A NOVEL FRAMEWORK FOR AUTOMATIC PASSENGER COUNTING

1Satarupa Mukherjee, 2Baidya Nath Saha, 3Iqbal Jamal and 4Richard Leclerc, 5Nilanjan Ray
1,2,5Department of Computing Sciences, University of Alberta, Edmonton, Canada
4Transportation Department, City of Edmonton, Edmonton, Canada
{1satarupa,2baidya,3ray1}@ualberta.ca, 3iqbaljamal@aqllmc.com, 4Richard.Leclerc@edmonton.ca

ABSTRACT

We propose a novel framework for counting passengers in a railway station. The framework has three components: people detection, tracking and validation. We detect every person using Hough circle when he or she enters the field of view. The person is then tracked using optical flow until (s)he leaves the field of view. Finally, the tracker generated trajectory is validated through a spatio-temporal background subtraction technique. The number of valid trajectories provides passenger count. Each of the three components of the proposed framework has been compared with competitive methods on three datasets of varying crowd densities. Extensive experiments have been conducted on the datasets having top views of the passengers. Experimental results demonstrate that the proposed algorithm performs well both on dense and sparse persons with different hair colors, hoodies, caps, long winter jackets, bags and so on. The proposed algorithm shows promising results also for people moving in different directions. The proposed framework can detect up to 30% more accurately and 20% more precisely than other competitive methods.

1. INTRODUCTION

Accurate passenger count is crucial for traffic management and decision making. In the past, passenger count was mostly done manually and was both labor and cost intensive. So, it is important to develop an automatic method for counting passengers. In the last few years, the effective development of automatic people-counting systems based on digital image processing techniques aroused considerable research interest [2], [3], [7], [8].

Two noteworthy attempts to develop automatic people counting system are as follows. Chan et al. [2] proposed a fast people counting method, where they first implemented approximate video frame segmentation using mixture of dynamic textures and then the correspondence between the features and the number of people per segment is learned through Gaussian process regression. Instead of detecting individual object instances, Lempitsky et al. [8] introduced a supervised learning framework for counting objects in an image posing the counting problem as estimating a continuous density function integral for which over any image region provides object count within the region. They start with dot-annotated training images and compute local appearances (SIFT-based visual dictionaries or a set of randomized trees). Then they learn a linear mapping that transforms the feature vector at each pixel to a density value.

These methods count total number of people frame-by-frame in a video independently taking into account average velocity of people. Consequently, these method are imprecise. In this paper, a novel framework is proposed to count the number of passengers in a Light Railway Transit (LRT) by counting each person individually. Local transportation agency at Edmonton urges automatic software for counting number of passengers travelled by LRT that can serve their future planning and decision making system. For achieving this goal, video cameras were mounted on the ceilings of the entrance and exit hallways of different LRT stations across the city during different hours of a day. In this article, we propose a Hough circle [5] based detection algorithm to detect people, then optical flow [6] based tracking algorithm is evoked for each person detected in a frame, and finally all trajectories are sieved through a spatio-temporal validation algorithm to verify whether the tracker follows a person correctly. The number of valid trajectories gives the number of people.

A noteworthy novelty in this framework is the introduction of self-validation technique after completing the detection and tracking algorithm. Most of the techniques available in the literature skip the validation step. Here, we emphasize that validation technique cannot be ignored since all the automatic object detection techniques developed till date produce a significant number of false alarms. Performance of an automatic object detection technique also becomes worse in a real passenger counting system like ours, where the top view of a scene is captured and very few important features are available for recognizing human body. On the contrary, recognition of the silhouette of a human body available at the front view is easier to detect by various shape based detection techniques like Histogram of Oriented Gradients (HOG) [3].

Extensive experiments have been carried out on datasets of varying crowd densities and people moving in different
directions. Results demonstrate that the performance of the proposed framework is acceptable for both sparse and dense crowds. Results of proposed techniques on the dataset with people moving in different directions also seem to be very promising. The entire method is described in section 2.

