

A real-time multiple vehicle tracking method for traffic congestion identification

Xiaoyu Zhang¹, Shiqiang Hu¹, Huanlong Zhang² and Xing Hu¹

¹ School of Aeronautics and Astronautics, Shanghai Jiao Tong University,
Shanghai, 200240, China
[e-mail: sqhu@sjtu.edu.cn]

² College of electric and information engineering, Zhengzhou University of Light Industry,
Zhengzhou, 450002, China
[e-mail: zh_lit@sjtu.edu.cn]

*Corresponding author: Shiqiang Hu

*Received July 17, 2015; revised November 3, 2015; accepted April 20, 2016;
published June 30, 2016*

Abstract

Traffic congestion is a severe problem in many modern cities around the world. Real-time and accurate traffic congestion identification can provide the advanced traffic management systems with a reliable basis to take measurements. The most used data sources for traffic congestion are loop detector, GPS data, and video surveillance. Video based traffic monitoring systems have gained much attention due to their enormous advantages, such as low cost, flexibility to redesign the system and providing a rich information source for human understanding. In general, most existing video based systems for monitoring road traffic rely on stationary cameras and multiple vehicle tracking method. However, most commonly used multiple vehicle tracking methods are lack of effective track initiation schemes. Based on the motion of the vehicle usually obeys constant velocity model, a novel vehicle recognition method is proposed. The state of recognized vehicle is sent to the GM-PHD filter as birth target. In this way, we relieve the insensitive of GM-PHD filter for new entering vehicle. Combining with the advanced vehicle detection and data association techniques, this multiple vehicle tracking method is used to identify traffic congestion. It can be implemented in real-time with high accuracy and robustness. The advantages of our proposed method are validated on four real traffic data.

Keywords: Track initiation technique, Multiple vehicle tracking, Track score, Ellipsoidal gate, Traffic congestion

This paper is jointly supported by the National Natural Science Foundation of China (Grant No.61374161, 61503173), China Aviation Science Foundation (Grant No. 20142057006).

1. Introduction

With the high-speed development of city construction, modern societies with well-planned road management systems and sufficient infrastructures for transportation still face the problem of traffic congestion. This results in loss of travel time, huge societal and economic costs. It has been recently realized [1] that mere infrastructure expansion cannot provide a complete solution for those problems because of economic and environmental reasons or in metropolitan areas, simply due to lack of space. An alternative would be to develop an intelligent transportation system which more efficiently uses of the existing infrastructure [2, 3]. Many approaches have emerged with the aim of limiting traffic congestion. One of approaches amounts to provide aggregate traffic information (such as velocity, the intensity of traffic volumes and lane occupancy rates) to transportation authorities, which in turn feed this information into advanced traffic management systems to take effective measures [4].

Detection via video processing technique is one of the most attractive new technologies as it offers opportunities for performing substantially more complex tasks and providing more information [5]. For example, video based traffic monitoring systems are easy to install and upgrade, which offer the flexibility to redesign the system and their functionality by simply changing the system algorithms. Furthermore, they provide a rich information source for human understanding. Therefore, developing real-time traffic parameter surveillance systems based on video aiming to derive reliable and robust traffic state information has attracted a lot of attention during the past decade [6, 7].

At present traffic congestion can be identified by classification method. Anto B. Chan and Nuno Vasconcelos [8] proposed a framework for the classification of visual processes that are best modeled with spatio-temporal autoregressive models. The proposed framework combines the modeling power of a family of models known as dynamic textures and the generalization guarantees, for classification, of the support vector machine classifier. This combination is achieved by the derivation of a new probabilistic kernel based on the Kullback-Leibler divergence between Gaussian-Markov processes. The kernels cover a large variety of video classification problems, including the cases where classes can differ in both appearance and motion and the cases where appearance is similar for all classes and only motion is discriminant. In addition, Anto B. Chan and Nuno Vasconcelos [9] proposed to model the entire motion field as a dynamic texture, which is an auto-regressive stochastic process with both a spatial and a temporal component. They utilize this representation to realize better performance under environmental conditions such as variable lighting and shadows without the requirement of segmentation and tracking. A drawback of those approaches is the large computational load in fitting the model might make this method impractical for application to real-time traffic monitoring [10]. Based on [9], Derpanis and Wildes [10] proposed a system, which achieves good performance, is amenable to computationally efficient realization. Alternatively, most extant approaches classify traffic videos using a combination of segmentation and tracking. In this paper, we focus on tracking all vehicles from entering to leaving on the road. Based on those requirements, the adopted methods must have the ability in track initiation and termination. Up to now, these methods can be grouped into two sets: vehicle recognition before tracking (VRT) methods and association based tracking (ABT) methods.

The VRT methods require recognition each vehicle on the road, then tracking them respectively. Sivaraman et al. [11] proposed a general active-learning framework for on-road

vehicle recognition and tracking method. After each vehicle is recognized, an extended condensation algorithm [12] is utilized to track multiple vehicles with vehicle entering and leaving in the field of view. Xia et al. [13] proposed an improved CAMShift approach to realize multi-vehicle tracking using video Cameras. It uses a background subtraction method for automatically initializing vehicle tracking. Then an improved CAMShift method is used to track each potential vehicle. Rad and Jamzad [14] proposed a real-time classification and tracking of multiple vehicles method in highways. They utilize sequence frame differencing, moving edge detection, hybrid approach that combines features from both contour based and region based methods to achieve vehicle recognition. Then Kalman filter is used to track each recognized vehicle. Those methods can perform well once the initializing tracks are right. However, the drawback of [11] lies in it needs off-line training step. In real world traffic scenes, the vehicles passing by are not only with various types of color, shape and style, but also with different views according to their distance from the processor. It is difficult to construct a detector that can detect all vehicles in real scenes. That is to say, not all of vehicle passing by can be tracked by this method. The drawback of [13, 14] are that they might track background which is detected by the detector as a vehicle. The reason behind them are that with the development of detection techniques, such as background subtraction algorithms [15-17], deformed part models [18], and histograms of sparse codes for object detection [19], perfect detection is still a challenging work in real application. In other words, miss detection and false alarm are inevitable. If the initialization is imprecise or even a false alarm, the adopted tracker would be problematic.

The ABT methods usually first localize objects in each frame and then link these object hypotheses into trajectories without any initialization labeling. Semertzidis et al. [20] proposed a video sensor network for real-time traffic monitoring and surveillance. In tracking unit, multiple hypothesis tracking (MHT) [21, 22] algorithm is implemented to realize vehicle recognizing and tracking from detection results. MHT has the mechanism of track initiation, maintenance and pruning. It allows the use of measurements that arrive in the future to resolve the uncertainty in the correct association of measurements and targets at present. However, the number of hypotheses in them grows exponentially over time and the required computational costs could render the implementation of MHT infeasible. Garcia et al. [23] proposed a visual feature tracking based on PHD filter. Comparing with the widely used MHT algorithm in multiple target tracking, Panta et al. [24] showed that the PHD filter outperforms MHT in tracking accuracy and computation. However, one limitation of PHD filter is that it is insensitive to birth target with unknown position [25, 26]. In other words, it needs a large time delay to initialize track of the birth target or even cannot initialize track them as the birth targets moving across the scene. Detection guided GM-PHD (D-GMPHD) filter [25] is proposed to relieve the insensitive of the PHD filter to birth target. It uses measurement information forming many track hypotheses at first. Then it confirms or deletes the track by a score based track initiation method. At last, the confirmed track hypothesis is used to give the state of birth target to GM-PHD filter.

Based on the characteristics of the D-GMPHD filter stated above, we introduce it into an image based multiple vehicle tracking. Motion of vehicle usually obeys constant velocity (CV) model. This property is used to identify the vehicle out of noisy measurements. Subsequently, the state of the vehicle is transmitted to GM-PHD filter as the birth target. Through this way, vehicles on the road can be tracked more timely than GM-PHD filter without vehicle recognition aided. The vehicle early tracking ability enables the traffic congestion identification in time, which can provide advanced traffic management with necessary traffic information.

