Symmetric Mask Wavelet Based Detection and Tracking of Moving Objects Using Variance Method

Resmi H.B\textsuperscript{a,*}, Deepambika V.A\textsuperscript{b}, M.Abdul Rahman\textsuperscript{c}

\textsuperscript{a}Department of ECE, LBS Institute of Technology for Women, Trivandrum, India
\textsuperscript{b}Department of ECE, LBS Institute of Technology for Women, Trivandrum, India
\textsuperscript{c}Kerala Technical University, Trivandrum, India

Abstract

Fast and accurate object tracking is very important for real time applications like video surveillance, traffic monitoring etc. In most of the conventional object tracking methods environmental changes, memory requirement and computation speed are the major constraints. This paper proposes an efficient object tracking method to compensate for all these challenges. Here a Symmetric Mask based on Discrete Wavelet Transform (SMDWT) has been used in order to compensate for illumination changes and low memory requirement. Symmetric Mask Discrete Wavelet Transform is a low complexity DWT where each filter is derived from the 2-D DWT of 5/3 integer lifting-based coefficients. For fast object tracking variance method is adopted where maximum nonzero pixel value is considered. The experimental results show that the proposed method yields better result on the basis of computational time, memory requirement, speed of operation and environmental changes than the conventional DWT based approach.

Keywords: Object detection; Object tracking; DWT, SMDWT; Frame differencing; Variance method.

*Corresponding author. Tel.: +917034987894
E-mail address: resmihb@gmail.com
1. Introduction

The intelligent video surveillance system has become an essential area of research topic in computer vision that tries to detect, recognize and track object from input video sequences. Successful motion detection in real time applications is a difficult task because of so many challenges which include illumination changes, fake motion, Gaussian noise in the background etc. that may lead to incorrect motion detection. The conventional approaches for motion detection include background detection\(^1\), temporal differencing\(^2\) and optical flow method\(^3\). In background subtraction approach the moving regions are detected by subtracting the current frame from background frame. It gives satisfactory motion mask data, but the problem with this approach is the dynamic scene changes due to lightening and other radiometric variations; therefore we need to frequently update the background frame. In temporal differencing approach the moving regions are extracted from consecutive frames of the video sequence. Even though it is suitable for dynamic environment, the problem with this approach is the incomplete motion extraction. In optical flow method, flow vector characteristics of the moving object over a time is used to detect the moving region. This method also gives satisfactory result but the computational complexity is very high. The most common problems with all of the above mentioned approaches are the sensitivity to illumination changes, noises, fake motion such as motion of tree leaves and memory requirement for the storage purpose and the speed.

DWT and Low resolution techniques like average filtering are used to reduce fake motions and computing cost. But the conventional DWT scheme has the disadvantages of complicated calculations. Moreover, if it uses LL3-band image, it may cause incomplete detection. To overcome this problem Symmetric Mask Based Discrete Wavelet Transform (SMDWT) is used. It preserves the quality of low resolution images better than that of the other low resolution methods\(^4\). Here spatial filtering is done using four different masks and masking is performed in two levels. Then the variance method is used for fast object detection and tracking.

The rest of the paper is arranged as follows: the proposed algorithm is described in section 2. In section 3 experimental result of the proposed algorithm is compared with conventional DWT based variance method. Section 4 provides the conclusion.

2. Proposed Method

For video compression and edge detection discrete wavelet transform is used. For extracting moving regions between the consecutive frames, frame differencing method is adopted. Finally for fast object detection and tracking variance method is used.

2.1. Discrete wavelet transform

Discrete wavelet transform is used in various fields such as signal processing, image processing, computer vision, video compression, biochemistry medicine etc. It gives a flexible multi-resolution image and decomposes the image into various sub-band images including low and high band frequencies.

