Background Modeling and Subtraction Based People Counting for Real Time Video Surveillance

Rakesh Kumar, Tapesh Parashar, Gopal Verma

***Abstract* - This paper presents a multiple people counting using only single camera, entering or exiting in a region of interest. Sigma-Delta background modeling and subtraction is used to segment the people from region. Background subtraction provides blobs as a result with connected component. CG of blobs was used to track and count the entry and exit of people. This paper also proposed method for counting two or more people based on blob size and distance between floor and camera. People will be counted if and only if they passed through the three regions consecutively. Experiments shows that proposed method is robust and provides approximate accurate counts.**

***Keywords* – Edge based Local Thresholding, Background Subtraction, Mathematical Morphological Processes, Connected Component.**

I. INTRODUCTION

People counting is usable in many applications. People counting and monitoring people flow pattern will be beneficial in developing market strategies and surveillance purpose. Many systems are available in a variety of implementation and technologies such as infrared rays, binocular camera and image sensor system. Different systems have different complexities.

Infrared and laser systems can't accurately count people entering simultaneously. Image processing systems have problems like quick illumination changes and occlusion, which inaccurately detect and count people. A network of image sensors can handle small crowds and count in real [1]. Although these systems have good reliability rate [2], these systems are too expensive because of its hardware architecture. Such systems will not be appropriate for small scale applications. If system is using a single camera, vision based systems are cheaper and with algorithms provide proper accuracy.

Automatic people counting using single camera is not a simple task, because of people detection problem. The problem is due to different shapes and sizes of people makes this challenging. Segan and Pingali [3] proposed a method which extracts and tracks pedestrian. Proposed system fails to deals in real time scenario because of its heavy algorithm. However, this system also assumed only simple movement of pedestrians. Huwer et al. [4] proposed a method of combining a temporal difference method with an adaptive background model subtraction scheme to deal the lighting changes. However these methods cannot adapt to quick image variations such as a light turning on or off.

Manuscript Received on November, 2012.
Rakesh Kumar, Director, KS Jain Engineering College, Modinagar, U.P., India.
Tapesh Parashar, Associate Professor, Vishveshwarya Institute of Technology, G.B.Nagar, U.P., India
Gopal Verma, Associate Professor, Sunderdeep Engineering College, Ghaziabad, U.P., India.

Although many researchers have focused on the background subtractions, few papers can be studied for foreground as moving object, shadow and ghost by combining the motion information. Another counting method was proposed by Terada [5] using a stereo camera which segments the human region and road region by using three dimensional data. Stauffer and Grimson [6] and Elgammal et al. [7] were the first in using advanced density estimation techniques on pixel-level; Gaussian Mixture Model (GMM) and non-parametric techniques respectively. No global difference models were used. In [8], Toyama et al. present the Wallflower back-ground estimation, which estimates multiple backgrounds (two in their experiments) and switches between the learned backgrounds when a large number of pixels are detected as foreground, for example due to different illumination conditions. This is a very pragmatic approach, but it does only provide a solution for two states, which also have to be trained in advance. In [9] adaptation is done by modified Kalman filtering for each pixel of the background image.

This paper proposes a real time people counting method with a single overheaded camera for server access or security inside the building. Assumptions are made that system will used for indoor applications and lighting conditions are satisfactory without big change. In section 2, the hardware architecture is explained. In Section 3, the algorithm for proposed method is explained. The results have presented in Section 4. Concluding remarks are given in Section 5.

II. HARDWARE ARCHITECTURE

Hardware architecture plays a vital role in people counting using a single camera. To successfully calculate the exact count in the indoor in region of interest near the security gate, single IP-Camera was hung from ceiling near the gate

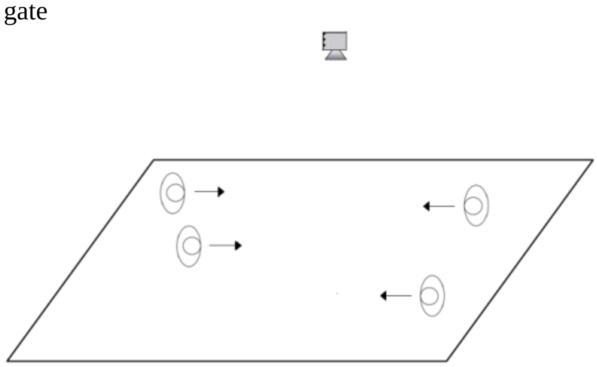

Figure 1: System Architecture

with proper lighting conditions so that passing person can be tracked in region of interest. Figure-1 shows a scene of the camera setup. The people are calculated from processed frames after capturing the images.

