A Robust Real Time People tracking and Counting incorporating shadow detection and removal

J. L. Raheja
Machine Vision Lab, DSG,
Council of Scientific & Industrial Research
(CSIR- CEERI), Pilani, Rajasthan, India

Sishir Kalita
Dept. of Electronics & Communication Engineering,
Tezpur University, Naapam
Tezpur, Assam, India

Pallab Jyoti Dutta
Dept. of Electronics & Communication Engineering,
Tezpur University, Naapam
Tezpur, Assam, India

Solanki Lovendra
Department of ECE,
BKBIET,
Pilani, Rajasthan, India

ABSTRACT
Video processing serves as a hidden treasure, rather a boon in disguise to surveillance system. The counting of people passing through a surveillance area is an important issue of this domain. It always relies on the process of background subtraction. The estimation of dynamic background model and the shadow removal are two main challenges of background subtraction. In this paper, a bi-directional people counting algorithm is proposed. To develop a robust counting system, Gaussian mixture model (GMM) is used to describe the background scene. But this algorithm does not provide a way to classify the shadows from the moving foreground objects. To achieve better performance, background model is upgraded by combining a Chromatic color model. This provides better improvement in moving objects detection by eliminating the shadows from foreground. A multi-class feature based tracking algorithm is applied for multiple tracking to handle occlusion problem. To improve the counting of individual in both directions a scheme is proposed, which is developed by a multi-level reverse tracking procedure. This proposed counting system seems to provide higher accuracy and better performance even under crowded situation and changing environmental situations. Experimental results show that high accuracy of bi-directional counting can be achieved if the density of the people access is low.

Keywords
Video Surveillance, People Counting, Gaussian Mixture Model, merge-split, Morphological Processing.

1. INTRODUCTION
Nowadays people counting system has become more and more important in many practical applications, such as video surveillance, pedestrian traffic management and tourists flow estimation. Many traditional approaches were employed for implementing this system. Researchers were also involved in Counting in earlier days, which were subject to errors. Later on, electronic devices such as lasers, infra-red counting systems were introduced. Some intractable problems of these systems are counting of every object that passes through its way and also it cannot track the objects precisely.A number of surveillance applications require detection and counting of people to ensure security, safety and site management. Examples include the length estimation of queue in retail outlets, the entry points monitoring, vending machines, ATM

Booths, parking and traffic lots etc. [1]. In contrast to traditional methods, computer vision-based approaches are more promising because they hold the advantage of non-intrusiveness and high flexibility. So, in present days, people counting system is done with the help of computer vision techniques and also video processing algorithms. The extraction of foreground from the background is the first and most important step in many applications. So, it is very much desirable to model the background more accurately and reliably so that it can be implemented in various strata of environment. In the past decades many background subtraction algorithms have been proposed including Running Gaussian Average [2], Temporal Median Filter [3], Mixture of Gaussians [4], Kernel Density Estimation (KDE) [5], and Kalman Filter [6]. Among the all techniques GMM is the most robust. Shadow is another factor that might affect the performance of foreground detection and make the problems more complicated. Most of background modelling method classifies shadows as foreground, which is the expected behaviour since shadows significantly differ from the background.

In this work, a real time bi-directional people counting and tracking method is proposed. To describe the background, GMM is used because of its robustness. To classify the shadow pixel from other non-background pixel, it is very much necessary to separate brightness from chromaticity component. In the horprasert computational color model [7], pixels brightness is separated from the chromaticity component and is integrated with the GMM. This is done by comparing the pixel value which is non-background against the current background component. After the completion of foreground detection some features are extracted, which are used in tracking and counting of each individual entering into the view area of the camera. From the analysis of the results obtained, it is clearedthat higher accuracy and better performance can be obtained even under crowded situation and dynamic background conditions.

This paper is organised into the following sections. Section 2 provides the literature survey of some related works on people counting system. Section 3 gives an insight into the proposed algorithm for counting. Experimental details and related resultsare included in Section 4. Finally, Section 5 concludes the paper.

2. LITERATURE SURVEY
In recent years, many computer vision based approaches for people counting and tracking to deal various applications are proposed. Researchers are trying to develop robust background modelling algorithm to recognize the background and track the moving object. Honglian Ma and Huchuan Lu et al. in [1] proposed a multiple people segmentation method based on the bi-directional projection histogram of grayscale of frame

