Advanced Motion Detection Technique using Running Average Discrete Cosine Transform for Video Surveillance Application

Ravi Kamble¹, Sushma Kejgir²

¹ Dept. of Electronics and Telecom. Engg. SGGS Institute of Engineering and Technolgy, Nanded - 431606, India.

Abstract — Object detection is always the first important step in video surveillance applications. This paper presents an automated moving object detection technique using background subtraction method. First, a background module generated from the video sequence effectively. Furthermore, the moving objects are detected by comparing current and background frame. Many background subtraction approaches are available but most of them are not illumination sensitive. The proposed method updated the current status of the background with sudden illumination change background in frequency domain using modified running average discrete cosine transform (RA-DCT). In order to produce higher accuracy for the motion detection, the proposed method also further includes a noise filter after which the binary motion detection mask can be computed. Experimental results demonstrate that the proposed method is faster and efficient as compared to the other existing methods.

Keywords — Motion detection, Background subtraction, Object tracking, Video surveillance.

I. INTRODUCTION

In recent years video surveillance aids in many aspects and become an important component by monitoring traffic at the signals, crowd at the arena, home security and many other social application [1], [2]. 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. Evolving from CCTV video surveillance, the IP camera surveillance system with Internet as the connection backbone is a trend in recent years. Now surveillance is going towards automatic way. The growth in this area is being driven by the increased availability of inexpensive computing power and image sensors, as well as the inefficiency of manual surveillance and monitoring systems. The major challenge of advanced video surveillance is the motion detection because it must be adaptive to different environmental and illumination conditions [3], [4], [5]. From motion detection the moving object is extracted. According to a survey of related literature motion detection approaches can be broadly classified as three categories temporal difference, optical flow and background subtraction [6]. Among that Back-ground subtraction is the most popular and less computational complexity method. Temporal difference approaches detect moving objects by calculating the differences between pixels in consecutive frames of a video sequence but it is illumination sensitive. Optical flow method is more computational complex and not easy to implement. This paper presents a novel background subtraction method using frequency domain discrete cosine transform. Furthermore, we also used noise filter and reduce the detection noise effectively for more accurate detection.

The remainder of this paper is organized as follows. Section 2 discusses some previous method of background subtraction. Section 3 describes our proposed method for motion detection by eliminating noise from previous methods. Section 4 is the quantitative and qualitative comparison of results for measuring accuracy.

II. LITERATURE SURVEY

Many algorithm have been developed previously for background subtraction like Simple Background Subtraction (SBS) method, Running Average (RA), Temporal Median Filter (TMF), Sigma Delta Estimation (SDE), Simple Statistical Difference (SSD) method [6], [7], and [3]. Some of them are adaptive to change in background intensity but not registered background efficiently. We summarized some of the techniques in following section.

A. Simple Background Subtraction Method

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].

\[
M(x, y) = \begin{cases} 
1, & \text{if } |B(x, y) - F(x, y)| > \tau \\
1, & \text{if } |B(x, y) - F(x, y)| \leq \tau 
\end{cases}
\]
Where \( \tau \) is some threshold value, if the absolute difference between background and current frame is greater than \( \tau \) 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

**B. Running Average 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,

\[
B_t(x, y) = (1 - \beta)B_{t-1}(x, y) + \beta F_t(x, y)
\]

(2)

Where, \( B_{t-1}(x, y) \) is the previous background frame and \( F_t(x, y) \) is the current frame and \( \beta \) is the weighting factor. If \( \beta \) is more that means less weight given to background frame and more to the current frame. If \( \beta \) is more than vice-versa. 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,

\[
M(x, y) \begin{cases} 
1, & \text{if } |B_t(x, y) - F_t(x, y)| > \tau \\
0, & \text{if } |B_t(x, y) - F_t(x, y)| \leq \tau
\end{cases}
\]

(3)

**C. Running Average with DCT**

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],

\[
d^{\beta}_{t,k}(x, y) = (1 - \beta)d^{\beta}_{t-1,k}(x, y) + \beta d_{t,k}(x, y)
\]

(4)

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

**D. Sigma Delta Estimation**

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,

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

(5)

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

\[
B_t(x, y) = B_{t-1}(x, y) + \text{sgn}(F_t(x, y) - B_{t-1}(x, y))
\]

(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.

\[
V_t(x, y) = V_{t-1}(x, y) + \text{sgn}(N \times \Delta - V_{t-1}(x, y))
\]

(7)

\[
M(x, y) = \begin{cases} 
1, & \text{if } \Delta(x, y) > V_t(x, y) \\
0, & \text{if } \Delta(x, y) \leq V_t(x, y)
\end{cases}
\]

(8)

**E. Temporal Median Filter**

Here, temporal averages of some consecutive frames are taken [10]. Each pixel in \( B_t(x, y) \) has two timers associated with it, \( T_1 \) and \( T_2 \). \( T_1 \) 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 \( T_2 \) is the short term timer, counts the number of frames for which the pixel differed from the value in \( B_t(x, y) \). If \( T_2 \) is greater than \( T_1 \) then the pixels of \( B_t(x, y) \) are updated by the new incoming frame \( F_t(x, y) \) and the short term timer \( T_2 \) can subsequently be reset to zero. Also if \( T_2 > \alpha \) (some tolerance) then \( T_1 \) is reset to \( \alpha \). 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**

The capability of extracting moving objects from a video sequence is a typical first step 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 give 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.
From the image sequence by applying RA-DCT algorithm as discussed previously it utilized discrete cosine transform (DCT) coefficients at block level to represent background and adapt the background by updating DCT coefficients.

\[ d_{x,y}^B (x, y) = (1 - \beta)d_{x,y}^B (x, y) + \beta d_{x,y} (x, y) \] (9)

