Data Fusion for Identity Estimation and Tracking of Centroid using Imaging Sensor Data

V.P.S. Naidu, Girija G., and J.R. Raol

National Aerospace Laboratories, Bangalore-560 017

ABSTRACT

Two aspects involved in automatic target recognition namely, (i) Location and identity estimation (LIE) of a target by fusing infrared (IR) and acoustic sensor data, and (ii) centroid tracking for target state estimation using IR sensor data are discussed in this paper. The LIE has been achieved using a combination of Bayesian fusion and one of the three search algorithms namely, metropolis hastings (MH), simulated annealing (SA) and gradual greedy (GG). It was observed that the performance of the GG search algorithms was better in terms of success rate which has been evaluated through Monte Carlo simulations. For tracking of the centroid, an algorithm, where the centroid of the gray level image is tracked using probabilistic data association filter, has been implemented. Simulated data results indicate good tracking performance of this algorithm. For robust tracking of centroid, the track from the imaging sensor was fused with the track from ground-based radar using state vector fusion. It was observed that fusion generates robust tracks even when there is data loss in one of the sensors.

Keywords: Data fusion, identity estimation, centroid tracking, centroid computation, imaging sensor data, automatic target recognition, metropolis hastings, simulated annealing, gradual greedy

1. INTRODUCTION

Automatic target recognition (ATR)[1] is one of the key components of present and future defence weapon systems to be used on autonomous as well as manned vehicle missions. ATR is extremely important for the safety and early warning of the perceived threat in air traffic control or air defence systems. One of the essential ingredients of ATR process involves automatic target acquisition, identification and tracking by processing a sequence of images. In such applications, suitable algorithms for registration, detection, classification, feature computation would be required for target location and identity estimation. Further, if the targets are moving, algorithms for tracking targets would involve aspects like segmentation, centroid computation and tracking using appropriate algorithms. In addition, if multiple sensors are present which give additional information, all the information could be combined in an appropriate way, not only to generate inference about the situation but also increase the confidence in the ATR process itself.

This paper presents two aspects involved in ATR: (i) Location and identity estimation (LIE) of a target by fusing infrared (IR) and acoustic data, and (ii) tracking of a centroid for target state estimation using IR sensor data. While the first aspect is handled using Bayesian fusion and search algorithms[2], the second aspect is solved using a centroid tracking algorithm[3] (CTA).

The IR sensor detects all the targets in its field of view (FOV) and generates images, which provide information of both the target location and identity. Acoustic sensor provides data which help determine the direction of the target. These sensors outputs or their extracted features cannot be fused directly since the information from the two sensors is disparate. Bayesian approach[2] enables fusion of such data by fusing the posterior probabilities of the sensor outputs. Inference about the scene namely the location and identity of objects, using Bayesian fusion is obtained by:

(i)   Formulating a prior distribution for the scene,

(ii) Constructing probability models for multiple-sensor data conditioned on the scene, and

(iii) Conducting unified inference about the scene using the posterior distribution of the scene for given sensor data.

Probabilities of the respective sensors are obtained as the product of the current likelihood and prior probability of the estimate of the scene. The method assumes that there exists a space of all the possible estimates where every estimate has a corresponding posterior probability. In this case, the posterior probability of a given estimate is obtained by fusing the data from the acoustic sensor and the IR sensor resulting in the posterior distribution. Since the ultimate goal is to obtain the optimal estimate or an approximation to an optimal estimate, the actual problem involves obtaining an estimate with maximum posterior probability from the posterior distribution. This is done using search algorithms.

Tracking the moving targets by using image data involves processing images from a target of interest and producing at each time step, an estimate of the target's current position and velocity vectors. Uncertainties in the target motion and in the measured values, usually modeled as additive random noise, lead to corresponding uncertainties in the target state. In this paper, the CTA that uses probabilistic data association filter (PDAF)[3,4] for the image centroid tracking problem is presented.

2. TARGET LOCATION AND IDENTITY ESTIMATION

In this section, details of data synthesis, mathematical models for the IR and acoustic sensors, sensor likelihood estimation for IR and acoustic sensors, Bayesian fusion for computation of posterior distribution and the search algorithms used are presented. Figure 1 shows the information flow diagram for target location and its identity estimation using Bayesian fusion and search algorithms.

2.1 Data Synthesis

The data scenario forms the input to the sensor mathematical models for generating the sensor data. For simplicity the scene X_s is considered to be 2-D. The targets are assumed to be at fixed orientations and of the same size. The same techniques could be applied to targets with different orientations. In case of IR sensors, the input to the sensor is the radiations emitted from the targets present in the FOV. The intensity of the image pixel represents the radiation from the targets. For acoustic sensor the input is the acoustic signals from the targets. Since acoustic sensor can detect only the direction of targets, the synthesised data for a mathematical model is a vector consisting of directions only.

The input to the sensors is synthesised by using the concept of marked spatial point process (MSP). If S denotes the continuum space, then the realization of a random point process on S is a set of points having coordinates in S. In a marked point process[5,6] each point in the set of points is a realisation of point process that is associated with some auxiliary variable. The continuum space could be one-, two- (area) or three-dimensional (volume). When the space considered is a domain of area or volume, the marked point process becomes an MSP process. For scenario simulation, auxiliary variables associated with the points are considered as the locations and identities of the targets. Four typical two dimensional scenes with two targets generated by a MSP process are:

$$X_s = (X_{s1}, X_{s2}, X_{s3}, X_{s4})$$

DATA SYNTHESIS

SCENARIO SIMULATION USING MSP

MATHEMATICAL MODELS OF INFRARED AND ACOUSTIC

FUSION AND ESTIMATION   (INFRARED AND ACOUSTIC SENSORS DATA)

GENERATE INITIAL ESTIMATE USING MSP $I=1$

GENERATE NEXT ESTIMATE USING PREVIOUS ESTIMATE

$I>1$

OBTAIN FUSED POSTERIOR PROBABILITY OF THE ESTIMATE USING BAYESIAN FUSION SCHEME

I   : ITERATION NUMBER
N   : TOTAL ITERATIONS
MH : METROPOLIS HASTINGS
SA : SIMULATED ANNEALING
GG : GRADUAL GREEDY

POSTERIOR PROBABILITIES AT I-1 AND I

ACCEPT OR REJECT THE ESTIMATE BASED ON ACCEPTANCE CRITERIA OF MH/SA/GG $I>1$

$I=N$

FINAL ESTIMATE

Figure 1. Flow diagram of target location and its identity estimation using Bayesian fusion.

