MJoSHT
Malaysian Journal of Science, Health & Technology
[mjosht.usim.edu.my]

USIM

Article

A New Motion Segmentation Technique Using Foreground-Background Bimodal

Ma'moun Al-Smadi[1,a], Khairi Abdulrahim[2], Rosalina Abdul Salam[1,b]

[1] Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai
Negeri Sembilan, Malaysia
E-mail: [a]ma_smadi@yahoo.com, [b]rosalina@usim.edu.my

[2]*Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai*
Negeri Sembilan, Malaysia
E-mail: khairiabdulrahim@usim.edu.my

***Abstract*— Vehicle detection is a fundamental step in urban traffic surveillance systems, since it provides necessary information for further processing. Conventional techniques utilize either background subtraction or foreground appearance-based detection, which involves either poor adaptation or high computation. The complexity of urban traffic scenarios lies in pose and orientation variations, slow or temporarily stopped vehicles and sudden illumination variations. In this work, a foreground-background bimodal is proposed to adapt for scene variation and complexity. Cumulative frame differencing and sigma-delta estimation are used to model foreground and background respectively. A correction feedback updates each model iteratively and recursively based on the detection mask of the other model. Variance update for sigma-delta estimation was limited to update background temporal activities, while cumulative frame differencing account for moving foreground by discarding limited background variations. Comparative experimental results for typical urban traffic sequences show that the proposed technique achieves robust and accurate detection, which improves adaptation, reduce false detection and satisfy real-time requirements.**

***Keywords*— Motion segmentation, Cumulative frame differencing, Sigma-delta filter, Vehicle detection.**

I. Introduction

Traffic surveillance and monitoring systems provide road users with valuable information, which require accurate and real-time parameter estimation. Information about traffic conditions improve road safety and utilization through assisting drivers and governments in rout selection and street planning respectively. Additionally, it helps in optimizing vehicle flow, which have economical and environmental benefits that reduce pollution emission and enhance life quality [1].

Intelligent transportation systems (ITS) use a variety of sensors to measure traffic flow. Conventional techniques use inductive loops, sonar or microwave detectors, which disturb traffic during the costly installation or maintenance process. In recent years, video cameras combined with computer vision techniques offer an attractive capability for data acquisition, since traffic videos provide more information about the traffic of vehicles. Such systems are easy to install, maintain and upgrade with relatively low cost and wide variety of applications especially in urban environments. Applications of video-based systems may include vehicle detection, classification and counting, speed measurement and incident detection. Thus, current technological trends in traffic monitoring and surveillance are oriented towards a video-based system [2].

Identifying moving objects (i.e. Vehicles) in a video sequence captured by a static camera is a fundamental and critical task in traffic surveillance and monitoring systems. Vehicle detection can be performed using either appearance-based techniques [1], which require prior knowledge and high computation, or motion segmentation techniques that include; frame differencing [3], background subtraction [4][5] and optical flow [6]. Motion segmentation and detection is often used in various applications to distinguish between moving foreground objects and stationary background scene. Accuracy and robustness of segmentation have a great importance in detection, recognition, tracking, and higher-level processing [7].

Many recent studies on background subtraction have been developed to detect moving objects, these studies can be classified into parametric, nonparametric and predictive techniques [8]. Parametric background modelling uses a single unimodal probability density function that model each background pixel. There are several techniques based on the above assumption such as; running Gaussian average [9], temporal median filter [10], sigma-delta filter [11], and Gaussian Mixture Model (GMM) [4], [12]. Running Gaussian average use Gaussian density function recursively to represent each pixel [9]. Approximate median estimates the background recursively based on the assumption that the pixel stays in the background for more than half of the period under consideration [10]. In [11], sigma-delta filter was used to update background intensity and variance. Intensity variance was used as a dynamic threshold to isolate foreground pixels from the estimated background [11]. GMM models each pixel as a mixture of two or more temporal Gaussians with online updated. The Gaussian distributions are estimated as either a more stable background process or a short-term foreground process [12]. GMM can adapt illumination variations and repetitive clutter with higher computation and memory requirements compared with standard background subtraction techniques [13].