The rest of the paper is organized as follows. Section 2 gives brief reviews of GM-PHD filter and track score based track initiation method. Section 3 presents the overall system for traffic congestion identification. Experimental results of the proposed system on real scenes are provided in Section 4. For illustration, Section 4 also discusses the performance of the proposed method, as compared with the original version of GM-PHD filter and Boosted particle filter.

2. GM-PHD filter and track score based track initiation technique

2.1 GM-PHD filter

In a multiple-target environment, the number of targets changes due to targets appearing and disappearing. Usually, not all of the existing targets are detected by the sensor. Moreover, the sensor also receives clutter not originating from any target. The objective of multiple-target tracking is to jointly estimate, at each time step, the number of targets and their states from a sequence of noisy and cluttered observation sets. The random finite sets (RFS) approach to multi-target tracking is an emerging and promising alternative to traditional association-based methods [27, 28]. In the RFS formulation, the collection of individual targets is treated as set-valued state, and the collection of individual observations is treated as a set-valued observation. Modeling set-valued states and set-valued observations as RFSs allows the problem of dynamically estimating multiple targets in the presence of clutter and association uncertainty to be cast in a Bayesian filtering framework. In order to alleviate the computational intractability in multi-target Bayes filter, PHD filter, a recursion that propagates the first-order statistical moment of the RFS of states in time is proposed [27]. Under the linear and Gaussian assumptions on the target dynamics and birth process, GM-PHD filter shows its excellent state estimation ability [28].

For an RFS X on \bar{X} with a probability distribution P , its first order moment is a nonnegative function v on \bar{X} , called the intensity, such that for each region $S \subseteq \bar{X}$,

$$\int |X \cap S| P(dX) = \int_S v(x) dx \quad (1)$$

Let v_k and $v_{k|k-1}$ denote posterior intensity and predicted intensity, respectively. Under the following three assumptions A.1-A.3, it can be shown that v_k can be propagated in time via the PHD recursion by Eqs. (2) and (3).

A.1: Each target evolves and generates observations independently of one another.

A.2: Clutter is Poisson and independent of target-originated measurements.

A.3: The predicted multiple-target RFS governed by predicted density is Poisson.

$$v_{k|k-1}(x) = \int p_{S,k}(\zeta) f_{k|k-1}(x|\zeta) v_{k-1}(\zeta) d\zeta + \int \beta_{k|k-1}(x|\zeta) v_{k-1}(\zeta) d\zeta + \gamma_k(x) \quad (2)$$

$$v_k(x) = [1 - p_{D,k}(x)] v_{k|k-1}(x) + \sum_{z \in Z_{k+1}} \frac{P_{D,k}(x) g_k(z|x) v_{k|k-1}(x)}{\kappa_k(z) + \int P_{D,k}(\zeta) g_k(z|\zeta) v_{k|k-1}(\zeta) d\zeta} \quad (3)$$

Where $\gamma_k(\cdot)$ is the intensity of the birth RFS at time k ; $\beta_{k|k-1}(\cdot|\zeta)$ is the intensity of the spawned RFS spawned at time k by a target with previous state ζ ; $p_{S,k}(\zeta)$ is the probability

that a target still exists at time k given that its previous state is ζ ; $P_{D,k}(x)$ is the probability of detection given a state x at time k ; $\kappa_k(\cdot)$ is the intensity of clutter RFS at time k .

Under the assumption of linear Gaussian multiple-target model, Vo and MA [28] derived a closed-form recursions for the weights and states of the constituent Gaussian components of the posterior intensity. It utilizes the state of each target to construct a Gaussian component with each measurement. If only one state of two targets is known, the state will be updated by two measurements respectively. If the derived state is in the vicinity of the target with unknown state, then the weight of the derived state will rise quickly. Otherwise, the weight of derived state will rise slowly. As the position of the birth target is not necessarily near the existing state of the target, GM-PHD filter is insensitive to birth target.

2.2 Score based track initiation method

In order to reduce the time needed to track all targets in the field of view, track initiation technique is used to boost the sensitivity of GM-PHD filter to birth target in the presence of spurious measurements. One of track initiation techniques named after detection guided GM-PHD (D-GMPHD) filter [25] is to construct or update track hypotheses with track score and use the track score to confirm, keep or delete each track hypothesis. The confirmed track hypothesis becomes the state of birth target to GM-PHD filter.

2.2.1 Track score in point target tracking

Track score L proposed by [29] is the log likelihood ratio of P_T to P_F . Where P_T is the probability of the received data is a true target, P_F is the probability of the received data is false alarm. L can be partitioned into a product of three terms, L_0 , L_K and L_S , which represent priori, kinematic and signal-related contributions, respectively. Given \mathcal{K} scans of data, $L(\mathcal{K})$ can be given by

$$L(\mathcal{K}) = \ln[L_0] + \sum_{k=1}^{\mathcal{K}} [L_K(k) + L_S(k)] \quad (4)$$

First, considering the kinematic term, assume a Gaussian distribution for the true target and a uniform distribution over the measurement volume, V_c , for the false target. Then,

$$L_K(k) = \ln \frac{e^{(-d(k)^2/2)} / \left[(2\pi)^{M/2} \sqrt{|S(k)|} \right]}{1/V_c} \quad (5)$$

$$P(k|k-1) = Q + FP(k-1)F^T, \quad S(k) = HP(k|k-1)H^T + R \quad (6)$$

$$d(k)^2 = \tilde{z}(k)^T S(k)^{-1} \tilde{z}(k), \quad \tilde{z}(k) = z(k) - HFx(k-1)$$

Where M is the measurement dimension. $S(k)$ is the measurement residual covariance matrix at time k , $d(k)^2$ is the normalized statistical distance for the measurement defined in terms of $S(k)$ and the measurement residual vector $\tilde{z}(k)$ at time k . F is the state transition matrix, H is the measurement matrix, $P(k-1)$ is the estimated process noise covariance at time $k-1$, Q is the process noise covariance, R is the observation noise covariance.

Next, considering the signal-related term, if a detection is received at time k , then

$$L_S(k) = \ln \frac{P_D}{P_{FA}} \quad (7)$$

Using Eq. (5) and Eq. (7) gives a recursive form for the computation of the track score, defined by Eq. (4):

$$L(k) = L(k-1) + \Delta L(k) \quad (8)$$

Where

$$\Delta L(k) = \begin{cases} \ln \frac{1-P_D}{1-P_{FA}} & \text{no track update at time } k \\ L_K(k) + \ln \frac{P_D}{P_{FA}} & \text{track up date at time } k \end{cases} \quad (9)$$

2.2.2 Score-based track confirmation and deletion

Sequential probability ratio test (SPRT), which is strictly proved by Wald [30] that needs a much smaller number of observations than all competitors, is used to classify track hypotheses according to their track scores. Given the upper and lower thresholds η_2 and η_1 , the alternatives to confirm the track hypothesis, delete the track hypothesis, or continue the track hypothesis at time k are shown below:

$$\begin{aligned} L(k) &\geq \eta_2; && \text{declare track confirmation} \\ \eta_1 &< L(k) < \eta_2; && \text{continue track} \\ L(k) &\leq \eta_1; && \text{delete track} \end{aligned} \quad (10)$$

Following the standard SPRT formulation, the thresholds are defined as $\eta_2 = \ln \frac{1-\beta}{\alpha}$,

$\eta_1 = \ln \frac{\beta}{1-\alpha}$. Where α and β are the probability of false track confirmation and true track

deletion, respectively. Note that the SPRT threshold values are chosen on the assumption that the initial track score is zero. The initial track score is entirely based on the first observation in the track. Thus, there is no kinematic contribution to the initial track score. As the signal-related datum is that a detection (or a miss) occurred, then the L at track initiation is

$$L(1) = \ln[P_0(H_1)] + \ln[P_D / P_{FA}] \quad (11)$$

Where $P_0(H_1)$ is the prior probability of measurement is true target. Thus, if an initial track score value is other than zero, this initial value should be add to η_1 and η_2 .

