COUNTING OF WAGONS IN NIGHT VIDEO BASED ON BACKGROUND SUBTRACTION

[1]Pritam Sirpotdar, [2]Anuja Nerkar, [3]Ajitkumar Deshmukh, [4]Prof. Dr. M. A. Joshi
[1,2,3]E&TC Department COEP, [4]Fellow IET COEP
Email:[1]pritamsir@yahoo.co.in, [2]anujanerkar@gmail.com, [3]ajitdeshmukh3@live.com, [4]maj.extc@coep.ac.in

Abstract— A novel approach for train wagon counting during night time is proposed in this paper. Different from traditional background subtraction and segmentation algorithms, this method subtracts only central reference line pixel values from the pixel values of the reference background at the same location. This increases the processing speed of the algorithm as well as efficiency of the system. The RGB pixel values are considered for background subtraction. The pixel difference values determine the segmentation of background and the foreground objects, which here in this case are railway wagons.

Index Terms—Segmentation, Background Subtraction, Object counting.

I. INTRODUCTION

In most of the computer vision related applications, detection and identification of moving object from a given video is critical task. Most common approaches include background subtraction method which is used in number of applications like surveillance systems, remote sensing, in biomedical fields, etc for monitoring and tracking purpose. It is mostly based on the prior background model and comparison of video frames with background model is carried out. Any deviation from background model is considered as the foreground object. Advantage of such types of method is that it is not dependent on the prior knowledge about shape and size of objects. Most of the background subtraction methods are sensitive to illumination changes. In many cases background is not stationary. It may change with time during a day or due to objects in background like trees, mountains or buildings. Lots of work is done in this area. Early approaches are based on Gaussians mixture model which depends on the distribution of pixels. Model stability and convergence problems were seen in such methods. Dar-Shyang Lee [1] proposed effective Gaussian Mixture Learning method which improves convergence rate and stability up to some extent. Memory consumption and processing time, these are some concerns related with the mixtures of Gaussian methods which may leads to errors. Probability based Robust Object Segmentation System for Background Extraction is developed by Chung-Cheng Chiu, Min-Yu Ku, and Li-Wey Liang [2]. Their algorithm detects objects under different illumination changes with reduction in memory consumption and minimizing processing time with further scope for including object tracking and recognition. Yumiba, Miyoshi, Fujiyoshi [3] developed an algorithm which is based on Spatio-Temporal Texture. Their method covers a dynamic background change and is based on *"Space-Time Patch features"*. Their method is robust against global and local background change and it is based on only detecting presence or absence of object but it doesn't reveal any information related to objects like position or size. To detect objects under sudden illumination changes, Cheng, Huang, and Ruan [4] proposed a system in which they have used three modules- background model module, illumination evaluation module and object detection module. Their algorithm is compared with approaches like MTD, SSD and MSDE. Their method calculates adaptive threshold for object segmentation with future scope for improving adaptive threshold for better detection. Another approach uses foreground adaptive background subtraction method where McHugh, Konrad, Saligrama [5] have used nonparametric background model along with foreground model based on small spatial neighborhood to improve sensitivity. They have also used Markov model to change labels to improve spatial coherence of the detections. This is mainly done in order to improve threshold of detection. However, proper initial labeling is needed for proper estimation. Recent approach includes work done by Shih-Chia Huang [6] for automatic surveillance system. Their extensive works based on detection of moving objects by using background modeling (BM) module, an alarm trigger (AT) module, and an object extraction (OE) module. For our proposed BM module, a unique two-phase background matching procedure is performed, AT module saves the effort by eliminating unnecessary examination of the entire background region, the OE module forms the binary object detection mask for detection of moving objects.

