Personnel detection using ground sensors

Raju Damarla and David Ufford
US Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783
rdamarla@arl.army.mil

Abstract: *In this paper we present a multi-modal multi-sensor fusion algorithm for the detection of personnel. The unattended ground sensors employed consist of acoustic, seismic, passive infrared (PIR) and video camera. The individual sensor data is processed and the probabilities of detection of a person are estimated. Then, the individual probability estimates are fused to estimate the overall probability of detection of a person. The confidence levels of each sensor modality are estimated based on a large set of data. The performance of the algorithm is tested on data collected in an unoccupied basement of a building with single and multiple people present.*

Key words: personnel detection, sensor fusion, acoustic detection, passive IR, change in scene, seismic detection.

1. Introduction: In urban operations or in concealed terrains such as caves, tunnels and sewers it is imperative to detect the presence of personnel prior to entering into the facilities. Personnel detection can be accomplished using several sensors of multiple modalities, namely, acoustic, seismic, passive infrared (PIR), magnetic, electrostatic, RF, chemical, video and thermal imaging. Each sensor modality has a unique role to play in the process, and together would result in providing robust intelligence, surveillance and reconnaissance (ISR) with low false alarms and high degrees of confidence. The description of these sensors is given below:

- **Acoustic** sensors used are the piezo electric microphones, which can be used to detect speech, sounds generated by machinery, etc.
- **Seismic** sensors are 3-axis sensors that can detect the vibrations in the ground. They are used to detect footfalls, vibrations caused by machines being operated, etc. Accelerometers can also detect vibrations in pipes that are produced by the flow of water.
- **RF** detectors can detect any RF activity such as the use of cell phones.
- **Magnetic (B-field)** sensors can be used to detect ferromagnetic materials carried by people, e.g., keys, firearms, and knives. These sensors may also detect the usage of computer monitors.
- **Electrostatic (E-field)** sensors can be used to detect the built-up electric charge on personnel. Together with magnetic sensor they can also detect electrical activity in the vicinity such as the usage of computer keyboards.
- **Chemical** sensors can be used to detect the presence of different kinds of chemicals in the atmosphere such as pheromones and household chemical vapors.

Unattended Ground, Sea, and Air Sensor Technologies and Applications IX
edited by Edward M. Carapezza, Proc. of SPIE Vol. 6562, 656205, (2007)
0277-786X/07/$18 · doi: 10.1117/12.723212

- **Passive infrared** devices are very inexpensive sensors that detect the nearby presence of a warm body, e.g., a human, within a cone shaped field of view.
- **Visible imagers** can capture color or grayscale video for human gait detection and object recognition.
- **Infrared imagers** can detect and localize hot bodies and warm surfaces, including the vents in tunnels. They can also provide thermal profiling of buildings, where warmer rooms are indicative of current or recent human inhabitation.

The goal is to detect personnel using a subset of sensors or all sensors. In this paper, we use acoustic, seismic, PIR and video sensors for detection of personnel. Each sensor detects personnel using algorithms specific to its modality, and their information is fused to give robust personnel detection capability with high degree of confidence with few false alarms compared to the personnel detection using any single modality.

Section 2 presents personnel detection algorithms used for each sensor modality. Section 3 describes the fusion algorithm to combine the information from all the sensors. We also present the results obtained using real data. Section 4 presents the conclusions and future directions.

2. Personnel Detection Algorithms for Acoustic, Seismic, PIR and Video

In this section, the algorithms used for processing the acoustic, PIR, and seismic signals and the overall logic used to combine the results from each individual algorithm will be discussed.

