International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 15489-15500

An Advanced H-Infinity Filtering Approach for Multiple Object Tracking in Video Sequences

Adinarayana Ekkurthi
Research Scholar, Department of Electronics and Communication Engineering
Acharya Nagarjuna University - College of Engineering and Technology (ANUCET),
Guntur, Andhra Pradesh, India.

Orcid Id: 0002-3731-0614

Dr. G.Sudhavani
Professor, Department of Electronics and Communication Engineering,
R.V.R & J.C College of Engineering, Guntur, Andhra Pradesh, India.

Orcid Id: 0000-0003-1242-1534

Abstract

Video target tracking is the way toward identifying the moving object and recognized the path of the tracking object. Single object tracking is the basic and simple way since it has distinctive fruitful systems to distinguish the single object. Yet, multiple object tracking is extremely opposing exertion in light of the fact that diverse objects have the comparative appearance. In this paper, we employ an advanced H∞ filtering for multiple objects tracking in the video sequences. In the beginning, we embrace the FDM (Frame Difference Method) and HOG (Histogram of Gradients) detection techniques to isolate the object region of the video frame. The color spatial feature, texture feature and edge oriented features are separated from the object region. The matching degrees of these features are registered by utilizing the Multi-feature fusion method. On the off chance that the matching degree is more prominent than the user-specified threshold value, the feature of the object is updated and the position of the object is identified. Otherwise again the new searching is reiterated and the matching degree is calculated for that new searching. An advanced H∞ filtering algorithm to establish object motion model, using the present items information to guess object's location, so that we can reduce the investigate capacity and search time of moving an object to achieve fast tracking. The H-infinity filter does not have any knowledge about the system model and the observation model; in any case, it has better execution and exactness.

Keywords: Multiple Object Tracking, FDM, HOG, Kalman filter, H-infinity filter.

INTRODUCTION

Object tracking [1, 2] has an essential angle in computer vision on account of its extensive variety of usage advance and prospects in industries, for example, intelligent human-computer interaction, video checking, and intelligent transportation. The position of moving target and the objective way is distinguished by utilizing video target tracking technique [3]. These days the single object tracking method is straightforward in light of the fact that quantities of systems are acquainted with distinguishing the single object. Be that as it may, the multiple objects tracking method is a testing errand in light of the fact that diverse objects have the comparative appearance [4]. When tracking multiple objects in a picture series, different complications such as occlusion, miss-detection, false detection, and abrupt camera motion often occurred [3]. There are two classes are utilized to defeat these challenges; they are deterministic methods and stochastic methods.

Deterministic methods typically track by performing an iterative search for the local maxima of a similarity cost function between the template image and the current image. The cost function widely used is the sum of squared differences (SSD) between the template and the current image such as in [4, 5]. More robust similarity measures have been applied and the mean-shift algorithm or other optimization techniques have been utilized to find the optimal solution [6-10]. Model-based tracking algorithms incorporate a priori information about the objects to develop representations such as skin complexion, body blobs, kinematic skeleton, silhouettes or layer information [7], [11-15]. Appearance-based approaches apply recognition algorithms to learn the objects either on some basis such as the eigenspace formed from observations or in kernel space [16-18]. On the other hand, the stochastic methods use the state space to model the underlying dynamics of the tracking system.

Video tracking systems face lots of challenges such as target occlusion, deformation, scaling, and illumination variation. A variety of filtering techniques has been implemented to overcome these challenges. The Kalman filter, known as the optimal linear quadratic estimator, is widely used in tracking systems. It utilizes present measurement and previous updated estimate, combining with recursive equations to predict and update the state at the current time step. More specifically, the Kalman filter requires prior knowledge of the process and measurement noises, so the optimal state estimator based on the given PDF of noise can be designed. However, in the real world, it is difficult to obtain the noise distribution model. Moreover, the Kalman filter only minimizes the expected value of the variance of estimation error. The particle filter, also known as sequential Monte Carlo [20], is the most popular approach which recursively constructs the posterior pdf of the state space using Monte Carlo integration. It has been developed in the computer vision community and applied to tracking problem and is also known as the Condensation algorithm [21]. In the process of H∞ filtering, the noises are assumed to be the worst case and the filter desires to minimize the worst-case estimation error [28]. The H∞ filter has better performance in robustness and accuracy when having no knowledge about system model and disturbance model. In addition, the H∞ filter has the similar recursive equations with the Kalman filter, which makes it highly efficient in practical applications.

In this paper we employ an advanced H∞ filtering algorithm to establish object motion model, using the current object's information to predict object's position, so that we can reduce the search scope and search time of moving an object to achieve fast tracking. By using H∞ filtering, there would be no assumption to the system disturbance as H∞ filtering possesses the characteristic of considering the worst case estimation error. Since H∞ filter is a recursive algorithm, only the previous time step and current state measurement are required for object tracking but no history observation is needed. Hence, there would be no requirement for high capacity of computational storage. In our proposed approach for object detection we adopt HOG detector and for predicting the next location of an object we employ advanced H∞ filtering algorithm.

The rest of the segment of the paper is depicted underneath. In section (3), the moving object detection technique is actualized. The modified H-infinity filter approach is delineated in section (4). The test result and the conclusion are examined in section (5) and (6).

LITERATURE REVIEW

Hui Li *et al*. [22] presented a tracking algorithm of multiple pedestrians based on particle filters in video frames. The method gets necessary value of the object and the background through extracting a priori knowledge thus to achieve multi pedestrian detection; it adopts color and texture features into particle filter to get better observation results and then automatically adjusts weight value of each feature according to current tracking background. During the process of tracking, the method processes rigorous occlusion condition to prevent drift and loss phenomena caused by object occlusion and associated detection results with particle state to propose discriminated method for object disappearance and emergence thus to achieve robust tracking of multiple objects. Practical proof and examination in video sequences demonstrate that proposed algorithm improves the tracking performance and has better tracking results.

Allan De Freitas *et al*. [23] proposed two solutions to the crowds tracking problem with a box particle filter approach and with a convolution particle filtering method. The developed filters can deal with the measurement origin uncertainty in a well-designed way, i.e. resolve the data association problem. For the box particle filter (PF) they derived a theoretical expression of the generalized likelihood function in the presence of clutter. An adaptive convolution particle filter (CPF) was also developed and the performance of the two filters was compared with the standard sequential importance resampling (SIR) PF. The pros and cons of the two filters are illustrated over a realistic scenario (representing a crowd motion in a stadium) for a large crowd of pedestrians. Accurate estimation results were achieved.

