Hussain Khan, S., Haroon Yousaf, M., Murtaza, F. y Velastin, S. A. (2019). Passenger Detection and Counting during Getting on and off from Public Transport Systems. *NED University Journal of Research,* 2(17), pp. 35-46.

DOI: https://doi.org/10.35453/NEDJR-ASCN-2019-0016

PASSENGER DETECTION AND COUNTING FOR PUBLIC TRANSPORT SYSTEM

Saddam Hussain Khan[1], Muhammad Haroon Yousaf[2], Fiza Murtaza[3], Sergio Velastin[4]

ABSTRACT

Implementing accurate and reliable passenger detection and counting system is an important task for the correct distribution of available transport system. The aim of this paper is to develop an accurate computer vision-based system to track and count passengers. The proposed passenger detection system incorporates the ideas of well-established detection techniques and is optimally customised for both indoor and outdoor scenarios. The candidate foreground regions (inside an image) are extracted in the proposed method and are described using the histograms of oriented gradient descriptor. These features are trained and tested using support vector machine classifier and the detected passengers are tracked using a filter. The proposed counting system is used to count passengers automatically when they pass through a virtual line of interest. Accuracies ranging 91.2 percent to 86.24 percent were found for passenger detection using the proposed passenger detection and counting system whereas relative counting errors varied ten percent to thirteen percent.

Keywords: computer vision; Gaussian mixture model; histogram of oriented gradients; line of interest; region of interest; detection and counting system.

1. INTRODUCTION

Passenger counting and detection is an important task for the traffic monitoring and utilisation of resources for a public transport system. It is important to monitor public traffic efficiently for a well-organised and cost-effective public transport system. Public transport companies provide information on the connecting routes which plays an essential role in the traffic monitoring system. The allocation of public vehicles to the various connection routes during normal and busy hours is essential for the management of a transport system. These connecting routes vary from time to time in different places. Statistical tests may require for getting a better estimation of the average number of vehicles needed for a particular route.

Traffic overload is more significant in the daytime as compared to night. More public vehicles are required on a route during busy hours. The correct distribution of available resources or vehicles over different routes is essential for better optimisation and management of a transport network. This requires continuous monitoring of the transport system such as counting the passengers during getting off and on the public transport. The analysis of this data to count the passengers can be carried out using a computer vision-based approach to avoid use of human efforts.

---

[1] PhD student, Computer Engineering Department, University of Engineering and Technology, Taxila, Pakistan, Ph. +92(0)51 924 8611, Fax: +92(0)51 9047406, Email: hengrshkhan822@gmail.com.

[2] Associate Professor, Computer Engineering Department, University of Engineering and Technology, Taxila, Pakistan, Email: haroon.yousaf@uettaxila.edu.pk.

[3] Assistant Professor, Computer Science Department, Women University Swabi, Pakistan, Ph. +92(0)938-221193, Email: fizamurtaza@yahoo.com.


Saddam Hussain Khan is a PhD student at Pakistan Institute of Engineering & Applied Sciences, Pakistan. He received his Bachelors from University of Engineering and Technology, Peshawar, Pakistan and Masters from University of Engineering and Technology, Taxila, Pakistan, respectively, in 2014 and 2017. His research interests include computer vision, deep learning and medical image processing.


Muhammad Haroon Yousaf is an Associate Professor in Computer Engineering Department at University of Engineering and Technology, Taxila, Pakistan. He is Senior Member of IEEE and a member of IEEE Signal Processing Society. He is an author of several research papers in international conferences and journals. His research interests include image processing and computer vision.


Fiza Murtaza is an Assistant Professor in Computer Science Department at Women University Swabi, Pakistan. She received her Bachelors from COMSATS Institute of Information Technology, Abbottabad, Pakistan in 2013. She received her Masters and PhD from University of Engineering and Technology, Taxila, Pakistan, respectively, in 2015 and 2019. Her research interests include artificial intelligence, computer vision, machine learning, human action recognition and temporal human action detection.


Sergio Velastin is Senior Research Scientist at Cortexica Vision Systems, UK. He is a Fellow of IET and Senior Member of IEEE where he was an elected member of the Board of Governors of the Intelligent Transportation Society (IEEE-ITSS). He is also an associate editor of the IET Computer Vision journal.

Developing reliable passenger detection and counting system is a critical issue. Counting passengers under a simple and controlled environment is relatively easy as compared to a complex environment. This is due to the fact that the density of passengers is high in a complex environment and realistic environmental conditions (such as high illumination, shadows, occlusions, complex interactions, lighting fluctuations, darkness, passenger viewpoint changes, environmental changes, and cluttered environment) may vary.

Two most reliable approaches commonly used for counting people include light and infrared directional sensor [1] and computer vision-based image and video sensing. The former approach is used for counting the passengers during 'Getting ON and OFF' from the public transport system. When the passenger crosses the sensor line, the light/infrared beam is broken and indicates the presence of a person. Counting people by using a sensor is advantageous in respect of resolution and cost. Other significant advantages include reduced size of network, easy installation and reliability. Nevertheless, the reliability of the passenger counting system through infrared sensors decreases in crowded situations due to variations in temperature, dust and smoke, and their high sensitivity to noise. To overcome this problem, visible light is used for counting purposes because visible light is less affected by electronic smog and provides good accuracy. One major deficiency of infrared or light sensors is that they are unable to differentiate between humans and non-humans. Therefore, a non-human object (such as an opaque body) will also be counted as human it is crosses the infrared beam. Similarly, these sensors are unable to count a group of people passing together through the light beam. Moreover, this system cannot easily determine the moving direction of the passengers while 'Getting ON or OFF'. These shortcomings of light or an infrared sensor can be addressed by analysing visual data using computer vision-based approaches.

