ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 10(17), DOI: 10.17485/ijst/2017/v10i17/111960, May 2017* ISSN (Online) : 0974-5645

Performance Evaluation of Crowd Analysis Algorithm using Modified GMM and Adaptive Thresholding

Navneet S. Ghedia[1*], C. H. Vithalani[2] and Ashish Kothari[3]

[1]Department of Electronics and Communication, Gujarat Technological University, Ahmedabad – 382424, Gujarat, India; navneet_ghedia@yahoo.com
[2]Department of Electronics and Communication, Government Engineering College, Rajkot, Rajkot – 360005, Gujarat, India; chvgec@gmail.com
[3]EC Department, Atmiya Institute of Technology and Science, Rajkot – 360005, Gujarat, India; amkothari.ec@gmail.com

Abstract

Objective: To evaluate the crowd densities in video scene under different constraints. For the crowded video analysis robust foreground detection methods are required to differentiate between moving or static foreground objects and static or dynamic background. Large number of foreground segmentation or motion segmentation approaches are available but only few can handle the various constraints like illumination variations, dynamic background partial or high level of occlusions. **Method:** We have proposed a modified Gaussian mixture model using adaptive thresholding. The proposed approach is implemented in MATLAB. **Findings:** Our proposed approach analyze all the aspects of the various backgrounds and foregrounds modelling and then compared their critical performance in terms of the PR curves and miss rate. The performance evaluation demonstrates considerable improvements in miss rate compared to traditional approaches. Our proposed method also shows significant improvements in Multi Object Detection and in Tracking Accuracy. **Application:** Our proposed approach analyzes the crowded scenes, especially handles outdoor environment. Optimized model parameters and adaptive thresholding makes it more robust to handle varying light conditions and partial occlusions.

Keywords: Background Model, Crowd Analysis, Foreground Model, Parametric Model

1. Introduction

Tracking of object precisely in dynamic and crowded scenes is a tough task due to occlusion between objects. Poor illumination and occlusions make it difficult to detect and track people in a crowded scene. Single camera system cannot handle such scenes for correct detection and tracking moving objects or peoples in crowded scenes for which multiview are required. Detection and tracking becomes much simpler task if an object is isolated i.e. it is neither occluded or occluding another object, because of their features[1]. In this paper, we offer an approach to detect and track objects in crowded scenes. A good number of videos used in this paper from the standard datasets, represent challenges like illumination variations, dynamic background and scene are suitably dense that partial or large amount of occlusions ensures that objects cannot be isolated. In this paper, we assume a camera view which is suitably crowded that partially or fully occluded and certainly that an object can be visually isolated. Figure 1 shows some of the crowded scenes that we used to evaluate our proposed approach. One thing that is noticeable that, only few objects are isolated and rest are in occlusion. In our proposed approach, we have used pixel based color model for individual object. We detect and track objects in single static camera with dynamic background. Under the inadequate visibility and clutter background it might be very difficult for the intelligent surveillance system to detect and track multiple moving people.

* Author for correspondence

Figure 1. Examples of crowded sequences with clutter background used to evaluate our approach.

Generally major part of the surveillance system is depends on the perception. One of the solutions of the said situation is that we might taken a help from the other view, Multi view may possibly avoid such discrepancies or occlusions. For the static single camera we must have to sense the environment and act accordingly[1].

Figure 2 shows general approach of the moving object detection and tracking. Object location, object silhouette, object classification and the activities carried out by the object is vital in scene analysis. The robust tracking depends on the robustness of the foreground detection. The role of the object detection and tracking algorithm is to estimate the trajectory path which compares against the ground truth, in subsequent frames. Under the assumption that our detection model is linear and the system noise and posteriori distributions are Gaussian in nature, among all tracking approaches prediction filter approach gives better accuracy[2].

