An Insight into the Algorithms on Real-Time People Tracking and Counting System

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

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

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

Solanki Lovendra
BKBIET, Pilani, India

ABSTRACT
People tracking and counting system is being widely used in modern days in the video surveillance for security purpose. In the recent, many algorithms have been proposed and developed in designing the system. But there always been confusions of using these algorithms based on their limitations and benefits. This paper provides a review on the methods that are used in the people tracking and counting system considering their limitations, benefits, speed and accuracy. In designing an efficient real time people tracking and counting system, designers can be effectively guided by this review work.

Keywords
object detection, background modeling, tracking, counting, counting lines.

1. INTRODUCTION
People tracking and counting system is of great importance for surveillance purpose. At past, people involved in the counting process and later with the development of electronic devices, the counting process was somewhat reliable. But both of them were not fully efficient, as there contained errors, and also tracking could not be done with those processes.

So, to sort out those difficulties, computer vision techniques have been used in the people counting system to give proper counting process and tracking. People tracking and counting process involves many phrases such as detection of moving objects, feature extraction, morphological processing, tracking and counting. For every process, there involves many algorithms to carry out the processes efficiently. Both the expert and the newcomer to this area can beconfused about the benefits and limitations of each step. This paper provides a thorough review of the main methods and their advantages and disadvantages based on speed and accuracy.
The design process of people counting system has been described as the block given below. Different researchers use different algorithms for their design. The blocks shown consist of the algorithms that are widely used in the tracking and counting process.

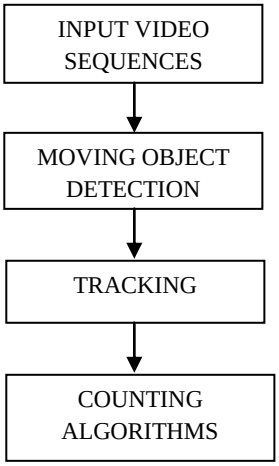

Fig. 1 People tracking and counting system

As seen from the above blocks, a short description of the people counting system can be given as that the real-time input video is taken from a camera. This video is further processes to detect to moving object. Then features are extracted from the processed video for tracking and counting purpose.

This review work is organized into five sections. Section 2 describes the background subtraction techniques. Various tracking algorithms are described in Section 3. Section 4 describes the counting process. In Section 5, a proposed model has been described based on the advantages and limitations of these algorithms. Conclusive remarks are described in Section 6.

2. MOVING OBJECT DETECTION
In a video sequence, one of the important steps is to detect the moving objectin the scene from static cameras.Several methods for performing moving the object detection have been proposed. All of these methods are used to estimate the background model effectively from the temporal sequence of the frames.In this section, different algorithms for detecting the moving objects have been discussed. Some of the widely used object detection techniques are as:

2.1 Frame Difference

Xiao Benxianet al. [1] used the frame difference method to detect the motion of the people moving in a video scene. His works also stated that this method computes less cost and the simplest motion detection technique.It is the simplest method for detection of moving object.It describes the detection of the foreground objects as the difference between the current frame and an image of the static background of the scene.

$$|frame_i - background| > Th \quad (1)$$

Where *i* represents the current i[th] frame, *background* is the static background frame taken at the beginning and *Th* is the threshold value. In this process, a first frame will be taken, which can be said as background. If the absolute difference is greater than threshold value, then the moving object in the frame will be detected, otherwise it will be neglected.

2.2 Running Gaussian Average

Gaussians are very useful in determining the background model. A proposal was given by Wren et al. in [2] to model the background independently at each (i,j) pixel location. Here a Gaussian probability density function (pdf) is being fitted ideally on the last n pixel's values. The following equation shows the update equation of the background model at the appearance of the frame at time *t*, in where cumulative average of the pixels is computed as:

$$\mu_t = \alpha I_t + (1 - \alpha)\mu_{t-1} \quad (2)$$

whereI_t is the pixel's current value, $\mu_t$the previous average andα is the learning parameter which can be considered as an empirical weight often chosen as a tradeoff between stability and quick update. For quick learning, αshould be in-between 0.5-1.0. As for each pixel, this consisted of two parameters (μ_t ,σ_t) instead of the buffer with the last *n* pixel values, where σ is the standard deviation of the Gaussian pdf. The following equation determines the condition for the foreground and background.

$$|I_t - \mu_t| > k\,\sigma_t \quad (3)$$

If this inequality holds, then at each frame time *t*, the I_t pixel's value can be classified as foreground pixel, otherwise $I_t$is classified as background. Kollerel at.[3] modified the running Gaussian average equation (3) as this equation unduely updated at the occurrence of foreground values. The modified equation is:

$$\mu_t = M\mu_{t-1} + (1 - M)(\alpha I + (1 - \alpha)\mu_{t-1}) \quad (4)$$

where M represents a binary value. If M is 1, it corresponds to the foreground value, and for background, it is 0. This approach is also known as *selectivebackground update.*

2.3 Temporal Median Filter

The performance of the temporal average filter is shown as better in the research works of many authors. A proposalto use the median value of the last n frames was put by Lo and Velastin et al. [4] as the background model. Even though the subsampling of *n* frames with respect to the original frame rate by a factor of 10, Cucchiaraet al. [5] clarified that an adequate background modeling can be provided by such median values.In order to increase the stability of the background model, [5] also proposed the computation of the median on a special set of values containing the last *n* subsampled frames and *w* times the last computed values.