Acoustic Algorithm: The algorithm first collects statistics with or without people present in the scene. The people are free to talk, walk or simply be silent. We primarily concentrated in the frequency band between 1 – 2000 Hz for these statistics. Four different statistics were collected, one for each band, namely, 1 – 500 Hz, 501 – 1000 Hz, 1001 – 1500 Hz, and 1501 – 2000 Hz. These statistics are estimated as follows: let $F^t = \{f_1^t, f_2^t, \cdots, f_{2000}^t\}$, represent the amplitudes of the spectra of the acoustic signal at instant 't', then the mean M_q^j of the spectral band is estimated as follows:

$$\begin{aligned} s_{t,q} &= \sum_{n=q+1}^{q+500} f_n^t, \quad \forall q \in \{0,500,1000,1500\} \\ M_q^j &= E\{s_{1,q}, s_{2,q}, \cdots, s_{t,q}\}, \quad \forall q \in \{0,500,1000,1500\},\ j \in \{0,1\} \\ \Sigma_q^j &= std.\,dev\,\{s_{1,q}, s_{2,q}, \cdots, s_{t,q}\}, \quad \forall q \in \{0,500,1000,1500\},\ j \in \{0,1\} \end{aligned} \tag{1}$$

where 'E' denotes the expected value and "*std. dev.*" denotes standard deviation and 'j=0' corresponds to the case with no people present and 'j=1' corresponds to the case with people present. We have four means, namely, $\{M_1^0, M_{500}^0, M_{1000}^0, M_{1500}^0\}$, corresponding to four bands and standard deviations $\{std_1^0, std_{500}^0, std_{1000}^0, std_{1500}^0\}$ for the case when no people were present in the scene, similarly we have the means and the standard deviations for the case when people are present. In general M_q^j and Σ_q^j are the representatives of energy level in each band and their variances. Let us denote the energy levels in N bands as

$$X = \{X_1, X_2, \cdots, X_N\}$$

where X_i is the energy in band 'i', and we assume the energy levels in each band are statistically independent and have the Gaussian distribution given by

$$p(X_i) = \frac{1}{(2\pi)^{1/2}|\Sigma_i|^{1/2}} \exp\left\{-\tfrac{1}{2}(X_i - M_i)^T \Sigma^{-1}(X_i - M_i)\right\} \tag{2}$$

where M_i and $\sum_i$ denote the mean and variance respectively and T denotes the transpose. Then the likelihood that a person is present or not is given by

$$p(X \mid H_j) = \prod_{i=1}^{N} p(X_i \mid H_j)\ p(H_j), \quad j = \{0,1\} \tag{3}$$

where H_0 and H_1 are the hypothesis correspond to a person is not present and a person present respectively. Then the posterior probability of human presence is given by

$$p(H_1 \mid X) = \frac{\prod_{i=1}^{N} p(X_i \mid H_1)\ p(H_1)}{\prod_{i=1}^{N} p(X_i \mid H_0)\ p(H_0) + \prod_{i=1}^{N} p(X_i \mid H_1)\ p(H_1)}. \tag{4}$$

Assuming $p(H_0) = p(H_1) = 0.5$, we can compute the posterior probability of a human present given *X*. If it exceeds a particular threshold value, we declare that a human is detected. The algorithm is given below:

Algorithm:

1. Let $s(t)$ corresponds to 1-sec data
2. $S = fft(s)$; is the Fast Fourier transform of the signal $s(t)$.
3. Compute the mean and variances for each band using equation (1)
4. Use equations (2) and (4) with appropriate means and variances for noise and statistics collected on people to compute the posterior probability $p(H_1 \mid X)$
5. Use the posterior probability for fusion and declare that a person is detected if $p(H_1 \mid X) > 0.6$.

