# Detection of Congested Traffic Flow during Road Construction using Improved Background Subtraction with Two Levels RoI Definition

1st Yoanda Alim Syahbana
*Graduate School of Engineering*
*Gifu University*
Gifu, Japan
yoanda@pcr.ac.id

2nd Yokota Yasunari
*Department of Electrical, Electronics and Computer Engineering,*
*Faculty of Engineering*
*Gifu University*
Gifu, Japan
ykt@gifu-u.ac.jp

***Abstract*—This study is aimed to detect traffic congestion that may occur during roadblocks of road construction. We improved the background subtraction method by considering Region of Interest (RoI) in the video frame to detect the congestion. The proposed method has experimented with video test material that shows traffic condition in the road construction site. The performance of the proposed method is evaluated using Confusion Matrix by comparing the result of the experiment with ground truth obtained visually. As a benchmarking process, the performance is also compared with the conventional background subtraction method. The result shows that the proposed method can achieve an accuracy of 83.2% for video from the first camera and 82.3% for video from the second camera. In comparison, the conventional background subtraction method only achieves 49.8% for video from the first camera and 0% for video from the second camera. Based on this evaluation, the proposed method can support implementation of efficient traffic control using adaptive traffic light that is equipped with camera.**

***Keywords*—traffic surveillance, traffic congestion detection, background subtraction, stationary foreground object, Region of Interest (RoI)**

## I. Introduction

Road traffic management plays an essential role in the smart city concept. The management collects traffic information and analyzes driving trends in the area. Based on the information, the management controls the signal of the traffic light to adapt to the fluctuation of the traffic situation. In addition, the management also provides information for the driver so that the driving time will be shortened and traffic congestion can be avoided. In terms of traffic safety, the management also ensures traffic accident monitoring and contributes to better comfortability for the road traffic environment.

The video tracking system is the most popular approach to understand traffic conditions in road traffic management [1]. It uses a camera for surveillance purposes. Video frame from the camera is processed to identify and track a moving object. The moving object is further classified as a vehicle, and movement of the vehicle is tracked frame by frame. Instead of using electronic sensors embedded in the roadside, a camera provides a non-invasive approach and is easy to install [2], [3].

The background of this study is to enable efficient traffic control using adaptive traffic light equipped with a camera as illustrated in Fig. 1. Mainly, this study is focused on one-sided alternating traffic sections such as during roadblocks. Camera records traffic activity and sends the captured scene to CPU. CPU performs decision-making process to control the traffic light based on traffic information. In order to implement the system, some objectives are important to be achieved.

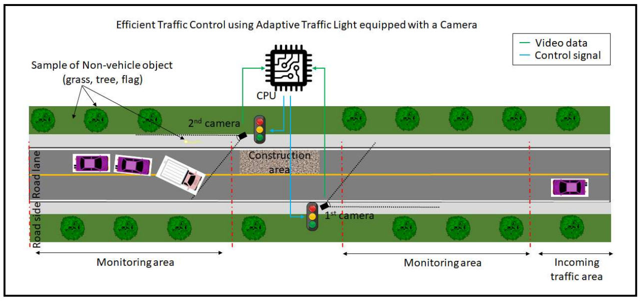

Fig. 1. Illustration of the road construction site.

First, the system needs to detect incoming traffic as early as possible in order sends control signal to the traffic light. Late detection of incoming traffic leads to late signaling of traffic light and this condition is dangerous for high-speed traffic. We had proposed method to achieve this objective in [4]. The result shows that the proposed method based on vanishing point concept has earlier detection of incoming traffic than detection using R-CNN method.

Second, the decision-making process is vital to control timing of traffic light signal based on traffic condition in monitoring area. Ideally, efficient traffic control is achieved when the traffic light signal can make minimum waiting time for both side of traffics. The waiting time is determined by length queuing traffic and waiting duration that can be defined as congestion level. However, since the monitoring area is an unpredictable outdoor environment, monitoring the traffic condition need to consider existence of non-vehicle object such as grass, tree, and flag. For example, existence of flag may masked the traffic position that cause error in traffic tracking. In addition, movement of non-vehicle objects can bias traffic detection and its condition (stopped or moving).