differencing image. The insufficiency of this system is to only use the projection information of gray value as the detection criterion. The error in counting may occur, if the projection information of other motion targets is similar to the human targets. A.J.Schofield et al. [8] proposed a method to count people in video images by using the neural networks. They use RAM based neural network classifier to identify section of background scene in each test image. Kenji Terada et al. in [9] proposed a counting method in which they use template process to detect the direction of the moving objects. From the moving direction information space time image generated. Then by counting the people data, the number of incoming and outgoing person is counted. Chao-Ho (Thou-Ho) Chen et al. in [10] proposed a scheme to count the number of people entering and leaving a bus by using a zenithal camera. They firstly divided the captured frame into many blocks then blocks with similar motion vectors regarded as belonging to same object. Hartono Septian et al. in [11] proposed a method to count number of people, where they first detect the person as blobs and represents as binary masks then a correlation based algorithm is applied to track the person in consecutive frames. Tsong-Yi Chen et al. in [12] proposed an algorithm of motion object detection and segmentation based on frame difference algorithm where the height and width of the detected people is taken as feature for counting. To detect and track moving people, Kim et al. in [13] proposed a real-time scheme, where they used bounding box to enclose each person. The approximated convex hull of each individual in the tracking area is obtained to provide more accurate tracking information. Javier Barandiaran et al. in [14] proposed a counting method in which Counting is performed by analysing an image zone composed by a set of virtual counting lines and frame difference method to detect the moving objects. All the above systems give almost satisfactory result when number of access people is low and the background is static. In case of dynamic background condition such as illumination variations, effects of moving elements of the scene (e.g. waving leaves) and slow-moving objects, a robust background modelling algorithm should be used. Huazhong Xu and Pei Lv et al. in [15] proposed a counting algorithm which is based on Head-shoulder Detection of human body. They used GMM to model the background. To model the background in their proposed counting method, Chunhui Zhao and Quan Pan et al.in [16] used nonparametric kernel density estimation algorithm and extracted the local features of the targets to track the person accurately and quickly. Bescos et al. in [17] proposed a counting method where they used a single zenithal camera to count people flow at an entrance to a store. To efficiently consider both lighting and texture information, they used DCT based segmentation, to deal with shadows, sudden changes in background illumination and sporadic camera motion due to vibration. In [18] Enwei Zhang et al. used single Gaussian model as moving objects detection method and normalize RGB to cope with the shadows. To classify the foreground pixel from the background Dar-Shyang LEE et al. in [19], proposed a Bayesian Framework based on the Mixture of Gaussian model. In recent years researchers are also giving interest in shadow removal aspect from the background. Many shadow removal algorithm is proposed in literatures. Fredrik Kristensen et al [20] used the MoG to model the background in YC_bC_r color space to cope with the shadow. In [21] Alessandro Leone et al. used photometric properties to detect shadows and regarded shadow as semi-transparent region which retains a reduced contrast representation of the underlying surface pattern and texture. After that they discriminated shadows from moving objects based on a Pursuit scheme using an over-complete dictionary. Our proposed counting method is based on Stauffer et al's [4] framework, the difference lies in the introduction of a shadow detection algorithm. The horprasert color model helps us to distinguish the shading background from the ordinary background and moving foreground objects. Apart from this, the inherent problem of merging and splitting, which is occurring due to touching with one another and separating from a heap to individual, also deal for nonlinear real time situation. In real time situation for bi-directional people counting, it is not sufficient to deal with previous frame only for refining the counting of number of objects. A multi-class feature based object tracking scheme is applied to overcome the multiple tracking problem in crowded scene. This reduced the inherent people-touching and overlapping problem when they are in motion. A multilevel reverse tracking method is proposed for better accuracy and refinement of the counting. A counting region is provided to constrain the counting process for lesser computational time.

3. THE PROPOSED METHOD

In our proposed counting method, one video camera is set on the top of the surveillance to capture the birds view motion of passing people. The proposed algorithm for automatic bi-directional counting for the pedestrian flow is described in Figure 1. In our approach we use GMM to extract the foreground region. A color model is used to classify the shadow region from the original background. After that morphological operation [22] is applied to remove the unwanted noise. Connected component labelling [22] method is used to extract the large blob. A few geometric features from the detected object is extracted which will be used in the tracking and counting process. The object's geometric features can include location, shape and centre of mass (centroid), etc. After feature detection, a multi-class object tracking scheme is used to remove multi-object tracking problem in occlusion. A fixed bounded region within the frame is used for the counting process. To achieve better accuracy, the proposed multilevel reverse tracking method is applied.

3.1. Moving Object Detection

The detection or extraction of moving objects from the video scene is initial and most important step in the field of video processing. Here, Gaussian mixture model is used to describe a pixel of background. This statistical background model allows multimodal background, thus providing robust adaption against repetitive motion of scene elements, slow-moving objects, and introducing or removing objects from the scene. In this model the value of a particular pixel over time is seen as a measurement, X_t of a stochastic variable.