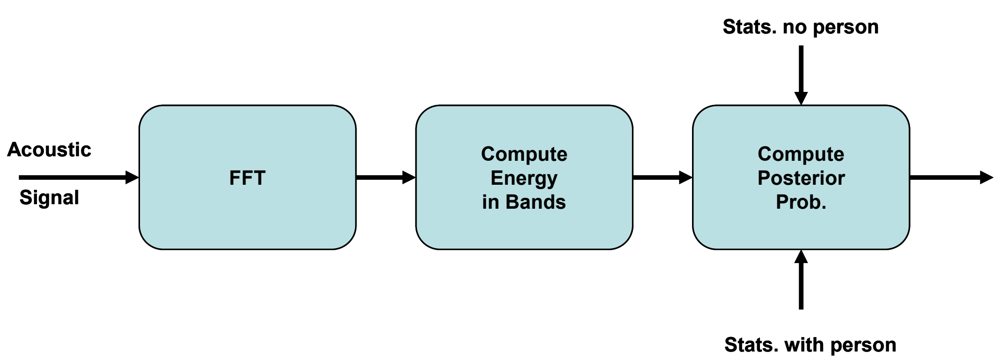

Figure 1: Block diagram for Acoustic Detection Algorithm

Programming aspect: To compute the posterior probability $p(H_1 | X)$ we used the "classify" function in MATLAB. The function takes three inputs, namely, (1) the vector that is to be classified, (2) set of vectors corresponding to the cases with and without personnel present and (3) the class of each vector in the second input. The function "classify" also gives the percentage of wrong classification – this information is used as the uncertainty parameter for fusion using Dempster-Shafer fusion algorithm in section 3.

Seismic Detection Algorithm: The purpose of the seismic sensor is to detect the footprints of a person walking by capturing the vibrations in the floor. A normal walk periodicity of a person or gait frequency is around 2 HZ [1, 2]. Hence, several algorithms try to extract the gait frequency and its harmonics in the seismic sensor data. In order to detect the gait frequency, we use six seconds of data (sliding scale) and take its absolute value to make the envelope of the signal more predominant. We then take the FFT of the absolute signal (FFT of the envelope). If more than one person is present in the scene, the joint gait frequency may be different and might generate multiple harmonics. As a result we select the amplitudes of the first 15 bins of FFT corresponding to 1 to 15 Hz. These fifteen amplitudes become the feature vector. Just as in the case of acoustic detection algorithm, we assume these features have the Gaussian distribution and use equation (3) for determining the $p(H_0 | X)$ and $p(H_1 | X)$ with appropriate mean and variance, which are then used in equation (4) to determine the posterior probability. Note in the case of seismic sensor the N in equation (4) is 15. Use the posterior probability for fusion and declare that a person is detected if $p(H_1 | X) > 0.6$. Once again, we used the "classify" function in MATLAB to compute the posterior probabilities.

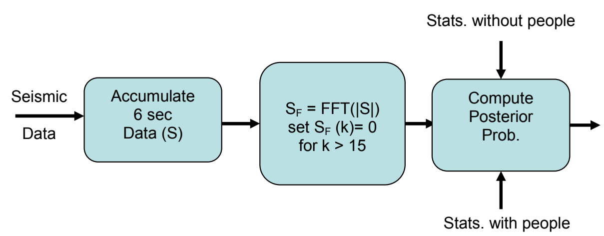

Figure 2: Seismic Detection Algorithm

PIR Detection Algorithm: PIR sensors are typically used as motion detectors. However, the output of the sensor is proportional to the heat radiated by the body of a person or an object. As a result, when a person comes into the range of a PIR sensor it generates an output. The PIR sensors used in this experiment are dual plate sensors where one plate gets positively charged and another gets negatively charged. It is possible to determine the direction of a person walking based on the output of the sensor depending on whether the output changed from positive to negative or vice versa. Since we are only interested in detecting whether a person is present or not, we used a simple threshold detector. Since this detector is in general a very robust sensor the probability of detection p is set to 1 if it exceeds a threshold of 0.3 volt.

Video Motion Detection: Another sensor used is a video camera to detect the presence of personnel. This is done by detecting a moving body in the video frames. The camera captured the scene in its field of view frame by frame every second. For motion detection, the current frame is compared to the previous frame. Timestamps at the bottom of the image are removed before this comparison. The images are then subtracted from each other on a pixel-by-pixel basis to find the difference. The resulting difference 'image' is filtered for noise. First, any small difference values are removed (set to 0). The difference image then undergoes median filtering, which removes the small elements of noise/illumination changes. Median filtering works by making each pixel equal to the median of the pixels within a small area. The small area is defined by a structural element, which for this algorithm was a 4x4 neighborhood. The total variance of the resulting image is computed, and if it exceeds a threshold, motion is detected. Whenever a motion is detected by the motion detection algorithm, the probability of detection of personnel is assigned a value of 1, probability of no detection and uncertainty are assigned a value of zero.

In the next section we present the fusion algorithm used and some results obtained using the fusion algorithm on real data.

