Journal of Computer Engineering, Electronics and Information Technology (COELITE)

JURNAL UPI

Journal homepage: https://ejournal.upi.edu/index.php/COELITE

Lost Centroid ID in CCTV Analytics for Slow Moving Object

Ulva Elviani[1*], *Hanavi*[2], *Fadhil Hidayat*[3]

[1]Universitas Pendidikan Indonesia, Indonesia
[2]Kementerian Luar Negeri Republik Indonesia, Indonesia
[3]Institut Teknologi Bandung, Indonesia
Correspondence E-mail: ulva@upi.edu

ABSTRACT

Accuracy is the most common problem experienced by traditional CCTV, especially people counting in locations where people are in the camera's field of view for a long time. This study proposes a method to address the lost centroid ID problem through a time delay mechanism in the object detection and tracking path using YOLO. This method can reduce the number of duplications and improve overall data reliability by maintaining object identities in predetermined frames, even when objects are temporarily undetected. Experiments were conducted on 100 CCTV video recording files in the Bandung train station area. The results showed that in locations in the slow-moving object category, such as waiting rooms, the accuracy increased significantly from 42% to 72% by applying a maxdisappeared threshold value of 200 frames. While at fast-moving object locations, a threshold of 40 frames increased the accuracy from 83% to 94%. This approach improves the performance of the people counting forecasting model for a more reliable surveillance system both statically and dynamically.

ARTICLE INFO

Article History:
Submitted/Received 10 Jun 2025
First Revised 17 Jun 2025
Accepted 27 Jun 2025
First Available online 28 Jun 2025
Publication Date 01 Oct 2025

Keyword:
CCTV Analytics,
Centroid ID,
People Counting,
Slow-moving object,
Video Surveillance.

1. INTRODUCTION

CCTV (Closed-Circuit Television) devices are still popular surveillance systems in public spaces. Although effective in recording and monitoring suspicious activity, CCTV, which works traditionally, only functions as a passive monitoring tool. Traditionally, this system requires officers to observe the footage and identify anomalies manually. These issues can cause delays in response and potential negligence. The analytical system embedded in CCTV or CCTV Analytics allows CCTV to monitor in real time and provide feedback to support the early warning system and decision-making process.

Computer vision and deep learning [1][2][3] enable CCTV to detect, classify accurately, and track objects [4]. The YOLO (You Only Look Once) algorithm [5] has been widely used in real-time object detection through video. This algorithm strongly detects objects with high processing speed [6]. Several studies, including those conducted by Barthélemy et al. [7] and Shakil et al. [8], have successfully detected individuals and monitored crowds in dynamic environments such as roads and pedestrian paths.

However, both studies still focus on fast-moving object environments, namely locations where individuals only cross the CCTV camera coverage area for a short time and do not stay in one spot for a long time. This problem is relatively more straightforward in the context of tracking and counting objects. In contrast, slow-moving object environments such as waiting rooms, service counters, or similar locations that require people to be in the camera's range for an extended period pose challenges in maintaining object tracking consistency [9], especially when there is temporary occlusion or stationary behaviour that causes the tracking algorithm to lose its centroid identity (ID) [10]. When the centroid ID is lost, the system assigns a new ID to the same object to be re-detected [1] so that the object is counted more than once [11]. Double-counting of objects significantly reduces the accuracy of the crowd [12] and will ultimately affect the reliability of the forecasting and decision-making models.

To overcome these problems, this study proposes an improved object detection mechanism by inserting a delay buffer to handle the loss of the centroid ID. The main idea of this study is to maintain the object identity for a specific period determined by the maxdisappeared parameter, even though the object is temporarily undetected [13]. For example, at a rate of 30 fps and a maxdisappeared threshold of 200 frames, the system will maintain the ID of a lost object for about 6.6 seconds before considering it permanently lost. The optimal delay value is determined through a trial-and-error approach based on analyzing CCTV footage at five locations in Bandung train stations. This mechanism helps prevent premature assignment of new IDs and reduces the risk of repeated object calculations.