Xingbo Wang *et al.* [24] proposed a new target tracking approach for wireless sensor networks (WSNs) by using the extended H-infinity filter. Initially, the extended H-infinity filter for nonlinear discrete-time systems was deduced through the Krein space analysis scheme. In the next phase of their approach, the proposed extended H-infinity filtering algorithm was applied to target tracking in wireless sensor networks. Finally, experiments were conducted through a small wireless sensor network test-bed. Both experimental and simulation results illustrated that the extended H-infinity filtering algorithm was more accurate to track a moving target in wireless sensor networks than using the extended Kalman filter in the case of having no knowledge of the statistics of the environment and the target to be tracked.

H. Morimitsu *et al.* [25] proposed a novel approach for exploiting structural relations to track multiple objects that may undergo long-term occlusion and abrupt motion. In their work, a colour-based particle filter was chosen as the single object tracker due to its simplicity and good results demonstrated in previous studies. They used a model-free approach that relies only on annotations given in the first frame of the video to track all the objects online, i.e. without awareness from upcoming frames. They initialized a probabilistic Attributed Relational Graph (ARG) from the first frame, which was incrementally updated along the video. Instead of using the structural information only to evaluate the scene, their proposed approach considers it to generate new tracking hypotheses. By that, their method was capable of generating relevant object candidates that were used to improve or recover the track of lost objects. Their proposed method was evaluated on several videos of table tennis, volleyball, and on the ACASVA dataset. The results showed that their approach was very robust, flexible and able to outperform other state-of-the-art methods in sports videos that present structural patterns.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 15489-15500
© Research India Publications. http://www.ripublication.com

Yunji Zhao and Hailong Pei [26] presented an algorithm which can detect and track many items, and modernize target representation automatically. The contributions of their paper as follow: Firstly, they also used color histogram(CH) and the histogram of orientated gradients(HOG) to represent the objects, model update are realized by Kalman filter and Gaussian model; secondly they used Gaussian Mixture Model(GMM) and Bhattacharyya distance to detect object manifestation. Particle filter with joint features and model update mechanism can improve tracking results. Experiments on video sequences demonstrate that the method presented in that paper can realize multiple objects detection and tracking.

Zhiyu Zhou *et al.* [27] proposed a novel object tracking method with the fusion of the extended Kalman particle filter (EKPF) and the least squares support vector regression (LSSVR). First, the observation value of Kalman filter was acquired with the cues of colour and movement features. The significance probability density function was generated by extended Kalman filter (EKF), which makes the distribution of particles approximately to the posterior probability. And then, a weighted plan was used to determine the weighted coefficient of LSSVR model, the robustness and sparseness of LSSVR modeling will thereby be improved. The efficient EKF features of tested samples served as training samples to establish the dynamic LSSVR model real-time in the next frame. Finally, the LSSVR was used to calibrate the tracking results of Kalman particle filter, such that the tracking object will always follow the correct motion trajectory. The experimental results showed that their method performs favourably against traditional Kalman particle filter with real-time performance and strong robustness.

H. Wang and S. Nguang [28], have developed a new video tracking method based on multi-feature fusion and H∞ filtering was proposed. They have extracted color spatial distribution feature, target contour feature and edge gradient histogram of video targets, and calculate the matching degree between the candidate feature and target feature. They have utilized these three features and calculate the final matching degree by using linear weighted fusion method. To boost the efficiency of features extraction and fusion, thread pool and multi-thread synchronization are adopted. A robust video target motion state estimation method based on H∞ filtering was presented to recursively estimate and predict the target state which can narrow the scope of features searching and then greatly improve the efficiency and accuracy of features extraction. Compared with the Kalman filter, the H∞ filter makes no assumptions in process and measurement noise but has the similar efficient recursive equations. Therefore when the noises are non-Gaussian distributed, the H∞ filter-based visual tracking systems have better performance in robustness.

M. Babaee *et al*. [29] have developed an approach that tracks super pixels instead of detection boxes in multi-view video sequences. Specifically, they have first extracted super pixels from detection boxes and then associate them within each detection box, over several views and time steps that lead to a combined segmentation, reconstruction, and tracking of super pixels. They have constructed a flow graph hand incorporate both visual and geometric cues in a global optimization framework to minimize its cost. Hence, they have simultaneously achieved segmentation, reconstruction, and tracking of targets in the video. Experimental results confirm that the existing approach outperforms state-of-the-art techniques for tracking while achieving comparable results in segmentation.

PROPOSED METHOD

Moving Object Detection

Moving Object detection is a critical undertaking in the video target tracking strategy. There are two procedures are utilized as a part of the moving object detection strategy. To distinguish the object, most importantly else, the target region is separated before tracking the objective. At to start with, takes the M frames from the back to back video arrangements. There are two procedures are utilized to distinguish the objective region; to begin with, embrace the frame difference method (FDM) to recognize the object region and from that point forward, the HOG (Histogram of Oriented Gradients) detection strategy is utilized to demonstrate the better consequence of the object region. The objective region extraction of the essential M frame appears in Fig. 1.

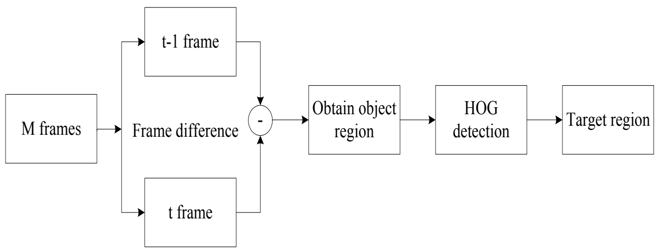

Figure 1: Target Region Extraction

Frame Difference Method (FDM)

FDM is one sort of strategy used to recognize the moving item by utilizing the contrast between the present frame image and the previous image from the information contained in each frame taken progressively in any of the images.

Figure 2: Background subtraction result

In the FDM technique, the background region of the image is subtracted then the video target is found and identify rapidly. The complexity of the two frames is given by,

$$V_n(u,v) = \begin{cases} 1, & |I_n(u,v) - I_{n-1}(u,v)| > t_T \\ 0, & otherwise \end{cases} \tag{1}$$

In formula (1), n is the current frame number, n is the previous frame number, I is the frame of the image, (u,v) is the frame coordinates and t_T is the threshold value of motion is to eliminate noise, V is the image that contains the information of the area in which motion is detected.