Therefore, this paper is aimed to report our research in supporting to achieve the second objective. We proposed a method to detect congestion conditions in the traffic. As test material, we use the same video used in [4] that shows the traffic controller managing the traffic flow near the road construction site. In this study case, only one traffic should be allowed at a time. The traffic stopped when the traffic controller raises the red flag. Then the traffic will continue when the traffic controller raises the white flag. To evaluate the proposed method's performance, we compare the timing of detected congestion from the proposed method with the timing of stopped traffic in the video. Timing of stopped traffic is obtained by watching every second of the video test material. Fig. 2 shows sample of captured scenes of the video test material.

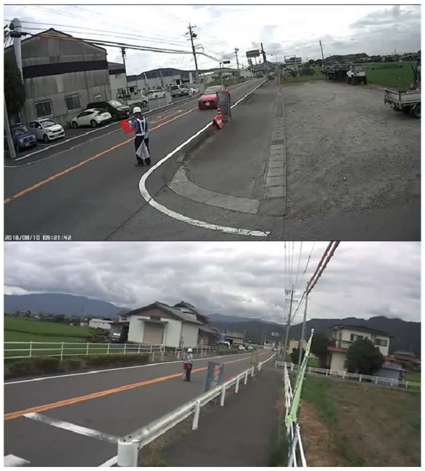

Fig. 2. Captured scene in the monitoring area.

The paper is organized into four sections, including this introduction section. Section 2 presents the design of the proposed method. It includes six essential steps of the proposed method. Section 3 explains the design of the experiment and the result of the method evaluation. Finally, Section 4 concludes the paper with a research conclusion and suggestions for future work.

## II. Proposed Method

### A. Design of the Proposed Method

Application for traffic detection in this study case requires a method to detect the traffic as Foreground Object (FO) in the scene. Most of the widely used methods are called background subtraction [5]. Background subtraction method can be further classified based on mathematical concept, machine learning concept, signal processing model, and classification model [6]. Basic techniques such as temporal median, temporal histogram, and filter are most employed for primary application. It is because the basic technique has low computation and minimum memory requirement.

This study case focused on a particular type of FO namely, Stationary Foreground Object (SFO). SFO is defined as an object that stops and remains static for several consecutive frames [7]. In this study case, SFO represents congested traffic that stops because the traffic controller raised a red flag as a stop signal.

Processing video frame to detect the congestion deals with some challenges. In this study case, the outdoor environment with a variation of sunlight exposure and wind that shakes shrub, grasses, trees, and flags influence the detection of traffic that passes the roadway. In addition, the passing traffic also moves with a variation of speed and direction. Using conventional background subtraction method [5], [8] leads to false detection of congestion.

Therefore, we improved this conventional method considering Region of Interest (RoI) existence in the video frame. The RoI improves the detection quality by focusing the detection in the roadway area only. This study proposed two levels RoI definition to focus the detection in roadway activity only. By this proposed RoI definition, unnecessary FO candidates such as cloud, shrub, grass, tree, and flag can occur. The first level of RoI is aimed to define the main region where traffic movement exists. This first level of RoI is defined in the video before the traffic control is applied. Then, the second level of RoI refines the first level of RoI to focus on the region where the traffic is stopped during the traffic control. This second level of RoI is defined in video when traffic control has been conducted.

Fig. 3 presents the design of the proposed method. First, the background subtraction process is performed for a sequence of video frames. This step detects FO in the video frame. Second, the first level RoI is defined based on the difference of FO from frame to frame. The difference represents the movement of the FO. Then, all of the FO movement is aggregated to obtain the region where the main movement exists in the captured scene. Third, background subtraction is performed once again, especially on RoI defined from the first step. Fourth, the FO detected from this background subtraction is further processed to obtain the second level RoI. In this step, FO that remains static for five sequential frames is defined as SFO. Fifth, the congestion metric is calculated to obtain the congestion timing of the SFO. Finally, resuming time is calculated by calculating the SFO difference from the sequential frame. Resume time is decided based on the maximum difference that represents the movement of the congested car.

