

# Optical Engineering

[SPIDigitalLibrary.org/oe](http://SPIDigitalLibrary.org/oe)

## **Pyroelectric linear array sensor for object recognition**

Srikant Chari  
Eddie L. Jacobs  
Divya Choudhary

# Pyroelectric linear array sensor for object recognition

Srikant Chari,<sup>a,\*</sup> Eddie L. Jacobs,<sup>b</sup> and Divya Choudhary<sup>c</sup>

<sup>a</sup>Independent Consultant, 3607, Highland Park Place, Memphis, Tennessee 38111

<sup>b</sup>University of Memphis, Department of Electrical and Computer Engineering, Memphis, Tennessee 38152

<sup>c</sup>Christian Brothers University, Department of Electrical and Computer Engineering, Memphis, Tennessee 38104

**Abstract.** This paper presents a proof of concept sensor system based on a linear array of pyroelectric detectors for recognition of moving objects. The utility of this prototype sensor is demonstrated by its use in trail monitoring and perimeter protection applications for classifying humans against animals with object motion transverse to the field of view of the sensor array. Data acquisition using the system was performed under varied terrains and using a wide variety of animals and humans. With the objective of eventually porting the algorithms onto a low resource computational platform, simple signal processing, feature extraction, and classification techniques are used. The object recognition algorithm uses a combination of geometrical and texture features to provide limited insensitivity to range and speed. Analysis of system performance shows its effectiveness in discriminating humans and animals with high classification accuracy. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.53.2.023101](https://doi.org/10.1117/1.OE.53.2.023101)]

Keywords: linear array sensor; pyroelectric detectors; object recognition.

Paper 131364 received Sep. 4, 2013; revised manuscript received Jan. 9, 2014; accepted for publication Jan. 10, 2014; published online Feb. 12, 2014.

## 1 Introduction

Perimeter protection and monitoring of national borders are challenging tasks. For example, the U.S.-Mexico border is about 3000 km long and mostly passes through uninhabited areas and rough terrain. The use of manpower to secure this region is very expensive. As a result, there has been growing interest in developing unattended ground sensor systems to address this problem. One approach is to deploy a large number of low cost sensing systems over the length of the boundary that needs to be protected. Since physically accessing the location of a sensor may be difficult and in many cases dangerous, sensors should be capable of operating over long periods without battery replacement. This necessitates the sensors to be capable of operating at low power levels and on resource starved computing platforms. A very common discrimination task in these scenarios is distinguishing between humans and animals. An alarm is typically needed only when a human is detected by the sensor. Most of the ground sensor literature dealing with such scenarios and constraints are typically nonimaging in nature. These sensors include acoustic sensors that measure sounds generated by objects, seismic sensors that measure ground vibrations, E-field sensors that measure static charges on objects, B-field sensors that measure ferromagnetic materials such as fire arms, and RF sensors that can detect cell phone activity.<sup>1</sup> Acoustic sensors have been used to classify different types of ground vehicles by fusing features associated with engine noise and other minor acoustic factors such as tire friction noise.<sup>2</sup> Seismic sensors have been used in conjunction with copula theory based hypothesis testing to detect human foot steps.<sup>3</sup> Wavelet transform followed by symbolic dynamics based feature extraction was applied to data generated from seismic sensors and a single element pyroelectric

infrared detector to distinguish between humans and animals.<sup>4</sup> Sparse dictionary learning techniques have also been applied to seismic sensors to classify human and human with animal categories.<sup>5</sup> Nonimaging target detections and tracking systems have also been designed by coherently fusing large number of fiber optic cables, creating a nonconventional optic to detect bright objects.<sup>6</sup>

Conventional imaging systems have been used extensively in intelligence, surveillance, and reconnaissance applications. These systems have typically been based on using two-dimensional (2-D) focal plane arrays. The disadvantage of such systems for applications of interest in this paper, namely deployment in inaccessible terrains, is power consumption and cost. If the discrimination task is narrowed down to specific classes such as distinguishing between humans, animals, and vehicles at ranges not exceeding a few tens of meters, high resolution, high bit depth imaging systems may not be required. This scenario is true especially in the terrains in which objects can only travel through specific routes consisting of narrow trails. An example of a low resolution, low bit depth image includes an active near-infrared optical trip wire style system.<sup>7-9</sup> The system consists of two vertical posts with one post lined with near-infrared transmitters and receivers. The other post, placed directly opposite, consists of reflectors. The system forms a vertical linear array of pairs of transceivers and reflectors. Each post is placed on the opposite sides of a bottleneck in a trail. When an object walks between these posts, it breaks the optical path of the beam between the transceiver and the reflector. The shape or the profile of the object is traced at the output of the system. This system has provided high classification accuracy in discriminating humans, animals, and vehicles. Another version of this system includes spreading the transceiver-reflector elements along a trail instead of being placed on posts. Being an active system, the power consumption of the trip wire style system is higher than

\*Address all correspondence to: Srikant Chari, E-mail: [srikchari@gmail.com](mailto:srikchari@gmail.com)

what is optimally required. Another disadvantage of this technique is that it requires the object to strictly follow the trail. These obstacles are overcome by the use of a passive linear array which does not require transceiver-reflector pairs and only requires that the object move in a transverse direction to the linear array. Fang et al. used a pyroelectric detector pair in conjunction with a Fresnel lens array to discriminate between individual humans registered in a data base and also reject unregistered humans.<sup>10</sup> Hao et al. used distributed wireless pyroelectric detector pairs to differentiate between different walking humans.<sup>11</sup> In this paper, a linear array of pyroelectric detectors is used. Initial design and testing of a linear array using such a system showed promising results.<sup>12</sup> Another aspect that differentiates the research reported here from the previous efforts is that, in this paper, a two class problems of discriminating humans versus animals is addressed. Further, a more comprehensive performance analysis is performed using a larger data base of objects. Section 2 of this paper provides details of the pyroelectric linear array (PLA) sensor along with an analytic form of the step response of the pyroelectric detector. Section 3 describes the human and animal signature characteristics as acquired by the sensor. Target detection using thermal signature characteristics of the pyroelectric detector are also presented in this section. Section 4 describes various feature extraction and classification algorithms used to classify humans and animals. Results and analysis of classification performance are reported in Sec. 5. In Sec. 6, conclusions drawn from this research and future efforts to make improvements in the sensor system are provided.