Histogram of Gradients (HOG) detection

After the background region subtraction, the HOG detection is utilized to detect the motion region in the image. The primary function of the HOG detector is utilized to examine the entire frame then pursuit the object in the video frame. This detection strategy is mostly influenced by background jumbled, so we need to apply the HOG detection technique after the background subtraction. At that point just it will detect the exact object else it will identify the false object region. At last, it will create the productive and precise identification. Fig 3, demonstrates the HOG detection result.

Figure 3: HOG detection result.

Feature Extraction

After the object region detection process, we have to detect the different objects precisely and powerfully. The multiple objects are detected based on the feature of the object. In the multiple targets tracking methods, the feature extraction is very important because the matching degree of each object is calculated from the detected image and the candidate image feature. Tracking of the single object gets an exact result in light of the fact that the single object detection method gets just the single feature, so it has better detection result. Two sorts of feature extraction methods are utilized to separate the feature of the object. One is HSV (Hue Saturation Value) color histogram and another one is the LBP (Local Binary Pattern) color histogram. But, LBP operator has a few unacceptable angles. In this paper, we have to utilize the HSV color histogram for detecting the multiple objects feature extraction method. To detect the multiple objects, we extract the color spatial distribution feature, target contour feature and edge gradient feature of the video target.

Color Spatial Distribution Feature Extraction

In multiple objects tracking strategy, each object has diverse color dispersions, so the color spatial distribution feature extraction method used to recognize the distinctive color of the each object. The color spatial information here and there makes perplexity to distinguish the object, so this strategy extricates both the color and pixel position information about the tracking object. The HSV color histogram technique is utilized to depict the color distribution since it is more useful than the RGB color histogram. The diverse color of the image is gotten by utilizing the pixels weight of the objective. The pixel weight is very high close to the objective centroids and the encompassing district has the low pixel weight. The HSV color histogram can be given by,

$$H(m) = \sum_{s=0}^{M}\sum_{w=0}^{N} \frac{((1-2(1-\alpha)\sqrt{(\lambda-\lambda_0)^2+(\Omega-\Omega_0)^2}}{\sqrt{\omega^2+h^2}} \times F(\psi(s,w),m)) \tag{2}$$

$$F(a,b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \tag{3}$$

In equation (2), ω and h is the width and tallness of the bounding region of objective, λ_0 and Ω_0 is the directions of the centroids. Consider the deliberate color index of a pixel is $\psi(\lambda,\Omega),\alpha$ the weighting coefficients, m is the color level ranging from 0 to 255. The level of coordinating between the reference and the candidate region is given by,

$$D_{HSV}(H_R,H_C) = 1 - \sqrt{1 - \frac{1}{256\sqrt{H_R H_C}} \sum_{i=1}^{256} \sqrt{H_R(i) H_C(i)}} \tag{4}$$

Where, $\overline{H} = \frac{1}{256\sum_{i=1}^{256} H(i)}$

In equation (4), H_R and H_C is the reference and candidate HSV color histogram.

Consider, $\eta(i)$ represents the pixel distribution vector in the HSV color histogram, it is given by,

$$\|\eta(i)\| = F(\psi(\lambda_j,\Omega_j),m_i \times \sqrt{(\sum_{j=1}^{\omega\times h}(\lambda_j-\lambda_0))^2+(\sum_{j=1}^{w\times h}(\Omega_j-\Omega_0))^2} \quad (5)$$

Where m_i is the color level of the i[th] bin? Suppose the reference pixel distribution vector corresponding to the i[th] bin of the HSV color histogram is $\phi(i)$. The Bhattacharyya coefficients of the two vectors can be represented as

$$d(i) = \frac{\eta(i)\ .\ \phi(i)}{\|\eta(i)\|\ .\ \|\phi(i)\|} \quad (6)$$

The degree of matching of the spatial color distribution vector is given by,

$$D_{HSV-PDV} = \mu_{HSV} d_{HSV} + \mu_{PDV} d_{PDV} \quad (7)$$

Where $d_{PDV} = \frac{1}{256}\sum_{i=1}^{256} d(i)$

In equation (7), μ_{HSV} and μ_{PDV} are the two coefficients. Here $\mu_{HSV} + \mu_{PDV} = 1$.

Target Contour and Edge Oriented Feature Extraction

In multiple objects tracking method, the video target mainly includes target contour, minimum bounding rectangle, and region. The target contour and the edge oriented feature extraction is a very important technique for the object detection and tracking method. The bounding rectangle and the centroids are calculated after the contour extraction. After the frame difference method, the target contour is extracted, once the target contour is extracted then calculates the centroids and the normalized histogram of the contour region.

After the FDM, the frame only contains the image target; it is represented as $f(x,y)$ and the width and height of the minimum bounding box of target contour are denoted W and H. The horizontal and the vertical coordinate ranges are taken from the target contour. Here, the horizontal coordinate ranges from $x_0 + W$ and the vertical coordinate range from $y_0 + H$. The centroids of the video target is given by,

$$x_p = \frac{\sum_{x=x_0}^{W+x_0}\sum_{y=y_0}^{H+y_0} x f(x,y)}{\sum_{x=x_0}^{W+x_0}\sum_{y=y_0}^{H+y_0} f(x,y)}\ ,\ y_p = \frac{\sum_{x=x_0}^{W+x_0}\sum_{y=y_0}^{H+y_0} y f(x,y)}{\sum_{x=x_0}^{W+x_0}\sum_{y=y_0}^{H+y_0} f(x,y)} \quad (8)$$

The centroid moment of the target contour can be given by,

$$M_c = \frac{\sum_{x=x_0}^{W+x_0}\sum_{y=y_0}^{H+y_0}(x-x_p)(y-y_p)f(x,y)}{\sum_{x=x_0}^{W+x_0}\sum_{y=y_0}^{H+y_0} f(x,y)} \quad (9)$$

The normalized histogram of the contour region can be given by,

$$\vec{V}_c = \frac{\begin{bmatrix}\sum_{j=0}^{N-1} X[j]-x_p & \sum_{j=0}^{N-1} Y[j]-y_p\end{bmatrix}^T}{\left(\sum_{j=0}^{N-1} X[j]-x_p\right)^2 + \left(\sum_{j=0}^{N-1} Y[j]-y_p\right)^2} \quad (10)$$

In equation (8), $x[j]$ and $y[j]$ are the coordinates of target contour, N is the number of contour pixel.