In our proposed method, object of interest is wagons i.e we are interested in counting no of moving wagons from a given video frames. Our aim is to track and count no moving wagons under different illumination condition especially during night time using background subtraction method. Most of the present algorithms useful for detecting moving object during day time. They are very much robust for detecting object when there is sufficient illumination but their performance is not very much accurate when there is insufficient light. During night time, background and object of interest both appears dark which leads to segmentation problem. Our work basically concentrated on this problem. Challenge is to develop such system to determine threshold which gives better results under circumstances mentioned above.

Paper is organized as follows: Following sections gives details about proposed algorithm along with results. Conclusion and future scope is given at last section followed by the references.

II. PROPOSED METHOD

A. Algorithm

The paper describes here an automated train monitoring system. To count the number of wagons in a freight train, an automated vision system is developed here. The train video of moving freight train from side view is captured as input to the system, which is further processed by using image processing techniques. The videos captured are during night and contain different imaging conditions i.e. cloudy, rainy etc. Also the background of the video can be plain or can contain clutter as well as it can be varying with time. The main challenges to the development of the system include illumination changes, varying background conditions and real time constraints on processing.

The method proposed in this paper considers only night time videos. The videos used for experiments consist of train wagons moving in backward or forward direction. The train speed changes abruptly. The train sometimes halts for any random time period. The background in the videos contain non-stationary objects such as tree leaves. Swaying of the tree leaves due to breeze also affects the RGB values of the background, thus indicating change in the foreground. Some of the videos contain noise due to improper illumination source and low resolution. The proposed algorithm is robust and addresses all the above mentioned environmental conditions and abrupt change in the train movements. The proposed algorithm is divided into five steps; first is video acquisition, second is background reference frame extraction. The third step is Central reference line detection. Next steps are RGB value extraction then background subtraction and the last step consists of wagon counting. The proposed method is explained in the following sub-section:

Video Acquisition:

First step is video acquisition. The videos containing the side view of railway wagons during night time is acquired first. The proposed method can be implemented for real time video processing. For experimental purpose, the pre-recorded video files are used in this paper. The acquired video sequence is first loaded into the database. The video file consists of consecutive image frames. As the system starts video acquisition, the width and height of the image frame are stored for scanning the frame. The image frames from the video sequence are then extracted and saved on the system. For each consecutive image frame, and for the total no. of rows and columns, the entire image frame is scanned and the pixel intensity values are stored in the database for further processing.

Background Reference Frame extraction:

In conventional background subtraction methods, the first image frame of the video is saved as the background. Alternatively, to make the system robust to dynamic or time varying background, first few image frames from the video are saved and then the average of all the pixels of the image can be taken and is considered as the reference background image. In the proposed method, the night time videos used as the database have poor illumination conditions. Thus, the background of the video does not change as compared to the foreground. As mentioned in the previous sub-section, the night time video contains the dynamic background like swaying tree leaves. To address the problem of such time-varying background, first 50 image frames of the video are acquired first and then for the total no. of rows and columns, these pixel intensity values at each pixel is stored. For every pixel location in the image frame, the summation of the intensity values is taken and then average of this addition is saved in the respective pixel location. Thus, a new image is obtained containing the average intensity pixel values of the first 50 frames. The new background reference image is obtained by the following averaging operation:

$$f'(i,j) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left[\left(\sum_{k=0}^{P-1} f(i,j) \right) \Big/ P \right] \quad (1)$$

This new image is taken as background reference image frame for further processing. The size of the reference frame is saved for scanning the image frames.

Central reference line detection:

In previously proposed background subtraction methods, the consecutive image frames are subtracted from the previous image frames or from the reference image frame. This increases the processing time, as to subtract entire consecutive image frame from the reference frame, we need to scan the entire image for total height and width of the image frame and then perform the image subtraction pixel-by-pixel. In our proposed method, instead of considering the entire image frame, only the pixels belonging to central reference line of the image are scanned for any foreground change. The main purpose of the algorithm is to find the gap between the consecutive wagons and then count the no. of wagons. The central reference line of the image frame scans the areas of railway wagons passing in the video. The value of the width of the background reference image frame is used for scanning the image horizontally and the value of the height of the reference image frame is used to get the height of the central reference line, as the central reference line is obtained by dividing the height by two. This central reference line is then used in the next step for RGB value extraction process.

RGB Value Extraction Process:

In the next step, the RGB values of the background reference image for the central reference line are stored in the class and then decomposed into three different R, G and B values for background subtraction. As the consecutive video frames appear and the railway wagons start to appear, there is change in the RGB value of the foreground. For each consecutive frame, the RGB value only for the central reference line is stored into a class. For each new frame, the RGB values are then decomposed into separate R, G and B values from the class. Here ends the RGB value extraction process.

Background Subtraction:

In previously proposed methods, first the binarization of the image frames is performed and then the new image frame pixel intensities are subtracted from those of the background frame. We have considered the RGB values of the video frames for background subtraction. In binarization, thresholding technique is used to assign the binary value to the pixel intensity. If the RGB value of the image frame is above certain threshold then it is assigned as 0 or 1. In this process of binarization, due to thresholding, the pixel intensity information may get lost. The pixel intensity values play vital role because of poor illumination conditions. To avoid it, in our paper we have considered only RGB values. The R, G and B values of the reference background frame are then subtracted from the respective R, G and B values of the consecutive video frames and then calculated differences are saved as R, G and B difference values. Even though the background remains considerably static, there may exist some noise or clutter in the background or there can be some illumination change due to light source. To tackle such scenarios, the threshold is set for the R, G and B difference values. If the difference values fall within that range, then the change in the foreground is considered as negligible and can be ignored.

This operation is performed as follows:

$$f'(i,j)(t+1) = \begin{cases} f(i,j)(t) & if\ |diff| < th \\ f(i,j)(t+1) & otherwise \end{cases} \quad (2)$$

Where 'th' stands for threshold value and the value of diff is calculated as:

$$diff = f(i,j)(t+1) - f(i,j)(t) \quad (3)$$

Wagon counting algorithm:

This sub-section explains the wagon counting algorithm used in the proposed method. Two separate difference values are stored from the background subtraction; one for the gap between the wagons and the other for the length of the wagon. The purpose behind two separate difference values is to evaluate the length of the gap between the wagons and to check the length of the wagon itself. Then on the basis of these lengths, the wagons are counted. If the difference in the background and foreground frame falls within the threshold, then the counter for the difference value for the gap is increased. And if the difference value is greater than the threshold, it indicates that there is change in the foreground and the train wagon has arrived. Such change in the foreground pixel is indicated by the yellow color on the central reference line. Also, the counter for the difference value for the length of the wagon is increased. Every time the counter for gap and wagon length is increased, the counters are checked for another threshold values. Once the counter for gap is reached, the next wagon is counted and added to the no. of bogies counted previously. Also, the gap counter is set to zero again. Similarly, once the counter for entire wagon length is reached, it is again set to zero to check for next gap between the bogies. For the night time videos, the main challenge is to deal with the background illumination. During night time, to illuminate the objects, the only source is the light source used along with the camera. Unlike for day time videos, the natural day light illuminates the foreground objects as well as the background. While processing the day time videos, the background may vary with time. Thus after extraction of first few frames, if the background varies, then this variation in background may be considered as change in the foreground, finally affecting the results. Thus, the algorithm considering only first few frames for the background subtraction fails in such cases. The same problem is encountered in the night time videos, where the light source illuminates the background as well as foreground properly. In such cases, if the RGB values of the foreground objects resemble with those of the background and fall within the pre-decided threshold, then there are chances of the foreground objects getting counted as the background. To address this problem, the reference background can be updated after pre-decided no. of frames.

III. EXPERIMENTS AND RESULTS

The video database of night time contains the side view of the railway wagons going in forward or backward direction with varying frame sizes. The resolutions of the frames are 1280 x 960, 1280 x 800. The frame rates are 5.03 FPS, 10 FPS. The pre-recorded videos are in .mp4 and .avi formats. The distance of the camera from train is varying in all videos. It is approximately 15-20 feet. The height of the camera is 10-15 feet from ground. The videos are captured with different background and environmental conditions with the light source for illumination purpose. Following table shows the details of the video database:

Video Name	Resolution
CTPS_Night_Forward_Stream	1280*960
CTPS_Night_Reverse_Stream2	1280*960
Panvel_NightStream1	1280 x 800
Panvel_NightStream2	1280 x 800

The quality of the results is checked by structural similarity (SSIM), peak signal to noise ratio (PSNR) and root mean square error (RMSE) values.
The quality of the image frames are checked by SSIM index. The SSIM checks for the structural similarity of the distortion free images with the noisy images. It is further extension of the traditional quality index measurement techniques i.e. PSNR and RMSE. The PSNR value calculates peak signal to noise ratio of the image in dBs. The high value of PSNR indicates the good quality of image frame, ideal value is infinity. As seen from the results, we get PSNR value in the range of 60dB. The RMSE assesses the quality of the resultant images. The smaller RMSE value better is the quality and we get very small range of RMSE values.
The SSIM is calculated on the basis of formula:

$$SSIM(i,j) = \frac{(2\mu_i\mu_j + c_1)(2\sigma_{ij} + c_2)}{(\mu_i^2 + \mu_j^2 + c_1)(\mu_i^2 + \mu_j^2 + c_2)} \tag{4}$$

Where μ represents the average value and σ represents the variance value and c is the factor for stabilization.
The RMSE is calculated as follows:

$$MSE = \frac{\sum\left(IMG_{orig} - IMG_{noisy}\right)^2}{m}$$

$$RMSE = \sqrt{MSE} \tag{5}$$

where the MSE is calculated by subtracting noisy image from original image and then normalizing it by factor 'm'. The RMSE value is calculated by taking square root of MSE.

The PSNR is calculated by following formula:

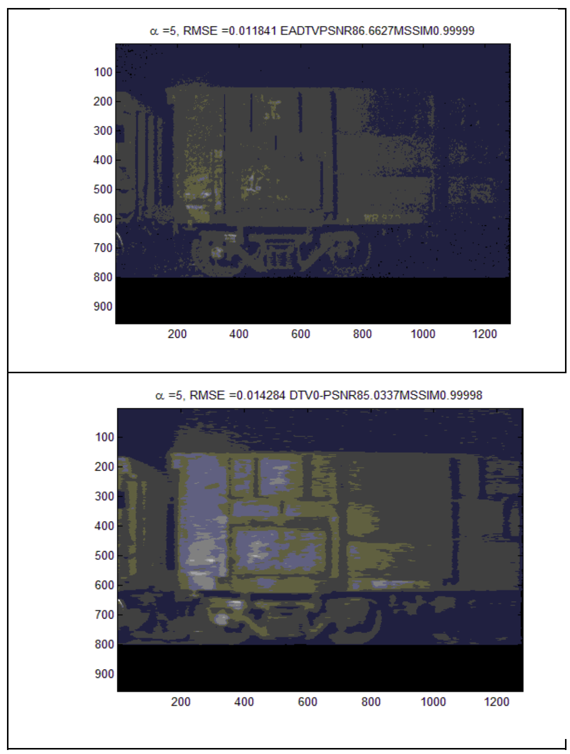
$$PSNR = 20\,log_{10}\left(\frac{MAX}{RMSE}\right) \tag{6}$$

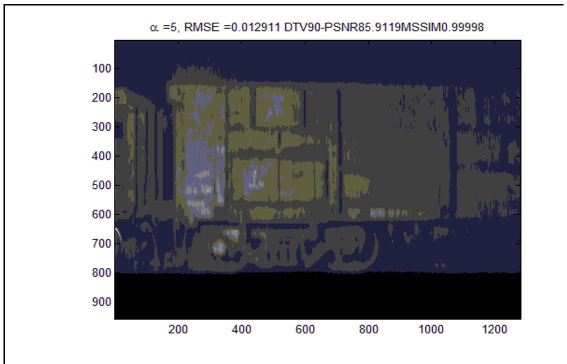
where Max represents the maximum illumination pixel intensity.

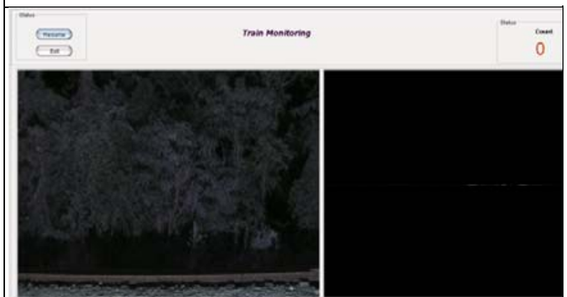
After applying the proposed algorithm, we are able to segment the railway wagons from the background and count them successfully and correctly. The quality index values of RMSE, SSIM and PSNR for the used night time video are tabulated below:

Video Name	RMSE	SSIM	PSNR
CTPS_Night_Forward_Stream1	0.0118	0.9999	86.66db

The results are as shown in the figure below:







Swaying tree leaves detected as foreground

Counted train wagons and foreground detection

IV. CONCLUSION AND FUTURE WORK

We have proposed and implemented video based automated system for train monitoring for counting the number of wagons of the train. This work presents a novel approach in detecting and counting the railroad cars in continuous train video images, which were recorded using a camera.

The proposed algorithm gives good results for the noisy night videos of railway wagons. The values of quality index parameters indicate same. The processing time for the algorithm is very fast as compared to traditional background subtraction algorithms. The same algorithm can be applied to the day time videos by adding the adaptive threshold algorithm. According to the illumination present in the video frames, the adaptive threshold is changed and the same algorithm can be implemented for both day and night time videos. This can be implemented as the extension of the method proposed in this paper.

REFERENCES

[1] Dar-Shyang Lee, "Effective Gaussian Mixture Learning for Video Background Subtraction", IEEE Transactions on pattern analysis and machine intelligence, vol. 27, no. 5, pp.827 - 832 May 2005

[2] Chung-Cheng Chiu, Min-Yu Ku, and Li-Wey Liang, "A Robust Object Segmentation System Using a Probability-Based Background Extraction Algorithm", IEEE Transactions on circuits and systems for video technology, vol. 20, no. 4, pp. 518 - 528 April 2010

[3] R.Yumiba, Miyoshi, M. Fujiyoshi, "Moving object detection with background model based on spatio-temporal texture", Applications of Computer Vision (WACV), 2011 IEEE Workshop, pp. 352 – 359, Jan. 2011

[4] Fan-Chieh Cheng, Shih-Chia Huang, Shanq-Jang Ruan , "Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection", IEEE Transactions on broadcasting, vol. 57, no. 4, pp. 794 – 801,December 2011

[5] J.Mike McHugh, Janusz Konrad, Venkatesh Saligrama,Pierre-Marc Jodoin, "Foreground-Adaptive Background Subtraction", IEEE Signal Processing letters, vol. 16, no. 5, pp. 390 - 393 May 2009.

[6] Shih-Chia Huang, "An Advanced Motion Detection Algorithm with Video Quality Analysis for Video Surveillance Systems", IEEE Transactions on circuits and systems for video technology, vol. 21, no. 1, pp.1 - 14 ,January 2011

[7] Qing-Jie Kong, Avinash Kumar, Narendra Ahuja, and Yuncai Liuy, "Robust Segmentation of Freight Containers in Train Monitoring Videos", 2009 IEEE.