Usage of Threshold Absolute Difference Algorithm and People Counting in a Crowded Environment

Minu.S , V.Cyril Raj(H.O.D CSE)

CSE Department, Dr.M.G.R.University
Chennai, India

ABSTRACT

People counting is a usual problem in visual surveillance. An accurate and real-time estimation of people in a crowded place can provide valuable information. Here video is inputted and gives the average number of people as output. The video input is separated to number of frames and some processing steps are performed on background subtraction results to estimate the number of people in a complicated scene, which includes people who are moving only slightly. A Threshold absolute difference algorithm is used here for background subtraction method. The extracted foreground image's pixels count is calculated and gives as input to the neural network. In learning phase, the people count is calculated by manually for test dataset. It is tested with remaining test cases by adjusting weight parameters to obtain relative to the target result.

Keywords

background subtraction, neural network, people counting.

1. INTRODUCTION

Automatic monitoring of the number of people in public areas is also important for safety control. This paper aims to develop an effective method for estimating the number of people in a complicated outdoor scene. In this scenario, occlusions exist everywhere with people walking, sitting, and standing. The video was taken by a static camera overlooking a large area. This scene is very common, but few results have ever been demonstrated for such events. The input video is segmented to a number of frames based on a frame rate and these frames are processed by foreground subtraction method. An absolute difference method is used here for background subtraction .The resulted foreground is binarized and found out the foreground pixel count. This foreground pixel count will be the input to the neural network and find out the average people count based on the trained data. A video of 400 frames are processed in the evaluation time and 50 frames used as the trained data which is manually counted to train network. Rest 350 frames are used as a test data which will count the average people based on the neural network training..

2. RELATED WORK

The work done by Ya-Li Hou, Student Member, IEEE, and Grantham K. H. Pang, Senior Member,IEEE was a method based on the neural network to estimate the number of people. They used a robust adaptive background estimation method based on the Gaussian Mixture Model(GMM) for extracting background mean. This foreground image binarized based on a threshold and produce foreground pixels. This foreground pixel count used as input to neural network for their people estimation. sudden increase or decrease of people resulted large variations in the people count.

The work done by Lijing Zhang and Yingli Liang was "Motion human detection based on back ground subtraction". They established a reliable background updating model based on statistical and use a dynamic optimization threshold method to obtain a more complete moving object. And then morphological filtering is introduced to eliminate the noise and solve the background disturbance problem. They focused only on the human location detection based on background subtraction.

The work done by T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments" was a model-based approach to interpret the image observations by multiple partially occluded human hypotheses in a Bayesian framework. They defined a joint image likelihood for multiple humans based on the appearance of the humans, the visibility of the body obtained by occlusion reasoning, and foreground/background separation. The optimal solution is obtained by using an efficient sampling method, data-driven Markov chain Monte Carlo (DDMCMC), which uses image observations for proposal probabilities. Knowledge of various aspects, including human shape, camera model, and image cues, are integrated in one theoretically sound framework.

The work done by V. Rabaud and S. Belongie, "Counting crowded moving objects," in Proc.IEEE Conf. Comput. Vis. Pattern Recog., 2006, pp. 705–711 was based on a highly parallelized version of the KLT tracker in order to process the video into a set of feature trajectories. While such a set of trajectories provides a substrate for motion analysis, their unequal lengths and fragmented nature present difficulties for subsequent processing. To address this, they proposed a simple means of spatially and temporally conditioning the trajectories. With this representation, they integrated it with a learned object descriptor to achieve a segmentation of the constituent motions. they presented experimental results for the problem of estimating the number of moving objects in a dense crowd as a function of time.

The work done by S.-Y. Cho, T. W. S. Chow, and C.-T. Leung, " A neural-based crowd estimation by hybrid global learning algorithm," was a neural-based crowd estimation system for surveillance in complex scenes at underground station platform is presented. Estimation is carried out by extracting a set of significant features from sequences of images. Those feature indexes are modeled by a neural network to estimate the crowd density. The learning phase is based on our proposed hybrid of the least-squares and global search algorithms which are capable of providing the global search characteristic and fast convergence speed. Promising experimental results are obtained in terms of accuracy and real-time response capability to alert operators automatically.

