Journal of Engineering and Applied Sciences 11 (3): 414-419, 2016
ISSN: 1816-949X

Background Subtraction in Urban Traffic Video Using Recursive Sigma-Delta Mixture Model

[1]Ma'moun Al-Smadi, [1]Khairi Abdulrahim, [1,2]Rosalina Abdul Salam and [1]Ahmad Alajarmeh
[1]Faculty of Science and Technology, Universiti Sains Islam
Malaysia (USIM), Negeri Sembilan, Malaysia
[2]Islamic Science Institute (ISI), UniversitiSains Islam Malaysia (USIM), Negeri Sembilan, Malaysia

Abstract: Motion segmentation is a fundamental step in urban traffic surveillance systems, since it provides necessary information for further processing. Background subtraction techniques are widely used to identify foreground moving vehicles from static background scene. Conventional techniques utilize single background model or Gaussian mixture model 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. To address these problems Sigma-Delta Mixture Model (SDMM) is proposed. Mixed distributions are updated dynamically based on matching and contribution in the two order temporal statistics. The constant amplification factor is replaced by weighted factor to update the variance rate over its temporal activity. The proposed technique achieve robust and accurate performance,which improves adaptation capability with balanced sensitivity and reliability, moreover, integerlinear operations enables the real-time capability.

INTRODUCTION

Modern trends in traffic monitoring systems are oriented toward video based systems which provide more information about the traffic of vehicles (e.g., count, speed, type, color, make and model). Such systems have relatively low coast with wide variety of applications especially in urban environments. Vehicle detection forms a basic step in video analysis for automatic traffic monitoring and surveillance systems (Piccardi, 2004). Thus, computer vision techniques used in vehicle tracking and classification require a robust and fast detection mechanism thatlocalize the moving vehicles as it appears in the scene or in each consecutive frames. There are two main categories for vehicle detection; firstly motion based techniques,which includes; frame differencing (Li and He, 2011), background subtraction (Manikandan and Ramakrishnan, 2013) and optical flow (Liu *et al*., 2013). Secondly, appearance based techniques (Gao *et al*., 2009) that extract edges, feature points or use prior knowledge to segment the vehicles in eachisolated frame.

Background modelling and subtraction is a common technique for detecting moving vehicles in videos captured by a static camera (Toral *et al*., 2009). The basic idea is to model the background scene and use it as a reference image to be compared with each consecutive frame to extract moving foreground. Therefore, it must represent the stationary background accurately and update it frequently to adapt variations in traffic or environmental conditions by taking into account motion discrimination, background complexity and illumination variations. Many recent studies on background subtraction have been developed to detect moving objects these studies can be classified into parametric, nonparametric and predictive techniques (Wan and Wong, 2008).

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 (Wren *et al*., 1997), temporal median filter (McFarlane and Schofield, 1995), sigma-delta filter (Manzanera and Richefeu 2007) and Gaussian Mixture Model (GMM) (Stauffer and Grimson, 1999).

Running Gaussian average useGaussian density functionrecursively to represent each pixel (Wren *et al*., 1997). The temporal median filter consists of non-recursive median and approximate median.Non-recursive technique finds the median for recent frames which require large memory storage and costly computation. Approximate median proposed in (McFarlane and Schofield, 1995) estimates the background recursively based on the assumption that the pixel stays in the

Corresponding Author: Ma'moun Al-Smadi, Faculty of Science and Technology, Universiti Sains Islam, Malaysia (USIM), Negeri Sembilan, Malaysia

background for more than half of the period under consideration. Manzanera and Richefeu (2007), 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. However, these approaches remain challenging for congested traffic or when the background undergoes sudden illumination variations (Manzanera and Richefeu, 2007).

GMMwas introduced by Chris stauffer and W.E.L Grimson in 1999 (Stauffer and Grimson, 1999). It 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 by evaluating its stability. If the pixel distribution is stable above threshold, then it is classified as background pixel. GMM can adapt illumination variations and repetitive clutter with higher computationand memory requirements compared with standard background subtraction techniques (Barcellos *et al*., 2015).