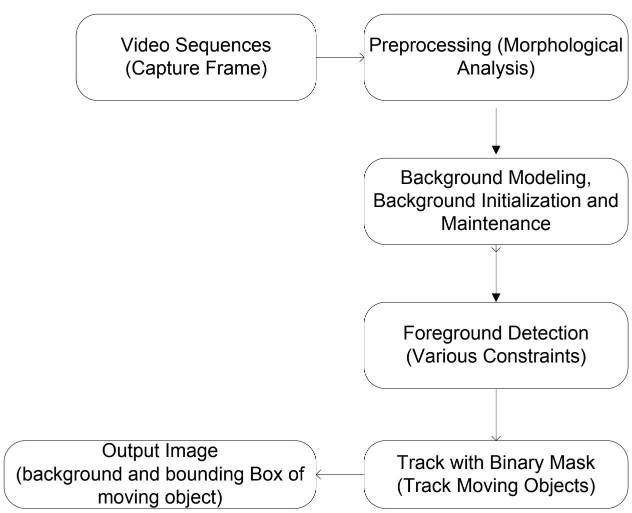

Figure 2. Block diagram of object detection and tracking.

Generally we are assuming that the static backgrounds are available in a video sequence but for the real time processing background frames are not available so, for the initial subsequent frames are required to generate the background. The next step is to preprocess the input frames, such as morphology, image resizing along with edge detection is required. Morphological approach will reduce noise in the moving object and fill the gaps in it. Dilation: Will convert each background pixel into foreground if it touches with the foreground. Erosion: Will convert each foreground pixel into background if it touches with the background and it helps to reduce the noise and fill gaps. Clutter and dynamic background constraints can be handle by estimating background modeling. Foreground detection using adaptive thresholding will segment the moving object. Finally the robustness of the surveillance system depends on how accurate it able to track object in every successive frame[3].

2. Background

Crowd analysis requires robust background modelling which helps to take certain challenges like dynamic and clutter background and similar appearance. Most of the researchers are using either pixel based background modelling or region based background modelling. Some are using non adaptive approaches; such approach fails to handle dynamic scenes for the real time analysis. In[4] it explains a novel approach for the motion estimation using pixel based Gaussian mixture model. They have used parametric approach and colour as a feature. The traditional GMM does not need to store input data. They have used multi model GMM for the dynamic scenes and their approach fails to detect object in sudden illumination. In[5] it has proposed the system which overcomes the demerits of[4] slower learning rate. In[6] it has proposed a colour and depth based segmentation approach. The proposed algorithm is suffered with slower learning rate because of the traditional MoG approach. In[7,8] proposed quantize based code book and traditional GMM approach for the moving object detection. It is a method which uses quantization/clustering pixel based technique to obtain a multimodal background model from long observation sequences. The codebook algorithm was anticipated to sample values over a long time. The algorithm works well on slow as well as fast moving objects. In[9] it has proposed used RGB background modelling for the real time moving object detection and used morphology for removing noise and blob labelling for real time moving object detection. They predict the velocity of the moving objects and detect it. In[10] they proposed Kernel Density Estimation to model the background distribution. KDE is a nonparametric region based approach which uses colour as a feature. It is also deal with multi modal backgrounds. It requires memory for the foreground detection. In[11] they have designed an improved algorithm to adaptively adjust the parameters and number of components of GMM. The algorithm can automatically adapt to the scene by choosing the number of components of pixel. In[12] they have proposed a method that computes the sample consensus of the background samples and estimates a statistical model of the background. In[13] they have proposed a background modelling method (SOBS) based on self-organization achieved by artificial neural networks. The proposed algorithm can handle dynamic background, gradual illumination changes and camouflage. It works well on stationary camera. In[14] they have proposed a random strategy in the field of background modelling to select values to build a sample-based estimation of the background. It is a nonparametric pixel based approach which uses colour as a feature. In[15] they suggested foreground estimation by creating pixel-based adaptive segmenter method.

3. Proposed Method

Crowd analysis requires vigorous detection and leads to a precise tracking. Our proposed approach estimates background which handles dynamic and clutter background.

It also handles different constraints like scenes, moving background, entering and leaving objects in a current frame, clutter background, complex object silhouette etc.

Figure 3 shows the proposed approach based on the modified GMM, the modifications are achieved in terms of the intrinsic and extrinsic. An intrinsic improvement significantly reduces false positives and hence increases the precision. An extrinsic improvement such as Pre processing and Post processing improves the performance evaluations of the proposed approach by reducing the false negatives and hence increases the recall. Our proposed algorithm works on background analysis, motion segmentation and a tracking. Our proposed algorithm use traditional Gaussian mixture model approach and we have modified and updated some of the parameter to handle different constraints.