Nonparametric techniques have more ability to handle arbitrary density functions, thus they are more suitable for complex functions that cannot be parametrically modelled. Kernel density estimation (KDE) is an example of such techniques [14]. It uses KDE to estimate the background probabilities of each pixel from many recent samples. Previous techniques are limited to smooth behaviour and limited variations, while KDE overcomes the problem of fast variations and nonstationary properties of the background. Another nonparametric technique is based on codebook model, which use a set of dynamically handled codewords to replace the parameters represented by a probabilistic function [15].

Finally, predictive techniques employ predictive procedures in predicting the dynamic state of each background pixel. Kalman filtering, Wiener filter, autoregressive models [16] and eigenbackground [17] are examples of such techniques.

The use of sigma-delta in background subtraction attract many researchers due to its computational efficiency [17]. Since it requires only basic integer arithmetic operations that include comparison, increment and absolute difference. The robustness of this technique is comparable with other unimodal statistical techniques that have higher computational cost.

Many improvements have been suggested to enhance this technique at the expense of computational complexity or memory requirement. In [18] Zipf-Mandelbort distribution was used to update the background according to the dispersion of the distribution. Spatiotemporal processing and multiple-frequency sigma–delta proposed in [11], improves the detection by removing non-significant pixels and using the weighted sum of multiple models with different updating periods. Another multi-model was introduced in [19], using a mixture of three distributions. They used a weight as a voting value of background models to sort the mixture according to higher and lower updating value. Confidence measurement was introduced in [20] and enhanced in [21]. They tied each pixel with a numerical confidence level that is inversely proportional to the updating period and used to control the booming of intensity variance. In [22] a hierarchical or bi-level sigma-delta filtering was introduced, which perform a conditional update that include the low level temporal update and high level spatial update. Selective and partial updates using global variance was applied in [23], which make a good balance between sensitivity and reliability at the expense of higher computation.

However, the sigma delta technique still facing many challenges, especially in urban environments, since it quickly degrades under slow or congested traffic conditions, due to the integration of pixel intensities from the moving foreground object into the background model [11]. Moreover, vehicles that stop on traffic light for a long period of time and start moving again produce false detection due to ghost and aperture effects.

All previous techniques use either a uni-model or multi-model to express the background pixels discarding the foreground variations, where the object of interest may lack motion in some cases. The selective update of background pixels only Contaminate foreground objects into background model and leave the other pixels outdated. On the other hand, appearance-based technique discard background scene and use prior knowledge at higher computational complexity to detect foreground objects even it lacks motion. Hence, considering both foreground and background pixels to model the scene will enhance the segmentation accuracy and improve detection capability especially in urban environments. Moreover, Video based traffic surveillance systems require massive amount of image processing that must be performed in real time.

The proposed technique aims to avoid post processing and computational complexity, while maintaining stopped objects as part of the foreground no matter how long the stop time gap. Simultaneously, the background model must adapt variations in the background scene or illumination conditions. Therefore, special attention must be paid in deciding when and how to update both foreground and background models to avoid misclassification of pixels and keep both models up-to-date.

In this work, foreground background bimodal segmentation technique with low computational requirement is considered for embedded system implementation. In general, it is preferable to minimize floating point computation, which can be achieved by extending the previously proposed cumulative frame differencing (CFD) [8] together with Sigma-delta filter in a bimodal segmentation technique. The use of recursive algorithm for CFD and median estimation provide a simple and fast computation at a low memory requirement. CFD is computed by adding the frame difference to CFD recursively while the running estimate of the median is incremented or decremented by one if the input pixel is above or below the estimation respectively. The sigma delta filter is also used to compute the time variance of background pixels, which is used as a dynamic threshold for segmentation.

The rest of this paper is organized as follows: The following section provides an overview of material and methods which include cumulative frame differencing sigma-delta background estimation and the proposed foreground background bimodal segmentation technique. Results and discussion are presented in section III. Conclusion and future work are discussed in section IV.