Now by taking absolute difference between current frame and background by RA-DCT the binary detected mask got as shown in figure 1.

\[ \Theta \ B. \ ] \text{Now morphological closing is the dilation followed by erosion i.e,} \]

\[ D(x, y) \ast B = (D(x, y) \oplus B) \Theta B \] (10)

and in opening the reverse that is erosion followed by dilation is done.

\[ D(x, y) \circ B = (D(x, y) \Theta B) \oplus B \] (11)

so closing connects objects that are close to each other tend to fill the holes. After that apply median filter of size $5 \times 5$ to remove the unwanted noise. The median filter is an effective method that can suppress isolated noise without blurring sharp edges. Specifically, the median filter replaces a pixel by the median of all pixels in the neighborhood.

\[ y[i, j] = \text{median}[F(i, j), (i, j) \in w] \] (12)

where $w$ represents a neighborhood centered around location $(i,j)$ in the image. So after registering background using RA-DCT and applying morphological and filtering operation we get more accurate results as shown in figure 5.

B. Tracking Detected Objects

After getting detected moving object mask using RA-DCT 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, along vectors 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. EXPERIMENTAL RESULTS

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 2.
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.

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 RADCT results shown in following figure. In this RADCT method each current frame is divided into 8 by 8 pixel blocks in the spatial domain, and then each block is transformed by DCT into a set of coefficients in the frequency domain to reduce spatial redundancy.

Fig. 3. RA-DCT result ws sequence (a) Current frame (b) Registered background (c) Detected result.

This RADCT method registers background fast and efficiently without any trailing behind the object. The result is shown in fig.4.

Fig. 4. RA-DCT result college sequence (a) Current frame (b) Registered background (c) Detected result.

Fig. 5. Ground truth and detected mask results of previous RA, SDE, TMF methods of database (1) WS (1616th) frame (2) Fountain (1190th) frame (3) College (127th) frame.

Now the results of RA-DCT and proposed method after applying filtering operation are shown in Figure 6. We are able to detect object accurately by eliminating noise effectively using noise filter.

Fig. 6. Detected results of sequence (a)WS (b) fountain (c) college using RA-DCT and PROPOSED method.

A. Accuracy Metrics

The performance evaluation of moving object detection and tracking methods is a crucial and tedious task. There are subjective and objective (or quantitative) methods to evaluate the performance of detection and/or tracking algorithm. 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.

\[
\text{recall} = \frac{tp}{tp + fn} \tag{13}
\]

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 conjunction with Precision, also known as positive prediction, that gives the percentage of detected true positives as compared to the total number of items detected by the method.

\[
\text{precision} = \frac{tp}{tp + fp} \tag{14}
\]

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,
\[ F_i = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \]  
(15)

And similarity is,

\[ \text{Similarity} = \frac{tp}{tp + fp + fn} \]  
(16)

<table>
<thead>
<tr>
<th>Frames</th>
<th>Metrics</th>
<th>Runn. Avg.</th>
<th>RA-DCT</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. WS</td>
<td>Recall</td>
<td>0.5014</td>
<td>0.8603</td>
<td>0.9119</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.6474</td>
<td>0.8486</td>
<td>0.9067</td>
</tr>
<tr>
<td></td>
<td>F1 test</td>
<td>0.5651</td>
<td>0.8544</td>
<td>0.8893</td>
</tr>
<tr>
<td></td>
<td>Similarity</td>
<td>0.3939</td>
<td>0.7458</td>
<td>0.8310</td>
</tr>
<tr>
<td>2. Fountain</td>
<td>Recall</td>
<td>0.7666</td>
<td>0.6783</td>
<td>0.8825</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.6356</td>
<td>0.8467</td>
<td>0.8190</td>
</tr>
<tr>
<td></td>
<td>F1 test</td>
<td>0.6950</td>
<td>0.7572</td>
<td>0.8314</td>
</tr>
<tr>
<td></td>
<td>Similarity</td>
<td>0.5326</td>
<td>0.6092</td>
<td>0.7422</td>
</tr>
<tr>
<td>3. College</td>
<td>Recall</td>
<td>0.8032</td>
<td>0.8672</td>
<td>0.9212</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>0.8359</td>
<td>0.8365</td>
<td>0.7391</td>
</tr>
<tr>
<td></td>
<td>F1 test</td>
<td>0.8192</td>
<td>0.8158</td>
<td>0.8191</td>
</tr>
<tr>
<td></td>
<td>Similarity</td>
<td>0.6938</td>
<td>0.7079</td>
<td>0.7236</td>
</tr>
</tbody>
</table>

Notice that the optimal value for F1 test is 1 (Precision =Recall = 1), and the worst value is 0 (Precision = Recall =0) based on these metrics we made comparison as shown in table 1.

The gray scale distribution for 500 Frames of water surface (WS) database is shown in figure 7. We see the average intensity is 120 on y axis and the object came in picture from frame 260 to 300, so used the running average filter for registering background.

After getting detected moving object mask using RA-DCT background subtraction method the next task is to track the detected object. Now the Standard Deviation of detected Regions is calculated. The Standard Deviation block computes the standard deviation of each row or column of the input, along vectors of a specified dimension of the input, or of the entire input.

Then by comparing standard deviation to some threshold value draw a rectangle around the detected mask. Rectangle function in matlab draws a rectangle with Position: ('Position',[x,y,w,h]). Draws the rectangle from the point x,y and having a width of w and a height of h. When the interest moving object still cannot be tracked, then the moving object is categorized as not interest moving object or another object and the tracking process is begun again from the beginning.

Furthermore, the detected objects can track using tracking methods as shown in Fig. 8.

In this paper, we present a novel method for moving object detection using efficient running average DCT method. Also by applying median filtering and morphological opening, closing operation 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