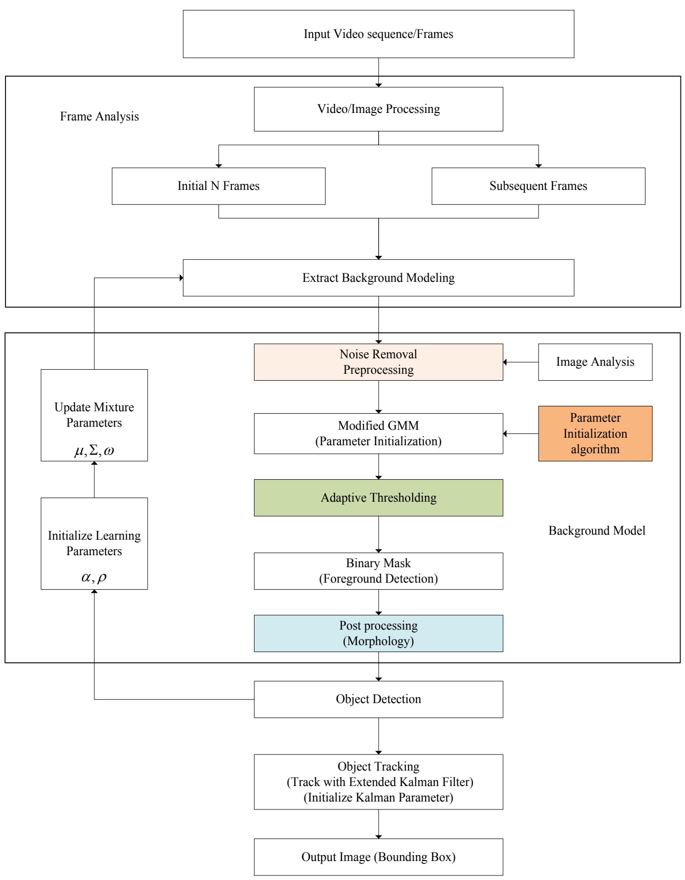

Figure 3. Proposed modified GMM algorithm.

3.1 Background Analysis

In a scene analysis background generally a non static due to flowing leaf's twinkling of water surface. Usually the Gaussian approach along with static or dynamic threshold will segment the foreground for the constant background. For the real time crowded analysis the system must take challenges of the dynamic and clutter background scenes. We have proposed multi modal Gaussian to handle moving background and other constraints efficiently. Generally our proposed algorithm works on parametric pixel by pixel approach rather than region based for the background modelling. For the background modelling our proposed algorithm uses parametric colour pixel by pixel approach if each pixel is considered from a static background and the illumination can be fixed with respect to time, single Gaussian would detect motion from the every successive frame.

$$F(X/\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

The single Gaussian handles small and gradual variations in background but fails to handle large and sudden illumination variations. For the real time crowded analysis multi model Gaussian is required to handle dynamic and clutter background. Every time mixture parameters are updated and it will maintain the model.

In a RGB color space, frame pixel can be characterized by intensities and its probability in the current fame is:

$$P(X_t)=\sum_{i=1}^{k}\omega_{i,t}.\eta(X_t,\mu_{i,t},\Sigma_{i,t}) \tag{2}$$

Where,

$\omega_{i,t}$ = weighted associate to current frame Gaussian.

k = no. of distributions.

$\mu_{i,t}$ & $\Sigma_{i,t}$ = mean and covariance matrix of the pixel intensities.

η = the Gaussian probability density function,

$$\eta(X/\mu,\Sigma)=\frac{1}{\sqrt{2\pi|\Sigma|}}e^{\left\{-\frac{1}{2}(X-\mu)^T\Sigma^{-1}(X-\mu)\right\}} \tag{3}$$

Each pixel is identified as a mixture of Gaussian and initializes the different mixture model parameters. The weight, the mean and the covariance matrix is initialized using an EM algorithm or Maximum a Posteriori (MAP) estimation[16].