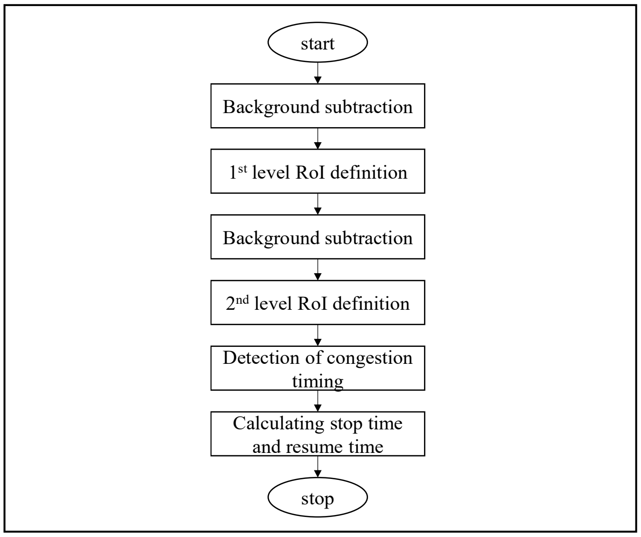

Fig. 3. Design of the proposed method.

The proposed method is designed based on the study case shown in Fig. 1. The main element of the method is using two levels of RoI to enclose the region where the congestion possibly happens. For different study cases with different conditions of the captured scene, the designed method can be adjusted. For example, other methods of FO detection other than background subtraction, such as Gaussian Mixture Model (GMM) or Neural Network, can be implemented with trade-off between accuracy and computation time. In addition, the thresholding process that is predefined in this method can be adjusted based on the condition of the captured scene. The design of the proposed method is also limited to processing traffic video recorded in daylight and fine weather.

### *B. Definition of Video Frame Variable*

Each video frame is represented as $\boldsymbol{I}(x,y,t)\in\{0,1,\ldots,255\}$ where $x\in\{1,2,\ldots,N_x\}$, $y\in\{1,2,\ldots,N_y\}$, and $t\in\{1,2,\ldots,T\}$ with $N_x$ and $N_y$ are width and height of video frame and $T$ is total of the video frame. This study downsamples video frames based on video frame rate to obtain frames for every second. Therefore, $t\in\{20/V_{fps},40/V_{fps},\ldots,T/V_{fps}\}$ where $V_{fps}$ is video frame rate.

### *C. Background Subtraction*

The Background subtraction method is used to detect FO in the evaluated video frame. First, the background frame is generated from the median value of the sequential frame before and after the evaluated video frame, $\boldsymbol{I}(x,y,t)$. Background frame is calculated using Eq. (1)

$$\boldsymbol{I_{BG}}(x,y,t) = \text{median}\big(\boldsymbol{I}(x,y,t-r),\ldots\boldsymbol{I}(x,y,t),\ldots,\boldsymbol{I}(x,y,t+r)\big) \quad (1)$$

where $\boldsymbol{I_{BG}}(x,y,t)$ and $r$ are generated background and range of sequential frame calculated using median, respectively. In this study, $r$ is predefined using Eq. (2)

$$r = \frac{T}{2} - 1 \quad (2)$$

where $r$ is defined based on observation from FO detection quality.

Background subtraction is calculated using Eq. (3)

$$\boldsymbol{I_{FO}}(x,y,t) = |\boldsymbol{I}(x,y,t) - \boldsymbol{I_{BG}}(x,y,t)| \quad (3)$$

where $\boldsymbol{I_{FO}}(x,y,t)$ is detected FO candidate from background subtraction calculation. The FO result includes noise from a variety of road textures and shadows under the car. In addition, the noise also can be a small insignificant object that is far away from the camera. The FO candidate result is filtered from noise using thresholding, blurring, and morphology process to remove the noise.