3. METHODOLOGIES

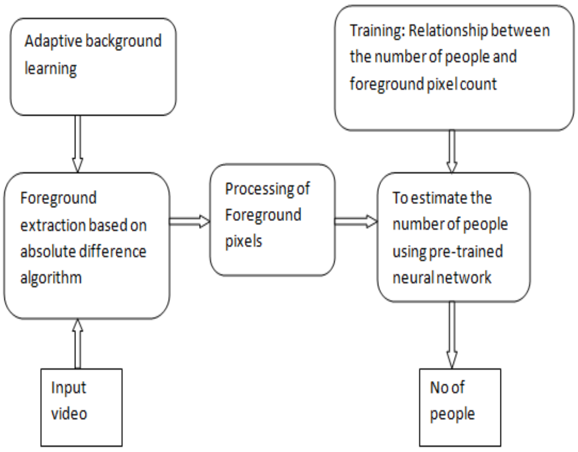

Fig1:System architecture

People always exhibit some movement whether they are standing or sitting. Motivated by this observation, it's possible to estimate the number of people by finding a relationship with the foreground pixels. A foreground image is obtained by subtracting the current frame(image) from the background frame (image). The background subtraction can be performed in different ways . In Ya-Li-Hou's paper they used GMM method(Gaussian Mixture Model) which produced a Gaussian distribution model of background image in HSV color space and generated the background mean. This background mean is subtracted with the next frame's mean which is resulted from its Gaussian distribution model. But In this paper an Absolute difference approach is implemented to perform background subtraction. The foreground image obtained is binarized based on a threshold to return the foreground pixels. Compute the all Foreground pixels from the foreground image and give that count as input to the neural network. It will find out the average number of people based on the pre-trained data set.

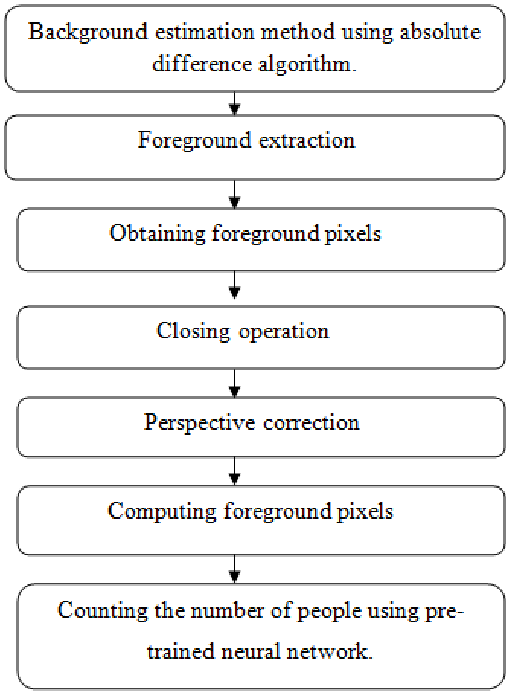

Fig2 :flow chart which shows the main procedures for counting.

A. Background Estimation Method

A robust background estimation method based on the Absolute difference algorithm is used here to find out the background.
The Algorithm involves :

- Finding out the starting frame and ending frame.
- Read first frame and convert it to gray image format. ('rgb2gray' command is used)
- Assign variable A= first frame's gray image
- Set a threshold value .(here 11 is used)
- Open a loop from starting frame to ending frame
  - do rgb to gray conversion .
  - Perform Absolute subtraction method between current frame and A. (use abs( ) method)
  - If the subtracted value of each pixel>threshold
    - Make visible that pixel in the foreground image.
  - Else update the pixel value to zero in the foreground image.
  - Change Value of A= current frame
  - Current frame= next frame.
- Returns the Foreground image.

Fig3: a)Typical scene to be processed. (b) Binary foreground image after foreground extraction for (a).

B. *Obtaining Foreground Pixels*

The foreground image obtained after extraction is Binarizing based on a threshold. A Foreground pixel is a pixel whose intensity difference between current image & background image greater than the threshold.

```
for j = 1:width
     for k = 1:height
        if (fr_diff(k,j)>thresh)
            fg(k,j) = A(k,j);
        else
            fg(k,j) = 0;
        end
     end
end
```

The above code shows the foreground pixel extraction based on a threshold.

Where `thresh` =11
`fr_diff` = foreground image pixels array after background subtraction.
`A` = Current gray image pixels array
`Fg` = foreground pixel array

C. *Perspective Correction*

The Size of an object varies linearly as a function of the y-coordinate of the image.The objects at different locations are brought to the same scale in this method.