Foreground detection using adaptive threshold:

First B_{back} Gaussian distributions from K no. of Gaussian distributions will be considered as the background model and B_{back} can be evaluated as:

$$B_{back}=\arg\min(\Sigma_{i=1}^{b}\omega_{i,t}\rangle T) \tag{4}$$

T is considered as the minimizing measure of estimating background. Particularly high threshold, foreground pixels with small colour differences will be misclassified and a lower threshold will result in unremovable noise. When using a single or a mixture of Gaussian models, the threshold for every pixel is a fixed multiple of its variance in which only temporal features are considered. Therefore, to attain satisfactory result we have to define threshold very precisely, particularly when foreground colour are similar to their surrounding background colour.

$$T_t=k_1(\sigma_p^2+k_2\sigma_s^2)/\mu_s \tag{5}$$

σ_p^2 = the local variance in the present frame.

σ_s^2 = the local variance in the subtracted frame.

μ_s = the local mean in the subtracted frame .

The background can be subtracted from the present frame will give us the subtracted frame.

3.2 Object Tracking

Object tracking with high precision in non-static and crowded scenes is a tough job due to occlusion among objects. The complexity can be more with the abrupt motion of the object, complex silhouette, illumination changes etc. Generally, tracking can be achieved after the motion segmentation but in some of the real time application it can be carried out along with the segmentation. Different kinds of recursive and non recursive tracking approaches are available. Among all approaches recursive Kalman approach will give somewhat more precise result[2].

3.3 Kalman Filtering

Object's correct position can be simply and promptly considered by means of mathematical analysis.

Figure 4 represents the simplified recursive Kalman filtering approach. Kalman is the mathematical approach which uses statistical equations and successive inputs.

Kalman filter is used to estimate the state of a linear system where the state is contained to be distributed by a Gaussian[2]. The prediction state of the Kalman filter uses the state model to predict the new state of the variable.

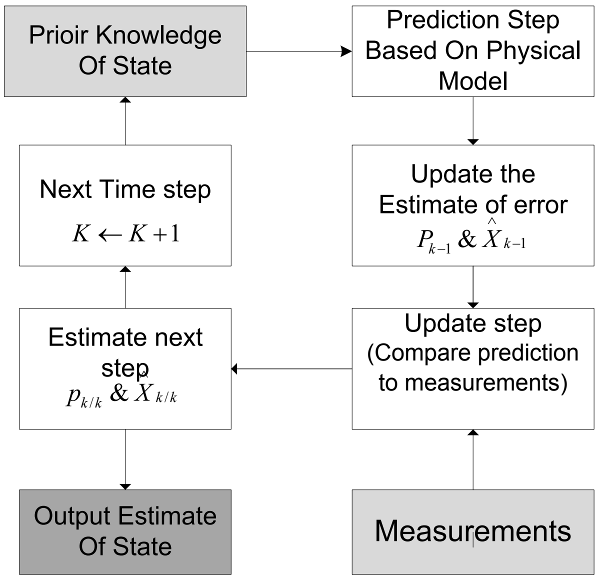

Figure 4. Block diagram of Kalman filter.

The Kalman filter estimates the entire process by means of feedback. The Kalman estimates the process state and obtains feedback in a form of measurement. Time update and measurement update are the two states of the Kalman. The time update is liable for the analysing the current state and the measurement update are liable for incorporating a new measurement into a prior estimate to get an improved posterior estimate[17].

Time update equation can also be considered as prediction Equations.

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \tag{6}$$

$$\bar{P}_k = AP_{k-1}A^T + Q \tag{7}$$

The measurement equation is recognized as correction Equations.

$$K_k = \bar{P}_k H^T (H\bar{P}_k H^T + R)^{-1} \tag{8}$$

$$\hat{x}_k = \hat{x}_k^- + K_k\left(z_k - H\hat{x}_k^-\right) \tag{9}$$

$$P_k = (I - K_k H)\bar{P}_k \tag{10}$$

4. Experiment Result

To evaluate our proposed algorithm, we have performed some of experiments on own datasets and standard video dataset PETS 2009[18] and ViSOR[19]. Proposed approach can be evaluated by various performance evaluation parameters such as miss rate, Multi Object Detection Accuracy (MODA) and Multi Object Tracking Accuracy (MOTA).

Figure 5 is a crowded standard sequence from PETS 2009[18] which is suffered with the clutter background and occlusions. Our proposed algorithm detects the crowded people under the different constraints. Figure 9 shows that our proposed algorithm is able to handle partial occlusion and give sufficient detection and tracking results. Figure 6 is also a crowded standard sequence ViSOR[19]. This sequence is suffered with the clutter background, occlusions and light variations. Our proposed approach is capable to detect and track the moving object against all constraints. Our proposed algorithm fails to detect and track objects which are fully occluded and those with similar appearance.