Nonparametric techniques have more ability to handle arbitrary density functions thus, they are more suitable for complex functions that cannot be parametrically modeled. Kernel Density Estimation (KDE) is an example of such techniques (Elgammal *et al*., 2000). It uses KDE to estimate the background probabilities of each pixel from many recent samples. Previous techniques are limited to smooth behavior 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 (Kim *et al*., 2005) in which a set of dynamically handled codewords are used to replace parameters represented by probabilistic function. A modified codebook technique with better performance was introduced and compared in (Ilyas *et al*., 2009).

Finally, predictive techniques employ predictive procedures in modeling and predicting the state dynamic of each background pixel. Based on Kalman filtering, background pixels were modeled in (Heikkila and Silven, 2004) by using a state space that combines pixel intensity and its spatial derivative. Additional predictive technique uses Wiener filter or autoregressive models (Toyama *et al*., 1999). Eignspace reconstruction or eigen background was also used in (Oliver *et al*., 2000) to model background with complex computation.

The use of sigma-delta in background subtraction attract many researchers due to its computational efficiency (Lacassagne *et al*., 2009). It requires only basic integer operations that includes comparison, increment and absolute difference. The robustness of this technique is comparable with other parametric techniques that have higher computational cost.

This study propose a new background subtraction technique using Sigma-Delta Mixture model SDMM. It does not require the use of color information, and the obtained results are strictly based on grayscale image processing. In this study, the mean update use the estimation for single matched distributions, while complete match update distributions with highest weight from the input frame. Moreover, a weighted amplification factor that depends on the matched background pixels is used to improve variance sensitivity. A more accurate and robust background model was achieved in complex background situations which is typical for urban traffic scenes. In addition it satisfies the real time requirements by using integer and linear operations only. The proposed technique combines the advantage of high computational efficiency of the basic sigma delta with the temporal recursive update to cope with background clutter and complexity.

Sigma-delta background estimation: 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 each pixel (Manzanera and Richefeu, 2007). The initial background estimation B_t is the first frame in the sequence $F_t$and the initial variance is set to zero ($V_t = 0$). After that, the image of absolute difference is computed as:

$$\Delta_t(x,y) = |F_t(x,y)\text{-}B_t(x,y)| \tag{1}$$

Background estimation is based on the increase or decrease of pixel intensity that iscalculated using thesign function (sgn) as:

$$B_t(x,y) = B_{t-1}(x,y) + \text{sgn}(F_t(x,y) - B_{t-1}(x,y)) \tag{2}$$

Thus, it can be considered as an approximation of frames median F_t. The time variance between pixels within consecutive frames is also calculated using the sign function as follows:

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

This will provide a measure of temporal activity with N used as an amplification factorin the range (Eq. 1-4). In other words, variance update will depend on the difference between the variation rate and temporal activity for each pixel. After that, the binary detection mask can becalculatedas shown in Eq. 4 in which intensity variance perform dynamic thre sholding to isolate foreground pixels from the estimated background pixels.

$$D_t(x,y) = \begin{cases} 1 \text{ if } \Delta_t(x,y) > V_t(x,y) \\ 0 \text{ if } \Delta_t(x,y) \le V_t(x,y) \end{cases} \quad (4)$$

A slight modification on the basic algorithm is done to improve background stability through selectively update background pixels only with relevance feedbackas:

$$B_t(x,y) = B_{t-1}(x,y) + \text{sgn}(F_t(x,y)B_{t-1}(x,y)).(1-D)_{t-1}(x,y)) \quad (5)$$

This modification preventscontamination of moving objectinto background model. In (Manzanera, 2007), a Zipf-Mandelbort distribution was used to update the background according to the dispersion of distribution. Spatiotemporal processing proposed in (Manzanera and Richefeu, 2007) improves the detection by removing non-significant pixels. The additional processing improves detection, yet adaption to complex scenes still an eventual problem. To overcome this problem, multiple-frequency sigma-delta was introduced in (Manzanera and Richefeu, 2007). In this technique, weighted sum was computed from multiple models with different updating periods. Another sigma-delta multi-model was proposed in (Abutaleb *et al*., 2009) using a mixture of three distributions. They used a weight as a voting mechanism to sort the mixture according to higher and lower updating value.

