Crowd Management Framework for Departure Control in Bus Transport Service using Image Processing

Adithi S[1], Mahanth Sai M[2], DhrithirhuthRajanna[3], Dr Rekha.N[4], K Rishika Ravi[5]
[1,2,3,5] *Department of Electronics and Communication, K.S. Institute of Technology, Bangalore, India*
[4]*Assoc Prof., ECE department, K.S. Institute of Technology, Bangalore, India*

Abstract— **Crowd detection is an important aspect of video surveillance. Video surveillance systems are one of the most modern methods for estimating the density of people in a given area for providing facilities and obtaining human statistics. Factors such as severe occlusions, scene perspective distortions in real time application make this task a bit more challenging. Image recognition and classification using Convolution Neural Networks (CNN) are the two popular approaches used in object recognition systems. CNN models are built to evaluate its performance on image recognition and detection datasets. This paper develops a prototype of an intelligent public bus management system based on collecting data from surveillance cameras, processing image frames to estimate crowd density, and sending messages to bus depot as needed. Besides image processing algorithms, model consists of camera, software and WIFI for wireless data transmission at the Bus Depot. This system prevents the overcrowding of passengers, provide security, report passenger density data and thereby organize an effective bus management.**

Index Terms: **Crowd Estimation, Video Surveillance, Convolution Neural Network (CNN), Occlusions, Image Processing, Image Segmentation.**

INTRODUCTION

In recent years, with the rapid development of technologies such as sensing, communication and management, improving the efficiency of traditional transportation systems through advanced technological applications is becoming more feasible. Therefore, intelligent transportation systems have gradually become a focus of transportation development around the world. Currently, many related applications exist in the field of bus information services for example, bus depots can utilize a dynamic information web page or a mobile app to inquire about the numbers of people waiting in the bus stops and send busses accordingly. If more comprehensive information is provided on existing bus information platforms, the quality of public transport services will be significantly improved. Thus, the number of passengers willing to use public transport will increase. Through intelligent traffic monitoring, bus depot managers can preview the allotment of buses to a particular bus stop in real time and then make a decision based on the additional information and evaluate the waiting time. Furthermore, the bus depots can manage vehicle scheduling based on this information thus, operational costs depending on whether service quality is degraded or not are reduced effectively.

LITERATURE SURVEY

People count system using multi-sensing application [1]: In this paper people count system has been developed using multi-element infrared sensor which is constructed using PbTi03 ceramics to set up a non-blocking and non-contact automatic real-time system which gives high accuracy. A pyroelectric array detector using human sensor information is employed to set out people count system for detection of passers and direction in which the people move in a 200cm wide door. A highly sensitive infrared array detector was fabricated using bulk ceramics. This system gives an accuracy of 95%.

Advantages: This system makes use of the infrared motion sensor which can detect motion both in day time and night time reliably. The sensors are smaller in size and affordable. This method is easy to install and maintain and offers good accuracy.

Disadvantages: The proposed system supports shorter range and hence its performance degrades with longer distance.

Device-free human detection using WiFisignals[2]:Wireless sensing represented by WiFi channel state information (CSI) enables various fields of applications such as person identification, human activity recognition, occupancy detection, localization and crowd estimation. The proposed system consists of three steps pre-processing, feature extraction and decision making. First, the CSI data is pre-processed using band pass filter and the band is preserved to human breathing frequency. Second, prediction is done using two features that is average of subcarriers and missing numbers over subcarriers. Lastly, a decision tree is built to find human presence and absence. In this method an AP with two antennas is used as transmitters and a laptop is used as a receiver. This system gives an accuracy of 64.3%.

Advantages: This system makes use of wifi signals which can travel through the wall, so it is not necessary for the person to be in line of sight. The cost of this method is less due to less hardware being used.

Disadvantages: This method doesn't enable counting in case of an obstruction or barrier. It possess high false alarm rate of about 46%.

Real-time crowd density estimation using images[3]: The crowd density in this system is measured in real time which is done based on textures from crowd image. The pictures at the input are segmented and sections are made based on crowd density. The sections are further processed with the help of a low pass filter depending on the images obtained previously from the incoming image sequences. The following are the steps involved in this algorithm first, the master processer divides the input image into n fragments (where n refers to the count of slave nodes in the bundle). Second, the slave processor is then given each of the fragmented image fragments. Third, the slave processer's job is to classify the texture of the object. In addition, the slave processor delivers the assigned parts to the master processor. Finally, the master assembles each fragment into a final texture.

