Information Sciences 278 (2014) 641–652

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Data fusion of radar and image measurements for multi-object tracking via Kalman filtering

Du Yong Kim [a], Moongu Jeon [b,*]

[a] School of Electrical, Electronic, and Computer Engineering, University of Western Australia, Australia
[b] School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, South Korea

ARTICLE INFO

Article history:
Received 7 October 2012
Received in revised form 29 October 2013
Accepted 13 March 2014
Available online 26 March 2014

Keywords:
Kalman filter
Data fusion
Multi-object tracking

ABSTRACT

Data fusion is an important issue for object tracking in autonomous systems such as robotics and surveillance. In this paper, we present a multiple-object tracking system whose design is based on multiple Kalman filters dealing with observations from two different kinds of physical sensors. Hardware integration which combines a cheap radar module and a CCD camera has been developed and data fusion method has been proposed to process measurements from those modules for multi-object tracking. Due to the limited resolution of bearing angle measurements of the cheap radar module, CCD measurements are used to compensate for the low angle resolution. Conversely, the radar module provides radial distance information which cannot be measured easily by the CCD camera. The proposed data fusion enables the tracker to efficiently utilize the radial measurements of objects from the cheap radar module and 2D location measurements of objects in image space of the CCD camera. To achieve the multi-object tracking we combine the proposed data fusion method with the integrated probability data association (IPDA) technique underlying the multiple-Kalman filter framework. The proposed complementary system based on the radar and CCD camera is experimentally evaluated through a multi-person tracking scenario. The experimental results demonstrate that the implemented system with fused observations considerably enhances tracking performance over a single sensor system.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Since the seminal paper of Kalman in 1960 [16], the Kalman filter has been the workhorse of various disciplines due to its properties of optimality in linear Gaussian systems and easy implementation for the real-world estimation problem. The literature shows that further research has been conducted to relax strict assumptions of the Kalman filter (e.g., prior knowledge of the system, Gaussian noises) that inhibit its applicability to real-world applications. Variants of the Kalman filter were rigorously investigated to overcome the practical issues; however, there are still several remaining issues for efficient implementation and development. In early days, Kalman filtering was usually applied to the aerospace science [25], navigation systems [29,33,36] and the mechanical system control [14]. Recent advances in theory and hardware extend its applicability to more diverse areas such as the process control in chemical engineering [15], bio-medical application [4,20], mobile robotics [6,7,17,18], image processing [10] and computer vision [9,12].

* Corresponding author.
E-mail addresses: duyongkim@gmail.com (D.Y. Kim), mgjeon@gist.ac.kr (M. Jeon).

http://dx.doi.org/10.1016/j.ins.2014.03.080
0020-0255/© 2014 Elsevier Inc. All rights reserved.

Among them, we focus on the very intuitive application of Kalman filtering, e.g., the object tracking, which has been implemented in many application areas. In this work, regarding two specific physical sensors we limit our discussion of related works to vehicle applications [19] and surveillance [34].

Conventional object tracking in the aerospace science utilizes the large-size radar system that returns point-wise measurements. However, recent electronics technology makes the radar systems small and cheap, enabling them to be used in commercial vehicle navigation systems [19,30]. In those systems, the radar measures relatively short range that can be utilized for pedestrians and adjacent vehicles detection to avoid collision. One shortcoming of this radar system is the low resolution in bearing angle information due to the wide beam width. In [30], authors proposed a combined system of a millimeter-wave radar and a CCD camera with a calibration method via homography. However, since their system simply combines detection data obtained from two sensors using the calibrated homography without considering noise model at all, it cannot track objects successfully when noises exist. In addition, spurious measurements from false targets are not considered so that many false tracks can be generated from background reflections.

There have been several studies about the tracking based on the single camera in surveillance which uses the special configuration of the camera to explain the relationship between the object size and its radial distance [34]. The multiple-camera system is also developed with overlapping or non-overlapping views [21] to reconstruct the 3D trajectories of objects. One of main tasks of the multiple-camera based tracking in surveillance system is the calibration of cameras to reconstruct 3D position of the object from 2D image planes [3]. Recent works tried to calculate the normal and inclination vectors to reconstruct 3D position of pedestrians from calibrated multiple cameras; however, the critical restriction of these system is that it can be used only for the fixed location set-up. Therefore, the multiple-camera system is seldom utilized for unconstrained applications (i.e., autonomous systems).

The main purpose of our research in this paper is to design a robust and cost-effective multi-object tracking system for autonomous agents which can be deployed in unconstrained environments. To this end, by adopting the sensor geometry [30] to combine a visual sensor with radar, we propose to employ Kalman filtering and data fusion technique underlying a Gaussian mixture form. As mentioned earlier, main differences between [30] and our work is that detection procedure of [30] is not able to circumvent sensor noises and false positive detections, whereas the proposed multi-object tracking system has such functionality that it can successfully track multi-object without failure.

Since a linear Gaussian state-space model (2) is considered in this paper, the Kalman filter is naturally used to guarantee the best linear minimum mean square error performance [1,16]; resulting in reduction of sensor noises. Additionally, we utilize data fusion technique which is embedded in the Kalman filtering framework to reduce the uncertainty originated from different sensor models and different types of sensor noises that are not taken into account in [30]. In tracking problems with multisensory observations, data fusion is important because different types of uncertainties are originated from individual sensors so that they affect the tracking performance differently. Here, we mean data fusion as the mathematical methodology to combine data collected from different sources with uncertainty models in order to provide more reliable information, which will be discussed in detail in Section 3.2. In addition, false positive detections (e.g., non-target originated measurements) are avoided substantially by incorporating the IPDA technique [23].

In the theoretical aspect, we extend the mixture of Kalman filter algorithm to use multisensory measurements for multi-object tracking that is implemented in the developed hardware module. A new tracker is called mixture of fusion Kalman filter because it fuses two independent observations from two physical sensors (i.e., radar, optical sensor) to construct the complementary system and tracks multiple-object under the Gaussian mixture structure.

The remainder of this paper is organized as follows. In Section 2, the Bayesian filtering problem is introduced and Kalman filtering equations are given for its solution. For the developed measurement system, radar and CCD sensors are also explained in Section 2. In Section 3, the proposed algorithm implemented in the system is subsequently explained in detail with the homography learning, data fusion technique, and multi-object tracking. Then, the proposed system is evaluated in real-world implementation of multi-person tracking in Section 4. Finally, conclusions are made in Section 5.

