International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 3 478 – 483

Advanced Moving Object Detection and Tracking for Video Surveillance

Patil Shital Manohar
Pursuing ME in Digital Electronics at
Gangamai College of Engineering, Nagaon ,
Dhule(M.S) 424005 (India).
e-mail:shitalpatil555@icloud.com

Anuradha B Pawar
Asst. Prof,Department of E&Tc Engineering,
Gangamai College of Engineering, Nagaon ,
Dhule(M.S) 424005 (India).
e-mail: er.anuradhapawar@gmail.com

Abstract—Moving object detection is a very crucial and challenging task in computer vision applications such as surveillance, vehicle and human tracking. Background subtraction is a preliminary technique widely used for the moving object detection. In this paper, an advanced automated moving object detection technique using background subtraction is proposed. The method uses running average wavelet transform (RAWT) for accurate registration of background from the video sequence. Furthermore, the moving objects are detected by comparing current and background frame. In order to produce higher accuracy for the object detection, the proposed method also further includes post-processing filter operation after which the binary object detection mask can be obtained. After moving object detection, tracking is performed. Experimental results demonstrate that the proposed method is faster and efficient as compared to the other state-of-the-art existing methods.

Keywords- *Object detection, background subtraction, wavelet transform, object tracking*

*****

I. Introduction

A. Motivation

In many video surveillance applications, it is important to extract foreground objects of interest, which can be vehicles, animals, person etc., from the background [1]. Many approaches, such as optical flow and background subtraction, have been developed to detect the moving objects, based on which the foreground and the background can be separated. Several sophisticated works for detecting moving objects have been presented for stationary camera. Many researchers have developed various techniques for object detection and tracking, but the challenges are still there. This gives us motivation to develop efficient and accurate surveillance system for the benefits of society [3]. In the analysis of human or traffic flow, the efficiency of a surveillance system significantly depends on its object detection capability. The appearance of objects in a video sequence can be detected using many techniques, which are generally classified into three main categories: 1) frame difference [2]; 2) optical flow [3]; and 3) background subtraction [4]. The background subtraction technique, is the most commonly used method for moving object detection, determines objects by extracting the difference between the current frame and the background. This technique consists of two stages: first is background estimation from the frame sequence and second is foreground extraction by subtracting the estimated background. This paper presents a novel background subtraction method using frequency domain background registration method [6]. The proposed method uses wavelet transform for estimating the accurate background.

Background subtraction is the most widely used approach for detecting the moving objects from a clean background. More seriously, the background may not be absolutely static in the video. Other than camera shake, the scene itself may contain frequent local motions, such as trees/grass waving in the breeze, which could be easily

confused as the foreground. Another problem in background subtraction is thresholding the difference image for foreground segmentation. To address these issues we proposed an efficient background subtraction method using wavelet transform in frequency domain. The proposed method is less computational complex with more accurate as compare to the other state-of-the-art methods in the literature.

B. Summary of major contribution

In this paper, we propose an efficient background subtraction technique using frequency domain approach for detecting the moving objects. With a complete background image for each frame, we can conduct background subtraction to identify the moving foreground objects. To address the local frequent motions in the background, we further develop a running average wavelet transform algorithm that can reduce the foreground false positives in background subtraction.

The major contributions of the paper are summarized as follows.

- The proposed method uses wavelet transform for the extraction of background accurately from the videos.
- The use of frequency domain technique for background registration gives efficient results without any trailing of pixels as compare to spatial domain.

- The computational complexity of the proposed methods is very less as compare to the state-of-the-art methods in the literature.

The rest of the paper is organized as follows. Section II discusses some related work using various methods of background subtraction. Section III describes our proposed method for moving object detection by eliminating noise from previous methods. Section IV is the results and discussion of the proposed method by measuring accuracy.Finally, we concluded the method in section V.