DWT can be given by:

\[ H_i = \sum_{m=0}^{k-1} x(2i - m).s_m(z) \tag{1} \]

\[ L_i = \sum_{m=0}^{k-1} x(2i - m).t_m(z) \tag{2} \]
where \( S_m(z) \) and \( T_m(z) \) are called wavelets filters, \( k \) is length of the filter, and \( i = 0 \ldots \lfloor n/2 \rfloor - 1 \). It gives the discrete wavelet representation of the discrete time signal. Here the input video sequence is decomposed into four components LL, LH, HL, and HH as shown in the Fig 1. The noise is suppressed in high frequency components. So the further processing is done on low frequency (LL) band.

\[ S_m(z) \] and \( T_m(z) \) are called wavelets filters, \( k \) is length of the filter, and \( i = 0 \ldots \lfloor n/2 \rfloor - 1 \). It gives the discrete wavelet representation of the discrete time signal. Here the input video sequence is decomposed into four components LL, LH, HL, and HH as shown in the Fig 1. The noise is suppressed in high frequency components. So the further processing is done on low frequency (LL) band.

Fig 1. Discrete wavelet transform

2.2 Symmetric mask based discrete wavelet transform

For accurately tracking the moving object, time and memory are the two constraints. Features such as short critical path, high speed operation, regular signal coding, and independent sub-band processing make SMDWT more efficient than the 2D-DWT. 2-D 5/3 integer LDWT is used in SMDWT. For increasing the computation speed, and reducing the complexity, spatial filtering is performed. In spatial filtering four masks, 3×3, 5×3, 3×5, and 5×5, are used where 3x3 is used for HH band, 5x3 for HL, 3x5 for LH and 5X5 for LL. Further processing is done on LL2 band. The coefficients of each sub-band mask are shown in Fig. 2.

Fig 2. (a) LL Mask Coefficient (b) HL Mask Coefficient (c) HH Mask Coefficient (d) LH Mask Coefficient

2.3 Background subtraction and frame differencing.

For steady state condition, background subtraction method is adopted. Here moving regions in the frames are not considered for reference or background frame. The reference frame is subtracted from the current frame to get the targeting object in the frame. In frame differencing method, the moving regions are extracted from successive frames of the video sequence. Moving regions are detected from this method. Background subtraction and frame differencing approaches can be represented as in equation 3 and 4

\[ BS(i,j) = F(i,j) - F(i,j)_{k=0} \quad (3) \]

\[ FD(i,j) = F(i,j) - F(i,j)_{k=1} \quad (4) \]

where \( i = 1, \ldots \), \( n \) and \( j = 1, \ldots, n \) \( k = 1, 2, 3 \ldots \), \( q \); \( q \) is the number of frames in the video. BS is the background frame and FD is the frame difference.
2.3. Binarization and OR operation

The process of conversion of gray image into black and white image is called binarization. The gray scale image consists of 256 intensity levels which are converted into two equal levels of either 0’s (dark) or 1’s (white). This can be done by taking the threshold value. For binarization, selection of threshold value is very important. From frame to frame the threshold value may change due to changes in the average frame intensity from one frame to another. Some parameters like mean $U$ and standard deviation $\sigma$ is used to calculate the global threshold value for a particular frame.

\[
U = \frac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} FD(i, j) \quad (5)
\]

\[
\sigma = \sqrt{\frac{1}{rc} \sum_{i=1}^{r} \sum_{j=1}^{c} (FD(i, j) - U)^2} \quad (6)
\]

\[
T = 0.06 \ast \sigma \quad (7)
\]

Here $FD(i,j)$ is the frame difference of successive frames, and $[r, c]$ is the size of the input image. The threshold value is then obtained as shown in equation (7) and the constant value 0.06 is obtained through trial and error method. After binarization, $OR$ operation is carried out in order to detect the object while moving, and when object stops suddenly. $OR$ operation can be given by:

\[
\{RV(i, j)\}_k = \{BS(i, j)\}_k \lor \{FD(i, j)\}_k \quad (8)
\]

$RV$ is the resultant frame of $OR$ operation. Since exact thresholding is not possible due to the presence of noise, some morphological operations are also done on the binarized frame. To remove the noise, pixel opening operation is done first, since object pixels also get eliminated during this process. Closing operation is done after this to retrieve the lost pixel. The result is shown in Fig 3.

