ICIC Express Letters
Part B: Applications
Volume **12**, Number **7**, July **2021**

ICIC International ©2021 ISSN 2185-2766
pp. **603–610**

AUTOMATIC COUNTING OF WAITING PERSON USING QUEUE AREA DETECTION OF PEOPLE BASED ON MACHINE LEARNING

Ningyuan Li*, Johei Matsuoka and Kazuya Tago

Graduate School of Bionics, Computer and Media Science
Tokyo University of Technology
1404-1 Katakuramachi, Hachioji City, Tokyo 192-0982, Japan
*Corresponding author: g211902137@edu.teu.ac.jp; { matsuokajh; ktago }@stf.teu.ac.jp

Received November 2020; accepted February 2021

Abstract. *An approach to measuring the number of people in the queue by detecting the person's queue area from information of the arbitrarily installed fixed camera using machine learning is proposed. There are existing studies that measure the number of people in the queue from the video, but these methods confuse passersby around the queue with people in the queue, so the number of people waiting is erroneously added. This is because the conventional methods count all people in the shooting area of the camera. In this paper, we propose a method which can reduce erroneous addition by detecting the queue area based on machine learning and excluding the passersby around the queue. The effectiveness of the proposed method has been confirmed by a comparative experiment with the conventional method using video from an actual fixed camera.*
Keywords: Machine learning, Queuing area detection, Queuing person counting, Image recognition, Object detection, Target tracking

1. **Introduction.** In recent years, as the population density of cities and the number of tourists are continuously increasing, the flow of people in public places is rapidly rising. For example, the reception areas of stores, stations, etc. are often crowded, and people will queue for a long time. Measuring the number of people in the queue is indispensable for improving the efficiency of area management and marketing in commercial facilities and transportation facilities. The area manager can optimize the space and staffing depending on the information about staying time and the number of people. Besides, it is also useful for facility users to rationally adjust the time according to the number of people in the queue and to make a better schedule. Therefore, it is highly desirable to realize a system that can easily measure the number of people in the queue.

Various methods have been proposed so far to measure the number of people in the queue. Using human labor to count is an example, but this method is expensive and time-consuming. In addition, the method of estimating the number of people waiting from the length of the queue is often used, but in the actual queue, the length of the queue is different from the number of people waiting because the spaces between persons are diverse [1]. As a method to automatically measure the number of people waiting, [2,3] measure the number of people waiting based on image processing. These methods use techniques such as background subtraction and optical flow, and they are difficult to detect the number accurately because of the difficulty in adapting to changes in angle and light. In addition, [4,5] use smartphones to estimate the number of people waiting based on applications and Wi-Fi/Bluetooth signals. These methods require the installation of an application and receiver equipment, and the measurement location is limited. Furthermore, the detection accuracy is also low because there are people who do not have any applications installed or who have multiple devices.

DOI: 10.24507/icicelb.12.07.603

There are two types of research on measuring the number of people in the queue based on machine learning from camera video, using person re-identification and person tracking technologies. In the method using person re-identification [6], two cameras are installed at the entrance and exit of the queue, and the number of people in the queue is measured through the identification of the same person. However, it is difficult to install two cameras when the measurement area is small. On the other hand, the person tracking method [7,8] measures the number of people waiting directly from a single camera. In contrast to person re-identification, the tracking method measures the number of people waiting by tracking a person in a single camera video.

By introducing machine learning techniques in these methods, the accuracy of the measurement is improved even when the person detection is difficult due to light changes or temporary shielding. In this paper as well, we adopt person tracking based on machine learning.

The person tracking method represented by the method of Wuthoo and Bedarkar [7] assumes a simple scene in which the shooting area of the camera shows only the queue. On the other hand, in the fixed cameras installed in transportation facilities and commercial facilities, which are the target of this research, many security cameras capture not only the queue area but also a wide range. Therefore, people and passersby who stop around the queue are counted as people in the queue, and thus the number of people waiting is incorrectly added.