II. The Material and Method

A. Basic Sigma-Delta Estimation (SDE)

The basic principle of this technique is the use of simple recursive non-linear operator based on sigma delta filter to estimate two orders of temporal statistics for every pixel in the frame [11]. Assuming that F_t is the current input frame at time t, B_t is the background model that is initialized using the first frame in the sequence (B_0=F_0) and the temporal variance estimator V_t, which represent the variability of pixel intensity and assumed to be initially zero (V_0=0). The sign function *sgn(x)* used to estimate the background and variance is defined as:

$$\text{sgn}(x)=\begin{cases}+1 & \text{if } x>0\\ 0 & \text{if } x=0\\ 1 & \text{if } x<0\end{cases} \tag{1}$$

The first step in the recursive estimation is to compute the image of absolute difference Δ_t as the initial differential estimate:

$$\Delta_t(x,y)=|F_t(x,y)\text{-}B_t(x,y)|. \tag{2}$$

The sigma-delta filter is also used to estimate the time variance for each pixel as a measure that represent its motion activity and used to distinguish whether the pixel is probably stationary or moving.

$$V_t(x,y)=V_{t\text{-}1}(x,y)+sgn\big(N\times\Delta_t(x,y)\text{-}V_{t\text{-}1}(x,y)\big) \tag{3}$$

Variance computation uses a multiple N (N=1-4) of the non-zero differences, which can distinguish pixels whose variation rate exceed its temporal activity significantly. In this way V_t will have the dimension of temporal standard deviation.

Finally, the binary detection mask is computed by comparing the absolute difference Δ_t with the variance V_t:

$$D_t(x,y)=\begin{cases}1 & if\ \Delta_t(x,y)>V_t(x,y)\\ 0 & if\ \Delta_t(x,y)\leq V_t(x,y)\end{cases} \tag{4}$$

The background pixels are selectively updated with relevance feedback using the sign function, which estimate the increase or decrease in the background pixel intensity as:

$$B_t(x,y)=B_{t\text{-}1}(x,y)+sgn\big(F_t(x,y)\text{-}B_{t\text{-}1}(x,y)\big) \tag{5}$$

Thus, it approximates the median of consecutive frame pixels. The use of relevance feedback prevents contamination of moving object into background model.

B. *Cumulative Frame Differencing (CFD)*

Cumulative frame differencing as proposed in [8], aims to model moving pixels of foreground objects. It can be defined by the recursive sum of consecutive frames difference. So, each pair of consecutive frames is subtracted, and the difference is added to the cumulative frame difference (CFD). Assume that F_t is the current frame and F_{t-1} is the previous one, CFD is calculated as follows:

$$CFD_t(x,y)=CFD_{t-1}(x,y)+(F_t(x,y)-F_{t-1}(x,y)) \tag{6}$$

Accordingly, pixel variations will be accumulated over time. Since the grayscale pixel value lie in the range (0-255), the difference between consecutive frames can be positive or negative. Thus, CFD values can range from -255 to 255, depending on the variation of grayscale intensity (increasing or decreasing). The limited variation in background pixel intensity will force its corresponding CFD to be very small and close to zero. The foreground pixel variation on the other hand, will be higher or unlimited, thus it will force CFD to be large in either positive or negative direction. The large variations correspond to foreground objects (i.e. Vehicles), either moving or stopped for a short or long time. Thus, it enables vehicle detection if it has motion history.

To discriminate foreground objects from background scene, a dynamic threshold is used. The standard deviation of the absolute CFD is estimated as a global variance threshold, it is multiplied by an experimentally estimated factor between 2 to 3 as:

$$Th(x,y)=2.5\times std(|CDF|) \tag{7}$$

After thresholding the detection mask $D_t(x,y)$is given 1 for pixels that have CFD greater than or equal the estimated global variance and 0 otherwise as:

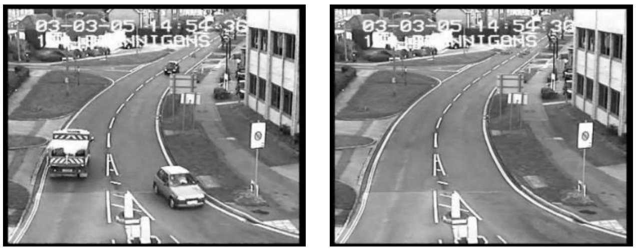
$$D_t(x,y)=\begin{cases}1 & CFD_t(x,y)\geq Th(x,y)\\ 0 & CFD_t(x,y)<Th(x,y)\end{cases} \tag{8}$$

C. *Proposed CFD-SDE bimodal*

This section describes the proposed technique which consists of three modules: initialization, foreground background modelling and segmentation. The foreground background modelling consists of two modules cumulative frame differencing and sigma-delta background estimation in a single CFD-SDE bimodal.

