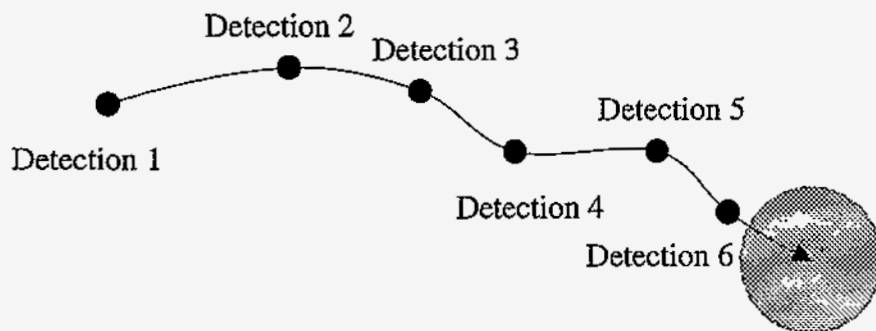


Figure 5. Threshold Reduction in Predicted Location

3. Performance Results

Detection Performance

The following section discusses several of the performance goals that were considered key issues for the AES detection and tracking algorithm.

Effect of Signal-to-Clutter Ratio

The detection and tracking algorithm which is under development for the AES system typically provides robust detection for targets in noisy or cluttered video with very low SCRs. The two-dimensional SCR used here is defined as the ratio of target signal power (T) to background signal power (B). It is computed as follows:

$$SCR = \frac{\sigma_T^2 + \mu_T^2}{\sigma_B^2 + \mu_B^2} \quad (2)$$

where σ is the standard deviation and μ is the mean. The statistics of the target and background are computed over local regions. The dimensions of the target region are one-third the dimensions of the background region and are centered within it (see Figure 6).

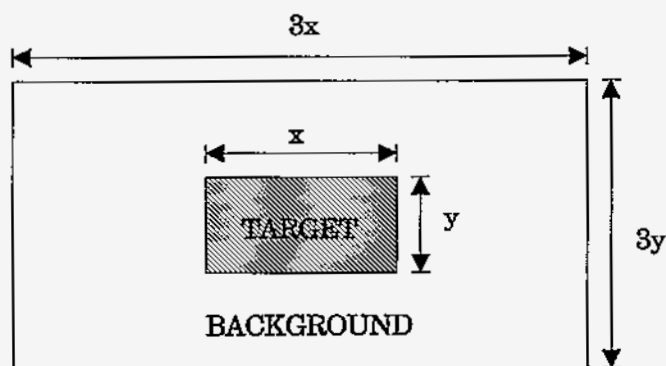


Figure 6. SCR Calculation Dimensions

Testing indicates that the AES detection and tracking algorithm reliably detects and tracks small targets (~ 20 pixels on target) with SCRs as low as 2 dB. Experimentation also indicates a rapid degradation in detection reliability when SCRs approach 1.5 dB. Studies, however, indicate that a typical human observer requires at least an SCR of 8.5 dB (2.7:1) for reliable assessment of non-moving targets in static imagery. This means that in many cases, the detection and tracking algorithm can detect and present targets to the operator that cannot be reliably observed unless multiple frames are provided. Even then, target recognition is improbable. Because of this, the data must be visually enhanced or presented to the operator in an innovative way to ensure efficient assessment.

If detection is needed at SCRs of 0 dB or less, more exotic techniques are likely to be required. Unfortunately, the computing requirements for 3-D approaches such as coherent integration and track-before-detect are prohibitive for AES applications due to the system's low cost and size requirements, at least at the time of this report.

System Adaptivity

The original motivation for the multilevel adaptivity of the detection and tracking algorithm was to achieve maximum detection sensitivity in every local region in the image. Since the parameters for this control system could not be set a priori for an unknown operating environment, it was necessary to achieve this automatically. Feedback was used to allow the system to strive to maintain certain predefined operating conditions which were general enough that they would basically apply to all deployment scenarios. This naturally resulted in a system that automatically adapts to nearly all reasonable operating conditions or image data quality without any operator setup or intervention.

Results from test data indicate that from initial compiled default parameters, the detection algorithm will fully adapt to a new deployment scenario regardless of sensor type (infrared or visible) or scene clutter statistics in approximately 120 frames or two minutes at the one frame per second AES frame rate. This provides the possibility of remote deployment applications where an operator is not available for setup and configuration of detection parameters.

Motion Types and Considerations

Many different types of motion have been considered during the development of this algorithm; these range from simple linear trajectories to complex near-random paths. The following is a list of motion types that have been investigated and considered.

Linear Motion: Consistent linear motion is tracked well by the detection and tracking algorithm. Currently, the algorithm must observe the target for 3 sequential frames to establish a good track and 5 to 6 frames to gain enough confidence in the track to signal an alarm. A maximum velocity is currently set within the algorithm to establish the upper velocity boundary. The higher the allowable velocity, the higher the likelihood that spurious detections can satisfy the requirements for purposeful motion.

Curving Motion: Smooth curving motion is tracked well by the detection and tracking algorithm under most circumstances. The MPEEM statistic uses independent horizontal and vertical predictions. Thus, at least one of the predictors usually returns a low prediction error. If the curve is very large, the system interprets it as a set of piecewise linear trajectories, which are tracked very well. If the curve radius is small, then it may not be considered purposeful unless it becomes somewhat helical. Helical motion eventually displaces enough total distance to be considered purposeful.

Intermittent Detection: If the motion sensor cannot detect a target for 3 sequential frames, the algorithm cannot establish a track on the target. In situations such as these it is difficult to mitigate the problem without accepting a higher false alarm rate or using two or more sensor phenomenologies. If the detections become intermittent after a track has been established, whether the algorithm can alarm on the track or even continue tracking the target will be based on the previous intermittency of the track.