The purpose of this paper is to propose a method for measuring the number of people in a queue, independent from the size of the shooting area, by using the video of fixed camera installation in much the same way as existing security cameras. In this paper, we propose a method which mainly focuses on detecting the specific area that people are queueing at and excluding nearby passersby to reduce the erroneous addition. Effectiveness of the proposed method has been confirmed by comparing the result by conventional method using the actual fixed camera video with the result by the proposed method. The comparison shows that the proposed method gives a closer number to the actual number of people waiting. As a result, the average accuracy of measurement is improved by 38.06%. We have achieved a highly accurate measurement in the number of people waiting in the queue, even in complex situations where there are many passersby around the queue.

Chapter 2 explains the system overview, related technology and system structure of the proposed method. Chapter 3 explains the experimental environment of the proposed method, displays the experimental results and analyzes them. Chapter 4 summarizes the analysis of the experimental results, and concludes the advantages and disadvantages of the proposed method. Also, the future prospects are described.

2. Methodology.

2.1. System overview. Figure 1 shows an overview of the proposed method. The white numbers in Figure 1 represent the tracking ID.

First, the target area is captured by a fixed camera from diagonally above, and the image is transmitted to the system. Next, the start of the queue is marked as the waiting point manually (the white dot in Figure 1). From this video, machine learning has been used by the system to recognize and track people. The system acquires everyone's movement, by locating and recording each person's position and movement speed. After that, density-based clustering is performed from the waiting point to detect the queue area. Passersby are deleted from the queue because of their high speed. Also, the distance between the person staying around the queue (ID-4 in Figure 1) and the approximate line of the queue is calculated. After that, the system deletes the person from the queue if the distance is large. Finally, the system displays the queue area as a line on the screen, measures the number of people in the queue area, and outputs the results.

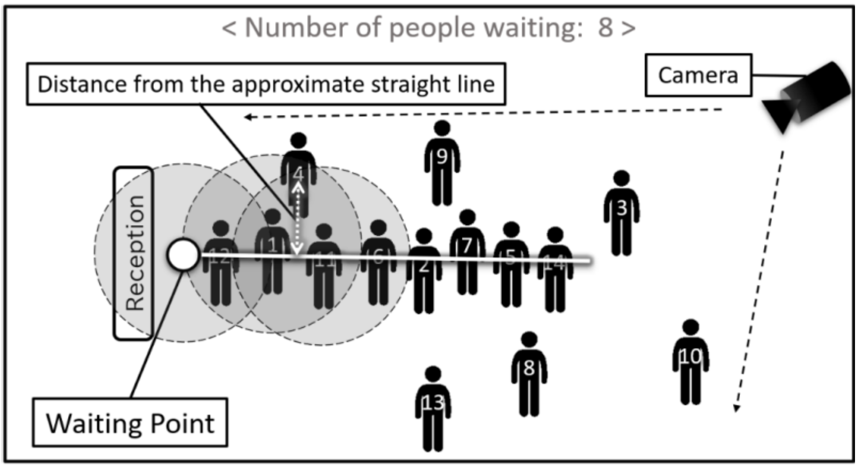

Figure 1. Overview of the proposed method

2.2. **Related technology.** Three technologies are required to implement the proposed method.

1) Object detection algorithm: Detect the position of people from the image.
2) Multi-target tracking algorithm: Obtain the movement paths of the passersby.
3) Clustering algorithm: Detect the person in the queue.

2.2.1. *Object detection algorithm.* Object detection is to detect the position and category of a defined object in the image. Currently, there are two types of object detection algorithms based on deep learning: 1-Stage, and 2-Stage. Overall, the 2-Stage method has high detection accuracy, but it requires a lot of calculation and processing time. The accuracy of the 1-Stage method is relatively low, but the amount of calculation is small and the speed is fast.

YOLOv4 is a 1-Stage object detection algorithm published in April 2020, which uses many of the latest detection technologies such as SPP and PANet, and has excellent detection accuracy and speed. Compared to other state-of-the-art object detection algorithms, YOLOv4 can process at twice the speed of EfficientDet with the same performance. Compared to YOLOv3, AP and FPS are improved by 10% and 12%, respectively [9]. Therefore, this study selected YOLOv4 as the object detection algorithm.