2.2.3 Vehicle recognition in image based tracking

In image based tracking, assume that there are N_k measurements at time k . The centroids of N_k measurements at time k are treated as the measurement set $Z = \{z^i(k)\}_{i=1}^{N_k}$. Hence the kinematic term in image based tracking is the same as point targets tracking. In radar application, P_D is defined the probability of a target that is declared, when target is in fact present. P_{FA} is defined as the probability of a target that is declared, when target is in fact not present [31]. Hence, P_D and P_{FA} in image based tracking can be obtained by (12)

$$P_D = \frac{TP}{TP + FN}, P_{FA} = \frac{FP}{FP + TN} \quad (12)$$

Where the number of true positives (TP), counts the number of correctly detected foreground pixels. The number of false positives (FP) counts the number of background pixels incorrectly classified as foreground. The number of true negatives (TN) counts the number of correctly classified background pixels. The number of false negatives (FN) counts the number of foreground pixels incorrectly classified as background.

Generally speaking, motions of vehicles on the road match to CV model, while the spurious measurements match to the random motion model. We form a set of track hypotheses based on CV model and measurements. Instead of prior information of vehicle shape and color, we recognize vehicle from the noised measurements by setting H and F in (6) according to CV model. Before describing the forming and updating mode of a track hypothesis, we give the state transition and measurement models.

Assume that the state $x(k) = [p_x(k), p_y(k), \dot{p}_x(k), \dot{p}_y(k)]$ of each vehicle at time k consists of position $(p_x(k), p_y(k))$ and velocity $(\dot{p}_x(k), \dot{p}_y(k))$, while the measurements of vehicles are a noisy version of the positions. Each vehicle follows a linear Gaussian dynamical model and the measurement model is linear Gaussian, i.e.,

$$x(k) = Fx(k-1) + Gw(k-1), z(k) = Hx(k) + v(k) \quad (13)$$

where F is the state transition matrix, $w(k-1)$ is the zero-mean Gaussian white process noise with standard deviation σ_w , H is the measurement matrix, and $v(k)$ is the zero-mean Gaussian white observation noise with standard deviation σ_v . F, G, H, Q, R are given by

$$F = \begin{bmatrix} I_2 & \Delta I_2 \\ 0_2 & I_2 \end{bmatrix}, G = \begin{bmatrix} \Delta^2/2 & 0 \\ 0 & \Delta^2/2 \\ \Delta & 0 \\ 0 & \Delta \end{bmatrix}, H = [I_2 \quad 0_2], Q = G\sigma_w\sigma_w^T G^T, R = \sigma_v\sigma_v^T \quad (14)$$

where I_n and 0_n denote, respectively, the $n \times n$ identity and zero matrices, Δ is the sampling period.

Given track hypotheses set $\{T^j(k-1)\}_{j=1}^{J_{k-1}}$ surviving from time $k-1$, measurements $\{z^i(k)\}_{i=1}^{N_k - N_k^0}$ that use for updating track hypotheses set, measurements $Z \setminus \{z\} = \{z_0^j(k)\}_{j=1}^{N_k^0}$ that do not use for updating track hypotheses set at time k , the vehicle recognition method can be described as follows:

Step 1. Update $\{T^j(k-1)\}_{j=1}^{J_{k-1}}$ to form new surviving track hypotheses set $\{T_s^\ell(k)\}_{\ell=1}^{J_{\beta,k}}$ by

Eq.(15). Meanwhile update track score set $\{L^j(k-1)\}_{j=1}^{J_{k-1}}$ to form a new surviving track

score set $\{L_s^\ell(k)\}_{\ell=1}^{J_{\beta,k}}$ according to Eq. (8).

$$\begin{aligned} P^j(k|k-1) &= Q + FP^j(k-1)F^T, S = R + HP^j(k|k-1)H^T, \\ K^j &= P^j(k|k-1)H^T S^{-1}, X^j(k|k-1) = FX^j(k-1) \end{aligned} \quad (15)$$

If there is $z^i(k)$ that falls into the ellipsoidal gate [29] of $T^j(k-1)$

$T^j(k-1)$ is updated according to Eq. (16)

$$\begin{aligned} X_S^\ell(k) &= X^j(k|k-1) + K^j(z^i(k) - HX^j(k|k-1)) \\ P_S^\ell(k) &= [I - K^jH]P^j(k|k-1) \end{aligned} \quad (16)$$

Else

$T^j(k-1)$ is updated according to Eq. (17)

$$X_S^\ell(k) = X^j(k|k-1), P_S^\ell(k) = P^j(k|k-1) \quad (17)$$

EndIf

Step 2. Create new track hypotheses set $\{T_\gamma^\ell\}_{\ell=1}^{J_{\gamma,k}}$ from $\{z_0^i(k)\}_{i=1}^{N_k^0}$ and $\{z_0^j(k-1)\}_{j=1}^{N_{k-1}^0}$ by two-point differencing method [32]. $T_\gamma^\ell(k)$ at time k is formed according to Eq. (18).

The corresponding new track score set $\{L_\gamma^\ell(k) = 0\}_{\ell=1}^{J_{\gamma,k}}$.

$$X_\gamma^\ell(k) = \begin{bmatrix} z_{0,x}^i(k) \\ z_{0,y}^i(k) \\ z_{0,x}^i(k) - z_{0,x}^j(k-1) \\ z_{0,y}^i(k) - z_{0,y}^j(k-1) \end{bmatrix}, P_\gamma^\ell(k) = \begin{bmatrix} R & 0 & R/T & 0 \\ 0 & R & 0 & R/T \\ R/T & 0 & 2R/T^2 & 0 \\ 0 & R/T & 0 & 2R/T^2 \end{bmatrix} \quad (18)$$

Step 3. Combining updated track hypotheses set and new formed track hypotheses set, track hypotheses set $\{T^\ell(k)\}_{\ell=1}^{J_k^i}$ at time k can be obtained by Eq. (19). Combining updated track scores and new formed track scores, track scores $\{L^\ell(k)\}_{\ell=1}^{J_k^i}$ at time k can be obtained by Eq.(20)

$$\{T^\ell(k)\}_{\ell=1}^{J_k^i} = \{T_\gamma^i(k)\}_{i=1}^{J_{\gamma,k}} \cup \{T_S^j(k)\}_{j=1}^{J_{\beta,k}} \quad (19)$$

$$\{L^\ell(k)\}_{\ell=1}^{J_k^i} = \{L_\gamma^i(k)\}_{i=1}^{J_{\gamma,k}} \cup \{L_S^j(k)\}_{j=1}^{J_{\beta,k}} \quad (20)$$

Step 4. Track hypotheses confirm, keep and delete according to Eq. (10). Only the track hypotheses kept are continuing tests. Together with track scores they are represented as follows:

$$\{T^\ell(k), L^\ell(k)\}_{\ell=1}^{J_k^i} \quad (21)$$

Step 5. The state and covariance of confirmed track hypotheses in step 4 are recognized as vehicles.

3. Traffic congestion identification method

In order to realize traffic congestion identification via video processing technique, the whole system mainly consists of four parts: a projective transformation, vehicle detection, multiple vehicle tracking, and traffic congestion identification. The block diagram of the proposed system in traffic congestion identification is shown in Fig. 1.

Traffic activity collected by a camera is a projection of three-dimensional plane space scenes in two-dimensional plane space. It is necessary to obtain the actual movement of vehicle to estimate the velocity of the vehicle. Projective transformation [33, 34] makes it possible to use the pixel distance to reflect the true distance in the real scene.

The vehicle detection method used here is ViBe [17], followed by morphological filtering. In this way, the need of prior information about the shape, color and other features of vehicles is avoided.