3. Fusion of Audio, Seismic, PIR and Video

There are several ways to fuse the information from multiple sensors, namely, Bayesian approach, Dempster-Shafer fusion, etc. Bayesian approach has some disadvantages. In particular, if any one of the sensors gives a low probability of detection, the overall probability goes down considerably even if other sensors have high probability of detection. For example, when a person is close to the maximum detection range of a seismic sensor and beyond, its probability of detection would be close to zero. In this paper we use Dempster-Shafer fusion [3, 4] to fuse the information from all the sensors.

Dempster-Shafer's method involves probability masses given by sensors to be distributed over a set. In this implementation, the set used is {A, B, AB}. 'A' represents the evidence for detection, 'B' represents the evidence against detection, and 'AB' represents the uncertainty or ignorance of the sensor.

Generally, with this scenario, the lack of evidence for detection does not signify that there is no person. For example, the acoustic sensor may not pick up anything if the target is intentionally being stealthy. To determine the uncertainty of a sensor, we compute the percentage of misclassification of training data by each detection algorithm and use it as the uncertainty for that sensor. We also adjust the probability of detection and probability of no detection appropriately so that the sum total of probability of detection, probability of no detection and the uncertainty equals 1.

To combine the results from two sensors (S_1 and S_2), the fusion algorithm uses the

Dempster-Shafer Rule of combination [4]: The total probability mass committed to an event Z defined by the combination of evidence represented by $S_1(X)$ and $S_2(Y)$ is given by

$$S_{1,2}(Z) = S_1(Z) \oplus S_2(Z) = K \sum_{X \cap Y = Z} S_1(X) S_2(Y) \tag{5}$$

where K, the normalization factor is:

$$K^{-1} = 1 - \sum_{X \cap Y = \phi} S_1(X) S_2(Y) \tag{6}$$

This is basically the sum of elements from the set of Sensor 1 who intersect with Sensor 2 to make Z, divided by 1- the sum of elements from S_1 that have no intersection with S_2.

The rule is used to combine all three probabilities (A, B, AB) of sensors S_1 and S_2. The resultant probabilities are combined with the probabilities of with the next sensor.

Data Collection: We collected data in an empty basement of a building with four acoustic, one seismic, four PIR sensors and one video camera to detect personnel as show in figure 3. The experiments were conducted with one person, 2 people, 3 people and 4 people walking the hallway while talking, singing, and in quite mode. The far left sensor location consists of only one video camera; the acoustic, PIR and seismic sensors are deployed as shown in the figure.

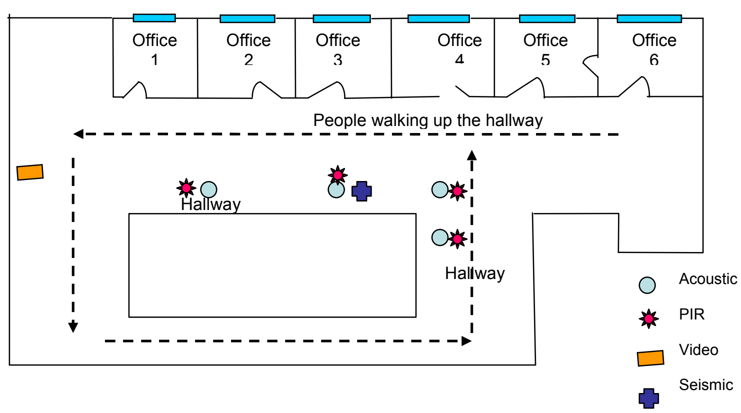

Figure 3: Experimental setup with sensor locations shown

From figure 3, it is clear that the sensors are not collocated. In fact they are placed along the hallway to cover most of the pathway. The orientation of the PIR sensors is such that they cover a large portion of the pathway. Video camera is placed at one side of the pathway so that it can see the entire path. Figure 4 shows a frame taken by the video camera. The seismic sensor is placed on the floor and tried to couple to the ground to the

Figure 4: Corridor experiment – Ground truth

best of the ability. In general seismic sensor should be buried for better coupling to the ground. Figure 5 shows typical outputs from acoustic, PIR and seismic sensors.

