MOVING OBJECT DETECTION USING BACKGROUND SUBTRACTION AND SHADOW REMOVAL FROM VIDEO

Chethan M N[1], S K Uma[2], B Ramachandra[3]

[1] *M.Tech student, Dept. of Computer science and engineering, PESCE Mandya, India*
[2] *Assistant Professor, Dept. of Computer science and engineering, PESCE Mandya, India*
[3] *professor and HOD, Dept. of Electrical and electronics engineering, PESCE Mandya, India*

ABSTRACT

Surveillance systems are widely used in various fields in recent years. To detect a moving object, a surveillance system utilizes background subtraction, which consists of obtaining a mathematical model of the static background and comparing it with every new frame from the video sequence. Background subtraction is commonly used to detect foreground objects in video surveillance. Traditional background subtraction methods are usually based on the assumption that the background is stationary. However, they are not applicable to dynamic background, whose background images that change over time. This paper proposes an adaptive Local-Patch Gaussian Mixture Model (LPGMM) as the dynamic background model for detecting moving objects from video with dynamic background. The shadow casted by the moving object is also detected. The shadow makes it difficult to detect the exact shape of object and to recognize the object. Therefore, the accurate detection of its exact shape by removing shadows have great influence on the performance of subsequent steps such as tracking, recognition, classification, and activity analysis. A cascading chromaticity difference estimator, brightness difference estimator, and spatial analysis are explored to discriminate the shadow and the moving object.

Keywords - Background Subtraction, Local-Patch Gaussian Mixture Model, Moving Object Detection, Dynamic Background, Shadow Removal, Chromaticity Difference, Brightness Difference, Spatial Analysis

I. INTRODUCATION

Moving object detection is one of the essential tasks in many computer vision applications, such as traffic monitoring and miltary. A typical and an efficient approach used to achieve such tasks is background subtraction. The idea behind background subtraction is to compare the current frame with a reference background model (reference model) which is learned and maintained in the background for a long time. Much effort has been devoted in developing efficient methods of moving object detection using background subtraction. Some of them estimated the probabilities of individual pixels belonging to background by using Gaussian Mixture Models (GMMs) [1] or labeled each pixel as foreground or background by Markov Random Fields (MRFs) [2].An improved GMM learning algorithm [3] was proposed to select an appropriate number of components for each pixel on-line, thus fully adapting to the scene. Elgammal et al. [4] proposed to utilize a general nonparametric kernel density estimation technique to build background model for detecting foreground objects. These methods work well under the assumption that the background scene is stationary. However, they are doomed to fail for the case of the dynamic scenes, which include repetitive motions like moving car, people walking etc. Several block-based methods were developed to overcome such problems, which usually divide an image into blocks and calculate block correlation [5] or block-specific features, such as the local binary pattern [6] histogram. However, these block-based approaches allow only coarse detection of the moving objects. Some recent methods proposed for dynamic background subtraction utilized not only the temporal information of a single pixel but also the spatial information of neighboring pixels. Li et al. [7] extracted foreground objects from a complex video under the Bayes decision framework. A Bayes decision rule was employed for classification of background and foreground from a general feature vector. Sheikh and Shah [8] also utilized a Bayes rule to build the background model based on an MRF framework to enforce the spatial constraint and obtained better results. Zhang et al. [9] proposed a spatial-temporal nonparametric background subtraction method to effectively handle dynamic background subtraction by modeling the spatial and temporal variations simultaneously.

In this paper, we propose a foreground detection algorithm from dynamic background videos. First, we propose a Local-Patch Gaussian Mixture Model (LPGMM) which is the extension of the GMM to represent local spatial distribution for each pixel. Fig.1 gives the foreground object extraction using LPGMM. Foreground object extraction have two phases. 1.Training phase 2.Testing phase. In training phase using image sequence build LPGMM background model (reference model). In testing phase new image sequence is compared with reference background model. If the observed pixel is matched to a Gaussian distribution and the mean difference falls within 2.5 times the corresponding standard deviation, then that pixel is classified as background and updated as background. Fig.1 shows the flow chart of the system. The shadow casted by the moving object is also detected. The shadow makes it difficult to detect the exact shape of object and to recognize the object. Therefore, the accurate detection of a moving object and the acquisition of its exact shape by removing shadows have great influence on the performance of subsequent steps such as tracking, recognition, classification, and activity analysis. Rest of this paper is organized as follows. In Section 2 LPGMM background modeling method. Section 3shows shadow removal method. Section 4 we show some experimental results. Section 5 shows conclusion.