Similarly, it will be difficult to detect passages accurately in an occluded environment when two or more persons cross a line within the same time or when two persons cross a line one after another without separation. Although a solution to these problems is to reset the time, the accurate selection of reset time is difficult. A decision in this case is needed on when the next person is to be counted. Although time depends on the moving velocity of the person, the slow-moving person will be counted multiple times when the time duration is small. The overlapping problem needs to be handled by segmentation. Foreground segmentation may give inaccurate results (such as false positives) due to overlapping and occlusion problems.

A computer vision-based method has been proposed in this paper to detect and count passengers during getting in and getting out of the train or bus. The PAMELA dataset [2] was used for the validation of the proposed passenger detection and counting system. This dataset contains real-time videos having illumination, overlapping, passenger orientation, complex interaction, shadow, darkness, occlusion and distortion present in images due to the curvature of the camera. The proposed detection method was compared with a pre-trained classifier trained on histograms of oriented gradients (HOG) and Haar [3, 4] type features. The results showed that the proposed detection method performs better as compared to the pre-trained classifiers.

2. BACKGROUND

Several techniques have been in use for counting persons in different environments and scenarios. These techniques can be divided into the following four categories: (a) edge analysis based techniques; (b) spatiotemporal techniques; (c) motion detection based techniques; (d) model-based techniques.

Edge analysis methods [5-7] have been used for tracking and detection of objects. The edge follows a particular shape. For example, human head corresponds to a circular shape. An edge filter (such as a Canny edge filter [8]), is applied to the object to extract edge information. Edge detection is important for the extraction of the skeleton of the silhouette. The main problem related to skeleton computation is its sensitivity to noise, although the noise from silhouettes can be reduced using smoothing filters.

Spatiotemporal methods [9, 10] are based on the selection of a line from the region of interest. The statistical model techniques are applied to estimate the number of passengers crossing the line. Although these techniques are fast, simple and easy to implement, these algorithms fail (in some cases) to provide accurate results for passengers counting, as the foreground generated from a stationary person is considered as multiple detected people in these techniques.

Motion detection-based techniques [11-13] follow two steps. Firstly, moving regions of objects and humans are detected in a video. Information about the trajectories and direction of a moving person are provided in the next step. These techniques recognise and count people who cross a virtual line. Trajectory and clustering algorithms can count passengers. Features of passengers are extracted and (based on these features) passenger direction and estimations are monitored. Due to complex environmental conditions (such as overlapping among passengers), the position of trajectories becomes complex and the accuracy of clustering is affected. The trajectories may include invalid trajectories, such as pixel motion of background of passengers. The selection and removal of these types of invalid trajectories is a difficult task. In addition, the detection system counts two persons as a single person when two or more passengers follow the same trajectory.

Finally, model-based techniques have implemented by the researchers [14-16] to find the region of interest of the images using prior knowledge. Prior knowledge is based on training a classifier or template matching. These types of algorithms require a learning database or a problem of model generalisation which limit their application. They may also be computationally complex in terms of time and resources. Passengers may be counted by seat occupation. Seat and passenger model-based detection and counting algorithms are also implemented in references [16, 17]. The use of pressure sensors has also been tried to count passengers which is expensive due to high hardware cost and pressure sensors [18]. This method is not unable to detect passengers if any person or object presses the seat. Counting people using face detection is another approach to detect passengers. This is a challenging problem as the face might not be detected if a human face is covered with a mask or hair. Nevertheless, model-based techniques are reliable and accurate for passenger detection, provided enough data for training such models are available.

The main problems faced by video processing-based approaches are occlusion, illuminations, overlapping, and shadow effects. The occlusion problem was minimised by Schousek [19] by using a fixed camera mounted overhead and looking straight down. Other approaches [21, 22] use stereo images captured by a pair of cameras to cope with occlusion and overlapping. Kilambi et al. [23] proposed a rectangular model-based detection in which a single camera is mounted on the ceiling of the door. A system to detect passengers was proposed by Kim et al. [24] which keeps tracks of moving passenger directions by tracking a bounding box.

Several pre-processing steps are required for passenger detection and counting systems, such as foreground/background subtraction. Illumination and colour similarity between the foreground and background are considered sensitive problems in this respect. If the background does not change from one video frame to another frame and the foreground is moving, simple frame difference computes moving foreground objects accurately. However, foreground/background subtraction algorithms do not give the appropriate result if the foreground objects are static. The adaptive background model for motion detection has been used in the literature to overcome this problem [25, 26]. A high frequency of human activity is also problematic for such kind of background modelling. The adaptive background can be affected by the problem of shadows. Another popular motion detection technique is recursive background modelling to handle frequent illumination [27, 28]. It works on the running average, although it tends to fail for sudden illumination changes.

The rest of the paper is organised as follows: section 0 presents the technical description of the proposed analytical techniques and approaches. Section 4 provides details of experimental results and comparative analysis. Finally, conclusions and recommendations are given in section 0.

3. PROPOSED MODEL

The proposed detection system is based on the existing detection algorithms which are incorporated to design a robust passenger detection system. The detected passengers are automatically counted using the system as they pass through a virtual line of interest (LOI). The methodology of the proposed method is based on the following four main steps: (a) pre-processing; (b) passenger detection; (c) passenger tracking; and (d) passenger counting. The aforementioned steps are shown in **Figure 1**. The detail of each step is provided in the forthcoming sections.

3.1 Pre-processing

Pre-processing is carried out in three steps. Region of interest (ROI) is selected from the input videos in the first step. Red, green and blue (RGB) video frames are converted to hue, saturation and value (HSV) colour space in the second step. Finally, the contrast of the video frames is enhanced using histogram equalisation to improve the visual quality of the frames.

3.1.1 Colour conversion

The input videos are composed of the RGB colour model. Since RGB colour space is the correlation of three-colour components, it can be disadvantageous to separate out the colour information. If illumination variations affect one layer, then the remaining layers will also be changed which causes instability. To overcome this problem of uncertainty that is caused by RGB colour space, HSV model is used [29, 30]. Moreover, RGB frames are highly sensitive to shadows which can create hurdles for the detection process. Therefore, these input video frames are transformed into HSV colour space (**Figure 2**) to separate the luminosity and chromaticity. This step is important for removing the illumination variations from video frames and to distinguish brightness variations from chromaticity. Only one grey channel (rich information channel) is supposed after colour conversion from RGB to HSV for further processes such as histogram equalisation.

3.1.2 Histogram equalisation

Histograms of digital images represent the total distribution of pixel values in images. Histogram equalisation is a widely used method in the field of object detection to normalise the illumination. The method has been applied in the presented paper for enhancing the contrast of video frames.

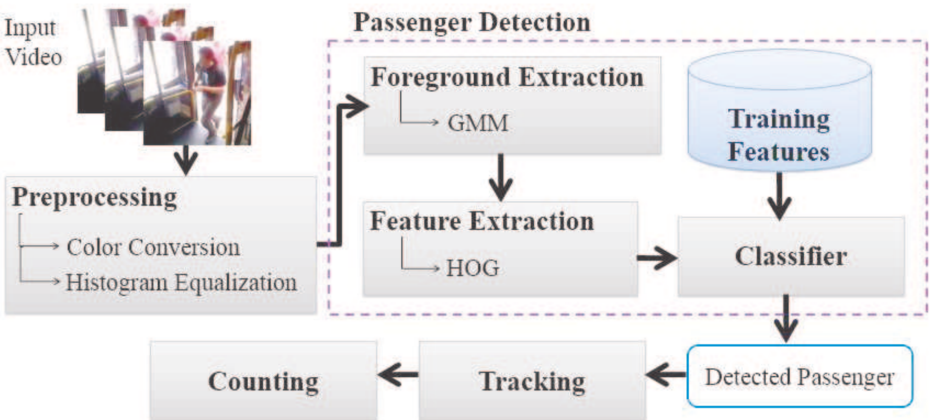

Figure 1. Proposed framework of passenger detection and counting system.

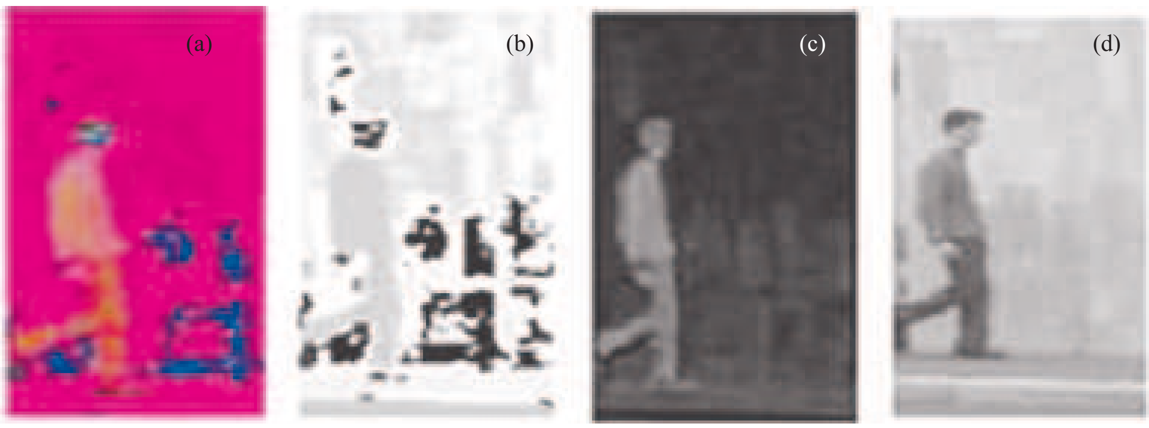

Figure 2. Transformation of input video frames: (a) breakdown of hue, saturation and value layer of HSV frame; (b) hue layer; (c) saturation layer; (d) value layer [28, 29].

3.2 Passenger Detection

The following three steps were performed for passenger detection in the presented paper: foreground extraction, feature extraction and classification. The detail of each step is provided in the forthcoming sections.

3.2.1 Foreground extraction

The Gaussian mixture model (GMM) has been used in the presented paper for foreground extraction. This model plays an important role in managing illumination effects due to its updating capability. Any change in light (in the static scene) can be tracked by using GMM as Gaussian parameters strongly depend on the variance and different Gaussian represent different colours. Unlike the study conducted by Friedman et al. [31], the background components have been determined in the presented paper by assuming that the background contains B highest probable colours. The probable background colours stay longer and more static. Each pixel {X1, ..., Xt} is modelled by a mixture of K Gaussian distributions. The probability of observing the current pixel value is given by Eq. (1)

$$p(X_N) = \Sigma_{i=1}^{K} W_{i,N}\,\eta(X_N, u_{i,N}, \Sigma_{i,N}) \tag{1}$$

where K is the distribution; $W_{i,N}$ is the weighed factor of i^{th} Gaussian mixture at time N; $u_{i,N}$ is mean value; and $\Sigma_{i,N}$ is the covariance matrix of i^{th} the Gaussian mixture at time N. Gaussian probability density function (η) is given by Eq. (2)

$$\eta(X_N, u, \Sigma) = 1/(2\pi)^{n/2}|\Sigma|^{1/2}\,\varrho - 1/2(X_N - \mu_N)^T\Sigma^{-1}(X_N - \mu_N) \tag{2}$$

K is determined by the available memory and computational power. The interpretation of each pixel in the frames is computationally expensive. Therefore, the covariance pixels are extracted for further processing from the covariance matrix which is given as $\Sigma_{i,N} = \sigma_k^2 I$.