Then the matching degree of the target contour is computed by,

$$D_{contour} = \beta_1\left(1-\frac{M_c(t-1)-M_c(t)}{M_c(t-1)}\right)+(1-\beta_1)\frac{\vec{V}_c^{\,T}.\vec{V}_c}{\|\vec{V}_m\|.\|\vec{V}_m\|} \quad (11)$$

In equation (9), M_c and V_c is the centroid moment and the normalized mean value vector of the target contour at the time t, $M_c(t-1)$ and $V_c(t-1)$ is the centroid moment and the normalized mean value vector of the contour at the time $(t-1)$.

If any changes occurred in the color and intensity of an image is detected by using the gradient level of an image. So, the edge gradient of the target is represented as $r\ (x,y)$ and the gray scale image is represented as $e(x,y)$. Then the gradient of the image is given by,

$$r_x(x,y) = e(x,y) - e(x-1,y) \quad (12)$$

$$r_y(x,y) = e(x,y) - e(x,y-1) \quad (13)$$

The norm of gradient at pixel (x,y) is given by,

$$N(x,y) = \sqrt{(r(x,y)-r(x-1))^2 + r(x,y)-r(x,y-1)^2} \quad (14)$$

We adopt the edge oriented gradient histogram (EOG) is compared to candidate contour with target contour is given by,

$$H_{EOG} = [b_1, b_2, b_3,\ \ \dots\ \ , b_9] \quad (15)$$

The whole part of the gradient is divided into nine equal parts, so, the angle of each part is 20 degrees. Here, the target contour is divided into four parts to obtain the accurate matching degree. The centroid of the video target is calculated from the origin coordinates of the target. The Bhattacharyya distance between the candidate contour histogram and the target contour histogram is denoted as H_c and as H_T.

$$H_{BHAT} = \sqrt{1 - \sum_{j=1}^{9} \sqrt{H_c(j) H_T(j)} / 9\sqrt{\overline{H}_c.\overline{H}_T}} \tag{16}$$

Where $\overline{H} = \frac{1}{9}\sum_{j=1}^{9} H(j), \quad H(j) = b_j$

The matching degree of the target contour is given by,

$$D_{HEOG} = (4 - (\sum_{j=1}^{9} H_{BHAT}(H_c, H_T)[j])) / 4 \tag{17}$$

Multi-Feature Fusion for Video Target

The color spatial feature, contour feature and edge gradient features are extracted from the already detected target region. These features are used to detect the video target accurately. To predict the accurate video target, the matching degree is estimated by using these three features. The matching degree of these three features are given by,

$$D_{mf} = \alpha_1 D_{colorspatial} + \alpha_2 D_{contour} + \alpha_3 D_{HEOG} \tag{18}$$

Where α_1, α_2 and α_3 are the weighted coefficients of the extracted features? D is denoted as matching degree. The feature matching matrix of the video target is given by,

$$\Gamma_t = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_1 N_{t-1} \\ g_{21} & g_{21} & \cdots & g_2 N_{t-1} \\ \vdots & & & \\ g\beta_t 1 & g\beta_t 2 & \cdots & g\beta_t N_{t-1} \end{bmatrix} \tag{19}$$

Where Γ_t is the square matrix, N_{t-1} is the number of video target at a time $t-1$, β_t is the number of measurement at a time t, g_{ij} is the entries. Here, $\{i = 1,2,\ldots,\beta_t\}$ and $\{j = 1,2,\ldots,N_{t-1}\}$. The matching degree between the video target j and the measurement i is calculated in [28]. This feature matching matrix equation is used to detect the number of objects appearing and disappearing. If β_t and N_{t-1} is equal, there is no video target is appear at time t. If N_{t-1} is less than the measurement β_t, the new target appears at the time t. If the number of video target at the time t is greater than the β_t number of video target is disappeared at the time t.

Modified H-infinity Filter for Motion state estimation

To gets the precise information about the object state such as position and velocity from the noisy estimations beginning from the single sources or multiple sources. The H-infinity filter is intended to discover the genuine location of the objects and moreover evaluate the speed of that object because the objects are moved by certain motion law. The position and velocity (speed of the object) are the two motion states of the tracking object. The dynamics state estimations are vital for recognizing the position and velocity of the object, so the representation of the dynamics is depicted by the following equations,

$$s_x(t) = s_x(t-1) + s_x(t-1)T \tag{20}$$

$$s_y(t) = s_y(t-1) + s_y(t-1)T \tag{21}$$

$$x_p(t) = x_p(t-1) + s_x(t-1)T + T^2 s_x(t-1)/2 \tag{22}$$

$$y_p(t) = y_p(t-1) + s_y(t-1)T + T^2 s_y(t-1)/2 \tag{23}$$

From the above equation, $s_x(t)$ and $s_x(t-1)$ is the horizontal velocity of the targeted object at a time t and $(t-1)$. $s_y(t)$, $s_y(t-1)$ is the vertical speed of the object at a time t and $(t-1)$. $x_p(t)$, $x_p(t-1)$ are the centroids of the targeted object at a time t and $(t-1)$ on the horizontal side, $y_p(t)$ and $y_p(t-1)$ is the centroids of the object at a time t and $(t-1)$ in the vertical side. T is the sampling time period.

The system state model can be represented by,

$$X_t = [x_p(t) \;\; y_p(t) \;\; s_x(t) \;\; s_y(t)] \tag{24}$$

The velocity $s_x(t)$ and $s_y(t)$ of the video target is computed by equating the above dynamic characteristics of the object.

$$s_x(t) = \frac{x_p(t) - x_p(t-1)}{T}, \quad s_y(t) = \frac{y_p(t) - y_p(t-1)}{T} \tag{25}$$

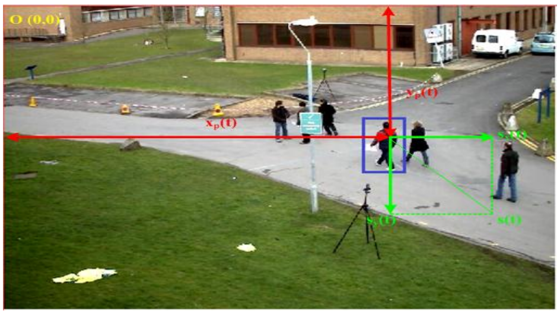

Figure 4: Motion state representation of Moving Target

In Fig (4), $s_x(t)$ is the horizontal speed of the target, $s_y(t)$ is the vertical speed of the moving target, $x_p(t)$ and $y_p(t)$ are the horizontal and vertical centroids of the moving target and

t is the time step. The origin of the image is set apart in the upper left corner of the image frame.

The motion state model and the observation model of the video target can be depicted as follows,

$$X_t = AX_{t-1} + u_t \tag{26}$$

$$y_t = CX_t \tag{27}$$

From the above equation, $A = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$. In the above equation, the motion state model and the observation model can be influenced by some indeterminate noises, for example, encompassing variety, jumbled background and motion resistance and so forth. So we have to alter the observation model and the motion state model of the video target.

$$X_t = AX_{t-1} + u_t + w_t \tag{28}$$

$$y_t = CX_t + v_t \tag{29}$$

Where, w_t is the process noise, v_t is the measurement noise and u_t is the extra process noise. A large portion of the object tracking framework generally utilized the Kalman filter to recognize the position and speed of the video target. To predict and update the position and velocity of the object, the Kalman filter requires the present estimation and past updates and furthermore, it requires the earlier information of the estimation and process noise, which are truly hard in this present reality. On account of most exceedingly bad noises, the Kalman filter does not limit the estimation blunder. So the modified H-infinity filter is utilized to limit the most pessimistic scenario estimation error. The H-infinity filter does not have any knowledge about the system model and the observation model; in any case, it has better execution and exactness. So the H-infinity filter is the powerful form of the Kalman filter.

The estimation error of the object is computed by using the following discrete-time dynamic system equations,

$$X_{t+1} = AX_t + w_t \tag{30}$$

$$Y_t = CX_t + v_t \qquad , \ t \in (0, N-1) \tag{31}$$

$$Z_t = L\ X_t \tag{32}$$

Where, w_t and v_t is the process noise and measurement noise respectively. A, C And L are the matrices with corresponding dimensions. $X_t \in R^n$ and $Y_t \in R^p$ is the state vector and the output measured vector. Z_t is the estimated output, suppose Z_t is $\tilde{Z}_t$, the estimation error is computed as $E = Z_t - \tilde{Z}_t$. Where, E is the estimation error, Z_t is the current position of the target and $\tilde{Z}_t$ is the estimate of the target position. The initial estimation error of the state is $E_i = F_0 - \tilde{F}_0$, here $\tilde{F}_0$ is the initial value of state estimation. If the process noise w_t and the measurement noise v_t are increases the estimation error E is also increased. So the cost function is designed to minimize the estimation error and the noises.

$$J = \frac{avg\|E\|_{Q_k}}{avg\left(\|w_t\|_{W_t} + \|v_t\|_{V_t}\right)} \tag{33}$$

Where, avg is the averages are taken from the sampling time T. W_t V_t and Q_t is the weighting matrices of the process noise, measurement noise, and the estimation error. The cost function pre-depicted the performance bound to avoid directly calculating the minimum value of the cost function. The estimation error problem is very difficult to solve. Consider the performance bound is $\frac{1}{\delta}$. If the performance bound $\frac{1}{\delta}$ increases, the cost function J is decreased. So the estimation error is minimized. Where J is the cost function and $\frac{1}{\delta}$ is the noise attenuation level (performance bound)? If the performance bound is impossible to measure or calculate, the H- infinity filter does not work.

When the noise w_t and v_t are in the worst case the H-infinity filter is to minimize the cost function J . To solve the optimal estimation problem, the h-infinity filter can be interpreted as a $\min\max$ problem.

$$\min_{\tilde{z}} \ \max_{(w_t, v_t, F_0)} \quad J = \frac{1}{2} avg\left[\|E\| - \frac{1}{\delta}\left(\|w_t\|_{W_t} + \|v_t\|_{V_t}\right)\right] \tag{34}$$

Where, E is the estimation error, $\frac{1}{\delta}$ is the noise attenuation level w_t and v_t is the process noise and measurement noise. The $\min\max$ problem can be solved by using the linear quadratic game approach. It is used to update the state estimate and also find out the gain of the H-infinity filter.

The updated state estimate is,

$$\tilde{X}_{t+1} = (A_t - H_t C_t)\tilde{X}_t + H_t Y_t \tag{35}$$

H-infinity filter gain is given by

$$H_t = p_t\left[I - \frac{1}{\alpha} L_t^T W_t L_t p_t + C_t^T V_t^{-1} C_t p_t\right]^{-1} C_t^T V_t^{-1} \tag{36}$$

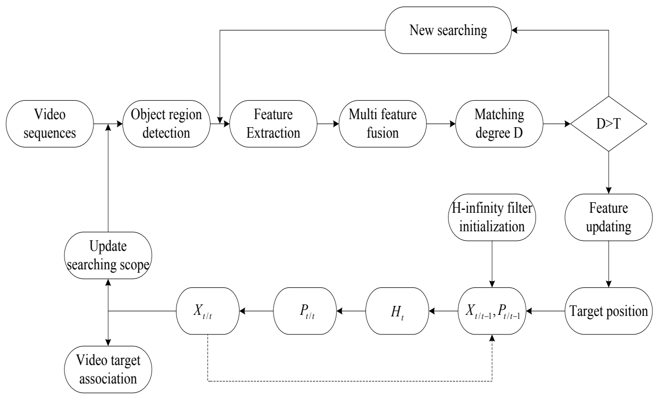

Figure 5: Block diagram of moving object tracking based on H-infinity filter

In this paper, the M video frames are taken from the comparing video groupings. The FDM method and the HOG detection methods are used to detect the target region of the object. At that point, the color spatial feature, target contour and edge gradient features are extracted from the video target, and afterward, the matching degrees between the features are estimated. The Multi feature fusion method is used to estimate the final matching degree of these features. On the off chance that the last evaluated matching degree D is greater than the user assigned threshold value T, the video target is accomplished and the object features are updated. If the matching degree is not as much as the threshold value, the same video frame is performed for new searching and the matching degree is estimated for that new searching features. The position measurements of the target are taken as the input of the H-infinity filter. The parameters of the target such as initial state estimation F_0, covariance p_t and the weighting matrices W_t, V_t and Q_t is estimated by using the H-infinity filter. To increase the searching speed, matching degree calculation and to avoid the global searching, the motion state of the target is estimated by using the H-infinity filter. If the moving target is occluded partially, the final matching degree D will be lessened, so we need to assess the threshold value estimation of the occlusion object. In the event that the moving target is completely occluded, the matching degree will be zero.