2.2.2. *Multi-target tracking algorithm.* Multi-target tracking is to identify the same target in different frames based on the detected position and image features of the target object, and to assign a unique tracking ID to the target object.

The predecessor of the Deep SORT [10] algorithm is SORT (Simple Online and Realtime Tracking [11]). The most important feature of SORT is that it is based on Faster R-CNN target detection method and uses the Kalman filter and the Hungarian algorithm to greatly improve the speed of multi-target tracking. However, SORT has the problem that when a person hides in the shadow of an object and reappears, a different ID is assigned. Deep SORT solves this problem by adding cascade matching and the new tracking path confirmation process, and using an AI model that compares the appearance similarity of targets to reduce tracking ID switching. This study uses Deep SORT multi-target tracking.

2.2.3. *Clustering algorithm.* A clustering algorithm is needed to obtain a group of people starting from the waiting point. The clustering method used in this study is created based on density-based spatial clustering of applications with noise (DBSCAN) [12].

DBSCAN is a density-based clustering algorithm proposed by Ester et al. in 1996 [12]. Given a set of points in some space, it groups together points that are closely packed together. Consequently, points with only nearby neighbors form a cluster. On the other

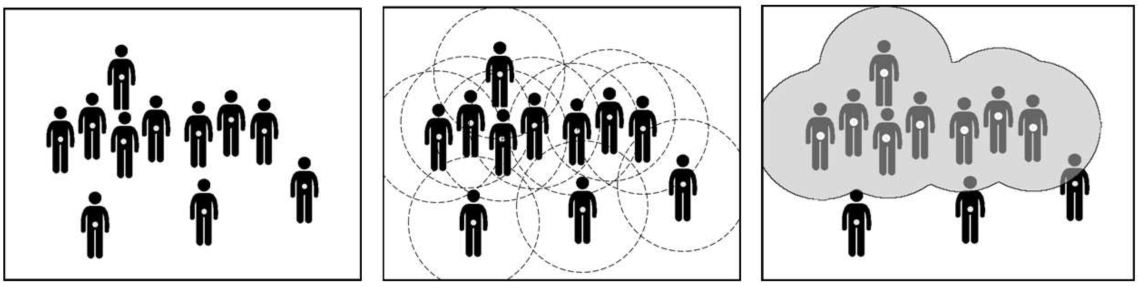

FIGURE 2. Density-based clustering – DBSCAN

hand, it marks as outliers points that lie alone in low-density regions whose nearest neighbors are too far away. As shown in Figure 2, the DBSCAN algorithm can find clusters of arbitrary shapes.

2.3. **System structure.** Figure 3 shows the structure of the proposed system. The input of the system is the video captured by a fixed camera and the position of the waiting point. The output is the position of the queue area and the number of people waiting. The entire system is divided into the following processes: queue detecting process, passerby deleting process, and people awaiting measuring process.