Confidence measurement was introduced in (Toral *et al*., 2009) and enhanced in (Vargas *et al*., 2010) to adapt slow motion and congested traffic. 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 (Lacassagne and Manzanera, 2009) a hierarchical or bi-level sigma-delta filtering was proposed which perform conditional update that includeslow level temporal update and high level spatial update. Selective and partial updates using global variance was developedin (Li and Yang, 2011) which provide a good balance between sensitivity and reliability at the expense of high computation.

However, sigma delta based techniques quickly degrades under slow or congested traffic conditions. This is due to the integration of moving pixel intensities into background model (Manzanera and Richefeu, 2007). Moreover, vehicles that stop for a while and start moving again produce false detection resulting from ghost and aperture effects.

MATERIALS AND METHODS

Sigma-delta mixture model: In the proposed technique each background pixel is modeled using a mixture of k distributions similar to the GMM and the work in (Manzanera and Richefeu, 2007) and (Abutaleb *et al*., 2009). Each distribution is characterized by sigma-delta mean estimation B_t, variance V_t and variable weight W_t maintained at time t. The delta is computed as the absolute difference between each pixel in the new frame and its correspondence mean estimation in each distribution as:

$$\Delta_t^k(x,y) = \left|F_t(x,y) - B_t^k(x,y)\right|$$

After that delta is compared with corresponding distribution variance to decide which of them match the background distribution as:

$$\text{Match}_t^k = \begin{cases} 1 & \Delta_t^k(x,y) \le V_t^k(x,y) \\ 0 & \text{otherwize} \end{cases} \quad (6)$$

A complete match score Match_t^k is computed from the matching value of each distribution Match_t^k:

$$\text{Match}_t^k = \text{Match}_t^1.\text{Match}_t^2...\text{Match}_t^k \quad (7)$$

Different from the previous techniques, weighted amplification factor is updated for each distribution based on its matching value in order to improve its sensitivity against background variations.Then, the match is used to compute the amplification factor for each distribution as:

$$N_t^k = \begin{cases} N_{t-1}^k + 1 & \text{Match}_t^k = 1 \\ N_{t-1}^k - & \text{Match}_t^k = 0 \end{cases} \quad (8)$$

Additional limitation is applied to limit its value remain in the range $((1 \le N_t^k \le 6))$. Thus, it can be used todetermine the sensitivity of local variance for each distribution. In this way, the measurement of pixels temporal activity against various distribution will depend on how often this pixel is being in the background. This will improve the detection of slow or temporary stopped vehicles. Unlike previous techniques where the amplification factor is fixed while here it is proportional to the cumulative matching index.Therefore, it will reduce the false detection of noisy pixels and improve detection.

If the pixel in the current frame matches any distribution, it is classified as a background pixel and the mean of the matched distribution is updated. Pixels that match all distributions are considered as complete match and the distribution with maximum weight is given the pixel value from the current frame as:

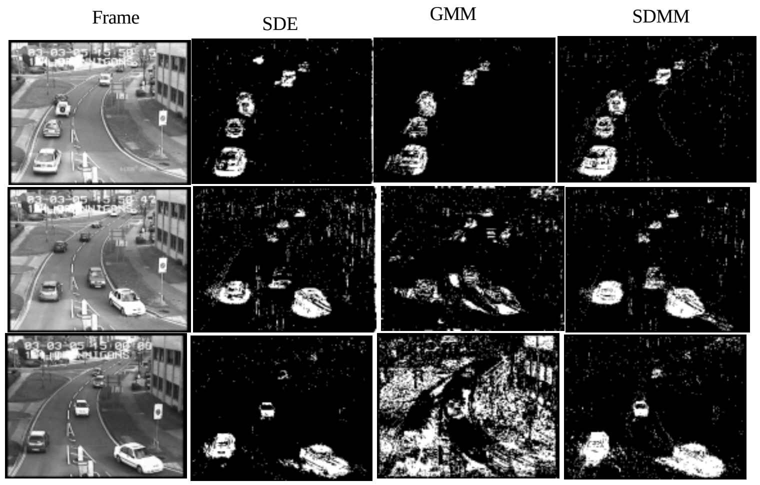

Fig. 1: PV_MEDIUM sequence from i-LDS comparative result of the three background subtraction techniques at frames 700, 1500, 8440, respectively