2. Kalman filter and developed measurement system

2.1. Dynamic system and Kalman filter

The fundamental formulation of the Kalman filter would be understood as a Bayesian filtering framework. Here, we introduce the basic Bayesian filtering framework and define the problem to be solved. Let $x_t \in \Re^{n_x}$ denote the state of object (e.g., position and velocity) at time t and the given sensor collects the observation $y_t \in \Re^{n_y}$ of the state x_t. In sequential Bayesian filtering, we want to construct $p(x_t|y_t)$, the probability density function (pdf) of x_t given y_t using two key equations.

$$\begin{aligned} &\text{Prediction}: \ p(x_{t+1}|y_{1:t}) = \int p(x_{t+1}|x_t)p(x_t|y_{1:t})dx_t \\ &\text{Update}: \ p(x_{t+1}|y_{1:t+1}) = \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|y_{1:t})}{\int p(y_{t+1}|x_{t+1})p(x_{t+1}|y_{1:t})dx_{t+1}} \end{aligned} \tag{1}$$

In practice, however, it is impossible to analytically obtain the complete pdf because of intractable integrals and arbitrary pdf forms. In 1960, Kalman proposed the optimal estimator using the minimum mean square error criterion under the linear

Gaussian assumption [16] that can be understood as the solution of the linear Gaussian Bayesian filtering problem. For the brief review, we introduce the governing equations of the Kalman filter.

Consider a linear dynamic Gaussian system as follows.

$$\begin{aligned} x_t &= F_t x_{t-1} + \xi_t, \\ y_t &= C_t x_t + \varsigma_t, \end{aligned} \tag{2}$$

where $F_t \in \Re^{n_x \times n_x}$ is the state transition matrix; $C_t \in \Re^{n_y \times n_x}$ is the observation matrix of the linear dynamic system; ξ_t and ς_t are system and observation zero mean noise vectors whose pdf are Gaussian, i.e., $\mathbb{N}(\xi_t; 0, Q_t)$ and $\mathbb{N}(\varsigma_t; 0, R_t)$, where Q_t and R_t are noise covariances, respectively. The initial state x_0 is also Gaussian i.e., $\mathbb{N}(x_0; \bar{x}_0, P_0)$ and mutually uncorrelated with noise sequences. Also, system and observation noises are mutually uncorrelated each other. The notation for Gaussian pdf of the random process x_t is given as $\mathbb{N}(x_t; \bar{x}_t, P_t)$ where $\bar{x}_t$ is mean vector and P_t is its covariance matrix. Then, following equations explicitly describe the recursion for Kalman estimate $\hat{x}_t$ and its covariance P_t as follows.

$$\begin{aligned} \text{Prediction}: \hat{x}_t^- &= F_t \hat{x}_t \\ P_t^- &= F_t P_t F_t^T + Q_t \end{aligned} \tag{3}$$

$$\begin{aligned} \text{Update}: \tilde{y}_t &= y_t - C_t \hat{x}_t^- \\ S_t &= C_t P_t^- C_t^T + R_t \\ K_t &= P_t^- C_t^T S_t^{-1} \\ \hat{x}_t &= \hat{x}_t^- + K_t \tilde{y}_t \\ P_t &= (I_{n_x} - K_t C_t) P_t^-, \end{aligned} \tag{4}$$

where P_t^- is the predicted error covariance, K_t is the Kalman gain and I_{n_x} is the $n_x \times n_x$ identity matrix. Note that this formulation is for the single object tracking formulation. In Section 3, we introduce the information form of Kalman filter which is mathematically equivalent to the standard Kalman filter (3) and (4) in order to propose data fusion and extend it to the multi-object, multi-sensor problem. System parameters used in the experiment will be described in detail in Section 4.

2.2. Radar measurement system

In this work, we developed a physical system by combining a radar and a CCD camera to get observations and detect moving objects more correctly and economically. Fig. 1 shows the radar module of the system that mainly measures the radial distance of objects. SiversIMA RS 3400 module is a frequency modulation continuous wave (FMCW) radar module [28]. Its center and sweep frequencies are 24.7 GHz, 1.5 GHz, respectively. In the experiments, we set the start-frequency of sweeping to 24 GHz and the stop frequency to 25.5 GHz. Because the Sivers IMA RS 3400 module is step frequency continuous wave (SFCW) radar, frequency separation between two contiguous frequency points of the sweeping is set to 1.6 MHz.

In the FMCW radar, the sensor output corresponds to the cosine of the phase difference between the echo signal and the radiated signal expressed as

$$s = \cos(\Phi), \tag{5}$$

where s is the output signal from the sensor and $\Phi = 2\pi 2d/\lambda$ is the phase difference between the echo RF signal and the radiated signal. Here, $2d$ is the round trip distance to the reflecting object and λ is the electrical wavelength of the RF signal. Then, Φ can be represented by the frequency of the RF signal, i.e., f_{RF} as

$$\Phi(f_{RF}) = 2\pi \frac{2d}{c} f_{RF}, \tag{6}$$

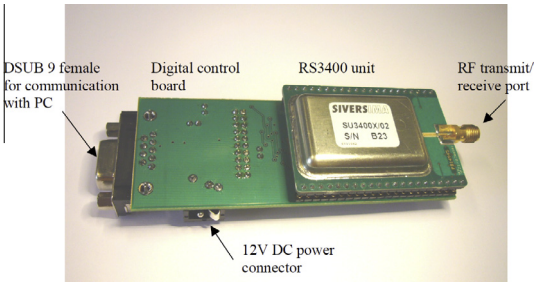


Fig. 1. The SiversIMA RS 3400 evaluation system.

where $\lambda = c/f_{RF}$, and c is the speed of light. From the FMCW property, a small value of d will create a slowly varying detection signal and a relatively long distant echo will return a quickly varying detection signal. We linearly increase the frequency from 24 GHz to 25.5 GHz for the bandwidth BW to be equal to 1500 MHz. Then the expression of the output signal can be represented in more specific form as

$$s(n) = \cos\left(\Phi_0 + 2\pi\frac{2d}{c}\frac{n}{N}BW\right), \tag{7}$$

where n is the frequency point index, $n = 0, 1, \ldots, N-1$ and N is the number of frequency points for the observation sequence, and Φ_0 is the phase value at the starting RF signal frequency of the sweeping. Then, we extract the distance information, i.e., the radial distance of the object from the simple fast Fourier transform (FFT) of the signal as

$$\Sigma(f) = fft(s(n)) = \frac{1}{2}\left(\Delta\left(f - \frac{2d * BW}{c}\right) + \Delta\left(f + \frac{2d * BW}{c}\right)\right), \tag{8}$$

where $f = 0, 1, \ldots, N-1$, denotes the normalized index in the transformed domain (distance or frequency domain) and $\Delta(\cdot)$ is Dirac delta function. By excluding negative frequency information, we finally detect range information as $d = fc/(2BW)$. Note that the range resolution is dependent upon the signal bandwidth only, not specific frequency. Based on specifications, the rage resolution of the system is about 0.1 m.

2.3. CCD measurement system

In the system we developed the radar module that is mainly used to collect the range information while the bearing information is not mainly considered because the used cheap radar sensor does not have good resolution in bearing angle. Thus, when the reflected signal from the object would be contiguous within certain bearing angle range, the radar has ambiguous measurements. In contrast, the CCD sensor can resolve close targets from the image observation. To compensate for the low resolution in bearing angle of the radar, we utilize the CCD camera to provide the x-, and z-position on the image plane. Here, we use the CCD with VGA sensor (640×480) pixel supporting 30 fps. On the image plane, we represent the position of the object as a bounding box and the center position of the bounding box is regarded as the point-wise measurement of the object.