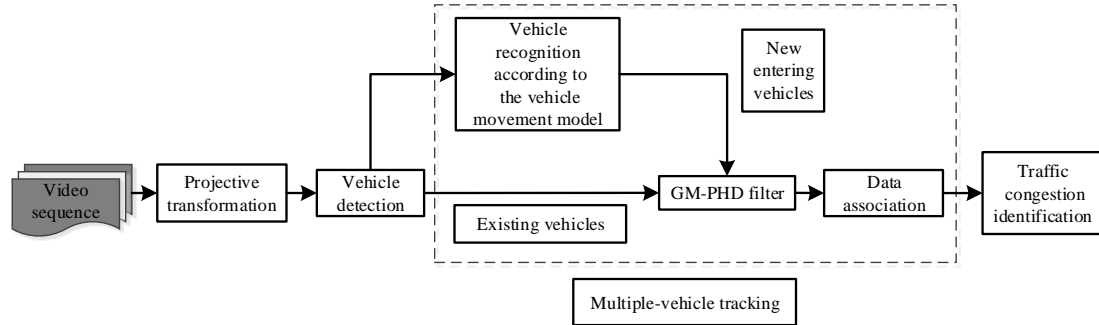


Fig. 1. Traffic congestion identification system

Vehicle recognition by their motion model is utilized to assist GM-PHD filter in tracking new entering vehicle earlier.

As D-GMPHD filter does not have the mechanism of trajectory, ellipsoidal gate [29] is used to form the trajectory for each vehicle. Ellipsoidal gate is a technique to provide multiple-target data association with a reliable basis. The major axis is direction of estimated movement. The ellipsoidal validation gate is optimal for a linear observation model with additive noise. The validity of a measurement $z(k)$ to state $x(k - 1)$ is determined from its residual with the predicted observation. The residual vector is shown in Eq. (6). If measurement $z(k)$ meets with Eq. (22), then $z(k)$ is used to associate with $x(k - 1)$.

$$d(k)^2 \leq \gamma \tag{22}$$

The threshold γ , for a measurement dimension M , can be computed efficiently since the $d(k)^2$ follows a chi-square probability density function. Thus, for the probability that $c\%$ of true association is accepted, γ is obtained from Eq. (23).

$$\frac{c}{100} = P\left(\frac{M}{2}, \frac{\gamma}{2}\right) \tag{23}$$

Where

$$p(a, b) = \frac{1}{\Gamma(a)} \int_0^b e^{-t} t^{a-1} dt \tag{24}$$

is the incomplete gamma function.

The velocity of the i th vehicle $vs_i(k)$ at time k is obtained based on its trajectory, i.e., the Euclidean distance of estimated position between two continues frames times with frame rate Fn . It is shown in Eq. (25).

$$vs_i(k) = \left\| Hx^i(k) - Hx^i(k-1) \right\|_2 \times Fn \quad (25)$$

The traffic congestion is identified by the average traffic speed in the detection area. In order to timely identify the traffic congestion, we only identify traffic congestion in the area, which is the exit road. Assume that there are N vehicles in the detection area and T_1 is the threshold to distinguish traffic congestion, the traffic congestion can be identified as follows:

$$\begin{cases} \frac{\sum_{i=1}^N vs_i(k)}{N} < T_1, & \text{traffic congestion} \\ \text{otherwise} & \text{no traffic congestion} \end{cases} \quad (26)$$

4. Experimental results and analysis

This section reports on a set of qualitative and quantitative experiments comparing D-GMPHD filter with GM-PHD filter and Boosted Particle Filter (BPF) [35]. BPF combines the strengths of Adaboost for object detection with those of mixture particle filters for multiple-object tracking. The combination is achieved by forming the proposal distribution for the particle filter from a mixture of the Adaboost detection in the current frame and the dynamic model predicted from the previous time step. As Adaboost detection needs many prior information about the vehicle, we use motion detection as an alternative. When one or more vehicles enter into the scene, they are detected by their motion and automatically initialized with an observation model. When one leaves the scene, the observation likelihood of it will drop rapidly. Hence the tracking will be terminated. The number of particles for each vehicle is 30. The size of all test images is resized to 320×240 .

Four video sequences representing traffic with and without congestion are used to test the effectiveness of the proposed method in traffic congestion identification. The first three test videos collected from a single stationary traffic camera in Shanghai, China. The fourth test video comes from NGSIM database [36]. Some important parameters used in all experiments are presented in Table 1. Where P_D and P_{FA} used in each test are obtained by Eq. (12). The measurement dimension of our tests is only the position information of the vehicle, so 2 is set. Other parameters can have a slight fluctuation. α and β determine the upper and lower thresholds for the track is confirmed, continued or deleted. In our tests, we do not hope that the clutter to be confirmed, α is set very low. We hope that the true track is confirmed instead of deleted. β is set low. More details can refer to [29]. σ_w is chosen according to the difference between the assumed motion model and the actual movement of vehicles. we set σ_w to 5 in both directions to ensure that the constant velocity model can adapt to the deceleration, acceleration and stop cases. σ_v is chosen according to the difference between measurements and actual positions of vehicles. σ_v can be set near to 0 so long as the detection results are good enough. However, this is often not true. σ_v is set to 15 in both directions here. As the tracked vehicle often exists in the next frame, P_S is set very high.

Table 1. The important parameter used in experiments

Parameter	Value
False track confirmation probability α	10^{-5}
True track deletion probability β	0.01
Standard deviation of process noise σ_w	$[5, 5]^T$
Standard deviation of measurement noise σ_v	$[15, 15]^T$
Measurement dimension M	2
Survival probability P_s	0.98

4.1 Qualitative evaluation

In this section, we use tracking results at some typical frames and trajectories of tracked vehicles for a period of time to show the tracking ability of three compared methods.

The first test is a traffic video without congestion. The surviving target is supposed at the top-center field of view for D-GMPHD filter and GM-PHD filter. The tracking results of the 24th frame are shown in Fig. 2 with different colored stars indicating the position estimations of vehicle centers, yellow rectangles indicating the regions where the targets are detected, the red number near vehicle indicating the vehicle speed and the yellow words at the top-left of the figure indicating the number of vehicles on the traffic. We only interest in the area which is within the red polygon, i.e., the vehicles only can be detected in the red polygon. In order to timely identify traffic congestion, we choose the exit road, which is marked with the green polygon to identify traffic congestion. In addition, we mark the jam in red and no traffic jam in green nearby the traffic congestion identification area to indicate the road condition. The colored rectangles in Fig. 2 (c) indicate tracking targets. It is obvious to see that six targets are false. From Fig. 2 (a) and (b), we can see that the D-GMPHD filter gives more complete trajectories of tracked vehicles than GM-PHD filter. In addition, the trajectory of the last vehicle is initialized and tracked by the D-GMPHD filter, while it is still treated as clutter by GM-PHD filter. Fig. 2 (c) indicates that BPF not only tracks all vehicles on the road, but also tracks some background areas which are treated as vehicles.