II. RELATED WORK

Many background subtraction techniques with different models or segmentation strategies have been proposed on motion detection. Simple background subtraction (SBS) method technique in which each pixel of the frame (B) distance from the background image is used to classify the pixel as foreground or background. In this method first assume that Background frame is available i.e. B(x, y) with us. Then for detecting moving object the current frame F (x, y) is subtracted from the background and by applying threshold function binary motion detection mask M (x, y) is obtained shown in equation (1) in which moving object is detected as follows[6].

$$M(x,y)=\begin{cases}1, \text{ if } |B(x,y)-F(x,y)|>\tau \\ 1, \text{ if } |B(x,y)-F(x,y)|\le\tau\end{cases} \quad (1)$$

Where τ is some threshold value, if the absolute difference between background and current frame is greater than τ then the object is present in current frame and in binary detection mask it is represented as 1, otherwise if difference is less, then it is represented as 0. But background frame is not always available with us so this is the disadvantage of SBS method

To overcome the disadvantage of simple background subtraction method go for running average method in which registered a background first from taking running average of some frames and then create Binary Mask. The averaging filter is as follows,

$$B_t(x,y)=(1-\beta)B_{t-1}(x,y)+\beta F_t(x,y) \quad (2)$$

Where, $B_{t-1}(x, y)$ is the previous background frame and $F_t(x, y)$ is the current frame and β is the weighting factor. If β is more that means less weight given to background frame and more to the current frame. If β is more than vice-verse. In this way after some frame a Background $B_t(x, y)$ is registered. Then by using that one can generate motion detection mask as follows,

$$M(x,y)\begin{cases}1, \text{ if } |B_t(x,y)-F_t(x,y)|>\tau \\ 0, \text{ if } |B_t(x,y)-F_t(x,y)|\le\tau\end{cases} \quad (3)$$

The same as above done in running average (RA) method but here in this methods utilize discrete cosine transform (DCT) coefficients at block level to represent background, and adapt the background by updating DCT coefficients [8]. Divide the image into 8X8 blocks if apply 8X8 DCT transform then take Running average of each 8X8 block of current and previous frame as follows [3],

$$d^B_{t,k}(x,y)=(1-\beta)d^B_{t-1,k}(x,y)+\beta d_{t,k}(x,y) \quad (4)$$

Where $d^B_{t,k}$ denotes the DCT coefficient vector for the k^{th} pixel block of the current frame at time t and β is the weight parameter. After getting $d^B_{t,k}$ binary mask can be detected easily. This method registered background fast and efficiently without any trailing behind the object

The problem with the Running Average method is, it doesn't take care of current frame intensity i.e. while registering background if the intensity in further frames is changing then background is not register accurate and based on that further Binary motion detection mask is not getting properly. So to maintain intensity updated throughout the frames. The sgn function is define as,

$$\text{sgn}(t)=\begin{cases}1, & \text{if } t>0 \\ 0, & \text{if } t=0 \\ -1, & \text{if } t<0\end{cases} \quad (5)$$

The sgn function for each frames [9], each pixel (x, y) is,

$$B_t(x,y)=B_{t-1}(x,y)+\text{sgn}(F_t(x,y)-B_{t-1}(x,y)) \quad (6)$$

Then time variance is calculated separately i.e. the variance between two pixels of each consecutive frame. Now by comparing time variance and frame difference binary motion detection mask is generated.

$$V_t(x,y)=V_{t-1}(x,y)+\text{sgn}(N\times\Delta_t-V_{t-1}(x,y)) \quad (7)$$

$$M(x,y)=\begin{cases}1, \text{ if } \Delta_t(x,y)>V_t(x,y) \\ 0, \text{ if } \Delta_t(x,y)\le V_t(x,y)\end{cases} \quad (8)$$

Here, temporal averages of some consecutive frames are taken [10]. Each pixel in B_t (x, y) has two timers associated with it, T_l and T_s. T_l is called the long term timer, which counts the number of frames that a pixel in B_t (x, y) has been steady in value. And T_s is the short term timer, counts the number of

 ISSN: 2321-8169
478 – 483

frames for which the pixel differed from the value in $B_t(x, y)$. If T_s is greater than T_l then the pixels of B_t (x, y) are updated by the new incoming frame F_t (x, y) and the short term timer T_s can subsequently be reset to zero. Also if $T_l > \alpha$(some tolerance) then T_l is reset to α. Temporal averaging produces poor results when movement occurs for long periods of time, or if an object stays in a position for a long period of time.