To detect the object as a bounding box from the given CCD camera image, first we utilize the bearing angle measurement of the radar. When the object is roughly detected from the radar measurements, a trained blob detector [5] is employed to obtain the bounding box as the representation of the object. From the bounding box detection in the image, we can resolve the ambiguity in bearing angle measurement of the radar. In this case, however, two sensors should be calibrated to be used complementarily because they are not in the same geometric space.

3. Proposed algorithm

3.1. Homography learning

Motivated from [30], we developed a multiple-object tracking system based on the combined observation (radar and CCD camera as shown in Fig. 2). We adopt the calibration method proposed in [30] to acquire the homography between the radar observation space and the CCD camera image space as described in Fig. 3.

Denote the radar and the camera coordinates by (x_r, y_r, z_r) and (x_c, y_c, z_c), respectively, and the coordinate in the image plane by (u, v). If we fix the radar plane in $y_r = 0$, then the relation between the radar coordinate and the image coordinate can be represented as

Fig. 2. Overall system and set up for outdoor experiments.

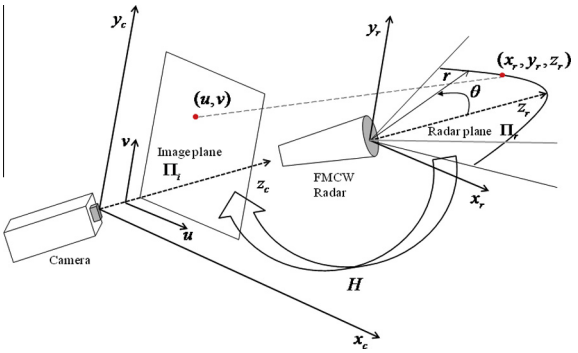


Fig. 3. Sensor geometry.

$$w\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} x_r \\ z_r \\ 1 \end{bmatrix}, \tag{9}$$

where $H = [h_{i,j}]_{i,j=1,2,3}$ is the 3×3 homography which represents the coordinate relation between the radar plane and the CCD image plane. Here, we assume the affine homography which means $h_{31} = h_{32} = 0$, $h_{33} = 1$. Then, $w = 1$ and two linear algebraic equations are obtained as

$$\begin{aligned} u &= h_{11}x_r + h_{12}z_r + h_{13}, \\ v &= h_{21}x_r + h_{22}z_r + h_{23}. \end{aligned} \tag{10}$$

After some algebra, we get the following linear system

$$a_x^T h = 0, \quad a_z^T h = 0, \tag{11}$$

$$\text{where } h = (h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, 0, 0, 1)^T, \quad a_x = (-x_r, -z_r, -1, 0, 0, 0, ux_r, vz_r, v)^T a_z = (0, 0, 0, -x_r, -y_r, vx_r, vz_r, v)^T. \tag{12}$$

This problem is known as the homogeneous linear least square problem which can be solved using Singular Value Decomposition (SVD).

Once we obtain the homography matrix as explained in (9)–(12), we need to consider two kinds of uncertainties: (1) sensor measurement error (i.e., signal noise), and (2) measurement conversion error (from polar to Cartesian). If these two uncertainties are critical, the estimated homography matrix can be unreliable, which leads to the tracking failure in the worst case. Beyond the work [30] that only learns homography, the novelty of the proposed algorithm is to consider a mathematical model of the uncertainty (e.g., sensor noises) as given in (13) and reduce them by using Kalman filtering and two sources of measurements will be fused using the data fusion technique that will be explained in Section 3.2. In the experiments, we will show that if the variance of the sensor noise for each sensor R_t^i is appropriately chosen, reliable tracking performance can be achieved. In a particular case, when the parameter R_t^i is not consistent, one can use a self-tuning technique to tune the parameters online [31] which is not considered in this paper.

After calibrating two sensors via the homography, collected observations from the radar and the CCD camera are converted onto the common plane from the learned sensor geometry. The resultant mathematical model of two sensors are represented as

$$y_t^i = C_t^i x_t + \varsigma_t^i, \quad i = 1, 2, \tag{13}$$

where $i = 1$ denotes the radar sensor and $i = 2$ does the CCD camera. The observation matrices for two sensors are given as $C_t^1 = [1\ 0\ 1]$ and $C_t^2 = [1\ 0\ 1]$, respectively. Note that $C_t^1 = [1\ 0\ 1]$ is the observation matrix for the radar measurement converted from the polar coordinate into the Cartesian coordinate plane.

In our system, y-coordinate representing the vertical sensor position is fixed, and measurement noises are considered as additive terms which represent aforementioned measurement uncertainties that contains (1) signal noise and (2) measurement conversion error. The fixed y-coordinate location is assumed to be high enough not to have severe ground clutters. But if the sparse ground clutters are dominant, they can be effectively reduced using the PDA based clutter rejection algorithm that will be detailed in Section 3.3. However, if the developed sensor system (i.e., radar + CCD camera) is moving in a y-coordinate direction there must be an additional block to estimate the attitude of the sensor system and this effect should be taken into account to calculate the homography. This problem known as the attitude estimation can be solved by employing another Kalman filter that estimates the attitude of sensor system based on the attitude measurements from the internal

measurement unite (IMU). This problem is well discussed in the previous research [32,33] and thus, remained as a straightforward extension.

3.2. Data fusion

As mentioned above, the object position in x–z plane is measured by two sensors. After calibration, we obtain two sets of measurements on the common plane. Considering the sensor resolution by modeling the sensor noise and errors in homography learning, we propose data fusion based on Kalman filtering technique to reduce uncertainties and obtain more accurate state estimates.

The main purpose of the data fusion technique is to combine measurements from multiple sensors that monitor common objects considering the uncertainty of individual sensor. It is a rule of thumb that collecting as many data as possible gives more reliable performance. However, just collecting data from different sources without efficient fusion rule can seriously degrade the quality of an estimate even with large amount of data [11,26,29,31,36].

Therefore, we propose to employ one of well-known data fusion methods called "information fusion" that is based on the information form of Kalman filter. The reason we chose information fusion method is that it is simple in use and the optimality [26]. In what follows, we first explain the centralized information fusion and then, the decentralized information fusion is subsequently provided. In addition, the functional equivalence between two data fusion schemes is explained with a simple proof.

In the centralized scheme, it is straightforward to consider an augmented measurement system by stacking individual measurement from all the sensor nodes as follows.

$$Y_t = C_t x_t + \varsigma_t, \tag{14}$$

where augmented measurement Y_t, measurement matrix C_t, and noise $\varsigma_t \sim \mathbb{N}(\varsigma_t; 0, R_t)$ are specified in (15).

$$\begin{aligned} &C_t = \left[\left(C_t^1\right)^T \ \left(C_t^2\right)^T\right]^T, \quad \varsigma_t = \left[\left(\varsigma_t^1\right)^T \ \left(\varsigma_t^2\right)^T\right]^T, \\ &R_t = diag(R_t^1, R_t^2), \quad Y_t = \left[\left(y_t^1\right)^T \ \left(y_t^2\right)^T\right]^T. \end{aligned} \tag{15}$$

Then, the information form of the Kalman filter provides the natural centralized data fusion estimate given by (16) and (17).