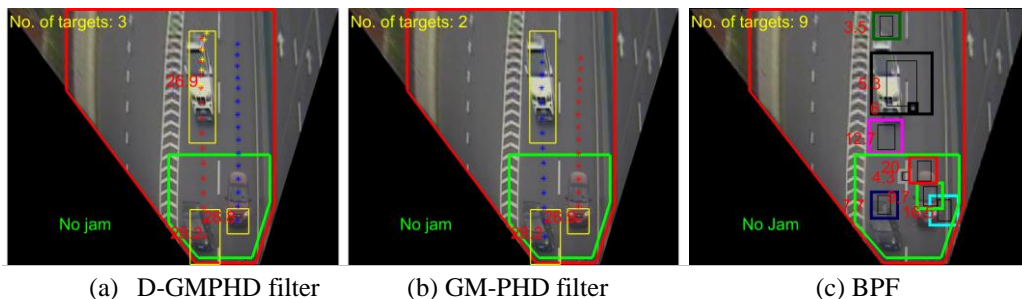


Fig. 2. Test 1: Tracking results of three methods at the 24th frame

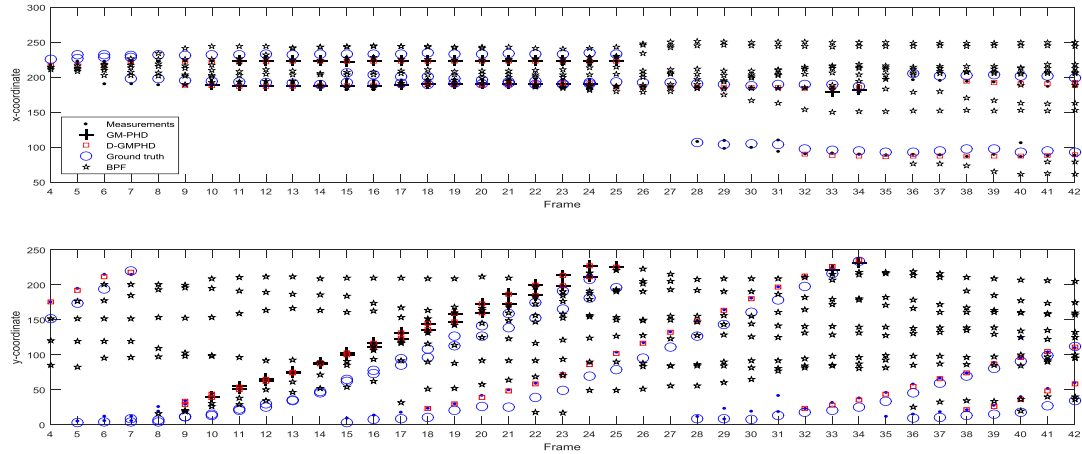
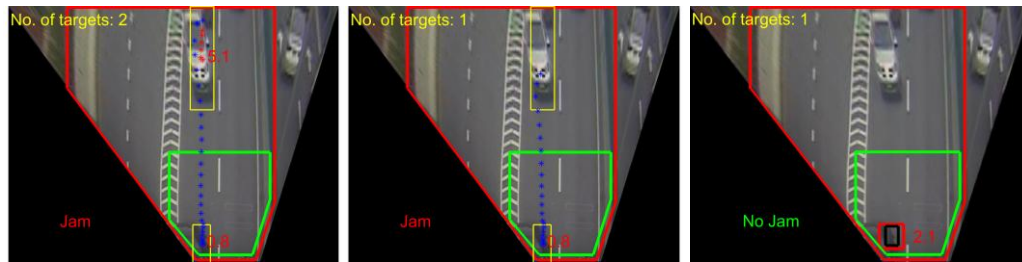


Fig. 3. Test 1: Tracking results of D-GMPHD, GM-PHD filter and BPF

Fig. 3 gives the position measurement results of the vehicle detection method, the ground truth positions of vehicles and track results of D-GMPHD filter, GM-PHD filter and BPF. For the vehicle appears at the 5th frame, the D-GMPHD filter can succeed in initializing track of the vehicle at the 9th frame, while the GM-PHD filter initializes track of it at the 10th frame. BPF initializes track of it at the 8th frame. For the vehicle appears at the 7th frame, D-GMPHD filter initializes track of it at the 9th frame. while the GM-PHD filter initializes track of the vehicle at the 11th frame. BPF initializes track of the vehicle at the 9th frame. For the vehicle appears at the 15th frame, D-GMPHD filter initializes track of it at the 18th frame. While GM-PHD filter initializes track of the vehicle at the 33th frame. BPF initializes track of the vehicle at the 17th frame. In addition, there are false positives tracked by BPF. For the vehicle appears at the 28th frame, D-GMPHD filter initializes track of it at the 32nd frame. However, GM-PHD filter fails to track of it. BPF initializes track of it at the 36th frame with great error. For the vehicle appears at the 36th frame, D-GMPHD filter initializes track of it at the 38th frame. However, GM-PHD filter fails to track it. BPF only initializes track of it at the 40th frame. In this case, all methods does not detect traffic congestion.

The second test is a traffic video with congestion, where the surviving target is supposed at the top-center field of view. All of three methods identify the traffic congestion except BPF. **Fig. 4** shows that GM-PHD still has a less complete trajectory than D-GMPHD filter, i.e., D-GMPHD filter initializes track of the white vehicle, while other methods still treat it as clutter. From **Fig. 5**, we can see that for the trajectory that starts from the 222th frame, D-GMPHD filter initializes track of it at the 225th frame. GM-PHD filter only initializes track of it at the 237th frame. BPF initializes track of it at the 227th frame. For the vehicle that appears since the 242th frame, D-GMPHD filter initializes track of it at the 245th frame, while GM-PHD filter just initializes track of it at the 270th frame. BPF initializes track of it at the 254th frame with a large tracking error. For the vehicle that appears at the 252th frame, D-GMPHD filter initializes track of it since the 255th frame, while GM-PHD filter initializes track it since the 271th frame. BPF initializes track of it since the 257th frame.



(a) D-GMPHD filter (b) GM-PHD filter (c) BPF

Fig. 4. Test 2: tracking results of three methods at the 250th frame

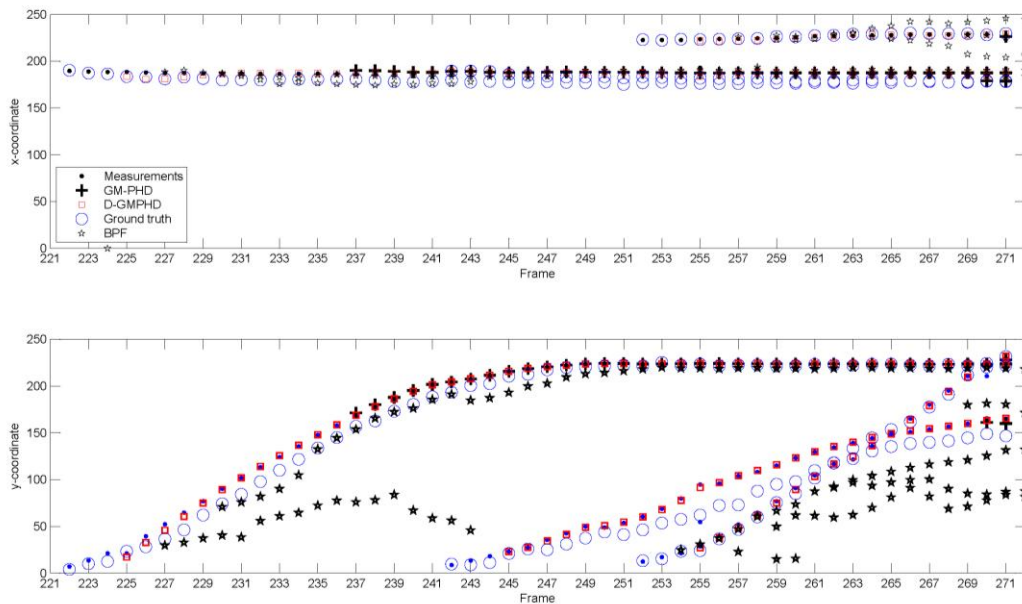


Fig. 5. Test 2: tracking results D-GMPHD, GM-PHD filter and BPF

The third test is also a traffic video with congestion, where the surviving target is supposed at the center-bottom field of view. **Fig. 6** shows that D-GMPHD filter has already identified the traffic congestion by the red words "jam" marked aside the parked vehicle, while GM-PHD filter still treats the vehicle as clutter. BPF produces many false targets while tracks the vehicle. According to Eq.(26), only D-GMPHD filter identifies the traffic congestion. Traffic congestion is identified based on the trajectory of it shown in **Fig. 7**. From **Fig. 7**, we can see that for the vehicle that appears at the 1th frame, D-GMPHD filter initializes track of it at the 5th frame, while GM-PHD filter just initializes track of it at the 20th frame. BPF can succeed tracking the vehicle appears since the 1th frame. However, it tracks the false positives since the 3th frame as well.