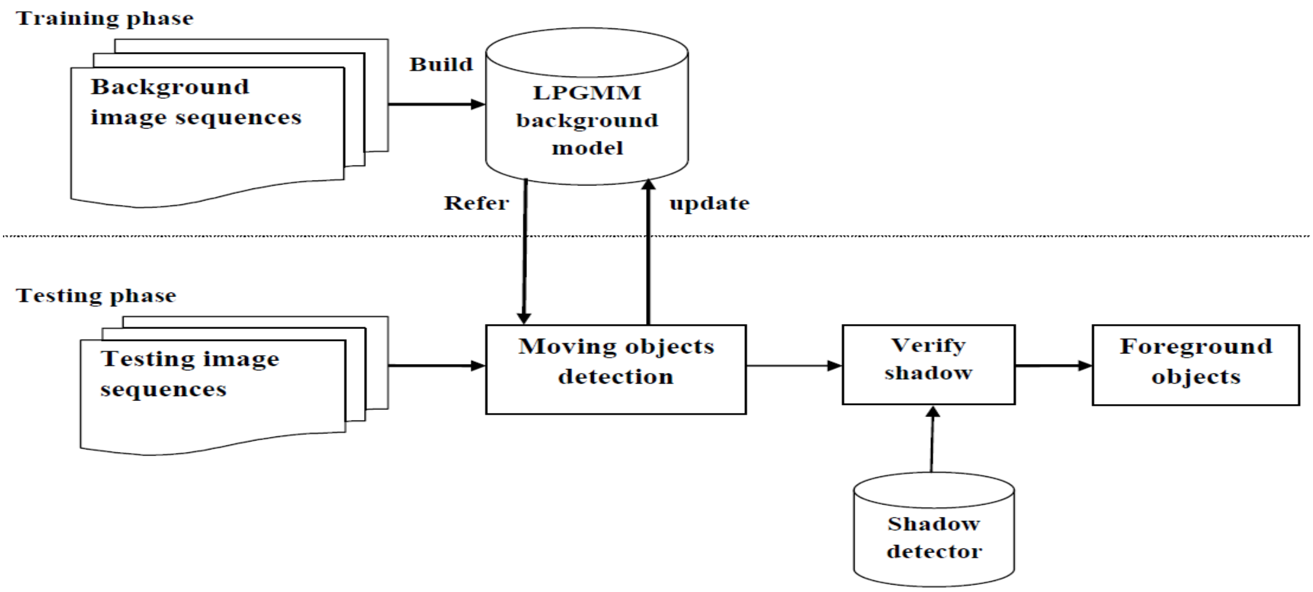


Fig.1The foreground object extraction using LPGMM

II. LPGMM BACKGROUND MODEL

The proposed method is a foreground object extraction technique for video surveillance system. Most of the traditional object detection approaches uses background is stationary. We consider the problem that the background is dynamic and usually change quickly overtime. Fig. 2 flow chart of the LPGMM model

Take 1^{st} frames(ex.5 frames) of the video as a training data for building the background model.

Extract recent history of a pixel (intensity and mean standard deviation) from a frame.

A. Intensity.

B. Mean and Standard deviation.

2.1 Intensity

Let the recent history of a pixel be $\{X_1,.....,X_N\}$, which is modeled by a mixture of K Gaussian distributions, and X be an intensity vector for R, G and B color channels. At time t, the probability density function at observing pixel value X_t is given by

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

Where η is the Gussian probability density function which is defined as

$$\eta(X_t,\mu,\Sigma)=\frac{1}{(2\pi)^{1/2}|\Sigma|^{1/2}} e^{-1/2(x_t-\mu_t)^T \Sigma^{-1}(x_t-\mu_t)} \tag{2}$$

The main difference between LPGMM and GMM [1] is that X_t,μ_t and σ_t for each pixel are vectors formed from observations from its local neighborhood instead of scalar values.

The covariance matrix Σ is defined as:

$$\Sigma=\begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{dxd}^2 \end{bmatrix} \tag{3}$$

Where d x d is the local patch size for each pixel, and σ_i denote the standard deviation for the i^{th} pixel. Let $f_t(x,y)$ be the intensity value of the observed pixel; $m_t(x,y)$ and $s_t(x,y)$ are the corresponding mean and standard deviation of the background model at time t in the original GMM [1]. For each observed pixel, set a window around the pixel as the center. Let the window size be d x d, we can extend the original X_t that is a single value in the original GMM model to a d^2-dimensional vector for each observed pixel as follows:

$$X_t(x,y)=[f_t(x-d',y-d')\ldots f_t(x,y)\ldots f_t(x+d',y+d')]^T \tag{4}$$

Where $d' = (d-1)/2$ is assumed to be an integer without loss of generality.