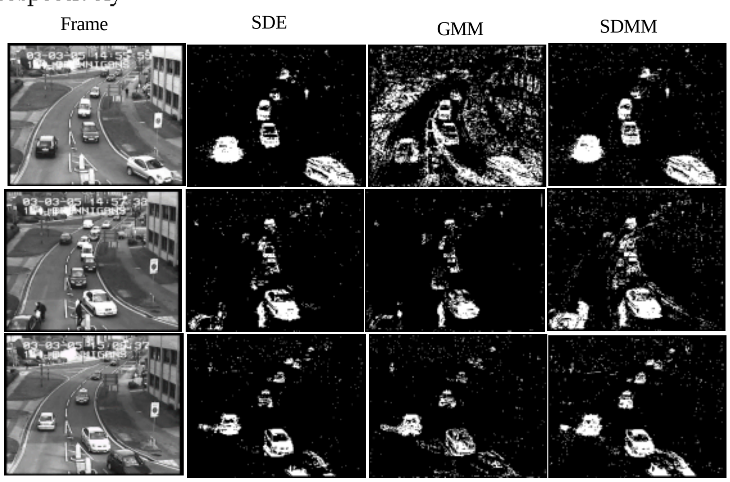

Fig. 2: PV_EVAL sequence from i-LDS comparative result of the three background subtraction techniques at frames 2220, 4540, 21200, respectively

$$B_t^k + \begin{cases} B_t^k + \mathrm{sgn}(F_t - B_{t-1}^k) & \text{if } \mathrm{Match}_t^k = 1 \\ F_t & \text{if } \mathrm{Match}_t^k = 1 \text{ and } W_t = \max(W_t^k) \end{cases} \quad (9)$$

The final detection mask of SDMM will be against the pixels that does not have a complete match.

$$D_t(x,y) = \begin{cases} 1 \text{ if } \mathrm{Match}_t^k = 0 \\ 0 \text{ if } \mathrm{Match}_t^k = 1 \end{cases} \quad (10)$$

If none of the k distributions match the current pixel intensity, then it will be considered as a foreground pixel.

RESULTS AND DISCUSSION

The experimental setup was conducted using MATLAB 2013a on an Intel corei5 2.3 Ghz Laptop with 8G RAM and Windows 10 platform. To evaluate the performance of the proposed technique, experiment was performed on twodataset. The first one is provided by the Home Office of the United Kingdom named i-LIDS (Image Library of Intelligent Detection Systems).

Sample frames of the PV_MEDIUM and PV_EVAL sequences with 576×720 pixels are used forevaluation. The second dataset is Traffic sequence showing an intersection at Rheinhafen, Karlsruhe with frame size of 688×565pixels.The selection take into account various situation that includes slow or temporary stopped vehicles, sudden illumination variations and variation in vehicle pose and orientation. To compare results, basic sigma-delta with relevance feedback and Gaussian mixture modelwere used as shown in Fig. 1-3.

In Fig. 1, PV_MEDIUM is used to evaluate background subtraction for three techniques that are the