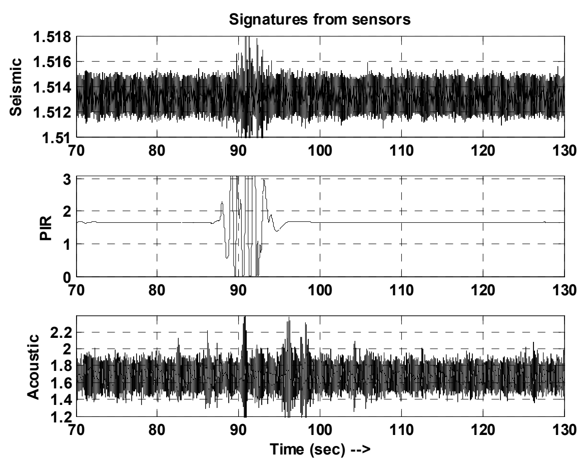

Figure 5: Output of Acoustic, PIR and Seismic Sensors

We used the acoustic, seismic, PIR and Video motion detection algorithms to detect the presence of personnel and computed the probability of detection, probability of no detection and probability of uncertainty for each sensor modality for the data. Figure 6 shows the probability of detection for all the four acoustic sensors and the result of the fusion using only acoustic sensors data. From figure 6, it is clear that the over all probability of detection has improved with fusion.