Advantages: This method facilitates estimation of human density in real time.

Disadvantages: Certain misclassifications are unavoidable in this system.

Pixel count based crowd density estimation[4]: Pixel counting systems uses geometric corrections that are usually overlooked by other methods. It shows correlation between ground plane and geometric correction. The effectiveness of this method depends on the result of the Background Subtraction. An additional imperative parameter is the perspective distortion which can be rectified using geometric correction. Initially it calculates scale of lower level plane and explores appraisal application. After use of robust segmentation algorithm mask determination is executed as the segmentation effects in false foreground regions. The Region of Interest (ROI) is carefully accounted and a mask is created. It may also be detected by accumulating foreground pixels, which has a 90% accuracy rate.

Advantages: The geometric adjustment of the ground plane is deduced from the expression.

Disadvantages: There is no automatic calibration.

Granular computing-based crowd segmentation[5]:It proposes a granular computing-based framework for crowd segmentation to validate the issue of crowd segmentation to be evaluated at various levels of granularity and to map issues to small problems. It demonstrates that by dissolving the correlation in the pixel granules, it is possible to convert the identical pixels into appropriate atomic structure granules. Each phase of GrC-based crowd segmentation (GrCS), which is based on granules, follows a non-identical degree of granularity. This is done to improve people's capacity to detect at various granular levels in order to map concerns to smaller, more manageable ones. To deal with diversity in gyration of collected microstructures, an expanded version of the Local Binary Patterns (LBP) operator named uniform patterns is utilised. Granularity skeletal structures are logically atomic portions in the frame that provide naturally separated regions between various human structures and the backgrounds. The fundamental goal of the atomic regions is to have a pixel total procedure that can adapt to different crowd scenarios. As a result, for robust crowd segmentation, this will be the ideal group varied structure in the scenario. This approach primarily analyses huge crowd images by extending GrC concepts to crowd issue segmentation at various granular levels.

Advantages: To deal with perspective distortion, fluctuating crowds, and dispersed backgrounds, this strategy is helpful in grouping blocks of the same pixels into a batch.

Disadvantages: When crowd situations are below par, granulated sight of various levels of granular is limited.

Depth driven people counting using deep region proposal network[6]:An overhead vertical Kinect sensor collects frames and performs head identification using depth pictures. It considers the impact of the amount and quality of positive anchors in the Region Proposal Network (RPN) on the performance of faster R-CNN and recommends a solution. The depth map of a scene is evaluated in hardware using Kinect sensors. To recognise heads on depth pictures, they used Faster R-CNN models and an RPN-alone model. A class-agnostic detector is utilised in the region proposal network. As a network input, a picture is fed, and as outputs, a set of region recommendations is generated. To create ideas, a small network glides over the convolutional feature map. For dealing with numerous scales, a variety of methods are described. The first method involves creating features maps from picture blocks and running the classifier at various sizes. The second method is to pass the feature map through several filter scales. To tackle the multi-scale problem, the RPN employs pyramids of several reference boxes, which differ from earlier techniques. The VGG-16 model is used as the backbone network in the original Faster R-CNN, which is regarded the baseline. In Faster R-CNN, the default RPN employs a foreground ratio of 0.5. If there are slightly more positive anchors in an image after computing the classification loss of a batch, they are filled with negative anchors. The default foreground ratio option causes the positive and negative anchors in the batch to be imbalanced since the number of positive anchors in a batch is generally smaller than the number of negative anchors. Negative samples are preferred by the network since they are more prevalent and would degrade performance.

Advantages: Runs at a frame rate of roughly 110 frames per second in real time.

Disadvantages: It cannot deal with situations where some items are closer to the sensor than the head.

Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [7]:Each phase in GrC-based crowd segmentation (GrCS), which is based on granules, follows a non-identical level of granularity. This is done to improve people's ability to detect at various granular levels in order to map concerns to manageable little problems. To deal with diversity in gyration of collected microstructures, an expanded version of the Local Binary Patterns (LBP) operator dubbed uniform patterns is utilized.

Advantages: To deal with perspective distortion, fluctuating crowds, and scattered backgrounds, this strategy is helpful in grouping blocks of the same pixels into a batch.

Disadvantages: When crowd scenes aren't up to par, granulated sight of various levels of granular is limited.