The key contribution of this paper is the development of a linear pyroelectric array sensor system. In conjunction with simple signal processing and classification algorithms, it is shown that the system can be used to effectively detect and discriminate humans and animals which makes it relevant for perimeter protection and trail monitoring applications.

## 2 PLA Sensor Design

In this section, first the specifications of the PLA sensor are provided. Next, the response of the pyroelectric detector is discussed in which the detector current and step voltage response of the pyroelectric detector are derived. These equations can provide insight into the characteristics of signals generated by the PLA system in sensing humans and animals.

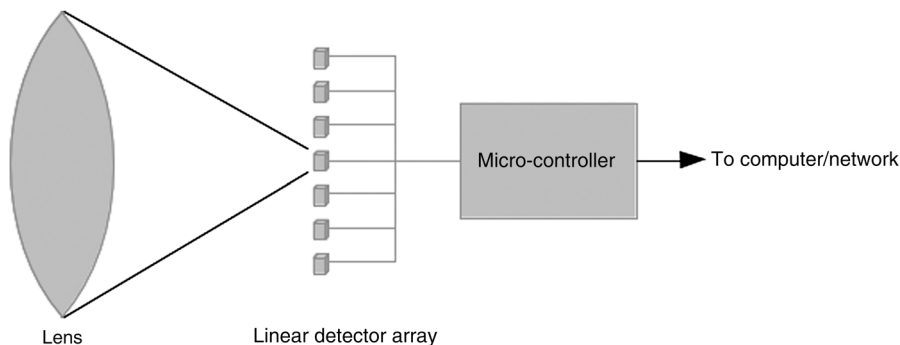


Fig. 1 Schematic of pyroelectric linear array (PLA) sensor.

## 2.1 Sensor Specifications

A schematic of the PLA sensor is shown in Fig. 1. Previous research indicated that 17 samples over a human, of height 2 m, provided sufficient information for accurate target discrimination.<sup>9</sup> The design process started with the choice of a Dias 128 element linear array of pyroelectric detectors. The size of each detector is  $90 \mu\text{m} \times 100 \mu\text{m}$  with a pitch of  $100 \mu\text{m}$ . A germanium lens with an F-number of 0.86 and focal length of 50 mm was chosen to provide a detector instantaneous field of view (IFOV) is  $1.8 \text{ mrad} \times 2 \text{ mrad}$ . As a result, the extent of the spatial sample at a range of 30 m is  $5.4 \text{ cm} \times 6 \text{ cm}$ , providing the required 17 samples over the height of a 2-m tall human. Finally, the lens and the linear detector array were packaged with a18F4550 pic micro-controller for A/D conversion and communication. The sensor system operates at a sampling rate of 20 Hz. A photo of the sensor package is shown in Fig. 2. The current prototype, which is a proof of concept system, operates on a standard 9 V battery that can power the PLA sensor for 4 h in continuous operation. It should be noted that the prototype is not intended for continuous operation. The sensor system is to be deployed in terrains where the monitored trails are used infrequently and signals present themselves for acquisition for only a few minutes a day. The PLA sensor will be a part of a sensor network with other low level sensors that will cue the PLA sensor to turn on and acquire data for short periods of time. For applications of interest such as trail monitoring, this infrared PLA sensor with one column of sensors has an advantage over conventional infrared 2-D focal plane array sensors in terms of power consumption.

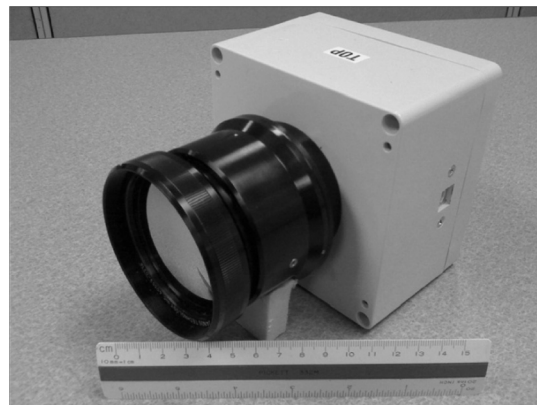


Fig. 2 Photo of PLA sensor.

**Table 1** Definitions for Eq. (1).

Symbol	Definition
$g_v$	Amplifier gain
$\alpha$	Detector absorptivity
$p$	Pyroelectric coefficient
$A_d$	Detector area
$C_e$	Combined amplifier/detector capacitance
$C_p$	Detector heat capacity
$\tau_E$	Electrical time constant
$\tau_T$	Thermal time constant

## 2.2 Pyroelectric Detector Response

The response of a pyroelectric detector was analyzed in a previous paper.<sup>13</sup> As shown there, the voltage response of a detector to an optical source whose temporally varying power on the detector is  $P_d(t)$  can be found from

$$v_d(t) = \frac{g_v \alpha p A_d}{C_e C_p} \left[ \int_0^\infty P_d(t - \xi) \exp\left(-\frac{\xi}{\tau_E}\right) d\xi - \frac{1}{\tau_T} \int_0^\infty \exp\left(-\frac{\xi}{\tau_E}\right) \left\{ \int_0^\infty P_d(t - \xi - \xi') \times \exp\left(-\frac{\xi'}{\tau_T}\right) d\xi' \right\} d\xi \right], \quad (1)$$

where the definitions of the various symbols are those shown in Table 1. Further detail may be found in Ref. 13.