III. DESCRIPTION OF PROPOSED APPROACH

In live streaming captured in hardware architecture as explained above, Shape of the People will be seen like the figure-2. Shapes contain certain fix shapes and sizes because of fix distance between the camera and floor. The person size, an input parameter to the algorithm was calculated from an automatic method. It can be calculated from the area covered by persons in monitored ROI. So min, max and mean size of a person should be calculated initially. Other one time activity is background generation. Then it should be updated frame to frame to handle the lighting change. Difference images are resulted after background subtraction and thresholding.

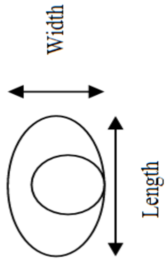

Figure 2: Human Shape from overheaded camera

People counting is a dynamic phenomenon in which people enters from one direction in ROI and exits from other side. There are different possibilities that two or more than two person can enter at a time. They can move in same or opposite direction. As shown in figure-3, person should be counted only if it crosses from three lines to confirm that it is coming from one direction and leaving from other. The proposed algorithm can be divided in the following phases:

- Background detection and subtraction
- Blob generation
- Blob tracking
- People counting

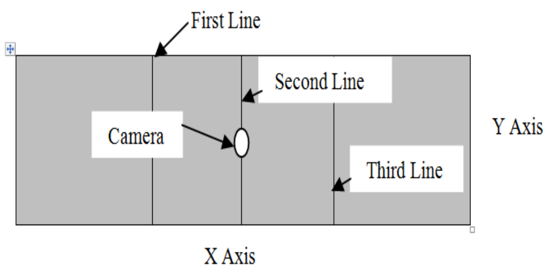

Figure 3: ROI with crossing lines

The proposed algorithm follows the steps shown below:

Algorithm: Main steps of the Algorithm

1. Capture video stream from overheaded camera.
2. Generate Background after hardware installation (one time activity).
3. Update background with slow speed to handle luminance change.
4. Subtract background from current frame.
5. Binarize subtracted frame using Edge based local thresholding.
6. Blob Correction using Morphological processes.
7. Blob tracking in Region of interest.
8. Count Blob if it passes from three lines

A. Sigma-delta Detection and Subtraction Model

Background subtraction is a very common technique for detecting moving objects from image sequences. It extracts moving objects as the foreground elements obtained from the difference image between each frame and known as background model of the scene. The sigma-delta background estimation is a simple method for background generation based on comparison and elementary increment/decrement. In this paper, authors have used the sigma delta method to detect the people based on motion detection.

Table-2 describes the basic sigma-delta algorithm from Manzanera et al. [10]. In sigma delta subtraction, the input signal is value of every pixel over time I_t, which gives first sigma delta operator B_t. Second operator variance V_t is computed with $|I_t-B_t|$. In this algorithm B_t and variance V_t are updated every frame, comparing with current image I_t and current difference Δ_t respectively. The only parameter to be adjusted is an amplification factor N with typical values between 1 and 4.

Algorithm: Steps of the Background estimation algorithm

Step 1: B_t estimation

$B_0 = I_0$ //Initialize Background Model B

$V_0 = 0$ // Initialize Variance V

for each pixel

if $B_{t-1} < I_t$ then $B_t = B_{t-1}+1$

else if $B_{t-1} < I_t$ then $B_t = B_{t-1}-1$

else $B_t = B_{t-1}$

Step 2: Δ_t estimation

for each pixel

$\Delta_t = |B_t - I_t|$

Step 3: V_t update variance

for each pixel

if $V_{t-1} < N * O_t$ then $V_t = V_{t-1} +1$

else if $V_{t-1} < N * O_t$ then $V_t = V_{t-1} - 1$

else $V_t = V_{t-1}$

B. Blob Generation

After Background subtraction, blobs were generated using edge based local thresholding method. Edge-Based local thresholding was chosen because of its good segmentation results. There are the possibilities of same background and people colour; this method maximizes the possibility of exact blob. The threshold value was calculated using edge density in local region. For all pixels, there are different thresholds. We can choose any mask of 3-by-3 or 5-by-5 or 7-by-7. If there is 5-by-5 mask then we can calculate thresholds by counting edges in that mask. If there are edges more than calculated threshold then give that pixel 255 otherwise give that pixel 0 value. There are two defined thresholds for max and min edge density and otherwise thresholds can be calculated by calculating mean of edges.

C. Morphological Operation

Once we got blob, there are the possibilities of some distorted edges. However the low-level morphological operations don’t guarantee that each person translates to exactly one blob, but we can get smooth and generic shape for persons. We are using edge based local thresholding, so we can get some spots inside the blob. In this step we have used opening and filling operations. After this step we will receive the filled and smooth blob.

