

## Webcam-Based Bus Passenger Detection System Using Single Shot Detector Method

Sigit Wasista\*<sup>1</sup>, Zulfan Indra Nugraha<sup>2</sup>, Setiawardhana<sup>3</sup>

<sup>1-3</sup>Departement of Informatics and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Indonesia

e-mail: \*<sup>1</sup>wasista@pens.ac.id, <sup>2</sup>zulfanugraha743@gmail.com, <sup>3</sup>setia@pens.ac.id

### Abstrak

*Bus merupakan salah satu transportasi yang banyak dipilih sebagai pendukung mobilitas masyarakat Indonesia. Mobilitas yang sering dijumpai selain sebagai transportasi umum, juga banyak dijumpai mobilitas kegiatan tur pariwisata untuk suatu kelompok perjalanan. Banyaknya tempat tujuan pariwisata yang dituju sehingga penumpang naik dan turun, membuat asisten supir bus atau ketua rombongan bekerja keras untuk memastikan jumlah penumpang yang naik ke dalambus sesuai dengan jumlah rombongan. Sering kali lamanya waktu yang dibutuhkan untuk memastikan ketepatan jumlah penumpang sebelum keberangkatan ke tujuan berikutnya. Cara konvensional tersebut mengakibatkan kemoloran jadwal tur pariwisata. Dalam proyek akhir ini penulis merancang sistem deteksi wajah penumpang bus berbasis webcam menggunakan metode Single Shot Detector (SSD) yang dapat memberikan informasi secara real time kepada supir bus, asisten supir bus atau ketua rombongan. Hasil yang didapatkan sistem mendapatkan capaian sebesar 95% dari total pembuatan sistem beserta pengujian pendeteksian wajah penumpang bus pada kondisi aktual menghasilkan rata-rata akurasi 77,5%.*

**Kata kunci**—Sistem Deteksi Wajah, Bus, Single Shot Detector, Webcam, Pemrosesan Gambar Digital

### Abstract

*Buses are one of the most widely chosen transportation methods to support the mobility of the Indonesian people. Mobility that is often found in addition to public transportation, is also often found in the mobility of tourism tour activities for a travel group. The number of tourist destinations to which passengers go up and down makes the assistant bus driver or group leader work hard to ensure that the number of passengers boarding the bus matches the number of groups. It often takes a long time to ensure the accuracy of the number of passengers before departure to the next destination. This conventional method results in the delay of the tourism tour schedule. In this research, the author designs a webcam-based bus passenger face detection system using the Single Shot Detector (SSD) method that can provide real-time information to bus drivers, assistant bus drivers or group leaders. The results obtained by the system obtained an achievement of 95% of the total system creation along with testing the detection of bus passenger faces in actual conditions resulted in an average accuracy of 77.5%.*

**Keywords**—Face Detection System, Bus, Single Shot Detector, Webcam, Digital Image Processing

## 1. INTRODUCTION

Traveling is one of the needs of some Indonesians. Based on a survey organized by Kompas Research and Development, it shows that the majority of respondents have traveled. From the data of the Ministry of Tourism, the number of domestic tourists in 2014 reached 251 million people, which means that the number of Indonesians traveling to tourist attractions is close to the population of Indonesia [1].

Tourist buses are the type of transportation that is often chosen by tourists, with the availability and scheduling system that is more flexible in setting schedules at tourist destinations. Seeing the demand from tourists regarding this transportation need, many private companies are competing to provide comfort and safety while still paying attention to the level of comfort based on the number of passengers [2].

Tourists as consumers of tourist transportation are not infrequently harmed by the transportation service provider with the existence of several problems that occur such as delays in tour departure schedules due to the lack of rapid delivery of information on the number of passengers on the tour bus in each passenger up and down, both when the rest schedule or visiting schedule at one tourist spot, due to manual counting by the assistant bus driver or group leader. In fact, it is inevitable that passengers are left behind, so that the tourism bus makes a U-turn to the place where the passengers are left, which results in further increasing the delay in the bus schedule [2].