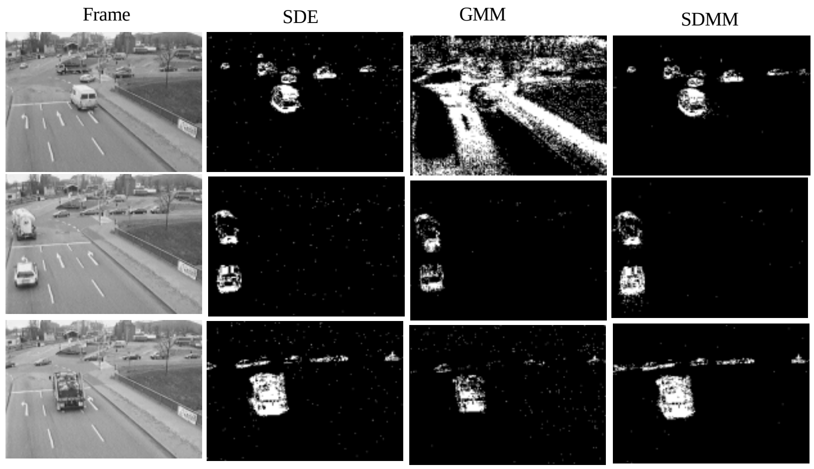

Fig. 3: Intersection sequence comparative result of the three background subtraction techniques at frames 170, 780, 950, respectively

basic sigma-delta, Gaussian mixture model and the mixture of sigma-delta respectively. The sample frames are 700, 1500, 8440. In this case, vehicles move at low speed in the first and second frame. As seen in the figure the slow motion degrade the detection in SDE and GMM with better detection result in SDMM while sudden illumination variation highly affect the background scene for GMM as shown in the last frame.

Next, the PV_EVAL sequence is evaluated. The main difficulty in this sequence is the temporary stopped vehicles and rapid illumination variations.Three frames have been selected to demonstrate the variations in the different techniques. In the case of temporary stopped vehicles, the variance of SDMM is still sensitive to the dense traffic areas as seen in Fig. 2 while it gives poor detection in the other techniques. Thus, the background will not be polluted with slow or temporary stopped vehicles.

Another urban sequence featuring light variation and slow motion is the intersection at Rheinhafen, Karlsruhe. It can be noticed from Fig. 3 that the proposed technique gives better description of the moving vehicles even at low speed.

In all three sequences, the comparative techniques either over detect or under detect the moving vehicles, on the other hand the proposed technique SDMM has a good balance which improves positive detection with limited fluctuation. Moreover, it can deal better with illumination variation and background vibrations.

CONCLUSION

Background modeling and subtraction technique using sigma-delta mixture model SDMM is proposed to detect foreground moving vehicles in urban environments. It performs background estimation using a mixture of sigma-delta filters with weighted amplification factor and complete matching update for the maximum weighted distributions.The proposed technique tries to maintain computational efficiency while achieving better robustness for typical urban traffic scenarios. Using variable amplification factor that take into account the matching distribution will improve variance update dynamically which in term improves detection of slow or temporary stopped vehicles. Moreover, distribution updates are performed based on the weight and matching to account for sudden illumination variation.

The proposed technique is primarily composed of linear operations with low complexity compared to other comparable techniques thus it will be more suitable for real time applications. To account for the limitations of complex scenarios such as clutter and longtime stopped vehicles, future work will address motion intensity modelto reduce the irrelevant background oscillationsand increase the directional motion difference.

ACKNOWLEDGEMENTS

This study presents a work that is supported by the Ministry Of Higher Education (MOHE), under the Research Acculturation Grant Scheme (RAGS/2013/ USIM/ICT07/2) and Science Fund Grant Scheme (USIM/SF/FST/30/30113).

REFERENCES

Abutaleb, M.M., A. Hamdy, M.E. Abuelwafa and E.M. Saad, 2009. FPGA-based object-extraction based on multimodal Ó-Ä background estimation. Proceedings of the 2nd International Conference on Computer, Control and Communication, February 17-18, 2009, Karachi, pp: 1-7.

Barcellos, P., C. Bouvie, F.L. Escouto and J. Scharcanski, 2015. A novel video based system for detecting and counting vehicles at user-defined virtual loops. Expert Syst. Applic., 42: 1845-1856.

Elgammal, A., D. Harwood and L. Davis, 2000. Non-parametric model for background subtraction. Proceedings of the 6th European Conference on Computer Vision Dublin, June 26-July 1, 2000, Ireland, pp: 751-767.