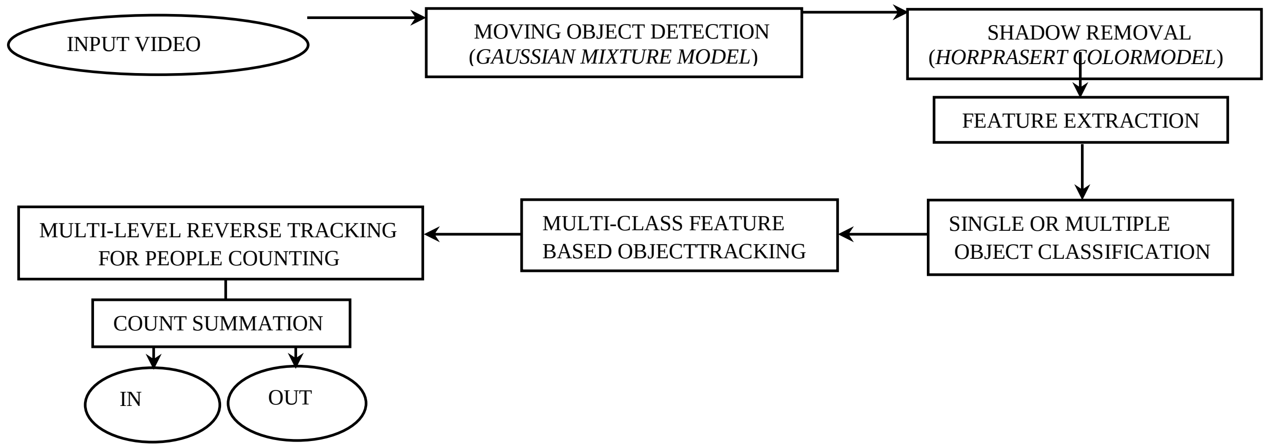


Fig. 1:System model for the proposed counting method.

At any time, besides the current measurement of X_t, the history, $M_t = \{X_1, X_2, \ldots X_{t-1}\}$, of that particular pixel is known [4]. Thus the recent history of a pixel can be modelled using mixture of K Gaussian distributions. Different colours are assumed to represents as different Gaussians. The probability to observe the current background pixel X_t is the weighted sum of the K distributions:

$$P(X_t) = \sum_{i=1}^{K} w_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

Where K is the number of Gaussian distributions, $w_{i,t}$ is the weight of the i^{th} distribution at time t and $\sum w_i = 1$, $\mu_{i,t}$ is the mean value of the i^{th} Gaussian at time t. $\Sigma_{i,t}$ is the covariance matrix of the i^{th} Gaussian in the mixture at time t. How longer a color stays in the scene is represented by the weight parameter w. η is the probability density function of i^{th} Gaussian, which is given by.

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = (1/((2\prod)^{n/2} |\Sigma_i|^{1/2})) * \exp(-1/2(X_t - \mu_t) \Sigma_i^{-1} (X_t - \mu_t) \quad (2)$$

K can be determined by computational and memory power. Here we take three Gaussian distribution to describe a pixel. To avoid the computational cost it is assumed that the variation of the R, G and B color channel is same. So the covariance matrix can be defined as:

$$\Sigma_{i,t} = \sigma_i^2 I \quad (3)$$

The algorithm assigns one GMM to each pixel of the image and updates the model parameters (mean value and variance) on-line. Every time when a new pixel is get it is checked with the already exist K Gaussian distributions. A match is found if $|X_{i,t} - \mu_{i,t}| < 2.5 * \sigma_i$. If none of the K distributions match the current pixel value, the least probable distribution is replaced with a distribution with the current value as its mean value, an initially high variance, and low prior weight. The weight is adjusted as following.

$$w_{k,t} = (1 - \alpha) w_{k,t-1} + \alpha M_{k,t} \quad (4)$$

Where α is the learning rate, $M_{k,t}$ is 1 if match is found and 0 if not match is found. The μ and σparameters for unmatched distributions remain the same. For the match distributions these parameters are updated as follows:

$$\mu_t = (1 - \rho) \mu_{t-1} + \alpha X_t \quad (5)$$

$$\sigma_t^2 = (1 - \rho) \sigma_{t-1}^2 + \rho (X_t - \mu_t)^T (X_t - \mu_t) \quad (6)$$

Where, $\rho = \alpha\eta(X_t | \mu_t, \sigma_t)$. To decide which Gaussian of the GMM represents the background, they are sorted by the value of w/σ. This value gets higher with less variance of the distribution and more times the distribution has been used. This leads to the Gaussians of the GMM being sorted in a list where the first more probably is background. Then the first B distributions are chosen as the background model, where

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^{b} w_k > T\right)$$

T is a threshold on how much data should be accounted to the background. If T is chosen low usually a unimodal background will be represented but with a higher T the probability of a multi-modal background increases.