The step response can be found by substituting  $P_d(t) = P_0 + \Delta P u(t)$  into Eq. (1), with  $P_0$  and  $\Delta P$  constants. The function  $u(t)$  is the Heaviside step function. The resulting step response is given by

$$v_{\text{step}}(t) = \frac{g_v \alpha p A_d}{C_e C_p} \Delta P \left[ \frac{\tau_T}{\tau_T - \tau_E} \exp\left(-\frac{t}{\tau_T}\right) - \exp\left(-\frac{t}{\tau_E}\right) \right]. \quad (2)$$

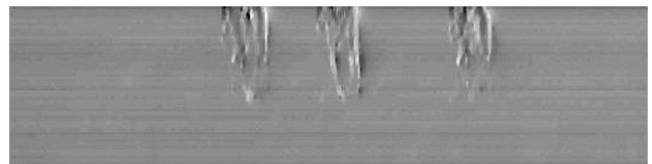
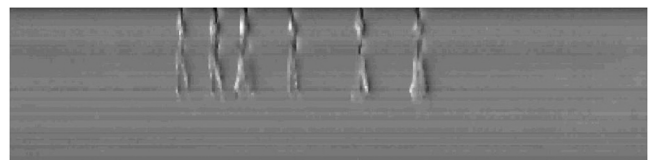
As will be further illustrated through examples in Sec. 3, the difference of exponentials in Eq. (2) explains the rapid change in voltage at the output of the detector as a hot object enters the field of view (FOV) of the sensor. Similarly, it explains the fact that as the hot body leaves the FOV of the sensor an opposite change in the output voltage is observed. For an object colder than its background, the voltage swing polarities are reversed. In summary, a body of uniform temperature would produce an output response only when it enters or leaves the FOV of the sensor.

## 3 PLA Signal Analysis

Data collection using the PLA sensor was performed in two geographically distinct locations. One location was near the U.S.-Mexico border which had an arid terrain with thorn bushes forming significant portions of the vegetation. The other location was a petting zoo near Memphis,

Tennessee, where the terrain was covered with grasses, trees, and rolling hills. The human category was represented by males and females with varying physical builds. The data collected in the petting zoo included animals such as miniature cows, small donkeys, and small ponies. The data collected in the Arizona site included large horses. This different set of animals in the two locations allowed for a diverse set for the animal class. The terrain and animal behavior did not allow for accurate speed measurements or range measurement to be made for each individual target. The speeds of the moving target are described simply by categorizing them as either walking or running and the targets were allowed to traverse at a range of 10 to 20 m from the sensor. It should be noted that the claim made in this paper, that the PLA system has limited speed and range insensitivity, is based on its ability to discriminate targets at the above specified range limits and speed categories. The objects moved in a direction that was transverse to the FOV of the sensor array. As the objects move through the FOV of the vertical linear array, the object thermal signature is traced out as a function of time over the 128 detectors. Thus each object generates  $D \times N$  data points. Here,  $D$  is the number of detectors on the linear array that sense the complete extent of the object and  $N = T \times R$ , where  $T$  corresponds to the period for which the object traverses in the FOV of the sensor and  $R$  is the sampling rate of the sensor. For visualization purposes, the  $D \times N$  data points can be displayed as images as in Fig. 3 which shows the output image of the PLA sensor with three horses. Figure 4 shows six humans and Fig. 5 shows a human followed by a miniature cow.

The process of object detection and segmentation involves first determining the presence of large sized objects in the image, next the suppression of isolated noise is implemented, and finally objects that may have fragmented are grouped back together. The initial process of detection is done using a statistical threshold. To determine the threshold for a given instant of time, a constantly updating buffer is used. For any instant of time, the buffer contains the previous  $N_b$  time samples from the output of each of the detectors. The mean  $\mu$  and standard deviation  $\sigma$  are then estimated for each detector at that time instant based on the data in the buffer. If the samples at a given instant of time have a value that is  $S\sigma$  away from the mean, then the output of the corresponding detector is marked as a potential target. A value of  $S = 1.5$  was determined to be an effective


**Fig. 3** Output of PLA sensor showing three horses.

**Fig. 4** Output of the 128 detectors versus time showing six humans.

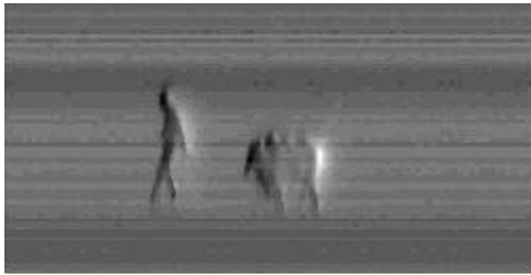


Fig. 5 Output of the 128 detectors versus time showing a human followed by a miniature cow.

value as a multiplicative constant for  $\sigma$ . The buffer is updated for the next time instant on a first-in first-out basis. When a target is moving in the FOV of the sensor, a large number of detector outputs in a given spatiotemporal neighbor exceed the  $S\sigma$  threshold. Following the  $S\sigma$  thresholding of individual detector outputs, a spatial temporal size threshold is used to suppress isolated noise pixels. Apart from noise, a single object may break into two or more fragments. These fragments need to be grouped together. The reason for this fragmentation is as follows. Consider Fig. 6 showing the output of detector number 26 in response to a human followed by a horse. As the human enters the FOV of the sensor, a large downward swing is observed at the output of the detector and a large upward swing is observed when the human leaves the FOV of the sensor, forming a bipolar pair. The bipolar pair is also observed at the output of the sensor in response to the horse but due to the larger width of the horse, the downward and upward swings have more separation in time. This signal behavior is linked to the discussion in Sec. 2.2 about detector characteristics. This property of the detector suppresses static objects in the background and only responds to the moving targets. However, if a target does not show variations in sections of its body, then the detectors do not respond. This can lead to fragmentation of an object. To overcome this problem, the pyroelectric detector response characteristics can be used. The regrouping of the fragments is done by tracking the bipolar swings that mark the entry and exit of the target in the FOV of the sensor. Following the grouping of objects, the images are processed for gray scale feature extraction. The image can also be binarized with segmented targets set to logic 1 and background set to logic 0 so as to extract binary features. This

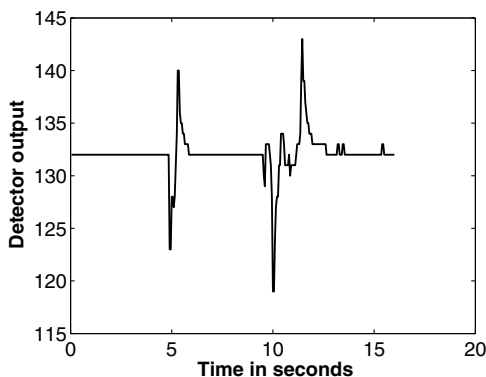


Fig. 6 Output of PLA sensor detector number 26 in response to a human followed by a horse.