Figure 5. Crowded sequences PETS 2009.

Figure 6. Crowded sequences ViSOR.

The result of Figure 5 and Figure 6 clearly indicate that our proposed algorithm gives good results against the clutter background and occlusions. The quantitative analysis shows more clearly about the validity of our proposed approach. We are comparing our proposed approach with the other similar approaches like traditional GMM[4] and KDE[10].

Figure 7 shows the comparative evalution for the influence of the false positives on true positives. It presents by means of Precision–Recall (PR) curve. We are comparing our prosed approach with the other similar approaches for the stadard videodataset PETS 2009. It indicates that the our proposed algorithm gives high recall and the optimized mixture model parameters reduces false negatives.

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{11}$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{12}$$

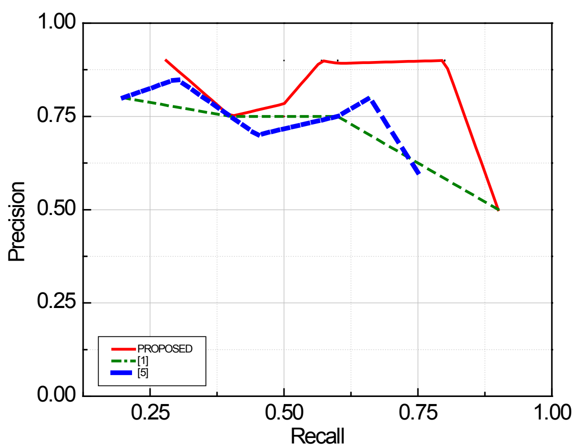

Figure 7. Precision-recall curve.

Figure 8 indicates comparative analysis for the false negatives with reference to the FPPI (False Positive Per Image) for PETS 2009 and ViSOR. We have compared our proposed approach with the other two approaches. Proposed approaches shows significant improvement in both the sequences. The quantitiative performance can be evaluated by a their detection and tracking accuracy.

$$MODA = 1 - \frac{F_n + F_p}{T_p + F_n} \tag{13}$$

$$MOTA = 1 - \frac{\sum_t m_t + fp_t + mme_t}{\sum_t g_t} \tag{14}$$

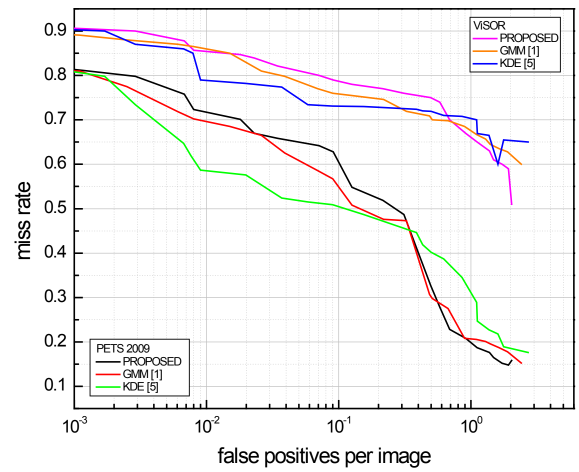

Figure 8. Miss rate against false positive per image.

Figure 9 indicates the comparative analysis of the standard sequences for the Multi Object Detection Accuracy (MODA) and Multi Object Tracking Accuracy (MOTA). MOD Accuracy gives the information regarding false positives with respect to miss rate or false negatives. MOT Accuracy depends on the total no. of false positives, negatives and the object mismatch errors.

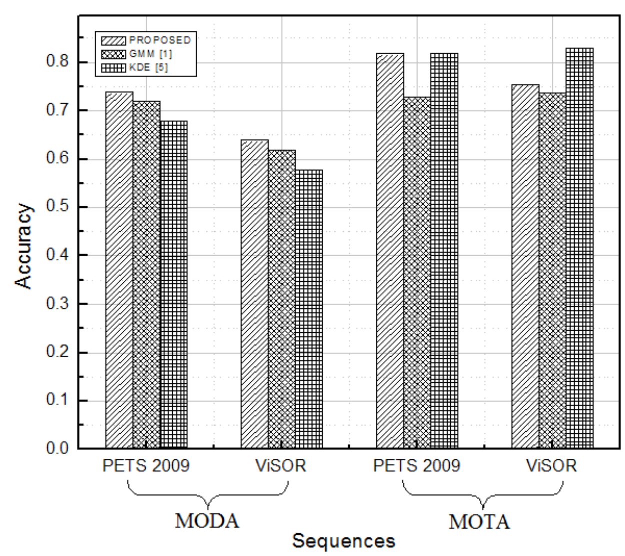

Figure 9. Multiple object detection and tracking accuracy.

