International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 8 (2017) pp. 1798-1806

Crowd Surveillance with the Video Captured by Handheld Devices

Chandra Prakash Sahu[1], Yashpal Yadav[2], Vikas[3], Madan Singh[4] and Rajesh Prasad[5]

[1,2,3,4] *Dept. of Computer Science and Engineering, SRM University (NCR Campus), India.*
[5] *School of Information Technology and Computing, American University of Nigeria, Nigeria.*

ORCIDs: [1] 0000-0002-9236-5271, [2] 0000-0002-1799-940X, [3] 0000-0002-5432-5156
[4] 0000-0003-3879-3213, [5] 0000-0002-3456-6980

Abstract

The Gaussian Mixture Model has been widely used to detect foreground in the rapidly changing environment but it fails in low light condition and deteriorates when the background is rapidly changing and it cannot be applied to the video captured by handheld until the frames are stabilized. The proposed work is focused on group detection and people counting in videos where the camera is handheld. We utilize the motion information for each pixel between current and previous frames to prepare a confidence matrix to correctly model background distribution. This work excludes the current distribution of the pixel if the match occurs; in case the pixel is foreground in confidence matrix, reducing the chance of false positive inclusion in background model estimation, and detects around 90% of the foreground pixels. Moreover, the height, width, area of the bounding box of the blobs detected and foreground pixel density are used to distinguish individual and groups, and people counting.

Keyword: Crowd density, Gaussian Mixture Model, monitoring, motion vector and people counting.

INTRODUCTION

The population growth and migration have led to increasing the traffic on already choked streets, malls, markets, religious places etc., affecting our well-being and comfort in these places and often the overcrowding results into mishaps as happened in Allahabad Kumbh Mela 2013[1] and in Love parade in 2010[2]. It is necessary to monitor the crowd density (i.e. individuals and groups), if the density of crowd is increasing then one take appropriate action. Having the information about the areas where Close Circuit Tele Vision (CCTV) camera is not installed, would further fill loopholes and strengthen security. The officials monitoring crowd only rely on their personnel deployed watching the place. They don't rely on the information from common people as this may be misleading. Providing video of the scene and then measuring the situation is reliable even by a common man. The crowd size estimated by the camera can be forwarded to the base stations, where authorities can decide to take proper action e.g. if a person captures the video of a location and forwards the information extracted (such as the place where it captured, the camera, the resolution and the number of people in the video appearing per second) from the video it will require much less bandwidth than a video. Though the same information can also be sent by the human being but it is often exaggerated or undermined depending on the viewers' perspective. Moreover, it is difficult to assess the number of people passing through a point per second. While determining it by using a mobile camera application is easy. This work can be applied to the CCTV videos which capture a scene having background in motion for example the swinging bridges.

Most of the works for people counting are focused on segmenting foreground from the video captured by the stationary camera. But it becomes more difficult if the camera itself is moving. Further, the detection system performance deteriorates if the motion of the camera is not steady. Now, two different kinds of motions are possible: one of the foreground object and other of the whole scene due to camera motion. These motions are jittery in nature. To extract the foreground from the video captured by moving camera, the motion compensation can be used.

This paper proposes an algorithm to determine the crowd size using the mobile camera captured video. So this can be utilized to develop a mobile application. The video captured by handheld (mobile camera or Handycam) is very jittery, and motion is not constant in comparison to the video obtained by wall mounted CCTV. The background information for every pixel is not valid. Some of the CCTV camera moves, e.g. fisheye camera have a background in motion, but the constant (the same frame is repeated over a time period) that makes background detection easier by just storing a separate background model for each frame position. So the requirement is to include the motion information of the background and foreground both. The motion of background includes several new pixels with no background data and excludes some of the pixels at the border. The background information in such cases, about the new pixels, is learned over a few frames.

The algorithm works in two stages first it determines foreground mask and then detects and counts people based on the parameters of the bounding box(BB). The first stage leads in two parallel streams. In the first stream, a confidence matrix is prepared using corner detection and motion compensation, and in second, the Gaussian Mixture Model (GMM) is used to determine the foreground mask. The feedback from confidence matrix in the form of foreground mask is used for background model estimation of GMM. In the second stage, the foreground is detected and the width, height and foreground pixel density of the BB are used to detect groups and count the number of people in the scene. The proposed work restricts itself to the small motion of camera in comparison to the foreground objects.