To overcome and as a preventive measure for possible errors due to lack of information on the number of passengers before departure, through the development of artificial intelligence and internet networks, various *Automated Process Control* (APC) technologies have been developed. Some of them use infrared methods, pressure sensors, or use *computer vision systems*. *Infrared* is the most widely used technology, but it is stated that this method has poor accuracy if the number of people passing through the sensor is too large. [3]

The number of passengers can also be calculated based on seat status by placing a pressure sensor on each seat. This method also does not guarantee the accuracy of the data provided, because the system will only detect based on weight (it could be that what is placed on the chair is goods and not people). This method also requires high *costs* due to the placement of *hardware* in each chair [4].

Based on the description of the problem above, the author chose the SSD (*Single Shot Detector*) method. SSD is one of the popular approaches from three approaches namely SSD, *You Only Look Once* (YOLO) and *Region- proposal Convolutional Neural Network* (RCNN). RCNN provides accurate results but a long process while *YOLO* provides fast results but less accurate than RCNN, therefore *SSD* is the best choice, which provides more accurate results than *YOLO* although not more accurate than *RCNN* and provides faster results than *RCNN* although not faster than *YOLO*, even so *SSD* also has several disadvantages such as not being very good at detecting small objects [5]. With this research, it is expected that the system can detect bus passengers through faces that have a high level of accuracy and have a high level of efficiency so as to minimize passengers who are left behind on tourism tours.

## 2. METHODS

In Figure 1 is a block diagram of the system to be designed. The diagram can be described as follows:

1. Video data input is the initial input in this research, where video images can be obtained from camera recordings installed on the bus, in a position at the front, so that all passengers can be seen clearly.
2. Using the camera hardware system, image acquisition is the process of obtaining images from video sources. This is the most crucial phase in this method's workflow because erroneous drawings will render the entire process ineffective. To ensure that machine vision

systems are able to interpret digital photographs of objects rather than the actual thing, it is crucial to obtain images with appropriate contrast and clarity.

3. Grayscale images have a pixel depth of 8 bits, or 256 degrees of gray. They are processed with the lowest intensity value representing black and the greatest intensity value representing white.
4. Proceeded with The process of altering an image's resolution or its horizontal and vertical dimensions is known as resizing. Reducing or increasing the resolution as necessary will change the resolution size. The device can then detect every face seen inside the video frame. After the face is discovered, a drawbox appears to indicate that the face has been detected.
5. The process of converting unprocessed data into numerical features that may be handled while keeping the information in the original data set is known as feature extraction. Compared to directly applying machine learning to the raw data, it produces better outcomes.
6. Face alignment recognizes the geometric structure of people's faces in digital photographs. The shape of facial features, such as the nose and eyes, is automatically determined based on face position and size. The face alignment algorithm is used to iteratively change a deformation model that encodes past knowledge about the shape or appearance of the face.
7. Face detection is a technique that recognizes a person's face from digital photos or objects acquired by a camera. Face detection's primary job is to identify faces rather than recognize and analyze them. Then, this technology can be used to monitor and track someone in realtime. Following the final phase, a complete face may be described, including the detection results in the bounding box, facial markers, and derived descriptors. The detection can then be displayed by creating a bounding box on the canvas and delivering a full face description, as used in facial recognition.