2.4 Mixture of Gaussians

The mixture of Gaussians was proposed by Stauffer C et al. [6].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 the model the value of a particular pixel over time is seen as a measurement, X_t of a stochastic variable. At any time, besides the current measurement of X_t, the history, $M_t = \{X_1, X_2, \ldots\ldots X_{t-1}\}$, of that particular pixel is known [1]. Thus the recent history of a pixel can be modeled using mixture of K Gaussian distributions. The probability to observe the current background pixel X_t is the weighted sum of the K distributions:

$$P(X_t) = \sum_{i=1}^{K} w_{i,t} * \eta(X_t, \mu_{i,t}, \mathbb{C}_{i,t}) \quad (5)$$

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. $\mathbb{C}_{i,t}$ is the covariance matrix of the i[th] Gaussian in the mixture at time t. η is the probability density function of i[th] Gaussian, which is given by:

$$\eta(X_t, \mu_{i,t}, \mathbb{C}_{i,t}) = (1/((2\prod)n/2\,|\mathbb{C}_i|\,1/2)) * exp(-1/2(X_t - \mu_t)\mathbb{C}_i - 1(X_t - \mu_t)) \quad (6)$$

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:

$$\mathbb{C}_{i,t} = \sigma_i^2 I \quad (7)$$

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 come 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.

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha M_{k,t} \quad (8)$$

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:

$$\mu_t = (1-\rho)\,\mu_{t-1} + \alpha X_t \quad (9)\sigma_t^2 = (1-\rho)\,\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \quad (10)$$

Where, $\rho = \alpha\eta(X_t \mid \mu_t, \sigma_t)$

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

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

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

2.5 Kernel Density Estimation

In statistics, kernel density estimation is a non-parametric way of estimating the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample.

Elgammal et al. in [7] proposed a new way of modeling of the background based on Kernel Density Estimation (KDE) on the last ***n*** background values. The sum of Gaussian Kernels centered in *n* background values, x_t, is used to derive background pdf as given:

$$P(x_t) = \frac{1}{n}\sum_{i=1}^{n}\eta\,(x_t - x_t, \sum_t) \quad (11)$$

Piccardi et al. [8] stated that this model seems to be dealing with a sum of Gaussians which is shown in equation (11). However, the main advantage is that each Gaussian describes just one sample data, with ***n*** in the order of 100, and ***Z;*** is the same for all kernels. If background values are not known, unclassified sample data can be used in their place; the initial inaccuracy will be recovered along model updates. Based on (11), classification of $x_t$as foreground can be straightforwardly stated if $P(x_t) < T$.

2.6 Eigen backgrounds

Nuria M. Oliver et al. in [9] built an eigen-space to detect the moving object by modeling the background. In his work, the eigen-space was modeled by taking a sample of N images of p pixels and computing the mean μ_b background images and covariance matrix C_b. Using PCA, dimensionality reduction of the covariance matrix is done to select the M largest eigenvalues matrix of ϕ_{Mb} size *M X p.* Each input image $I_i$can be projected to eigen spaces as which is expanded by the eigen backgrounds images $B_i = \phi_{Mb}X_i$ to model the static part of the scene. Hence, moving object was found out by computing and thresholding the Euclidean distance between the input image and projected image which can be shown as:

$$D_i = |I_i - B_i| > t \quad (12)$$

where*t* is a given threshold. This approach can compensate for the changes such a shadows, climatic change, light intensity etc. It can also offer less computational cost.

3. TRACKING ALGORITHMS

Object tracking is a problem of estimating the positions and other relevant information of moving objects in image sequences. Tracking uses different features such as center of gravity, color, HIS histogram etc. 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.

Tsong-Yi Chen et al. [10] used the center of gravity as a feature to track down the moving objects in the succeeding frames. According to their algorithm, they measured the Euler's distance of the center of gravity in the adjacent frames as shown in the equation 13.

$$Distance = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (13)$$

For the same object to be tracked this distance between the two successive frames must be minimum and less than a threshold and a bounding box is drawn around the object.

The Kalman filter can 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 [11].BerilSirmacek et al. [12] used the Kalman filter for thetracking purpose. Kalman filter helps in the prediction of person in a frame of an image sequence based on the information derived from the previous frames. It uses the parameters of the previous frame and based on the observation value of the current frame, it predicts the next frame.