This research's methodological basis lies in using the centroid ID as an object identity marker. Centroid tracking is a widely accepted approach in object counting systems, especially when facial recognition systems cannot be applied. Although face tracking provides more precise identification [14], its accuracy can decrease drastically when applied to conditions where the object is covered by specific attributes such as masks, head coverings, hair, or other obstacles [15]. Likewise, non-image processing [16], such as infrared sensors [17] or WiFi-based presence estimation methods [18][19][20], will experience obstacles in dense environments with overlapping objects. Therefore, this centroid tracking approach balances computational efficiency and tracking reliability.

However, the reliability of CCTV Analytics depends on the accuracy of the object-counting process. Inaccurate calculations can produce incorrect datasets, resulting in incorrect decision-making analysis. For example, overestimating or underestimating the number of

DOI: https://doi.org/10.17509/coelite.v4i2.85752
p- ISSN 2829-4157 e- ISSN 2829-4149

visitors in crowd management will impact resource allocation and safety and security planning [13]. Therefore, ensuring that each individual is not double-counted is important in producing valid analytical data. Based on these challenges, this study improves the detection and tracking pipeline by inserting a lost centroid ID-solving mechanism for slow-moving object environments. This solution is expected to improve the reliability of people counting, reduce double counting, and produce a clean dataset to support subsequent forecasting models. This study also bridges the practical needs of surveillance-based data analytics, especially in static or semi-static public spaces.

2. METHODS

The method of handling a lost centroid ID is to set the pause time when the system loses the ID. The pause time is determined by trial and error based on CCTV recordings at five Bandung station locations. The length of the pause time depends on the speed at which the system is running. If the speed is 30 fps with a maxdisappeared value of 200 frames, the system maintains the lost ID for 200/30 = 6.6 seconds. The flowchart of the pause time insertion in the YOLO algorithm can be seen in **Figure 1**. The figure generally explains the YOLO v3 algorithm, but is equipped with a tracking function and a proposed method for handling lost centroid ID. The lost centroid ID function insertion is placed on the shaded chart. The detection process begins with a video input resized to 416 x 416 pixels. Using a 3 x 3 kernel, convolution is performed to produce a new image. The new image is max-pooled to reduce the video input. If the output is not the same as 13 x 13 x 25, the convolution and max pooling processes are repeated. Meanwhile, if the output has met 13 x 13 x 25, then the thresholding score calculation and non-max suppression process are carried out. This stage is general in a YOLO algorithm for object detection.