EXPERIMENTAL RESULT

The exploratory results have been performed to approve the precise position and speed of the moving object. Our proposed algorithm is utilizing the present reality of video arrangements to evaluate the performance of the tracking. These present reality video successions are taken from the PETS dataset. The moving object detection is exceptionally troublesome in the PETS video groupings in light of the fact that these sorts of video arrangements have the confounded condition, cluttered background, object occlusion and so forth. The proposed approach will be implemented and experimented using Mat lab by exploiting the available benchmark video files. Comparison with existing similar approaches in terms of different parameters will be performed for justifying the performance of our proposed approach. The parameter setting of proposed multi-object tracking algorithm is shown in Table1.

Table 1: Parameter settings of proposed Multi-object tracking algorithm

Parameter	Video sequence 1	Video sequence2	Video sequence3
Total number of frames	950	175	1150
Number of training frames	25	20	35
Number of objects	6	7	8
Number of appearing objects	5	1	7
Number of disappearing objects	2	5	6

The performance of the tracking algorithm is evaluated by using the root mean square error and average root mean square error. The position error of the video target is given by,

$$positioner\,ror_i = \sqrt{\left(Z_t - \widetilde{Z}_t\right)^2 + \left(F_0 - \widetilde{F}_0\right)^2}$$

Where Z_t, F_0 the estimated position of the video target at a time t, and $\widetilde{Z}_t, \widetilde{F}_0$ is the real position at the time t. The average root mean square error is given by,

$$RMS = \frac{1}{Frames} \sum_{i=1}^{Frames} positioner\,ror_i$$

Whereas frame is the total number of tracked video sequence frames. The position error is known as the average root mean square error, which is seen as a measurement of test result error; smaller value indicates better tracking effect [22].The parameter setting of test video sequences are shown in Table2.

Table 2: Parameters of test video sequences

Video sequences	Number of Frames	Frame size	Bytes
Six pedestrians in street	950	768×576	331776
Seven pedestrians in sparse crowd	175	352×288	132710
Eight pedestrians in corridor	1150	384×288	304128

COMPARISON ANALYSIS

In the comparison analysis, the performance and tracking effect of our proposed method is compared with other existing methods such as Kalman filter [27]. The existing method and our proposed method used the same video sequences for evaluating which method has the better tracking result. A large portion of the object tracking framework generally utilized the Kalman filter to recognize the position and speed of the video target. To predict and update the position and velocity of the object, the Kalman filter requires the present estimation and past updates and furthermore, it requires the earlier information of the estimation and process noise, which are truly hard in this present reality. On account of most exceedingly bad noises, the Kalman filter does not limit the estimation blunder. So the modified H-infinity filter is utilized to limit the most pessimistic scenario estimation error. The H-infinity filter does not have any knowledge about the system model and the observation model; in any case, it has better execution and exactness. So the H-infinity filter is the powerful form of the Kalman filter.

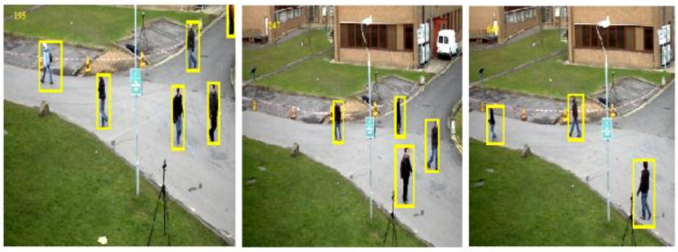

Figure 6 (a): Results of detecting the position of moving object (Video sequence 1)

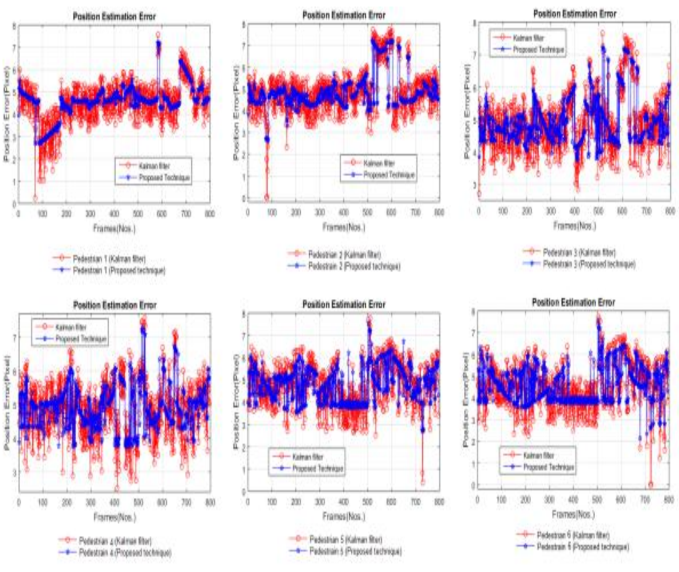

Figure 6 (b): Comparison of position error curve of our proposed method with another method (Video sequence 1)

Video Sequence 1

The required video arrangement for object recognition is taken from the CAVIAR and PETS databases. The position detection accuracy of our proposed method is tested by using this video sequences. In Fig 6 (a), the position of the target is recognized precisely. At the time of tracking, serious occlusion happens and the background of the video sequences is also changed; it appears in frames 195, 247 and 640. In the existing method, the position detection is very difficult at the time of background change. But, the background change does not affect our proposed method.

Table 3: Average Root Mean Square error for video sequence 1

	Kalman filter	Proposed method
Pedestrian 1	5.58	4.82
Pedestrian 2	5.92	4.99
Pedestrian 3	6.82	5.95
Pedestrian 4	5.85	5.65
Pedestrian 5	6.53	5.92
Pedestrian 6	6.28	5.92

The comparison of the position error of the detected target is appeared in Fig 6 (b). Here, our proposed method is contrasted with the existing Kalman filter method. The position error curve demonstrates that our proposed technique has better tracking result and preferred exactness over the Kalman filter method. The average root means square error up of the obvious object has shown up in Table 3. Our proposed procedure has the smaller average root mean square error stood out from the Kalman filter.