D. Comaniciu et al. [13] proposed an algorithm towards target representation and localization, the central component in visual tracking of non-rigid objects.Another tracking algorithm used is Mean-shift based object tracking. DorinComaniciu et al. [14] used the Mean-shift tracking algorithm for the non-rigid objects by a single camera. To find the target object similar to the present model, Mean-shift iterations are being employed. The similarity of these two models is expressed by Bhattacharyya coefficient.Given the predicted location of the target in the current frame and its uncertainty, the measurement task assumes the search of a confidence region for the target candidate that is the most similar to the target model. The similarity measure that is developed is based on color information.

At first, target localization is done, then Weighted Histogram Computation and then Distance Minimization using Mean-shift.The tracker employs two independent Kalman filters, one for each direction x and y. The target motion is assumed to have a slightly changing velocity modelled by a zero mean, low variance (0.01) white noise that affects the acceleration. The tracking process consists in running for each frame the mean shift based optimization which determines the measurement vector and its uncertainty, followed by the Kalman iteration which gives the predicted position of the

target and a confidence region. These entities are used in turn to initialize the mean shift optimization for the next frame.

Tao Liu et al. [15] proposed a mean shift object tracking algorithm. A novel weight to the given method is given, which improves the kernel function. The method is that the pixels which near the centre of object are given biggest weights, and the pixels which at the edge of the object are given by exponent distribution as a result of occlusion. In order to handle the occlusion, the occlusion detecting method based on sub-block detecting is also established. The novel sub-block detecting algorithm is that the tracking window is divided into two parts, including right and left, and the similarity measure is calculated respectively.

4. COUNTING ALGORITHMS

To count the number of people entering or leaving a room, different counting algorithms have been developed. For the counting process to be accurate, different techniques are being implemented. In this paper, some of the good counting algorithms are discussed.

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 modeling algorithm to recognize the background and track the moving object. Honglian Ma and Huchuan Lu et al. in [16] 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. [17] 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 [18] 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[19] 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.

Researchers use techniques such as single line counting or multiple lines counting for the counting process. Tsong-Yi Chen et al. [20] proposed a counting algorithm where a single camera is mounted at the top of the room. The algorithm used a single line in the screen as the counting zone. The persons crossing this line are counted as going in or coming out of the room. He took the object's perimeter and area as the most attainable features, for single image, for counting purpose. They have also put a suggestion for distinguishing the overlapping persons when it occurs in the frame. Analyzing the processed image, many features, such as the height, width, area, gravity center, contour, shape and pixels' number, can be obtained. From all these, only a few features of interest are selected to analysis.

As described in [20] if the number of pixels of a particular segmented image is greater than of a threshold value of pixels of one person, then there is a possibility of containing more than one person. A bounding box is drawn for one person and for two combined person, the box is divided into two rectangles.

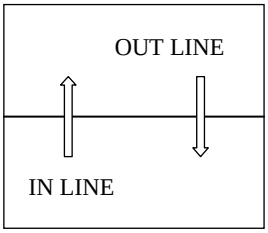

Fig. 2

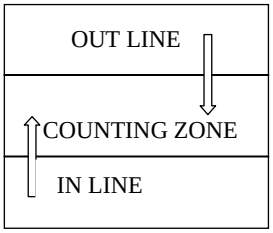

Fig. 3

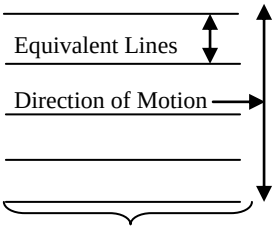

Fig. 4

Fig.2: Single line counting technique Fig.3: Two line counting technique. Fig.4: Multiple lines method

People whoever crossed this In Line as marked by the arrow head are counted as number of persons going in. Similarly, this is for the people going out after crossing the Out Line in the given direction. To detect and track moving people, Kim et al. in [21] 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. [22] used more than two lines, i.e. multiple lines to count the number of people going in or coming out within the range of the camera. The lines are virtual and placed perpendicularly to the direction of the motion of the people.The counting is performed by each line individually and the length must be bigger than half of a person's width. Pixels are getting accumulated when someone crosses the lines. To determine the direction of movement, people, Javier Barandiaran used Lucas-Kanade method around the counting zone.

5. A PROPOSED MODEL

All the algorithms for all the steps of a people counting system have been discussed in above. All these algorithms are widely used. But all these algorithms are not used considering their advantages, speed and limitations. Based on these properties, a model is proposed for the designing of people tracking and counting system.