3.2.Shadow Removal Algorithm

The GMM that we use to describe the pixels of the background cannot be used to identify the moving shadows from the original background. This is because no heuristic exists to label the Gaussian component as moving shadows. Shadows often represent pixel values different from the background distributions. The Horprasert color model gives good results in modelling the influence of shadows on underlying background pixel intensities. The idea behind the use of this color model is that it can separate the brightness and chromaticity components and make it compatible with mixture model.

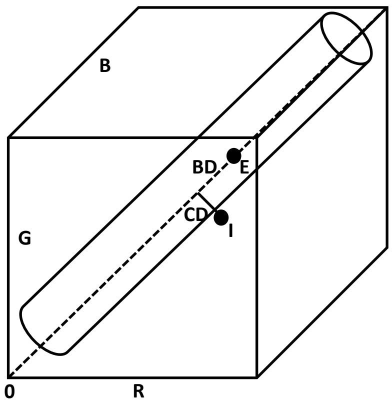


Fig. 2:The horprasert color model along with the Gaussian Mixture model.

A comparison of non-background with the current background is performed. A pixel is considered as shadow if the brightness component and chromaticity component are within some threshold. This color model along with the GMM can be viewed as shown in Figure 2. Here it is assumed that the RGB rectangle is the search space for the GMM. The Horprasert model assumes a cylinder within the RGB space and its axis is defined as a chromaticity line, which connects the pixel value and the origin. This model is based on empirical observation that shadowed pixels values tend to follow the expected chromaticity line. It separates distortions into Brightness and Chromaticity components, the axial distance between the expected intensity and the observed intensity is proportional to the variation in the brightness distortion (BD), and the radial distance proportional to the chromaticity distortion (CD).

$$\mathrm{BD} = \arg\min_{\alpha}(\mathrm{I} - \alpha\mathrm{E})^2 (7)$$
$$\mathrm{CD} = (\mathrm{I} - \alpha\mathrm{E}) \ (8)$$

Where E is the expected chromaticity line and I is the observed current pixel value. α represents the pixel's strength of brightness with respect to the expected value.For a pixel if BD is within some predefined threshold T_1 and $T_2$and the CD is less than some predefined threshold then it is considered as the shadow pixel. In our system we take T_1=0.5 and T_2=6.0.

3.3. Decision Rule to Segment the Detected People Pattern

If partial or complete occlusion occurs in the scene, it is very complicated to determine the number of actual persons. Based on analysis of the actual experiments, a segmentation rule is proposed. Height and width of bounding box of the detected people region and the number of pixels of this region is used to estimate the number of people containing in it. In the counting and the tracking process this gives basic information regarding the detected people regions. This single and multi-person classification is very essential to handle the merging and splitting condition.With the help of experiments, an Evaluation parameter, based on the above extracted features, is derived as shown in Equation (9).

$$\mathrm{E}_p(\mathrm{i}) = \alpha * \mathrm{A}(\mathrm{i}) + \beta * \mathrm{H}(\mathrm{i}) + \gamma * \mathrm{W}(\mathrm{i}) \ (9)$$

Where E_p is the Evaluation parameter; A, H and W are the number of filled pixels, height and width of the i^{th}detected people region respectively. The α, β and γ are the arbitrary parameters and $\alpha + \beta + \gamma = 1$. A threshold value T_p is obtained based on evaluation parameter for a typical size of one single people pattern.

This evaluation parameter estimates the number of people contained in one region. Practically, due to overlap between the people region, the area of n-objects will be smaller than total area of n single-objects. From the analysis of data during the experiment, a segmentation rule is derived in Equations (10)–(13).

$$\mathrm{P}_n = 1, \quad \text{if} \quad \mathrm{M}_p < \mathrm{E}_p \leq \mathrm{T}_p (10)$$
$$\mathrm{P}_n = 2, \quad \text{if} \quad \mathrm{T}_p < \mathrm{E}_p \leq 1.6\mathrm{T}_p (11)$$
$$\mathrm{P}_n = 3, \quad \text{if} \quad 1.6\mathrm{T}_p < \mathrm{E}_p \leq 2.8\mathrm{T}_p (12)$$
$$\mathrm{P}_n = 4, \quad \text{if} \quad 2.8\mathrm{T}_p < \mathrm{E}_p \leq 3.7\mathrm{T}_p (13)$$

Where P_n is the number of person and M_p is the minimum allowable quantity for a single person. By using this derived segmentation rule the single or multiple person classification is done. The above rule is based on the normal human orientation when they are in motion. After the estimation of the number of people in a mergedblob, a rectangular box based on a template of average human height and width on the individuals is applied. This is done using left-top to right-bottom approach.

3.4. Multi-Class Feature Based Multiple People Tracking Scheme

Object tracking provides the trajectories of the moving objects in a video scene. However, when occlusion occurs, it is very difficult to track multiple objects accurately. People may create group when they are walking and may split from a heap. With this consideration, the occlusion situation can be described by multi-class scenario for the better detection of merging and splitting condition. To achieve this, a feature based tracking scheme combined with this multi-class description of the occlusion scenario is applied. This scheme is very much helpful to overcome the problem of multiple tracking.

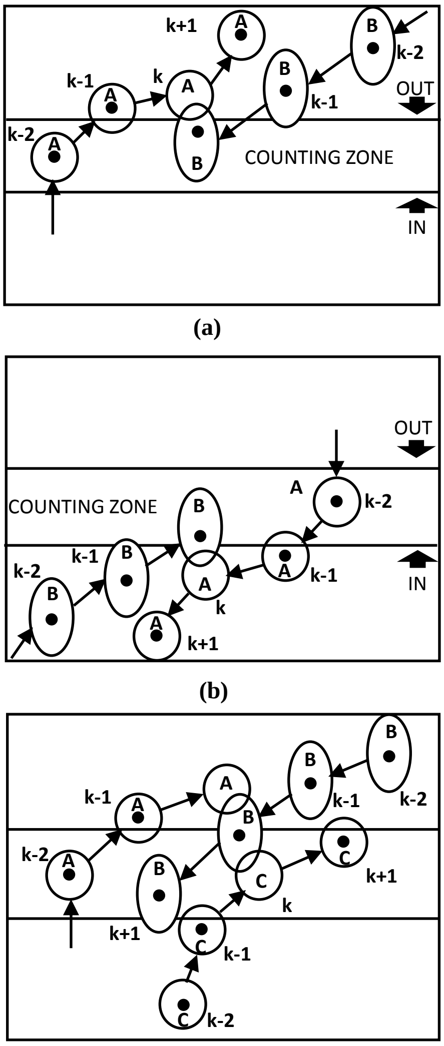
The Kalman filter is used to predict the future state of every object in the next frame. The Kalman Filter is a statistical method that involves an algorithm which provides an efficient recursive approach in estimating the states of process by minimizing the mean of squared error. Kalman filter uses elements of estimation theory to obtain the best unbiased estimator of a state of a dynamic system using the previous measurement [23].

A bounding box and its centre are used as the feature set to track the moving objects. A Similarity function is defined based on the above features to describe the closeness of two detected people regions in two consecutive image frames. If the horizontal and vertical coordinates of the centroid and the area of the i^{th} people region for the frame k are x_k^i, y_k^i and S_k^i respectively, then the distance between the j^{th} object of $(k\text{-}1)^{th}$ frame and the i^{th} object of the k^{th} frame is given by,

$$\mathrm{DIS}\,(\mathrm{i},\mathrm{j}) = ((x_k^i - x_{k-1}^j)^2 + (y_k^i - y_{k-1}^j)^2)^{0.5} (14)$$

Where the $x_{k-1}^j$and $y_{k-1}^j$are the horizontal and vertical coordinates of the j^{th} object of the $(k-1)^{th}$ frame. Then the area difference of the i^{th} object of the k^{th} frame and the j^{th} object of the $(k-1)^{th}$ frame can be derived as follows.

$$AR(i,j) = |S_k^i - S_{k-1}^j| \quad (15)$$



(a)

(b)

(c)

Fig. 3: Different nonlinear occlusion situations.

Where $S_k^i = l_k^i * h_k^i$, which represents the area of the rectangular window of i^{th} object of the k^{th} frame. Like this S^j_{k-1} represents the area of the rectangular window of j^{th} object of the $(k-1)^{th}$frame, l^i_k and h^i_k represent the length and height of the rectangular window respectively. From these above two parameters a Similarity function can be defined as follows.

$$C_f(i,j) = \alpha * DIS(i,j) + \beta * AR(i,j) \quad (16)$$

Where α and β are the arbitrary value and $\alpha + \beta = 1$. This Similarity function is used for the feature matching between the objects of two consecutive frames. The smaller value of the Similarity function for two objects of two consecutive frames represents more likely the same object.

By using the derived segmentation rule for the detected people region segmentation, the single or multiple person classification is done.

The occlusion scenario for merging and splitting can be viewed as follows.

Case1: A people pattern is detected and classified as single person and there is no feature match between the other tracked patterns of the previous frame, then this people pattern is new.

Case2:A people pattern is detected and classified as a group of people and there is no feature match between the other tracked patterns of the previous frame, then this group of people pattern appears for the first time.

Case3:A people pattern is detected and classified as single person and there is a feature match between one of the tracked single people patterns from the previous frame, and then this single person is tracked.

Case4: A people pattern is detected and classified as group of people and there is a feature match between one of the tracked group of people pattern from the previous frame, then this group of person is tracked.

Case5:A people pattern is detected and classified as single person and there is a feature match between one of the tracked groups of people pattern from the previous frame,then group of people is separating.

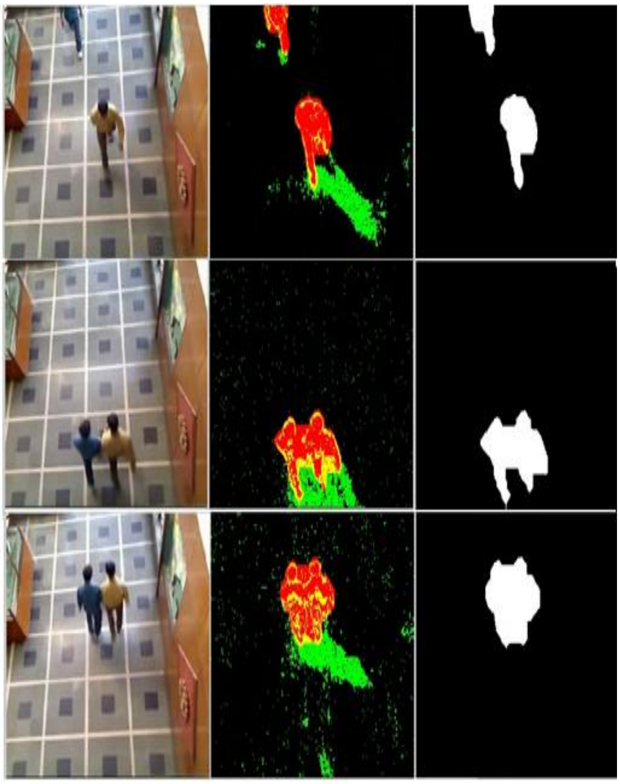
Case6:A people pattern is detected and classified as group of people and there is a feature match between one of the tracked single pattern from the previous frame, then occlusion is occurred.

3.4. People Counting And Refinement Based On Multilevel Reverse Tracking.

The counting of people is the main and the last step of the algorithm. From the feature based multi-class tracking scheme, the actual status of the people pattern in the current frame is now known. This information helps us to develop a reliable counting scheme. A fixed virtual region in the middle of the detection range of the camera is selected for the counting. The middle portion is preferred because a full human body becomes entirely visible within this region and provides a better counting result rather than using the multiple lines method at the edge of the detection range. The counting process will continue only when the objects enter in this counting zone. In conventional counting scheme, only the previous frame is taken into account for the counting process. But in real time case where occlusion is the inherent problem to degrade the performance of counting, it will be not sufficient to deal with only the previous frame. Here a multilevel reverse tracking method is proposed for better accuracy in the counting algorithm. According to this, the status of an object of the k^{th} frame is checked reversely with the $(k-1)^{th}$ and $(k-2)^{th}$ frames or sometimes on the higher level with the help of the Similarity function defined in Equation-16 to get accurate result. This algorithm also has the adoptability to sense the situation. If there is no occlusion in the video scene then the previous frame is only observed. This adoptability and the activation of process within the small counting zone, provides less computational time. Figure 3 shows a few nonlinear occlusion situations where multilevel reverse tracking is needed for accurate counting. From Figure 3, it is clearly visible that the counting cannot be done accurately in a proper direction by simply dealing with the previous frames. Figure 3(a) shows that two objects A and B are merged on the k^{th} image frame. In the $(k-1)^{th}$ frame the state of these two objects shows that the centroid of both are outside the counting zone. Then, because of the merging, the direction and counting status of A and B cannot be able to judge on the k^{th} frame by only observing the $(k-1)^{th}$ frame. Therefore, to count accurately in the proper direction, it is necessary to check the status on the $(k-2)^{th}$ frame also. From the overall observation, along with the $(k-2)^{th}$ image frame it is visible that object A, which is entering, is already counted and B is yet to count. The histories of all the objects are stored in a vector to compute the Similarity function between the two consecutive frames and the direction of each object. Thus this counting scheme is highly effective in case of and occluded situation. Figure 3(b) and 3(c) can also be described by this scheme only to obtain accurate counting.