2. PROPOSED METHOD

The dataset, on which the proposed framework is developed, consists of top views of passengers. Most of the previous efforts on automatic people counting system made attempt to detect human body by recognizing silhouette of a body from the front view of a person [3]. The advantage of working with top views is that, it does not have much occlusion and helps to track any person by a tracking algorithm fail-safe. However, the bottleneck of the top views is that it provides only a handful of features to work with and makes the task of automatic detection process cumbersome. Pose and gestures cannot be taken into account too as the front view of the person is not available. The illumination condition also plays an important role in the detection process. The input of the proposed algorithm is a time sequence of image frames of a top view of passengers and the output is the number of valid trajectories generated by the trackers that leads to the total number of the passengers. The entire framework consists of three sequential stages:

1) The object detection stage initiates the trajectories.
2) The object tracking step generates the trajectories.
3) The validation step checks whether a trajectory belongs to a person or not and counts the total number of valid trajectories.

2.1 Object Detection

Taking into account that there is circularity in the top view of a human body, the procedure has tried to capture the whole body of the person as a circular object. Initially, approximate median (AM) based background subtraction [10] was performed to remove the frames that did not have any people to speed up the process. Edge was detected on the entire image having motion with Canny’s edge detector [5]. On the edge image, circles having radii within 60 to 80 were detected using Hough circle method [5].

A square template was built around the center of the detected circle to represent the object and track it in the consecutive frames. Detection and tracking are performed at each frame. We used color information for distinguishing newly arrived persons and the persons already detected in the previous frame. For each frame, if the value of the Bhattacharyya coefficient between the color distribution of a newly detected and previously detected person becomes very high, then the newly detected person is evaded as he or she is considered already detected in the previous frame.

Apart from Hough Circle method, Histogram of Oriented Gradients (HOG) [3] method was also used for the detection process for a comparison. In our experiments, the performance of Hough circle method was better than HOG method as the top views of passengers had a parametric circular shape. Besides, passengers having different hair colors, wearing different types of hoodies, caps, long winter jackets, carrying bags etc. make the outer body an irregular shape that was very difficult to learn with supervised shape based object detection technique like HOG.

2.2 Object Tracking

After an object has been detected in a frame, the center of the object along with its template, are tracked in the consecutive frames using optical flow method [6]. In the optical flow based tracking method, the center position of a template at current frame is obtained by adding the average velocity of the pixels within each template with the center position of the template at previous frame. Tracking starts as a person is detected in a frame and continues as the person moves through the frame. Tracking ends when the person leaves the field of view. Apart from Horn-Schunck method, Lucas-Kanade method [9] and Brox et al. method [1] were also used for optical flow-based tracking. All these three methods perform well on the top views of the passengers. However, Horn-Schunck technique is the fastest and we use it in the proposed framework.

2.3 Object Validation

Hough or HOG based object detection technique generates two types of false alarms– (a) clutter detected as people; (b) duplicates: detecting different body parts of the same person. To eradicate these two issues we propose an approximate median (AM) based background subtraction method and measure the ratio of overlap of two trackers in the spatio-temporal domain for rejecting clutter and duplicate trackers respectively and we name the proposed validation framework as spatio-temporal validation (STV).

Let $T = \{T_1, T_2, ..., T_n, ..., T_p\}$ be a trajectory generated by a tracking algorithm on $p$ consecutive frames, where $T_i$ is the set of pixels contained by the tracker on $i$-th frame. Tracker $T_i$ trails a person correctly if $\forall i \in \{1,2,3,...,p\}$, $T_i$ is the output of AM for frame $I_i$. The value of each pixel $(x, y)$ in $T_i$ is either 1 or 0 if it belongs to foreground or background respectively and $K_i$ is the number of pixels contained by the tracker $T_i$ on $i$-th frame. For duplicate removal, tracker $A$ is considered as a duplicate of tracker $B$ if $(m) \geq \frac{m}{O}$, where, $m$ is the number of common frames for both tracker $A$ and $B$. The values of $L$ and $O$ are determined empirically that we demonstrate in the next section.

3. RESULTS AND DISCUSSIONS

The framework was tested on three different videos representing different crowd densities: very highly crowded,
moderately crowded and sparsely crowded videos. The total number of tested frames on these three videos was 2000. Additionally, the framework was tested on a video of 5000 frames where people are moving in different directions.