Meandering: Meandering-type motion, such as a cow or deer wandering in the field of regard, would cause motion which is really a variation of curving motion and is detected as such.

Random/Cyclic Motion: Random or cyclic motion, such as a swaying tree limb, can be a problem for the algorithm. In theory, the MPEEM metric will filter out this type of spurious movement because very little total displacement

occurs and the randomness will not allow an accurate prediction of the motion. Even though the system does not allow the motion itself to cause an alarm, it does continue to track the motion. A problem can occur when a random noise blip occurs within the radius of regard of the cyclic motion. This spurious point can assume the identity of the existing motion and then add its displacement to the existing measure of total displaced distance. In some cases, this has been shown to be significant enough to cause false alarms. Another problem has been seen in windy imagery when the detection and tracking algorithm is surveying a tree line. Swaying limbs on one end of the tree line can initiate tracks which become associated with swaying limbs on the tops of the trees. This association can keep recurring all along the tree line, eventually displacing enough distance to cause an alarm. At present, this is partially mitigated by adaptive thresholding, but can only be totally prevented by masking out the troublesome area.

Zigzag Motion: Zigzag motion is quite similar to random motion except that a significant displacement of the target occurs in a particular direction. The algorithm has very few problems tracking this erratic motion because the horizontal motion is predicted independently of the vertical motion. On a zigzag approach, the intruder may cause one predictor to operate with large errors, but the other should track fairly well on the intruder's position.

Hop-and-Stop Motion: Hop-and-stop-type motion is associated with nuisance alarms, such as those caused by jackrabbits, where the target makes a sudden movement and then pauses for several seconds. The algorithm cannot currently track this type of motion. Objects that show a small burst of movement and then do not move for a considerable period of time are assumed by the system to either be noise artifacts or non-purposeful motion such as that caused by swaying tree limbs. In either case, the motion is filtered out. The only hope of mitigating this problem is to attempt to discern a purposeful trend in the erratic motion with knowledge-based target analysis or to use primitive object recognition to identify the target causing the motion as a human intruder.

Computational Requirements

In its current form, one of the detection and tracking algorithm pipelines shown in Figure 1 requires approximately 12 million floating point operations to analyze a 512 x 480 video image. Of this amount, 75–80 percent of the processing requirements are due to the clutter suppression process. The clutter suppression algorithm also does not map itself efficiently onto general purpose microprocessors or digital signal processors. The clutter suppression algorithm's efficiency has been empirically determined for several hardware platforms at about 30–40 percent on average. For this reason, an application-specific integrated circuit (ASIC) is highly recommended for its implementation in a production system.

Each of the highlighted pipelines also requires about 2.5 megabytes (MB) of memory for a 512 x 480 image implementation. Of the 2.5 MB required, 0.5 MB is used for image storage, 1.5 MB is used by the adaptation and thresholding algorithms performing the image segmentation, and the remainder is used for program memory and other variable storage.

4. Future Directions

Currently, effort is underway to port the algorithm onto a dedicated hardware platform running a real-time operating system for implementation in the AES system. Work is also continuing on the unfinished portions of the system. Primarily, this effort entails completing the data fusion and target analysis blocks and integrating the radar processing pipeline as shown in Figure 1. An AES prototype is scheduled to be completed in FY96.

Research is also ongoing to further enhance the algorithm. The foremost effort is the mitigation of some of the problematic motion considerations discussed in this paper. In the final AES system, most of these deficiencies should be overcome by the use of the radar sensor, the fusion of all three sensor methodologies, and the inclusion of higher-level target analysis. The authors believe that the performance of the detection and tracking algorithm itself can be enhanced through the use of more sophisticated motion prediction, enhanced data association capabilities, and inclusion of a more robust registration process.

5. Acknowledgment

This research was supported by the Defense Nuclear Agency under the direction of the tri-services Physical Security Equipment Action Group and the Department of Energy Office of Safeguards and Security.

6. References

- H. D. Arlowe and D. Coleman, 1990, The Mobile Intrusion Detection and Assessment System (MIDAS), Sandia National Laboratories report SAND89-3016C. *IEEE Carnahan Conference on Security Technology*, 1990.
- D. A. Pritchard, 1990, An Infrared Imaging Area Sensor for Tactical and Physical Security Applications, *IRIS Specialty Group on Passive Sensors*. ERIM, Ann Arbor, MI, March 1990.
- D. A. Pritchard, 1995, System Overview and Applications of a Panoramic Imaging Perimeter Sensor, Sandia National Laboratories report SAND95-1055C, in *Proceedings of the 11th American Defense Preparedness Association Symposium on Security Technology*, June 1995.
- J. T. Vigil, 1992, An Evaluation of Exterior Video Motion Detection Systems, Volume 1: Intrusion Detection Tests, Sandia National Laboratories report SAND92-0108, 1992.
- D. A. Pritchard and S. A. Nichols, 1992, Advanced Exterior Sensor (AES) Phase One Final Report: Feasibility Analysis, Design Concept, and Cost Estimate, August 1990.
- B. K. P. Horn and B. G. Schunck, 1980, Determining Optical Flow, *Artificial Intelligence*, vol. 17, pp. 185-203, 1980.
- J. L. Wayman, et al., 1993a, Target Detection Using Image Motion Error Measure, in *Proceedings of SPIE Conference 1955 Signal Processing, Sensor Fusion, and Target Recognition*, Orlando, April 1993.
- J. L. Wayman, et al., 1993b, A Rapid Method for Target Motion Detection Using the Spatio-Temporal Constraint Equation, in *Proceedings of SPIE Conference 1955 Signal Processing, Sensor Fusion, and Target Recognition*, Orlando, April 1993.