Video Sequence 2

The video grouping 2 is gotten from the CAVIAR database. In the 5th frame, each one of the objects is recognized. The individual by walking vanishing shows up on the frame 136. In the 160th frame, only two objects are appearing. The position error curve of the perceived object is showing up in Fig 7 (b). Here, the individual by walking 2 and 3 has small tracking error in our proposed method. The pedestrian 1, 4, 5, 6 and 7 have the greater tracking error in the Kalman filter. The average root means square error up of the perceived object has shown up in Table 4. Our proposed methodology has the tinier average root mean square error compared to the Kalman filter.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 15489-15500

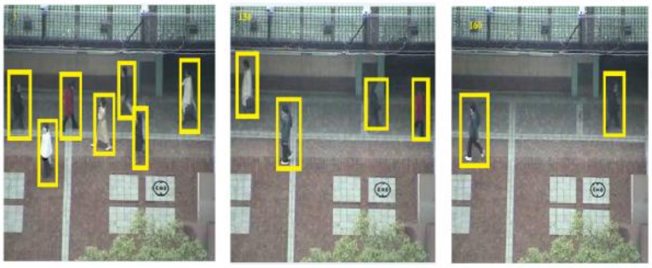

Figure 7 (a): Results of detecting the position of moving object (Video sequence 2)

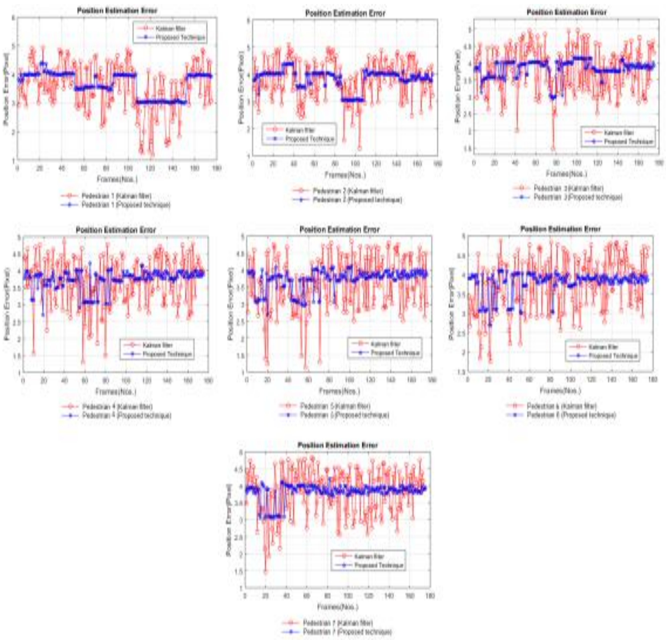

Figure 7 (b): Comparison of position error curve of our proposed method with Kalman filter method (Video sequence 2)

Table 4: Average Root Mean Square error for video sequence 2

	Kalman filter	Proposed method
Pedestrian 1	4.98	4.25
Pedestrian 2	4.97	4.17
Pedestrian 3	5.72	4.32
Pedestrian 4	4.83	4.07
Pedestrian 5	4.98	4.13
Pedestrian 6	4.85	4.08
Pedestrian 7	4.83	4.02

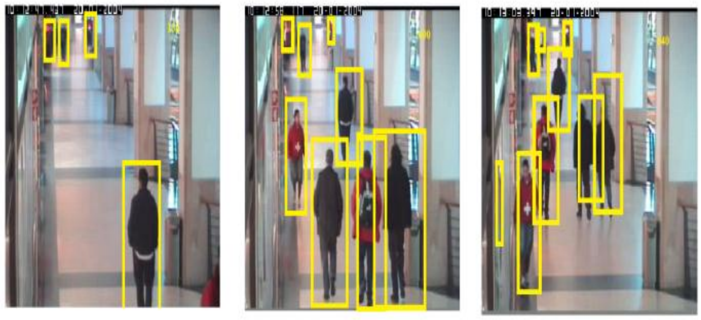

Figure 7 (a): Results of detecting the position of moving object (Video sequence 3)

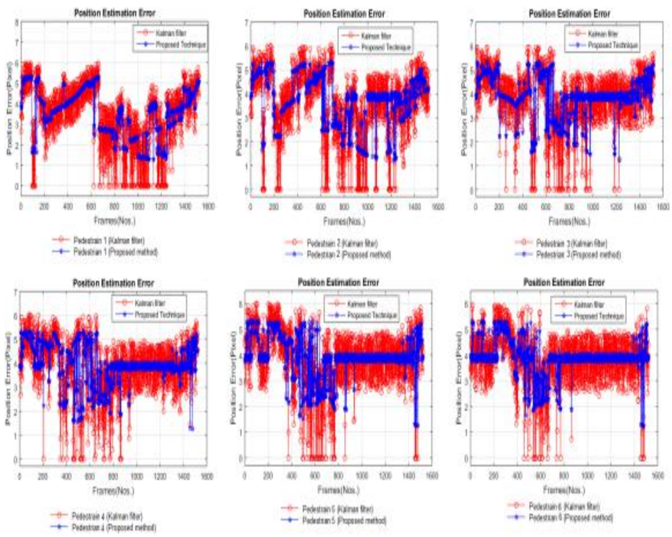

Figure 7 (b): Comparison of position error curve of our proposed method with Kalman filter method (Video sequence 3)

Video Sequence 3

The video sequence 3 is taken from the CAVIAR database appeared in FIG 8 (a). At first, just three people are showing up in the frame 136. At the time of tracking the moving object, extreme occlusion happens in the frame 400 and 640. The position error curve of the distinguished object is shown in Fig 8 (b). In our proposed method, Pedestrian 1, 3, 5 and 7 has the preferable tracking outcome and accuracy over the Kalman filter method. The average root means square error up of the apparent object has appeared in Table 5. Our proposed procedure has the smaller average root mean square error appeared differently in relation to the Kalman filter.

Table 5: Average Root Mean Square error for video sequence 3

	Kalman filter	Proposed method
Pedestrian 1	5.56	4.83
Pedestrian 2	6.03	5.89
Pedestrian 3	5.98	5.43
Pedestrian 4	5.98	4.89
Pedestrian 5	5.89	4.78
Pedestrian 6	4.87	4.53

CONCLUSION

An advanced H-infinity filter approach is balanced in this paper for multiple objects tracking in video groupings. The dedication of our work is recorded as takes after. (1) Before all else arrange, we embrace the FDM technique and the HOG strategy to separate the object region of the video frame. (2) We expel the color spatial feature, Edge oriented feature and texture feature from the object region and after that, the matching degree of these features are calculated by utilizing the Multi-Feature Fusion method. On the off chance that the matching degree is more conspicuous than the customer decided threshold value of the object, the feature of the object is updated and the position of the target is identified. Otherwise again the new searching is reiterated and the matching degree is calculated for that new searching. After updating the position of the target, the modified H-infinity filter is utilized to estimate the motion state of that target. The modified H-infinity filter is utilized to limit the most pessimistic scenario estimation error. The H-infinity filter does not have any knowledge about the system model and the observation model; in any case, it has better execution and exactness. So the H-infinity filter is the powerful form of the other existing method.