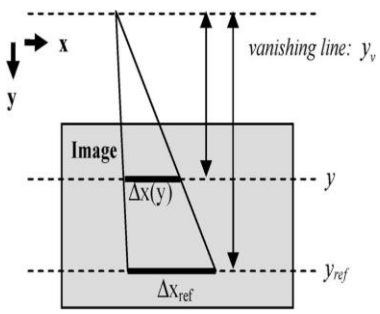

Fig 4: perspective correction

Equation (1) shows how to convert a scale at y to its scale at the reference location, $yref$. Fig. 4 is a simple illustration for (1). $\Delta x(y)$is the horizontal (vertical) scale of an object at y, and $\Delta xref$ is its horizontal (vertical) reference scale. $q(y)$ is the ratio for different locations

$$\Delta xref = \Delta x(y) * q(y) \text{ and}$$

$$q(y) = (yref - yv)/(y - yv) \text{ . (1)}$$

The extension of parallel lines intersects at a vanishing point,which lies on yv in the image. yv can be easily estimated using the same object at two different coordinates in (1).

D. *Computing Foreground pixels.*

Perspectively corrected foreground pixels are computed together using the following equation.

$$Npixel = \sum_{y=1:imgY} N(y) * q(y)$$

imgY = height of processing image
Npixel= total no of foreground pixels
N(y) = no of foreground pixels in yth row
q(y)= ratio for different locations.

E. *Closing operation.*

From the extracted foreground, will get solid blobs and some scattered pixels. Those solid blobs are from the moving people and scattered pixels are from the stationary crowd.
To reduce the difference between moving people and stationary people, a closing operation is employed.
Closing operation means , most areas occupied by people are covered with white pixels, while the other parts with black.Perspective effects also need to be considered during the closing operation.

F. *Counting the number of people using pre-trained neural network.*

Some manually annotated training images from a similar scene are needed to find the relationship between foreground pixel count & no of people. 3 methods are discussed below.

Method 1)-based on foreground pixels

$$M = f1(X)$$

Where M is the no of people and X is the no of foreground pixels after perspective correction and f1 is the manually annotated training set is used to ascertain the relationship f1.

Method 2) Based on Closed Foreground Pixels:

Let C be the number of foreground pixels after the closing operation and M be the number of people. The relationship between C and M will be found and used for estimation.

$$M = f2(C)$$

Method 3) Based on Both Foreground Pixels and Closed Foreground Pixels:

To keep more information about the original image, both foreground pixels and closed foreground pixels will be injected into the neural network. The relationship between the number of people and these two inputs is denoted as $f3$.

$$M = f3(X,C).$$

4. RESULTS AND ANALYSIS

Here Test data used was a video of 400 frames extracted from this. Training set consists of 50 images of that video ,which is counted manually. The test set is composed of the remaining 350 images. Image resolution taken was 768 * 576 . To increase the speed of people counting, all the images were resized to 256 $*$ 256 pixels.

Fig:5 : training of neural network

The above figure shows the training of neural network which produces a training set of 50 image frames.

Fig6: Testing video frames and shows the people count.

Fig6 shows the people counting of some random image frames and the average count of people based on three methods.

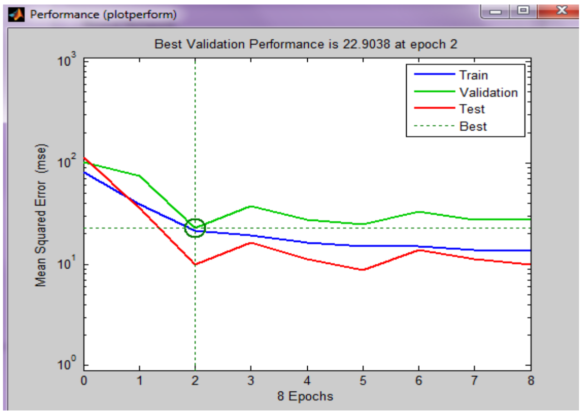

Fig 7: Performance graph while training, validation and testing.

The above graph shows the performance graph while training, validation and testing. The dotted lines show the best performance and it coincides at one point which shows the best validation performance. x axis represents the number of epochs and y axis represents the mean squared error(mse) , ie, the mean difference between the target output and network output.