1) Initialization: The initialization module consists of the basic sigma delta background estimation with relevance feedback. It is required to run the technique for a sufficient period of time T to achieve an accurate initial background model. Thus, if the frame rate is 25fps and the initialization time period is 6s then the starting M frames used for initialization will be the first 150 frames. Fig. 1 shows frame number 150 and the generated initial background model B0. After that, the bimodal uses the initial background model to initialize the cumulative frame differencing as:

$$CFD_0(x,y)=F_t(x,y)-B_0(x,y)) \tag{9}$$



(a) (b)

Fig. 1 (a) Frame number 150 and (b) Initial background model B0

2) Foreground background bimodal CFD-SDE bimodal: In the proposed technique, each pixel of the traffic scene can be considered as a bimodal distribution that contain a mixture of two independent unimodal distributions to represent foreground and background respectively. Thus, modeling foreground pixels using cumulative frame differencing and estimating background pixels using sigma-delta estimation will yield two independent representations of the scene.

One of the major drawbacks for the application of sigma-delta technique in urban traffic environments is that, the variance grows when vehicles pass over the background, which degrade detection because the threshold becomes too high. Thus, it is required to achieve a more selective update for the background and its variance. In the proposed technique, the variance is intended to represent the variability of background pixel intensities when no objects are over that pixel. Thus, Vt will not be updated for the pixels covered with moving foreground object. The selective update of the background and variance will prevent the growth of threshold value when vehicles pass over it as:

$$D_t^{BK}(x,y) = \begin{cases} 1 & if\ \Delta_t(x,y) > V_t(x,y) \\ 0 & if\ \Delta_t(x,y) \leq V_t(x,y) \end{cases} \tag{10}$$

Due to the fact that background pixels are fixed or having limited variations, it's corresponding CFD will remain limited and close to zero. Background pixels that have higher variation due to sudden illumination will be re-initialized using the sigma delta background model as:

$$CFD_t(x,y) = \begin{cases} CFD_t(x,y) \\ F_t(x,y) - B_{t-1}(x,y) \end{cases} \tag{11}$$

On the other hand, foreground pixels will have higher or unlimited variations that will force corresponding CFD far from zero in either positive or negative direction. These variations represent all moving vehicles together with slow or temporary stopped vehicles that was moving in earlier frames, which enables continuous detection of any vehicle even if it stopped for a long period of time, as long as it has an old motion history.

In order to isolate foreground pixel, a dynamic threshold value is also used. In this paper, foreground detection mask $D_t^{FG}(x,y)$is generated using sigma delta variance which is not affected by the foreground variation as in eq. (12). The dynamic threshold will keep detecting foreground object since it is not updated in the background model.

$$D_t^{FG}(x,y) = \begin{cases} 1 & CFD_t(x,y) \geq Vt(x,y) \\ 0 & CFD_t(x,y) < Vt(x,y) \end{cases} \tag{12}$$

The bimodal technique provides a balance between adaption to illumination or background variations and foreground detection without contaminating slow or stopped vehicle for any time period. Thus, it will keep detecting the vehicles until they start moving again.

III. Results and Discussion

The hardware platform used to implement the proposed technique is a Dell Laptop with an Intel Core i5 2.3 G Hz CPU and 8 GB RAM and Windows 10 platform. MATLAB 2013a software is used for the development and evaluation process. The videos were taken from i-LIDS (image library of intelligent detection systems) dataset.

To validate the proposed technique, it was compared with other approaches representative of the state of the art in terms of segmentation and detection, such as basic SDE with relevance feedback and GMM. Fig. 2 shows segmentation and detection results for SDE, GMM, CFD and CFD-SDE bimodal using three frame samples selected between frame 300 and 360. In this sample a vehicle has stopped for about 2 second (60 frames). The stopped vehicle starts vanishing in SDE and GMM, while CFD and CFD-SDE bimodal keep detecting the whole vehicle until it starts moving, with lower false detection in CFD-SDE bimodal.

Slow motion and illumination variation are shown in Fig. 3. CFD-SDE bimodal detect slow motion using CFD model and adapt illumination variation by CFD reinitialization. As compared to GMM the proposed CFD-SDE bimodal demonstrate an improved and accurate segmentation results.

Fig. 4 provides a performance comparison for parked vehicle using SDE, GMM, CFD and CFD-SDE bimodal.

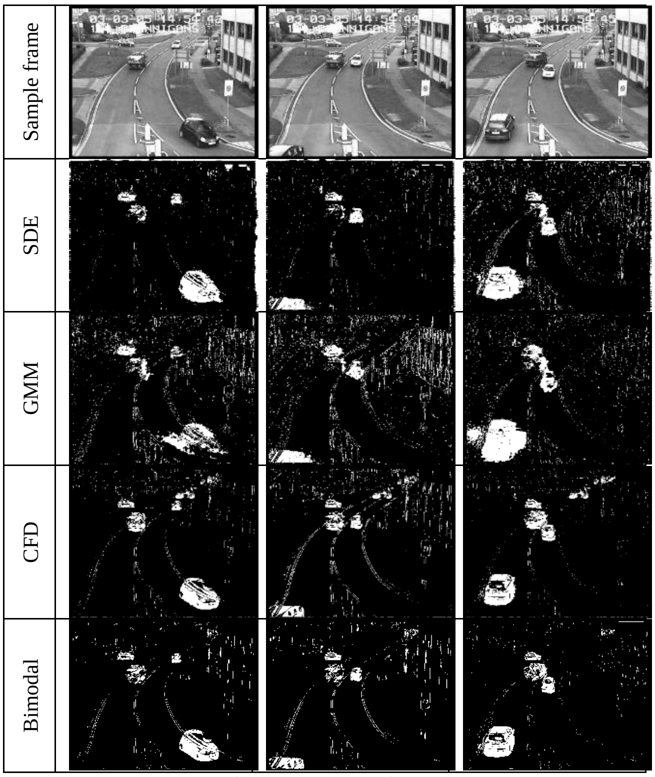


Fig. 2 Detection results. Left to right, sample frame, sigma-delta background subtraction, Gaussian mixture model, cumulative frame differencing and the detection mask of the proposed technique