Gao, T., Z.G. Liu, W.C. Gao and J. Zhang, 2009. Moving Vehicle Tracking based on SIFT Active Particle Choosing. In: Advances in Neuro-Information Processing, Koppen, M., N. Kasabov and G. Coghill (Eds.). Springer, Berlin, Heidelberg, ISBN: 978-3-642-03039-0, pp: 695-702.

Heikkila, J. and O. Silven, 2004. A real-time system for monitoring of cyclists and pedestrians. Image Vision Comput., 22: 563-570.

Ilyas, A., M. Scuturici and S. Miguet, 2009. Real time foreground-background segmentation using a modified codebook model. Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, September 2-4, 2009, Genoa, Italy, pp: 454-459.

Kim, K., T.H. Chalidabhongse and D. Harwood, 2005. Real-time foreground-background segmentation using codebook model. Real-Time Imag., 11: 172-185.

Lacassagne, L., A. Manzanera and A. Dupret, 2009. Motion detection: Fast and robust algorithms for embedded systems. Proceedings of the International Conference on Image Processing, November 7-10, 2009, Cairo, Egypt, pp: 3265-3268.

Li, K. and Y. Yang, 2011. A method for background modeling and moving object detection in video surveillance. Proceedings of the 4th International Congress on Image and Signal Processing, Volume 1, October 15-17, 2011, Shanghai, China, pp: 381-385.

Li, Q.L. and J.F. He, 2011. Vehicles detection based on three-frame-difference method and cross-entropy threshold method. Comput. Eng., 37: 172-174.

Liu, Y., Y. Lu, Q. Shi and J. Ding, 2013. Optical flow based urban road vehicle tracking. Proceedings of the 9th International Conference on Computational Intelligence and Security, December 14-15, 2013, Leshan, China, pp: 391-395.

Manikandan, R. and R. Ramakrishnan, 2013. Video object extraction by using background subtraction techniques for sports applications. Digital Image Processing, 5: 435-440.

Manzanera, A. and J.C. Richefeu, 2007. A new motion detection algorithm based on Ó-Ä background estimation. Pattern Recognit. Lett., 28: 320-328.

Manzanera, A., 2007. ó-ä Background Subtraction and the Zipf Law. In: Progress in Pattern Recognition, Image Analysis and Applications, Rueda, L., D. Mery and J. Kittler (Eds.). Springer, Berlin, Heidelberg, ISBN: 978-3-540-76724-4, pp: 42-51.

McFarlane, N.J.B. and C.P. Schofield, 1995. Segmentation and tracking of piglets in images. Mach. Vision Applic., 8: 187-193.

Oliver, N.M., B. Rosario and A.P. Pentland, 2000. A Bayesian computer vision system for modeling human interactions. IEEE Trans. Pattern Anal. Mach. Intell., 22: 831-843.

Piccardi, M., 2004. Background subtraction techniques: A review. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, October 10-13, 2004, IEEE, USA., pp: 3099-3104.

Stauffer, C. and W.E.L. Grimson, 1999. Adaptive background mixture models for real-time tracking. Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, June 23-25, 1999, Los Alamitos, CA., USA., pp: 246-252.

Toral, S., M. Vargas, F. Barrero and M.G. Ortega, 2009. Improved sigma-delta background estimation for vehicle detection. Electron. Lett., 45: 32-34.

Toyama, K., J. Krumm, B. Brumitt and B. Meyers, 1999. Wallflower: Principles and practice of background maintenance. Proceedings of the 7th IEEE International Conference on Computer Vision. September 20-27, 1999, Piscataway, NJ., USA., pp: 255-261.

Vargas, M., J.M. Milla, S.L. Toral and F. Barrero, 2010. An enhanced background estimation algorithm for vehicle detection in urban traffic scenes. IEEE Trans. Vehicular Technol., 59: 3694-3709.

Wan, Q. and Y. Wang, 2008. Background subtraction based on adaptive non-parametric model. Proceedings of the 7th World Congress on Intelligent Control and Automation, June 25-27, 2008, Chongqing, China, pp: 5960-5965.

Wren, C.R., A. Azarbayejani, T. Darrell and A.P. Pentland, 1997. Pfinder: Real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell., 19: 780-785.