III. PROPOSED METHOD

A. Moving object detection

The capability of extracting moving objects from a video sequence is a typical firststep in computer vision applications. The different approaches of background registration methods are describe in previous section. If background is good then based on that obtained binary mask is also good and correct. Now from the above different methods RA-DCT [8] is the best and computationally fast for background registration but due to some noise in the background mask the further accuracy Metrics gives poor results. So by applying some morphological operation and median filtering on background mask the noise is reduced and getting improved results of moving object detection. An overview of the proposed method can be illustrated in figure 1.

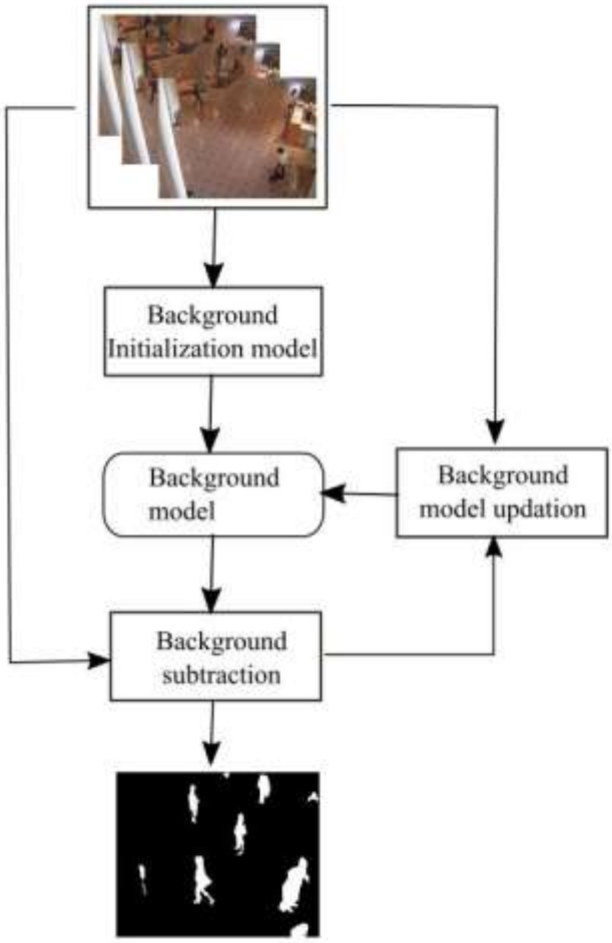

Fig.1 Overview of the proposed method considering mall database image.

The overview of the proposed method is shown above. The background initialization model content the available background with us. After few hundred frame the background updating performed using wavelet transform in frequency domain. The on level wavelet decomposition of an image on WS database is shown in Fig.3 using the decomposition scheme shown n Fig.2. Once the background is registered then, after taking difference of current frame with background frame we got the moving object detected mask as shown in Fig.1.

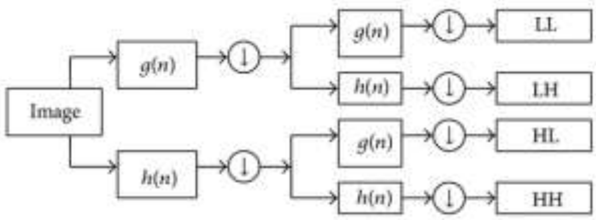

Fig.2. Low pass and high pass filter-bank for wavelet decomposition of the image.

Two dimensional DWT can be applied to decompose an image into four sub-images approximation and three details.The discrete wavelet transform was adopted to solve the problem of fake motions in the background.The proposed technique utilizes the 2D DWT to overcome the drawbacks of small/slowly moving objects and illumination variations.

Fig. 3 Output of wavelet transform on WS database.

From the image sequence by applying RA-WT algorithm as discuss previously it utilized discrete wavelet transform (DWT) coefficients at block level to represent background, and adapt the background by updating DWT coefficients.

$$d_{t,k}^{B}(x,y) = (1-\beta)d_{t-1,k}^{B}(x,y) + \beta d_{t,k}(x,y) \quad (9)$$

Where $d_{t,k}^{B}$ denotes the WT coefficient vector for the kth pixel block of the current frame. Now by taking absolute difference

 ISSN: 2321-8169
478 – 483

between current frame and background by RA-WT the binary detected mask got as shown in figure1.