Time Update;

$$\begin{aligned} &\hat{x}_t^- = F_t \hat{x}_t, \\ &P_t = F_t M_t F_t^T + Q_t, \end{aligned} \tag{16}$$

Observation Update;

$$\begin{aligned} &D_t = C_t^T R_t^{-1} C_t, \\ &B_t = C_t^T R_t^{-1} Y_t, \\ &M_t = \left(P_t^{-1} + D_t\right)^{-1}, \\ &\hat{x}_t = \hat{x}_t^- + M_t\left[B_t - D_t \hat{x}_t^-\right], \end{aligned} \tag{17}$$

where D_t and B_t represent the contribution terms of the state and information matrix, respectively.

From the parallelization of the contribution terms, the mathematically equivalent fusion filter can be obtained with a summation of individual terms as (18) and (19),

$$D_t = C_t^T R_t^{-1} C_t = \sum_{i=1}^{2} \left(C_t^i\right)^T \left(R_t^i\right)^{-1} C_t^i, \tag{18}$$

$$B_t = C_t^T R_t^{-1} Y_t = \sum_{i=1}^{2} \left(C_t^i\right)^T \left(R_t^i\right)^{-1} y_t^i. \tag{19}$$

From the mathematical equivalence, the optimality of the decentralized fusion algorithm (16)–(19) is guaranteed [8].

Another data fusion method called decentralized fusion can be described by the weighted sum observation form as follows.

$$C_t = \left[\sum_{i=1}^{2} \left(R_t^i\right)^{-1}\right]^{-1} \sum_{i=1}^{2} \left(R_t^i\right)^{-1} C_t^i, \quad R_t = \left[\sum_{i=1}^{2} \left(R_t^i\right)^{-1}\right]^{-1}, \quad Y_t = \left[\sum_{i=1}^{2} \left(R_t^i\right)^{-1}\right]^{-1} \sum_{i=1}^{2} \left(R_t^i\right)^{-1} C_t^i. \tag{20}$$

In the proposed algorithm, the information fusion filter (16)–(19) is used as a basic framework and the decentralized data fusion form (20) has been applied.

Note that if observation matrices of sensors are identical, then, the centralized fusion scheme and the decentralized fusion scheme are functionally equivalent as shown in the following simple proof.

Proof. (18) and (19) represent the contribution term calculation in the information Kalman filtering. This implies that the augmented measurement form of (15) can be transformed to the summation of contributions from each sensor. Thus, if (18) and (19) with (20) are functionally equivalent to (18) and (19) with (15), the centralized fusion and decentralized fusion are equivalent. This reasoning is represented by the following equality and it holds if observation matrices are identical, i.e., $C_t^1 = C_t^2$. □

$$\sum_{i=1}^{2}\left(C_t^i\right)^T\left(R_t^i\right)^{-1}C_t^i = \left[\left(\sum_{i=1}^{2}\left(R_t^i\right)^{-1}\right)^{-1}\sum_{i=1}^{2}\left(R_t^i\right)^{-1}C_t^i\right]^T\sum_{i=1}^{2}\left(R_t^i\right)^{-1}C_t^i,$$

$$\sum_{i=1}^{2}\left(C_t^i\right)^T\left(R_t^i\right)^{-1}y_t^i = \left[\left(\sum_{i=1}^{2}\left(R_t^i\right)^{-1}\right)^{-1}\sum_{i=1}^{2}\left(R_t^i\right)^{-1}C_t^i\right]^T\sum_{i=1}^{2}\left(R_t^i\right)^{-1}y_t^i,$$

In summary, data fusion procedure in this paper can be described in following steps.

(1) Collect the raw measurements from two physical sensors (radar and CCD camera).
(2) Convert the raw measurements onto the x–y–z coordinates, respectively.
(3) Learn the homography between the radar and CCD camera coordinate.
(4) Perform the decentralized data fusion using (20).
(5) Calculate the contribution terms of parallelized information fusion algorithm (18) and (19).
(6) Calculate the estimate and covariance using (16) and (17) with contribution terms.

3.3. Multiple-object tracking using IPDA

The main task of the proposed tracking system is to detect the positions of multiple objects and track their trajectories with the multi-sensory measurements. Intuitively, one may think that the remaining task is simply assigning the information fusion filters (16)–(20) to individual tracks. However, it is not that straightforward in practice because we have to deal with false positive detections (i.e., observations originated from clutters) and track management in a principled manner.

In the proposed work, we adopt the multi-object integrated probabilistic data association (MIPDA) to resolve two aforementioned problems in the multiple-object tracking [22]. This is the multiple-object extension of IPDA [23] which is known as one of most efficient data association algorithms under clutter environments. And the automatic management of tracks (object appearance and disappearance), is handled using the target existence probability.

To make the paper self-contained, we summarize MIPDA algorithm for our system. First, we introduce notations for track management and data association for each track τ as follows.

- χ_t^τ the discrete event of object existence at time t where $\chi_t^\tau = 1$ for the target existence and nonexistence for $\chi_t^\tau = 2$, respectively.
- $\theta_{t,j}^\tau$ the event that the jth gated observation is object-originated observation and all others are clutter-originated at time t.
- $\theta_{t,0}^\tau$ the event that all gated observations are not detected at time t.
- $\hat{Y}^{\tau,t} = \left\{\hat{Y}_1^\tau, \hat{Y}_2^\tau, \ldots, \hat{Y}_t^\tau\right\}$ a set of composite observations inside the gate of track τ up to time t, where

$$\hat{Y}_t^\tau = \left\{\hat{Y}_{t,j}^\tau\right\}_{j=1,\ldots,m_t} = \left\{\hat{y}_{t,j}^{\tau,1}, \hat{y}_{t,j}^{\tau,2}\right\}_{j=1,\ldots,m_t} = \left\{\hat{y}_t^{\tau,i} \middle| \left(\hat{y}_t^{\tau,i} - C_t^i\hat{x}_t^{\tau-}\right)^T\left(S_t^i\right)^{-1}\left(\hat{y}_t^i - C_t^i\hat{x}_t^{\tau-}\right) \leqslant \gamma^{\tau,i}, i = 1,2\right\}. \tag{21}$$

In (21), S_t^i is the residual covariance defined in (4); $\gamma^{\tau,i}$ is the gating threshold for track τ; m_t is the number of composite gated measurement sets for track τ which contains a pair of the radar and the CCD measurements described in (13).