2.2 Mean and standard deviation

The mean μ_t and standard deviation σ_t are extended to d^2-dimensional vectors for each background pixel, i.e.

$$\mu_t(x,y)=[m_t(x-d',y-d')\ldots m_t(x,y)\ldots m_t(x+d',y+d')]^T \tag{5}$$

$$\sigma_t(x,y)=[s_t(x-d',y-d')\ldots s_t(x,y)\ldots s_t(x+d',y+d')]^T \tag{6}$$

These vectors represent the local spatial information in a local neighborhood for each pixel. When a new frame is processed, we first check if the color values for each pixel are matched to any of the K Gaussian distributions. Then mean difference between $X_t(x,y)$ and $\mu_t(x,y)$ is computed as follows:

$$D(X_t(x,y),\mu_t(x,y)) = \sum_{-d'\leq i,\, j \leq d'} | f_t(x+i\,,y+j) - m_t(x+i,y+j)| \quad (7)$$

If the observed pixel is matched to a Gaussian distribution and mean difference falls within 2.5 times the corresponding standard deviation, then that pixel is classified as background. The equation to classify the pixel is given by

$$D(X_t(x,y),\mu_t(x,y)) \leq 2.5\sum_{-d' \leq i\,,\; j \leq d',} \sigma_t(x+i\,,y+j) \quad (8)$$

In the updating process, we update $\sigma_t(x,y)$ and $\mu_t(x,y)$ for those pixels to classify background as follows:

$$\mu_t(x,y) = (1-\rho)\mu_{t-1}\big((x,y)+\rho_t(x,y)\big) \quad (9)$$

$$\sigma_t^2\,(x,y) = \big[\big(1-\rho\big)\,\sigma^2_{t-1}(x,y) + \rho\big[\big(X_t(x,y)-\mu_t(x,y)\big)\big]^2 \quad (10)$$

Where $\sigma_t^2\,(x\,,y)$ - Element-wise square function . ρ - is the learning rate between 0 and 1.

III. SHADOW REMOVAL

To remove shadow 1st foreground object is extracted by using background subtraction method, then shadow removal technique is used. A cascading chromaticity difference estimator, brightness difference estimator, and spatial analysis are explored to discriminate the shadow and the moving object. Estimated the background by using a metrically trimmed mean and also including a local spatial coherence for foreground detection.

Let us denote the position of a pixel x = (x, y), and the intensity of each pixel p(x) is expressed as I (x), which is defined as follows:

$$I\,(x) = [I^{\,R}\,(x), I^{\,G}\,(x), I^{\,B}\,(x)]^T \quad (11)$$

Where I R(x), I G(x) and I B(x)denote the intensity of the red, green, and blue component, respectively. For notational simplicity, we will use superscript K representing R,G or B. Suppose $[I_1(x), I_2\,(x),\ldots, I_T\,(x)]$ be T image frames used in the training stage, and M(x) be the image containing the temporal median of each pixel x. For $0 \leq \alpha < 1$, the α -metrically trimmed mean $\lambda_\alpha^k(\,x)$ of each RGB component for each pixel is obtained by computing the temporal average of $I_t^k(x)$, disregarding the largest $[\alpha T\,]$ deviations from the median (here, [.] denotes the greatest integer function). Formally, let us consider an ordering of the differences $|I_t^k\,(x)-M^k\,(x)|$ and define an integer function $1 \leq f\,(x,t) \leq T$ that returns the position of $|\,I_t^k\,(x)-M^k\,(x)|$ in such ordering. The α -metrically trimmed mean is $\lambda_\alpha^k(\,x)$ given by

$$\lambda_\alpha^k(\,x)=\frac{1}{T-[\alpha T]}\;\Sigma_{t\in s\alpha(x)}\;I_t^k(x) \quad (12)$$

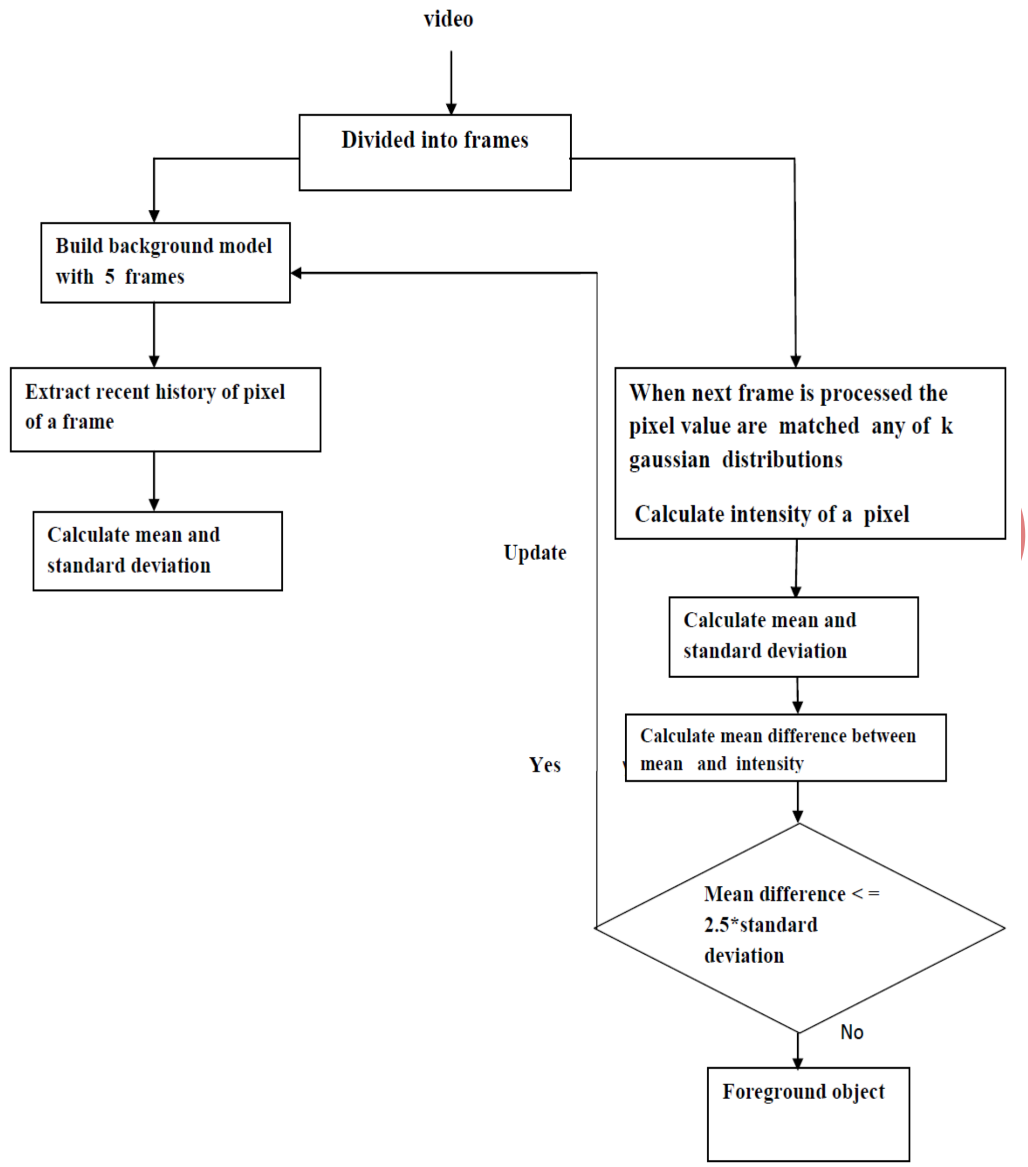


Fig. 2 flow chart of the LPGMM model

Where S (x) {t : f (x, t) ≤ T- [α T]} When α = 0 , the trimmed mean is exactly the average, and as α approaches one, the trimmed mean approaches the median value. In this work, we set α = 0.3 experimentally. From this point on, $\lambda^K(x)$ will denote the α -metrically trimmed mean value for pixel x using α = 0.3. A scale parameter (such as the standard deviation) is also important to evaluate the spread of the noise around the actual background value. A robust scale estimator is given by the mean absolute deviation (MAD), defined as

$$MAD^k(x)=\text{median}_{\,t=\{1....T\}}\{|I_t^k(x)-M^k(x)|\} \qquad (13)$$

If additive Gaussian noise is assumed, the relation $\sigma^K(x)=1.4826MAD^K(x)$ provides an estimate of the standard deviation of each RGB component for each pixel. Pixels belonging to the foreground are probably far from the estimated mean of the distribution. Foreground pixels usually do not appear isolated in an image, they tend to appear in blobs, so we analyze a small neighborhood around each pixel. A pixel x is assigned to the foreground if

$$\sum_{u\in\Omega(x)} w(u)\,|I_t^k(x) - \lambda^K(u)| > h \sum_{u\in\Omega(x)} w(u)\sigma^k(u) \quad (14)$$

Where Ω (x) is a small neighborhood centered at x, w(u) is a weighting mask for each pixel u = (u,v), and h controls the maximum allowed deviation from the mean w.r.t. the standard deviation. In this work, w is the weighted average mask, whose central point has weight 4, vertical and horizontal neighbors weight 2, and diagonal neighbors weight 1.We set a 3×3 neighborhood for Ω and h = 3. In practical applications, the detection of foreground pixels using [3] may produce some isolated pixels, or holes in the interior of valid objects. To overcome this problem, we apply sequentially an opening and a closing morphological operator.

Feature extracted for shadow removal

A. Chromaticity difference estimator.

B. Brightness difference estimator.

C. Spatial analysis for shadow verification.

The drawback of background subtraction techniques is the undesired detection of shadows as foreground objects. A cascading process is proposed to detect shadows from foreground pixels. We discard the local relation estimator and change the strategy of the spatial adjustment step. The shadow detection process is described as follows.