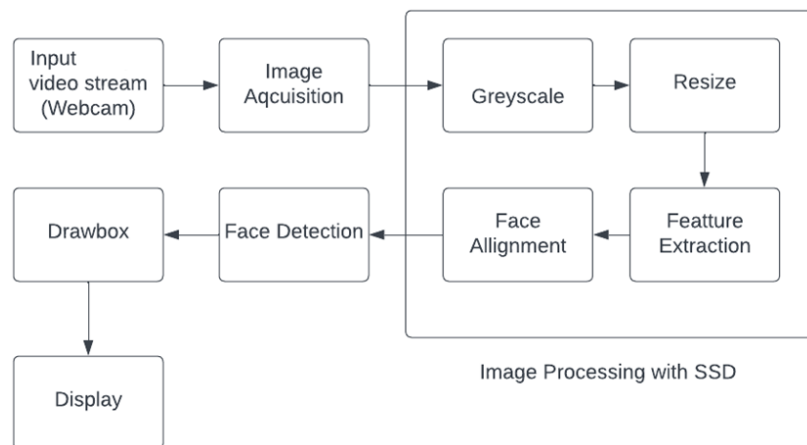


Figure 1 Block Diagram of System

### 2.1 Capture Image

At this stage, images will be taken by activating the videostream function with script 1.

```

// try to access users webcam and stream the images
// to the video element
const stream = await navigator.mediaDevices.getUserMedia({ video: { } })
const videoEl = $('#inputVideo').get(0)
videoEl.srcObject = stream
  
```

Script 1 Access users webcam

Then it will open the webcam display that has been integrated on the computer. Figure 2 is a webcam display that has been accessed to the web and has been integrated with the computer.



Figure 2 webcam display

## 2.2 Image Acquisition

The first stage of this system is *image* acquisition, which is taking images from various sources such as *datasets* or *images*, which are generated by a *webcam* integrated with a computer and will later be processed *training* and data validation using a *single deep neural network* technique. Illustration of the placement of this detection system can be illustrated in Figure 3 and 4.

## 2.3 Processing Single Shot Detector

Single Shot Detector (SSD) is a method used to detect certain objects in images or videos using a single deep neural network and Single Shot Detector (SSD) is one of the most popular object detection algorithms used because it has better accuracy and speed of image processing when compared to other methods such as FasterR-CNN, YOLOv1. Figure 5 is the architecture of the Single Shot Detector (SSD).

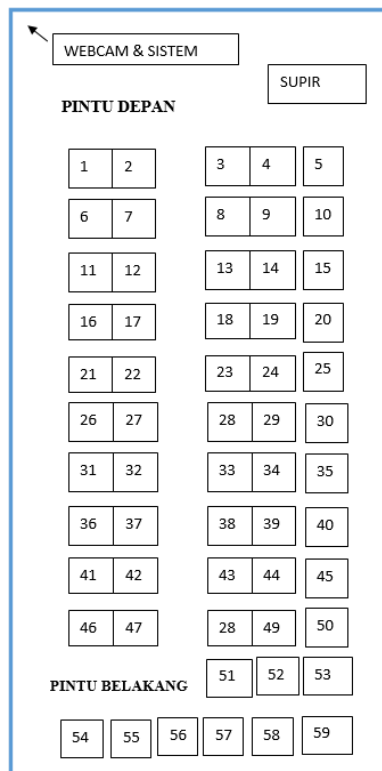


Figure 3 Bus Plan

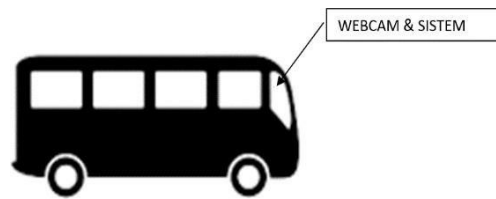


Figure 4 Bus Illustration

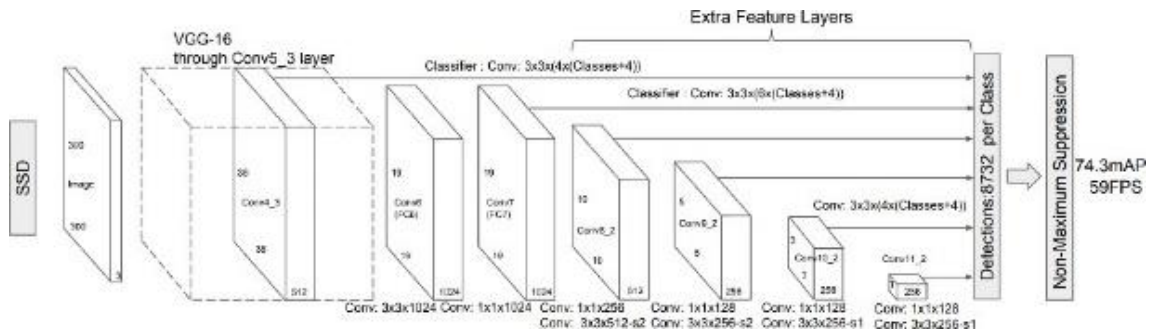


Figure 5 Single Shot Detector (SSD) Architecture

In the face recognition and identification process based on face recognition using the SSD method, there are also some simple programming paths where the face recognition process starts, the system stops when the condition is met, and the program resumes when the condition is not met. repeating the initial process until the system detects the face. The author's face recognition program uses Visual Studio Code and XAMPP and works online. The algorithm used by the author is Single Shot Detector.

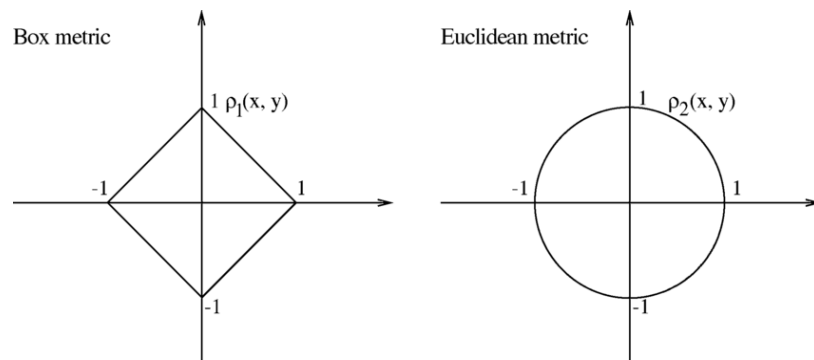
#### 2.4 Face Alignment

Face alignment is the following phase. In this instance, a modest yet precise 68-point facial landmark predictor from Face-api.js is employed. The entire model fits in web contexts and weighs 200 kilobytes. The expected 68-point landmarks are then aligned with conventional landmarks using this procedure. A face alignment example is shown in Figure 6.

The face has then been positioned for facial recognition by the system. Architectures like Single Shot Detector (SSD) are used to calculate facial descriptors in order to conduct facial recognition. A vector with 128 values that depicts a face in a 128-dimensional vector space is called a face descriptor. Furthermore, the euclidean metrics demonstrate the similarities between several vectors. In Figure 7, the euclidean metric is displayed.

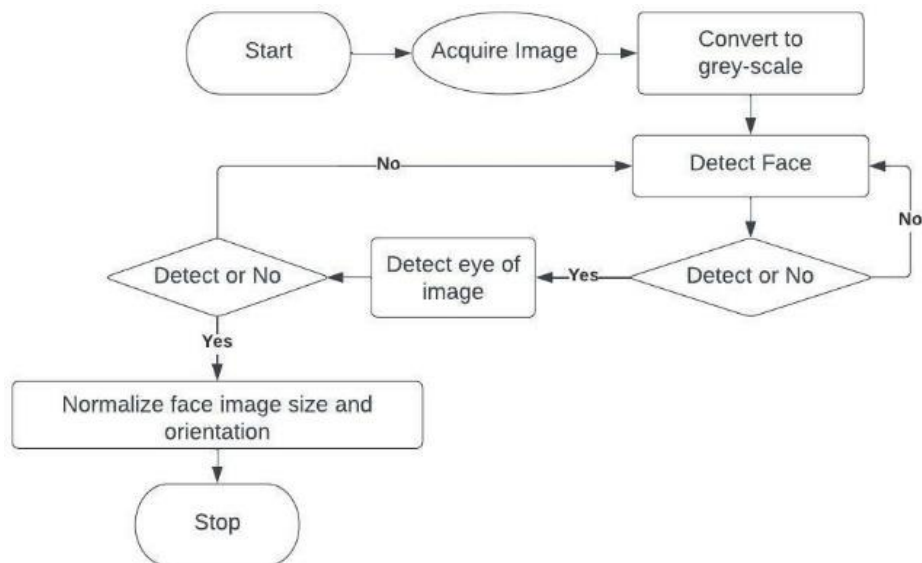


Figure 6 face alignment example

Figure 7 Metric *Euclidean*

## 2.5 Face Detection and Tracking

An essential first step in identifying human faces is face detection, which divides the image into two halves. The program will have very big computing limits because the first half will recognize faces and the second part will represent non-face areas. As a result, in order to reduce computation in an active program, one must ascertain a calculation limit that reduces the calculation area. In addition, face detection performs the flexible localization of the face region, enabling it to recognize all inputs inside the frame area and distinguish between faces and non-faces [4][15]. In this work, the Single Shot Detector (SSD) technique is used for face detection. Now, the flowchart will be displayed in Figure 8.

Figure 8 Flowchart of *Face Detection and Tracking*

After completing the face detection and tracking procedure, the process of detecting faces in the bus is depicted in Figure 9. The method of monitoring unique features, such points and contours, in a facial image is known as feature extraction and tracking. The nose, lips, eyes, eyebrows, and other facial features are among those that have been removed. The accuracy of face detection will also be high if the computation on the facial contour is successful in yielding a decent value. Face contours are identified by scanning the first frame of the video, where the first pixel representing the skin color that is identified is displayed as the first contour point (either the left or right side of the head).

Figure 10 is the outcome of feature extraction on the face. When identifying human facial emotions in real time, two factors must be considered: accuracy and efficiency. Face detection and tracking, feature extraction and tracking, feature reduction, and finally separation are all necessary processes.



Figure 9 Face Detection on the Bus



Figure 10 Face feature extraction results

## 2.6 Drawbox counter

At this stage the program will provide a drawbox in which there are faces of the bus passengers. Then all drawboxes that appear on the system will be totaled. An example of a drawbox that appears on the display can be shown in Figure 11.

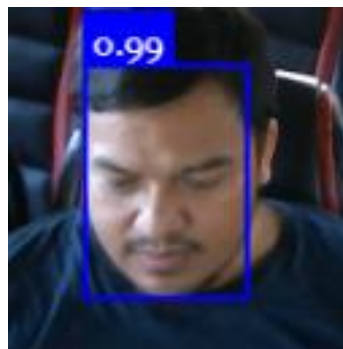


Figure 11. drawbox

Drawboxes will rise if a large number of drawboxes emerge on the system display, and the results of the total drawboxes that appear will also be presented on the system display, as will the accuracy level of the passenger face detection system.

### 3. RESULTS AND DISCUSSION

#### 3.1 Testing real time face detection based on time and conditions

At this stage the author has done face detection. The test is based on predetermined time and conditions. The test results are shown in Tables 1, 2, Figures 12, and 13.

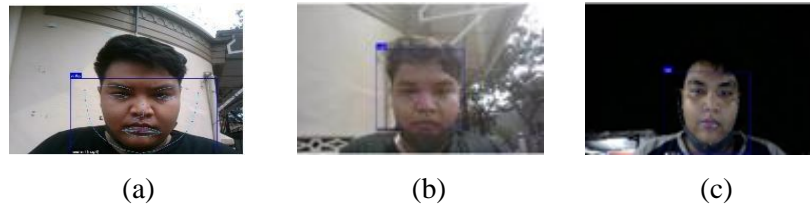


Figure 12 (a) in the morning at a distance of 30 cm, (b) in daylight with a distance of 30 cm, and (c) at night with a distance of 30 cm

Table 1 Test results based on time

No	Identity	Result
1	Figure (a)	Detected
2	Figure (b)	Detected
3	Figure (c)	Detected

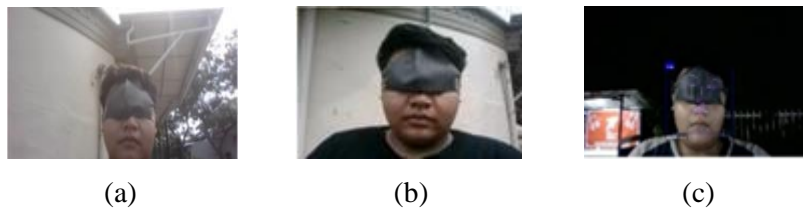


Figure 13 (a) in the morning and obstructed objects, (b) in daylight and obstructed objects, and (c) at night and obstructed objects

Table 2 Test results based on conditions

No	Identity	Result
1	Figure (a)	Detected
2	Figure (b)	Detected
3	Figure (c)	Detected

Based on test Tables 1 and 2, it is concluded that the system can perform face detection with variable brightness conditions in the morning, afternoon and night, and can distinguish the detection of faces that are detected intact or partially covered faces.

#### 3.2 Testing face detection based on seat distance

At this stage, the author has tested face detection using objects that resemble faces to find out the system can distinguish between objects that resemble humans and human faces.

Based on Table 3 and Figure 14, it can be seen that the system cannot detect objects that resemble objects and can only detect human faces so that it does not affect the performance of the system.



Table 3 Test results based on seating position

No	Identity	Result
1	Figure (a)	Detected
2	Figure (b)	Detected
3	Figure (c)	Detected
4	Figure (d)	Detected
5	Figure (e)	Detected
6	Figure (f)	Detected
7	Figure (g)	Detected

### 3.3 Testing face detection under actual conditions

At this stage, the author has tested face detection with the original conditions, namely on a bus with many passengers.

Based on Figure 15 and Table 4 the detection of bus passengers by detecting faces through cameras installed towards passengers in Table 4 with a total of 5 samples resulting in an average of 77.5%. This test was conducted on 16 passengers calculated from the actual camera's point of view. From these tests, errors are still found when there are still passenger faces that have not been detected perfectly. This is due to the condition of passengers on the bus who often move and/or do not point to the camera. This can be minimized by instructing all passengers to look at the camera as final data before ensuring departure by the assistant bus driver or group leader. In addition, the placement of cameras with a viewing angle that can cover all passenger faces will provide more optimal data.