$$X_{s1} = \begin{bmatrix} 37 & 82 \\ 84 & 91 \\ 1 & 2 \end{bmatrix}$$

$$X_{s2} = \begin{bmatrix} 21 & 94 & 34 & 64 & 98 & 55 \\ 107 & 67 & 70 & 21 & 107 & 105 \\ 1 & 2 & 1 & 1 & 1 & 2 \end{bmatrix}$$

$$X_{s3} = \begin{bmatrix} 32 & 108 & 54 & 21 \\ 30 & 51 & 104 & 87 \\ 1 & 2 & 1 & 1 \end{bmatrix}$$

$$X_{s4} = \begin{bmatrix} 94 & 43 & 75 & 22 \\ 106 & 101 & 66 & 63 \\ 2 & 1 & 2 & 1 \end{bmatrix}$$

In these matrices, each column represents a target with the first 2 rows giving the *x, y* coordinates of the target location and the third row gives the identity of the target (1 or 2). It can be observed that in each case, total number of targets is varying (except the last two), also the positions and identities of the targets are random. The synthesised scenario data for the space X_s is shown in Fig. 2.

2.2 Mathematical Models for IR and Acoustic Sensors

The mathematical model for the IR sensor is: If Y_1 is an image as in Fig. 2, *q* denotes the locations of targets contained in the true scene, and *z* is a typical pixel in the image Y_1, then

$$Y_1(z) = P(z) + R(z) + D(z) + B(z) \qquad (1)$$

where, $P(z)$ is the Poisson mean at *z* which is obtained by the convolution of the ideal image I_o with the point spread function *h* of the IR camera:

$P(z)$= poisson [I(z)], where $I=I_o+w$ and $R(z)$ is readout noise at *z*; $D(z)$ accounts for a dead pixel at *z*; and $B(z)$ is black current at *z*.

For simplicity only the $P(z)$ contribution to Y_1 is considered in the simulations. Random noise with a known standard deviation (σ_1) is added to the generated image to account for the readout noise. Figure 3 shows the output of the IR sensor for a typical scenario from which it is clear that information on both target identity and location can be obtained from the IR sensor data.

Mathematical model for acoustic sensor is:

The acoustic signal received by a linear array of *m* equally spaced sensors at time *t* is

$$Y_2(t) = \sum_{i=1}^{p} A_i(t)d(\theta_i) + w(t) \qquad (2)$$
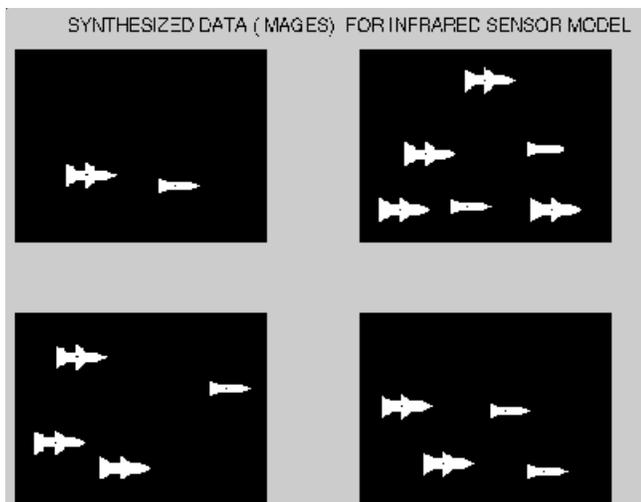
Figure 2. Synthesised images generated using the elements X_s.
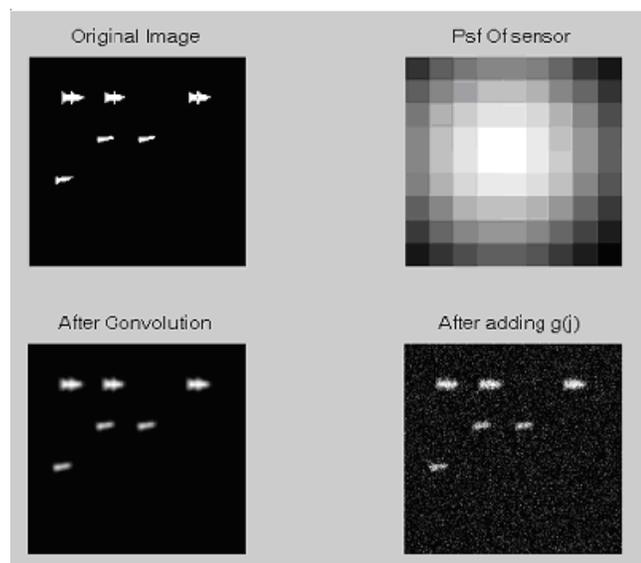
Figure 3. Generation of infrared image from the model X_s.

where p is number of target in the scene; $w(t)$ is the measurement noise at time t, A_i is the amplitude of the signal from i^{th} target; and $d(\theta_i)$ is the direction vector associated with i^{th} target and is given as

$$d(\theta_i) = [1, e^{-j\pi\cos(\theta_i)}, ..., e^{-j(m-1)\pi\cos(\theta_i)}]^T$$

Figure 4 shows the synthesised data for acoustic sensor model with the position of acoustic sensor at $a_x, a_y = (0,150)$. The simulated scene consists of three targets along the directions of 84°, 45° and

45°. The direction of targets is wrt acoustic sensor location on which the signals from the target shown with white lines strike. The response of acoustic sensor which is obtained by plotting amplitude for θ ranging from 1° to 90°, is also shown in the Figure. It is observed that the response of acoustic sensor shows peaks along the directions of the target, since the three targets are present along two directions 45° and 84°, two main peaks can be seen along those directions.

