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

Ma'moun Al-Smadi, Khairi Abdulrahim, Rosalina Abdul Salam and Ahmad Alajarmeh
Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Negeri Sembilan, Malaysia
Islamic Science Institute (ISI), Universiti Sains 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, integer linear 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 cost 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 that localize 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 each isolated 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 Richalet 2007) and Gaussian Mixture Model (GMM) (Stauffer and Grimson, 1999).

Running Gaussian average uses Gaussian density function recursively 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 Richiieu (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 Richiieu, 2007).

GMM was 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 computation and 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). Eigenspace 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 advantages 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 Richiieu, 2007). The initial background estimation $B_0$ is the first frame in the sequence $F_0$ and the initial variance is set to zero ($V_0 = 0$). After that, the image of absolute difference is computed as:

$$
\Delta_i(x,y) = |F_i(x,y) - B_{i-1}(x,y)|
$$

(1)

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

$$
B_i(x,y) = B_{i-1}(x,y) + sgn(F_i(x,y) - B_{i-1}(x,y))
$$

(2)

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

$$
V_i(x,y) = V_{i-1}(x,y) + sgn(N \times \Delta_i(x,y) - V_{i-1}(x,y))
$$

(3)

This will provide a measure of temporal activity with $N$ used as an amplification factor to 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
be calculated as shown in Eq. 4 in which intensity variance perform dynamic thresholding to isolate foreground pixels from the estimated background pixels.

\[
D_i(x, y) = \begin{cases} 
1 & \text{if } \Delta_i(x, y) > V_i(x, y) \\
0 & \text{if } \Delta_i(x, y) \leq V_i(x, y) 
\end{cases}
\]  

(4)

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

\[
B_i(x, y) = B_{i-1}(x, y) + \text{sgn}(F_i(x, y))B_{i-1}(x, y)(1 - D_{i-1}(x, y))
\]  

(5)

This modification prevents contamination of moving objects into background model. In (Manzanera, 2007), a Zipf-Mandelbrot 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 includes slow level temporal update and high level spatial update. Selective and partial updates using global variance was developed in (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 \( \mu_i \), variance \( V_i \), and variable weight \( W_i \) 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:

\[
\Delta_i^k(x, y) = |F_i(x, y) - \mu_i^k(x, y)|
\]

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

\[
\text{Match}_i^k = \begin{cases} 
1 & \Delta_i^k(x, y) \leq V_i^k(x, y) \\
0 & \text{otherwise}
\end{cases}
\]  

(6)

A complete match score \( \text{Match}^k \) is computed from the matching value of each distribution \( \text{Match}_i^k \):

\[
\text{Match}^k = \text{Match}_1^k \cdot \text{Match}_2^k \cdots \text{Match}_k^n
\]  

(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:

\[
N_i^k = \begin{cases} 
N_{i-1}^k + 1 & \text{Match}_i^k = 1 \\
N_{i-1}^k - \text{Match}_i^k & \text{otherwize}
\end{cases}
\]  

(8)

Additional limitation is applied to limit its value remain in the range \((1 \leq N_i^k \leq 6)\). Thus, it can be used to determine 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_MEDİUM 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.

\[
B_k^t + \begin{cases} 
B_k^t + \text{sgn}(r_t^k - B_k^c) & \text{if Match}_k^t = 1 \\
F_t & \text{if Match}_k^t = 1 \text{ and } W_t = \max(W_k^t) 
\end{cases} 
\]  

(9)

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

\[
D_t(x,y) = \begin{cases} 
1 & \text{if Match}_k^t = 0 \\
0 & \text{if Match}_k^t = 1 
\end{cases} 
\]  

(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 2015a on an Intel core i5 2.3 Ghz Laptop with 8G RAM and Windows 10 platform. To evaluate the performance of the proposed technique, experiment was performed on two dataset. 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_MEDİUM and PV_EVAL sequences with 576×720 pixels are used forevaluation. The second dataset is Traffic sequence showing an intersection at Rheinhausen, Karlsruhe with frame size of 688×565 pixels. The selection took into account various situations 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 model were used as shown in Fig. 1-3.

In Fig. 1, PV_MEDİUM 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 Rheinhsafen, 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 updates dynamicaly which in turn 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 model to reduce the irrelevant background oscillations and 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**


419