TRACKING AND COUNTING HUMAN IN VISUAL SURVEILLANCE SYSTEM

Ms. Kavita P. Mahajan¹, Prof. S. V. Patil²
¹(Research Scholar, ME E&TC, JTMCOE, Faizpur, kavitamahajan85@gmail.com)
²(Asst. Prof, E&TC Dept, JTMCOE, Faizpur, svpatil67@yahoo.co.in)

ABSTRACT

The moving object detection is one of the challenging problems in the visual surveillance system, especially when illumination changes and shadow exists. The proposed system first separates the frames in a video. Then background subtraction method is used for moving object detection. This method can be used on gray scale image format as well as on binary image format. Human detection process is carried out on both the formats and the format with best result is carried further for tracking and counting the humans present in a video. After detection of objects, the feature extraction is performed to get more knowledge about the objects. Then tracking of objects is performed according to their size. Finally counting of humans is performed to get the total number of people in a video. The accuracy of detection obtained with this system is up to 100%. Further this system requires very less time to track and count the objects in the video.

Keywords: Background subtraction, binary, blobs, counting, RGB.

1. INTRODUCTION

Video surveillance has been in used from long ago to monitor the security sensitive areas such as banks, department stores, highways, crowded public places and borders. Many algorithms and different methods have been proposed for this purpose. But the complexity increases with the problems encountered in the video.

Chih-Chang Chen et al. [1] used a dynamic background subtraction module to model light variations and to determine pedestrian objects from a static scene. The background model has been built adaptively using pixel gray level values. Yanling Wang et al. [2] used the background subtraction method to separate background and foreground objects. A new self-adaptive background approximating and updating algorithm is proposed. The background model is updated using a temporal low-pass filter. Osama Masoud et al. [3] developed a new method for tracking and counting pedestrians in real time system using a single camera. The background subtraction method and thresholding is used to produce the differential image. The threshold value is obtained by examining an empty background and measuring the maximum fluctuation of pixel values during this training period. J. L. Raheja et al. [4] proposed a bidirectional people counting algorithm. The Gaussian mixture model is used to describe a pixel of background. Multimodal background is used to provide robust adaptation.
against repetitive motion of scene elements, slow moving objects and introducing or removing objects from the scene. These methods have many computational problems which require large processing time. Some methods have less computational activities but the results obtained are not perfect. These results contain some errors in the output. Thus a simple method is proposed to track and count human in visual surveillance system. Foreground segmentation is the first important step in tracking [5]. This method detects the moving objects by using background subtraction method which detects objects robustly and accurately [6]. Then tracking is performed by extracting some features of blobs. Finally counting is done by using appropriate variables. The Fig.1 below shows the architecture of the proposed system.

Figure 1: Proposed system architecture

2. PROPOSED SYSTEM

The moving object detection is the main task in any visual surveillance system. There are different object detection techniques like background subtraction, frame difference, optical flow etc. In this work, the background subtraction method is used to detect the foreground objects because it is simple to implement and gives high accuracy [7].
2.1. Background Subtraction Method

In this method, every frame from the video clip is subtracted from the background frame. Thus differential image is generated. Then threshold is applied to this differential image to obtain the foreground objects. This method can be applied to gray level image as well as binary image. Thus we performed this method on both the image formats to compare the results. The format with the best output is selected to track and count the people.

First the color image is converted into the gray level format. Then the background model is built by adding first few frames and taking average of that addition [1]. Hence each background pixel can be updated by:

\[
B_m(x, y) = \frac{1}{m} \sum_{i=0}^{m} F_i(x, y) 
\]  

where \( m \) is the index of current frame, \( F_i(x, y) \) denotes the pixel gray level value on \((x, y)\) and \( B_m(x, y) \) denotes the background pixel gray level value calculated from previous frames[1]. Here we get the background image.

Then a differential image is generated by subtracting the background image from the current frame. The threshold value for detecting the foreground regions need to be determined accurately. Thus an initial threshold is set by averaging the pixel values of the differential image. Then means of pixels belonging to the foreground and background regions are separately calculated and denoted as \( \mu_o \) and \( \mu_B \) respectively as:

\[
\mu_o = \frac{\sum_{(i,j) \in \text{objects}} F(i, j)}{\# \text{object pixels}} 
\]

\[
\mu_B = \frac{\sum_{(i,j) \in \text{background}} F(i, j)}{\# \text{background pixels}} 
\]

Then a threshold \( T \) is calculated by:

\[
T = \frac{\mu_o + \mu_B}{2} 
\]

This threshold \( T \) is used to detect the foreground objects from the differential image.