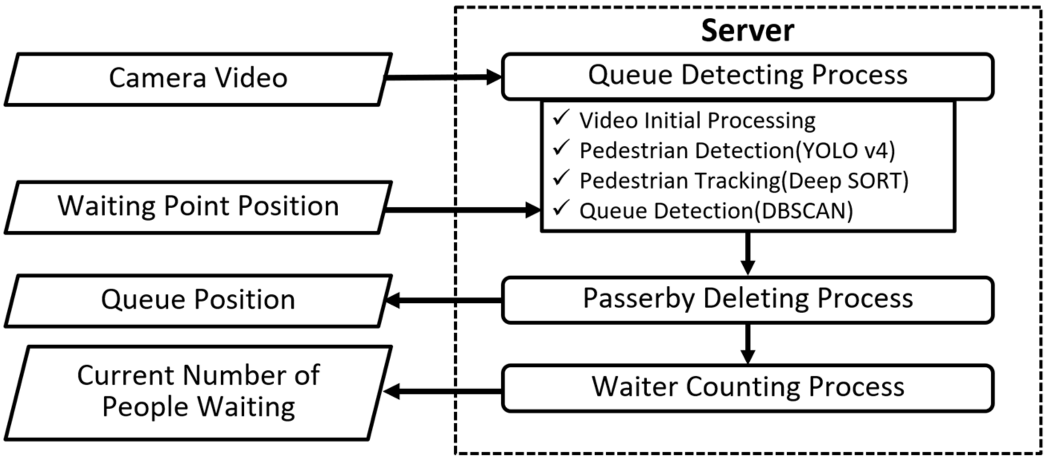

FIGURE 3. System structure

2.3.1. *Queue detecting process.* First, the captured video image is processed by the queue detecting process. This process divides the video image into frames and analyzes them. Next, the YOLOv4 object detection algorithm is used to detect the person from the image of each frame and record the person's position and the area of the bounding box. In addition, the position and bounding box information of each person are passed to the Deep SORT multi-target tracking algorithm. This algorithm combines image information, analyzes and records the characteristics of each person, assigns a unique ID to each tracking target, and tracks the person. At the same time, the system also records each person's current position and movement path for subsequent processing. After that, the queue detection is performed. The system uses the input waiting point position and combines the position information of each person to execute the density-based clustering algorithm. In this clustering method, as shown in Figure 1, the waiting point is set as the center of the circle, and the waiting person within the search range is searched. If the number of people in the search range is two or more, the search is repeated centering on the target waiting person. This allows the system to detect a tentative queue and obtain a group of people starting from the waiting point.

2.3.2. *Passerby deleting process.* In the passerby deleting process, we perform two processes, speed-based deletion and distance-based deletion, for the passerby around the queue.

Speed-based deletion: To delete passerby around the queue and people leaving the queue, the system computes the movement speed of each person gathered around the waiting point using the movement path information for each person obtained by the tracking algorithm. Since the detection algorithm is sometimes unstable and the position of a person in the neighboring frames may shift, we record the average of the moving distance (pixels) per 10 frames as the moving speed. Since the average moving speed of passerby around the queue is faster than the people in the queue, set a speed threshold to delete the fast-moving person from the queue.

Distance-based deletion: People around the queue are not always moving, so for example, the staff who stays around the queue does not belong to it. In order to delete these persons from the group, the distance-based deletion is necessary. The system uses the information on the person who is remained from deletion process of the speed-based approach. And the system calculates approximate line for the current queue based on the position of each person, computes the distance of each person in the group to the approximate straight line, and the person who is far away is deleted from the queue. Therefore, the system obtains the number of people waiting in the queue that is closer to reality.

The system parameters are determined based on the resolution of the input video and the size of the person on the video. If the resolution of the input video is 1920×1080 pixels and the average size of the person's bounding box is about 120×350 pixels, the search range of the clustering is set to 350 pixels based on the long side of the person's bounding box. In the deleting process, the threshold for speed-based deletion is set to 30 pixels per 10 frames, and the maximum distance for distance-based deletion is set to 120 pixels based on the short side of the person's bounding box.

2.3.3. *Waiter counting process.* After going through the passerby deleting process, the system uses the remaining person's information in the queue to calculate the number of people waiting. In addition, the system redraws the approximate line of the queue and displays the queue area on the output video.

3. **Experiment.**

3.1. **Experimental environment.**

Figure 4 shows the fixed camera video used in the experiment. This video includes a queue for baggage inspection at the airport. The resolution of the video is 1920×1080, the frame rate is 25 fps, and the length is 195 seconds. This video includes not only the scene where there are few people around the queue as shown in Figure 4(a), but also the scene where there are passersby around and people passing through the queue as shown in Figure 4(b). For comparison, we use two methods: the

(a) (b)

FIGURE 4. Camera video used in the experiment – (a) Scene A; (b) Scene B

method based on Wuthoo and Bedarkar [7] that measures all people on the screen using person recognition and tracking (conventional method), and the method that introduces queue area detection and passerby deletion (proposed method).