2.3 Sensor Likelihood Computation and Bayesian Fusion

Bayesian sensor fusion methodology fuses the IR sensor and acoustic sensor outputs by fusing their posterior probabilities. The posterior probabilities for the sensor outputs are their likelihood functions. The IR camera can detect all the aspects of a scene X and the data $Y_1(1), Y_1(2), ..., Y_1(j)$ are assumed to be conditionally independent for a given X. These assumptions lead to a Poisson likelihood function for infrared sensor data Y_1 for a given X.
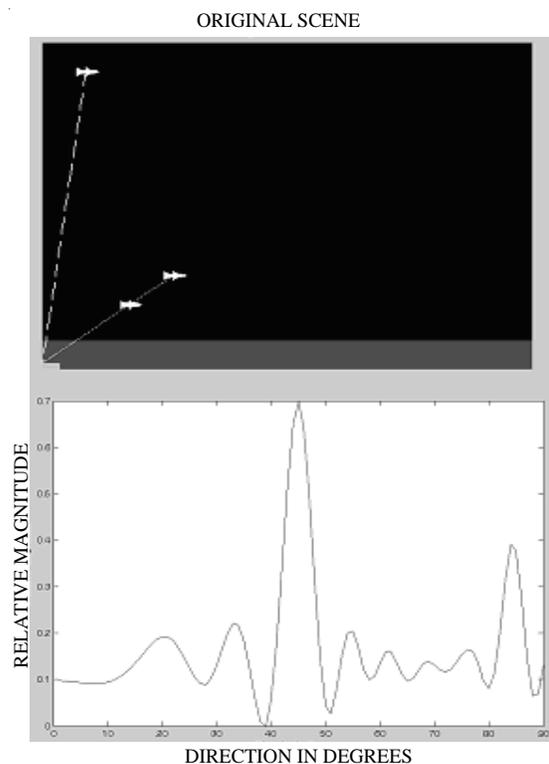
Figure 4. Simulated scene and corresponding response of acoustic sensor.

$$L_1(Y_1 | X) = \prod_{k=1}^{K} \frac{I(z_k)^{Y_1(z_k)} e^{-I(z_k)}}{Y_1(z_k)!} \quad (3)$$

The above equation is simplified using a Gaussian likelihood function[2]:

$$L_1(Y_1 | X) = \frac{1}{z} \exp\left(-\frac{1}{2\sigma_1^2} \|Y_1 - X\|_F^2\right) \quad (4)$$

where $\| \|_F$ is the Frobenius norm of a matrix, and σ_1^2 is the variance of Y_1 and Z is a normalising constant.

If the acoustic sensor output is represented by a data vector Y_2, the complex Gaussian likelihood functions of Y_2 for a given X is[2]:

$$L_2(Y_2 | X) = \frac{1}{z} \exp\left(-\frac{1}{2\sigma_2^2} \|Y_2 - X\|^2\right) \quad (5)$$

where $\| \|$ is norm of a matrix.

Applying Bayes' rule and assuming conditional independence of the sensor outputs Y_i's for a given X, the posterior probability is given by:

$$P(X | Y_1, Y_2) \propto L_1(Y_1 | X) L_2(Y_2 | X) P(X) \quad (6)$$

where $P(X | Y_1, Y_2) \equiv P'$ is posterior distribution and $P(X)$ is the prior probability distribution. Thus, the Bayesian fusion involves only multiplying the posterior probabilities of the respective sensors.

After obtaining the a posteriori probability distribution using Bayesian fusion, the target identity and location are obtained by searching for appropriate state X_s in the state space χ_s. In this paper, the performance of three search algorithms namely, metropolis hastings (MH)[1,7], simulated annealing (SA)[8] and gradual greedy (GG)[2] are evaluated for target location and identity estimation for the typical scene (Fig. 2).

3. CENTROID TRACKING ALGORITHM

This section presents the mathematical models for data generation for tracking, segmentation and centroid estimation and the tracking algorithm.

In general, typical characteristics of the target obtained by motion recognition or by object (pattern) recognition methods are used in associating images to the target being tracked. Motion recognition characteristics of a target are its location, velocity and acceleration (i.e. state vector) which could be generated using data from successive frames (inter-scan level). Object (pattern) recognition characteristics are its geometric structure (shape, size), energy level distribution (i.e. different gray level in the image) in one or more spectral bands which are obtained using image data at the intra scan level. The CTA combines both object and motion recognition methods for practical target tracking from imaging sensors. The CTA implementation involves the conversion of the data from the image obtained from the IR camera into a binary image by applying upper and lower threshold limits for the "target layers". The binary target image is then converted to clusters using nearest neighbor criterion. If the target size is known, the information is used to set limits for removing those clusters that differ sufficiently from the size of the target cluster to reduce computational complexity. The centroid of the clusters is then calculated and this information is used for tracking the target.

3.1 Synthetic Image Generation

The mathematical basis for generation of synthetic image[9] is briefly described below:

Consider two-dimensional array of pixels

$$m = m_\xi \text{ x } m_\eta \quad (7)$$

where each pixel is represented by a single index $i = 1, ..., m$ and the intensity I of pixel i is given by

$$I_i = s_i + n_i \quad (8)$$

where, s_i is the target intensity and n_i is the noise intensity in pixel i, which is assumed to be Gaussian with zero mean and covariance σ^2.
The total target-related intensity is given by:

$$s = \sum_{i=1}^{m} s_i \quad (9)$$

If the number of pixels covered by the target is denoted by m_s, then the average target intensity over its extent is given by:

$$\mu_s = \frac{s}{m_s} \tag{10}$$

The average pixel SNR (over the extent of the target) is

$$r' = \frac{\mu_s}{\sigma} \tag{11}$$

Using Eqns (7) - (11), the synthetic images in a frame can be generated by using the following inputs:

(i) Target pixel intensity (s_i): $N(\mu_t,\sigma_t^{\,2})$

(ii) Noise pixel intensity (s_i): $N(\mu_n,\sigma_n^{\,2})$

(iii) Target: Rectangle (base NX and height NY),

(iv) Position of the target in each scan: (x-position and y-position)

In order to simulate the motion of the target in the frame, kinematic models of target motion are used. Constant velocity kinematic model is used for generation of the data which determines the position of the target in each scan.

State Model:

$$X(k+1) = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} X(k) + \begin{bmatrix} \dfrac{T^2}{2} & 0 \\ T & 0 \\ 0 & \dfrac{T^2}{2} \\ 0 & T \end{bmatrix} w(k) \tag{12}$$

where state $X(k) = [x_p,x_v,y_p,y_v]$, T = sampling period, and $w(k)$ is the zero mean Gaussian noise with variance Q.

Measurement Model:

$$z(k+1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} X(k+1) + v(k+1) \tag{13}$$

where $v(k)$ is the centroid measurement noise that is zero mean Gaussian noise with covariance matrix:

$$R = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \tag{14}$$

Both process noise and centroid measurement noise are assumed to be uncorrelated.