B. Object tracking

After getting detected moving object mask using RA-WT background subtraction method the next task is to track the detected object. Standard Deviation of detected Regions is calculated. The Standard Deviation block computes the standard deviation of each row or column of the input, alongvectors of a specified dimension of the input, or of the entire input. The Standard Deviation block can also track the standard deviation of a sequence of inputs over a period of time. Then by comparing standard deviation to some threshold value draw a rectangle around the detected mask.

IV. RESULTS AND DISCUSSION

Experimental results for moving object detection using the proposed approach have been produced for several image sequences. Here, we describe three different sequences water surface (WS), fountain and college. Background registered for each sequence is shown in figure 4.

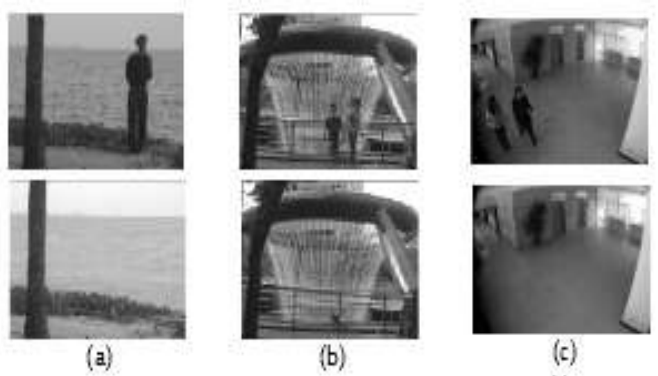

Fig. 4 Registered background of sequence (a) WS (b) Fountain (c) College.

The ground truth and obtained results of RA, SDE and TMF methods are shown in figure 3. Here, for WS sequence used 1400 to 1616 frames for background and detect object at 1616th frame. For fountain sequence used 1000 to 1190 frames while for college used 1 to 127 frames.

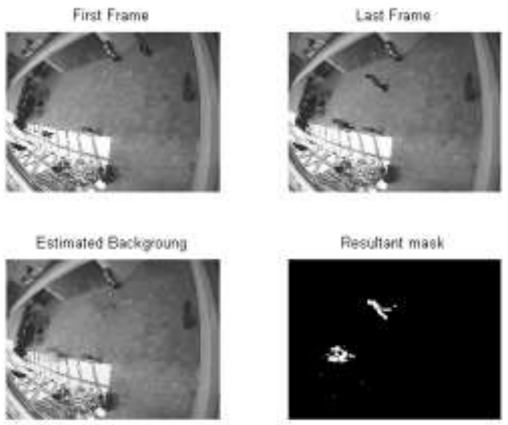

Fig.5. Running average method results.

For a background subtraction frame work in compressed domain, which models background directly from compressed video using DCT coefficients and is able to extract moving objects at the pixel resolution. Some RAWT results shown in following figure 6. In this RAWT method each current frame is divided into 8 by 8 pixel blocks in the spatial domain, and then each block is transformed by WT into a set of coefficients in the frequency domain to reduce spatial redundancy.

This RAWT method registers background fast and efficiently without any trailing behind the object. The result is shown in fig.6.The performance evaluation of moving object detection and tracking methods is a crucial and tedious task. There are subjective and objective (or quantitative) methods toevaluate the performance of detection and/or tracking algorithm.

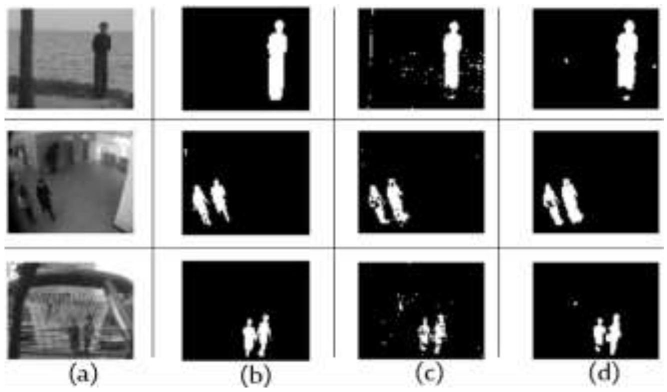

Fig. 6 Results on WS, college and fountain database. (a) Simple background subtraction (b) Running average (c) RAWT method.