The K distributions depend on fitness value (W_k/σ_k) of what. The B distribution in Eq. (3) is used to model the background scene.

$$B = argmin_b\,(\Sigma_{j=1}^{b}\,w_j > T) \tag{3}$$

The threshold (T) of each background pixel is the minimum fraction of the background model. It is the minimum prior probability that shows that the background is in the scene.

At times segmentation results are not satisfactory [32] which requires certain operations to be performed to refine the results. For example, some segmented images may contain degraded edges, extra regions, discontinuities or uncompleted boundaries which can be corrected by performing a morphological operation [33]. Morphological operations (opening, closing or filling) [34, 35] and connected component analysis have been performed in the presented paper to enhance the results of GMM, as illustrated in **Figure 3**.

3.2.2 Features extraction

HOG [4, 37] is extracted as a descriptor on the detected foreground regions using GMM. HOG descriptors operate on localised cells and are invariant to geometric and photometric transformations. It is implemented by using the following characteristics: grey scale without gamma correction; gradient filter [-1 0 1] without smoothing; 8×8 blocks of pixels of four 4×4 pixel cells (for default 16×16 pixels with four 8×8 cells of pixels); linear-gradient voting for the intervals of 0 deg to 180 deg on nine orientation bins; non-overlapping blocks (for default half of block overlap); L2-norm is used for normalisation. Each block contains four cells, so four histograms (nine orientation bins)

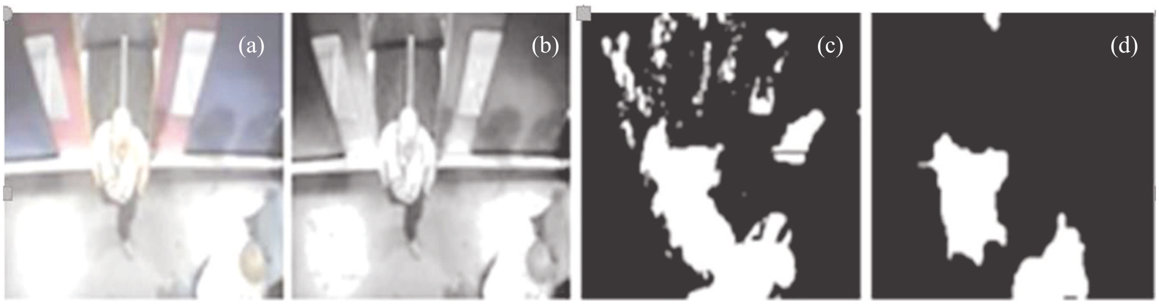

Figure 3. Video processing steps for passenger detection: (a) red, green and blue video frames; (b) colour conversion and historgram equalisation; (c) foreground region extraction; (d) morphological operation applied on extracted region.

have been concatenated for each block into a row vector that contains thirty six dimensions. The resultant descriptor is created by concatenating the all block histograms into a single row vector of the size of 36×total number of blocks in a single image. In the proposed work Each image of 120×120 pixel is divided into 8×8 pixels blocks in the presented paper and blocks for each image were obtained. As a result, the final descriptor has eight thousand one hundred dimensions (225×36).

3.2.3 Classification

The successfully extracted features are classified with a reliable classifier [38]. Linear one-versus-one support vector machine (SVM) classifier was used in this paper to recognise if the detected region was a passenger. SVM [37, 39] is known as supervised learning classifier which is associated with learning models that use data for regression and classification. It uses the concept of decision planes which defines the boundaries of decisions. HOG features were utilised for training and testing the SVM classifier.

3.3 Passenger Tracking

After the extracted foreground regions are classified into passengers and non-passenger regions, these are tracked in consecutive frames. Kalman filter [40-42] was applied for tracking the detected passengers. The detected passenger is ROI and its centroid (mean points of ROI along both x-axis and y-axis) is extracted in each frame. The tracking information is concerned with the motion of the passenger. The Kalman filter [40-42] estimates coordinate of human points in every frame of an image sequence. The input to the Kalman filter is centroids, and the radius of the stated vector and measurement vector. Passenger tracking was performed by prediction from the previous frame and verification of passenger existence at the predicted location.

3.3.1 Passengers counting

A vision-based counting system has been proposed in this paper to count passengers from the video. An LOI [40] has been defined to detect the passenger during getting on and/or off the transport vehicle. The counter is incremented as the detected passengers cross LOI. LOI can be defined manually to be more or less perpendicular to the main direction of motion, as the camera is fixed. The counter is incremented by one when the bottom of the bounding box of the detected passenger touches LOI. In order to know the direction of the motion of each blob, the algorithm compares the position of the bounding box in the current frame with the position of the bounding box in the previous frame.