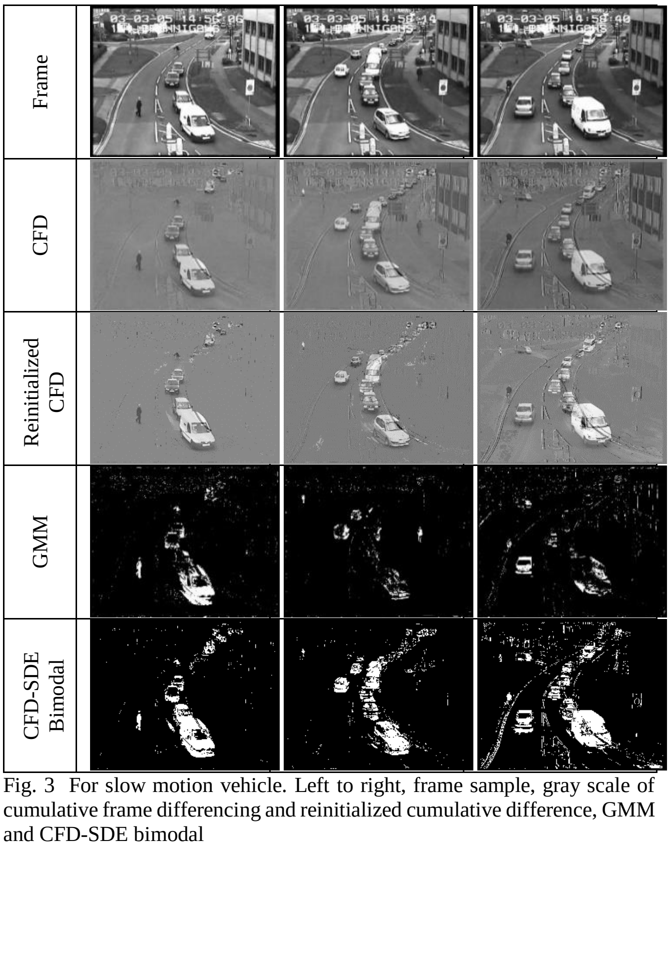


Fig. 3 For slow motion vehicle. Left to right, frame sample, gray scale of cumulative frame differencing and reinitialized cumulative difference, GMM and CFD-SDE bimodal

For SDE and GMM techniques the parked vehicle will vanish with time as seen in frame 24500. On the other hand, the detection mask of CFD and CFD-SDE bimodal keep detecting the parked vehicle until frame 25650 after about 100 second. Therefore, slow or temporary stopped vehicles are detected clearly, parked vehicles can be detected regardless of the long parking time.

The main difficulty the selected frame sequence is the slow motion and temporary stopped vehicles. The test results show that the proposed technique outperform the comparative techniques in many ways. Moreover, it can deal better with illumination variation.

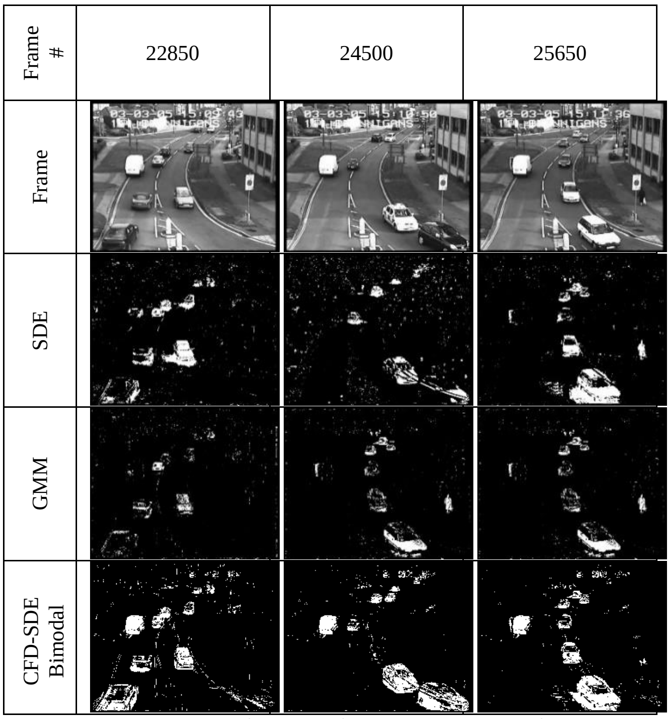


Fig. 4 Detection results. Left to right, frame number, sample frame, sigma-delta background subtraction, Gaussian mixture model and the detection mask of the proposed technique CFD-SDE bimodal

IV. Conclusion

CFD-SDE bimodal segmentation is proposed to detect vehicles in urban environments. It combines cumulative temporal motion with background estimation to classify pixels into the foreground or background based on dynamic thresholding criteria. The proposed technique tries to maintain computational efficiency, while achieving better robustness for typical urban traffic scenarios. Combining motion history with background estimation improve detection of slow or temporary stopped vehicles. Moreover, the use of background model to reinitialize CFD will account for the sudden illumination variation.

The proposed technique uses simple and efficient arithmetic computation as compared with background subtraction techniques, thus it will be more suitable for real time applications.

References

[1] T. Gao, Z.G. Liu, W.C. Gao and J. Zhang, "Moving vehicle tracking based on SIFT active particle choosing," *In International Conference on Neural Information Processing, Springer, Berlin, Heidelberg,* pp. 695-702, Nov. 2008.

[2] M. Al-Smadi, K. Abdulrahim and R. Abdulsalam, "Traffic surveillance: A review of vision based vehicle detection, recognition and tracking," *International journal of applied engineering research*, vol. 11(1), pp. 713-726, 2016.