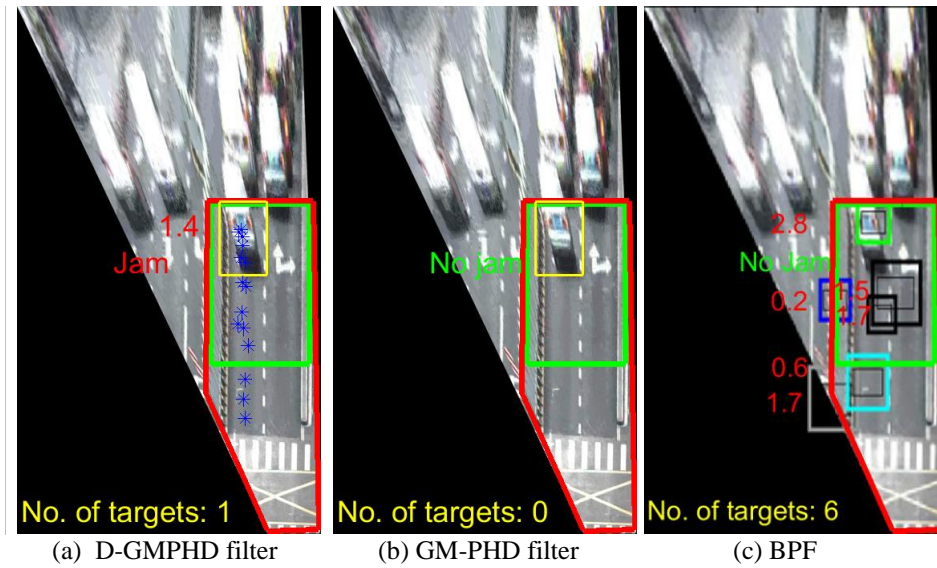


Fig. 6. Test 3: tracking results of three methods at the 19th frame

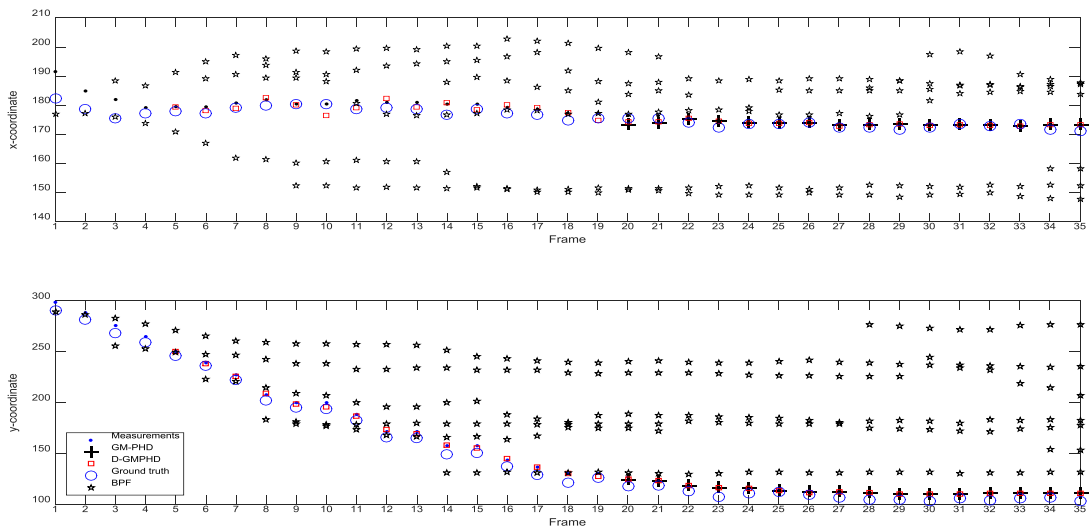


Fig. 7. Test 3: tracking results of D-GMPHD, GM-PHD filter and BPF

The fourth test is a traffic video with congestion, where the surviving target is supposed at the center of the intersection. Fig. 8 shows that all three methods can identify the traffic congestion. However, GM-PHD filter only tracks one vehicle on the road, while D-GMPHD filter tracks all of the vehicles in the interested area of the road. In other words, GM-PHD filter can not track the vehicles with No. 144 and No. 159. In this case, BPF performs even better than other methods. It does not track the false positives that cannot be terminated. In addition, it initializes track of vehicles as the detection results are obtained. From Fig. 9, we can see that there are clutter but no miss detection. The clutters appear as the vehicle detection method treats some parts of one vehicle to be background. As they are separated from each other, one vehicle will cause several measurements. For example, the detector treats the vehicle appeared from the 553th frame as 2 and 3 vehicles at the 588th, 591th, 592th and 594th frames, respectively. However, the D-GMPHD filter still treats them as one and has a good

tracking results. For the vehicle appears at the 553th frame, D-GMPHD filter initializes track of it at 557th frame. While GM-PHD filter initializes track of it at 594th frame. BPF shows its advantage as it initializes track of it 1 frame later. For the vehicles appear at the 603th and 635th frame, GM-PHD filter fails to initialize track of them until the 649th frame. BPF can initialize track of them once they appear. For the vehicle appears at the 603th frame, the D-GMPHD filter initializes track of it at the 604th frame. For the vehicle appears at the 635th frame, the D-GMPHD filter initializes track of it at the 638th frame.

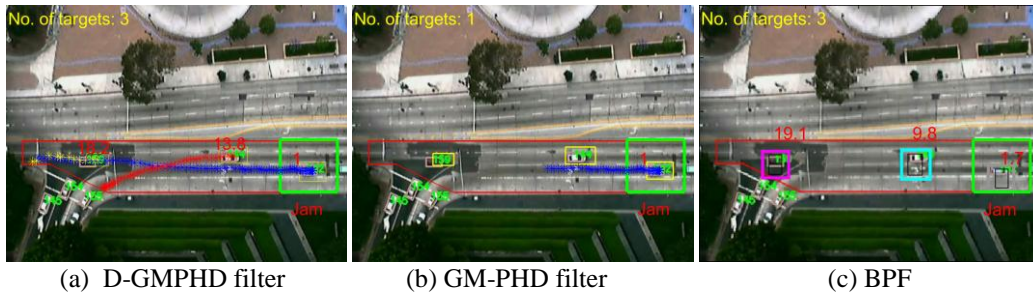


Fig. 8. Test 4: tracking results of three methods at the 649th frame

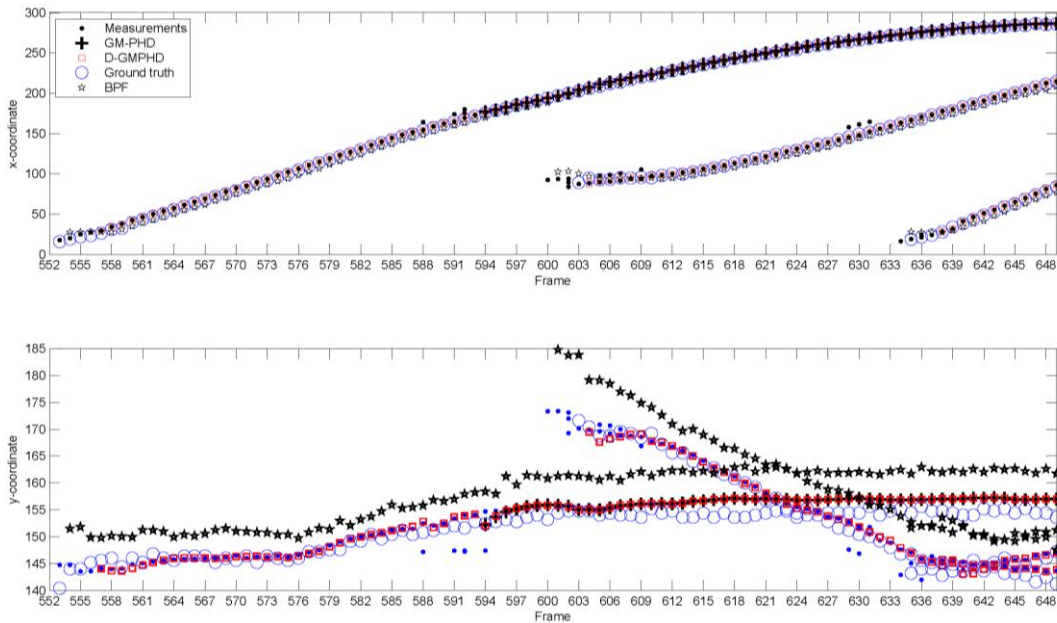


Fig. 9. Test 4: tracking results of D-GMPHD, GM-PHD filter and BPF

4.2 Quantitative evaluations

In this section, the overall performances for all three methods are summarized by early tracking ability, performance evaluation and time complexity analysis.