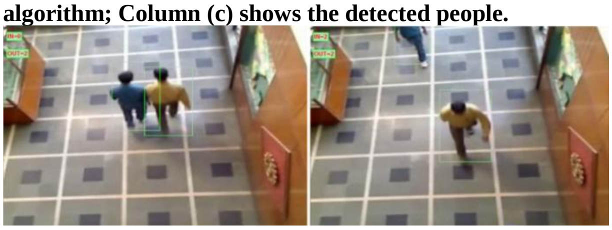
4. EXPERIMENTTAL RESULTS AND ANALYSIS

The experimental results and the performance evaluations are discussed in this section. Indoor as well as outdoor video sequences are used to realize proposed people-counting system with a colour video camera kept 6 meters above the projectionarea. The surveillance area of the captured video is 320×240 pixels (i.e., frame size). Figure 4 shows the results of background subtraction and shadow removal algorithm which gives a better performance inthe illumination variation indoor scene. Column 4(a) shows three random frames of the captured video scene. Column 4(b) shows the results of background subtraction.Here red pixels represent the foreground region and green pixels represent the shadow region. Some pixels are



(a) (b) (c)

Fig. 4: The results of background subtraction; Column (a) displays the original frames; Column (b) shows the results after the application of GMM and shadow removal algorithm; Column (c) shows the detected people.



(a) (b)

Fig. 5:Results of the counting system for the indoor scene.

(a) (b)

Fig. 6: Results of the counting system for the outdoor scene.

falsely represented as shadows which can be ignored. The resultant detected objects are displayed in column 4(c).The indoor video sequences are captured in the entrance room of an office Figure 5 and Figure 6 shows the result for counting and tracking in indoor and outdoor scene respectively. The estimated numbers of entering and leaving persons are shown in the left-up corner of the frames the both cases. For detailed evaluation, the directions (unidirectional & bi-directional) and speed (normal & fast) are taken as the attributes for indoor and outdoor cases to evaluate the performance. Table I shows the evaluation results for the indoor scene when there is no occlusion occurred in the frames. Table II shows the evaluation results when occlusion is exist in certain frames of the indoor video scene. Table III shows the evaluation results when occlusion is exist in certain frames of the outdoor video scene. From the table I it iscleared that our system provides very high accuracy in case of normal situation when there is no occlusion in the captured video sequences. As situation become more complicated, erroneous judgments occur in the counting.

Table I: Evaluation results for the indoor scene when there is no occlusion occurred in the frames.

direction	speed	No of people	Count No	Accuracy (%)
Uni.	normal	40	39	97.50
	fast	39	37	94.87
Bi.	normal	39	37	94.87
	fast	49	46	93.88

Table II: Evaluation results of the indoor video scene.

direction	speed	No of people	Count No	Accuracy (%)
Uni.	normal	34	32	94.12
	fast	40	37	92.50
Bi.	normal	53	49	92.45
	fast	43	38	88.37

Table III: Evaluation results of the outdoor video scene.

direction	speed	No of people	Count No	Accuracy (%)
Uni.	normal	34	32	94.12
	fast	40	37	92.50
Bi.	normal	50	47	94.00
	fast	43	37	86.05

From the tables II and III, it is clear that the accuracy of the counting system is above 94% for the case of unidirectional, normal speed in case of both the indoor and outdoor scenario. The accuracy is reduced when the people are in fast motion. For the fast motion case, the people pattern cannot be recognized sometimes due the camera's frame rate and thus the count may be mistaken. In case of bi-directional normal situation the system provides better accuracy than the other conventional People CountingSystems. Experimental results show the proposed method provides a dramatic enhancement in the moving object detection aspect to deal with changing background environment. Also it is able to ensure an efficient and robust tracking with merge and split of multi-object. If the occlusion is very low then the performance of our system very high but in case of the high occlusion the performance of the system maydecreases.

5. CONCLUSIONS

This paper presents a bi-directional people tracking and counting algorithm for the people flow in changing background conditions for indoor and outdoor environments. The idea to combine the GMM with a chromatic color model improves the system more effective to detect the shadows.A new counting method is proposed to handle the occlusion problem. This provides an accurate and reliable counting of people who pass through the detection area of the camera. This system can be very much helpfulin video surveillance application.It is robust and efficient in handling real-life video sequences of different scenarios. Sometimes, when people are moving with very fast speed,it seems to be difficult for tracking and appeared in few frames only; these are the most significant reason of confusing the counting. In addition, the clothes of people may puzzle the tracking for each person. The work can be further extended to improve the accuracy of counting byincreasing the frame rate process and using a human detection approach.

6. ACKNOWLEDGEMENT

This research is being carried out at Central Electronics Engineering Research Institute (CEERI), Pilani, Rajasthan, India as a part of project "Supra Institutional Project on Technology Development for Smart Systems". Authors would like to thank the Director, CEERI for his active encouragement and support.

7. REFERENCES

[1] Honglian, M., Huchuan, L. Mingxiu, Z., "A Real-time Effective System for Tracking Passing People Using a Single Camera", Proceedings of the 7thWorld Congress on Intelligent Control and Automation, Chongqing, China, pp. 6173-6177, 2008.

[2] Wren, C., Azarhayejani, A., Darrell, T., Pentland, A. P., "Pfinder: real-time tracking of the human body", IEEE Trans. on Pattern Analysis. And Machine Intelligence, vol. 19, No. 7, pp. 780-785, 1997.

[3] Lo, B.P.L. Velastin, S.A., "Automatic congestion detection system for underground platforms", Proc. Of Int. Symposium on Intelligent Multimedia, Video and Speech Processing (ISIMP2001), pp. 158-161, 2001.

[4] Stauffer, C., Grimson, W. E. L., 1, "Adaptive Background Mixture Models for Real-Time Tracking", Proceedings of conference on Computer Vision and Pattern Recognition (Cat. No PR00149), IEEE Computer Society Vol. 2, pp. 246-25, 1999.

[5] Elgammal, A., Hanvood, D., Davis, L.S., "Nonparametric model for background subtraction", Proc. Of European Conf. on Computer Vision (ECCV 2000), pp. 751-767, 2000.

[6] Koller, D., Weber, J., Huang, T., Malik, J., Ogasawara, G., Rao, B., and Russell, S., "Towards Robust Automatic Traffic Scene Analysis in Real-time", Proc. Of Int. Conf. on Pattern Recognition (ICPR'94), pp. 126-131, 1994.

[7] Horprasert, T., Harwood D., Davis, L.S., "A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection", in IEEE FRAME-RATE WORKSHOP of International Conf. on Computer Vision (ICCV'99), 1999.

[8] Schofield, A.J., Mehta, P.A., Stonham, T.J., "A System for Counting People in Video images using Neural Networks to identify the Background scene", Journal of Pattern Recognition, Vol. 29, Issue no. 8, pp. 1421-1428, 1996.

[9] Terada,K., and Kurokawa, N., "A Method of Counting the Passing People by Using the Method of the Template Matching", IAPR Workshop on Machine Vision Applications, Makuhan, Chiba. Japan, pp. 498-501, 1998.

[10] C.H. Chen, Y.C. Chang, T.Y. Chen, D.J. Wang, "People Counting System for Getting In/Out of a Bus Based on Video Processing" IEEE Computer Society, Eighth International Conference on Intelligent Systems Design and Applications, pp. 565-569, 2008,.

[11] Hartono, S., Ji T., Yap-Peng, T., "People Counting by Video Segmentation and Tracking", 9th international Conference on Control, Automation, Robotics and Vision, pp. 1-4, 2006.

[12] Tsong-Yi, C., Chao-Ho, C., Da-Jinn, W., Tsang-Jie, C., "Real-Time Counting Method for a Crowd of Moving People", Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 643-646, 2010.

[13] Kim, J. W., Choi, K. S., Choi, B. D., Ko, S. J., "Real-time Vision-based people counting system for security door", International Technical Conference on Circuits/Systems Computers and Communications, pp. 1416-1419, 2002.

[14] Javier, B., Berta, M., Fernando, B., "Real-Time People Counting Using Multiple Lines", IEEE Computer Society, Ninth International Workshop on Image Analysis for Multimedia Interactive Services, pp. 159-162, 2008.

[15] Huazhong, Xu, Pei, L., Lei, M., "A People Counting System based on Head-shoulder Detection and Tracking in Surveillance Video", International Conference on Computer Design and Applications, pp. VI-394 –VI-398, 2010.

[16] Chunhui, Z., Quan, P., Stan, Z.L., "Real Time People Tracking and Counting in Visual Surveillance", Proceedings of the 6th World Congress on Intelligent Control and Automation, Dalian, pp. 9722-9724, June 21 – 23, 2006.

[17] Bescos, J., Menendez, J. M., Garcia, N., "DCT Based Segmentation Applied to a Scalable Zenithal People Counter", IEEE International Conference on Image Processing (ICIP), vol. 3, pp. 1005-1008, 2003.

[18] Enwei, Z., Feng, C., "A Fast and Robust People Counting Method in Video Surveillance", International Conference on Computational Intelligence and Security pp. 339-343, 2007.

[19] Dar-Shyang, L., Jonathan, J.H., Berna, E., "A Bayesian framework for Gaussian mixture background modeling", International Conference on Image Processing, pp.973-976, 2003.

[20] Fredrik, K., Peter, N., and Viktor, O., "Background Segmentation beyond RGB", in Proceedings of Asian Conf. on Computer Vision (ACCV 2006), Hyderabad, India, pp. 602-612, 2006.

[21] Alessandro, L., Cosimo, D., Francesco B., "A Shadow Elimination Approach in Video-Surveillance Context", Pattern Recognition Letters 27, pp. 345-355. 2006.

[22] Gonzalez, R. C., and Woods, R. E., "Digital Image Processing", Prentice Hall, 2nd Edition 2007.

[23] Cazan, I., "Kalman Filters," available at, www.colby.edu/math/program/cazan-honors.pdf, pp. 1–2, 2011.