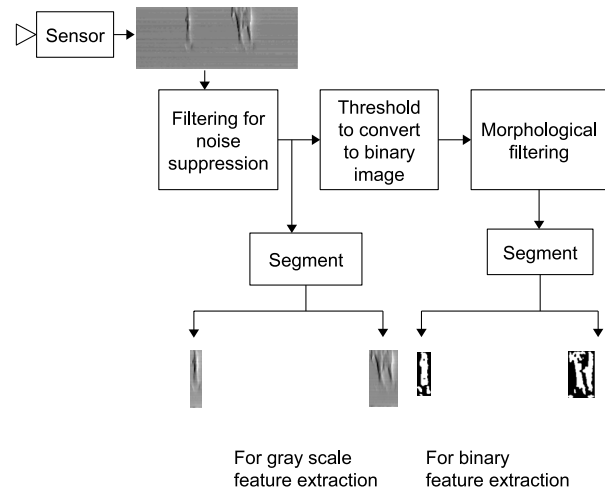


Fig. 7 PLA sensor signal processing.

process is described in Fig. 7. The process of gray scale and binary feature extraction is discussed in Sec. 4.2.

#### 4 Object Recognition

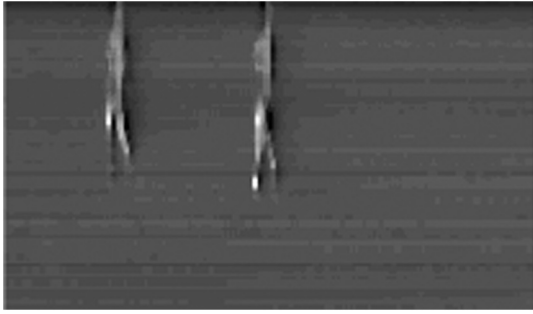
Once the presence of a potential target has been established through the process described in Sec. 3, the next step is to categorize the object as a human or an animal. This is a two step process. In the first step, signal characteristics or features are extracted from the object signal. In the second step, the extracted features are processed by classification algorithms. Two types of features were extracted, namely, the height-to-width ratio and energy in frequency bands at the output of Gabor filters. The features are then processed by a classification algorithm to determine whether a test object is human or animal. The classification algorithms include logistic regression, decision trees, and Gaussian mixture models (GMMs). The following combinations of features extraction techniques and classification algorithms were applied.

- Decision tree based classification algorithm which uses both the height-to-width ratio and the Gabor features.
- Logistic regression using height-to-width ratio.
- GMM based classifier using height-to-width ratio.

In the following section, first the feature extraction techniques are presented followed by a discussion about the three classification algorithms.

##### 4.1 Height-Width Feature Extraction

The motivation behind measuring height and width of objects for the problem of interest is that the height of a human is much greater than his/her width, while the difference between the height and width of an animal, walking at slow speeds, is comparatively smaller. These two features are measured from the binary image of an object. Height is measured as the largest dimension of the object along the vertical direction. This is done by identifying the row in the image where the first occurrence of “ON” pixels takes place and the row where the last set of “ON” occur. Difference between the row numbers is used as the measure of height. Similarly, to measure width, the column in the image where the first



**Fig. 8** Height and width measurements of a human and an animal.

occurrence of “ON” pixels takes place and the column where the last set of “ON” occur are identified. Difference between the column numbers is used as the measure of width. Figure 8 shows an example of the height-width feature of a human and a large horse. Although this feature has been shown to be very effective in discriminating humans from animals, it is not perfect. Since the sensor array is sampled at regular time intervals, the faster an object moves the more compressed it becomes in the horizontal direction. As a result, fast moving animals can have height-width features that are similar to humans. This problem is addressed in Sec. 4.3.3.

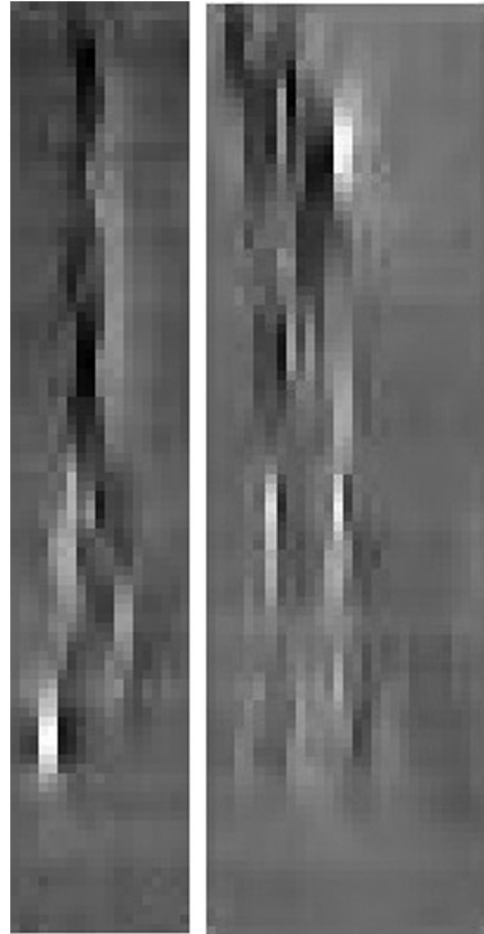
## 4.2 Gabor Feature Extraction

Gabor filters have been extensively used to discriminate objects based on texture.<sup>14</sup> Figure 9 shows an example of a human and a running animal with similar height-to-width ratio but different gray scale textures. To exploit textural difference between humans and animals for classification, each object in gray scale is filtered by a bank of Log-Gabor filters and the energy at the output of the filters computed. The energy values are used as features of the object and these features are processed by classifier to identify if the object is an animal or a human. Log-Gabor filter, a variation of the standard Gabor filter, is given by the product of its radial and angular components as shown in