Furthermore, the detected objects can track using tracking methods as shown in Fig. 7.

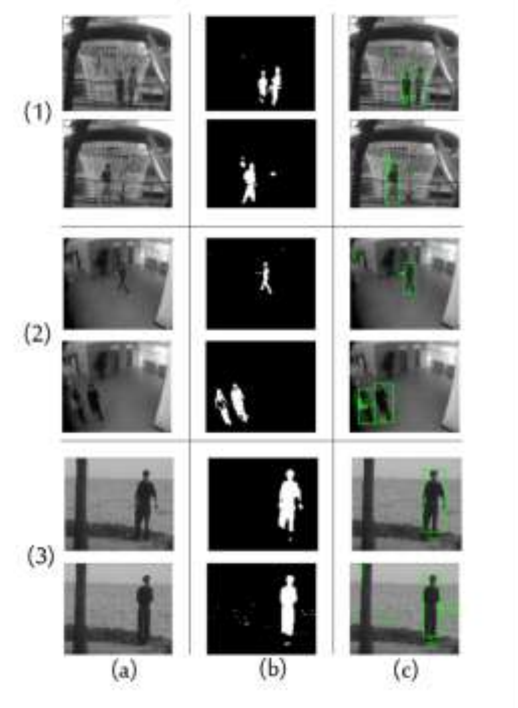

Fig. 7 Result of Tracking after object detection from frames on different databases (1) Fountain (2) Mall (3) WS.

TABLE 1. RESULTS ON WS DATABASE

Metrics	SBS	Runn. Avg.	RA-DCT	Proposed
Recall	0.5676	0.5014	0.8867	**0.9054**
Precision	0.6123	0.6474	0.8324	**0.9012**
F1 test	0.6345	0.5651	0.8600	**0.8921**
Similarity	0.5143	0.3939	0.7153	**0.8421**

TABLE 2. RESULTS ON FOUNTAIN DATABASE

Metrics	SBS	Runn. Avg.	RA-WT	Proposed
Recall	0.6523	0.7346	0.6783	**0.8785**
Precision	0.6433	0.6356	**0.8467**	0.8240
F1 test	0.5832	0.6130	0.7572	**0.8435**
Similarity	0.5912	0.5556	0.6092	**0.7676**

TABLE 3. RESULTS ON COLLEGE DATABASE

Metrics	SBS	Runn. Avg.	RA-WT	Proposed
Recall	0.7776	0.8032	0.8672	**0.9232**
Precision	0.6346	0.8359	**0.8365**	0.7391
F1 test	0.6540	0.8192	0.8158	**0.8291**
Similarity	0.5240	0.6938	0.7079	**0.7336**

TABLE 4. RESULTS ON MALL DATABASE

Metrics	SBS	Runn. Avg.	RA-WT	Proposed
Recall	0.5233	0.8321	0.8312	**0.9022**
Precision	0.6432	0.8234	0.8233	**0.8591**
F1 test	0.6123	0.8101	0.8238	**0.8201**
Similarity	0.5091	0.6623	0.7079	**0.7334**

For measuring accuracy different metrics are used namely Precision, Recall, F1 test, and Similarity [6], [11]. Recall, also known as detection rate, gives the percentage of detected true positives pixel as compared to the total number of true positives in the ground truth.