5. Conclusion

Performance evolutions shown in sequences of the intense crowed occlusions are relatively common. Our proposed approach is capable of tracking multiple objects in cutter background and complex background and it is preferably appropriate for the partially occluded objects. For the fully occlusion our algorithm will fail to detect and track objects. There are numerous tentative approaches available to handle fully occlusion and among all, the multi view approaches may give better accuracy. For handling the dense crowd, we must have to increases the camera views. The robustness of the proposed algorithm is proved using the approaches and challenging datasets.

6. References

1. Khan SM, Shah M. Tracking multiple occluding people by localizing on multiple scene planes. PAMI. 2009; 31(3):505–19. PMid: 19147878. Crossref
2. Yilmaz A, Javed O, Shah M. Object tracking: A survey. ACM Journal of Computing Surveys. 2006; 38(4):1–45. Crossref
3. Ghedia NS, Vithalani CH, Raval M, Parmar K. A novel approach to detect foreground in video sequences based on mixture of Gaussians. IJSRD. 2015; 3(6):1220–4.
4. Stauffer C, Grimson WEL. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000 Aug; 22(8):747–57.
5. Bowden R, Kaewtrakulpong P. An improved adaptive background mixture model for real-time tracking with shadow detection. AVBS01; 2001. p. 1–5.
6. Harville M. A framework for high-level feedback to adaptive, per-pixel, mixture-of-Gaussian background models. 7th European on Computer Vision. 2002; 3:37–49.
7. Kim K, Chalidabhongse TH, Harwood D, Davis L. Real-time foreground-background segmentation using codebook model. Real-Time Imaging. 2005; 11:172–85. Crossref
8. Kim K, Chalidabhongse T, Harwood D, Davis L. Background modelling and subtraction by codebook construction. Int Conf on Image Processing; 2004. p. 3061–4. PMid: 15107464. PMCid: PMC452564.
9. Kim JS, Yoem DH, Joo YH. Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems. IEEE Transaction on Consumer Electronics. 2011 Aug; 57(3):1165–70. Crossref
10. Elgammal A, Hanvood D, Davis LS. Non-arametric model for background subtraction. Proc ECC; 2000 Jun. p. 751–67.
11. Zivkovic Z, Heijden FV. IEEE Transactions on pattern analysis and machine. PAMI. 2004; 26(5):138–41.
12. Wang H, Suter D. A consensus-based method for tracking: Modelling background scenario and foreground appearance. Pattern Recognition. 2007 Mar; 40(3):1091–105. Crossref
13. Maddalena L, Petrosino A. A self-organizing approach to background subtraction for visual surveillance applications. IEEE Trans Image Process. 2008 Jul; 17(7):1168–77. PMid: 18586624. Crossref
14. Barnich O, Droogenbroeck MV. Vibe: A powerful random technique to estimate the background in video sequences. IEEE Int Conf on Acoustics, Speech and Signal Processing; 2009. p. 945–8. Crossref
15. Hofmann M, Tiefenbacher P, Rigoll G. Background segmentation with feedback: The pixel-based adaptive segmenter. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognition Workshops; 2012 Jun. p. 38–43. Crossref
16. Reynolds D. Gaussian mixture models. 244 Wood St., Lexington, MA 02140, USA: MIT Lincoln Laboratory; 2007. p. 1–5.
17. Welch G, Bishop G. An introduction to the kalman filter. Technical Report; 2006. p. 1–16.
18. Ferryman J. Performance evaluation of tracking and surveilliance. PETS Winter; 2009. p. 1–8.
19. Vezzani R, Cucchiara R. Video Surveillance Online Repository (ViSOR). An integrated framework in Multimedia Tools and Applications. Kluwer Academic Press. 2010; 50(2):359–80.