**Available Online at www.ijcsmc.com**

**International Journal of Computer Science and Mobile Computing**

**A Monthly Journal of Computer Science and Information Technology**

**ISSN 2320–088X**

***IJCSMC, Vol. 4, Issue. 6, June 2015, pg.858 – 861***

**SURVEY ARTICLE**

# Survey on Moving Object Detection Techniques

**Hitesh A. Patel[1], Paresh M. Tank[2]**
[1]Computer Engineering Department, B & B Institute of Technology, India
[2]Computer Engineering Department, B & B Institute of Technology, India
[1] hitehspatelbbit@gmail.com; [2] pareshtank@gmail.com

***Abstract— Surveillance system can detect, classify and track the moving objects. Moving object detection is first and most important part in the field of security. The performance of the system mainly depends on the object detection quality. To detects the moving objects in the video objects. Here, the survey of object detection methods viz. Temporal Frame Differencing[6], Running Gaussian Average[11] ,Approximate Median[7], Mixture of Gaussian[10] and Kernel Density Estimation[3] is presented.***

***Keywords — Visual surveillance, Temporal frame differencing, Approximate median ,Kernel density estimation, Running Gaussian average, Mixture of Gaussian.***

## I. INTRODUCTION

This Visual surveillance systems have been in long use to monitor security sensitive areas. The advances in computing power, availability of large-capacity storage devices and high speed network infrastructure paved the way for cheaper, multi sensor visual surveillance systems. Visual surveillance systems offer three main challenges - fast, reliable and robust algorithms for moving object detection, classification and tracking. For this task, we need to first detect the objects of interest by segmenting them from the background, and then track them across different frames while maintaining the correct identities[2].

The object detection algorithms are useful for many application domains such as criminology, sociology, statistic, traffic accident detection and military applications[9].

In visual surveillance system first requirement is to detect moving objects from videos. For moving object detection many methods are available in the literature viz. Temporal Frame Differencing(TFD), Running Gaussian Average, Approximate Median(AM), Mixture of Gaussian[10] , and Kernel Density Estimation(KDE) whose descriptions are given below.

## II. TEMPORAL FRAME DIFFERENCING

Temporal frame differencing method presented by Lipton [6], detects moving object by taking difference of consecutive frames (two or three) in a video sequence. Temporal frame differencing is having very low computation cost. It is very adaptive to dynamic environments, but generally does a poor job of extracting all relevant feature pixels. The main problem with temporal frame differencing is that pixels interior to an object with uniform intensity aren't included in the set of "moving" pixels [1]. Another problem is, it cannot detect object that remain motionless for long period.

Lipton et al. [6] has proposed temporal frame differencing method. In this method frame at *t-1* time has been considered as background frame. The difference of current frame and background frame has been calculated. If absolute difference calculated is greater than the threshold value than the pixel has been considered as foreground pixel otherwise as background pixel. Equation for two-frame differencing is given bellow.

*© 2015, IJCSMC All Rights Reserved* 

$$|I_t(x,y) - I_{t-1}(x,y)| > \tau \qquad (1)$$

Pixels that satisfying equation 1 has been considered as foreground pixels.

## III. RUNNING GAUSSIAN AVERAGE

Wren et al**.** in [11] have proposed to model the background independently at each (i,j) pixel location. The model is based on ideally fitting a Gaussian probability density function (pdf) do on the last *n* pixel's values. In order to avoid fitting the pdf from scratch at each new frame time, t, a running (or on-line cumulative) average is computed instead as: After classifying all foreground pixels, morphological closing and opening operation are used to eliminate the small-sized regions. This technique is sensitive to dynamic changes i.e. when stationary objects uncover the background or sudden illumination changes occur[8].

$$\mu_t = \alpha I_t + (1 - \alpha)\mu_{t-1} \qquad (2)$$

Where ***I*,** is the pixel's current value and μ, the previous average, α is an empirical weight often chosen as a trade- off between stability and quick update. Although not stated explicitly in [11], the other parameter of the Gaussian pdf, the standard deviation $\sigma_t$**,** can be computed similarly. In addition to speed, the advantage of the running average is given by the low memory requirement: for each pixel, this consists of the two parameters ($\mu_t$ ,$\sigma_t$) instead of the buffer with the last *n* pixel values. At each T frame time, the $I_t$ pixel's value can then be classified as a foreground pixel following condition satisfied[8].

$$|I_t - \mu_t| > K\,\sigma_t \qquad (3)$$

## IV. APPROXIMATE MEDIAN

Approximate median was presented by McFarlane and Schofield in [7]. The first stage of image segmentation is image differencing. Each successive frame is subtracted from a time-averaged reference image and the difference image is thresholded to determine the foreground pixels.

Approximate median method does a much better job at separating the entire object from the background. This is because more slowly adapting background incorporates a longer history of the visual scene, achieving about the same result as if we had buffered and processed N frames[5]. Median filtering has been shown to be very robust and to have performance comparable to higher complexity methods and it costs not much more in computation and storage than Temporal frame differencing [12].

The approximate median method works as such: if a pixel in the current frame has a value larger than the corresponding background pixel, the background pixel is incremented by 1. Likewise, if the current pixel is less than the background pixel, the background is decremented by one. In this way, the background eventually converges to an estimate where half the input pixels are greater than the background, and half are less than the background. The background is calculated as follows[12].

```
IF (It(x,y) > Bt(x,y)
      Bt(x,y) = Bt(x,y) + 1;
ELSEIF (It(x,y) < Bt(x,y)
      Bt(x,y) = Bt(x,y) - 1;
END;
```

Finally the absolute difference of pixel's current frame value and background frame value has been calculated. If calculated difference is found to be greater than the threshold value than the pixel has been considered as foreground pixel otherwise it has been considered as background pixel. This procedure has been repeated for every frame.

$$|I_t(x,y) - B_t(x,y)| > \tau \qquad (4)$$

## V. MIXTURE OF GAUSSIAN

Stauffer and Grimson [10] presented a novel adaptive online background mixture model that can robustly deal with lighting changes, repetitive motions, clutter, introducing or removing objects from the scene and slowly moving objects.

In this technique, the values of each pixel are calculated as a mixture of Gaussians usually three to five Gaussians are used. The values of an individual pixel (e. g. scalars for gray values or vectors for color images) over time is considered as a "pixel process" and the recent history of each pixel in frame, $\{X_1 \ldots\ldots X_t\}$, is modelled by a mixture of *K* Gaussian distributions. The probability of observing current pixel value is given by following equation [10].

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \textstyle\sum_{i,t}) \qquad (5)$$

Where $\omega_{i,t}$ is an estimate of the weight (what portion of the data is accounted for this Gaussian) of the i[th] Gaussian ($G_{i,t}$) in the mixture at time *t*, $\mu_{i,t}$ is the mean value of $G_{i,t}$ and $\sum_{i,t}$ is the covariance matrix of $G_{i,t}$ and $\eta$

is a Gaussian probability density function and it is given by equation 5. K is determined by the available memory and computational power. Currently, from 3 to 5 are used. Also, for computational reasons, the covariance matrix is assumed to be of the form[10]:

$$\sum_{k,t} = \sigma_k^2 I \qquad (6)$$

For equation 5 to become a model of the background alone, a criterion is required to provide discrimination between the foreground and background distributions. In[10]**,** it is given like this: first all the distributions are ranked based on the ratio between their peak amplitude, $\omega i$ and standard deviation $\sigma i$**.** The assumption is that the higher and more compact the distribution, the more is likely to belong to the background. Then, the first B distributions in ranking order satisfying with T an assigned threshold, are accepted as background[8].

$$\sum_{i=1}^{B} \omega_i > T \qquad (7)$$

## VI. KERNEL DENSITY ESTIMATION

Kernel density estimation is a non parametric background subtraction technique for density estimation in which a known kernel density function is averaged across the observed data points to create a smooth approximation. A variety of kernel functions with different properties have been used in the literature. Typically the Gaussian kernel is used for its continuity, differentiability, and locality properties. Elgammal in [3] has applied kernel density estimation to handle situations where the scene background is not completely static, but contains small motions such as moving tree branches and bushes or illumination changes. On the other hand, it is able to suppress false detections that arise due to small camera displacements.

According to Elmammal [3] the model uses pixel intensity (color) as the basic feature for modeling the background. The model keeps a sample of intensity values for each pixel in the image and uses this sample to estimate the density function of the pixel intensity distribution. Therefore, the model is able to estimate the probability of any newly observed intensity value. The model can handle situations where the background of the scene is cluttered and not completely static but contains small motions that are due to moving tree branches and bushes. The model is updated continuously and therefore adapts to changes in the scene background[3].

A particular nonparametric technique that estimates the underlying density, avoids having to store the complete data, and is quite general is the kernel density estimation technique.
In this technique, the underlying pdf is estimated as[3].

$$f(x) = \sum_i \alpha_i K(x - x_i) \qquad (8)$$

Let $x_1, x_2, \ldots\ldots, x_n$ be sample of intensity values for a pixel. Given this sample, an estimate of the pixel intensity pdf at any intensity value using kernel density estimation can be obtained. Given the observed intensity $x_t$ at time *t* ,probability of this observation is estimated as[3].

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^{N} K_\sigma (X_t - X_i) \qquad (9)$$

Where $K_\sigma$ is a kernel function with bandwidth $\boldsymbol{\sigma}$. This estimate can be generalized to use color features by using kernel products as[3].

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^{N} \prod_{j=1}^{d} K_{\sigma j} \left(X_{tj} - X_{ij}\right) \qquad (10)$$

Decision on *K* depends on the available memory and computational power. where $x_t$ is *d*- dimensional color feature and $K_{\sigma j}$ is a kernel function with bandwidth $\sigma_j$ in the *j*th color space dimension. If chosen kernel function *K* to be Gaussian, then the density can be estimated as[3].

$$P_r(x_t) = \frac{1}{N} \sum_{i=1}^{N} \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{\left(x_{tj} - x_{ij}\right)^2}{2\sigma_j^2}} \qquad (11)$$

Using this probability estimate, the pixel is considered to be a foreground pixel if $Pr\ (x_t) < th$ where the threshold *th* is a global threshold over all the images that can be adjusted to achieve a desired percentage of false positives. Here, the estimate is based on the most recent N samples used in the computation. Therefore, adaptation of the model can be achieved simply by adding new samples and ignoring older samples[3].