First, $\boldsymbol{I_{FO}}(x,y,t)$ that consist of $I_R(x,y,t)$, $I_G(x,y,t)$, and $I_B(x,y,t)$ is converted to grayscale image. Grayscale conversion is calculated using Eq. (4)

$$I_{FOgray}(x,y,t) = \frac{I_R(x,y,t)+I_G(x,y,t)+I_B(x,y,t)}{3} \quad (4)$$

where $I_{FOgray}(x,y,t)$ is grayscale conversion result from $\boldsymbol{I_{FO}}(x,y,t)$. To remove noise caused by road texture and shadow variation, $I_{FOgray}(x,y,t)$ with low intensity value lower than 30 is removed from $I_{FOgray}(x,y,t)$. The selection of 30 as a threshold value is predefined in this study. Removal of this type of noise is calculated using Eq. (5)

$$I_{FO}(x,y,t) = \begin{cases} 1, I_{FOgray}(x,y,t) > 30 \\ 0, otherwise \end{cases} \quad (5)$$

where $I_{FO}(x,y,t)$ is detected FO after $I_{FOgray}(x,y,t)$ thresholding. In addition, the morphology process, including blurring and dilation, is also applied to remove the noise.

### *D. First Level RoI Definition*

Since RoI's purpose is to localize the main region where the traffic movement exists, this study uses the frame difference method to obtain moving FO. This step has also been used in [4] for the different research objective as previously mentioned in Section 1. Frame difference is calculated using Eq. (6)

$$I_{FD}(x,y,t) = |I_{FO}(x,y,t) - I_{FO}(x,y,t+1)| \quad (6)$$

where $I_{FD}(x,y,t)$ is detected moving FO. Then, $I_{FD}(x,y,t)$ from all $t$ is summed up to obtain a single frame that shows the number of the region where moving FO existed.

In addition, thresholding is also applied to filter regions with little movement. The selection of 20 as a threshold value is predefined in this study. The process is calculated using Eq. (7)

$$ROI_{first}(x,y) = \begin{cases} 1, \sum_{t=1}^{T} I_{FD}(x,y,t) > 20 \\ 0, otherwise \end{cases} \quad (7)$$

where $ROI_{first}(x,y)$ represents the first level RoI. Among the existing region in $ROI_{first}(x,y)$, the largest region is defined as the first level of RoI because it covers most roadway areas. Fig. 4 shows $ROI_{first}(x,y)$ results from the first camera and second camera.

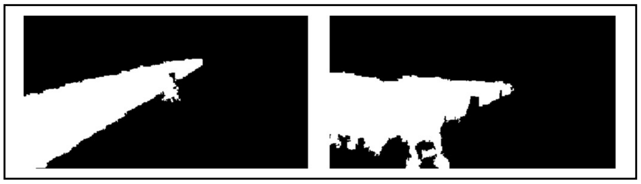

Fig. 4. First level of RoI from the first camera (left figure) and the second camera (right figure).

### *E. Second Level RoI Definition*

The second level of RoI refines the first level of RoI to focus on the region where the traffic is stopped during the traffic control. Background subtraction processes are repeated, and $I_{FO}(x,y,t)$ result is masked using $ROI_{first}(x,y)$. The masking process is calculated using Eq. (8)

$$I_{FO}(x,y,t) = I_{FO}(x,y,t) \cap ROI_{first}(x,y) \quad (8)$$

where $I_{FO}(x,y,t)$ is redefined as a masked result of the process to avoid notation complexity. This study defines SFO as $I_{FO}(x,y,t)$ that remains unchanged for five frames. The SFO is calculated using Eq. (9)

$$I_{SFO}(x,y,t) = I_{FO}(x,y,t-2) \cap I_{FO}(x,y,t-1) \cap I_{FO}(x,y,t) \cap I_{FO}(x,y,t+1) \cap I_{FO}(x,y,t+2), \quad (9)$$

where $I_{SFO}(x,y,t)$ is detected FO that remain static for five sequences of the frame.

Similarly, (7) is modified and used to obtain the second level of RoI. The process is calculated using Eq. (10)

$$ROI_{second}(x,y) = \begin{cases} 1, \sum_{t=1}^{T} I_{SFO}(x,y,t) > 10 \\ 0, otherwise \end{cases} \quad (10)$$

where $ROI_{second}(x,y)$ represent the second level RoI. $ROI_{second}(x,y)$ shows the region where SFO mainly exists, including stopped vehicles and standing traffic controllers. Refining this result, the selection of RoI in this step is based on the width of the RoI candidate. As observed in this study, the width of the stopped vehicle is wider than the width of the standing traffic controller. Fig. 5 shows $ROI_{second}(x,y)$ results from the first camera and second camera. The white region in the figure represents the second level of RoI.

