



AUTOMATIC PEOPLE TRACKING AND COUNTING SYSTEM FOR VIDEO SURVEILLANCE

Mr. Shiva Kumar A, Eslabs Pvt Ltd, Bangalore, India.

Mr. Depesh Kumar, Dhanush InfoTech, Hyderabad, India.

Abstract

In this paper a Computer vision based Tracking and Counting System is proposed which can deal with object occlusion effectively. First an Adaptive Background Estimator which uses pixel averaging technique is employed to obtain a background image and is updated periodically to reduce the effect of illumination and environmental changes. Foreground extractor with a combination of Background Subtraction and Frame differencing is used to obtain a foreground mask. Then using a blob analyzer, object properties like area, centroid etc. is found. A tracker is employed to track the characters using its centroid and size. Counting is done based on the position of centroid and object tagging. Occlusion effect can be minimized by considering the area of an object to detect merged states. The main Objective is to obtain a low computational system which can be possibly used in real time tracking.

Keywords: *Occlusion, Background Estimator, Foreground Extraction, blob, centroid.*

Introduction

PC helped following is important to consequently track and screen the question exercises in video observation. The fundamental issue in following individuals when they are moving in gathering is impediment impact i.e., one question will block the perspective of different items. Thus checking gets to be distinctly troublesome. In this paper, a question following and tallying plan is proposed which can diminish the impact of impediment and enhance the precision of including items a video reconnaissance framework. Routine plans are accessible to

track the human articles. It incorporates recognizing people utilizing highlights like shading, forms, facial points of interest and so forth. Already numbering was done utilizing distinctive sensors like infrared, laser bar sensors and so on. In any case, these strategies are wasteful when individuals are moving in a gathering. A portion of the past works are specified beneath. In a review Wren et al. [1] utilized the components of little blob to track the single human in an indoor situation. A human body was see as a blend of a few blobs individually speaking to different

body parts, for example, head, middle and four appendages. Marcenaro et al [2] utilized Kalman channel for calming the impact of element impediment. The straight channel was utilized to perform label following by coordinating shape highlights. Lien et al [3] proposed a multi-mode strategy to enhance exactness and proficiency for following different focuses in a crowded scene. The items were thought to be in any of six modes like new, leaving, blended, split states and so on. Shape of a gathering is acquired and a system called format working is utilized to include the quantity of individuals the gathering. In another review, Haritaoglu et al [4] recommended the utilization of shapes of human bodies to separate amongst union and split states for packed conditions in open air scene.

Jang and Choi [5] proposed utilizing a dynamic format to manufacture described territorial and auxiliary elements of a question powerfully in light of the data of shape, surface, and shading and edge highlight of the district. Fehr et al. [6] utilized a developed Kalman channel as a part of conjunction with various foundation division techniques to enhance the precision of numbering. In target following plans, the Kalman channel was utilized for anticipating directions of individuals' stream in the back to back casings.

Routine techniques for tallying individuals are, utilization of sensors as said by Lefloch [7]. It incorporates infrared light radiating diodes (LEDs), laser shafts, warm sensors and so forth. These strategies are savvy however have real downsides like when

individuals are moving in gathering these techniques can't gauge precisely. Likewise warm sensors won't work in high temperatures. The above mentioned schemes use features like color, shape and contours to track the targets which rely on object segmentation, change in illumination, occlusion etc. We have to consider the object states like merge and split in high crowded scenes. Here a scheme is proposed where an adaptive. Background Estimation method is used to find the background image. Foreground extraction is done using the background image to obtain the foreground regions. The blobs obtained undergo some morphological operations like erosion and dilation to get a better structure. Later a tracker is used to track and find the merge and split states. A better result in counting is obtained by considering the area and centroid of the objects.

Proposed scheme

The block diagram of the proposed scheme is shown in fig.1. First the input video is preprocessed to remove unwanted noises and to obtain frames. The Background estimator sub-block in below figure estimates the background using first 50 or 100 frames by using pixel averaging technique. For segmenting the foreground regions, the foreground extractor uses background subtraction technique. The segmented objects undergo morphological operations in the next block. During this phase, the noise due to background subtraction is removed. A clean object is obtained after few operations like erosion, dilation and filling of holes are performed.

In blob analysis block, properties like centroid, area and bounding box are found for the labeled image. Then a tracker which is similar to overlap tracker is used. In a modified overlap tracker [8] centroid distances between the objects is considered for finding the merged and split states. The proposed tracker makes use of object area to detect the merged and split states. The objects will be tagged and counted by drawing a virtual line for in and out of objects. When objects cross the virtual line counter will be incremented.

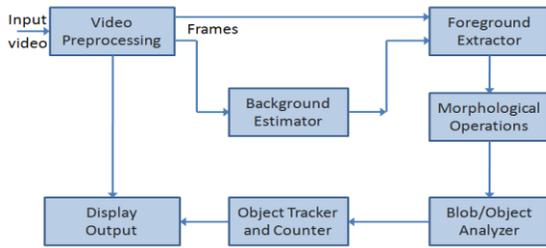


Figure 1. Block Diagram

The steps involved in the proposed scheme are shown in the flowchart below in section VII.

Background Estimator

The Background Estimator uses a background model which is robust against environmental changes in the background and less sensitive to illumination changes. Here an adaptive background model is built by exploring the consistency of pixel gray-level values of consecutive frames. Assume $F_m(x, y)$ denote the gray-level pixel value of the m -th frame, and $B_m(x, y)$ present the corresponding background pixel's gray-level value calculated from previous frames. Each Background pixel is updated using the

following equation.

$$B_m(x, y) = \frac{1}{m} \sum_{i=0}^{m-1} F_i(x, y) \cdot \quad (1)$$

In Eq. 1 'm' is the index of current frame and also denotes the accumulated frames for background pixel averaging. For initially building background model, the starting 50 or 100 frames in a video sequence is used. Later the background image is obtained by taking the averaged value of the pixels and their related background ones. The background image can be updated periodically. This reference image is used further for foreground extraction.

Foreground Extraction

The foreground extractor compares the input video frame with reference image obtained from background model. Foreground regions are extracted using the below mathematical equation. The reference frame is subtracted

$$FG_m(x, y) = \begin{cases} I_m(x, y), & \text{for } I_m(x, y) - B_m(x, y) > \text{Thresh.} \\ 0, & \text{elsewhere} \end{cases}$$

from current frame to get a foreground mask. A foreground mask is a binary image.

In the above equation, $I_m(x, y)$ and $B_m(x, y)$ are gray scale images. $I(x, y)$ is the current frame and $B(x, y)$ is background frame. Here 'm' is the index of current frame. Threshold is some value. By using this equation foreground regions are extracted from the frame. Results of foreground extraction are

shown below. Images on left side of figure 2 shows original frame and images on right side shows the foreground regions extracted.

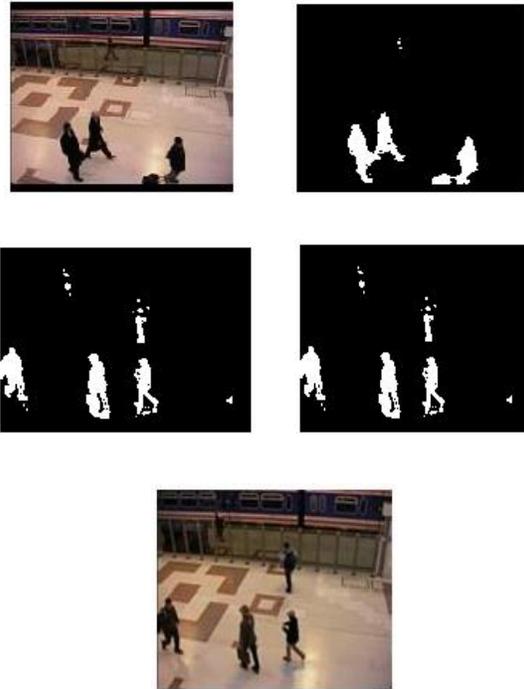


Figure 2: Foreground Extraction.

Morphological Operations

The morphological operation is performed on output video signal coming from Foreground extractor sub-block. The frames are processed in such way that all erroneous blobs residing in the image are eliminated and only the correctly detected blob is retained. The main morphological operations used in this scheme are the *erosion* and *dilation*. The basic morphological operations, erosion and dilation, produce contrasting results when applied to either grayscale or binary images. Erosion shrinks image objects while dilation expands them.

Characteristics of Erosion

- Erosion generally decreases the sizes of objects and removes small anomalies by subtracting objects with a radius smaller than the structuring element.
- With gray scale images, erosion reduces the brightness (and therefore the size) of bright objects on a dark background by taking the neighbourhood minimum when passing the structuring element over the image.
- With binary images, erosion completely removes objects smaller than the structuring element and removes perimeter pixels from larger image objects.

Characteristics of Dilation

- Dilation generally increases the sizes of objects, filling in holes and broken areas, and connecting areas that are separated by spaces smaller than the size of the structuring element.
- With grayscale images, dilation increases the brightness of objects by taking the neighborhood maximum when passing the structuring element over the image.
- With binary images, dilation connects areas that are separated by spaces smaller than the structuring element and adds pixels to the perimeter of each image a proper structuring element has to be determined for erosion and dilation.

In this method each frame was dilated and eroded with a square structuring element. Results are shown below.

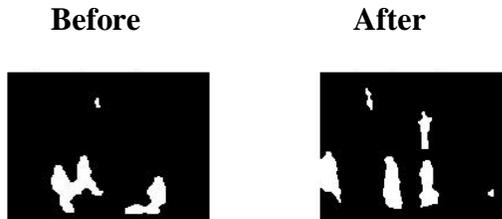


Figure 3: Results of Morphological Operations

Blob Analysis

The main functionality of blob analysis sub-block is to determine the properties of blobs and the maximum number of blobs that are present in each frame. The Properties include area, centroid, Bounding Box etc. of each object or blob in the frame. These properties will be used by the tracker to track a particular blob. Objects of very small area can be neglected because those small blobs cannot be a human. Also a filter can be applied either in morphological sub-block or in this block to eliminate small noises. A minimum size that a blob must have can be specified in order to be correctly tracked. Thus, any blob that doesn't fulfil this condition will not be tracked. Mainly centroid of each blob is used to find their positions.

Flowchart

The steps involved in the proposed tracking and counting system is shown in the flowchart below. The frames go through these stages before the tracking and counting result is displayed.