A.**Chromaticity difference estimator**

The Fig.3 shows the flow chart of Chromaticity difference estimator. To discriminate the shadow pixel and the object pixel, we define the chromaticity difference $CD^k(\mathbf{x})$ as follows:

$$CD^k(\mathbf{x}) = I_s^k(x) / \|I_{s(x)}\| - I_b^k(x) / \|I_b(x)\| \quad (15)$$

Where the subscripts s and b represent the shadow and the background respectively. ||Is (x) and $I_b(x)$|| are the norm of Is(x) and $I_b(x)$ respectively. For every pixel x in the set of moving pixels, we calculate $CD^K(x)$. According to the assumptions in [7], $CD^K(x)$ of the shadow pixels has Gaussian distribution, and its mean is close to zero, because shadow pixel is darker than background. So, a pixel which is brighter than the background cannot be a shadow pixel, and it conclude that it is a moving object pixel. $CD^K(x)$ of the moving object pixel has an unknown distribution that depends on the object. Thus, we can determine that a pixel is a moving object pixel if $CD^K(x)$ is far from zero. To reduce computation, we only use pixels which satisfy the following condition.

$$\mathbf{-0.2 \le CD^K(x) \le 0.2}$$

While estimating the mean m_{CD}^k and the standard deviation σ_{CD}^k .Then, m_{CD}^k and σ_{CD}^k can be estimated as follows:

$$m_{CD}^k = 1/N_M \sum_{p(x)\in M} CD^K(x) \quad (16)$$

$$(\sigma_{CD}^k)^2 = 1/N_M \sum_{p(x)\in M} (CD^K(x) - m_{CD}^k)^2 \quad (17)$$

Where N_M is the set of pixels which satisfy $-0.2 \le CD^K(x) \le 0.2$ in the set of moving pixels. Using the estimated m_{CD}^k σ_{CD}^k and , we calculate the threshold α_h^k and α_l^k with reliability of 95%

$$\alpha_h^k = m_{CD}^k + 1.96 * \sigma_{CD}^k \tag{18}$$

$$\alpha_l^k = m_{CD}^k + 1.96 * \sigma_{CD}^k \tag{19}$$

Where: h - is high threshold.

l - is low threshold.

After calculating α_h^k and α_l^k the class of the pixel is determined by:

$$\begin{cases} p(x) \in s1, \text{if} \quad \alpha_l^k < CDk(x) < \alpha_h^k \\ p(x) \in 0, \quad \text{otherwise} \end{cases} \tag{20}$$

Where S_1 is the 1st candidate set of shadow pixels, and 0 is the candidate set of object pixel.

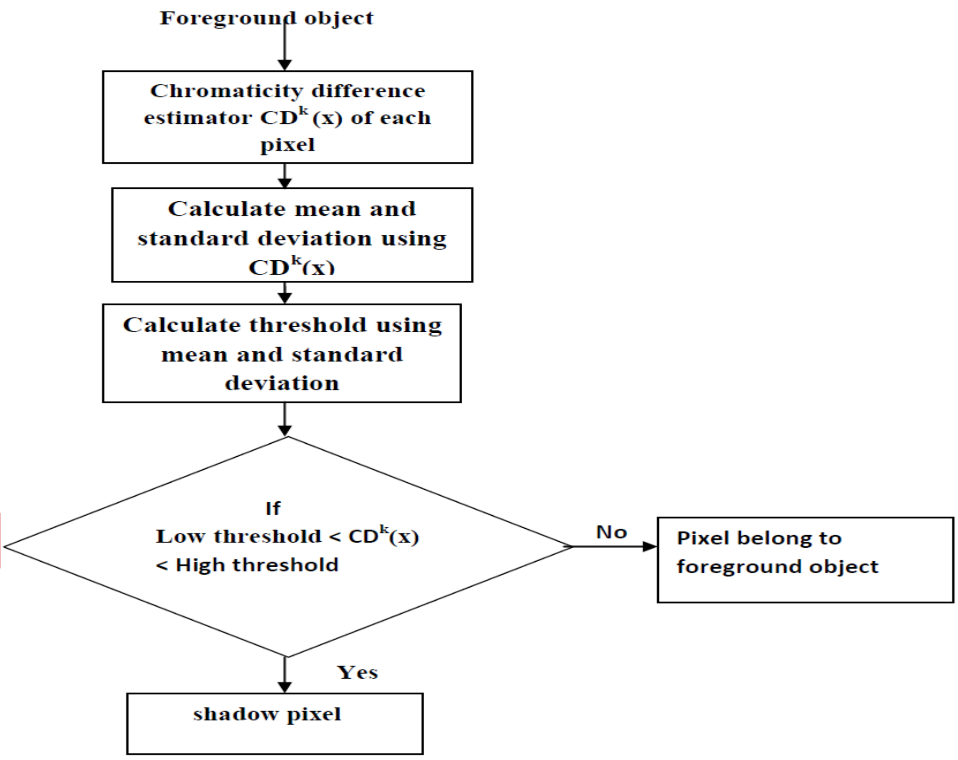


Fig. 3 Flow chart of Chromaticity difference estimator

B. Brightness difference estimator

After extracting chromaticity difference estimator, there still exist many moving object pixels in the 1st candidate set of shadow pixels. This shadow pixels can be removed by Brightness difference estimator. The Fig.4 shows the flow chart of Brightness difference estimator.

The brightness difference estimator separates moving object pixels from the 1st candidate set of shadow pixels using brightness difference BD^K (x) of all pixels in the 1st candidate set of shadow pixels. BD^K (x) can be defined as

$$BD^k(\mathbf{x}) = I_s^k(x) \,/\, I_b^k(x) \qquad (21)$$

The distribution of BD^K (x) of the shadow pixels is Gaussian such that the mean is m_{BD}^k and the standard deviation is σ_{BD}^k . We estimate m_{BD}^k and σ_{BD}^k as follows:

$$m_{BD}^k = 1/\,N_s \sum_{p(x)\in M} BD^K(x) \qquad (22)$$

Where N_S is the number of pixels in S1.Using the estimated m_{BD}^k and σ_{BD}^k, we calculate the threshold β_h^k and β_h^k as follows: with reliability of 95%.

$$\beta_h^k = m_{BD}^k + 1.96 * \sigma_{BD}^k \qquad (23)$$

$$\beta_h^k = m_{BD}^k - 1.96 * \sigma_{BD}^k \qquad (24)$$

Where: h - is high threshold, l - is low threshold.

After calculating β_h^k and β_h^k , the class of the pixel is determined by

$$\begin{cases} p(x) \in s1, \text{if} \quad \alpha_l^k < CDk(x) < \alpha_h^k \\ p(x) \in 0, \quad \text{otherwise} \end{cases} \qquad (25)$$

Where S2 is the 2nd candidate set of shadow pixels, and 0 is the candidate set of object pixels.

C. Spatial Analysis for Shadow Verification

By the previous two estimators, the set of moving pixels are divided into the 2nd candidate set of shadow pixels and the candidate set of object pixels. However, two types of shadow detection errors may commonly occur, namely shadow detection failure and object detection failure. To improve the accuracy of shadow detection, a post processing spatial analysis is added for shadow confirmation. Then analysis is done to confirm the true shadows as well as the true objects according to their geometric properties.

The Fig.5 shows the flow chart of spatial analysis for shadow verification. In order to break the weak connection between shadow regions, we apply an opening morphological operator to the shadow mask. Then a flood-fill operation is used to fill the holes in shadow regions. The foreground consists of shadow regions and object regions. If the shadow candidate is a true shadow, less than a half of the boundary should be adjacent to the boundaries of object regions. Thus we can use the boundary information of a shadow candidate region to confirm whether the shadow is a true shadow or not. The outer pixels E_S of each candidate shadow mask M_S can be calculated as follows:

$$E_S = M_S \oplus S_E - M_S \qquad (26)$$

Where $\oplus$ is the dilation operator and E_S is the structure element of 3×3 square. If the proportion of object pixels in E_S is less than 50%, M_S is determined to be a shadow region otherwise, M_S is an object region.

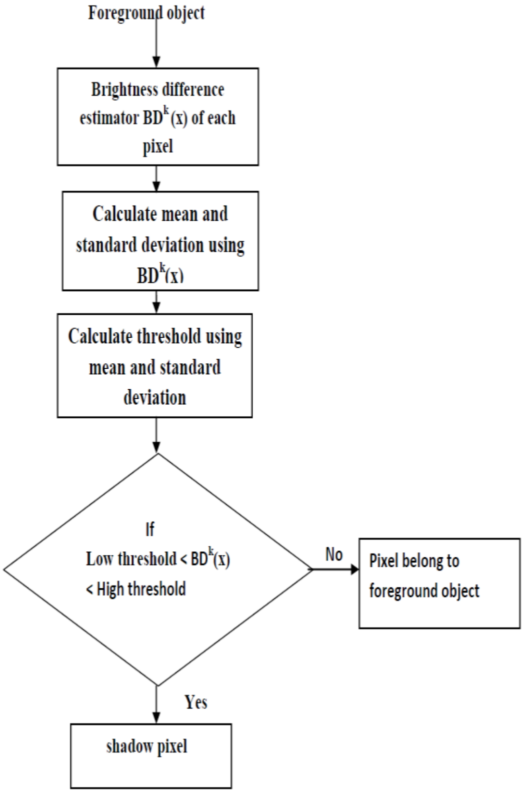


Fig. 4 Flow chart of Brightness difference estimator

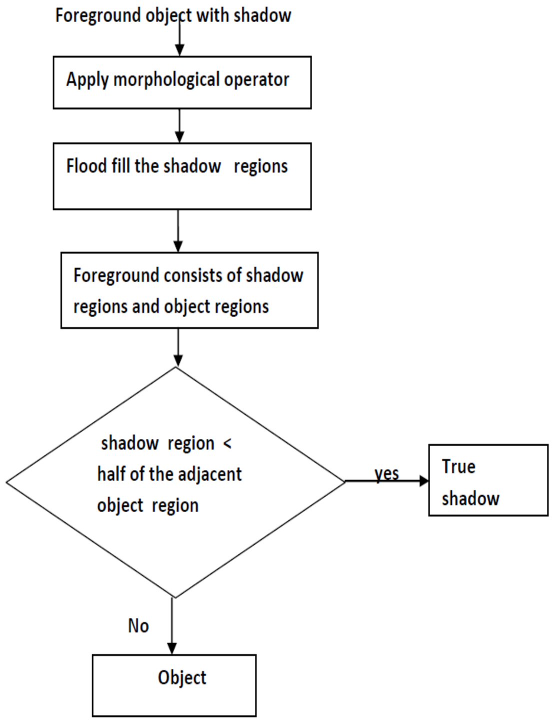


Fig.5 Flow chart of spatial analysis for shadow verification

IV. EXPERIMENTAL RESULTS

4.1 LPGMM model

In our experiment, the algorithm is tested on two video sequences (traffic control, railway station) of size 720x576 and 320x240. The experiments gives good results for both the video sequence, we obtain good results with the some parameter settings. The video sequence “Traffic control” is depicted. There is a car moving in the scene, and it causes intensity variation of many pixels over time. The reason GMM could have more influences caused by noises is because GMM does not use the spatial information but LPGMM exploits it explicitly to construct the background model. The second video sequence “railway station” contains people moving. We also evaluated the Qualitative result for two test videos described in Table 1. The experimental results on video containing vehicle and human is shown in Fig 6.

Test video	Method	TP	FN	TPR
Traffic control	LPGMM	50	10	83.3
Railway station	LPGMM	48	12	81.6

Table 1: Qualitative results of LPGMM

We have tested with 1second video for traffic control and railway station we have calculated the sensitivity (TPR- true positive rate) for the above said videos using formula

$$\mathbf{TPR = TP / (TP + FN)} \tag{27}$$

Where: TP – True Positive, FN- False Negative

(a,1) (a,2)

(b,1) (b,2)

(c,1) (c,2)

Fig 6. (a) input frame (b) moving object detection frame (c) foreground object extraction frame

4.2 Shadow removal

The proposed algorithm is tested on two video sequences (Highway I, Campus) which can be downloaded from http://cvrr.ucsd.edu/aton/shadow. First 100 frames are used for background training for Highway I sequence; while for Campus sequence, From the Fig 7 shows the snapshot of shadow removal method.

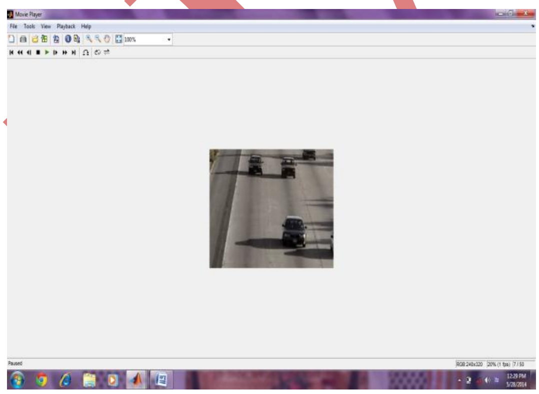
To evaluate the performance of the proposed shadow removal method quantitatively, we calculate the shadow detection rate and shadow discrimination rate. The shadow detection rate η and shadow discrimination rate ξ are defined as follows:

$$\eta = TP_S / (TP_S + FNs) \tag{28}$$
$$\xi = TP_f / (TP_f + FN_f) \tag{29}$$

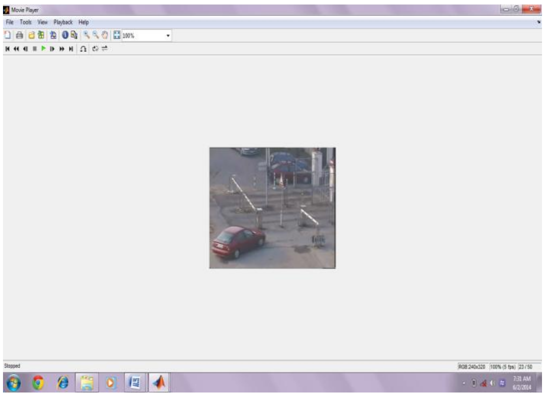
Where TPs and TP_f are the number of pixels which are determined correctly as shadow pixels and object pixels, in order. FNs is the number of errors in which a shadow pixel is defined as an object pixel, are FN_f is the number of false detection which identified an object pixel as a shadow pixel. Table 2 shows the results of the proposed method. From the table,

Methods	**Highway I**		**Campus**	
	η (%)	**ξ (%)**	**η (%)**	**ξ (%)**
Shadow removal	**81.1**	**88.89**	**87.7**	**92.67**

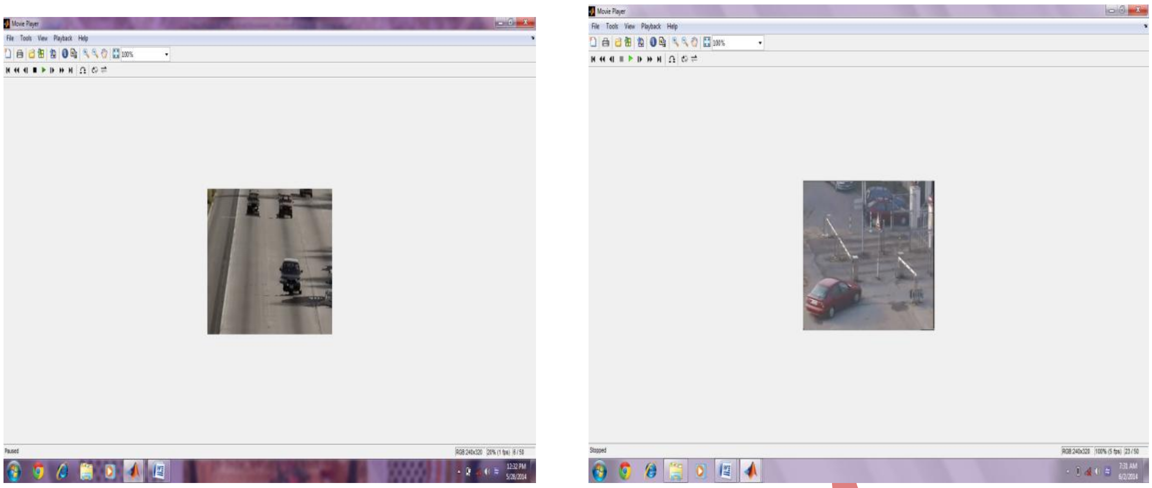
Table 2: Qualitative results of shadow removal



(a,1)



(a,2)



(b,1) **(b,2)**

Fig7 (a) input frame (b) frame after removing shadow

V. CONCLUSION

In this paper, we proposed a **LPGMM** model is used detect moving object from surveillance videos with dynamic background, which consider the local spatial information for each pixel. This approach is used for object detection with background subtraction and shadow removal in the RGB color space. For background subtraction, the metrically trimmed mean is used as a robust estimate of the background model, and the MAD (mean absolute deviation) is adopted as a scale estimate. The shadow and the moving object are discriminated by cascading two estimators which use the properties of chromaticity and brightness. The spatial information is utilized to verify the true shadow regions. Our experimental results show the good performance of the proposed method. For future work, texture and edge information will be adopted as features to further improve the shadow detection accuracy. Classification method can be applied after the extraction of foreground object for further analysis of moving object whether it is human or vehicle.

References

[1] C. Staufer, W.E.L. Grimson, "Adaptive background mixture models for real-time tracking," CVPR, Vol. 2, pp. 246-252,1999.

[2] N. Paragios and V. Ramesh, "A MRF-based approach for real time subway monitoring," CVPR, vol. 1, pp. 1034-1040, 2001

[3] Z. Zivkovic and F. van der Heijden, "Recursive unsupervised learning of finite mixture models," IEEE Trans. PAMI, Vol.26, No.5, pp. 651–656, 2004.

[4] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," ECCV, 2000.

[5] T. Matsuyama, T. Ohya, H. Habe, "Background subtraction for non-stationary scenes," ACCV, pp. 662–667, 2000.

[6] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," IEEE Trans. PAMI, Vol. 28, pp. 657-662, 2006.

[7] L. Li, W. Huang, I.-Y.H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," ACM Intern. Conf. Multimedia, 2003.

[8] Y. Sheikh and M. Shah, “Bayesian modeling of dynamic scenes for object detection,” IEEE Trans. PAMI, Vol. 27, pp. 1778-1792, 2005.

[9] S. Zhang, H. Yao, and S. Liu, “Spatial-temporal nonparametric background subtraction in dynamic scenes,” *ICME*, 2009

[10]L. Cheng and M. Gong, “Real-time background subtraction from dynamic scenes,” *ICCV*, 2009.

[11]S.-S. Huang, L.-C. Fu, and P.-Y. Hsiao Region-level motionbased background modeling and subtraction using mrfs,” IEEE Trans. Image Processing, vol. 16, no. 5, pp. 1446 –1456, 2007.

[11] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," Computer Vision and Image Understanding, vol. 95, pp. 238-259, August 2004.

[12] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting Moving Shadows: Algorithms and Evaluation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 918-923, July 2003.