D. Connected Component

Connectivity between pixels is a fundamental concept simplifies the definition of numerous objects such as regions and boundaries. Connected components labeling scans the labeled array and groups its pixels into regions based on its connectivity and label. Once all groups have been determined, each pixel is labeled with a region number according to the region it was assigned to. Connected component labeling was principled on scanning an image, pixel-by-pixel from top to bottom, left to right in order to identify connected pixel regions. After applying connected component labeling, to track the blob the following properties for each blob should be stored: blob area, current and previous two CG locations in (x, y).

E. Tracking and Counting

This final step has two steps, first one is tracking and second one is counting. Now After processing above steps, we have one blob or more blobs. The Parameter required for starting the tracking and counting is person size bound the upper bound only. To track and count those blobs we are using their CGs (Centre of Gravity). Each CG should be tracked in ROI which shown in Figure-3 with three tracking lines. Count should be increased in that direction only if CG passed through all three lines and leaves that ROI. Count should be managed in both the direction Left-to-Right and Right-to-Left.

First and third lines represent the entry/exit line respectively. If there is need we can interchange status for the entry and exit line. Tracking performed by computing the distance between all the blobs from the vector of the previous frames and the vector of the current frame. If blobs size is greater than person size bound then count for that blob will be two. There are some following possibilities:

- Blob of single person entering from entry/exit gate and exiting from exit/entry gate.
- One blob for two persons entering from entry/exit gate and exiting from exit/entry gate.
- Two blobs for two different persons entering from entry/exit gate and exiting from exit/entry gate.
- One blob appears from entry/exit gate and other blob appears from exit/entry gate and combines at middle line for some frames.

In proposed algorithm, there is need to track and count the blobs with their CGs (x, y) coordinates. The counting was performed by tracking the change in (x, y).

IV. RESULTS

Proposed system was initialized with server access control system, which was responsible to assure the one entry to one card swap. The camera was located at the same height as the top of the door. The automatic counting was performed for bidirectional flow. System was tested with different scenarios and multi 24*7 hours testing time. The real time counting was performed with 3.06 GHz I3 machine and system was running at 15 fps on 640*480 pixels images. The system achieved 99% accuracy rate for estimating people flow.

V. CONCLUSION

In this paper, we have proposed a robust algorithm for multiple people counting using only single optical camera entering or exiting a region of interest. The system can handle the situation, when a blob goes back and forth from region of interest before exiting to opposite side of entry. If person is not identified in a frame than it considers previous and succeeding frames data. Experiments shows that proposed method is robust & provides approximate accurate counts. The applicability of the method in a surveillance environment was quantitatively demonstrated with a live feed taken in a constrained environment.

REFERENCES

1. S. Stillman, R. Tanawongsuwan, and I. Essa, “Tracking multiple people with multiple cameras”, College of Computing, GVU C enter, Georgia Institute of Technology Atlanta, GA 30332-0280 USA
2. H. Goldstein, “People counting using image processing”, Technion - Israel Institute of Technology, Department of Computer Science, Advanced Programming Lab B. July 1999.
3. Segen, J. and Pingali, S., “A Camera-based System for Tracking People in Real Time”, IEEE Proc. of Int. Conf. Pattern Recognition, 1999, Vol.3, pp.63-67.
4. Stefan Huwer and Heinrich Niemann, “Adaptive Change Detection for Real-Time Surveillance Applications”, Third IEEE International Workshop on Visual Surveillance - VS-2000, Dublin, Ireland.
5. K. Terada, D. Yoshida, S. Oe, J. Yamaguchi, “A counting method of the number of passing people using a stereo camera”, IEEE Proc. o f Industrial Electronics Conf., Vol. 3, pp.1318-1323, 1999.
6. C. Stauffer and W. Grimson, “Adaptive background mixture models for real-time tracking,” presented at the IEEE Computer Society Conf. Computer Vision and Pattern Recognition, June 1999.
7. A. M. Elgammal, D. Harwood, and L. S. Davis, “Nonparametric Model for Background Subtraction”. In Proc. Of the 6th European Conf. on Computer Vision-Part II, pages 751–767, 2000.
8. K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, “Wallflower: Principles and practice of background maintenance”, In Int. Conf. on Computer Vision, volume 232, 1999.
9. C. Ridder, O. Munkelt, and H. Kirchner, “Adaptive Background Estimation and Foreground Detection using Kalman Filtering”, In O. Kaynak, M. ¨Ozkan, N. Bekiro˘glu, and ˙I. Tunay, editors, Proceedings of International Conf. on recent Advances in Mechatronics, ICRAM’95, pages 193–199, Istanbul, TURKEY, Aug. 1995. UNESCO.
10. J. Richefeu, A. Manzanera, “robust and computationally efficient motion detection algorithm based on sigma-delta background estimation,” in ICVGIP, IEEE, 2004.
11. Rafael C. Gonzalez, Richard E. Woods, “Digital Image Processing”, Pearson Education, Third Edition, 2009.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication