Abstract—Video segmentation is a process of dividing a movie into meaningful segments. It helps in the process of the detection of moving objects within a scene which play a vital role in many application such as Surveillance, Safety, Traffic monitoring and Object detection, etc.. Especially, Background subtraction methods are widely used for moving object detection in videos. In this paper, a new method has been proposed for object detection using background subtraction and thresholding based segmentation algorithms. Experimental results proved that the proposed method achieved high accuracy rate than other existing techniques.

Keywords—Background Subtraction, Threshold based segmentation, Object detection, Video segmentation.

I. INTRODUCTION

Object detection and segmentation of moving objects in video streams is an essential process for information extraction. Surveillance system uses video cameras to monitor the activities of targets (human, vehicle, etc.) in a scene [1]. In order to obtain an automatic motion segmentation algorithm that can work with real images there are several issues that need to be solved, particularly important are: noise, missing data and lack of a priori knowledge. One of the main problems is the presence of noise. For some applications, the noise level can become critical. There are three conventional approaches to moving object detection: background subtraction, temporal differencing and optical flow [2].

Background subtraction is one of the most popular methods for novelty detection in video streams. Background Subtraction generates a foreground mask for every frame. This step is simply performed by subtracting the background image from the current frame. When the background view excluding the foreground objects is available, it becomes obvious that the foreground objects can be obtained by comparing the background image with the current video frame [3]-[4]. Moreover, a background image can be elegantly used to determine the foreground objects by comparing the input frame with the background image and marking the differences as foreground objects. This technique is commonly known as background subtraction or change detection [5].

In segmentation process each image/frame is splitted into a set of non-overlapping uniform connected regions such that any two adjacent ones are not similar. It is a difficult task due to the complexity and diversity of images and moving objects. Influencing factors range from illuminating, contrast and frames. Most of these segmentation algorithms are based on similarity, difference and particularly, can be divided into different categories: threshold, template matching, region growing, edge detection and clustering [6].

A. Related Work

This section presents an overview of existing background subtraction techniques and also analysis the different approaches that provide relevant work for the sudden illumination changes in background subtraction algorithms.

Qiong Wu et al. (2008) described about a design which can automatically digitize synchronized video sequences without the need for further temporal or geometric processing. There are, however, two shortcomings with our hardware design. First, our hardware can automatically recognize the foreground object only if it is within the effective distance of the IR source, and this distance acts like a plane dividing foreground and background. Therefore, the user may need to move the IR source around and find the best position by observing whether the IR image yields a good foreground MASK. Second, if an object appears closer than the foreground it will also be captured [8].

S. Maludrottou et al. (2009) proposed a corner-based background segmentation for real-time applications. The algorithm successfully classifies corners extracted from a video sequence as pertaining to background or foreground after a joint application of Fuzzy ARTMAP neural networks and spatial clustering. The performances have been evaluated using different corner extraction algorithms and clustering techniques [9].

Deepak Kumar Rout et al. (2013) proposed the inter-plane correlation between three consecutive R, G and B planes by using a correlation function. The correlation matrix obtained is then used to construct a segmented image which gives a rough estimate of the object. The segmentation of the correlation plane is done by a threshold. This threshold selection is made adaptive to the video sequence considered. This segmented plane along with the moving edge image is then taken into
consideration to improvise the correct classification of the moving object in the video [10].

Anaswara S Mohan et al. (2014) described two methods for detection and segmentation of moving objects in videos. First method is for object detection using back ground subtraction and second method for segmentation using two approaches i.e. thresholding and edge detection. Simulation results demonstrated that the proposed technique can successfully extract moving objects from various sequences. Sometimes the boundaries of the extracted object are not accurate enough to place them in different scenes, which require a nearly perfect boundary location [11].

Zhou Wei et al. (2015) presented a foreground-background segmentation algorithm for video sequences dealt with slow lighting changes by slowly adapting the values of the Gaussians. It also dealt with multi-modal distributions caused by shadows, swaying branches, and other troublesome features of the real world. By dynamically adjusting the parameters and the number of Gaussian components, the computation cost reduced greatly. Combining two-way matching method based on frame difference thoughts with a series of image filtering methods, the method can extract the moving objects exactly which is superior to the traditional method [12].

The rest of the paper is organized as follows: Section II describes about the proposed method. The performance of the proposed method is demonstrated in Section III. Finally, the conclusion of the proposed method is presented in Section IV with future enhancement.

II. THE PROPOSED METHOD

The proposed method aims at extracting the moving objects in an input image from their background. The method is based on using background subtraction algorithm for separating moving objects from their background. The proposed method has three phases. In first phase the input video is read and it is converted into frames stored in a file. The second phase is segmentation which is a critical step in image analysis, pixel as a unit of observation to working with objects (or parts of objects) in the image, composed of many pixels. If segmentation is done well then, all other stages in image analysis are made simpler. There are two general approaches to segmentation, termed thresholding, edge based methods and region-based methods. In thresholding, pixels are allocated to categories according to the range of values in which a pixel lies.

- The boundaries between adjacent pixels in different categories have been superimposed in white on the original image.
- Finally, region-based segmentation algorithms operate iteratively by grouping together pixels which are neighbours and have similar values and splitting groups of pixels which are dissimilar in value.

The moving object is determined by taking the difference between the background image and the input image. Background subtraction finds moving objects information by subtracting background model. For gray video stream, only intensity (lightness) and for colour video stream, HSI (Hue-Saturation-Intensity) colour space background model is used. The HSI system separates colour information of an image from its intensity information, and has a good capability of representing the colours of human perception. The foreground object is detected by taking the difference of current frame and background as given in equation (1),

\[ |f(i) - b(i)| > T \]  \hspace{1cm} (1)

where, \( f \) is an original frame, \( b \) is a background of an original frame and \( T \) is threshold value.

To update background image that is not fixed and estimated background is just the previous frame. It evidently works only in particular conditions of objects and the frame rate using very sensitive to the threshold,

\[ |f(i) - f(i-1)| > T \]  \hspace{1cm} (2)

The running average of the background image is calculated as follows,

\[ B_{i+1} = \alpha F_i + (1-\alpha)B_i \]  \hspace{1cm} (3)

\( \alpha \) is the learning rate 0.05.

The image is segmented into object and background pixels as describe above creating two sets,

\[ G1 = \{f(m,n) : f(m,n) > T\} \text{(object pixel)} \]  \hspace{1cm} (4)

\[ G2 = \{f(m,n) : f(m,n) \leq T\} \text{(background pixels)} \]  \hspace{1cm} (5)

Where, \( f(m,n) \) is the values of pixel located to the \( m^{th} \) column , \( n^{th} \) row and \( T \) is threshold value.

The average of each set is computed as,

\[ m1 = \text{avg}(G1) \]  \hspace{1cm} (6)

\[ m2 = \text{avg}(G2) \]  \hspace{1cm} (7)

A new threshold is created that is the average of \( m1 \) and \( m2 \)

\[ T = (m1 + m2)/2 \]  \hspace{1cm} (8)

The image is segmented into object and background pixels as describe above creating two sets, \( G1 \) is object pixel and \( G2 \) is background pixel. Average value to be compute the each set and average values to be stored \( m1 \) and \( m2 \). New threshold value found out by calculation the average of \( m1 \) and \( m2 \). Apply the background mask and detect to the object. Finally, the third phase is get the extract object from the original frames and all frames are segmented then stored to the all segmented frames. Finally, all segmented frames convert into new segmented video file.

A. The Proposed algorithm

The Summarization of the proposed method is given below.

Phase I
Step1: Read an input video.
Step2: Convert the video file into frames.

Phase II
Step3: Detect the foreground object by taking the difference of current frame and background.
Step4: Update the background image that is not fixed.
Step5: Segment an image into object and background pixels using as given in equations (4&5).
Step6: Compute the average of each set of frames using as given in equations (6&7).
Step7: Apply the threshold into the average of \( m1 \) and \( m2 \).
Step8: Repeat step 4 and the new threshold is computed in step6 until the new threshold matches the previous solution.

Phase III
Step9: Display and stored the segmented frames.
Step10: Convert the segmented frames into video.
B. Proposed flow diagram

![Diagram](http://www.ijritcc.org)

**Fig1. Process of Proposed Method**

The proposed video segmentation using background subtraction method is shown in fig 1. Graphical representation of this diagram contains ten blocks to perform the operations and get the input video, convert a video to frame, compare to pixel, find the frame difference, apply threshold value, apply to background mask, detect the object, segmented frame and segmented video.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, the experimental result of the performance of the proposed method is evaluated using various evaluation metrics such as, TP, TN, TPR, TNR, DICE, Accuracy, MAE and Jaccard distance is presented. Performance metrics of TP is true positive, FP is false positive, TN is true negative and FN is false negative are using the following equations,

A. True Positive and True Negative:

The true positives(TP) value is the number of the corners correctly assigned to the foreground, the true negatives(TN) value is the number of the corners correctly detected as background.

\[
TP = \frac{TP}{TP+FP} \quad (9)
\]

\[
TN = \frac{TN}{TN+FN} \quad (10)
\]

B. False Positive and False Negative:

The false positives(FP) and the false negatives(FN) metrics represent the number of corners incorrectly labelled as, respectively, foreground and background.

\[
FP = \frac{FP}{TP+FP} \quad (11)
\]

\[
FN = \frac{FN}{FN+TN} \quad (12)
\]

C. True Positive Rate:

Sensitivity or true positive rate or recall is the percentage of true positive pixels and its formula is,

\[
TPR = \frac{TP}{P} = \frac{TP}{TP+FN} \quad (13)
\]