One major issue that needs to be addressed when using kernel density estimation technique is the choice of suitable kernel bandwidth (scale). Theoretically, as the number of samples reaches infinity, the choice of the bandwidth is insignificant and the estimate will approach the actual density. Practically, since only a finite number of samples are used and the computation must be performed in real time, the choice of suitable bandwidth is essential. Too small a bandwidth will lead to a ragged density estimate, while too wide a bandwidth will lead to an over-smoothed density estimate. Since the expected variations in pixel intensity over

© 2015, IJCSMC All Rights Reserved 

time are different from one location to another in the image, a different kernel bandwidth is used for each pixel. Also, a different kernel bandwidth is used for each color channel[3].

To estimate the kernel bandwidth $\sigma j^2$ for the *j*th color channel for a given pixel, median absolute deviation over the sample for consecutive intensity values of the pixel can be computed. That is, the median *m* of $|x_i - x_{i+1}|$ for each consecutive pair $( x_i , x_{i+1} )$ in the sample is calculated independently for each color channel. The motivation behind the use of median of absolute deviation is that pixel intensities over time are expected to have jumps because different objects (e.g., sky, branch, leaf, and mixtures when an edge passes through the pixel) are projected onto the same pixel at different times. Since deviations between two consecutive intensity values have been measured, the pair $( x_i , x_{i+1} )$ usually comes from the same local-in-time distribution, and only a few pairs are expected to come from cross distributions (intensity jumps). The median is a robust estimate and should not be affected by few jumps[3].

If local-in-time distribution is Gaussian $N( \mu, \sigma^2)$, then the distribution for the deviation $( x_i - x_{i+1} )$ is also Gaussian $N ( 0, 2\sigma^2 )$ . Since this distribution is symmetric, the median of the absolute deviations is equivalent to the quarter percentile of the deviation distribution[3]. That is given as follows,

$$Pr(N(0, 2\sigma^2 ) > m) = 0.25 \qquad (12)$$

and therefore the standard deviation of the first distribution can be estimated as

$$\sigma = \frac{m}{0.68\sqrt{2}} \qquad (13)$$

Since the deviations are integer gray scale (color) values, linear interpolation is used to obtain more accurate median values[3].

## CONCLUSION

This paper provided an overview of the recent research on object detection methods. The state of art in contemporary work has been thoroughly studied. The methods reviewed can handle object detection in indoor and outdoor environments, under changing illumination conditions.

## REFERENCES


[1] Collins R.T. et.al. ," A system for video surveillance and monitoring: VSAM final report". Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.Dedeoglu, Yigithan, " Moving object detection, tracking and classification for smart video surveillance." Diss. bilkent university, 2004.

[2] Elgammal A, Harwood D, Davis L., Non-parametric model for background subtraction, in: *European* Conference on Computer Vision, Dublin, Ireland, June 2000.

[3] Elgammal A., Duraiswami ramani, Harwood D, Davis L., "Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance", Proceedings of the IEEE, vol. 90, no. 7, july 2002.

[4] Kumari Deepa, Shamik Tiwari, Deepika Gupta, Raina," Analysis on Adaptive Moving Objects via Robot Vision Implementations by Detection Techniques", International Journal of Scientific & Engineering Research, Volume 3, Issue 4, April-2012

[5] A. J. Lipton, H. Fujiyoshi, and R.S. Patil. Moving target classification and tracking from real-time video. In Proc. of Workshop Applications of Computer Vision, pages 129–136, 1998.

[6] McFarlane N, Schofield C., “Segmentation and tracking of piglets in images”, British Machine Vision and Applications, BMVA 1995, pages 187-193, 1995.

[7] Piccardi,M. “Background subtraction techniques: a review.” Systems, Man and Cybernetics, 2004 IEEE International Conference, Vol 4, pp.3099, 2004.

[8] Prasad Kauleshwar, Sharma Richa and Wadhwani Deepika," A Review on Object Detection in Video Processing", International Journal of u- and e- Service, Science and Technology Vol. 5, No. 4, December, 2012.

[9] Stauffer C., Grimson W.E.L," Adaptive background mixture models for real-time tracking" in: Computer Vision and Pattern Recognition, Santa Barbara, CA, June 1998.

[10] Wren C., A. Azarhayejani, T. Darrell, and A.P. Pentland, “Pfinder: real-time tracking of the human body,” IEEE Trans. on Patfern Anal. and Machine Infell., vol. 19, no. 7, pp. 780-785, 1997.

[11] *www.eetimes.com/General/PrintView/4017685- United States.*

[12] http://groups.inf.ed.ac.uk/vision/caviar/caviardata1/

[13] http://www.hitech-projects.com/euprojects/cantata/ datasets_cantata/ dataset.htm.


*© 2015, IJCSMC All Rights Reserved* 