RELATED CONCEPTS

The most popular background model is GMM [3], which updates the background information for every new frame recursively using k-means approximation. The background information over a number of frames is included with a learning rate α which varies in the range from 0 to 1. It is kept very small and fixed for experiments. Lee et al[4] proposed a different strategy to detect the foreground. They divided the segmentation in to two approaches. In the first approach, they estimate the distribution of all observations, as a Gaussian mixture and then to estimate the probability with which each Gaussian distribution in the mixture is part of the background. Though, the problem of selecting a particular value for learning rate and background segmentation remain unsolved. Prati and Andrea [5] used background modeling and motion segmentation for surveillance. They use k (generally 3 to 5) Gaussian distribution to include variations in the lighting condition of the background and describe the change in illumination of the scene.

In [6] Zivkovic uses a different strategy to adapt changes in light by adding new samples and excluding the older ones. They use a fixed learning rate α to limit the influence of the old data. Moreover, this method uses a more heuristic thresholds featured by the others, leading to slow convergence. Lee et al [7] used a recursive filter for training the mixture and introduced a new variable learning rate for each components of the mixture. Though, the computational complexity of the newer approach was still high as more variables needed to be set for each frame in for computation. Bouwmans et al [8] published a survey on GMM background estimation methods and compared the original GMM with its derivatives. Gorur et al [9] devised a window weight scheme to update the background model, which was suitable to hardware implementation and reduce time complexity. The Expectation-Maximization method used to update GMMs may converge to local maxima, if the main mode stretches. Thus over dominating weaker sections, results in decline of proper detection. Evangelio et al[10] proposed splitting of Gaussians into GMM to solve this problem, by avoiding over dominating modes through an updated formula for calculating the weight function. They show an improvement over the adaptive GMM background subtraction proposed by Zivkovic[6]. Though, the low light condition and motion of foreground and background pixels still affects the accurate background model estimation and thus the selection of foreground pixels. The repositioning of the background pixels in the scene will nullify the learned model for each pixel if motion compensation is not done.

Foresti et al [11] proposed a moving camera surveillance system. They used background compensation to detect moving object in the scene with a mounted camera having steady motion in comparison to the mobile camera. Sheikh et al [12] proposed background modeling by single probability density function. The foreground is detected by using previous detection information of the object and the background model. This model is capable of handling the nominal motion of camera due to strong wind, the ground trembling and also the cyclic motion of the camera. Though there is the motion of both background and foreground but the scene doesn't change as in the case of the mobile camera. Szolgey et al [13] proposed a foreground detection approach using a wearable camera, which produces a video with abrupt motion. They use motion compensation using kernel matching on a reduced image plane. Valestin et al [14] used a reference image with only background data to classify pixels belonging to either foreground as people or background (i.e. nobody) and established a functional relationship between people count and pixel count representing them manually for the measurement of crowd density. Ma et al [15] used background removal and derive a linear relation between the number of pixels and number of persons by applying the geometric correction. In their work, they make an assumption that the number of foreground pixels is proportional to the number of people, which can cause a serious error if people occluding each other. The relationship between the feature histograms and the number of the pedestrian is learned from labeled training data. Chen et al[16] use a two-stage segmentation technique. First, the shadow and background are eliminated from the current frame by taking the difference from previous frames and then the BBs are used to determine the number of people and the direction. They use three thresholds number of pixel per person(TSP), the width of a BB(TW) and height of the BB(TH). If the number of pixels an object occupying is less than TSP, then this object is not counted and hence, removed from calculations. Objects with more pixels than the TSP are assumed as representing one or more objects and then these objects' BB are divided into equal size boxes and then again compared. If the number of pixels is more than 150% of the number of pixel per person then this BB is divided into two sub-regions else it is assumed as one object.

Jung et al [17] used a camera on the robot to detect moving objects in the scene. They use feature sets to detect the motion in two frames and compensated the current frame accordingly. The locations of the human in the scene are detected by Expectation-Maximization (EM) algorithm with adaptive particle filter. The similar approach used by Jung et al [18], for human-robot interaction system. They used motion compensation and disparity segmentation for this purpose. Szolgay et al [13] proposed a more robust method to segment the foreground using a wearable camera, that have abrupt changes in motion than the wall mounted moving camera. They use hierarchical block matching with some basic techniques for motion compensation to reduce the motion blur, which performs better in comparison to the pixel based methods. Though, they reduce the number of pixels in the image plane to apply kernel based estimators, in order to decrease the computational complexity. But it remains computationally expensive in comparison to our proposed method of selecting a single motion vector for all the pixels, from a set of motion vectors in matched corner points in the pair of frames. But our method cannot perform better if used independently from GMM.