The background subtraction can be applied to binary image format. In this method, the background is assumed to be static, so it does not change with the number of frames. Thus the third frame is considered as a background image. Then remaining frames are checked if their RGB components are same as the background image or not. If they are same then that frame is treated as the background frame. Then a differential image is generated by subtracting the binary background image from the current frame. Simultaneously a threshold is applied to the differential image to detect foreground objects [8].

The difference equation is given by:

\[
D_k(x, y) = |f_k(x, y) - b_k(x, y)| 
\]
Then this difference is thresholded as:

\[
R_k(x,y) = \begin{cases} 
1, & Dk(x,y) > T \\
0, & Dk(x,y) < T 
\end{cases}
\] (6)

Here \(R_k(x,y)\) is the detected foreground object. The threshold used for this detection should be accurate because accuracy of \(T\) has directly impact on the quality of thresholding [9][10].

2.2. Post Processing

By comparing the results obtained from both the approaches, it is clear that the binary image format gives the more accurate results. Thus the results obtained with this approach are taken for tracking and counting of people [11].

It is possible that the image will contain some noise in it even though the human is detected in foreground. This noise should be removed to get the correct output. Thus the morphological operation i.e. dilation operation is performed on the detected objects. Thus the foreground objects eventually become bigger and the holes in the object due to noise will become smaller or disappear.

2.3. Feature Extraction

The group of cells that pertain to a single target is called as a blob. This group of cells can be find out by connected component labeling. The different features related to thus blob are derived to get more information about the detected foreground object [12].

These features includes,

2.3.1. Number of blobs: It indicates the number of objects present in that particular frame.

2.3.2. Size of blob: Size of blob means the total number of pixels for the foreground objects. Only objects having size above a certain threshold (e.g. 200 pixels) are kept for tracking to eliminate small objects.

2.3.3. Centroid of blob: It is equal to the mean of all the foreground pixels composing the whole object [13].

2.3.4. Average color of blob: The average color of a blob can be calculate as:

\[
\text{Avg Color} = \frac{\sum R-Components}{\text{Length of object}} + \frac{\sum G-Components}{\text{Length of object}} + \frac{\sum B-Components}{\text{Length of object}}
\] (7)

By observing these features, we can get idea about the objects in a video clip. These features are very useful in tracking and counting the human in a video scene.
2.4. Tracking and Counting Human

After detecting the foreground objects, the main task is to track these objects in a video scene. The goal of this algorithm is to track the entire deformable object successfully in the video [14]. These objects should be tracked in a complete video without any disturbance.

The main difficulty in tracking is the discontinuity in the detected objects. Due to movement of human in the video, it is possible that some parts of the human will appear separate. Also during background subtraction process, some errors can occurs. Thus legs or hands could appear as the separate objects from the body. It can be also caused by any noise or shadows present in the video. Due to this, the parts of body of a single object could be get tracked separately in a video. Thus error will occur in counting also.

To improve the results in such situation, a separate rectangle has been plotted with object detection. This is done by using the subplot function in matlab. Thus one rectangle will cover whole body of an object detected in background subtraction method. This is performed by initializing two separate variables as sp (starting point) and ep (ending point). Then if-else and while loops are applied to these variables with necessary conditions. Thus each object will get tracked completely in a video clip.

First the non-zero element of the detected objects is found out which indicates the object is present. Then the locations of this non-zero element are estimated. Thus four points indicating minimum and maximum x and y co-ordinates of the object are obtained. Then the condition to locate human object is as given below:

\[
(y_{\text{max}} - y_{\text{min}}) < (x_{\text{max}} - x_{\text{min}})
\]  

(8)

where, 
- \(y_{\text{max}}\) = max coordinate of the object in y-axis
- \(y_{\text{min}}\) = min coordinate of the object in y-axis
- \(x_{\text{max}}\) = max coordinate of the object in x-axis
- \(x_{\text{min}}\) = min coordinate of the object in x-axis

If this condition gets satisfied then a white rectangle around the object is drawn by converting color pixels to white. This rectangle or bounding box of human will remain throughout the video.

Then counting of people is performed by using necessary variables and conditions. These conditions find out the previous and current value of pixel in the detected image. If human object is present that pixels will be represented by 1. Other pixels are represented by 0. Thus a condition such as, if \(\text{pre} = 0\) and \(\text{cur} = 1\), is used in processing and counting can be performed.