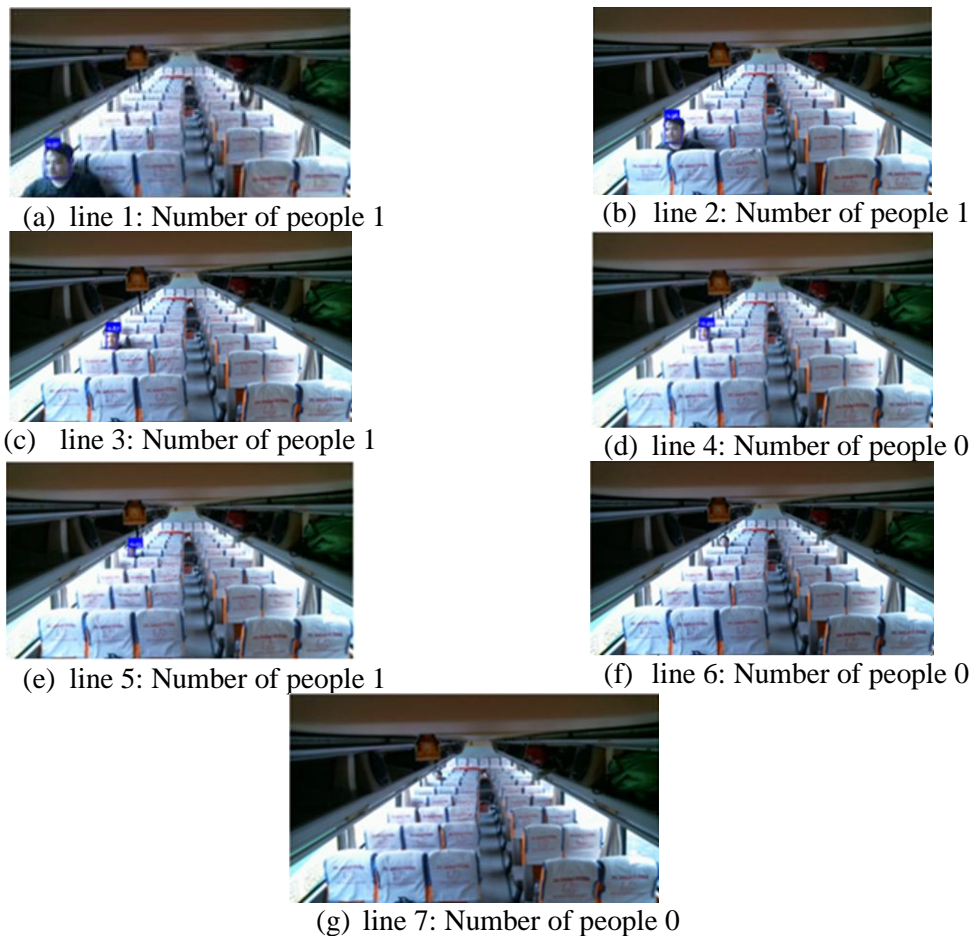


Figure 14 a) line 1, (b) line 2, (c) line 3, (d) line 4, (e) line 5, (f) line 6, and (g) line 7

Table 4 Test results based on actual conditions

No	Identity	Accuracy
1	Figure (a)	75 %
2	Figure (b)	81,25 %
3	Figure (c)	81,25 %
4	Figure (d)	75 %
5	Figure (f)	75 %



Figure 15 (a) condition 1, (b) condition 2, (c) condition 3, (d) condition 4, and (e) condition 5

#### 4. CONCLUSIONS

This research shows that the system can detect faces with an accuracy of 77.5% based on the results of the data that has been done. Face detection uses features extracted in the form of eyebrows, eyes, mouth, nose and others. The bus passenger detection system can be applied. For further development, it can provide information on the number of passengers in bus operations in several sessions.

#### REFERENCES

- [1] JUNAEDI, Muhamad Asyudin; HIDAYAT, Rahmat. Analisis preferensi konsumen dalam memilih bus pariwisata (Studi kasus pengguna bus pariwisata masyarakat Bandung tahun 2018). *eProceedings of Applied Science*, 2018, 4.3.
- [2] Putrawan, I., Sarjana, I., & Sukranatha, A. (2019). TANGGUNG JAWAB PERUSAHAAN ANGKUTAN WISATA DALAM HAL TERJADI KETERLAMBATAN KEBERANGKATAN DAN KECELAKAAN LALU LINTAS (Study Kasus Pada PT. Sanjaya Kusuma Transport). *Kertha Semaya : Journal Ilmu Hukum*, 7(11), 1-15.