[3] Q.L. Li, and J.F. He, "Vehicles detection based on three-frame-difference method and cross-entropy threshold method," *Computer Engineering*, vol. 37(4), pp. 172-174, 2011.

[4] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40(7), pp. 1726-1740, 2018.

[5] M. Piccardi, "Background subtraction techniques: a review," *In IEEE international conference on Systems, man and cybernetics*, vol. 4, pp. 3099-3104, IEEE. Oct. 2004.

[6] Y. Liu, Y.Lu, Q. Shi, and J. Ding, "Optical flow based urban road vehicle tracking," *In International Conference on Computational Intelligence and Security (CIS)*, 9th pp. 391-395, IEEE. Dec. 2013.

[7] R. Manikandan, and R. Ramakrishnan, "Video object extraction by using background subtraction techniques for sports applications*," Digital Image Processing*, vol. 5(9), pp. 435-440, 2013.

[8] M. Al-Smadi, A. Khairi, and R. Abdulsalam, "Cumulative frame differencing for urban vehicle detection," *In First International Workshop on Pattern Recognition*, vol. 10011, pp. 100110G. International Society for Optics and Photonics, 2016.

[9] C.R. Wren, A. Azarbayejani, T. Darrell and A.P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.19(7), pp. 780-785, 1997.

[10] N.J. McFarlane and C.P. Schofield, "Segmentation and tracking of piglets in images," *Machine vision and applications*, vol. 8(3), pp. 187-193, 1995.

[11] A. Manzanera, and J.C. Richefeu, "A new motion detection algorithm based on Σ–Δ background estimation," *Pattern Recognition Letters*, vol. 28(3), pp. 320-328, 2007.

[12] C. Stauffer and W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," *In cvpr* pp. 2246, IEEE. Jun. 1999.

[13] P. Barcellos, C. Bouvié, F.L. Escouto, and J. Scharcanski, "A novel video based system for detecting and counting vehicles at user-defined virtual loops," *Expert Systems with Applications*, vol. 42(4), pp. 1845-1856, 2015.

[14] A.Elgammal, D. Harwood, and L. Davis, June. "Non-parametric model for background subtraction," *In European conference on computer vision, Springer, Berlin, Heidelberg*. vol. II751- 767, pp. 751-767, 2000.

[15] K. Kim, T.H. Chalidabhongse, D. Harwood and L. Davis, "Real-time foreground–background segmentation using codebook model," *Real-time imaging*, vol. 11(3), pp. 172-185, 2005.

[16] K. Toyama, J. Krumm, B. Brumitt and B. Meyers, "Wallflower: Principles and practice of background maintenance," *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, pp. 255-261, 1999.

[17] N.M. Oliver, B. Rosario and A.P Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 22(8), pp. 831-843, 2000.

[18] A. Manzanera, "σ-δ background subtraction and the Zipf law," *In Ibero American Congress on Pattern Recognition, Springer, Berlin, Heidelberg,* pp. 42-51, Nov. 2007.

[19] M. M. Abutaleb, A. Hamdy, M. E. Abuelwafa and E. M. Saad, "FPGA-based object-extraction based on multimodal Σ-Δ background estimation," *2nd International Conference on in Computer, Control and Communication,* IEEE, pp. 1-7, 2009.

[20] S. Toral, M. Vargas, F. Barrero, and M.G. Ortega, "Improved sigma-delta background estimation for vehicle detection," *Electronics letters*, vol. 45(1), pp. 32-34, 2009.

[21] M. Vargas, J.M. Milla, S.L. Toral and F. Barrero, "An enhanced background estimation algorithm for vehicle detection in urban traffic scenes," *IEEE Transactions on Vehicular Technology*, vol. 59(8), pp. 3694-3709, 2010.

[22] L. Lacassagne, A. Manzanera, and A. Dupret, "Motion detection: Fast and robust algorithms for embedded systems," *16th IEEE International Conference on Image Processing (ICIP),* pp. 3265-3268, Nov. 2009.

[23] K. Li and Y. Yang, "A method for background modelling and moving object detection in video surveillance," *4th International Congress on Image and Signal Processing (CISP),* vol. 1, pp. 381-385, Oct. 2011.