$$G(f, \theta) = \exp \left\{ \frac{[-\log(\frac{f}{f_0})]^2}{2\sigma_f^2} \right\} \exp \left\{ -\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2} \right\}, \quad (3)$$

where  $f_0$  is the radial center frequency and  $\theta_0$  is the orientation of the filter. The parameters  $\sigma_f$  controls the scale bandwidth and  $\sigma_\theta$  controls the angular bandwidth of the filter.<sup>15,16</sup>

A total of 24 Log-Gabor filters, corresponding to six orientations and four frequencies, were used with  $\sigma_f = 0.65$  and  $\sigma_\theta = 1.5$ .<sup>17</sup> The orientation angles used are 0, 30, 60, 90, 120, and 150 deg. The highest frequency of the filter bank is chosen so that it is less than the normalized Nyquist frequency. The energy content in each of the 24 band pass images of an object is measured.<sup>18</sup> This is done by squaring each pixel, summing them, and then finally dividing this sum by the total number of pixels in the band pass image.<sup>19</sup> A total of 24 features, which measure frequency content of the object in the corresponding 2-D spatial frequency bands, are extracted for each object.



**Fig. 9** Sample of images of human and running animal showing gray scale texture difference between the objects.

## 4.3 Classification Algorithms

Three classification algorithms were tested for their ability to use the features extracted from an object to classify them as human or animal. Three supervised learning algorithms were implemented in MATLAB. In supervised learning, the data set is partitioned into a training set that is used to learn parameters of the classifier model and a testing set that is used to evaluate the performance of the classifier. Logistic regression, GMM, and decision tree using Mahalanobis distance based classification algorithms were tested for classification accuracy. It should be noted that during the training phase, estimation of the parameters of the GMM requires the use of expectation maximization (EM). The EM algorithm is computationally intensive. However, for the application of interest, when the system will actually be deployed in the field, the training will be done offline. Once the mixture parameters are estimated, the calculation of posterior probability using GMM, which is not computationally challenging, can be done online. These statements on the difference between computational cost during training and testing are also valid for Mahalanobis distance and logistic regression as well.

### 4.3.1 Logistic regression

In this research, a variation of the linear regression technique, referred to as the logistic regression, is used to predict the

class  $y$  given the target feature vector  $x = [x_1, x_2, \dots, x_n]$ . Here,  $n$  is the number of features used to represent a target. Logistic regression provides a probabilistic interpretation to the prediction,  $y$ , by limiting it between 0 and 1.<sup>20</sup> The relationship between the predicted class  $y$  and the target features  $x = [x_1, x_2, \dots, x_n]$  is given by the logit function as shown in

$$y = h(\beta^T x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}}. \quad (4)$$

The task, here, is to estimate the weighting coefficients  $\beta = [\beta_1, \beta_2, \dots, \beta_n]$  using features from targets in the training set. The coefficients,  $\beta$ , can be estimated using techniques such as gradient descent that minimizes the cost function given by<sup>21</sup>

$$\begin{aligned} J(\beta) = & -\frac{1}{(m+p)} \sum_{i=1}^{m+p} \{-y^i \ln[h(\beta^T x^i)] \\ & - (1 - y^i) \ln[1 - h(\beta^T x^i)]\} \\ & + \frac{\lambda}{2(m+p)} \sum_{j=1}^n \beta_j^2. \end{aligned} \quad (5)$$

In Eq. (5),  $x^i$  and  $y^i$  correspond to the feature vector and training label of the  $i$ 'th target in the training set,  $\beta_j$  corresponds to the coefficients of the  $j$ 'th feature,  $m$  represents the number of targets in the human training set,  $p$  represents the number of targets in animal training set, and  $\lambda$  is the regularization factor. The training label  $y^i$  can be set to 0 for animal class and 1 for the human class.

In implementing the gradient descent algorithm, the following equations are used to compute the gradient of the cost function with respect to  $\beta_j$ : for  $j > 0$ ,

$$\frac{\partial J(\beta)}{\partial \beta_j} = \left\{ \frac{1}{(m+p)} \sum_{i=1}^{m+p} [h(\beta^T x) - y^i] x_j^i \right\} + \frac{\lambda}{(m+p)} \beta_j \quad (6)$$

for  $j = 0$ ,

$$\frac{\partial J(\beta)}{\partial \beta_j} = \left\{ \frac{1}{(m+p)} \sum_{i=1}^{m+p} [h(\beta^T x) - y^i] x_j^i \right\}. \quad (7)$$

Once  $\beta$  has been estimated, the value of  $y$  can be predicted for the class of a test target using Eq. (4). If the predicted value of  $y$  for a test target is  $>0.5$ , the target is assigned to the human class, else the target is assigned to the animal class.

### 4.3.2 Gaussian mixture model

Finite mixture models (FMM)<sup>22</sup> can be used to represent the distribution of a random variable  $X$  as a weighted sum of a finite number of constituent distributions as shown in

$$p(X|\Theta) = \sum_{k=1}^K w_k p_k(X|\theta_k), \quad (8)$$

where  $K$  represents the total number of constituent distributions forming the mixture,  $X$  represents the random variable

(features in this case),  $p(X|\Theta)$  represents the finite mixture,  $p_k(X|\theta_k)$  represents the  $k$ 'th constituent distribution, and  $w_k$  and  $\theta_k$  represent the mixing coefficient and the parameters of the  $k$ 'th constituent distribution.