Crowd motion estimation and motionless detection in subway corridors by image processing[8]:This method is appropriate for spatial time related surfaces with a constant phase to derive the two motion components. A set of Gabor filters is convolved with the image, and the displacement vectors are obtained using this method. The downside is that it takes a long time. This method is based on the idea that the brightness of a light source is proportional to its size. Two filters, spatial filtering and contextual filtering, were used to process the segmentation in a good context. The usage of temporal filtering is employed.

Advantages: Motion-based and motionless detection are both carried out.

Disadvantages: The working pace is inefficient.

Histograms of oriented gradients for human detection[9]:While the security motive for crowd description algorithms is undeniable, the research field currently faces a number of significant challenges: To begin with, despite the fact that high definition cameras are becoming increasingly affordable, many existing CCTV cameras still record at low resolutions like as 352240 or 640x480. Because of the requirement to operate in a continuous operation 24 hours a day, 7 days a week, it's difficult to provide acceptable image quality (contrast, brightness), and because there are more cameras to monitor at once, the overall processing time should be maintained low (i.e. ideally real-time capabilities). Furthermore, when crowd density increases, the number of pixels describing human decreases, making typical person detectors such as histograms of individuals hard to use.

Advantages: The method described here allows for the detection of busy areas and their segmentation, as well as the counting of people in certain areas.

Disadvantages: In circumstances where the camera shakes, background subtraction-based procedures can cause issues (e.g. pole-mounted outdoor camera under windy weather conditions).

Statistical and Structural Approaches to Texture[10]:Textural descriptors were offered as a method for estimating crowd density. The technique's first step is to classify each pixel of the input image into one of the texture classes that have been previously identified. A self-organizing map (SOM) neural network performs the classification utilizing feature vectors built of texture descriptors derived from co-occurrence matrices and generated using a wxw window centred in the pixel to be classified.

Advantages: The variance of crowd density calculations for each class was extremely tiny, and their means were the expected values, therefore these results are quite good.

Disadvantages: It takes a long time to classify all of the pixels in an image.

Estimation of crowd density in surveillance scenes based on deep convolutional neural network[11]:In large-scale image identification, object recognition, and segmentation, deep convolutional neural networks (ConvNet) are commonly utilized. Deep ConvNets estimates population densities by extracting picture characteristics directly and mapping them to crowd density on three levels: low, medium, and high. The overall number of people in each range, as well as the number of ranges, may vary depending on the application and the field's unique features. The neural network is trained using frame samples from the train subset, which are divided into five categories based on the amount of people in the image: Very-low, Low, Medium, High, and Very-high. The output of the neural network is separated into 5 levels of crowd density by quantifying the estimated crowd density. The test subset's classification accuracy is used to assess performance. A new crowd dataset of subway scenes with over 160K photos is utilized to test the accuracy of the crowd density estimate approach in this deep ConvNets method. This method's experimental findings show that it has the best accuracy of 91.73 percent on average, and that it can perform better in real applications.

Advantages: The accuracy of cross-scenes is better evaluated using 160K density annotated images

Disadvantages: This method doesn’t employ rough people counting function.

Human count estimation in high density crowd images and videos[12]: This approach calculates the density and number of individuals in photos. The density of the crowd may fluctuate across the viewing area due to changes in the crowd pictures. To solve this problem, photos are divided into tiny patches with the same size, which are referred to as patches, and the number of patches is calculated. In terms of head count, confidences, and absolute errors from the specified patch, data is taken from several sources. The authors also used cascade training of head pictures, which included a set of bounding boxes that covered all of the positions and orientations of human heads, to increase count accuracy. Three systems are combined in this method: head detection, Fourier analysis, and feature extraction. The HOG-based feature descriptor is used to detect heads. Using edge detection and intensity gradients, this approach is utilized to distinguish the local item and outline it. After that, the picture is divided into small spatial sections known as patches, for which a 1-D histogram of gradient directions or pixel edge orientations is created. In high-density crowd photos, the human head appears as little dots. To solve this difficulty, the picture is subjected to Fourier analysis, which is highly accurate in recognizing human heads. To gather information about placements and big variations in intensity levels, the Fourier transform of all the patches is obtained. Fourier analysis is accurate for crowded patches, but not for uncrowned patches. To solve this problem, patch confidence is calculated and combined with the determined Fourier model results. Once the number of patches has been determined, the eSVR model is complete and used to train the patches against ground truth.