REFERENCES

[1] Y. Sui and L. Zhang, "Visual Tracking via Locally Structured Gaussian Process Regression", *IEEE Signal Processing Letters*, vol. 22(9), pp. 1331-1335, 2015.

[2] T. Zhou, X. He, K. Xie, K. Fu, J. Zhang and J. Yang, "Robust visual tracking via efficient manifold ranking with low-dimensional compressive features", *Pattern Recognition*, vol. 48(8), pp. 2459-2473, 2015.

[3] S. Bae and K. Yoon, "Confidence-Based Data Association and Discriminative Deep Appearance Learning for Robust Online Multi-Object Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-1, 2017.

[4] P. Athanasopoulos and D. Albright, "A Perceptual Learning Approach to the Whorfian Hypothesis: Supervised Classification of Motion", *Language Learning*, vol. 66(3), pp. 666-689, 2016.

[5] J. Barrett, H. Garcke, and R. Nürnberg, "Phase Field Models Versus Parametric Front Tracking Methods: Are They Accurate and Computationally Efficient?", *Communications in Computational Physics*, vol. 15(02), pp. 506-555, 2014.

[6] K. Han, "Image object tracking based on temporal context and MOSSE", *Cluster Computing*, 2017.

[7] G. Walia, S. Raza, A. Gupta, R. Asthana and K. Singh, "A novel approach of multi-stage tracking for precise localization of target in video sequences", *Expert Systems with Applications*, vol. 78, pp. 208-224, 2017.

[8] V. Karavasilis, C. Nikou and A. Likas, "Real-time visual tracking using a spatially weighted von Mises mixture model", *Pattern Recognition Letters*, vol. 90, pp. 50-57, 2017.

[9] G. Wu and W. Kang, "Exploiting superpixel and hybrid hash for kernel-based visual tracking", *Pattern Recognition*, vol. 68, pp. 175-190, 2017.

[10] J. Wang, H. Zhu, S. Yu and C. Fan, "Object tracking using color-feature guided network generalization and tailored feature fusion", *Neurocomputing*, vol. 238, pp. 387-398, 2017.

[11] J. Jeong, T. Yoon, and J. Park, "Mean shift tracker combined with an online learning-based detector and Kalman filtering for real-time tracking", *Expert Systems with Applications*, vol. 79, pp. 194-206, 2017.

[12] A. Dehghan and M. Shah, "Binary Quadratic Programming for Online Tracking of Hundreds of People in Extremely Crowded Scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-1, 2017.

[13] L. Hou, W. Wan, K. Lee, J. Hwang, G. Okopal and J. Pitton, "Robust Human Tracking Based on DPM Constrained Multiple-Kernel from a Moving Camera", *Journal of Signal Processing Systems*, vol. 86, (1), pp. 27-39, 2015.

[14] I. Markovic and I. Petrovic, "Bayesian Sensor Fusion Methods for Dynamic Object Tracking - A Comparative Study", *Automatica – Journal for Control, Measurement, Electronics, Computing and Communications*, vol. 55(4), 2014.

[15] B. Liu, "Monocular 3D Hand Tracking Using Particle Filter with Partitioned Sampling", *Journal of Information and Computational Science*, vol. 10(15), pp. 4733-4742, 2013.

[16] K. Ayusawa, Y. Ikegami and Y. Nakamura, "Simultaneous global inverse kinematics and geometric parameter identification of human skeletal model from motion capture data", *Mechanism and Machine Theory*, vol. 74, pp. 274-284, 2014.

[17] M. Gianni, F. Ferri, M. Menna and F. Pirri, "Adaptive Robust Three-dimensional Trajectory Tracking for Actively Articulated Tracked Vehicles", *Journal of Field Robotics*, vol. 33(7), pp. 901-930, 2015.

[18] A. Altaf and A. Raeisi, "Presenting an Effective Algorithm for Tracking of Moving Object based on Support Vector Machine", *Indian Journal of Science and Technology*, vol. 8, no. 17, 2015.

[19] A. Mazinan, "Automated probabilistic estimation framework in real-time simultaneous objects tracking", *Computers & Electrical Engineering*, vol. 40(8), pp. 186-203, 2014.

[20] J. Blevins, "Sequential Monte Carlo Methods for Estimating Dynamic Microeconomic Models", *Journal of Applied Econometrics*, vol. 31(5), pp. 773-804, 2015.

[21] J. Wu, S. Hu, and Y. Wang, "Adaptive multi-feature visual tracking in a probability-hypothesis-density filtering framework", *Signal Processing*, vol. 93(11), pp. 2915-2926, 2013.

[22] H. Li, Y. Liu, C. Wang, S. Zhang and X. Cui, "Tracking Algorithm of Multiple Pedestrians Based on Particle Filters in Video Sequences", *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1-17, 2016.

[23] A. De Freitas, L. Mihaylova, A. Gning, D. Angelova and V. Kadirkamanathan, "Autonomous crowds tracking with box particle filtering and convolution particle filtering", *Automatica*, vol. 69, pp. 380-394, 2016.

[24] X. Wang, H. Zhang, X. Jiang, and Y. Yang, "Target tracking based on the extended H-infinity filter in wireless sensor networks", *Journal of Control Theory and Applications*, vol. 9(4), pp. 479-486, 2011.

[25] H. Morimitsu, I. Bloch and R. Cesar-Jr, "Exploring structure for long-term tracking of multiple objects in sports videos", *Computer Vision and Image Understanding*, 2016.

[26] Y. Zhao and H. Pei, "Model Update Particle Filter for Multiple Objects Detection and Tracking", *International Journal of Computational Intelligence Systems*, vol. 5(5), pp. 964-974, 2012.

[27] Z. Zhou, D. Wu, and Z. Zhu, "Object tracking based on Kalman particle filter with LSSVR", *Optik - International Journal for Light and Electron Optics,* vol. 127(2), pp. 613-619, 2016.

[28] H. Wang and S. Nguang, "Robust Video Target Tracking Based on Multi-feature Fusion and H∞ Filtering", *International Journal of Computer and Communication Engineering*, vol. 5(2), pp. 79-98, 2016.

[29] M. Babaee, Y. You and G. Rigoll, "Combined segmentation, reconstruction, and tracking of multiple targets in multi-view video sequences", *Computer Vision and Image Understanding*, vol. 154, pp. 166-181, 2017.