GMM is a specific case of FMM where all the constituent distributions are Gaussian distributed. In this research, GMMs are used to represent the probabilistic distribution of features of the two classes. One GMM is used to model the features of the human class and another is used to model the features of the animal class. Modeling the features using GMM requires the estimation of  $w_k$  and  $\theta_k$ . EM technique is used to estimate the GMM using features extracted from the training set.<sup>23</sup> After the GMMs have been estimated for each class, the classification of a test target is done as follows. The posterior probability of a target is computed using the GMMs for each class based on target features. The target is assigned to the class for which the target has highest posterior probability.

The following describes the EM technique of estimating the GMM for a particular class. The EM algorithm is an iterative technique. The steps below describe the  $t$ 'th iteration for a particular class. In Eq. (9),  $x^i$  is the feature vector associated with  $i$ 'th target in the training set in a particular class.  $N$  is the number of training targets for the class under consideration. First, a value of  $K = 1$  is picked. The choice of  $K = 1$  creates a GMM with only one constituent distribution, which is same as a conventional Gaussian distribution.

Step 1: The posterior probability of the  $k$ 'th constituent distribution given the  $i$ 'th data sample is computed using Bayes' rule

$$P(\theta_{k,n}|x^i) = \frac{P(\theta_{k,t})P(x^i|\theta_{k,t})}{P(x^i)} = \frac{w_{k,n}P(x^i|\theta_{k,t})}{\sum_{k=1}^K w_{k,n}P(x^i|\theta_{k,t})}. \quad (9)$$

Step 2: Mixing coefficient of the  $k$ 'th constituent distribution is calculated for the use in the  $(t+1)$ 'th iteration is calculated using

$$w_{k,t+1} = P(\theta_{k,t}) = \frac{1}{N} \sum_{i=1}^N P(\theta_{k,t}|x_i). \quad (10)$$

The parameters of the  $k$ 'th constituent distribution,  $\theta_{k,n+1}$ , for the  $(n+1)$ 'th iteration are obtained by solving the

$$\sum_{i=1}^N P(\theta_k^t|x_i) \frac{\partial [P(x_i|\theta_k^{t+1})]}{\partial \theta} = 0. \quad (11)$$

The above two steps are repeated iteratively until parameters of the model converge. This process of estimating GMMs is repeated by  $K$ . To determine the value of  $K$  that best represents the data, the model selection Akaike's Information Criterion (AIC) is used. Given a set of models under evaluation, the AIC metric value can be used to identify the best model for the data.<sup>24</sup> In this case, each model under evaluation is an FMM with a particular value of  $K$  and the model with the lowest AIC value is chosen. AIC is given by

$$\text{AIC} = -2 \ln[P(x|\hat{\theta})] + 2Q, \quad (12)$$

where  $\hat{\theta}$  are the estimated parameters of the model and  $Q$  is the number of parameters of the statistical model. The AIC equation is composed of two parts, the negative log-likelihood that measures how well the model under test fits the data and the number of parameters which penalizes a model for over fitting. AIC selected  $K = 2$  for both the GMM modeling the animal class and the GMM representing the human class.

**4.3.3 Decision tree**

The idea behind using a decision tree for classifying humans and animals is as follows. Analysis of the data indicates that the geometric feature based on height-to-width ratio is very effective in classifying walking and running humans against walking animals. The feature, in some cases, is not effective in classifying humans against running animals, since the height-to-width ratio of humans and running animals can be similar. To address this issue, the two class problems are split into a three class problems: humans, animals, and running animals. A decision tree with two nodes is employed as shown in Fig. 10. At the first node, humans and animals are classified using the height-to-width ratio. At this node, if the classifier labels the test object as an animal, no further processing is done. However, if the classifier labels the object as a human at this node, then Gabor features are extracted from the object and the decision making process is transferred to the second node. The classifier at this node, based on the Gabor features, classifies the object as either humans or running animals.

At each node of the tree, the decision of which path to traverse is determined by the Mahalanobis distance classifier. The Mahalanobis distance is a statistical measure of how far

a test sample is, with respect to a particular class described by its class mean and class covariance. If  $X_t$  represents feature vector of a test target, then the Mahalanobis distance between the test sample and the  $i$ 'th class is given by<sup>25</sup>

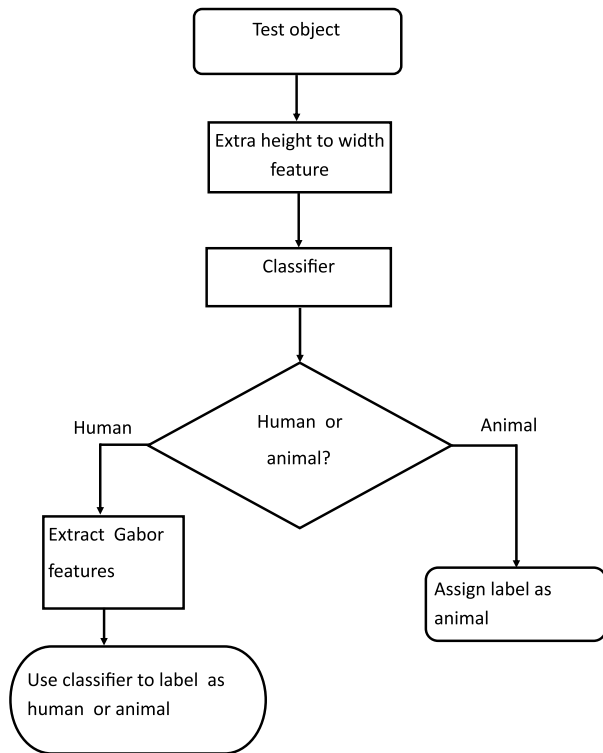
$$D_i = (X_t - \mu_i)C_{X_i}^{-1}(X_t - \mu_i)^T, \tag{13}$$

where  $\mu_i$  and  $C_{X_i}$  are the mean vector and covariance matrix of the  $i$ 'th class, respectively. These parameters are estimated from the training samples of the  $i$ 'th class. The test sample  $X_t$  is assigned to the class with the lowest Mahalanobis distance value.

**5 Results**

A total of 315 human profiles and 182 animal signatures were collected using the pyroelectric sensor at two locations: near the U.S.-Mexico border in Arizona (ArzData) and at a petting zoo near Memphis (PzData). The number of samples at each location for the three data collection sites is shown in Table 2. In a seminal paper comparing accuracy estimation techniques, Kohavi showed that stratified  $K$ -fold cross-validation with  $K = 10$  is an extremely effective tool for accuracy estimation.<sup>26</sup> This technique is used in this paper to evaluate the classification performance of the system. In this approach, the total data set was partitioned into 10-folds. Each fold contained randomly chosen data points from each of the data collection sites. In each fold, the percentage of members from the two classes is made equal to that of the overall data set. The classification rate is computed by holding out one group for testing while the rest of the nine groups is used for training the classifiers. This process is repeated over all the groups and the classification rates are averaged to compute the overall classification rate.

The feature extraction and classification algorithms described in Sec. 4 were applied to the data. Classification rates are shown in Table 3. It is observed that the decision tree approach that uses both the height-to-width ratio feature



**Fig. 10** Object recognition algorithm using decision tree for classifying humans against animals.

**Table 2** Number of data samples collected using pyroelectric sensor at various sites.

	ArzData location 1	ArzData location 2	PzData
Human signatures	140	145	41
Animal signatures	73	84	41

**Table 3** Comparison of classification rates.

Feature extraction	Classifier	Classification rate (%)
Height-to-width ratio and Gabor features	Decision tree	94
Height and width	Logistic regression	87
Height-to-width ratio	GMM	84
Gabor features	Mahalanobis distance	72



**Table 4** Confusion matrix for decision tree classifier.

	Human predicted (%)	Animal predicted (%)
Human	95.4	4.6
Animal	8.6	91.4

and the gabor texture feature achieves the highest classification rate of 94%. Logistic regression using the height-to-width ratio provides a classification accuracy of over 87% and GMM classifier using height-to-width ratio classifies objects with an accuracy of over 84%. The Mahalanobis distance classifying based on the Gabor features has a classification rate of over 72%. Table 4 shows the confusion matrix for the decision tree-based approach. The main cause of error was the incorrect grouping of fragments of objects during the object detection phase. There were a total of 16 running animals in the data set. The decision tree approach that uses both height-to-width ratio and Gabor features to achieve speed independence, classifies the running animals at a rate of 98%. The use of dimensionality reduction techniques such as linear discriminant analysis<sup>27</sup> only provided marginal improvement in the classification accuracy. Algorithmic implementation of wavelet packet-based feature extraction was also done,<sup>28</sup> however, its classification performance was significantly lower and its results are not reported here.

Another technique of measuring classification accuracy is to separate the training and testing data based on data collection sites. In this performance measurement, in one round, the Arizona site is used to provide the training samples and the petting zoo data are used for testing. In the next round, the petting zoo data are used as training samples and the Arizona site data are used for testing. The results of the two rounds are then averaged to compute the overall performance. In this type of performance measurement, the highest classification rate of 71% was obtained using the decision tree classifier. To understand the reason for lower performance, in comparison to the  $k$ -fold cross-validation technique, the classifier boundaries in the feature space were analyzed. It was noticed that the classifier boundaries determined based on one site did not provide good partitioning of data for the classes of the other sites. The animal features from the Arizona site had significant differences from the features of animals of the petting zoo site. This issue could be attributed to the fact that the category of animals in the Arizona site was very different from those in the petting zoos. Due to the lack of diversity in the individual data sites, the classifier models over fitted during training and performed poorly during testing. In comparison, the better performance under the  $K$ -fold based evaluation is because the  $K$ -fold data partitioning provided for a good diversity in the training set. This allowed for the classifier models to generalize and hence provided high classification rates of over 94%. However, the lower performance in the site based evaluation is acknowledged and this research, as a result claims only limited insensitivity instead of invariance to range and speed. Further, the results indicate that more robust features are needed to address this issue. Increasing the temporal and sampling capabilities of the sensor can also alleviate the problem. It should be noted that the sensor is a proof of

concept system and these issues are to be addressed in future research on this topic.

## 6 Conclusions

This paper presents a proof of concept sensor system that uses a linear array of pyroelectric detectors for moving object recognition. The response of the detectors inherently suppresses the static background and only responds to moving objects in the image. The use of this system in trail monitoring is demonstrated by its ability to distinguish between humans and animals when object motion is transverse to the FOV of the sensor. This system can be used as a cost-effective alternative to conventional 2-D focal plane array sensors that use detectors such as microbolometers. A data collection effort was undertaken near the U.S.-Mexico border and at a petting zoo to acquire signatures of humans and a variety of animals. Simple noise mitigation, object detection, feature extraction, and classification are applied for discriminating the two classes of interest. Various object features and classification algorithms are compared based on their ability to accurately classify objects. Using  $K$ -fold cross-validation, it is demonstrated that a decision tree based classifier that uses a combination of geometric characteristics and texture features can be used to discriminate between humans and animals with high classification rates. Several tasks are foreseen in the future to make the system more practical and accurate for field deployment. This includes implementing the software on low resource computing platforms for real-time performance and incorporating this system into a multimodal sensor network. To improve accuracy, it is hypothesized that fusion of features from the PLA sensor with sensors such as acoustic and seismic sensors will further increase the discrimination rates. A sensor employing a two column arraies is currently being developed. The time of appearance of an object in each of the columns can be used to estimate the speed of the objects. With the speed estimated in using the two column approaches, a speed normalized height-to-width ratio can then be used for classifying targets, which computational is less intensive than using a combination of Gabor features and conventional height-to-width ratio. Further, a more sensitive detector array is being used for the development of a long range PLA with the ability to discriminate well beyond the 30 m range limit of the current PLA system. This long range PLA system will provide greater sensor to object stand off distance thus expanding the utility of the sensor beyond trail monitoring into other tactical and military applications. The higher sensitivity detector system can reduce the fragmentation effect at the sensor output that has also been a source of classification errors.