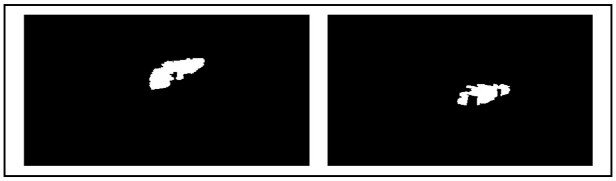

Fig. 5. Second level of RoI from the first camera (left figure) and the second camera (right figure).

The first level of RoI is defined using the video before the traffic control is applied. Therefore, movement activity is well captured. On the contrary, the second level of RoI is defined using the video after the traffic control is applied. Following these two processes, the rest of the video is processed using the background subtraction method. In addition, $I_{FO}(x,y,t)$ result is masked using $ROI_{second}(x,y)$. The masking process is calculated using Eq. (11)

$$I_{FO}(x,y,t) = I_{FO}(x,y,t) \cap ROI_{second}(x,y) \quad (11)$$

where $I_{FO}(x,y,t)$ is redefined as a masked result of the process to avoid notation complexity. Then, (9) is used to calculate masked $I_{SFO}(x,y,t)$ as a static FO that stopped for minimum of five frames.

### F. Detection of Congestion Timing

In order to obtain timing information of significant SFO, $I_{SFO}(x,y,t)$ is summed up for all of its $y$. The aim is to obtain the maximum horizontal width that is defined as the traffic congestion candidate. This approach is selected because most of the traffic movement is mainly progressed in the horizontal width of the video frame. This process is calculated using Eq. (12)

$$length_{SFO}(x,t) = \sum_{y=1}^{N_y} I_{SFO}(x,y,t) \quad (12)$$

where $length_{SFO}(x,t)$ is the length of SFO. Then, timing of the congestion candidate is calculated using Eq. (13)

$$congestion(t) = \begin{cases} \sum_{x=1}^{N_x} length_{SFO}(x,t), length_{SFO}(x,t) > 20 \\ 0, otherwise \end{cases} \quad (13)$$

where $congestion(t)$ is a metric to measure the congestion candidate.

### G. Calculating Stop Time and Resume Time

The timing represents the stop time of the front-most vehicle. However, it does not precisely provide information on the resume time. Therefore, frame difference is applied again for all $I_{SFO}(x,y,t)$ inside the timing. It also aimed to check whether the candidate is valid congestion or only slowed traffic. The process is calculated using Eq. (14)

$$I_{FD-SFO}(x,y,t) = |I_{SFO}(x,y,t) - I_{SFO}(x,y,t+1)| \quad (14)$$

where $I_{FD-SFO}(x,y,t)$ is absolute difference of $I_{SFO}(x,y,t)$. Then, $I_{FD-SFO}(x,y,t)$ is further processed to determine the congestion state of the traffic. The congestion state is calculated using Eq. (15)

$$congestion_{sum}(t) = \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} I_{FD-SFO}(x,y,t) \quad (15)$$

where $congestion_{sum}(t)$ is the congestion state that represents how the $I_{SFO}(x,y,t)$ changed inside the timing.

## III. Experiment and Evaluation

### A. Video Test Material

Video test material is obtained from In-Luck Company. This company provides services of security, including traffic control in road construction. The video test material shows two captured scenes from both sides of the roadway. It shows traffic flow that will pass the construction site controlled by the traffic controller. Captured scene from the first and the second camera shows the roadway and the traffic controller. The scene also shows the environment surrounding the road, such as the parking area and rice field. In addition, the captured scene also shows a waving flag on the roadside. These non-vehicle objects such as shrubs, grasses, trees, flags, and their movements influence traffic detection.

Sample of video test material is selected from the first and the second camera. The video has been synchronized to align the recorded duration between the cameras. Four sample videos recorded from 09:10 a.m. to 09:30 a.m from each captured scene from the first and the second camera are used for the experiment and evaluation. Each of the videos has a total of $T$=300 frames. The videos have $N_x$=640 and $N_y$=360 with $V_{fps}$=20.

### B. Experiment Setup

The proposed method is implemented using Matlab R2020b (64 bit) with the operating system Microsoft Windows 10 Pro. The method has experimented on Intel core i9 processor with 32 GB memory. This study also utilizes parallel processing to optimize the processing time of the video frame.