The accuracy of the number measurement of each method is evaluated using the following Equation (1). #(Ground Truth) represents the actual number of people in the queue measured manually, and #(Count) represents the number of people in the queue output by the system.

$$1 - \frac{|\#(\text{Ground Truth}) - \#(\text{Count})|}{\#(\text{Ground Truth})} \times 100(\%) \tag{1}$$

In addition, in the person detection and tracking algorithm, the conventional method and the proposed method use the same pre-training model and conduct experiments in the same hardware environment. The CPU used is Intel Core i7-10700, the memory capacity is 16 GB, and the GPU used is NVIDIA GeForce RTX 2060 SUPER.

3.2. **Results and analysis.** In the experiment, the processing speed of the entire system reached 7 fps. Figure 5 shows the experimental video processed by the proposed method. The gray point on the video is the waiting point. The recognized and tracked person is surrounded by a gray bounding box, and the tracking ID is displayed as a white number in the upper left. The system outputs queue position is displayed on the video as a gray line, and the people in the queue determined by the system are marked with a white dot. Figure 5(a) shows that the passerby with tracking ID-30 is approaching the queue and trying to pass through the queue. Since this person does not belong to the queue, the system does not put a white dot mark as shown in Figure 5(b). Also, although the person with tracking ID-28 is standing around the queue, he does not belong to the queue, it can be confirmed that the system made the correct processing and did not mark the white dot. From the above, the effectiveness of the proposed method is confirmed.

(a) (b)

Figure 5. Experimental video processed by the proposed method – (a) Scene A; (b) Scene B

The experimental results of each method are shown in Figure 6. The statistical frequency of the number of people waiting used for the experimental results is once every 3 seconds. In Figure 6, the horizontal axis is time and the vertical axis is the number of people. The dotted line is the result of the conventional method, the black line is the result of the proposed method, and the double dotted line is the actual number of people waiting measured manually. The measurement result of the proposed method is closer to the actual number of people waiting than the conventional method. It is considered that the reason is that the erroneous addition by passersby around the queue was effectively reduced.

Table 1 shows the accuracy of waiting people measurement within the entire experimental video time. From Table 1, the average accuracy of the proposed method exceeds 90.00%, reaching 93.74%. At the same time, the average error is less than 1.0 person,

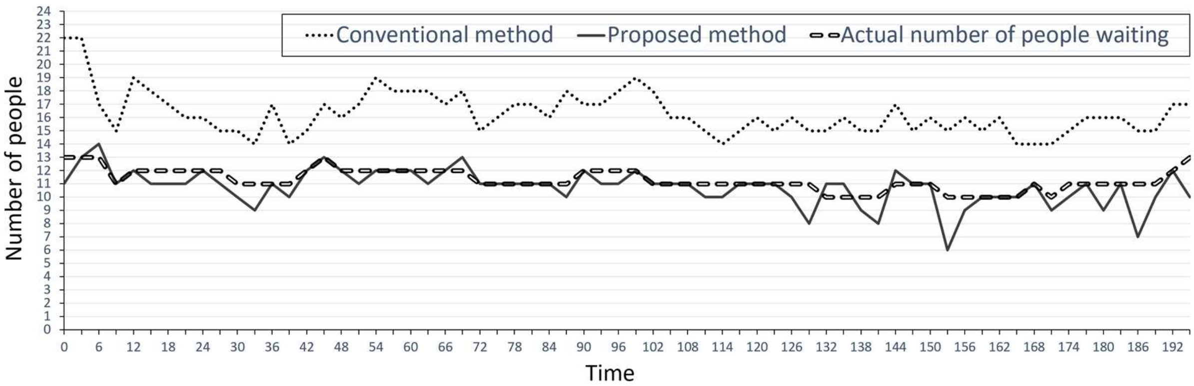

Figure 6. Experimental results – Comparison of the number of people measured

Table 1. Experimental results – Comparison of measurement accuracy

	Conventional method	Proposed method
Average error (person)	5.0	0.7
Average accuracy (%)	55.68%	93.74%

which is 0.7 person. Compared with the conventional method, the average accuracy of measurement is improved by 38.06%. From the above, it can be confirmed that the proposed method shows sufficiently high accuracy.

