IJETCR
Open Access

ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 2; Issue 6; Page No. 50-54

Background Subtraction Techniques

Mr. Manojkumar. M. Gawali[1], Prof. Dipti R. Pawar[2]

[1] Student, Pune University, Sinhgad College of Engineering, Vadgaon (BK), Pune, 411041.Maharashtra, India

gawalimanojkumar@gmail.com

[2] Professor, Pune University, Sinhgad College of Engineering, Vadgaon (BK), Pune, 411041.Maharashtra, India

dvshitole.scoe@sinhgad.edu

ABSTRACT

Background subtraction is one of the important and useful step for moving object detection, especially in the domain of video surveillance. There are various methods have been developed over the recent years. This paper gives survey of the recent approaches which concern to statistical background modeling techniques. These background subtraction techniques have benefits and limitations in terms of noise, illumination change. To overcome this problem, this paper provides a review of Background subtraction methods and comparison mainly based on three factors speed, memory requirements and accuracy.

Key Words: Background modeling, Kernel Density Estimation, Mixture of Gaussians, Single Gaussian

INTRODUCTION:

Visual surveillance is very important research topic in the last few years due to its growing importance in security, law enforcement and military applications [1]. More number of surveillance cameras was installed in security sensitive areas such as banks, train stations, highways, and borders. Background subtraction is a widely used approach for detecting moving objects in videos from camera [4, 5]. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, is nothing but background model. Background subtraction is one of the most popular methods to detect objects in video frames [10]. One way to calculate background from background image which doesn't include any moving object but in some of the cases background is not available. In many cases, it can always be changed under critical situations like illumination changes, noise [3].The term BS includes many of methods that aim to distinguish between foreground and background areas in video sequences [14]. Over the past years, various BS methods have been developed with each of them having its own characteristic, strength and weakness. Evaluation allows identifying those characteristics and aid to focus on the remaining problems.

1. DIFFERENT TECHNIQUES FOR BACKGROUND SUBTRACTION:

There are various methods exist for performing background subtraction, and those methods are:

A.Single Gaussian
B.Temporal median filter
C.Mixture of Gaussians
D.Adaptive Gaussian mixture model
E.Kernel density estimation (KDE)
F.Sequential KD approximation
G.Co occurrence of image variations
H.Eigen backgrounds

A. Single Gaussian:

Wren et al. [l] proposed to model the background independently at each pixel location (i, j). The model is based on Gaussian probability density function on the last n pixel's values. In order to avoid fitting the pdf from scratch at each new frame time t+1, the mean and the variance are updated as follows:

$\propto_{t+1} = (1-\alpha)\propto_t + \square\alpha . I_t$ (1)

${}^{\sigma 2}t+1 = (1-\alpha)^{\sigma 2}t \, \square\alpha . (I_t - \propto_{t+1})\,(I_t - \propto_{t+1})^T$ (2)

Where I_t is the pixel current value, $\propto_t$ is the previous average, σ_t is the previous variance and α is the learning rate. The foreground detection is made as follows: if $| \propto_{t+1} - I_t | < T$, the pixel is classified as background otherwise the pixel is classified as foreground. The single Gaussian (SG) is suited for indoor scenes where there are moderate illumination changes.

B. Temporal Median filter:

Temporal median filter method was proposed by Lo and Velastin [2] in the year 2001. This model uses the median value of the last n frames as the background model. The main disadvantage of a temporal median approach is that it requires buffer with the recent pixel values. Moreover, the median filter does not accommodate for a rigorous statistical description.

C. Mixture of Gaussian:

Mixture of Gaussians method was proposed by Stauffer and Grimson [3] in the year 2000. Stauffer and Grimson found very useful case for multi-valued background model for multiple background objects. In the context of a traffic surveillance system, Friedman and Russel [4] proposed to model each background pixel using a mixture of three Gaussians corresponding to road, vehicle and shadows. This model is initialized using an EM algorithm. Then, the Gaussians are manually initialized: the darkest component is labeled as shadow, in the remaining two components; the one with the largest variance is labeled as vehicle and the other one as road. This remains as it is for all the process giving lack of adaptation to changes over time. For the foreground detection, each pixel is compared with each Gaussian and it is classified according to it corresponding Gaussian. The maintenance is made using an incremental EM algorithm for real time consideration. Stauffer and Grimson [3] generalized this idea by modeling the recent history of the color features of each pixel {X1,..., X t } by a mixture of K Gaussians. We remind below the algorithm.