Advantages: Very low normalized mean and standard deviation values.

Disadvantages: There is no proper time constraint taken into consideration during the estimation.

Determining optical flow [13]-This approach is based on the idea that a moving point's brightness remains constant throughout time. The suggested system uses a modified Horn and Schunk technique, in which the velocity vectors vary only little between successive photos. These data are used to derive a global crowd motion direction. Two filters, spatial filtering and temporal filtering, are employed to process the segmentation in an appropriate context. In motionless detection, a module looks for all the regions where motion is present. Stationery individuals can be found by removing the relevant locations and filtering the results. Only when there is no motion the computation take place. As a result, with each occlusion, the stop duration is delayed. To avoid occlusion, the rate of occlusion is collected for a certain frequency, and the duration is adjusted using the data. This system's accuracy is estimated to be approximately 93 percent.

Advantages: It efficiently determines the motion of the objects on a constant basis.

Disadvantages: The power consumption is higher, and the reliability is deteriorating with time.

Image Crowd Counting Using Convolutional Neural Network and Markov Random Field[14]:This method uses the properties of image recognition, object detection, and segmentation of images borrowed from deep convolutional networks. The convolutional neural network Markov Random Field (CNNMRF) is used to calculate the number of people in a scene. Images are first decomposed into overlapping patches (overlapping), and then using Deep CNN, the features are taken from the overlapping and highly correlated patches. The number of people approaching the patch is the same and may vary significantly from place to place, depending on the vehicle and trees in the photo. Markov random fields (MRFs) are used in local patches to smooth the count results and make the count equal to ground truth. In particular, the CNNMRF-based method is used to get the number of people in an image from different locations.

Advantages: Due to the overlapped patches separated strategies, the nearing original counts are highly correlated.

Disadvantages: For a crowd exceedingly more than thousands this method shows high error rate.

Automated Solutions for Crowd Size Estimation[15]:This article outlines the main approaches to automatic crowd estimation to help you do this. Selection of a specific deployment scenario. Instead of reviewing a particular technology, the solution has three different theoretical understandings, as in existing research papers. The approach provided (computer vision, WSN, internet data / mobile phone based approach). In addition, a detailed description of the strengths and weaknesses of each approach is provided in terms of accuracy. Compatibility with various operating conditions. However, it is worth noting that there is no crowd. The estimation system itself depends on several aspects of technical and environmental factors. It needs to be taken into consideration. Automated systems reduce or eliminate the need for human interaction in the process. It accelerates Reduce process, labour costs, eliminate human error.

Automated systems are being developed for this purpose Complete difficult tasks for humans. Automated systems also have some limitations, including: High initial cost of the system, limited level of intelligence for specific scenarios and research, and Sometimes development costs.

Advantages: One of the most commonly used technique. Good accuracy depending on algorithm and environment. Accuracy will be high and it is able to distinguish human and other objects.

Disadvantages: Very expensive to achieve good coverage and high reliability. Cannot be used if an accurate counting system is needed.

PROPOSED METHOD & CONCLUSION

A high-resolution camera is used to capture the frames from a video. Since the surveillance cameras are already installed in the platforms the step becomes easier to implement in real time. Once the image is captured it is processed using an image processing algorithm. The platform used to process the image is OpenCV-python.

Once the processing is completed the count obtained is categorized into 3 classes of low, medium and high. Based on the division appropriate messages are sent to the controller using the Twilio as the medium of communication. The controller will take the necessary actions and inform the drivers of that particular bus. For crowd detection we are using Convolution Neural Network such as Tensorflow where crowd detection is achieved by mapping those visual features. Tensorflow’s Object Detection API provides pre-trained models for object detection, which can recognise roughly 90 different classes (objects), with people being one of them. To achieve real time application, we make use of raspberry pi along with integrated camera to capture the live data and to obtain the output via a digital screen this Model is trained with the help datasets.

The mentioned system was created to meet the need for human crowd estimation. The estimation and control of crowds in bus stops can lead to good bus management that provides passengers with a comfortable journey. According to the study, there was a need to design a new efficient system, and this project is an attempt to meet that requirement. The numerous estimating strategies available are time-consuming and produce unsatisfactory outcomes in practise. The proposed system is much simpler and easier to comprehend. The results were satisfactory, and they may be implemented on a limited scale in real-time scenarios.