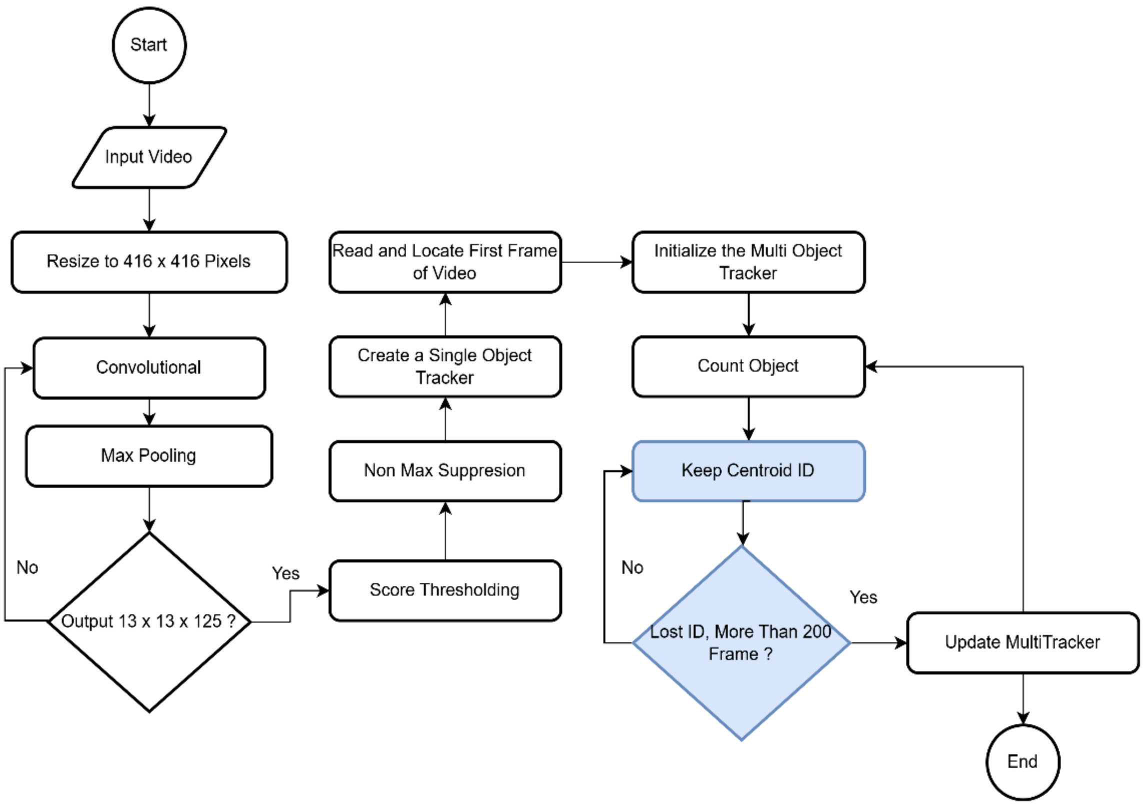

Figure 1. Lost Centroid ID solving flowchart

DOI: https://doi.org/10.17509/coelite.v4i2.85752
p- ISSN 2829-4157 e- ISSN 2829-4149

Research improvements are in the shaded section. When the object has been successfully detected, tracking is carried out to determine the object's location and provide labelling as a centroid ID number. The counting process occurs every time the system produces a new ID. Therefore, when a lost ID occurs, a pause time setting function is inserted so the system does not immediately update the 17 new IDs. The system will update the new ID if 200 frames have been running, starting from when the system loses the ID.

3. RESULTS AND DISCUSSION

Lost centroid ID is when an object loses detection and simultaneously loses its centroid ID. ID numbering is used to identify each detected and counted object [9]. Therefore, if an object is in the camera's range for a long time and the lost ID occurs repeatedly [11][1], it can cause low counting accuracy [12] and interfere with forecasting results. The lost centroid handling method is demonstrated by sensing 100 video files and presenting screenshots of the detection and counting results. The data used in this demonstration comes from CCTV footage of Bandung Train Station. The description of the dataset and the things involved in the data collection process are in **Table 1**.

Table 1. Data Understanding

No	*Aspects*		*Descriptions*
1.	Describe dataset	Number of video files	100 video files
		Duration per file	± 1 Hour
		Video date	27 – 31 October 2020
		Video time	13.00 – 17.00 WIB
		Channels	5 Channels / Location
		Video format	. Dav
2.	Counting technique	Up and down	Channel 3 and Channel 5
		Stay/All object	Channel 2, Channel 4, and Channel 6
3.	Data collection	Dataset validation	CCTV Stasiun KA Bandung
		Data model	Pretrained Yolo
		Storage duration	Per 5 Seconds

This process uses trial and error to find the optimal pause time to maintain the centroid ID. The pause time is adjusted to each location based on the length of visitor interaction in a room. Locations oriented as paths (fast-moving objects) are given a maxdissappeared value of 40 frames, and locations that function as waiting rooms or services (slow-moving objects) of 200 frames. In addition, determining the pause time also considers the processing speed (fps) when the detection and counting process takes place. The list of test values to find the

DOI: https://doi.org/10.17509/coelite.v4i2.85752
p- ISSN 2829-4157 e- ISSN 2829-4149

most optimal maxdissappeared value running on a system with a speed of 28 fps is shown in **Table 2**.

Table 2. Maxdissappeared Testing Value

Locations	***30***	***40***	***70***	***100***	***150***	***200***	***240***
Entrance Waiting Room							
South Gate Route							
Exit Waiting Room							
South Entrance							
North Entrance							

Based on the maxdisappeared value test in **Table 2.** The location of the north entrance waiting room, the north exit waiting room, and the north entrance are slow-moving object locations tested with maxdisappeared values of 70, 100, 150, 200, and 240. This location was not tested with values 30 and 40 because the resulting pause time is too small, which is 1.43 seconds, so there is still the potential for the same object to be counted repeatedly. The pause time is obtained by dividing the maxdisappeared value by speed.

Based on the test results at the slow-moving location adjusted to the speed and movement of the number of passengers, the most optimal maxdisappeared value is 200 frames, or equivalent to 7.14 seconds. The location of the south door path and the south entrance are fast-moving object locations tested with maxdisappeared values of 30 and 40. This location was not tested with maxdisappeared values of 70, 100, 150, 200, and 240 because this location only functions as a path, so that visitors are not long in the CCTV range. Based on the test results, the most optimal maxdisappeared was chosen at 40 frames or 1.43 seconds.

The test results prove that handling lost centroids has been proven to minimize counting the same objects repeatedly, especially in slow-moving locations such as the north entrance waiting room, the north exit waiting room, and the north entrance. The lost centroid ID handling method was evaluated at two location points: the north entrance waiting room, which represents the slow-moving location, and the south door path, which represents the fast-moving location. Ground-truth evaluation was carried out by comparing the number of manual calculations with 51 system calculations.

The number of visitors who interacted at the north waiting room location for 5 minutes based on human vision observations was 29 people. In **Figure 2,** shown results of computer vision counting before the lost centroid modification counted 68 people with an accuracy of 42%. After the lost centroid modification, computer vision managed to count 40 people with an accuracy of 72%. Thus, modifying the lost centroid ID at the slow-moving location (north waiting room entrance) increased the counting accuracy by 30.5%. The increase in counting accuracy also occurred at the fast-moving location (south entrance lane). Based on observations, the number of visitors who interacted at the south entrance lane location for 5 minutes was 99 people. Before the lost centroid modification, 120 people were counted with an accuracy of 83%. After the lost centroid modification, it became 105 people with an

DOI: https://doi.org/10.17509/coelite.v4i2.85752
p- ISSN 2829-4157 e- ISSN 2829-4149

accuracy of 94%. Thus, modifying the lost centroid at the south entrance lane increased the counting accuracy by 11%.

Figure 2. Visual test results in the waiting room north of the exit

4. CONCLUSION

The proposed lost centroid ID method, which applies a delay time of up to 200 frames at the location of slow-moving objects, has proven effective in simultaneously minimizing repeated calculations of the same object. Based on the experimental results of slow-moving and fast-moving objects, accuracy is significantly increased by 30.5% in the waiting room area and by 11-11% on routes with heavy traffic levels. This approach can increase the reliability of people counting on CCTV Analytics, especially for forecasting and strategic decision-making in public spaces.

5. ACKNOWLEDGMENT

The author would like to thank the Indonesian University of Education as the institution that has supported this research and would like to thank the Bandung Institute of Technology for being willing to facilitate the development of this research and provide support as well as providing the technical resources needed in the implementation and evaluation of the proposed method.

6. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. Authors confirmed that the paper was free of plagiarism.

7. REFERENCES

[1] Narayanan, S. J., Perumal, B., Saman, S., & Singh, A. P. (2020). Deep Learning For Person Re-Identification In Surveillance Videos. Studies in Computational Intelligence, 865, 263 – 297. https://doi.org/10.1007/978-3-030-31760-7_9

[2] Liu, Z., Chen, Y., Chen, B., Zhu, L., Wu, D., & Shen, G. (2019). Crowd Counting Method Based on Convolutional Neural Network With Global Density Feature. IEEE Access, 7, 88789–88798. https://doi.org/10.1109/ACCESS.2019.2926881

[3] Bisht, A., Mishra, U., & Saraf, S. (2024). People Analytics Using Deep Learning. 2nd IEEE International Conference on Advances in Information Technology, ICAIT 2024 - Proceedings. https://doi.org/10.1109/ICAIT61638.2024.10690501

[4] Lu, W., Wang, G., Wang, H., & Vondrick, C. (2020). Deep Learning for Multi-object Tracking: A survey. Pattern Recognition, 106, 107420. https://doi.org/10.1016/j.patrec.2020.107420

[5] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 779-788.