### C. Experiment Result

Fig. 6 shows $congestion(t)$ as a result of (13) for the video test material that has been processed by the proposed method. The vertical axis of Fig. 6 represents the congestion metric that shows the length of congested traffic. This metric along the horizontal axis represents the congestion, especially for the initial stop time of the congested traffic. As highlighted by the red ellipse, congested traffic is detected from the start of the video until frame $t$=25 in the first camera. After the congested traffic is detected in the first camera, the proposed method also detects congested traffic in the second camera from $t$=14 to $t$=50. Therefore, this timing pattern shows that the roadway is only accessed one at a time.

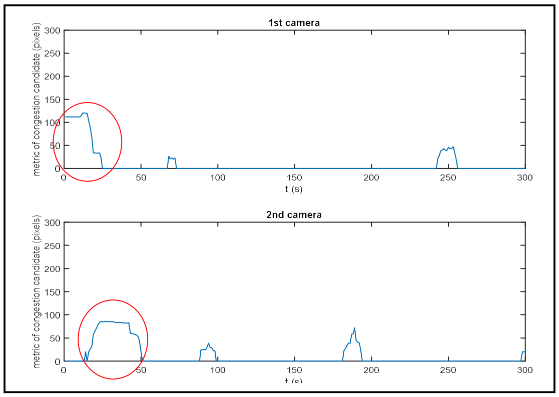

Fig. 6. Candidate of congestion timing.

Following the result discussed in Fig. 6, the proposed method processes the result to detect the resume time of the congested traffic. Eq. (14) and Eq. (15) are used to determine the resume time. For example, Fig. 7 shows the result of $congestion_{sum}(t)$ for $t$=1 to $t$=25 from the first camera. The peak of $congestion_{sum}(t)$ highlighted by the red circle represents the resume time of the congested traffic. In this example, $t$=17 is the resume time.

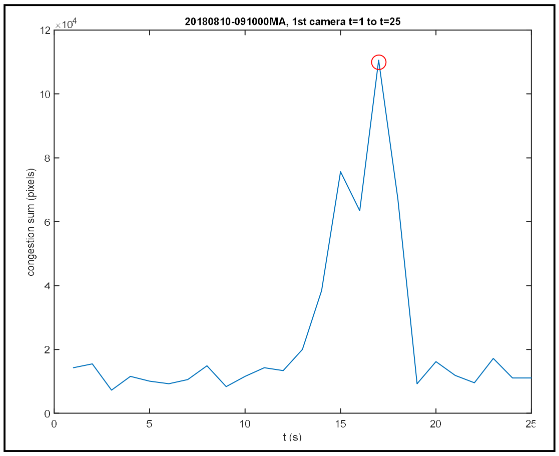

Fig. 7. Result of resume time calculation from the proposed method.

## D. Evaluation

This study uses Confusion Matrix to evaluate the performance of the proposed method. Confusion Matrix compares congestion timing from the proposed method with congestion timing from the ground truth. The ground truth is obtained by watching the video visually and record the timing manually. Table I shows the rule of Confusion Matrix that is used to evaluate the performance of the proposed method. Confusion Matrix compares congested and uncongested timing between the proposed method and the ground truth.

TABLE I. RULE OF CONFUSION MATRIX

| | | **Proposed Method** | |
|---|---|---|---|
| | | **Congested** | **Uncongested** |
| **Ground Truth** | **Congested** | True Positive (TP) | False Negative (FN) |
| | **Uncongested** | False Positive (FP) | True Negative (TN) |

Table II summarizes the result of the Confusion Matrix calculation from each sample of video test material that has been averaged.

TABLE II. RESULT OF CONFUSION MATRIX ELEMENT (%)

| | | TP | FP | FN | TN |
|---|---|---|---|---|---|
| Proposed method | First camera | 74 | 0.15 | 16.65 | 9.12 |
| | Second Camera | 82.25 | 17.75 | 0 | 0 |
| Conventional background subtraction method | First camera | 31 | 43.15 | 7.675 | 18.12 |
| | Second Camera | 0 | 100 | 0 | 0 |

Based on the rule, the accuracy of the proposed method is calculated using

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \qquad (16)$$

where $Accuracy$ is metric to evaluate performance of the proposed method.