- [3] ADMAJA, Yosia Pradeska. *TA: Sistem Penghitung Jumlah Pengunjung di Restoran Menggunakan Kamera Berbasis Single Shot Detector (SSD)*. 2021. PhD Thesis. Universitas Dinamika.
- [4] Chen Z., K. R. (2019). Real Time Object Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility. *Proceedings of the 2019 Eighth International Conference on Emerging Security Technologies (EST)*, 1–6.
- [5] F. Li, F.-W. Yang, H.-W. Liang, and W.-M. Yang, “Automatic Passenger Counting System for Bus Based on RGB-D Video,” *Adv. Eng. Res.*, vol. 117, pp. 209–220, 2017, doi: 10.2991/eeeis-16.2017.29.
- [6] SUSILO, Agus; HENDRIYAWAN A, M. S. *RANCANG BANGUN SISTEM PENDETEKSI PELANGGARAN PADA JALUR BUSWAY BERBASIS CITRA DIGITAL MENGGUNAKAN METODE SSD (SINGLE SHOT MULTIBOX DETECTOR)*. 2019. PhD Thesis. University of Technology Yogyakarta.
- [7] SUKUSVIERI, Andrianto. *TA: Implementasi Metode Single Shot Detector (SSD) untuk Pengenalan Wajah*. 2020. PhD Thesis. Universitas Dinamika.
- [8] A. S. A. Nasir, N. K. A. Gharib, and H. Jaafar, “Automatic Passenger Counting System Using Image Processing Based on Skin Colour Detection Approach,” in *International Conference on Computational Approach in Smart Systems Design and Applications, ICASSDA 2018*, 2018, pp. 1–8, doi: 10.1109/ICASSDA.2018.8477628.
- [9] Review: SSD — Single Shot Detector (Object Detection). (2019). Retrieved 27 July 2022, from <https://towardsdatascience.com/review-ssd-single-shot-detector-object-detection-851a94607d11>
- [10] B. Yuliandra and Tjokorda Agung Budi, “People Counting menggunakan Extended CAMSHIFT dan Fitur Haar-like People Counting using Extended Camshift and Haar-like Features,” in *eProceedings of Engineering*, 2015, pp. 1–10.
- [11] ANDRE, Julfikar Ali. Sistem *Security Webcam* dengan menggunakan Microsoft Visual Basic (6.0). *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 2016, 1.2: 48-60.
- [12] LESTARI, RINI PUJI. *STUDI ANGKA KUMAN UDARA BUS TRANSJOGJA DAN BUS ANGKUTAN DALAM KOTA DI KOTA YOGYAKARTA*. 2010. PhD Thesis. Poltekkes Kemenkes Yogyakarta.
- [13] S. H. Khan, M. H. Yousaf, F. Murtaza, and S. Velastin, “Passenger Detection and Counting for Public Transport System,” *NED Univ. J. Res.*, vol. XVII, no. 2, pp. 35–46, 2020, doi: 10.35453/nedjr-ascn-20190016.
- [14] Chen Z., K. R. (2019). Real Time Objek Detection, Tracking, and Distance and Motion Estimation based on Deep Learning: Application to Smart Mobility. *Proceedings of the 2019 Eighth International Conference on Emerging Security Technologies (EST)*, 1–6.
- [15] PUTRO, M. Dwisnanto; ADJI, Teguh Bharata; WINDURATNA, Bondhan. Sistem Deteksi Wajah dengan Menggunakan Metode Viola-Jones. In: *Seminar Nasional “Science, Engineering and Technology*. 2012.
- [16] DESHMUKH, Shubhada; PATWARDHAN, Manasi; MAHAJAN, Anjali. Survey on real-time facial expression recognition techniques. *Iet Biometrics*, 2016, 5.3: 155-163.
- [17] D. Takacs, A. Lorincz and J. A. Laszlo, High Quality Facial Expression Recognition in Video Streams using Shape Related Information only, 2010.
- [18] Feature Extraction - MATLAB & Simulink. (2023). Retrieved 4 July 2023, from <https://www.mathworks.com/discovery/feature-extraction.html>

- [19] KE, Qihong, et al. Computer vision for human-machine interaction. In: *Computer Vision for Assistive Healthcare*. Academic Press, 2018. p. 127-145.
- [20] SYAFRUDIN, Sukri, et al. Prediksi Status Pinjaman Bank dengan Deep Learning Neural Network. *Jurnal Khatulistiwa Informatika*, 2021, 7.2: 130-135.