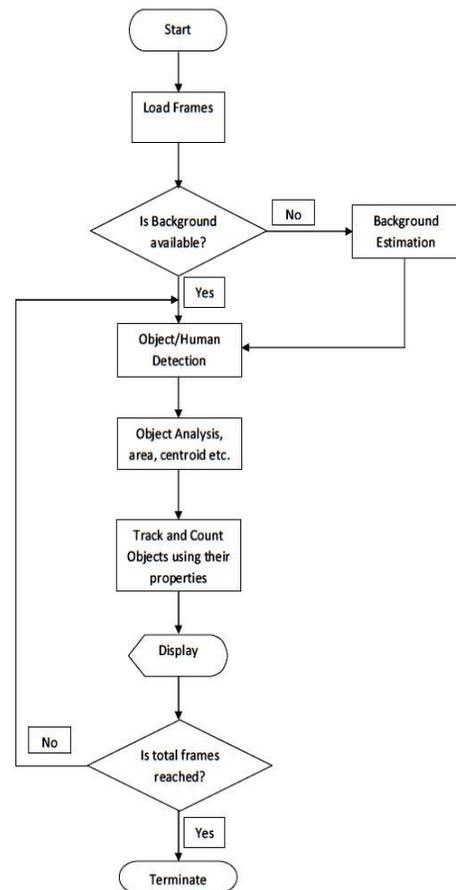


Figure 4: System flowchart

Tracking and Counting

In this sub-block the objects which satisfy certain criteria are further tracked. The tracker uses the object area and spatial coordinates of the characters to identify the merged and split states. Lists are created to group old blobs and new blobs separately. Each object is tagged with a unique identifier. Using the spatial co-ordinates of characters the objects are tracked in consecutive frames until objects move out of scene. For counting, virtual lines are drawn in frame to detect in and out movement of blobs. The line acts as a virtual gate for objects. In 2D world, when an object

approaches a camera the area of the object increases. The occlusion effect can be reduced using this theory. By initially making an observation on areas of individual objects, we can approximately find the area of a single human blob. Based on this area calculation we can decide whether the blob has one or more humans in it. The results of tracking are shown in figure 4. In the figure each object is tagged with a unique identifier.

Experimental Results

The proposed scheme is experimented on video clips obtained from public testing data sets of PETS 2006 [10]. These video scenes were shot in train station. MATLAB is used for implementation of the proposed system.



Figure 5: Tracking and Counting

The result of foreground region extraction is shown in figure 2. Figure 3 shows the result of morphological operations and figure 4 shows the tracking and counting result. In most of the previous works kalman filter is

used for tracking. The disadvantage of using kalman filter is, we need a bank of filters to track multiple objects resulting in more number of computations. Due to this, time required for execution will be very high compared to proposed scheme. Therefore proposed scheme is better than kalman filter technique while tracking multiple objects. Applications of tracking and counting include security management, people management etc. in public places like shopping malls, railway stations etc.

Conclusion

In this work, an automatic computer-aided scheme is proposed to track and count people in a stationary video scene. An adaptive background model is used to construct a background which is less sensitive to environmental changes. Foreground regions are segmented out by comparing the frames with reference frame. To identify the foreground characters, the position and size is considered.

For tracking characters, a special tracker is used which tracks the objects based on area and centroid of characters. This tracker can reduce the effect of occlusion by analyzing the area and detect the merged state of characters. Counting is done by considering a line drawn as a virtual gate, when the objects cross the line, count will be incremented. The direction of objects moving is considered for counting. Therefore the main objective of reducing occlusion effect is fulfilled.

References

1. C.R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland, Pfindex: real-time tracking of the human body, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19 (7) (1997) 780-785.
2. L. Marcenaro etc., "Multiple object tracking under heavy occlusions by using Kalman filters based on shape matching,"
3. C.-C. Lien, Y.-L. Huang and C.-C. Han, "People counting using multi-mode multi-target tracking scheme," *Procc of IEEE IHH-MSP*, pp. 1018-1021, 2009.
4. I. Haritaoglu, D. Harwood and L. S. Davis, "W4:real-time surveillance of people and their activities," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 22, pp. 809-830, 2000.
5. D-S. Jang, H-I. Choi, Active models for tracking moving objects, *Pattern Recognition*, 33 (7) (2000) 1135-1146.
6. Duc Fehr etc., "Counting people in groups," *Proceedings of AVSS*, pp.152~157, 2009.
7. Damien LEFLOCH, "Real-Time People Counting System Using Video Camera".
8. Chih-Chang Chen; Hsing-Hao Lin; Chen, O.T.-C. "Tracking and counting people in visual surveillance systems", *Acoustics, Speech and Signal Processing (ICASSP)*, 2011 *IEEE International Conference*, May 2011.
9. Vincent van der Tuin "Computer Aided security surveillance".
10. V. Van der Tuin, *Computer-aided Security Surveillance Design of the Quo Vadis Object Tracker*, Master's thesis, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, 2007.
11. A.Yilmaz, O. Javed and M. Shah, "Object tracking: a survey," *ACM Computing Surveys*, vol. 38, no. 4, article 13, Dec. 2006.

PETS 2006 dataset website:
<http://www.pets2006.net/>.