Principle: In this, each pixel is characterized by its intensity in the RGB color space. Then, the probability density function of observing the current pixel value is considered given by the following formula in the multidimensional case:

$$P(X_t) = \sum_{i=1}^{k} w_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (3)$$

Where the parameters are K is the number of distributions, $w_{i,t}$ is a weight associated to the ith Gaussian at time t with mean $\mu_{i,t}$ and standard deviation $\Sigma_{i,t}$. η is a Gaussian probability density function. Once the parameters initialization is made, a first foreground detection can be made and then the parameters are updated. Firstly, Stauffer and Grimson [3] used as criterion the ratio $r_j = \omega_j / \sigma_j$ and ordered the K Gaussians following this ratio. This ordering supposes that a background pixel corresponds to a high weight with a weak variance due to the fact that the background is more present than moving objects and that its value is practically constant. The first B Gaussian distributions which exceed certain threshold *T* are retained for a background distribution

$$B = \sum_{i=1}^{k} w_{i,t} > T \quad (4)$$

D. Adaptive Mixture of Gaussians:

This is up-gradation of Mixture of Gaussian model proposed by Stauffer and Grimson. It handles:-

A.Lighting changes

B.Repetitive motion from cluster

C.Long term scene changes

Each pixel is modeled separately [4, 5, 6] by a mixture of K Gaussians

$$P(I_t) = \sum_{i=1}^{k} w_{i,t} \cdot \eta(I_t, \mu_{i,t}, \Sigma_{i,t}) \quad (5)$$

Where K = 4 in [9] and K = 3…5 in [10]. In [6, 10], it is assumed that $\Sigma_{i,t} = \sigma^2_{i,t} . I$

The background is updated, before the foreground is detected, as follows:

1. If I_t matches component i, i.e., I_t is within λ standard deviations of $\mu_{i,t}$, then the ith component is updated as follows:

$$w_{i,t} = w_{i,t-1} \quad (6)$$

$$\mu_{i,t} = (1-\rho)\mu_{i,t-1} + \rho I_t \quad (7)$$

$$\sigma^2_{i,t} = (1-\rho)\sigma^2_{i,t-1} + \rho(I_t - \mu_{i,t})^T(I_t - \mu_{i,t}) \quad (8)$$

2. Components which I_t don't match are updated by

$$w_{i,t} = (1-\alpha)w_{i,t-1} \quad (9)$$

$$\mu_{i,t} = \mu_{i,t-1} \quad (10)$$

$$\sigma^2_{i,t} = \sigma^2_{i,t-1} \quad (11)$$

3. If I_t does not match any component, then the least likely component is replaced with a new one which has $\mu_{i,t} = I_t, \Sigma_{i,t}$ large and $w_{i,t}$ low. After the updates, the weights $w_{i,t}$ are renormalized. The foreground is detected as follows so higher importance gets placed on components with the most evidence and lowest variance, which are assumed to be the background.

E. Kernel density estimation (KDE):

To deal with dynamic backgrounds like water rippling, waving trees and camera jitter, Elgammal et al. [7] proposed to estimate the probability density function for each pixel using the kernel estimator K for N recent sample values {x_1 , x_2,..., x_N } taken consecutively in a time size window W as follows:

$$P(X_t) = \frac{1}{N}\sum_{i=1}^{N} K(X_t - X_i) \quad (12)$$

Where K is the kernel estimator function which is taken as a Normal Gaussian function *N*(O,∑)

Elgammal et al. [7] assumed that the different color channels are independent with different kernel bandwidths, then the kernel function bandwidth is as follows:

$$\Sigma = \begin{vmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{vmatrix} \quad (13)$$

Elgammal et al. [7] detected the foreground using the probabilities and a threshold T as follows:

If $P(X_t)$<T then the pixel classified as foreground otherwise it is classified as background pixel. At this step, a binary mask is obtained. Then, to perform foreground pixel detection, the parameters must be updated. For this, Elgammal et al. [7] used two background models for object detection: a short term one and a long term one. These two models having different objectives: The short term models modify quickly and it allow very sensitive detection. KDE model having most recent N background sample values. The sample is updated using a selective maintenance mechanism, where the decision is based on the foreground object. The long term model is very effective when object is moving slowly. This model consists of N number of sample pixels taken from a much larger window in time. The sample is modified by using a non selective maintenance mechanism.

F. Sequential KD approximation:

Mean-shift vector techniques have recently been developed for various pattern recognition problems such as image segmentation and tracking [8, 9]. The mean-shift vector is an effective gradient technique able to detect the main modes of the true probability density function directly from the sample data with a minimum set of assumptions. However, it has a very high computational cost since it is an iterative technique and it requires a study of convergence over the whole data space. As such, this technique not immediately applicable to model background probability density function at the pixel level. These problems have been solve in recent approaches by [10], Piccardi and Jan. [6] Han et al. compared the probability density function obtained with their method against that of a KDE approach over a 500-frame test video, finding a low mean integrated squared error in the order of 10^{-4}; this justifies the name of sequential Kernel Density approximation (SKDA) that the authors gave to their method.

G. Co-occurrence of image variations:

Seki et al. in [12] try to go across the idea of mere chronological averages by exploiting spatial co-occurrence of image variations. Their main statement is that neighboring blocks of pixels belonging to the background should experience similar variations over time. This method proves hue for blocks belonging to a same background object, it will evidently not hold for blocks at the border of distinct background objects. The method in [12] can be summarized as follows:

Rather than working at pixel resolution, it works on blocks of N x N pixels treated as an N^2 component vector. This trades off resolution with better speed and stability.

Learning phase:

1. For each block, a certain number of time samples is acquired; the temporal average is fast computed and the differences between the samples and the average are called the image variations;

2. The N^2 x N^2 covariance matrix is calculated, with respect to the average and an eigenvector transformation is applied reducing the dimensions of the image variations from N^2 to K.

H. Eigen background:

The approach proposed by Oliver et al. in [13] is also based on Eigen value decomposition, but this time applied to the whole image irrespective of blocks. The method in [13] can be summarized as follows:

Leaning phase:

1. A samples of n images is achieved, each image with ρ pixels; the average image, μ_b is then computed and all images mean is subtracted.

2. The covariance matrix is calculated and the hest M eigenvectors stored in an eigenvector matrix.

Classification phase:

Every time a new image, I, is inserted, it is projected onto the Eigen space as

$$I' = \phi_{Mb}\,(I - \mu_b) \quad (14)$$

Foreground points are detected at locations where

$|I - I'|$

2. PERFORMANCE ANALYSIS:

This section presents a comparative performance analysis based on: -

A. Speed

B. Memory requirements

C. Accuracy.

Table I. Comparative Analysis

Method	*Speed*	*Memory*	*Accuracy*
Single Gaussian	1	1	L / M
Temporal Median filter	n_s	n_s	L / M
Mixture of Gaussian	m	m	H
Kernel Density Estimation	n	n	H
Sequential KD approximation	m+1	m	M / H
Co occurrence of image variation	$8n/N^2$	nK/N^2	M
Eigen Background	M	n	M

A. Speed:

The SG is the fastest method compared to all techniques because it uses threshold and the background maintenance just adapts the mean and the variance. Its complexity depends on N for the initialization. So time complexity as O (1). The median filter has a similar classification cost, but model update can be approximated as linear in the number of samples, n_s. The time complexity for median filter can be stated as O (n_s). The Mixture of Gaussians method has O (m) complexity. KDE model computes its value in the Gaussian kernels centered on the past n frames, So time complexity O (n). The SKDA method has O (m + 1) complexity, where m is the number of modes of the approximated probability density function. Time complexity for the co-occurrence of image variations O $(8*(n+L^4+L)/N^2)$, where n is accounted for searching the nearest neighbors amongst the n variations, L^4 is the estimated cost for computing the interpolation coefficients for applying them to the current block. Finally, the Eigen background method has an approximate complexity per pixel of O (M), where M is the number of the best eigenvectors.

B. Memory requirement:

For the statistical methods, the memory complexity per pixel is the same as the computational complexity. At classification time, reconstructive approaches require a memory complexity per pixel O (P), with P the number of the best eigenvectors used. However, at training time these methods require allocation of all the N training images, with an O (N) complexity.