## Acknowledgments

The authors would like to thank John Hutchinson and Josh Gabonia of the U.S. Army Night Vision and Electronic Sensors Directorate, Ft. Belvoir, Virginia, who fabricated the sensor used in this work and Mr. Harry McClellan of EOIR Technologies, Inc. who provided support for data collections. The authors are also grateful for the support of Thyagaraju Damarla, Nino Srour, and Ronnie Sartain of the U.S. Army Research Laboratory, and to Jeremy Brown, Forrest Smith, and Jason Brooks of the University of Memphis for their assistance with the data collections

performed in this work. The work presented in this paper was conducted with the financial support of a U.S. Army Research Laboratory cooperative agreement (Contract W911NF-10-2-0071) and through funding from the U.S. Army Night Vision and Electronics Sensors Directorate through a subcontract through EOIR Technologies, Inc. (Purchase Order P4002151).

## References

1. R. Damarla and D. Ufford, "Personnel detection using ground sensors," *Proc. SPIE* **6562**, 656205 (2007).
2. B. Guo, M. Nixon, and T. Damarla, "Acoustic information fusion for ground vehicle classification," in *2008 11th Int. Conf. on Information Fusion*, pp. 1–7, IEEE, Cologne, Germany (2008).
3. A. Sundaresan et al., "A copula-based semi-parametric approach for footstep detection using seismic sensor networks," *Proc. SPIE* **7710C** (2010).
4. X. Jin et al., "Target detection and classification using seismic and PIR sensors," *IEEE Sens. J.* **12**(6), 1709–1718 (2012).
5. N. Nguyen, N. Nasrabadi, and T. Tran, "Robust multi-sensor classification via joint sparse representation," in *2011 Proc. of the 14th Int. Conf. on Information Fusion (FUSION)*, pp. 1–8 (2011).
6. J. Geary et al., "Dragonfly directional sensor," *Opt. Eng.* **52**(2), 024403 (2013).
7. D. J. Russomanno et al., "Sparse detector sensor: profiling experiments for broad-scale classification," *Proc. SPIE* **6963**, 69630M (2008).
8. D. Russomanno, S. Chari, and C. Halford, "Sparse detector imaging sensor with two-class silhouette classification," *Sensors* **8**(12), 7996–8015 (2008).
9. D. Russomanno et al., "Near-IR sparse detector sensor for intelligent electronic fence applications," *IEEE Sens. J.* **10**(6), 1106–1107 (2010).
10. J. Fang et al., "A pyroelectric infrared biometric system for real-time walker recognition by use of maximum likelihood principal components estimation method," *Opt. Express* **15**(6), 3271–3284 (2007).
11. Q. Hao, F. Hu, and Y. Xiao, "Multiple human tracking and identification with wireless distributed pyroelectric sensor systems," *IEEE Syst. J.* **3**(4), 428–439 (2009).
12. W. E. White, III et al., "Real-time assessment of a linear pyroelectric sensor array for object classification," *Proc. SPIE* **7834**, 783403 (2010).
13. E. L. Jacobs et al., "Pyroelectric sensors and classification algorithms for border and perimeter security," *Proc. SPIE* **7481**, 74810P (2009).
14. T. Randen and J. Husoy, "Filtering for texture classification: a comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(4), 291–310 (1999).
15. D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Am. A* **4**(12), 2379–2394 (1987).
16. W. Wang et al., "Design and implementation of log-Gabor filter in fingerprint image enhancement," *Pattern Recognit. Lett.* **29**(3), 301–308 (2008).
17. P. D. Kovsi, "MATLAB and Octave Functions for Computer Vision and Image Processing," *Centre for Exploration Targeting, School of Earth and Environment, University of Western Australia, Australia* <http://www.csse.uwa.edu.au/~pk/research/matlabfnfs/>.
18. A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognit.* **24**(12), 1167–1186 (1991).
19. R. Jenssen and T. Eltoft, "Independent component analysis for texture segmentation," *Pattern Recognit.* **36**(10), 2301–2315 (2003).
20. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).
21. C. Zhang and Z. Zang, *Boosting-Based Face Recognition and Adaptation*, Morgan & Claypool, San Rafael, California (2011).
22. G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York (2000).
23. W. Martinez and A. Martinez, *Computational Statistics with Matlab*, CRC Press, Boca Raton, Florida (2001).
24. H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Second Int. Symp. on Information Theory*, Vol. 1, pp. 610–624 (1992).
25. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley (2000).
26. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Int. Joint Conf. on Artificial Intelligence*, pp. 1137–1143, Morgan Kaufmann, San Francisco, California (1995).
27. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic (1990).
28. T. Chang and C. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. Image Process.* **2**(4), 429–441 (1993).

**Srikant Chari** received his PhD from the University of Memphis in 2007. From 2007 to 2009, he was a postdoctoral fellow at the University of Memphis. From 2009 to 2013, he held the position of research assistant professor at the University of Memphis. He is currently an independent consultant specializing in digital signal processing, computer vision, image quality metrics, machine learning, and sensor networks.

**Eddie L. Jacobs** received BS and MS degrees in electrical engineering from the University of Arkansas in 1986 and 1988, respectively, and a Doctor of Science in electrophysics from George Washington University in 2001. From 1989 to 2006, he was an engineer with the U.S. Army Night Vision and Electronic Sensors Directorate, Fort Belvoir, Virginia, where he led a team of engineers and scientists developing models of the performance of passive and active imaging sensors. He is currently an associate professor in the Department of Electrical and Computer Engineering at the University of Memphis.

**Divya Choudhary** received her PhD degree from the University of Memphis in 2009. She is currently an assistant professor in the Department of Electrical and Computer Engineering at Christian Brothers University. Her research interests include wireless communication, statistical signal processing, and machine learning.