Rosten et al [19] proposed a new corner detection method named a feature from accelerated segment test and show the effectiveness of this method in establishing a correspondence between two consecutive frames of a video for tracking the foreground objects. They prove the effectiveness of the corner matching. In our work, we have used the FAST feature to establish a correspondence between two frames, and reposition the background model for the pixels in the current frame.

PROPOSED WORK

The proposed algorithm starts with extracting frames from a video captured by handheld and produces the count of people in the current frame. Figure 1 shows the Foreground Detection Model.

It works in two stages. In the first stage the foreground mask is determined using the motion compensation and GMM and in the second stage, the foreground mask is used to detect the foreground pixels and count the people present using the BB information. The first stage works in two parallel streams first are based on motion compensation to generate a confidence matrix, and in the second stream, the GMM generates a mixture of Gaussians for each pixel. The background model is created by GMM, with the feedback of confidence matrix for the current frame. In the second stage, the final foreground mask produced by the first stage is used to detect the foreground in the current frame and mark the foreground area by a BB. The BB is further considered to detect individual and groups in the frame, and finally, produce the total number of people in the frame.

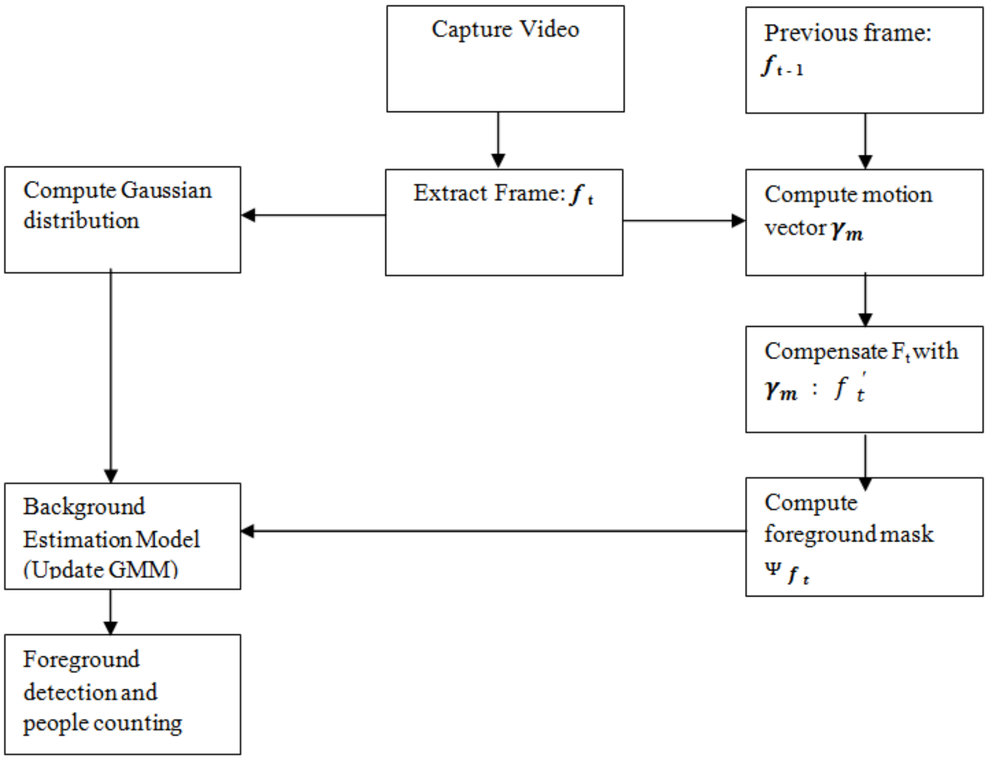

Figure 1. Foreground Detection Model

The detection of the foreground is based on the GMM, which has the capability to include all the major changes in the background distribution, for example: change in illumination, shades, reflection and objects occlusion over each other. This strength becomes a weakness in absence of high-quality video or in low light. The GMM computes Gaussian distribution over each pixel and then matches with the previous distributions of the background available. On matching, it updates the model by including the new distribution and otherwise denoting it as a foreground pixel. The idea behind the proposed work is to have a confidence over the selection of distributions as background.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 8 (2017) pp. 1798-1806

Figure 2. Current frame of escalator dataset

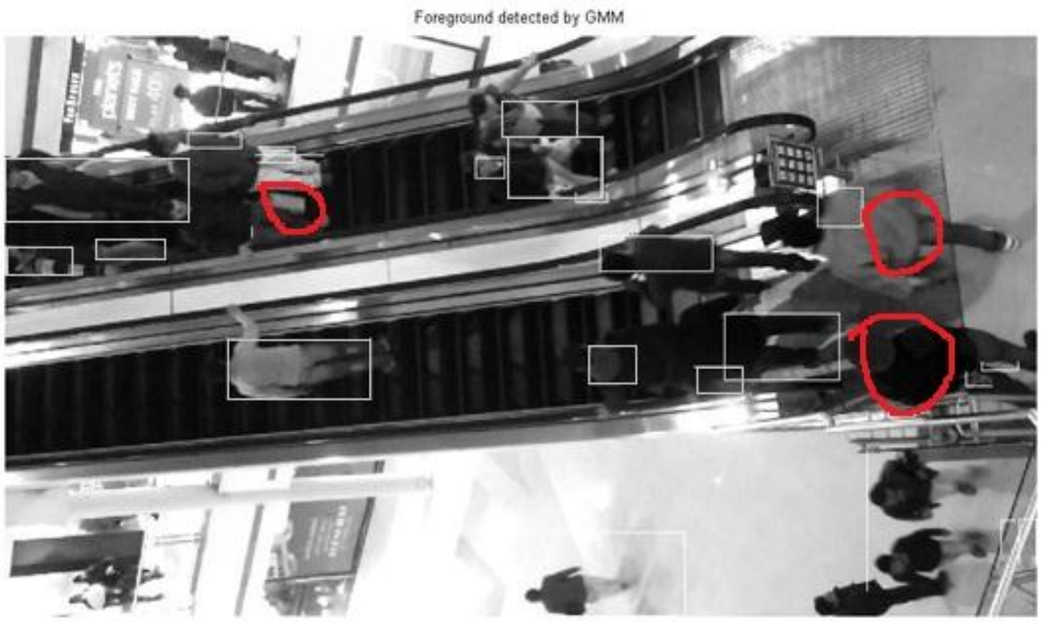

Figure 3. Foreground detected by GMM

Figure 2 shows the current frame *f*t and Figure 3 shows the foreground detected by GMM. Many of the foreground pixels in Figure 3 shown by a red marker which belongs to the people moving on escalator are not being detected as foreground, though, only a portion marked as a BB. Under the marked area, the foreground pixels are confused with the background as at these positions the reflection of light, the shadow of people are changing rapidly and so being included in the background model with GMM.

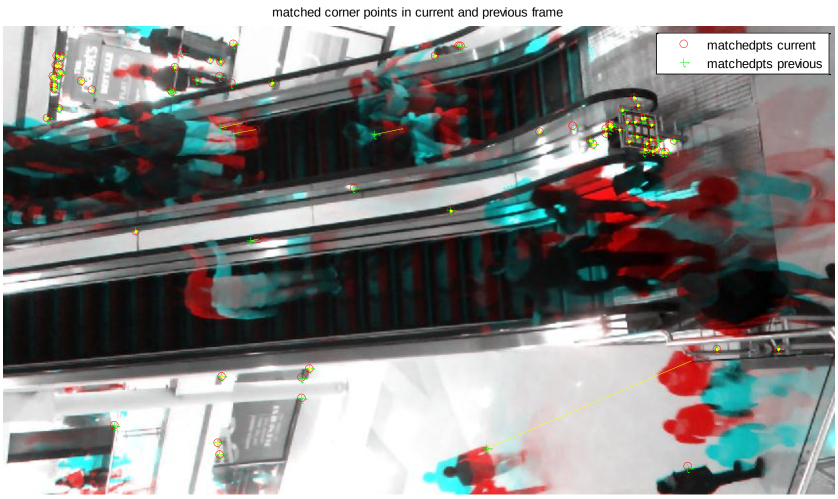

Figure 4. Matched corner points

Motion Compensation

The video captures by people directly using handheld need to be stabilized and for that purpose, the correspondence between important points of the two frames needs to be calculated. This has been done by first detecting FAST features and then matching these features and finally warping by transforming the new frame over old one. The transformation factor is calculated by using the correspondence between matched points. Figure 4 showing the matched corner points in the two frames under consideration, f_t, and f_{t-1}.