3.2 Segmentation and Centroid Estimation Technique

Particle segmentation is used to separate the target (object of interest) from background, when target is not fully visible[3]. It is assumed that the pixel intensities are discretised into 256 gray levels. Particle segmentation is done in two steps:

(i) The gray level image is transformed into binary image using lower and upper threshold limits of the target. These thresholds of target are determined using the pixel intensity histograms from the target and its surroundings, and

(ii) The detected pixels are grouped into clusters with nearest neighbour technique. The gray image $\text{Im}(i,j)$ is converted into binary image with intensity $\beta(i,j)$ by a hard limit on the intensity:

$$\beta(i,j) = \begin{cases} 1 & I_L \leq \text{Im}(i,j) \leq I_U \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where I_L and I_U are the lower and upper threshold limits of the target intensity.

The detection probability of the pixel (i, j) can be defined as:

$$\begin{aligned} P\{\beta(i,j)=1\} &= p(i,j) \\ P\{\beta(i,j)=0\} &= 1 - p(i,j) \end{aligned} \tag{16}$$

where $p(\text{i},\text{j}) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_{I_L}^{I_U} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$, considering the gray image I (i,j) as having a Gaussian distribution with mean μ and variance σ^2.

The binary image is then grouped into clusters using the nearest neighbour technique. A pixel is considered as belonging to the cluster only if the distance between this pixel and at least one other pixel of the cluster is less than the proximity distance (d_p). The d_p is chosen as:

$$\sqrt{\frac{1}{p_t}} < d_p < \sqrt{\frac{1}{p_v}} \tag{17}$$

where p_t and p_v are detection probabilities of target and noise pixels respectively.

By choosing the proximity distance as in Eqn (17), fewer noise clusters are obtained. The d_p affects the size, shape and number of clusters obtained by clustering. The centroid of the cluster is determined using:

$$x_{n_c} = \frac{\displaystyle\sum_{k=1}^{N} x_{n_k} I_k}{\displaystyle\sum_{k=1}^{N} I_k} \tag{18}$$

where x_{n_k} is the nth co-ordinate of pixel k, N is the number of pixels in the cluster, k is the pixel index, and I_k is k^{th} pixel intensity.

3.3 Image Centroid Tracking Algorithm

Centroid tracking in the presence of clutter is achieved using the nearest neighbour Kalman filter (NNKF) and probabilistic data association filter (PDAF)[4]. In general gating and data association enable tracking in multi target scenario. Gating helps in deciding if an observation (which includes clutter, false alarms and electronic counter measures) is a probable candidate for track maintenance or track update. Data association is the step to associate the measurements to the targets with certainty when several targets are in the same neighbourhood. The NNKF or PDAF is necessary for the centroid tracking application because in the neighbourhood of the predicted location for the target centroid during tracking, several centroids could be found due to splitting of the target cluster or due to noise clusters.

3.4 Fusion of IR Sensor Data and Ground-based Radar Data

For robust tracking in the presence of measurement loss, the track from the imaging sensor is fused with the track from ground-based radar using state vector fusion. It improves the tracking performance and reduces the uncertainty in the target state estimation. Flow diagram for fusion of tracks from imaging sensor and ground-based radar is shown in Fig. 5. The tracks which are state vector estimates from the IR sensor (Track *i*) and ground based radar (Track *j*) and their covariance matrices at scan k are shown as[10]:

Track i : $\quad \hat{x}_i(k), \hat{P}_i(k)$

Track : $\quad \hat{x}_j(k), \hat{P}_j(k)$ $\tag{19}$

The fused state estimate is given by:

$$\hat{x}_c(k) = \hat{x}_i(k\,|\,k) + \hat{P}_i(k\,|\,k)\hat{P}_{ij}(k)^{-1}$$
$$\left[\hat{x}_j(k\,|\,k) - \hat{x}_i(k\,|\,k)\right] \tag{20}$$

The combined covariance matrix associated with the estimate of Eqn. (20) are given by

$$\hat{P}_c(k) = \hat{P}_i(k\,|\,k) - \hat{P}_i(k\,|\,k)\hat{P}_{ij}(k)^{-1}\,\hat{P}_i(k\,|\,k) \tag{21}$$

where $\hat{P}_{ij}$ is cross covariance between $\hat{x}_i(k\,|\,k)$ and $\hat{x}_j(k\,|\,k)$, and is given by

$$\hat{P}_{ij}(k) = \hat{P}_i(k\,|\,k) + \hat{P}_j(k\,|\,k) \tag{22}$$

4. RESULTS AND DISCUSSION

4.1 Performance Evaluation of the LIE Algorithm

Figure 6 shows the plot of posterior energy (PE) for MH, SA and GG algorithms for 1000 iterations. It can be seen that the PE ultimately decreases in all the three cases, which indicates that the algorithms tends towards the solution. The PE is computed using the following:

PE = –log (posterior probability *y*)

For comparison of the three algorithms, the following parameters are evaluated:

(a) Success rate (*sr*)

(b) *X* position error (*xpe*) and *Y* position error (*ype*)

(c) Id error (*ide*): Number of targets misclassified

(d) Target count error (*tce*)
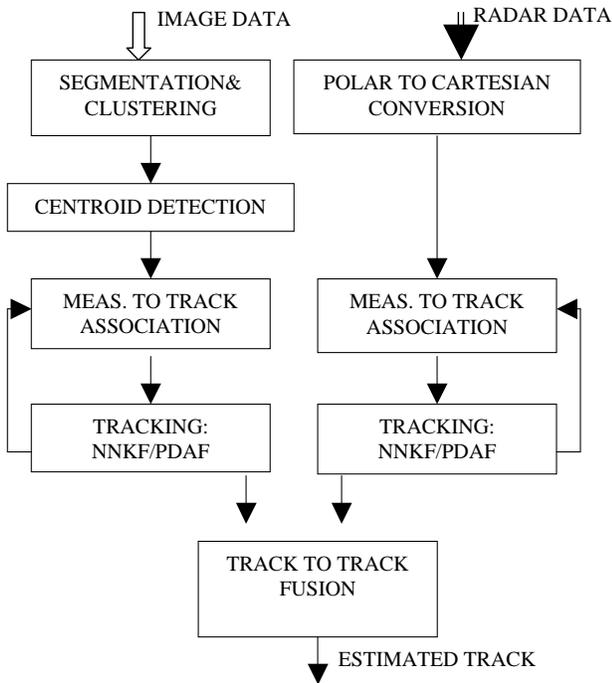
Figure 5. Fusion of data from imaging sensor and ground-based radar.