4. DISCUSSION ON RESULTS

The PAMELA metropolitan train and bus datasets [2] were employed in the presented paper for performing the experiments. In the former dataset, the passengers getting in and out of the train have been captured from the camera mounted outside the train, whereas the camera is mounted inside the bus in the latter dataset (**Figure 4**). A total of eight videos (1, 2, 3, 4, 9, 10, 11, and 12) for getting in and seven videos (5, 6, 7, 8, 13, 14, and 15) for getting out scenarios are available in the PAMELA metropolitan train dataset [2]. These videos are in moving picture experts group format, with a resolution of 352×288 pixels and with a frame rate of 25 frames/sec. PAMELA bus dataset [2] contains three videos (1-3) in which passengers are getting on the bus as shown in **Figure 4**. The experiments were performed on MATLAB using the aforementioned datasets and the results from these experiments have been presented in the forthcoming sections.

Figure 4. View of PAMELA dataset [2] where left and middle frames are from PAMELA metropolitan train dataset during getting on and off, respectively and right frame is from the PAMELA bus dataset.

4.1 Passenger Detection

The passenger detection was carried out using the steps explained earlier. The input video frames are converted into HSV colour space. The visual quality of each input frame was improved by performing histogram equalisation and the foreground regions in each frame were extracted using GMM. These candidate regions were described using the HOG descriptor to perform the classification. The detected passengers (**Figure 5**) are tracked using a Kalman filter [40-42] and counted as they cross LOI.

The detection results have been summarised in **Table 1**. The results in **Table 1** are compared with ground truths (GT), and accuracy, sensitivity and false-negative rate were used as evaluation parameters. The SVM classifier was trained using video 8 and 2 and was tested with the remaining videos. This process was repeated for training the classifier using videos 8 and 3, 13 and 2, 13 and 3 iteratively. It is noted in **Table 1** indicate that the proposed algorithm achieves higher accuracy when the videos from both scenarios (video 13 and 3 for getting on and off, respectively) are used for training. The proposed method provided ninety one percent accuracy for passenger detection by using the SVM classifier.

4.2 Comparative Analysis of Passenger Detection

A comparison of the proposed detection approach has been made with the following two existing three detection approaches: (a) HOG for normal view [2]; (b) Viola-Jones [41, 43]; and (c) HOG with a pre-trained classifier [44]. The pre-processing steps are the same for both of these techniques. HOG features are extracted from the input videos In HOG with a pre-trained classifier, the latter is classified using an SVM classifier that is pre-trained on the INRIA person dataset [45]. The results for the passenger detection using this technique are shown in **Figure 6**. Haar features are extracted from input videos for Viola-Jones [41, 43] and testing was performed on a pre-trained cascade classifier on the VGG human pose estimation dataset. The camera is mounted at the top of the scene in the VGG human pose estimation dataset to overcome the shortcomings of occlusion. Upper body detection improves the robustness as well as reduces occlusion which is also a challenging task. The upper body detection regions include the head, face and shoulders of passengers as shown in **Figure 7**.

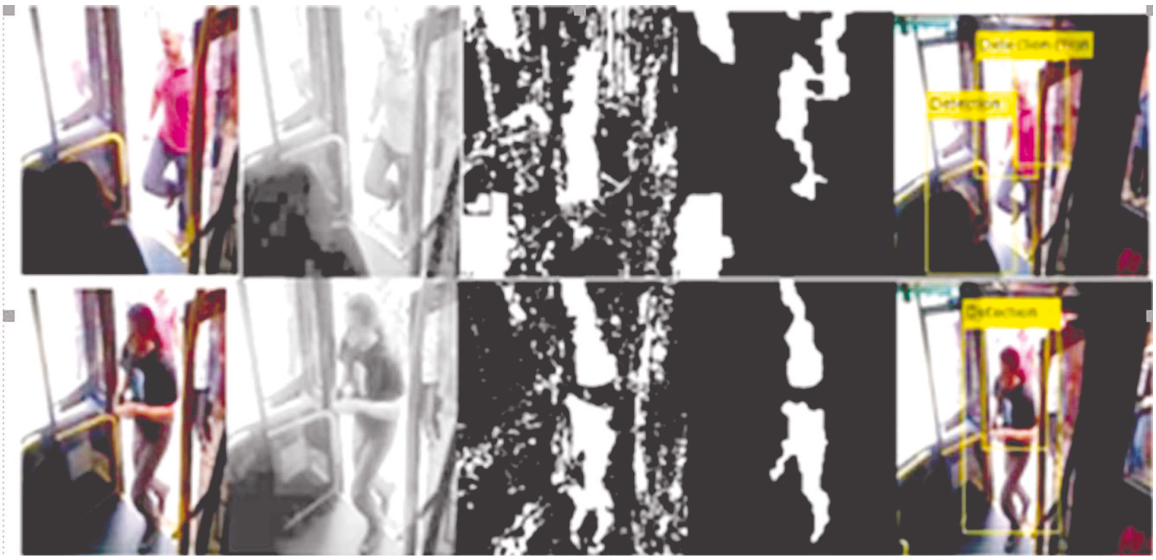

Figure 5. Video processing steps for passenger detection: (a) red, green and blue video frames; (b) pre-processing; (c) foreground extraction using Gaussian mixture model; (d) morphological operations on foreground; (e) detected passenger.