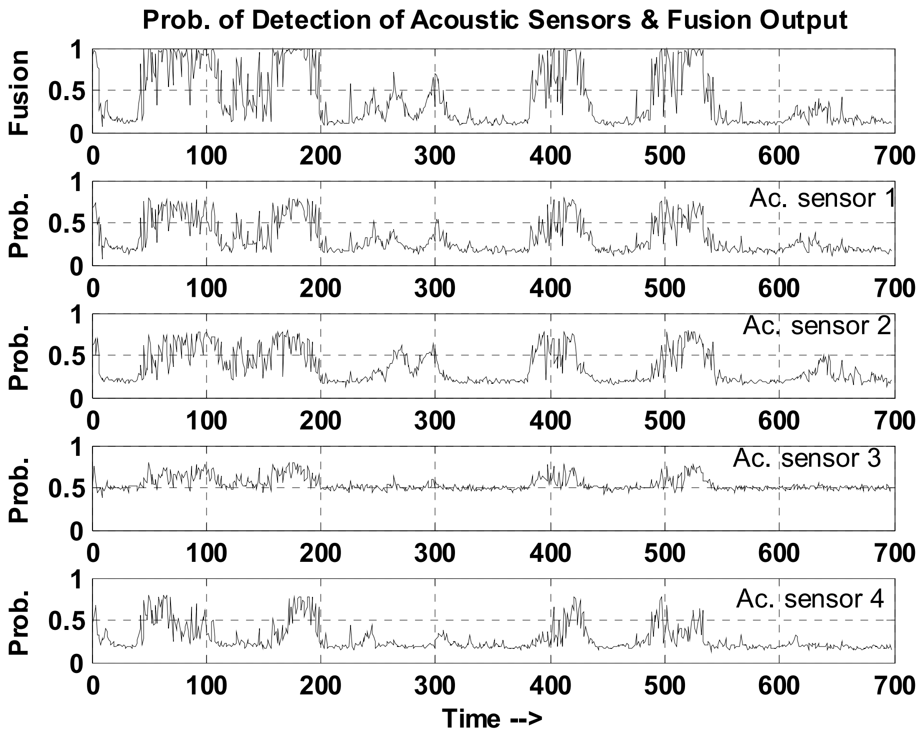

Figure6: Results of fusion using acoustic sensors

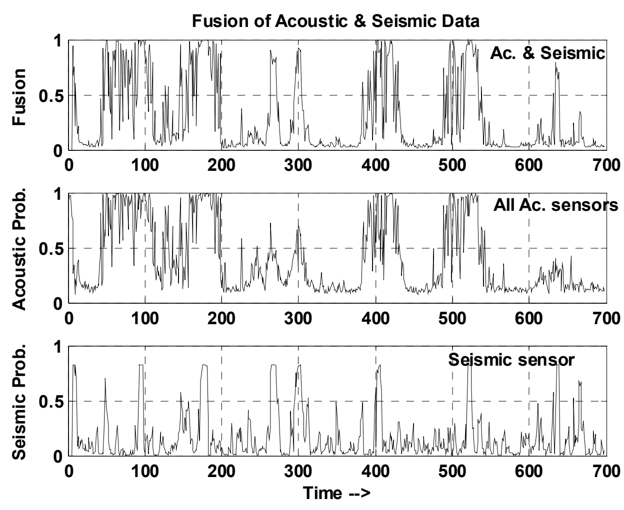

Figure 7: Probability of detection using acoustic and seismic data

In figure 7 we show the results of fusion of acoustic and seismic data. Seismic sensor detects the footprints of a person. From figure 7 we notice around 300 sec time period, the seismic sensor has high probability of detection compared to the acoustic sensors.

This is the case where the people were not talking and hence has low probability of detection. We also notice that the overall probability of detection after fusion is improved. Figure 8 shows the detections by individual sensor modality. As expected the

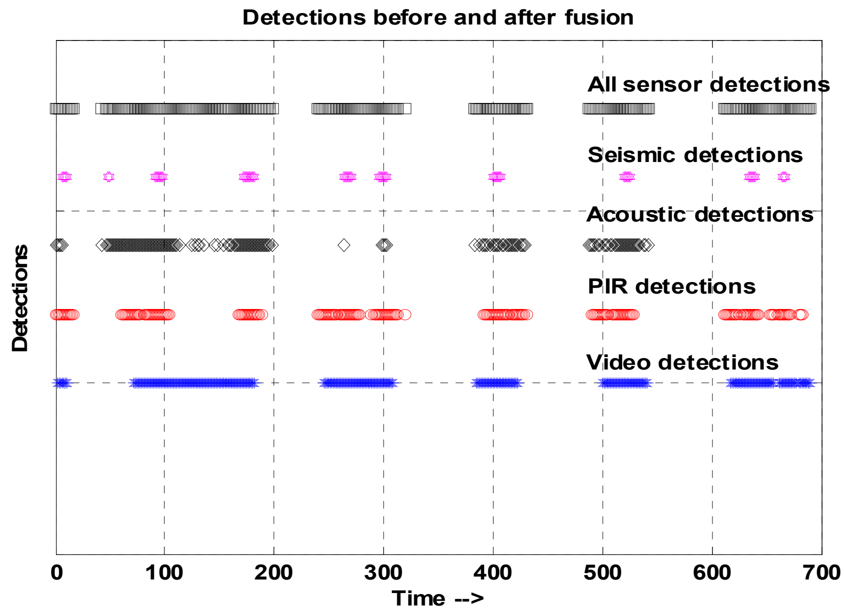

Figure 8: Detection before and after fusion

detections by the seismic sensor are minimum compared to other modalities. Reliable detections are obtained using PIR and video sensors. Acoustic sensors work best when some people are conversing.

4 Conclusions: In this paper we presented technique to fuse multi-modal multiple sensor data fusion and showed that the fusion improves the performance. In general video and PIR sensor have the best capability for detecting people. However, they require line of sight. Acoustic sensor can detect even when the people are not in line of sight. Acoustic sensor can be defeated if one goes into stealth mode. Hence, a variety of sensor modalities are required to detect people robustly. We found Dempster-Shafer fusion works better compared to the Bayesian approach when the sensors are distributed over a wide area.

References:

[1] T. Raju Damarla, et all, "Personnel detection algorithm for disposable sensors using multi sensor data," *Proc. MSS Battlefield Acoustic & Magnetic Sensing Symposium*, Laurel, Maryland, 2005.

[2] K. M. Houston and D. P. McGaffigan, "Spectrum Analysis Techniques for Personnel Detection Using Seismic Sensors", *Proc. of SPIE,* Vol. 5090, 2003, pp. 162-173.

[3] D. L. Hall and J. Llinas, *Handbook of Multisensor Data Fusion*, CRC Press: Boca Raton, FL, 2001.

[4] L. A. Klein, Sensor and Data Fusion – A Tool for Information Assessment and Decision Making, SPIE Press, Bellingham, Washington, USA 2004.