![Fig 3. Binarized frame](image)

2.4. Variance method

It is used to detect and track the moving region. In this step the variance values of the entire row and column of the frame is calculated. It gives image intensity variations corresponding to each row and column. Variance is given as:

\[
\sigma^2 = \frac{1}{n^2} \sum_{i=1}^{n} RV^2 (i,j) \quad (9)
\]

Where, $RV$ is the sub image and $n$ is the number of pixels in the sub image. First, the intersection point of row and column where the value of variance is maximum is considered. Then a window is formed around it and the same process is carried out in the window resulting in a new intersection point. The window is shifted around the new intersection point until same point is obtained as intersection in successive iteration. Here the process converges to a
point. Finally, for showing the location of the object in the input frame, a bounding box is placed around the object. This will help in tracking the object in the video. The above steps are summarized in a flow chart, as shown in Fig.4.
3. Experimental Result

This section gives the comparative result of the various methods done. The tests were carried out on an Intel(R) Core(TM) i3 processor with 2.10GHz clock frequency and 4GB of RAM. CAVIAR\textsuperscript{11} project at INRIA labs database\textsuperscript{12} are used. The details about the videos are given in Table1. Comparing the result of SMDWT and DWT from the Fig (7) and Fig (8), we can see that the objects are accurately detected using SMDWT method where three objects are detected from the input frame; but in the case of DWT method only partial results are obtained, leaving all other objects eliminated from the frame. So DWT may cause incomplete detection. For tracking, we calculate the variance of the image where the pixel with maximum nonzero value is considered. The computational time of SMDWT is also lesser compared to DWT as shown in Table 2.
4. Conclusion

A combined SMDWT and variance method has been used for object tracking. Since SMDWT is insensitive to illumination changes, an accurate detection can be achieved; here the memory requirement is also less. The use of variance method for tracking increases the computational speed to a great extent. The comparative results show that proposed method is much better than the conventional approach in detection and tracking time. On account of these advantages, the proposed method is very suitable for real time applications such as video surveillance, traffic monitoring system etc.

References


Table 1. Video Details

<table>
<thead>
<tr>
<th>Videos</th>
<th>Size of Frame</th>
<th>Number of Frames</th>
<th>Duration of Video (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browse1.mpg</td>
<td>384X288</td>
<td>1040</td>
<td>41</td>
</tr>
<tr>
<td>Walk1.mpg</td>
<td>384X288</td>
<td>610</td>
<td>24</td>
</tr>
<tr>
<td>Rest_InChair.mpg</td>
<td>384X288</td>
<td>1007</td>
<td>42</td>
</tr>
<tr>
<td>LeftBag.mp4</td>
<td>384X288</td>
<td>1439</td>
<td>56</td>
</tr>
<tr>
<td>Women.mp4</td>
<td>480X640</td>
<td>201</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 2. Comparison on the basis of computation time.

<table>
<thead>
<tr>
<th>Videos</th>
<th>SMDWT-Variance Method</th>
<th>DWT-Variance Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browse1.mpg</td>
<td>2.17</td>
<td>4.49</td>
</tr>
<tr>
<td>Walk1.mpg</td>
<td>2.9</td>
<td>5.14</td>
</tr>
<tr>
<td>Rest_InChair.mpg</td>
<td>2.23</td>
<td>3.02</td>
</tr>
<tr>
<td>Walk2.mp4</td>
<td>2.7</td>
<td>5.05</td>
</tr>
<tr>
<td>Women.mp4</td>
<td>5.2</td>
<td>6.91</td>
</tr>
</tbody>
</table>

11. For video database: http://groups.inf.ed.ac.uk/vision/ CAVIAR/ CAVIARDATA1/