3. EXPERIMENTAL RESULTS

In order to analyze robustness and effectiveness of the implemented system, some experimental results under different conditions are carried out on this system. These videos are captured with a Sony cyber shot digital camera with 12 Megapixel resolutions. The size of a frame in the video is 180 X 320.

In first video, three people are present in a single frame. Generally, it is difficult to track the multiple people in a single frame. But this system shows the accurate results and successfully tracks the three people in video. The counting of people is also performed correctly. The total time taken to perform this analysis is 50sec.
The second video contains multiple people but shadows are also present in the video. Also illumination appears at the background due to reflection of light. Due to these disturbances in the video, the error could be appeared in background subtraction process. In figure 3, the image shows the result of this process. The object in a frame is appeared broken in three parts. Thus tracking and counting of people also could be wrong. But this system shows good results with background subtraction process. The tracking is also carried out successfully. The video contains nine numbers of people, but this system shows total count as ten. The total time required to perform this experiment is 60sec. Thus this system shows good results in case of shadows and illumination.

In third video, shadows and illumination are present. But at the same time occlusion of two people is also present. Again due to occlusion and other disturbances, it is possible that an error will occur. During experiment, this system successfully carried out the moving object detection. Then tracking is also performed nearly accurate. While counting it shows some error, but it is near to the perfection. The video contains two people while the result shows the count as three. The total time required for this experiment is 50sec.
There experiments results are tabulated to show the processing time, accuracy of the system and detection error. The results are shown for the three videos shot under different conditions.

Table 1: Accuracy of detection and detection error for different videos

<table>
<thead>
<tr>
<th>Videos</th>
<th>Actual Objects</th>
<th>Detected Objects</th>
<th>Accuracy of detection</th>
<th>Detection Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video-1</td>
<td>3</td>
<td>3</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>Video-2</td>
<td>9</td>
<td>10</td>
<td>90%</td>
<td>0.11</td>
</tr>
<tr>
<td>Video-3</td>
<td>2</td>
<td>3</td>
<td>70%</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The first table shows the actual number of objects, number of detected objects, accuracy of detection and detection ratio. The results show that this system is accurate and detection ratio is also small.

The detection ratio is given by $D$ as:

$$D = \frac{T-C}{T}$$  \hspace{1cm} (9)

where $T$= ground truth (actual number of people)

$C$= Number of detected people

Table 2: Time required for processing of different videos

<table>
<thead>
<tr>
<th>Videos</th>
<th>Time for object detection (Sec)</th>
<th>Time for tracking (Sec)</th>
<th>Time for Counting (Sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video-1</td>
<td>55</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>Video-2</td>
<td>70</td>
<td>85</td>
<td>89</td>
</tr>
<tr>
<td>Video-3</td>
<td>28</td>
<td>35</td>
<td>37</td>
</tr>
</tbody>
</table>

The second table shows the total time required for the analysis of this system. The total time has been divided into time required for people detection, time taken for tracking and time taken for counting of people. These results have shows that this system requires very less time for counting people.

4. CONCLUSION

Moving object detection, tracking of detected objects and counting of these objects are difficult procedures in the presence of shadows and illumination. The proposed system performs well in these scenarios. This system uses background subtraction method to detect the foreground objects. These results are processed for further use. Then different features related to the objects are extracted. Then tracking is performed according to the size of each object and finally total count of objects is shown.

The experimental results show that this method is simple to implement. It is robust and very effective. The tracking and counting is successfully performed in the presence of shadows, illumination and occlusion. This system can achieve 100% result in absence of shadows. Also
it can count multiple numbers of people present in a video. It requires very less time for all
the processing. This system can be efficiently used to monitor people in public places like
banks, hospitals, government offices, shopping malls etc.

REFERENCES

People Tracking and Counting Incorporating Shadow Detection and Removal”, International
2, May 2011.
Surveillance”, School of Computer Science, University of Windsor, Canada.
Detection and Tracking in Videos”, Journal of Aeronautics and Space Technologies, Vol. 4,
[13] Damien Lefloch, Fauzi Cheikh, Jon Hardeberg, Pierre Gouton and Romain Clemente,
“Real-Time People Counting System Using a Single Video Camera”, Real Time Image
[14] Qian Zhang, King Ngan,”Segmentation and Tracking Multiple Objects Under Occlusion