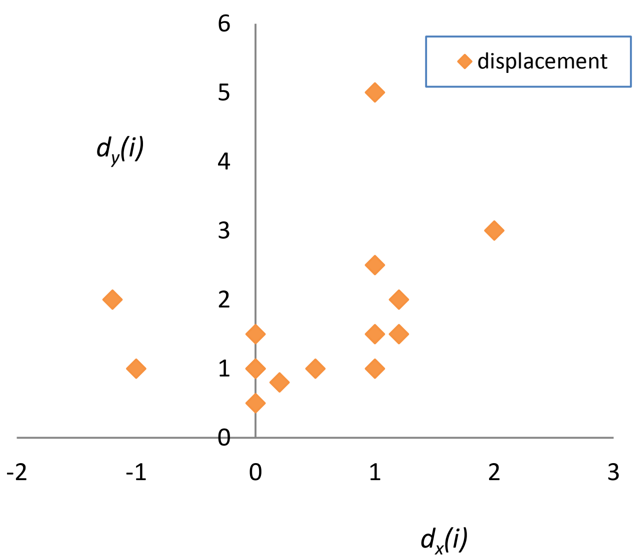

Figure 5. Motion vector for corresponding corner points

Detection and tracking of the corners in the consecutive frames and preparing a foreground/ background mask can be done through the following process-

The video captured by handheld is shaky and there is a small difference in all the frames whether some object is moving or not, though some objects in the scene might be moving in any direction. This produces a difference in the frames ft and ft-1 of background and foregrounds both. The background shows the small difference and foreground shows larger if the foreground is not moving in the opposite direction of background. The motion vector (displacement) D_i, produced by matching the FAST feature in the frames is small for background in comparison to the foreground.

Figure 5 shows the motion vector between corner points in terms of displacements, in the two frames ft and ft-1. It shows a large number of displacements is around the smaller values of $d_x(i)$ and $d_y(i)$ produced by matching corner points in the two frames.

$$d_x(i) = x_{f_t}(i) - x_{f_{t-1}}(i) \tag{1}$$

$$d_y(i) = y_{f_t}(i) - y_{f_{t-1}}(i) \tag{2}$$

Where $d_x(i)$ and $d_y(i)$ are the displacements between the matched corner points in frame f_t and f_{t-1}. A cluster of smaller displacements produces the motion vector for background movement. We use the mean of this cluster as minimum displacement of frame, γ_m, to compensate frame f_{t-1} over f_t .The transformation produces a new frame $f_t{}'$.

$$f_t{}' = \gamma_m + f_{t-1} \tag{3}$$

$$\forall \gamma_m = [\tau_x \quad \tau_y]$$

τ_x and τ_y are displacements in x and y directions.

Figure 6. Compensated frame $f_t^{'}$.

Figure 6 shows the newly generated frame $f_t^{'}$ by transforming f_{t-1} with the translation vector γ_m.

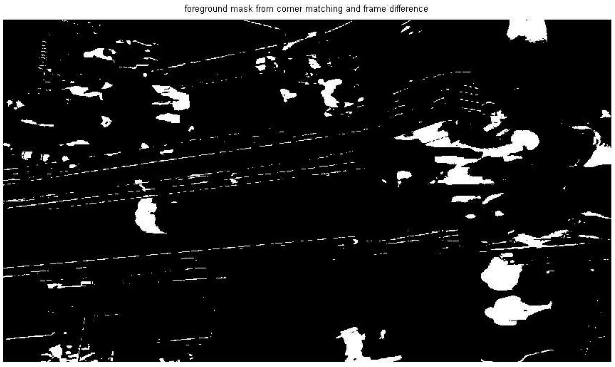

Figure 7. Foreground mask by compensation and difference

The frame difference after compensation is used further to generate foreground mask in Figure 7.

$$\psi_{f_t} = \left[f_t^{'} \sim f_t \right] \quad (4)$$

The foreground mask is created by the following steps to classify pixels—

i. Create a reference frame of the same size
ii. Initialize it with zeros (Background)
iii. If any pixel in the second frame shows motion, mark the pixels as foreground
else mark them as background

Gaussian Mixture Model

GMM is a good model to detect the foreground in the highly changing environment, it provides more measures to clearly distinguish the change in illumination level by reflection, shadow and object motion. The proposed algorithm is based on the work of Zivkovic (2004) and Stauffer (1999) for generating the GMMs for each pixel of the frame and matching the distribution of the pixel of the current frame with the previous values. If the matching is found with great observation and less variance, then the pixel is assumed as background otherwise foreground. The probability of observing a given pixel value Xt at time t is estimated as:

$$P(X_t) = \sum_{i=1}^{k} \left(\omega_{i,t} * \eta \left(X_t, \mu_{i,t}, \Sigma_{i,t} \right) \right) \quad (5)$$