$$recall = \frac{tp}{tp + fn} \qquad (13)$$

Where, tp is the total number of true positives pixels, fn is the total number of false negatives, and (tp + fn) indicates the total number of pixels present in the ground truth. Recall alone is not enough to compare different methods, and is generally used in

$$precision = \frac{tp}{tp + fp} \qquad (14)$$

Here, fp is the total number of false positives, (tp + fp) indicates the total number of detected pixels from the output mask. Moreover, we considered the metric, also known as Figure of Merit or F-measure, that is the weighted harmonic mean of Precision and Recall,

$$F_1 = \frac{2 * recall * precision}{recall + precision} \qquad (15)$$

And similarity measure is given in Eq. (16) as follows,

$$Similarity = \frac{tp}{tp + fp + fn} \qquad (16)$$

V. CONCLUSIONS

In this paper, we summaries the development of various background modeling techniques for moving object detection. Also, we proposed a novel background subtraction technique using wavelet transform. Furthermore, by applying median filtering and morphological opening, closingoperation the proposed approach can increase the accuracy of object detection. The obtaining results are compared with the other existing methods available in the literature using different accuracy metrics. Moreover, the proposed method comparably faster and accurate for detecting moving objects for video surveillance.

REFERENCES

[1] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based meth-ods for action repre-sentation, segmentation and recognition," Computer Vision and Image Understanding, vol. 115, no. 2, pp. 224 – 241, 2011.

[2] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "*A survey of video processing techniques for traffic applications*," Image and Vision Com-puting, vol. 21, no. 4, pp. 359 – 381, 2003.

ISSN: 2321-8169
478 – 483

[3] S.-C. Huang and F.-C. Cheng, "*Motion detection with pyramid structure of background model for intelligent surveillance systems*, " Engineering Applications of Artificial Intelligence, vol. 25, no. 7, pp. 1338 – 1348, 2012.

[4] R. Zhang and J. Ding, "*Object tracking and detecting based on adaptive background subtraction*," Procedia Engineering, vol. 29, no. 0, pp. 1351 – 1355, 2012.

[5] Y. Cao, Z. Lei, X. Huang, Z. Zhang, and T. Zhong, "*A vehicle detection algorithm based on compressive sensing and background subtraction*," {AASRI} Procedia, vol. 1, no. 0, pp. 480 – 485, 2012.

[6] S.-C. Huang, "*An advanced motion detection algorithm with video quality analysis for video surveillance systems*," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 21, no. 1, pp. 1–14, 2011

[7] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "*Background and foreground modeling using nonparametric kernel density estimation for visual surveillance*," Proceedings of the IEEE, vol. 90, no. 7, pp. 1151–1163, 2002.

[8] W. Wang, L. Yang, and W. Gao, "*Modeling background and segmenting moving objects from compressed video*," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 5, pp. 670–681, 2008.

[9] A. Manzanera and J. C. Richefeu, "*A new motion detection algorithm based on sigma-delta background estimation*," Patt. Recog, vol. 28, no. 0, p. 320328, 2007.

[10] B. Shoushtarian and N. Ghasem-Aghaee, "*A practical approach to realtime dynamic background generation based on a temporal median filter*," J. Sci. Islamic Republic Iran, vol. 14, no. 4, p. 351362, 2003.

[11] L. Maddalena and A. Petrosino, "*A self-organizing approach to background subtraction for visual surveillance applications*, " Image Process-ing, IEEE Transactions on, vol. 17, no. 7, pp. 1168–1177, 2008.

[12] R. Babu, K. Ramakrishnan, and S. Srinivasan, "*Video object segmentation: a compressed domain approach*," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, no. 4, pp. 462–474, 2004.

[13] L. Li, W. Huang, I.-H. Gu, and Q. Tian, "*Statistical modeling of complex backgrounds for foreground object detection*," Image Processing, IEEE Transactions on, vol. 13, no. 11, pp. 1459–1472, 2004.

[14] L. Maddalena and A. Petrosino, "*A self-organizing approach to background subtraction for visual surveillance applications,* " Image Process-ing, IEEE Transactions on, vol. 17, no. 7, pp. 1168–1177, 2008.

[15] J. Bescos, "*Real-time shot change detection over online mpeg-2 video*," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, no. 4, pp. 475–484, 2004.

[16] M. Kim, J. G. Choi, D. Kim, H. Lee, M.-H. Lee, C. Ahn, and Y.-S. Ho, "*A vop generation tool: automatic segmentation of moving objects in image sequences based on spatio-temporal information*," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 9, no. 8, pp. 1216–1226, 1999.