Based on Confusion Matrix calculation, the proposed method can achieve an accuracy of 83.2% for video from the first camera and 82.3% for video from the second camera. In comparison, processing video using the conventional background subtraction method only achieves 49.2% for video from the first camera and 0% for video from the second camera. The existence of a waving flag causes a 0% result for video from the second camera detected as SFO.

The performance of the proposed method has been experimented on the captured scene, as shown in Fig. 1. Therefore, the design of the proposed method is fine-tuned with the study case and has some limitations. The proposed method's design has not considered a variation of captured scene conditions influenced by camera position, angle, and lens. For example, a high position capturing camera provides a different angle to capture traffic congestion. In this condition, conversion of world space to camera space is required. In addition, the captured scene is also influenced by the perspective projection that also needs to be considered. The study case also shows a straight road lane without an intersection. The existence of more than one road lane, intersection and curved road will influence estimated RoI and calculation of Eq. (15).

# IV. CONCLUSION AND FUTURE WORK

## A. Conclusion

This study has proposed a method to detect congested traffic that is happened during road construction. The proposed method improved the existing background subtraction method with two levels of RoI definition. The improvement aims to minimize the influence of non-vehicle objects in the captured scene and focus the detection on roadway areas only. The experiment of the proposed method has been conducted using video test material obtained from the In-Luck company. Performance evaluation using Confusion Matrix also has been performed. The result shows that the proposed method achieves better accuracy in detecting congestion than the conventional background subtraction method. Based on the benchmarking result, non-vehicle object movement such as waving flags influences congestion detection. In conclusion, the proposed method can support road traffic management in detecting traffic congestion, especially in temporary surveillance such as in road construction.

## B. Future Work

It is interesting to evaluate the further performance of the proposed method in a variety of environmental conditions, such as roads with more than one road lane, road intersection, and different weather conditions. It is also intriguing to implement other FO methods with a combination of two levels of RoI definition. In addition, it is also possible to continue to other objective of the research as mentioned in Section 1 that is designing a decision making process to achieve efficient traffic control using adaptive traffic light.

## References

[1] Y. B. B. Pethe, “An implementation of Moving Object Detection , Tracking and Counting Objects for Traffic Surveillance System,” 2011, doi: 10.1109/CICN.2011.28.

[2] M. Fathy and M. Y. Siyal, “A window-based image processing technique for quantitative and qualitative analysis of road traffic parameters,” *IEEE Trans. Veh. Technol.*, vol. 47, no. 4, pp. 1342–1349, 1998, doi: 10.1109/25.728525.

[3] D. M. Ha, J. M. Lee, and Y. D. Kim, “Neural-edge-based vehicle detection and traffic parameter extraction,” *Image Vis. Comput.*, vol. 22, no. 11, pp. 899–907, 2004, doi: 10.1016/j.imavis.2004.05.006.

[4] Y. A. Syahbana and Y. Yokota, “Early Detection of Incoming Traffic for Automatic Traffic Light Signaling during Roadblock using Vanishing Point-Guided Object Detection and Tracking,” in The SICE Annual Conference 2021, Tokyo, Japan, 2021.

[5] A. Sobral and A. Vacavant, “A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos,” *Comput. Vis. Image Underst.*, vol. 122, pp. 4–21, 2014, doi: 10.1016/j.cviu.2013.12.005.

[6] B. Garcia-Garcia, T. Bouwmans, and A. J. R. Silva, “Background subtraction in real applications: Challenges, current models and future directions,” *Comput. Sci. Rev.*, vol. 35, p. 100204, 2020, doi: 10.1016/j.cosrev.2019.100204.

[7] C. Cuevas, R. Martínez, and N. García, “Detection of stationary foreground objects: A survey,” *Comput. Vis. Image Underst.*, vol. 152, pp. 41–57, 2016, doi: 10.1016/j.cviu.2016.07.001.

[8] T. Bouwmans, “Traditional and recent approaches in background modeling for foreground detection: An overview,” *Comput. Sci. Rev.*, vol. 11–12, no. October, pp. 31–66, 2014, doi: 10.1016/j.cosrev.2014.04.001.