4.2.1 Early tracking ability

The quantitative comparison of D-GMPHD filter with GM-PHD filter and BPF in track initiation of each vehicle is given in Table 2. In each test, the values indicate the frames needed to initialize tracking of the vehicles after they appear. From Table 2, we can see that

GM-PHD filter needs more time delay in tracking the birth target than D-GMPHD filter. The biggest difference between D-GMPHD filter and GM-PHD filter can reach to 37 frames. In test 2 the biggest difference between them is 24 frames, considered the frame rate of this video is 8 frames/second. Hence D-GMPHD filter can even obtain 3 seconds earlier than GM-PHD filter in tracking vehicles in the traffic. That will leave more time for advanced traffic management systems to take effective action. From **Table 2**, BPF performs better in some tests. The reason behind this is that BPF treats detection results, which are not overlapped with existing tracking, directly as new entering vehicles. The result why it does not perform well in all tests is that some vehicles have appearance changes inter-frame, which cause the observation likelihood of them drop rapidly. The track can persist only when the appearance of the vehicle is stable.

Table 2. Frames needed by three methods to track the vehicle after it appears

		D-GMPHD filter	GM-PHD filter	BPF
Test 1	Target 1	4	5	3
	Target 2	2	4	2
	Target 3	3	18	2
	Target 4	4	-	8
	Target 5	2	-	4
Test 2	Target 1	3	15	5
	Target 2	3	28	12
	Target 3	3	19	5
Test 3	Target 1	4	16	1
Test 4	Target 1	4	41	1
	Target 2	1	-	0
	Target 3	3	-	0

4.2.2 Performance evaluation

In order to give a more convinced comparison, two intuitive metrics [37] described below are used.

(1) The multiple object tracking precision (MOTP):

$$MOTP = \frac{\sum_{i,k} d_k^i}{\sum_k c_k} \quad (27)$$

Where c_k is the number of matches found for time k . A matching is defined as the distance between the object and tracker hypothesis within the threshold T . The distance d_k^i between the object and its corresponding hypothesis is calculated based on each of these matches. The MOTP is the total error in estimated position for matched object-hypothesis pairs over all frames, averaged by the total number of matches made. It shows the ability of the tracker to estimate precise object positions, independent of its skill at recognizing object configurations, keeping consistent trajectories, and so forth. The performance will get better with the value decreases.

(2) The multiple object tracking accuracy (MOTA):

$$MOTA = 1 - \frac{\sum_k (m_k + fp_k + mme_k)}{\sum_k g_k} \quad (28)$$

Where m_k , fp_k , g_k and nme_k are the number of misses, of false positives, and of mismatches, respectively, for time k . The MOTA accounts for all object configuration errors made by the tracker, false positives, mismatches, over all frames. It gives a very intuitive measure of the tracker's performance at detecting objects and keeping their trajectories, independent of the precision with which the object locations are estimated. The performance will get better with the value increases.

The MOT comparison among the three methods is shown in **Table 3**. Under MOTP metric, GM-PHD filter performs best except for test 4. However, this superiority of GM-PHD filter to D-GMPHD filter can be ignored as the maximum difference between them cannot reach to 1 pixel for all tests. Under MOTA metric, D-GMPHD filter performs best as the minimum value for all tests is still above 72%. This is all due to the earlier tracking ability of D-GMPHD filter. The values of other methods except BPF at test 4, are all below 60%. Though BPF shows best in test 4, the superiority of BPF to D-GMPHD filter is not obvious as the difference between them only 3%. As BPF treats every measurement that is not overlapped with existing tracking as new trajectory, it performs well in test 4 in which fewer noisy measurements and appearance changes exist. Overall, the superiority of D-GMPHD filter is the best choice for multiple vehicle tracking on the road among the comparison methods.

In order to further improve the D-GMPHD filter in tracking with the big miss detection or large number of vehicle tracking, the GMPHD filter in the D-GMPHD filter can be seamlessly replaced by the GM-CPHD filter. The PHD recursion propagates cardinality information with only a single parameter (the mean of the cardinality distribution), and thus, it effectively approximates the cardinality distribution by a Poisson distribution. Since the mean and variance of a Poisson distribution are equal, when the number of targets present is high, the PHD filter estimates the cardinality with a correspondingly high variance. To address this problem, Cardinalized PHD is proposed by Mahler [38] and implemented by [39, 40]. However, PHD has the computational complexity of $O(mn)$, while CPHD has the computational complexity of $O(m^3n)$ [41]. n is the number of targets in the scene and m is the number of observations in the current measurement set. Hence if the tracking performance of GM-PHD filter can meet with the requirement of the application, there is no need to use GM-CPHD filter, and vice versa.

Table 3. Evaluating performance (MOTP in pixels)

Approach	Test 1		Test 2		Test 3		Test 4	
	MOTP	MOTA	MOTP	MOTA	MOTP	MOTA	MOTP	MOTA
D-GMPHD filter	9.15	72.34%	11.98	90.10%	4.59	88.57%	2.22	94.97%
GM-PHD filter	8.18	27.66%	11.84	34.65%	4.43	54.29%	2.65	35.22%
BPF	17.21	-114.89%	20.68	40.59%	24.17	-2.86%	7.77	98%

4.2.3 Time complexity analysis

The proposed system contains four parts: a projective transformation, vehicle detection, multiple target tracking and traffic congestion identification. Processing time mainly costs on vehicle detection and multiple target tracking.

- (1) Vehicle detection results are achieved by ViBe [17]. Barnich et al. state that it can nearly reach 200 frames per second for 640×480 pixels wide images with three color channels on the platform (2.67 GHZ Core i7 CPU, 6GB of RAM, C implementation).

- (2) **Table 4** shows frame rate on a Pentium Dual-Core CPU @ 2.6GHZ with Matlab implementation. The size of all test images is normalized to 320×240 pixels for fair comparing with BPF. D-GMPHD filter can reach to 6.67 frames/second at least.
- (3)

Table 4. Frame rate (average frames per second) of three methods for all tests

Approach	Test 1/ Test 2	Test 3	Test 4
D-GMPHD filter	6.67	7.78	9.71
GM-PHD filter	7.18	9.82	10.06
BPF	1.43	2.83	7.38

5. Conclusion

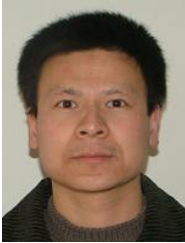
This work presents a real-time method for video based traffic congestion identification. The proposed technique includes projective transformation, vehicle detection, multiple vehicle tracking, and traffic congestion identification. After all vehicles are detected on the road, the adopted tracking method should have the ability in tracking them from their appearance to their disappearance. GM-PHD filter can track an unknown and time-varying number of targets in the presence of data association uncertainty, clutter, noise, and detection uncertainty. However, GM-PHD filter is intensive to birth target with unknown position. Vehicle recognition based on its movement model is used to relieve the shortage of GM-PHD filter at first. Then GM-PHD filter and ellipsoidal gate are used to form the trajectory of each vehicle. And then the speed of each vehicle can be obtained through the trajectory of each vehicle. At last the traffic congestion can be identified by the average traffic speed. Experimental results on real data indicate the advantage of the proposed method in giving a more complete trajectory of each vehicle in the field of view. The identified traffic congestion can provide the advanced traffic management systems with earlier and reliable basis to take further measures.

6. References

- [1] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: an overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271-281, December, 2002. [Article \(CrossRef Link\)](#)
- [2] L. D. Baskar, B. Schutter, J. Hellendoorn, and Z. Papp, "Traffic control and intelligent vehicle highway systems: A survey," *IET Intelligent Transport Systems*, vol. 5, no. 1, pp. 38-52, March, 2011. [Article \(CrossRef Link\)](#)
- [3] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 4, pp. 271-288, August, 1998. [Article \(CrossRef Link\)](#)
- [4] G. Marfia and M. Rocchetti, "Vehicular congestion detection and short-term forecasting: a new model with results," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 7, pp. 2936-2948, September, 2011. [Article \(CrossRef Link\)](#)
- [5] B.-F. Wu and J.-H. Juang, "Real-time vehicle detector with dynamic segmentation and rule-based tracking reasoning for complex traffic conditions," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 5, no. 12, pp. 2355-2373, December, 2011. [Article \(CrossRef Link\)](#)
- [6] G. Wang, D. Xiao, and J. Gu, "Review on vehicle detection based on video for traffic surveillance," in *Proc. of IEEE Conference on Automation and Logistics*, pp. 2961-2966, September 1-3, 2008. [Article \(CrossRef Link\)](#)
- [7] Y. Wang and S. Hu, "Chaotic Features for Traffic Video Classification," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 8, no. 8, pp. 2833-2850, August, 2014. [Article \(CrossRef Link\)](#)

- [8] A. B. Chan and N. Vasconcelos, "Probabilistic kernels for the classification of auto-regressive visual processes," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 846-851, June 20-25, 2005. [Article \(CrossRef Link\)](#)
- [9] A. B. Chan and N. Vasconcelos, "Classification and retrieval of traffic video using auto-regressive stochastic processes," in *Proc. of IEEE Intelligent Vehicles Symposium*, pp. 771-776, June 6-8, 2005. [Article \(CrossRef Link\)](#)
- [10] K. G. Derpanis and R. P. Wildes, "Classification of traffic video based on a spatiotemporal orientation analysis," in *Proc. of IEEE Workshop on Applications of Computer Vision*, pp. 606-613, January 5-7, 2011. [Article \(CrossRef Link\)](#)
- [11] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 267-276, June, 2010. [Article \(CrossRef Link\)](#)
- [12] E. B. Koller-Meier and F. Ade, "Tracking multiple objects using the condensation algorithm," *Robotics and Autonomous Systems*, vol. 34, no. 2, pp. 93-105, February, 2001. [Article \(CrossRef Link\)](#)
- [13] J. Xia, W. Rao, W. Huang, and Z. Lu, "Automatic multi-vehicle tracking using video cameras: An improved CAMShift approach," *KSCE Journal of Civil Engineering*, vol. 17, no. 6, pp. 1462-1470, September, 2013. [Article \(CrossRef Link\)](#)
- [14] R. Rad and M. Jamzad, "Real time classification and tracking of multiple vehicles in highways," *Pattern Recognition Letters*, vol. 26, no. 10, pp. 1597-1607, July, 2005. [Article \(CrossRef Link\)](#)
- [15] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 246-252, June 23-25, 1999. [Article \(CrossRef Link\)](#)
- [16] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-time imaging*, vol. 11, no. 3, pp. 172-185, June, 2005. [Article \(CrossRef Link\)](#)
- [17] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709-1724, June, 2011. [Article \(CrossRef Link\)](#)
- [18] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627-1645, September, 2010. [Article \(CrossRef Link\)](#)
- [19] X. Ren and D. Ramanan, "Histograms of sparse codes for object detection," in *Proc. of Conference on Computer Vision and Pattern Recognition*, Portland, Oregon, USA, pp. 3246-3253, 2013. [Article \(CrossRef Link\)](#)
- [20] T. Semertzidis, K. Dimitropoulos, A. Koutsia, and N. Grammalidis, "Video sensor network for real-time traffic monitoring and surveillance," *IET Intelligent Transport Systems*, vol. 4, no. 2, pp. 103-112, June, 2010. [Article \(CrossRef Link\)](#)
- [21] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843-854, December, 1979. [Article \(CrossRef Link\)](#)
- [22] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine* vol. 19, no. 1, pp. 5-18, January, 2004. [Article \(CrossRef Link\)](#)
- [23] F. Garcia, A. Prioletti, P. Cerri, A. Broggi, A. de la Escalera, and J. Armingol, "Visual feature tracking based on PHD filter for vehicle detection," in *Proc. of IEEE Conference on Information Fusion*, pp. 1-6, July 7-10, 2014. [Article \(CrossRef Link\)](#)
- [24] K. Panta, B.-N. Vo, S. Singh, and A. Doucet, "Probability hypothesis density filter versus multiple hypothesis tracking," in *Proc. of Defense and Security*, pp. 284-295, April 12, 2004. [Article \(CrossRef Link\)](#)
- [25] Y. Wang, Z. Jing, S. Hu, and J. Wu, "Detection-guided multi-target Bayesian filter," *Signal Processing*, pp. 564-574, February, 2012. [Article \(CrossRef Link\)](#)
- [26] B. Ristic, D. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive target birth intensity for phd and cphd filters," *IEEE Transactions on Aerospace and Electronic Systems* vol. 48, no. 2, pp. 1656-1668, April, 2012. [Article \(CrossRef Link\)](#)

- [27] R. P. S. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152-1178, October, 2003. [Article \(CrossRef Link\)](#)
- [28] B. N. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4091-4104, November, 2006. [Article \(CrossRef Link\)](#)
- [29] S. S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Artech House, Boston, 1999. [Article \(CrossRef Link\)](#)
- [30] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, pp. 117-186, 1945. [Article \(CrossRef Link\)](#)
- [31] J. D. Echard, "Estimation of radar detection and false alarm probability," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, no. 2, pp. 255-260, March, 1991. [Article \(CrossRef Link\)](#)
- [32] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. Wiley-Interscience, Canada, 2001. [Article \(CrossRef Link\)](#)
- [33] I. Carlbom and J. Paciorek, "Planar geometric projections and viewing transformations," *ACM Computing Surveys (CSUR)*, vol. 10, no. 4, pp. 465-502, December, 1978. [Article \(CrossRef Link\)](#)
- [34] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*, 3rd ed. Thomson, Toronto, 2008. [Article \(CrossRef Link\)](#)
- [35] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little, and D. G. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *Proc. of European Conference on Computer Vision*, pp. 28-39, May 11-14, 2004. [Article \(CrossRef Link\)](#)
- [36] NGSIM. NGSIM [Online]. Available: <http://ngsim-community.org>
- [37] B. Keni and S. Rainer, "Evaluating multiple object tracking performance: the CLEAR MOT metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1-10, February, 2008. [Article \(CrossRef Link\)](#)
- [38] R. Mahler, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1523-1543, October, 2007. [Article \(CrossRef Link\)](#)
- [39] B. T. Vo, B. N. Vo, and A. Cantoni, "Analytic implementations of the cardinalized probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3553-3567, July, 2007. [Article \(CrossRef Link\)](#)
- [40] M. Ulmke, O. Erdinc, and P. Willett, "Gaussian mixture cardinalized PHD filter for ground moving target tracking," in *Proc. of IEEE Conference on Information Fusion*, pp. 1-8, 2007. [Article \(CrossRef Link\)](#)
- [41] R. P. Mahler, *Statistical multisource-multitarget information fusion* vol. 685. Artech House, Boston, 2007. [Article \(CrossRef Link\)](#)



Xiaoyu Zhang is a PhD Candidate in the School of Aeronautics and Astronautics at Shanghai Jiao Tong University, Shanghai, China. He received his BE degree in Metallic Materials Engineering and MS degree in Detection Technology and Automatic Equipment from Hebei University of Science and Technology, Shijia Zhuang, China, in 2006 and 2009, respectively. His research interests include object detection, visual tracking.



Shiqiang Hu received his BS degree from Hebei University of Science and Technology, Shijia Zhuang, China and MS and PhD degrees from Beijing Institute of Technology, Beijing, China, in 1991, 1998, and 2002, respectively, all in Electronics and Information Technology. Currently, he is a professor at the School of Aeronautics and Astronautics, Shanghai JiaoTong University, China. His research interests include intelligent information processing, information fusion, nonlinear filtering, and aerospace control.



Huanlong Zhang received his BS degree in Automation and MS degree in Computer Application from Henan University, Kaifeng, China, in 2004, 2007, respectively. He received his PhD degree in Control Theory and Control Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2015. Currently, he is a teacher lecturer in Zhengzhou University of Light Industry. His research interests include pattern recognition and computer vision.



Xing Hu is a Ph. D Candidate in the School of Aeronautics and Astronautics at Shanghai Jiao Tong University. He received his BE degree in computer science and technology from Anhui University of Technology, Maanshan, China, in 2007. He received his MS degree in computer science and technology from Guangxi Normal University, Guilin, China, in 2010. His research interests include video anomaly detection and video scene analysis.