D. True Negative Rate:

Specificity or False Positive rate or fallout is the percentage of false positive pixels and its formula is,

\[
TNR = \frac{TN}{N} = \frac{TN}{TN+FP} \quad (14)
\]

E. Dice Co-efficient:

Dice co-efficient is a similarity measure mostly used to processing performance of segmentation algorithms which has a predefined ground truth information. It is calculated using the formula,

\[
DICE = \frac{2 * TP}{(FP + TP) * (TP + FN)} \quad (15)
\]

F. Accuracy:

Accuracy is the percentage of correct data retrieval. It is calculated by dividing the number of pixel with true positive plus true negative pixel over the total number of pixels in the frames. The following equation displays the calculation of accuracy,

\[
Accuracy = \frac{TP+FN}{P+N} \quad (16)
\]

G. Mean Absolute Error:

The mean absolute error (MAE) is used to measure how close forecasts or predictions are to the eventual outcomes.

\[
Mean \text{ Absolute Error} = \frac{1}{n} \sum_{i=1}^{n} |e_i| \quad (17)
\]

The mean absolute error is an average of the absolute errors \(|e_i| = |f_i - y_i|\)

where, \(f_i\) is the prediction and \(y_i\) is the true value.

H. Jaccard-coefficient:

Jaccard Co-efficient is used to calculate the similarity between the two set of images and it also measures the variation or dissimilarity between two images. The Jaccard index, also known as Intersection over Union and the Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets. This distance is a metric on the collection of all finite sets.

\[
\text{Jaccard coefficient} = \frac{A \cap B}{A \cup B} \quad (18)
\]

\[
\text{Jaccard dist} = 1 - \text{J}(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \quad (19)
\]

where ‘A’ is the non zero pixel element in ground truth image and ‘B’ is the non zero pixel element is segmented image.

In proposed method, detected targets are accurate and the recall is higher than other algorithms. Table I demonstrates the performance of the proposed method based on the standard
segmentation metrics discussed above. Table 1 demonstrates the higher performance of the moving object detection and segmentation using background subtraction for the Man.avi video file which gives higher values on Accuracy(0.9723), DICE(2.5963), TNR(0.9476), TPR(0.9447), MAE(0.0313), Jaccard_dist(-1.9028) and Funny.avi video file for lower values on Accuracy(0.8933), DICE(2.0234), TNR(0.8242), TPR(0.7867), MAE(1.9390), Jaccard_dist(1.6796). It could be observed that the good result for Man.avi files when compared to others.

**TABLE I. PERFORMANCE ANALYSIS OF THE PROPOSED METHOD**

<table>
<thead>
<tr>
<th>Performance Metrics</th>
<th>Input Videos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Funny.avi</td>
</tr>
<tr>
<td>TP</td>
<td>0.7776</td>
</tr>
<tr>
<td>TN</td>
<td>0.9885</td>
</tr>
<tr>
<td>TPR</td>
<td>0.7867</td>
</tr>
<tr>
<td>TNR</td>
<td>0.8242</td>
</tr>
<tr>
<td>DICE</td>
<td>2.0234</td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.8933</td>
</tr>
<tr>
<td>MAE</td>
<td>1.9390</td>
</tr>
<tr>
<td>Jaccard_Distance</td>
<td>-1.6796</td>
</tr>
</tbody>
</table>

**TABLE II. COMPARATIVE ANALYSIS OF EXISTING METHODS AND PROPOSED METHOD**

<table>
<thead>
<tr>
<th>Methods</th>
<th>Performance Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TNR</td>
</tr>
<tr>
<td>Robust Foreground detection</td>
<td>0.81</td>
</tr>
<tr>
<td>A Moving Object detection</td>
<td>0.85</td>
</tr>
<tr>
<td>A Foreground-background Segmentation</td>
<td>0.89</td>
</tr>
<tr>
<td>Scene Analysis for Object Detection</td>
<td>0.9288</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>0.9476</td>
</tr>
</tbody>
</table>

From the observation in Table 2, the proposed method gives a high Specificity (0.9476), Sensitivity (0.9447) with low True Positive (0.7227). Despite Specificity, Sensitivity and true positive of other methods are equally high as shown in Table 2. The standard values for Sensitivity (1), Specificity (1) and True Positive (1) [13]. The proposed method achieves the Sensitivity value (0.94), Specificity value (0.94) and True Positive value (0.72). It gives a segmentation result in high quality of the images and videos, which is used to detect the extract Object from image and clearly the background subtraction.

Visual representation of the proposed method has been tested using three different videos like Man.avi, funny.avi and Dog.avi. The proposed method has achieved by detecting moving objects accurately in all video sequences and the segmentation results are shown below fig (2).

**REFERENCES**