<table>
<thead>
<tr>
<th>HOG</th>
<th>Hough Circle</th>
</tr>
</thead>
</table>

**Low**

**Medium**

**High**

![Fig. 1: Results of HOG and Hough circle based detection.](image)

Visual results of HOG and Hough based detection algorithms on low, medium and dense crowds are demonstrated in Fig. 1 and quantitative comparison in terms of Accuracy, Recall, Precision and F-measure between

<table>
<thead>
<tr>
<th>Optical flow based tracking algorithms</th>
<th>Horn-Schunck</th>
<th>Lucas-Kanade</th>
<th>Broxet al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s) taken between two consecutive frames</td>
<td>1.32</td>
<td>9.67</td>
<td>15.81</td>
</tr>
</tbody>
</table>

**Table 1: Running time of three tracking algorithms.**

We chose Horn-Schunck [6] method for tracking in the proposed framework as it seems to be faster than Lucas-Kanade [9] and Brox et al. [1] method. Average time taken to track a person between two consecutive frames of resolution 480x640 by these three methods implemented in Matlab on a desktop (Intel duo 2 core processor, 2 GHz and 4 GB RAM) is illustrated in Table 1.

<table>
<thead>
<tr>
<th>Value of L</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>83</td>
<td>88</td>
<td>96</td>
<td>83</td>
<td>71</td>
</tr>
<tr>
<td>F-measure</td>
<td>90</td>
<td>93</td>
<td>97</td>
<td>97</td>
<td>79</td>
</tr>
</tbody>
</table>

**Table 2: Accuracy and F-measure for different values of threshold parameter, L**

Proposed validation step described in section 2.3 enhanced the performance of proposed method. Before validation, the recall and precision of the system was 100% and 70% respectively. After passing the entire trajectories through validation process, the recall and precision of the system was 96% and 90% respectively. Table 2 shows accuracy and F-measure values for different values of threshold parameter, L regarding outlier rejection of proposed STV on 100 randomly selected images. We choose the value of L as 0.6 since both accuracy and F-measure value is the highest among all other values of L as shown in Table 2. Similarly we have chosen the value of O, for duplicate tracker removal as 0.1.

We have compared proposed STV with two other spatio-temporal features, motion histogram (MH) [4] and spatio-
temporal gradient (STG) [7] and their comparisons are illustrated in Fig. 3. Fig. 3 shows that proposed STV outperforms over MH and STG. Both MH and STG are supervised validation algorithm and first 50% of the total frames were used as training and the remaining 50% were used as testing. Here, it is noted that Hough was used as detection and Horn-Schunck was used as tracking while evaluating the performances of MH, STG and STV. People entering from any of the four borders of the frame and moving in any direction were detected and tracked successfully using proposed framework as shown in Fig. 4. Accuracy, Recall, Precision and F-measure of proposed framework on 5000 frames are demonstrated in Table 3.

<table>
<thead>
<tr>
<th>Recall</th>
<th>Precision</th>
<th>Accuracy</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>97.4</td>
<td>94.2</td>
<td>93.5</td>
<td>95.8</td>
</tr>
</tbody>
</table>

Table 3: Performance of proposed framework on 5000 frames with people moving in different directions.

4. CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework to count the total number of people in a railway station. The framework has three components: object detection, object tracking and object validation. The object detection step detects a person as he enters the image frame, the object tracking step tracks the person as he moves through the frame and as a result, a trajectory is generated. The trajectories are then analyzed in the validation step and the number of valid trajectories is counted. The number of valid trajectories gives the number of people. The algorithm performs well on dense as well as sparse crowds. The algorithm succeeded for various types of passengers: persons having different hair colors, wearing hoodies, caps, long winter jackets, carrying bags etc. The proposed framework is also smart enough for tackling people entering in the frame from any direction.

Our future work includes speeding up of the framework using a multi-resolution image pyramid and extending the framework to various camera views: gazing, frontal etc.

ACKNOWLEDGEMENTS

The work was supported by MITACS internship award, AQLMC Inc. Edmonton, City of Edmonton, NSERC Discovery Grant and University of Alberta, Canada.

5. REFERENCES