The performance of the algorithms is evaluated for a typical scenario consisting of two targets. Tables 1 to 2 give a comparison of the different parameters for the three algorithms with the average of 25 Monte Carlo simulations. It is clear that the performance of these search algorithms is almost identical. However, from Table 1 it is seen that in terms of sr, the GG algorithm seems somewhat better than other search algorithms.

4.1.1 Tracking Performance

The performance of CTA is evaluated in terms of the following[11]:

(i) The percentage fit error (PFE) (x or y)

$$PFEx = 100 * \frac{norm(x - \hat{x})}{norm(x)} \quad (23)$$

(ii) The root mean square position error

$$RMSPE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}{2}} \quad (24)$$

(iii) The root sum square position error

$$RSSPE = \sqrt{(x - \hat{x})^2 + (y - \hat{y})^2} \quad (25)$$

Similar formulae are applicable for velocity errors.

(iv) The state errors with $\pm 2\sqrt{P_{i,i}}$ bounds (26)

where P is the state error covariance matrix, where x and y are the measurements, $\hat{x}$ and $\hat{y}$ are the estimated target locations in x and y coordinates, respectively.
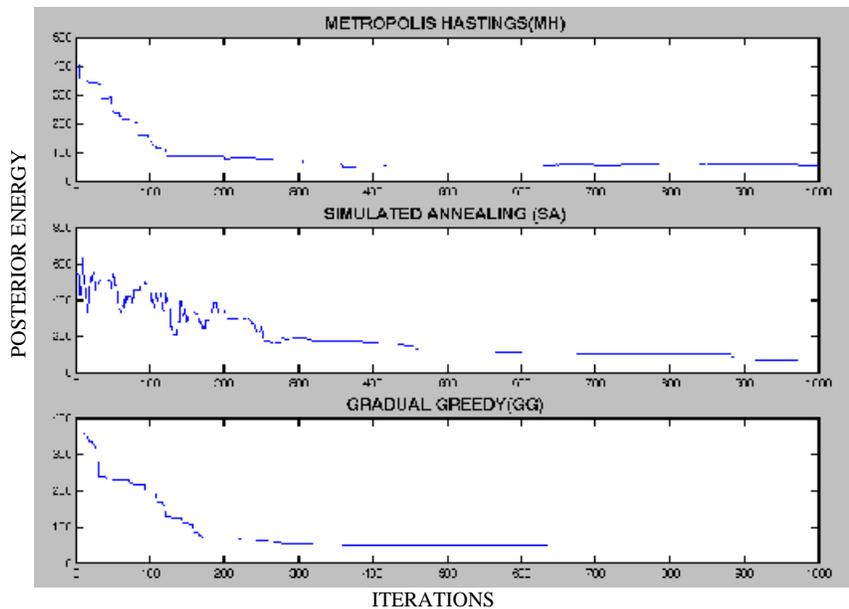
Figure 6. Posterior energy convergence for the MH, SA and GG algorithms.

4.1.2 Performance Evaluation of CTA

A 2-D of 64 x 64 pixels is considered for the image. A 2-D array of pixels, which is modelled as white Gaussian random field with a mean μ_t and variance σ_t^2 is used to generate a rectangular target of size (9 x 9). The image is converted into binary image using the upper (I_U=110) and lower (I_L=90) limits of a target layer [Eqn (15)], and then grouped into clusters by the nearest neighbor technique using the optimal proximity distance (d_p=2).

The initial state vectors of target 1 and target 2 in the image frame are:

Track 1: $\begin{bmatrix} 10 & 0.99 & 10 & 0.99 \end{bmatrix}^T$, and

Track 2: $\begin{bmatrix} 20 & 0 & 60 & -0.99 \end{bmatrix}^T$

The total number of scans is 50. The image frame rate is one frame/s. The target and noise parameters used in this simulation are as follows[3]:

Target pixel intensity = $N(100, 10^2)$, and

Noise pixel intensity = $N(50, 50^2)$.

The centroid of each cluster is calculated and used for state estimation in the measurement update part of the PDAF filters to track the target in clutter. The PDAF algorithm includes track initiation and track deletion features which are essential in multi target tracking in clutter[5]. Figure 7 shows the frame, which includes the estimated and true data of two targets in clutter. The frame shows the background clutter and the two synthetic target images at 50th scan. Figures 8(a) and 8(b) show estimated positions and velocities compared with the true values. It is observed that the estimated and true tracks are well matched. The state errors with their bounds, RSSPE and RSSVE for the two targets are shown in Figs 9(a) and 9(b). Percentage fit error in *x*- and *y*-positions and RMSPE for two targets are within the acceptable range as shown

Table 1. Performance of the algorithms for different inputs (from 25 Monte Carlo simulations)

Algorithm	Scenario I			Scenario II			Scenario III			Scenario IV		
	Su	Fa	Sr	Su	Fa	Sr	Su	Fa	Sr	Su	Fa	Sr
MH	24	1	96	25	0	100	25	0	100	6	19	24
SA	25	0	100	23	2	92	22	3	88	21	4	84
GG	24	1	96	25	0	100	25	0	100	25	0	100

Su: Success; Fa: Failure; Sr: Success rate

Table 2. Statistics for parameters *xpe, ype, ide, and tce*

Scenario	Algorithm	*xpe* mean ± std	*ype* mean ± std	*ide* mean ± std	*tce* mean ± std
	MH	0.4 ± 2	0.48 ± 2.4	0.04 ± 0.2	0 ± 0
	SA	0 ± 0	0 ± 0	0 ± 0	0 ± 0
I	GG	0.16 ± 0.8	0.8 ± 4	0.04 ± 0.2	0 ± 0
	MH	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	SA	1.16 ± 4.02	1.28 ± 4.44	0 ± 0	0 ± 0
II	GG	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	MH	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	SA	0 ± 0	0 ± 0	0 ± 0	0.12 ± 0.33
III	GG	0 ± 0	0 ± 0	0 ± 0	0 ± 0
	MH	0.88 ± 0.67	0.2 ± 0.5	0 ± 0	0 ± 0
IV	SA	0.24 ± 0.72	0.04 ± 0.2	0 ± 0	0 ± 0
	GG	0 ± 0	0 ± 0	0 ± 0	0 ± 0

in Table 3. State errors and root sum square errors in position and velocities respectively using 25 Monte Carlo simulations are shown in Fig. 10. The state errors are within the theoretical bounds and the root sum square errors are fraction of a pixel, which shows the robustness of the algorithm.

The true, estimated and fused trajectories are shown in Fig. 11 from which it is clear that the fused trajectory matches the true trajectory. A measurement loss in imaging sensor is simulated from 15 s to 25 s and in the ground-based radar from 30 s to 45 s. Track extrapolation has been
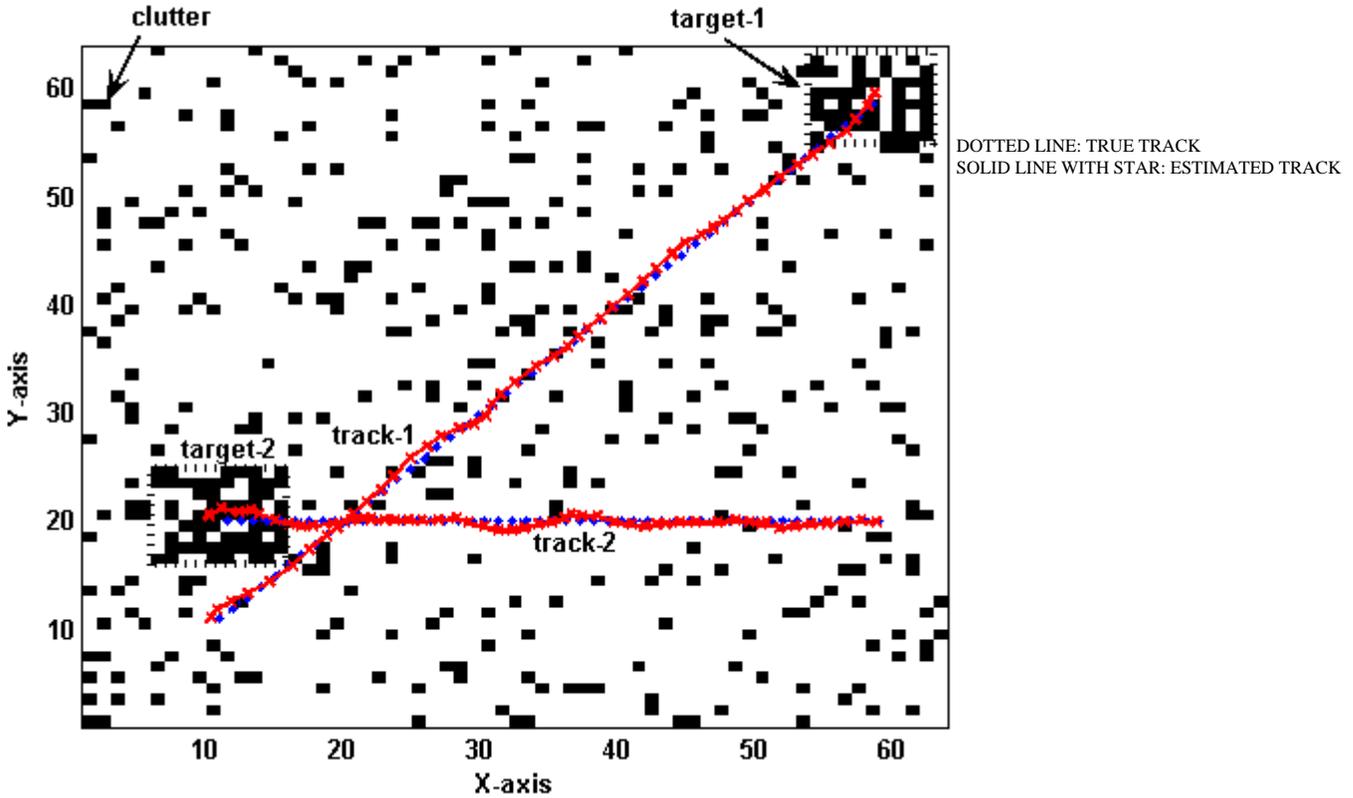
Figure 7. Tracking of two targets in presence of clutter.
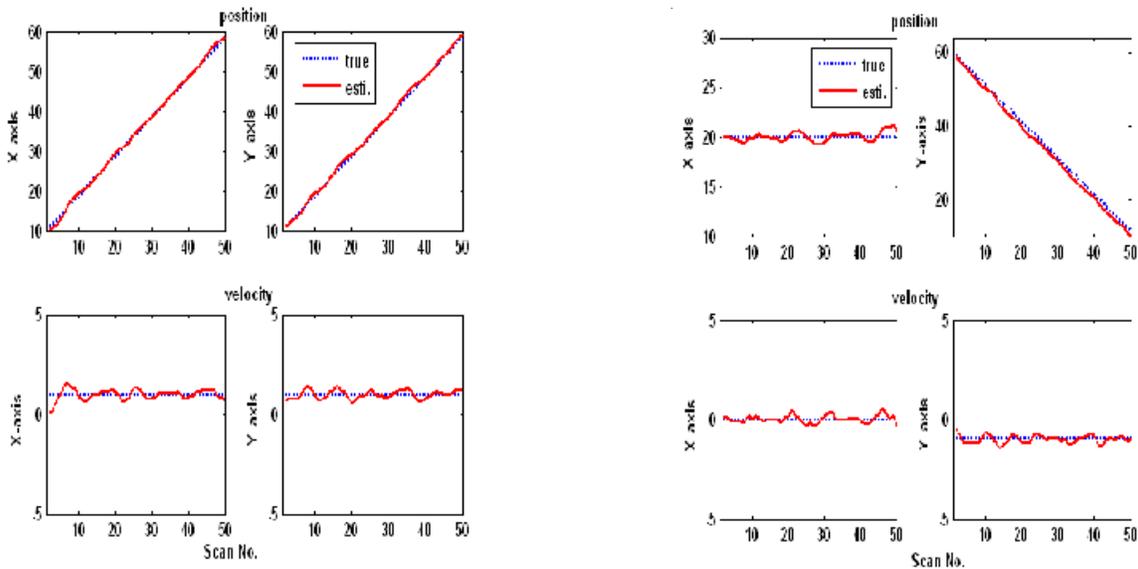
Figure 8. (a) True and estimated positions and velocities for: (a) track 1, and (b) track 2.

done during these periods. Track deviation can be observed in Fig. 11 during these durations. Percentage fit error in *x* and *y* directions and root mean square

Table 3. Percentage fit error and root mean square error with radar, CTA and fusion

$N_D = 50$ samples	Measure-ment loss	PFEx	PFEy	RMSPE	RMSVE
Radar	30 - 45 s	2.56	2.59	1.39	0.24
IR	15 - 25 s	4.29	2.86	1.98	0.37
Combined (fused)		1.28	1.17	0.65	0.41

Table 4. Percentage fit error and root mean square errors

	PFEx	PFEy	RMSPE	RMSVE
Track 1	1.34	1.41	0.52	0.25
Track 2	2.19	3.39	0.92	0.2

error in position and velocity for before and after fusion are shown in Table 4. From Fig. 12 and Table 4, it is observed that fusion of tracks gives better results when there is a measurement loss in either of the sensors thereby demonstrating the robustness and better accuracy achieved because of fusion.

(a) (b)
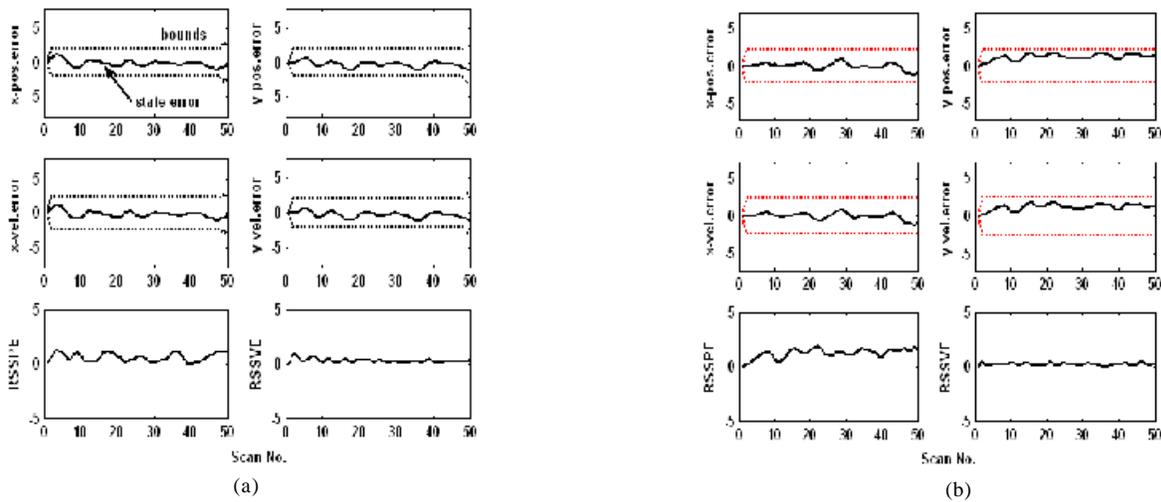
Figure 9. State errors in position and velocity, RSSPE and RSSVE for: (a) target 1, and (b) target 2.
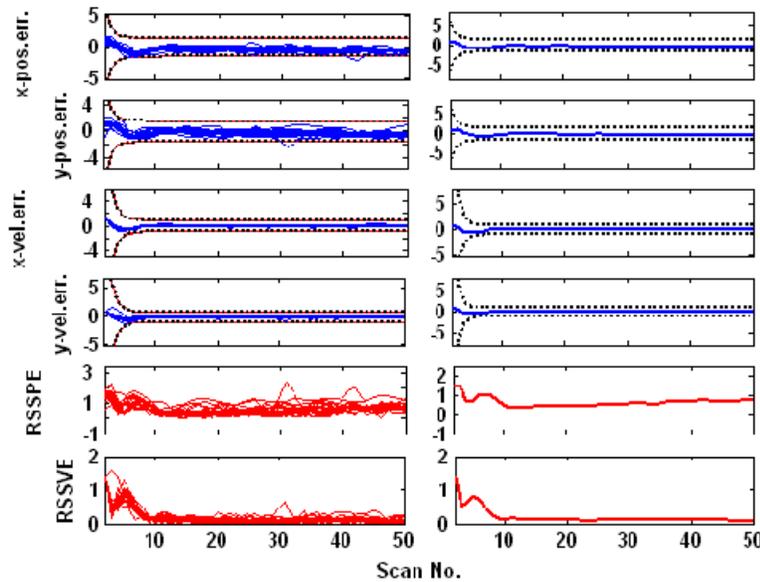
Figure 10. State errors, RSSPE, and RSSVE using 25 Monte Carlo simulations (column 1) and mean for the same (column 2).

(a)

(b)

(c)
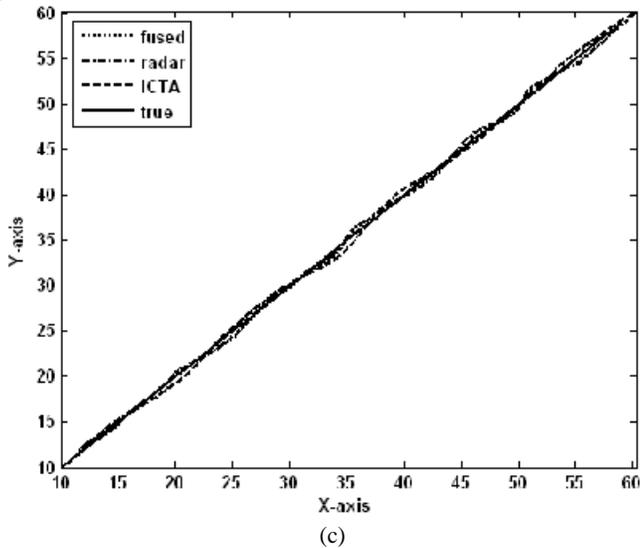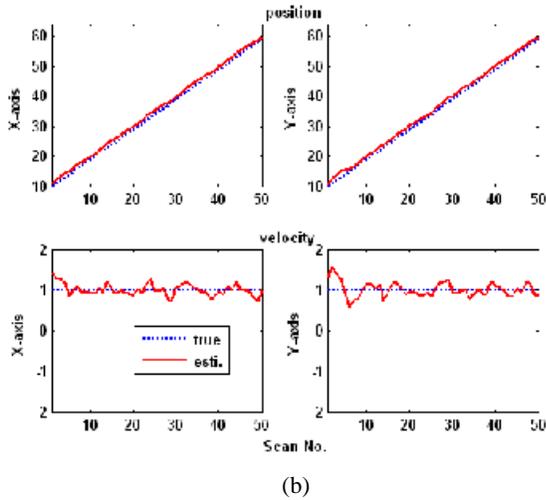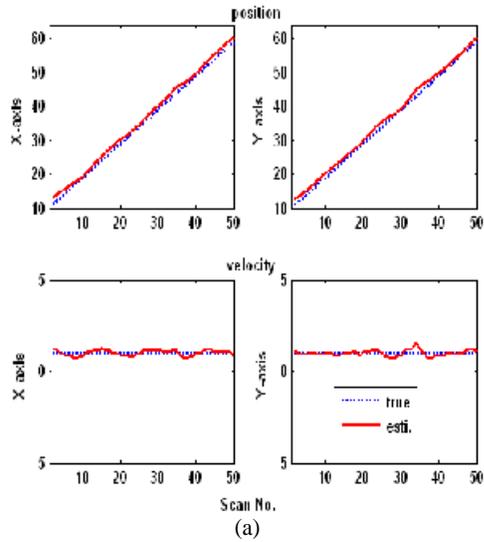
Figure 11. True and estimated states from: (a) imaging sensor (b) ground-based radar, and (c) individual and fused trajectories.
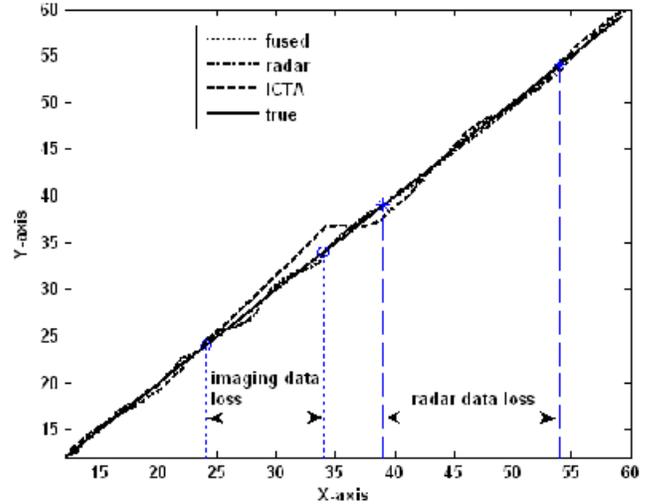
Figure 12. Individual and fused trajectories in presence of measurement loss.

5. CONCLUSIONS

Two important aspects involved in ATR, namely, estimation of identity and location of targets and tracking of the centroid of targets using IR sensor data are addressed in this paper. For the purpose of estimation of target location and identity, data from the IR sensor is fused with the data from an acoustic sensor using Bayesian fusion method. Marked spatial point process and sensor mathematical models (IR and acoustic) have been adopted to generate the scenarios. In Bayesian fusion method, the likelihood functions of the two sensors are used to obtain the posterior probabilities. From the posterior probability distribution, the target identity and location are established using search algorithms. On the average the performance of the search algorithms is almost similar. However, GG search algorithm is slightly better in terms of success rate.

The CTA has been implemented for accurate target tracking based on the data obtained from imaging sensors, when the target is not fully visible during tracking. Using segmentation technique the gray level image has been converted into binary image and reduced into clusters using nearest neighbour (NN) criteria. Tracking filters (PDAF) were employed for state estimation using centroid measurement of clusters. It is observed that the CTA performed very well even in the dense clutter produced by the segmentation and clustering stage. The state estimates

obtained using the CTA algorithm are fused with the state vector obtained by processing the ground based radar data which is also tracking the target to improve robustness of tracker. It is observed that fusion generates robust tracks even when there is data loss in one of the sensors.

REFERENCES

1. Bir Bhanu. Automatic target recognition: State of the art survey. *IEEE Trans. Aero. Elec. Syst.*, 1996, **22**(4), 1872-873.

2. Smith, Michale J. Bayesian sensor fusion a framework for using multi-model sensors to estimate target location and identities in a battlefield scene. Florida State University, 2003. PhD Dissertation.

3. Oron, E.; Kumar, A. & Shalom, Y. Bar. Precision tracking with segmentation for imaging sensor. *IEEE Trans. Aero. Elect. Syst.*, 1993, **29**, 977-87.

4. Shalom, Y. Bar & Fortmann, T.E. Tracking and data association. Academic Press, New York, 1988.

5. Descombes, Xavier & Zerubia, Josiane. Marked point process in image analysis. *IEEE Signal Process. Maga.*, 2002, 77-84.

6. Synder, D.L. Random point processes, Ed. 2. John Wiley & Sons, New York, 1995.

7. A. Lanterman, A.; Miller, M & Snyder, D. General Metropolis-Hastings jump diffusion for automatic target recgnition in infrared scenes. *Optical Engineering*, 1997, **36**(4), 1123-137.

8. Rich, Elaine & Knight, Kevin. Artificial intelligence, Ed. 2, Tata Mcgraw-Hill Publishing Co. Ltd, New Delhi, 1991.

9. Kumar, A.; Shalom, Y. Bar & Oron, E. Precision tracking based on segmentation with optimal layering for imaging sensor. *IEEE Trans. Anal. Mach. Intelli.*, 1995, **17**, 182-88.

10. Chang, K.C.; Saha, R.K., & Shalom, Y. Bar. On optimal track-to-track fusion. *IEEE Trans. Aero. Elec. Syst.*, 1997, **33**, 1276-278.

11. Naidu, V.P.S.; G, Girija & Raol, J.R. Evaluation of data association and fusion algorithms for tracking in the presence of measurement loss. AIAA Guidance, Navigation, and Control Conference & Exhibit, Austin, USA, August 2003. Paper No. AIAA-2003-5733.

Contributors

Mr V.P.S. Naidu obtained his ME (Medical Electronics) from the Anna University, Chennai, in 1997. He is working at National Aerospace Laboratories (NAL), Bangalore, as Scientist since December 2001. His area of interest are: multi-sensor data fusion for target tracking, image registration, etc.

Dr (Ms) Girija G. obtained her PhD from the Bangalore University in 1996. She is presently, working as a Scientist at the NAL, Bangalore. Her areas of research are: Modelling, parameter estimation of aerospace vehicles, and multi-sensor data fusion. She is a recipient of *NAL Foundation Day Award for Research* (1995). She has published about 50 research papers.
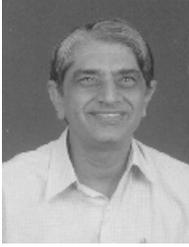
Dr J.R. Raol obtained BE and ME both from the MS University, Baroda, in 1971 and 1974, respectively, and PhD from the McMaster University, Canada in 1986. He is Head of the Flight Mechanics and Control Division of NAL. His current activities include modelling, identification, multi-sensor data fusion, fuzzy systems, genetic algorithm, and neural networks. He has 100 publications to his credit.