C. Accuracy:

The methods with a background model based on a single Gaussian value can guarantee adaptation to slow illumination changes; but it cannot cope with multi-valued background distributions. For the approximation of a multi-modal distribution, both techniques such as parametric and non-parametric methods have been applied successfully. For, both the Mixture of Gaussians and KDE approaches can model well background probability density in general cases. In addition, in [14] the proposed KDE temporal model is complemented by having fast time complexity, spatial correlation and a combination of blind and selective update. Mean-shift effectively model a multimodal distribution However, their computational cost is very high. In the SKDA method, they are used only in an initial stage. Examples of the accuracy achievable by the method based on the co-occurrence of image variations can he found in [12]. We experimented the Eigen background method with a training set with n = 20 recent images and M = 3 Eigen backgrounds. The quality of results was good but seemed to significantly depend on the images used for the moving object. In [13], however, the authors report good results with lower computational complexity than a Mixture of Ganssians approach training set.

3. CONCLUSION:

In this paper, we have presented statistical background techniques. This allows comparing the statistical BS methods complexity in terms of speed, memory requirements and accuracy. Simple methods such as the running Gaussian average and median filter offer acceptable accuracy while achieving a high frame rate and both method requires less memory. Methods such as Mixture of Gaussians and KDE prove very good model accuracy. KDE has a high memory requirement. SKDA is an approximation of KDE which proves almost as accurate, but it reduces memory requirement by an order of magnitude and this method had low time complexity. Methods such as the co-occurrence of image variations and the Eigen backgrounds explicitly address spatial correlation. Co-occurrence of image variation and Eigen background offer good accuracy against reasonable time and memory complexity.

4. REFERENCES:

1. Wren, Christopher Richard, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. "Pfinder: Real-time tracking of the human body." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19,no. 7(1997):780-785.
2. Lo, B. P. L., and S. A. Velastin. "Automatic congestion detection system for underground platforms." In *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, pp. 158-161. IEEE, 2001.
3. Stauffer, Chris, and W. Eric L. Grimson. "Adaptive background mixture models for real-time tracking." In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2. IEEE.1999.
4. Friedman, Nir, and Stuart Russell. "Image segmentation in video sequences: A probabilistic approach." In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pp. 175-181. Morgan Kaufmann Publishers Inc.,1997. Conference on Networking and Services, page 45, Washington, DC, USA, 2007.
5. Grimson, W. Eric L., Chris Stauffer, Raquel Romano, and Lily Lee. "Using adaptive tracking to classify and monitor activities in a site." In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pp. 22-29. IEEE, 1998.

6. Ivanov, Yuri, Chris Stauffer, Aaron Bobick, and W. E. L. Grimson. "Video surveillance of interactions." In *Visual Surveillance, 1999. Second IEEE Workshop on,(VS'99)*, pp. 82-89. IEEE, 1999.

7. Elgammal, Ahmed, David Harwood, and Larry Davis. "Non-parametric model for background subtraction." In *Computer Vision—ECCV 2000*, pp. 751-767. Springer Berlin Heidelberg, 2000.

8. Lo Comaniciu, Dorin, and Peter Meer. "Mean shift: A robust approach toward feature space analysis." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24, no. 5 (2002): 603-619.

9. Comaniciu, Dorin. "An algorithm for data-driven bandwidth selection." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25, no. 2 (2003): 281-288.

10. Piccardi, Massimo. "Background subtraction techniques: a review." Systems, man and cybernetics, 2004 IEEE international conference on. Vol. 4. IEEE, 2004.

11. Han, Bohyung, Dorin Comaniciu, Ying Zhu, and Larry S. Davis. "Sequential kernel density approximation and its application to real-time visual tracking". Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 1186-1197. 30.7(2008).

12. Seki, Makito, Toshikazu Wada, Hideto Fujiwara, and Kazuhiko Sumi. "Background subtraction based on co-occurrence of image variations." In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2, pp. II-65. IEEE, 2003.

13. Oliver, Nuria M., Barbara Rosario, and Alex P. Pentland. "A Bayesian computer vision system for modeling human interactions." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22, no. 8 (2000): 831-843.

14. Gallego, Jaime, Montse Pardas, and Gloria Haro. "Bayesian foreground segmentation and tracking using pixel-wise background model and region based foreground model." In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pp. 3205-3208. IEEE, 2009.