The accuracy of the mentioned system ranges from 80 to 90%. The frequency of buses is controlled, it promotes travel convenience. The detecting method is complicated since the photos recorded must be pre-processed in order to produce correct results. The field of image processing for human detection is continually evolving, and no 100 percent results have yet been achieved. When the number of persons increases, however, the accuracy of the results declines. The accuracy ranges from 12 to 18 people, but as the number increases, the accuracy drops.

This could be due to the camera's resolution or the region covered by the camera. Furthermore, because the system is still in development, the chances of achieving 100 percent accuracy are slim.More field study is required to acquire complete results.The current system is confined to a single bus stop. This can be expanded to all of the bus stops along a specific route, and the data can be transmitted over the cloud. Bus frequency can be regulated by taking an average count from all bus stops.

REFERENCE

[1] Hashimoto, Kazuhiko, Katsuya Morinaka, Nobuyuki Yoshiike, Chihiro Kawaguchi, and Satoshi Matsueda. "People count system using multi-sensing application." In Solid State Sensors and Actuators, 1997. TRANSDUCERS'97 Chicago., 1997 International Conference on, vol. 2, pp. 1291-1294. IEEE, 1997.

[2] Li, Chu-Chen, and Shih-Hau Fang. "Device-free human detection using WiFi signals." In Consumer Electronics, 2016 IEEE 5th Global Conference on, pp. 1-3. IEEE, 2016.

[3] Marana, AparecidoNilceu, Marcos Antonio Cavenaghi, Roberta SpolonUlson, and F.L.Drumond."Real-time crowd density estimation using images." In International Symposium on Visual Computing, pp. 355-362. Springer, Berlin, Heidelberg, 2005.

[4] Ma, Ruihua, Liyuan Li, Weimin Huang, and Qi Tian. "On pixel count based crowd density estimation for visual surveillance." In Cybernetics and Intelligent Systems, 2004 IEEE Conference on, vol. 1, pp. 170-173. IEEE, 2004.

[5] Kok, Ven Jyn, and Chee Seng Chan., 2017 "Grcs: Granular computing-based crowd segmentation." IEEE transactions on cybernetics 47, no. 5: 1157-1168.

[6] Song, Diping, Yu Qiao, and Alessandro Corbetta., 2017 "Depth driven people counting using deep region proposal network." In Information and Automation (ICIA), 2017 IEEE International Conference on, pp. 416-421.

[7] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa., 2002 "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." IEEE Transactions on pattern analysis and machine intelligence 24, no. 7: 971-987.

[8] Bouchafa, Samia, Didier Aubert, and Salah Bouzar. "Crowd motion estimation and motionless detection in subway corridors by image processing." In Intelligent Transportation System, 1997. ITSC'97., IEEE Conference on, pp. 332-337. IEEE, 1997.

[9] Navneet Dalal and Bill Triggs, “Histograms of oriented gradients for human detection,” in International Conference on Computer Vision and Pattern Recognition (CVPR 2005), 2005, pp. 886–893.

[10] Haralick, R. M., "Statistical and Structural Approaches to Texture", Proceedings of the IEEE, vol. 67(5), pp. 786-804, 1979.

[11] Pu, Shiliang, Tao Song, Yuan Zhang, and Di Xie., 2017 "Estimation of crowd density in surveillance scenes based on deep convolutional neural network." Procedia Computer Science111: 154-159.

[12] Chauhan, Vandit, Santosh Kumar, and Sanjay Kumar Singh. "Human count estimation in high density crowd images and videos." In *Parallel, Distributed and Grid Computing (PDGC), 2016 Fourth International Conference on*, pp. 343-347. IEEE, 2016.

[13] Horn, Berthold KP, and Brian G. Schunck., 1981 "Determining optical flow." Artificial intelligence 17, no. 1-3: 185-203

[14] Han, Kang, Wanggen Wan, Haiyan Yao, and Li Hou., 2017 "Image Crowd Counting Using Convolutional Neural Network and Markov Random Field." arXiv preprint arXiv: 1706.03686.

[15] Muhammad Waqar Aziz1,2, Farhan Naeem3,Muhammad Hamad Alizai4, and Khan Bahadar Khan2 "Automated Solutions forCrowd Size Estimation" Reprints and permission: sagepub.com/journalsPermissions.nav DOI: 10.1177/0894439317726510