Where k is the number of distributions, $\eta \left(X_t, \mu_{i,t}, \Sigma_{i,t} \right)$ is a normal density of mean $\mu_{i,t}$, and $\Sigma_{i,t}$, is the covariance matrix of ith Gaussian. The components are sorted according to their relevance and background model contains first B components which have a minimum threshold in weight such that

$$B = arg\ min_k \left(\sum_{k=1}^{B} \omega_k > T \right) \quad (6)$$

Where B<= K and T is predefined threshold indicating the minimum portion of the data that should be assumed to be background. For the background pixels the observation is strong while the variants are low and for the object pixels that occlude the background, the distribution will not match to the existing distributions, which will result either in the creation of the new Gaussian or the increase in the variance of an existing distribution. The variance of the moving object is expected to remain larger than a background pixel until the moving object stops. To model this, a method is used to decide which portion of the mixture model represents the background process best. For this purpose, the distributions are ordered on the ratio of evidence and the variance. Here the most probable background distribution remains at the top and the order. The lower order models are updated or replaced.

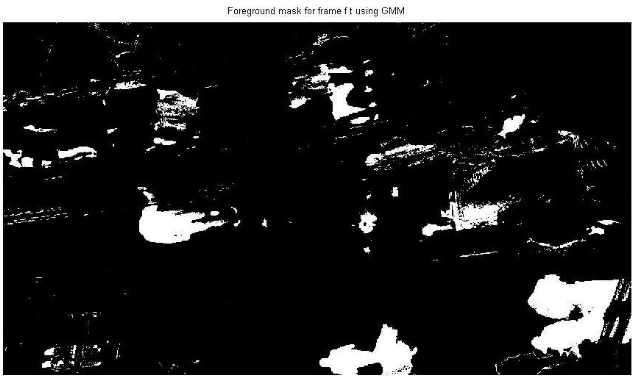

Figure 8. Foreground mask for frame f_t using GMM

The recent history of each pixel $\{x_1, x_2 \ldots.. x_t\}$ is modeled by a mixture of k-Gaussians. The probability of observing the distribution of recently observed values of each pixel in the scene is characterized by a mixture of Gaussians. Every new pixel value is checked against the existing k- Gaussians until a match is found. If none of the k-distributions matched with the current value, the least probable distribution is replaced with a distribution with the current value as a mean value, an initially high variance, and low weight. Then the weights $\omega_{k,t}$ of k distributions at time t are adjusted as-

$$\omega_{k,t} = \begin{cases} (1-\alpha).\,\omega_{k,t-1} + \alpha\,.\,m_{k,t}\,, \forall\ \psi_{i,j,f_t} = 0 \\ \omega_{k,t}\,, \qquad\qquad \forall\ \psi_{i,j,f_t} = 1 \end{cases} \qquad (7)$$

Where α is the learning rate and $m_{k,t}$ is 1for the model which matched and 0 for the remaining models, ψ_{i,j,f_t} is the confidence value of the pixel (i,j) in the foreground mask after motion compensation. After this approximation the, weights are normalized. Mean and variance for unmatched distribution remain same. The parameter for the distribution which matches the new observation and have confidence value 0 for the pixel, are updated as follows-

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho\, X_t \qquad (8)$$

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

Where $\rho = \alpha\, \eta(X_k|\mu_k\,, \alpha_k)$

All the pixels in foreground mask, ψ_{ft}, which are marked as foreground, i.e. $\psi_{i,j,f_t} = 1$ in the corner matching process, they are not included as background distribution in GMM. The values τ_x and τ_y are the number of columns and rows the frame moved and the same number of columns and rows are added to the frame in opposite direction and for these new pixels the, GMM is applied from initial phase. Figure 8 show the foreground mask generated by using GMM for the current frame and Figure 9 depicting the foreground mask generated by our approach for the same frame. In our method, almost 90% of the foreground objects are detected.

Foreground Detection and People Counting

The foreground mask generated by the modified GMM in the second stage, as shown in Fig. 9, is now used to detect the foreground pixels in the current frame, and the x, y coordinates of the pixels connected together are used to detect the BB around the foreground regions. Then width and height of the BB and ratio of foreground/background pixels in the BB are used to count the number of people in the current frame. The method proposed by Chen et al (2012), is used determine the number of people in each BB. Thus the output is the total number of people and groups in the scene. It starts determining the height H_i, width W_i of the BB and the number of foreground/background pixels (NF_i, NB_i in each box). Reject the BB that have very less foreground pixel in comparison to TSP and divide the BB that have 1.5 times TH/TW. Each part again compared with 3 thresholds till they completely accepted or rejected. Finally, the count of BB gives a number of people in the scene.

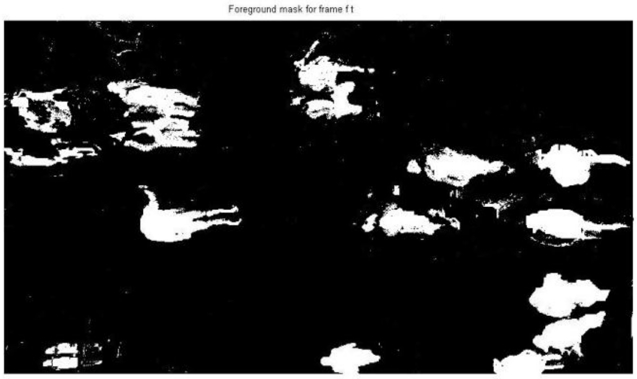

Figure 9. Foreground mask for frame f_t generated by proposed work

EXPERIMENT AND ANALYSIS

The model works on a video of the escalator in a mall, captured by the mobile camera. The resolution of the camera is 720p. Frame interval of 10 (this frame interval was chosen to minimize the chances that consecutive images have the same people at the same location). By measuring the motion vector over every pair of the frames and then compensating the current frame stabilizes the video. The corner detection approach has been used to determine a motion vector. Among all the motion vectors available, the number of the motion vector for the background is much more than the number of motion vectors for the foreground, concentrated in the lower region and almost all have same values. Using this motion vector for compensating all the pixels places the background pixels in the frame f_{t-1} corresponding to the position in f_t. This stabilized frame is then processed against the current frame to generate the confidence matrix which is serving as foreground mask. This makes our method of foreground segmentation faster in comparison to the works of Szolgay et al. Our model does not compare kernel in the search region as in the case of block matching, which is very costly computationally, and this foreground information is used as feedback to correct the GMM.

There are some people at the top of escalators that have not been detected in Figure 2. This happens because of the inclusion of some false positives in the Gaussian mixture. The information from confidence matrix prepared by motion compensation is used in updating GMM background model, thus avoiding false inclusion of the wrong background distribution. This modified GMM is now used to generate foreground mask which is used with the current frame to detect the foreground objects. The Gaussian distribution of each pixel is matched with a mixture of k Gaussians; a match is defined as a pixel value within 2.5 standard deviations of a distribution. The learning rate α is kept 0.05. The foreground mask in Figure 9 shows the result of our method, where more number of pixels are detected as foreground in comparison to the GMM mask in Figure 8 based on Zivkovik (2004).

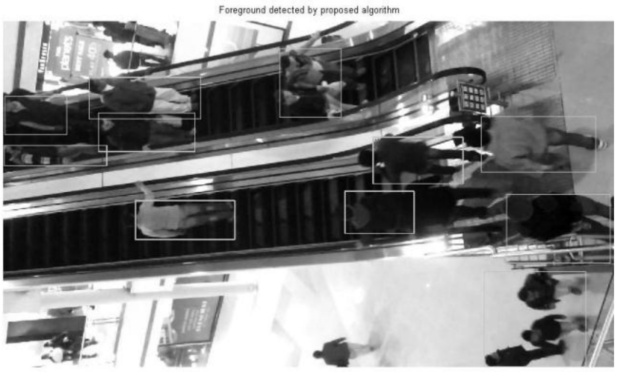

Figure 10. Output of proposed work

If a static object is present in the scene and does not add to the background model, for example, the moving steps can be detected as foreground objects due to different illumination at different locations, leading to accumulated errors in the foreground estimation, thus it may affect the result in tracking. By motion tracking, it can soon be determined if some of the objects are not moving. Motion tracking using FAST considers foreground object to only those points that are present in a pair of frames and steps of the escalators have different illumination level when it's horizontal and when vertical. So no two frames can have a step with same feature points. Figure 7 shows the mask obtained by motion compensation. It presents a clear difference from the mask generated by GMM in Figure 8 without using motion compensation. Figure 9 is the result of using motion compensation as feedback for updating GMM. Figure 10 shows the output of our work using the BB around the people and group detected in the current frame.

CONCLUSION AND FUTURE WORK

The proposed method improves foreground detection scheme in highly changing environments, especially handling the situation where the camera is not stable. The confidence matrix used as feedback to background model excludes the Gaussian distribution of pixel which matched with the Gaussian mixture for the pixel but marked as foreground in confidence matrix. This results in the reduction of the chance of false positive inclusion, i.e. false background distribution inclusion in the background model. Our approach is faster than kernel matching method, proposed by Szolgey et al (2011), to compensate frames and detect foreground, where the kernel size or the search region affects the time complexity of foreground detection. Our approach does not search for motion compensation but just compensates with the most probable motion vector for the whole frame, thus reducing the number of comparisons in the search region for kernel matching.