On the other hand, the graph after 126 seconds in Figure 6 shows that the number of people waiting output by the proposed method is less than the actual number of people waiting. Looking at the experimental video of the relevant part, the middle of the queue is sometimes interrupted, as shown in Figure 7. This prevents the person behind from entering the search area of the person in front and the person behind is not measured, so the system outputs fewer people than the actual number of people waiting. To address this issue, it is necessary to improve the queue area detection and adjust search range parameters.

Figure 7. Experimental video processed by the proposed method – Scene C

4. **Conclusions.** In this paper, we proposed a waiting person measurement method that introduced queue area detecting process and passersby deleting process based on machine learning, and it reduces erroneous addition by people who do not belong to the actual queue. As a result, we have achieved a highly accurate measurement of the number of people waiting in the queue, even in complex situations where there are many passersby around the queue.

This time, we found that the system is difficult to handle when the queue is interrupted. It is necessary to improve the queue area detection. In addition, we have realized the detection of single straight-line queuing in this paper, but we think that the detection of multiple queues and the curved queue should be considered as future tasks.

As a future development, we are also considering predicting the future number of people waiting by combining factors such as weather and date based on the historical data of number of people waiting.

REFERENCES

[1] D. Yanagisawa, Y. Tanaka, R. Jiang, A. Tomoeda, K. Ohtsuka, Y. Suma and K. Nishinari, [The length of the queue = the number of people waiting] is it correct?, *Reports of RIAM Symposium*, vol.21ME-S7, no.33, pp.210-215, 2010 (in Japanese).

[2] H. Zhang, J. Huang, Z. Song and M. Yang, Detection method of the number of people of indoor queues based on background difference method, *Electronic Measurement Technology*, vol.42, no.14, pp.123-126, 2019.

[3] R. Shbib and M. Merhi, Queueing detection system using an integrated pedestrian flow and optical regression, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol.10, no.12, pp.75-94, 2017.

[4] J. Wang, Y. Wang, D. Zhang, L. Wang, C. Chen, J. W. Lee and Y. He, Real-time and generic queue time estimation based on mobile crowdsensing, *Frontiers of Computer Science*, vol.11, no.1, pp.49-60, 2017.

[5] F.-J. Wu and G. Solmaz, Are you in the line? RSSI-based queue detection in crowds, *2017 IEEE International Conference on Communications (ICC)*, pp.1-7, 2017.

[6] A. R. Lejbølle, B. Krogh, K. Nasrollahi and T. Moeslund, One-to-one person re-identification for queue time estimation, *2020 IEEE International Conference on Image Processing (ICIP)*, pp.1706-1710, 2020.

[7] D. Wuthoo and K. Bedarkar, Intelligent queue management system at airports using image processing and machine learning, *International Research Journal of Engineering and Technology (IRJET)*, vol.6, no.10, 2019.

[8] L. Wang, J. Gui, Z.-M. Lu and C. Liu, Fast pedestrian detection and tracking based on ViBe combined HOG-SVM scheme, *International Journal of Innovative Computing, Information and Control*, vol.15, no.6, pp.2305-2320, 2019.

[9] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *arXiv 2020*, arXiv: 2004.10934, 2020.

[10] N. Wojke, A. Bewley and D. Paulus, Simple online and realtime tracking with a deep association metric, *2017 IEEE International Conference on Image Processing (ICIP)*, pp.3645-3649, 2017.

[11] A. Bewley, Z. Ge, L. Ott, F. Ramos and B. Upcroft, Simple online and realtime tracking, *2016 IEEE International Conference on Image Processing (ICIP)*, pp.3464-3468, 2016.

[12] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, *KDD*, vol.96, no.34, pp.226-231, 1996.