In the moving object detection algorithm, it cannot be said that all the algorithms are good. The simplest of them is frame difference method. One of the main reasons is that of its modest computational load. Another is that the background model is highly adaptive to the changes in background, as it depends solely on the previous frame, which makes it faster

than any other methods. The frame difference method has also the capability to subtract out extraneous background noise (such as waving of leaves), much better than the more complex approximate median and mixture of Gaussians methods.

For the tracking systems that are discussed, Kalman filter tracking is one of the simplest and fastest tracking algorithms that is used in this system. The advantage of Kalman filter-based techniques lies in their efficiency and in the high accuracy that can be obtained. It is predictive and adaptive as it looks forward with an estimate of the covariance and mean of the time series one step into the future, whereas in mean-shift algorithm, the implementation in software takes a long time which is about 15 fps.

For the counting process, the two line counting process can be found more reliable compared to others. In this system, two virtual lines are drawn such that the distance from bottom edge to the IN LINE and from top edge to the OUT LINE is greater than the bounding box of the person entering or leaving the room. Using all these algorithms, a real time people counting system can be designed based on their accuracy, speed etc.

6. CONCLUSION

In this paper, a review of the algorithms that are widely used in real-time people tracking and counting system have been described. This can help many of researchers to select the appropriate algorithm in designing the system in real-time field based on the advantages and limitations. We have also proposed a system for real time tracking and counting which may be useful to the researchers. Based on the proposed system given above, it gives less computational time and also tracking can be done effectively in real-time purpose.

7. ACKNOWLEDGEMENT

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

8. REFERENCES

[1] Xiao Benxian, Lu Cheng, Chen Hao, Yu Yanfeng and Chen Rongbao, "Moving object detection and recognition based on the frame difference algorithm and moment invariant features", 27th Chinese Control Conference (CCC), pp. 578-581, 2008.

[2] C. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time Tracking of the Human Body", IEEE Trans. on Pattern Anal. And Machine Intell., Vol. 19, No. 7, pp. 780-785, 1997.

[3] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B.Rao and S. Russell, "Towards Robust AutomaicTraffic Scene Analysis in Real-Time", Proc. ICPR'94, pp. 126-131, Nov. 1994.

[4] B.P.L. Lo and S.A. Velastin, "Automatic Congestion Detection System for Underground Platforms", Proc. ISIMP 2001, pp. 158-161, May 2001.

[5] R. Cucchiara, C. Grana, M. Piccardi and A. Prati, "Detecting Moving Objects, Ghosts and Shadows in Video Streams", IEEE Trans. On Pattern Anal. and Machine Intell., Vol. 25, No. 10, pp. 1337-1442, 2003.

[6] Stauffer C, Grimson W. E. L., "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. 246-252, June 1999.

[7] A. Elgammal, D. Harwood and L. S. Davis, "Non-Parametric Model for Background Subtraction", Proc. ECCV 2000, pp. 751-767, June 2000.

[8] M. Piccardi, "Background subtraction techniques: a review", IEEE International conference on Systems, man and Cybernetics, pp. 3099-3104, 2004.

[9] N. M. Oliver, B. Rosario and A. P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions", IEEE Trans. on Pattern Anal. and Machine Intell., Col. 25, No. 11, pp. 1499-1504, 2003.

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

[11] Raul Rojas, "The Kalman Filter", available on www.robocup.mi.fu-berlin.de/buch/kalman.pdf, pp. 1-7.

[12] BerilSirmacek and Peter Reinartz, "KALMAN FILTER BASED FEATURE ANALYSIS FOR TRACKING PEOPLE FROMAIRBORNE IMAGES", Proc. ISPRS XXXVIII.

[13] D. Comaniciu, V.Ramesh and P.Meer, "Kernel Based Object Tracking",IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 5, 564-575, 2003.

[14] D.Comaniciu and P.Meer, "Mean-shift Analysis and Applications", The Proc. of the 7th IEEE International Conference on Computer Vision, Vo. 2, pp. 1197-1203, 1999.

[15] Tao Liu, Xiaping Cheng, "Improved Mean Shift Algorithm for Moving Object Tracking", 2nd International Conference on Computer Engineering and Technology(ICCET), Vol. 1, pp. 575-578, 2010.

[16] Honglian Ma and Huchuan Lu Mingxiu Zhang, "A Real-time Effective System for Tracking Passing People Using a Single Camera", Proc. Of the 7th World Congress on Intelligent Control and Automation, June 25-27, 2008, Chongging, China, pp. 6173-6177.

[17] A. J. Schofield, P. A. Mehta, T. J. Stonham, "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.

[18] 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.

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

[20] Tsong-Yi Chen, Chao-Ho Chen, Da-Jinn Wang and Tsang-Jie Chen, "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.

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

[22] Javier Barandiaran, Berta Murguia and Fernando Boto, "Real-Time People Counting Using Multiple Lines", IEEE Ninth International Workshop on Image Analysis for Multimedia Interactive Services, pp. 159-162, 2008.