Our approach detects around 90% of the people in the crowd as shown in Figure 10 and marks the group with the same accuracy. It produces a more accurate foreground mask in comparison to the method proposed by Zivkovik (2004) when applied to the video captured by the mobile camera.

REFERENCES

[1] Allahabad stampede kills 36 Kumbh Mela pilgrims. *Reuters*. 2013-02-11. Retrieved 2013-02-12.

[2] Helbing, D. and Mukerji, P. (2012). Crowd disasters as systemic failures: Analysis of the Love Parade Disaster. *EPJ Data Science* 1:7 (2012).

[3] Stauffer C., and Grimson W. (1999). Adaptive background mixture models for real-time tracking. *In Proceedings CVPR*, pp. 246.252.

[4] Lee, D. S., Berna, E., and Hull J. J. (2002). Segmenting people in meeting videos using mixture background and object models. *Advances in Multimedia Information Processing-PCM 2002*, no. ISBN 978-3-540-00262-8, pp. 393–401, 2002.

[5] Prati, A., Mikic, I., Trivedi, M.M., and Cucchiara, R.(2003). Detecting moving shadows: algorithms and evaluation. *In Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 25, no.7, pp.918-923, July 2003.

[6] Zivkovic, Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. *In Pattern Recognition, ICPR 2004, Proceedings of the 17th International Conference on,* vol.2, no., pp.28-31 Vol.2, 23-26.

[7] Lee, D.S., (2005). Effective Gaussian mixture learning for video background subtraction. *In Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 27, no.5, pp.827-832, May 2005 doi: 10.1109/TPAMI.2005.102

[8] Bouwmans T., Baf F. E., and Vachon, B. (2008). Background modeling using mixture of Gaussians for foreground detection-a survey. *Recent patents on Computer Science*, Vol. 1, pp. 219-237.

[9] Gorur, P., and Amrutur, B. (2011). Speeded up Gaussian Mixture Model algorithm for background subtraction. *Proceedings of 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS),* 2011, pp.386-391.

[10] Evangelio, R.H., Patzold, M., and Sikora, T. (2012). Splitting Gaussians in Mixture Models. *Proceedings of 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*,pp.300-305, 18-21 Sept. 2012.

[11] Foresti, G.L., and Micheloni, C. (2003). A robust feature tracker for active surveillance of outdoor scenes. *Electronic Letters of Computer Vision and Image Analysis, 2003*, 1(1):21–34.

[12] Sheikh, Y., and Shah, M. (2005). Bayesian modeling of dynamic scenes for object detection. *IEEE Trans Pattern Analysis and Machine Intelligence, 2005,* 27(11):1778–1792.

[13] Szolgay, D., Benois, P.J., Megret, R., Gaestel, Y., and Dartigues, J.F. (2011). Detection of moving foreground objects in videos with strong camera motion. *Pattern Analysis and Applications*, 14:311–328.

[14] Yin, J., Velastin, S., and Davies, A.(1995). Image Processing Techniques for Crowd Density Estimation Using a Reference Image. *Proceedings of 2nd Asian Conference on Computer Vision*, 1995, pp 489-498.

[15] Ma, R., Li, L., Huang, W., and Tian, Q. (2004). On pixel count based crowd density estimation for visual surveillance. *IEEE Conf. Cybernet. Intell. System,* 2004, pp 170-173.

[16] Chen C. H., Chen, T.Y., Wang, D.J., and Chen T-J. (2012). A cost effective people counter for a crowd of moving people based on two stage segmentation. *Journal of Information Hiding and Multimedia Signal Processing,* 3(I): 2073-4212, 2012, Vol 3, no 1 pp 12-23.

[17] Jung, B., and Sukhatme, G.S. (2004). Detecting moving objects using a single camera on a mobile robot in an outdoor environment. *In: the Conference on Intelligent Autonomous Systems*, Amsterdam, pp. 980–987.

[18] Jung, H.A., Cheolmin, C., Sooyeong, K., Kilcheon, K., and Hyeran, B.(2008). Human tracking and silhouette extraction for human– robot interaction systems, *Pattern Analysis and Applications (published online)*, June 2009, Vol. 12, issue 2, pp 167-177.

[19] Rosten, E., and Drummond, T. (2005). Fusing Points and Lines for High-Performance Tracking. *Proceedings of the IEEE International Conference on Computer Vision*, Vol. 2 (October 2005): pp. 1508–1511.