Then, the joint posterior probability density $p\left(x_t^\tau, \chi_t^\tau | \hat{Y}^{\tau,t}\right)$ can be decomposed into two conditional probabilities by the product rule as

$$p\left(x_t^\tau, \chi_t^\tau | \hat{Y}^{\tau,t}\right) = P\left\{\chi_t^\tau | \hat{Y}^{\tau,t}\right\}p\left(x_t^\tau | \chi_t^\tau, \hat{Y}^{\tau,t}\right), \tag{22}$$

where $P\left\{\chi_t^\tau | \hat{Y}^{\tau,t}\right\}$ is the object existence probability calculated via prediction with a Markov chain transition probabilities and the update equations given in (23) and (24).

$$\text{Prediction}: \; P\left\{\chi_t^\tau | \hat{Y}^{\tau,t-1}\right\} = \pi_{1,1}P\left\{\chi_t^\tau | \hat{Y}^{\tau,t-1}\right\} + \pi_{2,1}\left(1 - P\left\{\chi_t^\tau | \hat{Y}^{\tau,t-1}\right\}\right) \tag{23}$$

$$\text{Update}: \; P\left\{\chi_t^\tau|\widehat{Y}^{\tau,t}\right\} = \frac{(1-\delta_t^\tau)P\left\{\chi_t^\tau|\widehat{Y}^{\tau,t-1}\right\}}{1-\delta_t^\tau P\left\{\chi_t^\tau|\widehat{Y}^{\tau,t-1}\right\}} \tag{24}$$

where $\pi_{1,1} \triangleq P\{\chi_t^\tau = 1|\chi_{t-1}^\tau = 1\}$ and $\pi_{2,1} \triangleq P\{\chi_t^\tau = 2|\chi_{t-1}^\tau = 1\}$ are the predefined transition probabilities between the binary states of the existence. δ_t^τ is the data association factor given in (25)

$$\delta_t^\tau = P_D P_G\left(1 - \sum_{j=1}^{m_t}\frac{\Lambda_{t,j}^\tau}{\rho_{t,j}^\tau}\right), \tag{25}$$

where P_D and P_G are the detection probability and the gating probability; $\rho_{t,j}^\tau$ is the given clutter density; and $\Lambda_{t,j}^\tau$ is the predicted measurement density defined as

$$\Lambda_{t,j}^\tau \triangleq p\left(\widehat{Y}_{t,j}^\tau|\chi_t^\tau, \widehat{Y}^{\tau,t-1}\right). \tag{26}$$

We approximate the predicted measurement density (26) by product of the Kalman filter innovation sequence specified in (27).

$$\Lambda_{t,j}^\tau \approx \mathbb{N}\left(\tilde{y}_{t,j}^1; \hat{y}_{t,j}^1 - C_t^1\hat{x}_t^{\tau-}, S_t^1\right)\mathbb{N}\left(\tilde{y}_{t,j}^2; \hat{y}_{t,j}^2 - C_t^2\hat{x}_t^{\tau-}, S_t^2\right), \tag{27}$$

where $\hat{x}_t^{\tau-}$ is the predicted estimate of track τ.

The object existence probability provides the track management capability that is not inherently given to the classical PDA framework. In the implementation, the track is created using the known initialization algorithm, namely the two-step initialization (TSI) [2]. Once track is created, the object existence probability is calculated as described in (23) and (24) recursively, and the track deletion is decided when this value is below a certain threshold that is set to 0.008 in our experiment.

Then, IPDA approximates the object state $\hat{x}_t^\tau$ and its covariance P_t^τ by using the weighted summation representation as following

$$\hat{x}_t^\tau = \sum_{j=0}^{m_t}\beta_{t,j}^\tau\hat{x}_{t,j}^\tau, \tag{28}$$

$$P_t^\tau = \sum_{j=0}^{m_t}\beta_{t,j}^\tau\left[P_{t,j}^\tau + \left(\hat{x}_{t,j}^\tau - \hat{x}_t^\tau\right)\left(\hat{x}_{t,j}^\tau - \hat{x}_t^\tau\right)^T\right], \tag{29}$$

where $\hat{x}_{t,j}^\tau$ and $P_{t,j}^\tau$ are obtained using the jth gated measurement set via the data fusion technique given in the previous section, and $\beta_{t,j}^\tau$ is the associated weight of gated measurement set defined by (30).

By utilizing the object existence variable χ_t^τ, we integrate the track management process into the data association as given in (30),

$$\beta_{t,j}^\tau \triangleq P\left\{\theta_{t,j}^\tau|\chi_t^\tau, \widehat{Y}^{\tau,t}\right\} = \frac{P\left\{\chi_t^\tau, \theta_{t,j}^\tau|\widehat{Y}^{\tau,t}\right\}}{P\left\{\chi_t^\tau|\widehat{Y}^{\tau,t}\right\}} = \frac{1}{1-\delta_t^\tau}\begin{cases}1 - P_D P_G, & j = 0\\ P_D P_G\frac{\Lambda_{t,j}^\tau}{\rho_{t,j}^\tau}, & j > 0\end{cases}. \tag{30}$$

So far we have discussed the data fusion and multiple-object tracking framework for the proposed algorithm. Fig. 4 describes the whole process of the proposed work. Specifically, the flow of the overall system is described by following steps:

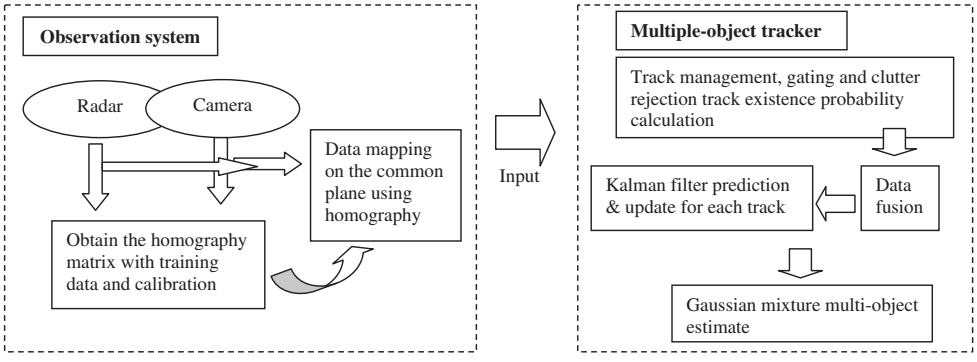


Fig. 4. Overall system block diagram.

(1) Using a set of reference data (radar and camera), calculate the homograph matrix in (9) with the least square method explained in (10)–(12).
(2) From the homography, the radar and camera measurements are mapped onto the common observation plane.
(3) Multi-object tracker block initializes or terminates tracks and excludes clutters from the object existence probability and PDA technique in (21)–(30).
(4) For each track, data fusion and track estimation are performed via information Kalman prediction and update steps as described in (16)–(20).
(5) Gaussian mixture of Kalman estimate is finally established to represent multi-object state estimate.

The proposed algorithm is based on individual IPDA filter for each track combined with the specified track management block and the data fusion technique embedded in the information Kalman filtering. Compared to other multi-object trackers such as JPDA filter [2] or PHD filter [25], IPDA reduces the computational time thanks to the track quality measure, i.e., object existence probability. In the experimental results, we include the computation time comparison with the PHD filter that is known to be the state-of-the-art multi-object tracker.

4. Experimental results

Our proposed system is applied to tracking three objects that appear and disappear independently. A set of the ground truth for them is displayed in Fig. 6 as red$***$ lines. Our experiments are done under controlled environment where objects are moving based on the predetermined scenario. Therefore, the ground truth is obtained from the simulated position based on the exactly known scenario. But the scenario is not known to the sensors and filters.

Observations are collected from the proposed system as described in Section 3. Note that we observe and estimate positions of objects in 2D space assuming that y-coordinate is fixed. In the experiments, we focus on two performance comparisons. One is the comparison of the proposed system with the radar-only system, and the other is the comparison of the tracking algorithms embedded in the proposed system. Before discussing the experimental results, we describe the mathematical model of the dynamic system and related parameters.

4.1. Model description and filter parameter settings

The dynamic system (2) for each object is specified with

$$F_t = diag(F_1, I_2, F_1), \quad F_1 = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \tag{31}$$

$$Q_t = diag(Q_1, 0_2, Q_1) \quad Q_1 = q \begin{bmatrix} T^3/3 & T^2/2 \\ T^2/2 & T \end{bmatrix}, \tag{32}$$

where T is the sampling interval and the system process noise parameter q is set to 0.1. I_2 and 0_2 are 2×2 identity and zero matrices, respectively. F_t and Q_t represent the block matrices for the dynamic system and noise covariance, respectively. The given model (31) and (32) is conventionally used in tracking literature; called Ground Moving Target Indicator (GMTI) model [13,29]. Note that the GMTI model describes the position and velocity of moving object for each coordinate. Here, the velocity state is not measured but regarded as noise terms as given in (32). Because we do not measure the velocity using the sensor, the measurement matrix only considers positions.

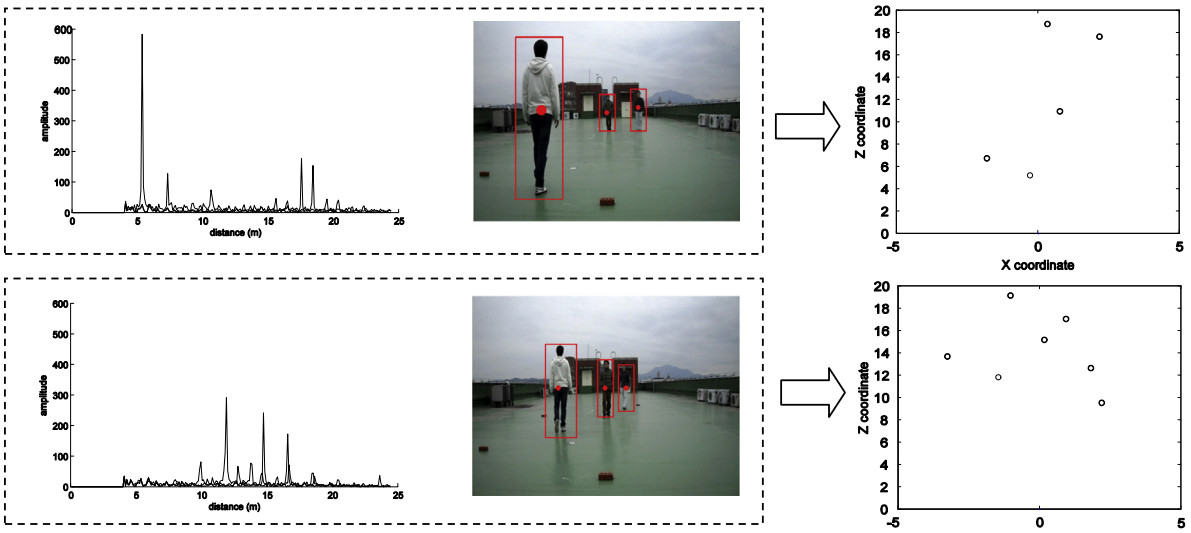
Originally, radar observations are on the polar coordinate, thus, the measurement matrix of the radar sensor, $C_t^1 = [1\ 0\ 1]$, is obtained by using the coordinate transformation from the polar to the Cartesian. In the experiments, the observation noise parameters for each sensor are set as $R_t^1 = diag(5, 0.1)$ and $R_t^2 = diag(0.5, 0.5)$. P_D and P_G are set to 0.9 and 0.865, respectively.

For the proposed multi-object tracker, we assume that the dynamic system and the noise model are given as a priori as mentioned before. The system noise is assumed to be Gaussian which contains information about the maximum speed of the object. The measurement noise covariances are obtained by experiments, and the gating threshold is set to the value as suggested in the literature [1]. The track deletion threshold is set to 0.008 in the experiments.

Once the parameters are decided, it is sufficient to guarantee the reliable performance. That is because sensors are stable and the maximum velocity (speed) of the object is predictable based on the sampling interval (or scanning time) of the system. However, the performance of the tracker can be degraded when severe clutters of non-uniform distribution are involved. In such situations, an adaptive clutter intensity approximation technique can be applied [24].

4.2. Discussion about experimental results

Fig. 5 illustrates range signals from the radar, detections from CCD and combined sets of these two types of signals mapped onto the common plane. The displayed detections in Fig. 5 can be considered as the result of [30] that is the motivation of the complementary observation system. Note however that it is not sufficient to show the reliable tracking



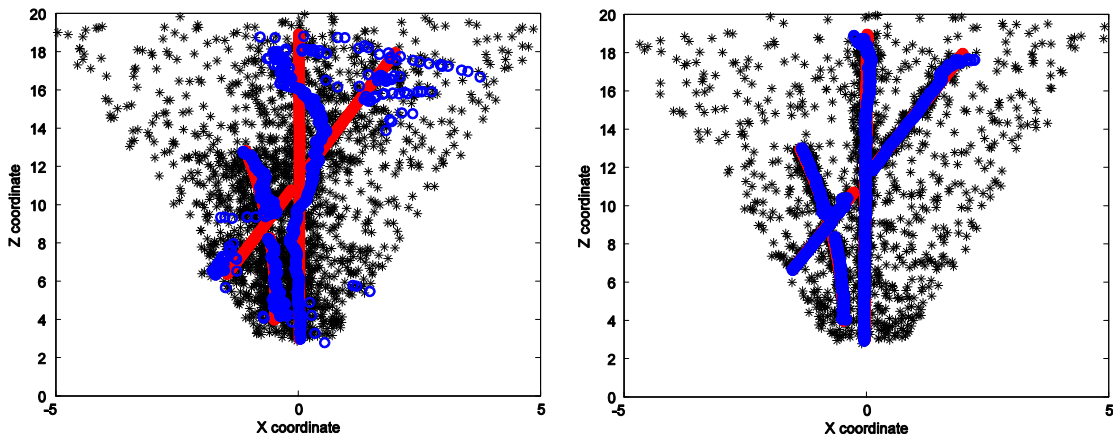
(a) radar detections (b) CCD detections (c) mapped detections

Fig. 5. Observations from two sensors and its display in the common plane including clutters.

performance because the detection results contain uncertainty from sensor noises and ambiguity in data association for multi-object tracking.

As can be seen in Fig. 5a, range measurements of radar are corrupted by sensor noise and clutters. Clutters are originated from the multi-path effect from fences and the ground. Detected CCD measurements are displayed in Fig. 5b with bounding box representation and its centroid. In this case, it is not possible to perceive exactly how far objects are located if the CCD is not calibrated. Fig. 5c displays the combined measurements using estimated homography. From those inputs, the designed filter rejects clutters and associates detections using IPDA technique as suggested in Section 3.3. The blob detector used in our system may have false positive detections when the feature values have not sufficient discriminating power. However, certain amount of false positive detections can be moderately rejected using the IPDA technique.

For the quantitative comparison we use the particular metric for multi-object tracking called the Optimal Sub-Pattern Assignment (OSPA) distance [27] instead of the mean square errors (MSE). Note that the OSPA distance is the overall performance index which contains the localization error (i.e., position error) and the cardinality error (object number error) simultaneously. The reason for using the OSPA distance is that MSE is valid only when the object existence is exactly known in advance.



(a) radar-only tracking system (b) proposed tracking system

Fig. 6. Trajectory comparison (red: ground truth, blue: estimated trajectory, black star: observations). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

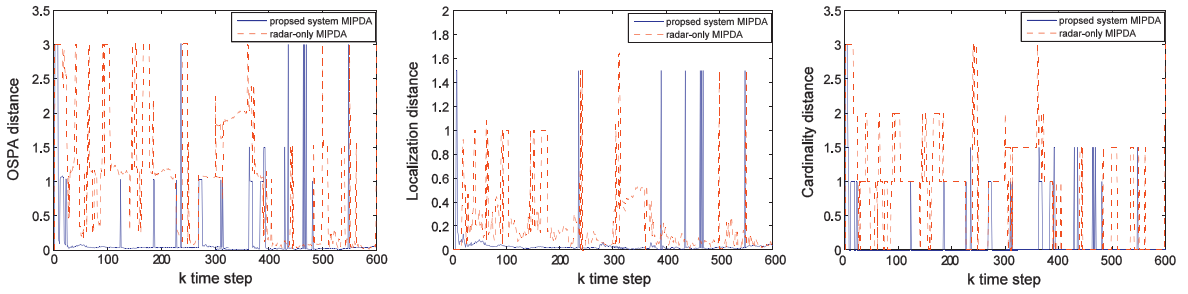


Fig. 7. Comparison between the radar-only system and the proposed system using OSPA distance, localization distance, and cardinality distance.

4.2.1. Performance comparison of the proposed system with radar-only system

To clearly show the advantages of the proposed system, we compare its performance with that of radar-only system. As can be seen in Fig. 6, the comparison of estimated trajectories results show that the proposed system outperforms the radar-only system due to the complementary observations based on the data fusion technique.

We display the comparison of overall performance with respect to the OSPA distance in Fig. 7. In Fig. 7a, we see that the proposed system is superior to the radar-only system. Note that the OSPA distance is important performance metric of multiple object tracking because it simultaneously calculates the localization error and the cardinality error, thus, track losses are taken into account in the error metric.

The resultant OSPA distance of the proposed system has significant low (accurate) and stable values compared to the radar-only system. For the clarity, the localization distance and the cardinality distance are also displayed in Fig. 7b and c, respectively. They also confirm that the proposed system outperforms the radar-only system in both aspects.

The proposed complementary observation with MIPDA gives an accurate estimate for the number of objects due to the correct data association and clutter rejection. Thus, the correct track management and the efficient clutter rejection significantly enhance the tracking performance of the proposed system. It is obvious that it also outperforms the vision-only system because vision-only system cannot inherently measure the radial distance of objects.

4.2.2. Performance comparison of trackers embedded in the proposed system

Finally, we make the performance comparison with recently proposed multi-object tracker based on Gaussian Mixture PHD (GMPHD) [25]. GMPHD is known to be the full Bayesian multi-object tracker which does not require the data association. This method regards multi-object states as single meta-state over the whole state space. The experimental results for the same scenario with respect to the OSPA distance are illustrated in Fig. 8a. We also confirm from the results that the proposed system is superior to the GMPHD not only in the accuracy but also in computational time per scan where the elapsed time are given as follows: GMPHD: 0.1 s, Proposed MIPDA: 0.002 s, respectively. The reason for this result is from the gating technique used in the MIPDA algorithm. GMPHD also can be improved by using the gating technique as suggested in [35], however, track identity information is not inherently provided in GMPHD filter. The measured computation time can be thought of as the execution time of the global system and the elapsed time of our system shows that it is real-time. Note that our system is implemented in C++ using the standard PC (Intel Core 2 Duo, 2.67 GHz, 3 GB RAM).

Compared to the results of radar-only system in Fig. 7, GMPHD achieves little performance improvement; because the cardinality error is not reliable as can be seen in Fig. 8b. The unreliable cardinality information leads to the degradation in the overall performance as represented in the OSPA distance.

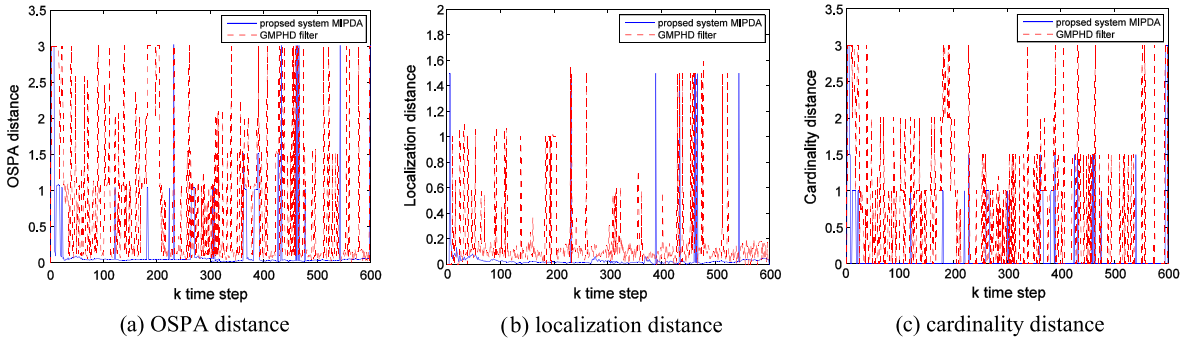


Fig. 8. Comparison between the GMPHD filter and the proposed system using OSPA distance, localization distance, and cardinality distance.

5. Conclusion

In this paper, we developed hardware integration combining radar and CCD sensor measurements for a multiple-object tracking system. The complementary observation system is designed by using homography and data fusion techniques. By doing so, the low resolution of bearing angle of radar is compensated by CCD observation. Then, the multiple-object states are accurately estimated via Gaussian mixture of IPDA filter called multi-object IPDA (MIPDA). The performance of the proposed system is evaluated with the real-world experiments and test results verify that the system is viable for the real-world implementation.

Acknowledgments

This work was financially supported in part by the Pioneer Research Center Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (Grant no. 2012-0009462), and by the National Research Foundation of Korea under Grant NRF-2013R1A1A2A10012587.

References

[1] Y. Bar-Shalom, X.R. Li, Multitarget–Multisensor Tracking: Principles and Techniques, YBS publishing, Storrs, CT, 1995.
[2] Y. Bar-Shalom, G.A. Waston, Automatic tracking formation in clutter with a recursive algorithm, in: Proceedings of 28th Conference on Decision, Control, Florida, December, 1989, pp. 1402–1408.
[3] T. Bucher, C. Curio, J. Edelbrunner, C. Igel, D. Kastrup, I. Leefken, G. Lorenz, A. Steinhage, W.V. Seelen, Image processing and behavior planning for intelligence vehicles, IEEE Trans. Industr. Electron. 50 (1) (2003) 62–75.
[4] M.D. Butala, R.A. Frazin, Y. Chen, F. Kamalabadi, Tomographic imaging of dynamic objects with the ensemble Kalman filter, IEEE Trans. Image Process. 18 (7) (2009) 1573–1587.
[5] F. Chang, C.-J. Chen, C.-J. Lu, A linear-time component-labeling algorithm using contour tracking technique, Comput. Vis. Image Underst. 93 (2) (2004) 206–220.
[6] S.Y. Chen, Kalman filter for robot vision: a survey, IEEE Trans. Industr. Electron. 59 (11) (2012).
[7] H. Cho, S.W. Kim, Mobile robot localization using biased chirp spread spectrum ranging, IEEE Trans. Ind. Electron. 57 (8) (2009) 2826–2835.
[8] C. Chong, K. Chang, S. Mori, Distributed tracking in distributed sensor networks, in: Proceedings of the American Control Conference, Seattle, 1986.
[9] Tinne De Laet, Herman Bruyninckx, Joris De Schutter, Shape-based online multitarget tracking and detection for targets causing multiple measurements: variational Bayesian clustering and lossless data association, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 2477–2491.
[10] Eric Dedrick, Daniel Lau, A Kalman-filtering approach to high dynamic range imaging for measurement applications, IEEE Trans. Image Process. 21 (2) (2012) 527–536.
[11] Zili Deng, Peng Zhang, Wenjuan Qi, Jinfang Liu, Yuan Gao, Sequential covariance intersection fusion Kalman filter, Inf. Sci. 189 (2012) 293–309.
[12] F. Fleuret, J. Berclaz, R. Lengagne, P. Fua, Multi-camera people tracking with a probabilistic occupancy map, IEEE Trans. Pattern Anal. Mach. Intell. 30 (2) (2008) 267–282.
[13] A. Gelb (Ed.), Applied Optimal Estimation, The MIT Press, 1974.
[14] G.C. Goodwin, H. Haimovich, D.E. Quevedo, J.S. Welsh, A moving horizon approach to networked control system design, IEEE Trans. Autom. Control 49 (9) (2004) 1427–1455.
[15] E.L. Haseltine, J.B. Rawling, Critical evaluation of extended Kalman filtering and moving-horizon estimation, Ind. Eng. Chem. Res. 44 (8) (2004) 2451–2460.
[16] R.E. Kalman, A new approach to linear filtering prediction problems, ASME J. Basis Eng. 82 (1960) 34–45.
[17] Chanki Kim, Rathinassamy Sakthivel, Wan Kyun Cung, Unscented FastSLAM: a robust and efficient solution to the SLAM problem, IEEE Trans. Robot. 24 (4) (2008) 808–820.
[18] J.M. Lee, K. Son, M.C. Lee, J.W. Choi, S.H. Han, M.H. Lee, Localization of a mobile robot using the image of a moving object, IEEE Trans. Ind. Electron. 50 (3) (2003) 612–619.
[19] M.-S. Lee, Y.-H. Kim, An efficient multitarget tracking algorithm for car applications, IEEE Trans. Ind. Electron. 50 (2) (2003) 397–399.
[20] S.J. Lee, Y. Motai, M. Murphy, Respiratory motion estimation with hybrid implementation of extended Kalman filter, IEEE Trans. Industr. Electron. 59 (11) (2012).
[21] Y. Li, B. Wu, R. Nevatia, Human detection by searching in 3D space using camera and scene knowledge, in: Proceedings of the 17th International Conference on Pattern Recognition, USA, 2008.
[22] D. Musicki, R. Evans, Joint integrated probabilistic data association-IJPDA, in: Proceedings of the 5th IEEE Conference on Information Fusion 2002, Annapolis, Maryland, USA, July 2002.
[23] D. Musicki, R. Evans, S. Stancovic, Integrated probabilistic data association, IEEE Trans. Autom. Control 39 (6) (1994) 1237–1241.
[24] D. Musicki, B.L. Scala, Multi-target tracking in clutter without measurement assignment, in: Proceedings of the 43rd IEEE Conference on Decision and Control, Bahamas, 2004.
[25] S.A. Pasha, B.-N. Vo, H.D. Tuan, W.-K. Ma, A Gaussian mixture PHD filter for jump Markov system models, IEEE Trans. Aerosp. Electron. Syst. 45 (3) (2009) 919–936.
[26] B.S.Y. Rao, H.F. Durrant-Whyte, J.A. Sheen, A fully decentralized multi-sensor system for tracking and surveillance, Int. J. Robot. Res. 12 (1) (1993) 20–44.
[27] D. Schumacher, B.T. Vo, B.N. Vo, A consistent metric for performance evaluation of multi-object filters, IEEE Trans. Signal Process. 56 (8) (2008) 3447–3457.
[28] SiversIMA RS 3400. <http://www.siversima.com/products/fmcw-sensor-front-ends/>.
[29] H. Song, V. Shin, M. Jeon, Mobile node localization using fusion prediction-based interacting multiple model cricket sensor network, IEEE Trans. Industr. Electron. 59 (11) (2012).
[30] S. Sugimoto, H. Tateda, H. Takahashi, M. Okutomi, Obstacle detection using millimeter-wave radar and its visualization on image sequence, in: Proceedings of the 17th International Conference on Pattern Recognition, UK, 2004.
[31] S. Sun, Optimal and self-tuning information fusion Kalman multi-step predictor, IEEE Trans. Aerosp. Electron. Syst. 43 (2) (2007) 1400–1417.
[32] S.-H.P. Won, W.W. Melek, F. Golnaraghi, A Kalman/particle filter-based position and orientation estimation method using a position sensor/inertial measurement unit (IMU) hybrid system, IEEE Trans. Ind. Electron. 57 (5) (2010) 1787–1798.
[33] Xiufeng Xe, Yang Le, Wendong Xiao, MEMS IMU and two-antenna GPS integration navigation system using internal adaptive Kalman filter, IEEE Aerosp. Electron. Syst. Mag. 28 (10) (2013) 22–28.
[34] X. Xu, B. Li, Adaptive Rao-Blackwellized particle filter and its evaluation for tracking in surveillance, IEEE Trans. Image Process. 16 (3) (2007) 838–849.
[35] Hongjian Zhang, Zhongliang Jing, Hu Shiquiang, Gaussian mixture CPHD filter with gating technique, Signal Process. 89 (2009) 1521–1530.
[36] Zebo Zhou, Yong Li, Junning Liu, Gun Li, Equality constrained robust measurement fusion for adaptive Kalman-filter-based heterogeneous multi-sensor navigation, IEEE Trans. Aerosp. Electron. Syst. 49 (4) (2013) 2146–2157.