[6] Chua, D. E. P., Recto, K. H. A., & Mayuga, G. P. T. (2023). Real-Time Human Detection and Tracking System: A Novel Comparative Study of Centroid Tracking, Single Shot Detection and YOLO Algorithms. 2023 1st International Conference on Advanced Engineering and Technologies, ICONNIC 2023 - Proceeding, 97–102. https://doi.org/10.1109/ICONNIC59854.2023.10467636

[7] Barthélemy, J., Verstaevel, N., Forehead, H., & Perez, P. (2019). Edge-Computing Video Analytics For Real-Time Traffic Monitoring In A Smart City. Sensors (Switzerland), 19(9), 16–17. https://doi.org/10.3390/s19092048

[8] Shakil A., S., & Kureshi, D. A. K. (2020). Object Detection and Tracking using YOLO v3 Framework for Increased Resolution Video. International Journal of Innovative Technology and Exploring Engineering, 9(6), 118–125. https://doi.org/10.35940/ijitee.e3038.049620

[9] Li, Y., Qu, L., Cai, G., Cheng, G., Qian, L., Dou, Y., Yao, F., & Wang, S. (2023). Video Object Counting with Scene-Aware Multi-Object Tracking. Journal of Database Management, 34(3). https://doi.org/10.4018/JDM.321553

[10] Wojke, N., Bewley, A., & Paulus, D. (2017). Simple Online And Realtime Tracking With A Deep Association Metric. 2017 IEEE International Conference On Image Processing (ICIP), 3645-3649.

[11] Meng, C., Zhang, Y., & Hou, Q. (2023). Person Re-identification Based on Centroids Triplet Loss. ACM International Conference Proceeding Series, 306–311. https://doi.org/10.1145/3641584.3641630

[12] Putra, M. P. K., & Wahyono. (2021). A Novel Method for Handling Partial Occlusion on Person Re-identification using Partial Siamese Network. International Journal of Advanced Computer Science and Applications, 12(7), 313 – 321. https://doi.org/10.14569/IJACSA.2021.0120735

[13] Zhan, B., Wang, D., Li, J., Wei, X., & Zhang, X. (2022). A Survey on Deep Learning-based Crowd Counting. IEEE Transactions on Artificial Intelligence, 3(6), 920-938.

[14] Ullah, R., Hayat, H., Siddiqui, A. A., Siddiqui, U. A., Khan, J., Ullah, F., Hassan, S., Hasan, L., Albattah, W., Islam, M., & Karami, G. M. (2022). A Real-Time Framework for Human Face Detection and Recognition in CCTV Images. Mathematical Problems in Engineering, 2022. https://doi.org/10.1155/2022/3276704

[15] Nguyen, K.-D., Nguyen, H. H., Le, T.-N., Yamagishi, J., & Echizen, I. (2024). Analysis of Fine-Grained Counting Methods for Masked Face Counting: A Comparative Study. IEEE Access, 12, 27426–27443. https://doi.org/10.1109/ACCESS.2024.3367593

[16] Kouyoumdjieva, S. T., Danielis, P., & Karlsson, G. (2020). Survey of Non-Image-Based

Approaches for Counting People. IEEE Communications Surveys and Tutorials, 22(2), 1305–1336. https://doi.org/10.1109/COMST.2019.2902824

[17] Violatto, G., & Pandharipande, A. (2020). Anomaly Classification in People Counting and Occupancy Sensor Systems. IEEE Sensors Journal, 20(12), 6573–6581. https://doi.org/10.1109/JSEN.2020.2976547

[18] Zhang, H., Zhou, M., Sun, H., Zhao, G., Qi, J., Wang, J., & Esmaiel, H. (2021). Que-Fi: A Wi-Fi Deep-Learning-Based Queuing People Counting. IEEE Systems Journal, 15(2), 2926–2937. https://doi.org/10.1109/JSYST.2020.2994062

[19] Ibrahim, O. T., Gomaa, W., & Youssef, M. (2019). CrossCount: A Deep Learning System for Device-Free Human Counting Using WiFi. IEEE Sensors Journal, 19(21), 9921–9928. https://doi.org/10.1109/JSEN.2019.2928502

[20] Choi, H., Fujimoto, M., Matsui, T., Misaki, S., & Yasumoto, K. (2022). Wi-CaL: WiFi Sensing and Machine Learning Based Device-Free Crowd Counting and Localization. IEEE Access, 10, 24395–24410. https://doi.org/10.1109/ACCESS.2022.3155812

DOI: https://doi.org/10.17509/coelite.v4i2.85752
p- ISSN 2829-4157 e- ISSN 2829-4149