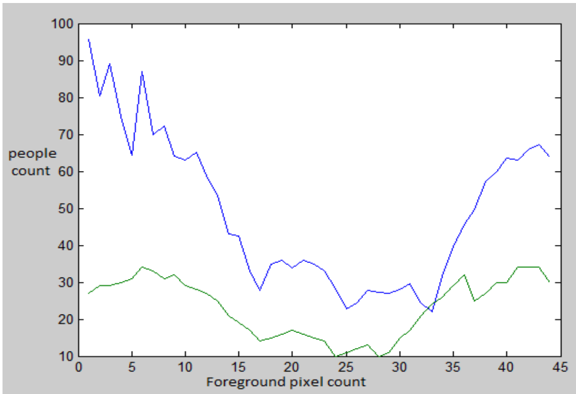

Fig 8: multiple curve Graph shows the relationship between foreground pixel count and the people count.

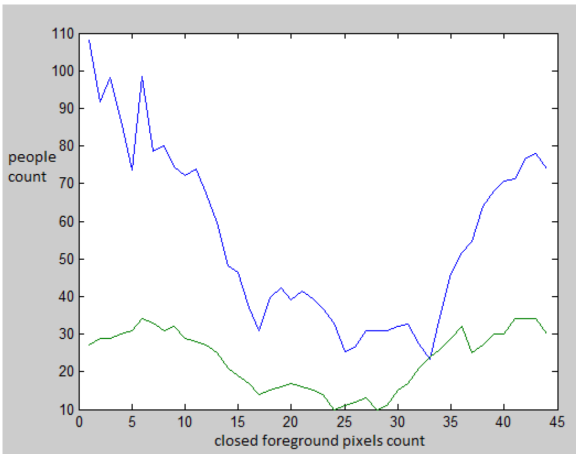

Fig 8: multiple curve Graph shows the relationship between closed foreground pixels count and the people count.

5. CONCLUSIONS

In this paper, foreground pixel estimation is carried out after background extraction of frames using Absolute difference method. Closing operation over foreground pixels used for considering stationary people. The best estimation results, with a 14% average error, were achieved when foreground pixels and closed foreground pixels ,are learned in a neural network.

6. REFERENCES

[1] Ya-Li Hou, *Student Member, IEEE*, and Grantham K. H. Pang, *Senior Member, IEEE*"People Counting and Human Detectionin a Challenging Situation"

[2] Lijing Zhang and Yingli Liang "Motion human detection based on back ground subtraction".

[3] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments"

[4] V. Rabaud and S. Belongie, "Counting crowded moving objects," in *Proc.IEEE Conf. Comput. Vis. Pattern Recog.*, 2006, pp. 705–711

[5] S.-Y.Cho, T. W. S. Chow, and C.-T. Leung," A neural-based crowd estimation by hybrid global learning algorithm,"

[6] R. Ma, L. Li,W. Huang, and Q. Tian, "On pixel count based crowd density estimation for visual surveillance," in *Proc. IEEE Conf. Cybern. Intell.Syst.*, 2004, pp. 170–173.

[7] H. Celik, A. Hanjalic, and E. A. Hendriks, "Towards a robust solutionto people counting," in *Proc. IEEE Int. Conf. Image Process.*, 2006,pp. 2401–2404.

[8] P. Kilambi, O. Masoud, and N. Papanikolopoulos, "Crowd analysis at mass transit site," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2006, pp. 753–758.

[9] A. N. Marana, S. A. Velastin, L. F. Costa, and R. A. Lotufo, "Estimation of crowd density using image processing," in *Proc. IEE Colloq. Image Process. Security Appl.*, 1997, pp. 11/1–11/8.

[10] A. N. Marana, L. Da Fontoura Costa, R. A. Lotufo, and S. A. Velastin, "Estimating crowd density with Minkowski fractal dimension," in *Proc.Int. Conf. Acoust., Speech, Signal Process.*, 1999, pp. 3521–3524

[11] H. Rahmalan,M. S. Nixon, and J. N. Carter, "On crowd density estimation for surveillance," in *Proc. Inst. Eng. Technol. Conf. Crime Security*, 2006,pp. 540–545.

[12] X. Li, L. Shen, and H. Li, "Estimation of crowd density based on wavelet and support vector machine," *Trans. Inst. Meas. Control*, vol. 28, no. 3, pp. 299–308, Aug. 2006.

[13] D. Kong, D. Gray, and T. Hai, "A viewpoint invariant approach for crowd counting," in *Proc. Int. Conf. Pattern Recog.*, 2006, pp. 1187–1190.

[14] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Proc.IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–7.