Table 1. Passenger detection percentage accuracy and sensitivity for PAMELA metropolitan train dataset [2]

Training video	Test video scenarios	Accuracy	Sensitivity
8, 2	ON/OFF	75.62	84.2
8, 3	ON/OFF	88.75	91.0
13, 2	ON/OFF	78.12	86.2
13, 3	ON/OFF	91.21	93.0
Average	ON/OFF	83.43	88.6

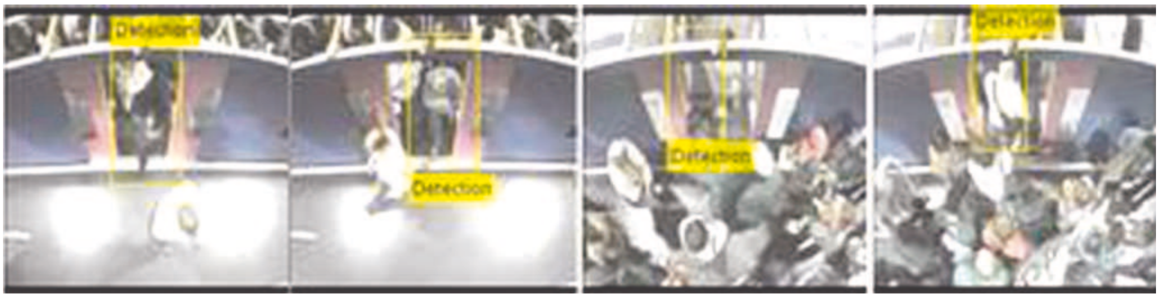

Figure 6. Passenger detection results using histograms of oriented gradients with pre-trained classifier.

Figure 7. Cascade or Viola-Jones [41, 43]algorithm for upper body passenger detection: (a) red, green and blue frame; (b) preprocessed frame; (c) final passenger detection.

Table 2. Comparisons of proposed passenger detection approach with the other approaches on PAMELA metropolitan train dataset [2]

Methods	Accuracy	Sensitivity
Proposeddetection system	91.21	93.0
Viola-Jones [41, 43]	69.00	75.4
HOG+SVM [50]	63.50	44.2
HOG with a pre-trained classifier	58.12	43.0
Viola-Jones [41, 43] and HOG (Combined)	77.60	79.8

Table 3. Comparisons of proposed passenger detection approach with existing approaches on PAMELA bus dataset [2]

Methods	Accuracy	Sensitivity
Proposeddetection system	86.24	80.2
Viola-Jones [41, 43]	69	75.4
HOG+SVM[50]	68.75	64.2
HOG with pre-trained classifier	66.73	52.7

Tables 2 and **3** show that the proposed approach provides better results for human detection in terms of accuracy and sensitivity as compared to the existing algorithms. HOG with a pre-trained classifier provides lower accuracy as the classifier is trained with full-body images whereas the PAMELA dataset [2] contains images of the passengers from the top view. Although HOG with a pre-trained classifier can detect whole body region, the entire human body region is not cleared due to occlusion in the PAMELA dataset [2].

4.3 Passengers Counting

It is clear in the above that the proposed passenger detection approach provides accurate results. The counting of passengers is carried out with the help of LOI. An automated counter is incremented when a passenger passes through LOI which observes the position of centroid of each detected passenger. LOI is manually selected as the placement and dimensions of the door are the same for the chosen dataset [3].

The relative error (η) for the evaluation of passenger counting is given by Eq. (4)

$$\eta = |\text{Total passengers counted} - \text{Total passengers in GT}| / \text{Total Passengers in GT} \quad (4)$$

$\eta > 0$ if the number of passengers is less or greater than GT

Tables 4 and **5** show η for passenger counting results for PAMELA metropolitan train and bus datasets [2], respectively. Results in **Table 4** show that extra passengers are detected and counted. This is due to the reason that the persons stopping nearby LOI are counted multiple times. As a result, the number of detected passengers is greater than GT passengers. **Figure 8** shows the passenger detection process. Note that the blue line in **Figure 8** represents LOI, the yellow box represents the boundary box and the integer value is a counter.

False results exist in both detection and counting. False results in counting mostly occur due to multiple counting or when a passenger is not detected due to occlusion, complex interaction or shadow. Sometimes the shadow is detected as human as it is moving along with the human body. The opening of the door gives false detection also as the foreground region is extracted on the motion-based method. All these reasons lead to false detection results. **Figure 9** shows false passenger detection results.

The pose of the passenger during getting on and off is different at times. Therefore, the counter does not work when detected passengers cross LOI. Similarly, if a passenger takes a long time than usual (for example the passenger in the second row of **Figure 10** who is reading a newspaper while staying near the door on LOI) before passing through LOI it is counted multiple times.

Table 4. Passenger counting results for PAMELA metropolitan train dataset [2] trained on videos 3 and 13

Video	Scenario	Ground truth	Counted	Relative error (η) (%)
1	OFF	50	54	8.0
2	OFF	50	53	6.0
4	OFF	50	55	10.0
5	ON	27	29	7.4
6	ON	27	28	3.7
7	ON	27	28	3.7
8	ON	27	31	14.8
9	OFF	50	59	18.0
10	OFF	51	59	15.6
11	OFF	50	60	20.0
12	OFF	50	56	12.0
14	ON	27	33	22.2
15	ON	27	35	29.6
			Average η	13.2

Table 5. Passenger counting results for PAMELA bus dataset [2] trained on video 3

Video	Ground truth	Counted	Relative error (η) (%)
1	10	11	10.0
2	10	9	10.0
		Average η	10.0

Figure 8. Passenger counting: top row shows counting results for PAMELA [2] metropolitan train; bottom row shows counting results for PAMELA [2] bus dataset.

Figure 9. False detection results.

Figure 10. False multiple counting.

5. CONCLUSIONS

Computer vision-based passenger detection and counting system has been proposed in this paper. The passengers are detected under different challenging scenarios such as curved curvature camera position, illumination, complex interaction, overlapping, occlusion, shadow, darkness and distortion effects. The proposed algorithm was tested using the existing dataset which provided good results with an accuracy up to ninety three percent. It was noted that the passengers who are moving toward or away from the top-mounted camera and changing dimensions of the passenger body size affects the accuracy of the detection system.

The proposed system can be extended to dynamic environments using moveable cameras. The use of the proposed detection system can be extended in security applicatins such as activity recognition. Similarly, the counting system can be used for measuring economic activites such as in a shopping mall.

REFERENCES

[1] Amin I, Al-Habaibeh F, Junejo A, Taylor J, Parkin RM. Automated People-counting by Using Low-resolution Infrared and Visual Cameras. Measure 2008;41(6):589-599.

[2] Belloc M, Velastin SA, Richard J, Fernandez D, Jara M. Detection of People Boarding/Alighting a Metropolitan Train using Computer Vision. In: 9th International Conference on Pattern Recognition Systems (ICPRS), Valparaíso, Chile: 2018.

[3] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. In: International Conference on Computer Vision & Pattern Recognition (CVPR'05) (Vol. 1). San Diego, CA, USA: 2005. p. 886-893.

[4] Pham M-T, Cham T-J. Fast Training and Selection of haar Features Using Statistics in Boosting-based Face Detection. In: IEEE 11th International Conference on Computer Vision. Rio de Janeiro, Brazil: 2007

[5] Bozzoli M, Cinque L, Sangineto E. A Statistical Method for People Counting in Crowded Environments. In: 14th IEEE International Conference on Image Analysis and Processing (ICIAP 2007). Modena, Italy: 2007. p. 506-511.

[6] Yu S, Chen X, Sun W, Xie D. A Robust Method for Detecting and Counting People. In: IEEE International Conference on Audio, Language and Image Processing. Shanghai, China: 2008. p. 1545-1549.

[7] Qi B, Kang L, Banerjee S. A Vehicle-based Edge Computing Platform for Transit and Human Mobility Analytics. In: Proceedings of Second ACM/IEEE Symposium on Edge Computing. Silicon Valley, CA, USA: 2017.
[8] Bao P, Zhang L, Wu X. Canny Edge Detection Enhancement by Scale Multiplication. IEEE Trans Pat Anal Mach Intel 2005;27(9):1485-1490.
[9] Barandiaran J, Murguia B, Boto F. Real-time People Counting Using Multiple Lines. In: Ninth IEEE International Workshop on Image Analysis for Multimedia Interactive Services. Klagenfurt, Austria: 2008. p. 159-162.
[10] Hejin Y. A Bus Passenger Flow Estimation Method based on Feature Point's Trajectory Clustering. In: IEEE International Conference on Intelligent Computing and Intelligent Systems (Vol. 1). Xiamen, China: 2010. p. 426-430.
[11] Xu XW, Wang ZY, Liang YH, Zhang YQ. A Rapid Method for Passing People Counting in Monocular Video Sequences. In: IEEE International Conference on Machine Learning and Cybernetics (Vol. 3). Hong Kong, China: 2007. p. 1657-1662.
[12] Xin Y, Sun G, Wu Q. A Preprocessing Method for Tracking and Counting Pedestrians in the Bus Video Monitor. In: 6th IEEE International Conference on Industrial Informatics. Daejeon, South Korea: 2008. p. 1689-1693.
[13] Escolano CO, Billones RK, Sybingco E, Fillone AD, Dadios EP. Passenger Demand Forecast Using Optical Flow Passenger Counting System for Bus Dispatch Scheduling. In: IEEE Region 10 Conference (TENCON). Singapore, Singapore: 2016. p. 1875-1878.
[14] Sidla O, Lypetskyy Y, Brandle N, Seer S. Pedestrian Detection and Tracking for Counting Applications in Crowded Situations. In: IEEE International Conference on Video and Signal Based Surveillance. Sydney, Australia: 2006. p. 70-70.
[15] Boreiko O, Teslyuk V. Structural Model of Passenger Counting and Public Transport Tracking System of a Smart City. In: XII IEEE International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH). Lviv, Ukraine: 2016. p. 124-126.
[16] Alodhaibi S, Burdett RL, Yarlagadda PK. Framework for Airport Outbound Passenger Flow Modeling. Procedia Eng 2017;174:1100-1109.
[17] Schousek TJ. Vehicle Occupant Restraint with Seat Pressure Sensor. Google patents, 1995.
[18] Faber P. Seat Occupation Detection Inside Vehicles. In: IEEE Southwest Symposium on Image Analysis and Interpretation. city, country: 2000. p. 187-191.
[19] Kocamaz MK, Gong J, Pires BR. Vision-based Counting of Pedestrians and Cyclists. In: IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Placid, NY, USA: 2016. p. 1-8.
[20] Cleopatra WS, Venkatesh S, Tan T. Passenger Monitoring in Moving Bus Video. In: 10th IEEE International Conference on Control, Automation, Robotics, and Vision. Hanoi, Vietnam: 2008. p. 719-725.
[21] Terada K, Yoshida D, Oe S, Yamaguchi J. A Method of Counting the Passing People by Using the Stereo Images. In: Proceedings of IEEE International Conference on Image Processing (Cat. 99CH36348) (Vol. 2). Kobe, Japan: 1999. p. 338-342.
[22] Kilambi P, Ribnick E, Joshi AJ, Masoud O, Papanikolopoulos N. Estimating Pedestrian Counts in Groups. Comp Vis Image Understand 2008;110(1):43-59.
[23] Kim JW, Choi KS, Choi BD, Ko SJ. Real-time Vision-based People Counting System for the Security Door. In: International Technical Conference on Circuits/Systems Computers and Communications. Phuket, Thailand: 2002. p. 1416-1419.
[24] Stauffer C, Grimson WE. Adaptive Background Mixture Models for Real-time Tracking. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149) (Vol. 2). Fort Collins, CO, USA: 1999. p. 246-252.
[25] Chen M, Wei X, Yang Q, Li Q, Wang G, Yang MH. Spatiotemporal GMM for Background Subtraction with Superpixel Hierarchy. In: IEEE Trans Patt Anal Mach Intel 2018;40(6):1518-25.
[26] Bouwmans T, Silva C, Marghes C, Zitouni MS, Bhaskar H, Frelicot C. On the Role and the Importance of Features for Background Modeling and Foreground Detection. Comp Sci Rev 2018;28:26-91.
[27] Tavakkoli A, Nicolescu M, Bebis G. An Adaptive Recursive Learning Technique for Robust Foreground Object Detection. In: Proceedings of the International Workshop on Statistical Methods in Multi-image and Video Processing (in conjunction with ECCV06). Graz, Austria: 2006.
[28] Saravanan G, Yamuna G, Nandhini S. Real-time Implementation of RGB to HSV/HSI/HSL and its Reverse Color Space Models. In: IEEE International Conference on Communication and Signal Processing (ICCSP). Melmaruvathur, India: 2016. p. 462-466.
[29] Cucchiara R, Grana C, Piccardi M, Prati A, Sirotti S. Improving Shadow Suppression in Moving Object Detection with HSV Color Information. In: IEEE Intelligent Transportation Systems. Proceedings (Cat. No. 01TH8585). Oakland, CA, USA: 2001. p. 334-339.
[30] Friedman N, Russell S. Image Segmentation in Video Sequences: A Probabilistic Approach. In: Proceedings of the Thirteenth Conference on Uncertainty in artificial intelligence. Morgan

Kaufmann Publishers Inc, USA, 1997. p. 175-181.
[31] Ahmad N, Hussain S, Ahmad K, Conci N. Computer Vision based Room Interior Design. In: Eighth International Conference on Machine Vision (ICMV 2015). (Vol. 9875). Barcelona, Spain: 2015. p. 98751G.
[32] Albiol A, Mora I, Naranjo V. Real-time High-density People Counter Using Morphological Tools. IEEE Trans Intel Transport Sys 2001;2(4):204-218.
[33] Lefloch D, Cheikh FA, Hardeberg JY, Gouton P, Picot-Clemente R. Real-time People Counting System Using a Single Video Camera. In: Real-Time Image Processing (Vol. 6811). California, United States: 2008. p. 681109.
[34] Liang C-W, Juang C-F. Moving Object Classification Using Local Shape and HOG Features in Wavelet-transformed Space with Hierarchical SVM Classifiers. App Soft Comp 2015;28:483-497.
[35] Ahmed U, Khan A, Khan SH, Basit A, Haq IU, Lee YS. Transfer Learning and Meta Classification Based Deep Churn Prediction System for Telecom Industry. Cornell University Working Paper, 2019 (accessed on dd-mm-year). Available from http://arxiv.org/abs/1901.06091
[36] Lee Y. Support Vector Machines for Classification: A Statistical Portrait. In: Statistical Methods in Molecular Biology, Springer, 2010. p. 347-368.
[37] Patel HA, Thakore DG. Moving object tracking using Kalman filter. Int J Comp Sci Mob Comp 2013;2(4):326-332.
[38] Li SE, Li G, Jiaying Y, Chang L, Bo C, Jianqiang W, Keqiang L. Kalman Filter-based Tracking of Moving Objects Using Linear Ultrasonic Sensor Array for Road Vehicles. Mech Sys Sig Process 2018;98:173-189.
[39] Chai J, Wu C-h, Zhao C, Chi H-L, Wang X, Ling BW-K, Teo KL. Reference Tag Supported RFID Tracking Using Robust Support Vector Regression and Kalman Filter. Adv Eng Inform 2017;32:1-10.
[40] Ma Z, Chan AB. Counting People Crossing a Line Using Integer Programming and Local Features. IEEE Trans Circuit Sys Video Tech 2016;26(10):1955-1969.
[41] Hsieh C-C, Liou D-H. Novel Haar Features For Real-time Hand Gesture Recognition Using SVM. J Real-Time Imag Process 2015;10(2):357-370.
[42] ElMaghraby A, Abdalla M, Enany O, ElNahas MY. Hybrid Face Detection System using a Combination of Viola-Jones Method and Skin Detection. Int J Comp App 2013;71(6):15-22.
[43] Vikram K, Padmavathi S. Facial Parts Detection Using the Viola-Jones Algorithm. In: 4th IEEE International Conference on Advanced Computing and Communication Systems (ICACCS). Coimbatore, India: 2017. p. 1-4.
[44] The MathWorks, Inc. Detect Upright People Using HOG Features, 2013 (updated 3 May 2013; accessed on 1 October 2018). Available from https://www.mathworks.com/help/vision/ref/vision.peopledetector-system-object.html.
[45] Visual Geometry Group, University of Oxford. VGG Human Pose Estimation Dataset, 2013 (updated 15 May 2013; accessed on 1 October 2018). Available from https://www.robots.ox.ac.uk/~vgg/data/pose/