Vehicle Tracking System in Drone Imagery with YOLOv5 and Histogram

Jehwan Choi, Seongbo Ha, Youlkyeong Lee, Kanghyun Jo
Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Ulsan, Korea
jhchoi@islab.ulsan.ac.kr sbha@islab.ulsan.ac.kr yklee@islab.ulsan.ac.kr acejo@ulsan.ac.kr

***Abstract*—In this study, we introduce a vehicle tracking system for drone imagery, utilizing the real-time object detection network YOLOv5 to get vehicle location and cropped images. The system analyzes the segmented regions' histograms, compares them with previous frames, and identifies identical objects for tracking. The algorithm is designed to compare objects within a specific radius using coordinate information, enhancing histogram comparison efficiency. The MOTA (Multi-Object Tracking Accuracy) showed 90%, but the limited environment of data usage and experiments must be considered. The findings suggest that the real-time performance of the vehicle tracking system can be applied in various fields such as traffic control, vehicle management, and accident response.**

***Index Terms*—Vehicle tracking system, Intelligent surveillance system, Histogram**

I. Introduction

Urban areas often suffer from severe traffic congestion. When traffic congestion is severe, many citizens feel inconvenienced, and the probability of traffic accidents increases. One solution to alleviate traffic congestion is the introduction of a Vehicle Tracking System (VTS). VTS is a system that applies Multi-Object Tracking (MOT) specifically to vehicles. There are two main reasons why VTS is efficient in urban areas. First, by operating a vehicle tracking system in real-time through cameras, it is advantageous in understanding traffic flow. Utilizing the gathered information, traffic flow can be dispersed smoothly, and accidents can be prevented. Second, it is advantageous in identifying the cause of an accident when it occurs and is effective in tracking criminal vehicles. Additionally, we apply VTS to drone footage. Utilizing VTS with drones is more versatile than using Closed-circuit Television (CCTV). There are three reasons why vehicle tracking must be done using drones. First, drones have a higher altitude and a wider field of view than CCTV, allowing for the analysis of a broad area at once. Second, unlike CCTV, which is typically fixed to buildings, drones can move freely, making them effective. Third, their free mobility allows for rapid response to specific events.

For the implementation of the system described above, we use the One-stage detection model YOLOv5 [1] to obtain object detection results and track multiple vehicles using video sequences and histograms. YOLOv5 possesses both fast speed and high accuracy. By using the past five frames and comparing the vehicles and histograms around each object's location, we determine the presence of the same object and implement VTS by assigning unique IDs.

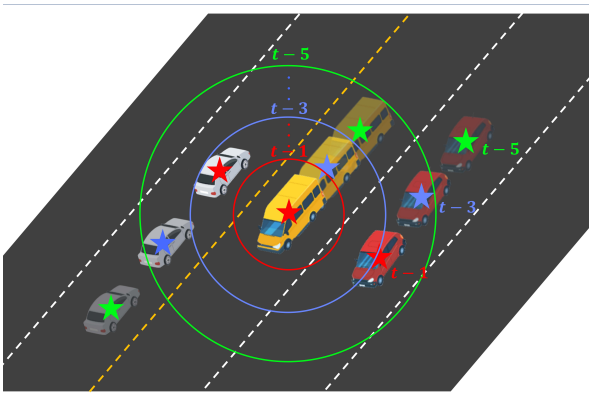

Fig. 1: Schematic Image of Fixed Radius Exploration Process for Utilizing Past Frames in Histogram Comparison Candidate Search. Red color represents $frame_{t-1}$, blue color represents $frame_{t-3}$, yellow color represents $frame_{t-5}$.

II. Proposed Method

A. Object Matching

For assigning a unique ID for vehicle tracking, the same object is identified in consecutive frames. To increase accuracy, the method described in Figure 1 is used. Where t and $t - n$ are the current point in time and n points in time before the present, respectively. The histogram of the currently detected object is compared with the vehicles detected in similar locations in the previous 5 frames to determine whether they are the same object. The vehicle's movement speed was detected to specify the range, and an appropriate radius range is designated.

B. Overall Framework

Object tracking is mainly used in videos. In this paper, the VTS network also takes video as input. The overall framework is illustrated in Fig. 2. The input video is connected to the object detection network, YOLOv5. When a video is input, the YOLOv5 network splits it into 30 FPS (Frame Per Second) for processing. The object detection network outputs a vector $\vec{V}$=(x_1, y_1, x_2, y_2, $confidence$, $category$) representing the detected object's coordinates (x_1, y_1, x_2, y_2), accuracy, and category. If 10 objects are detected in a single frame, 10 vectors are generated. However, since the order of object prediction and the number of detected objects are

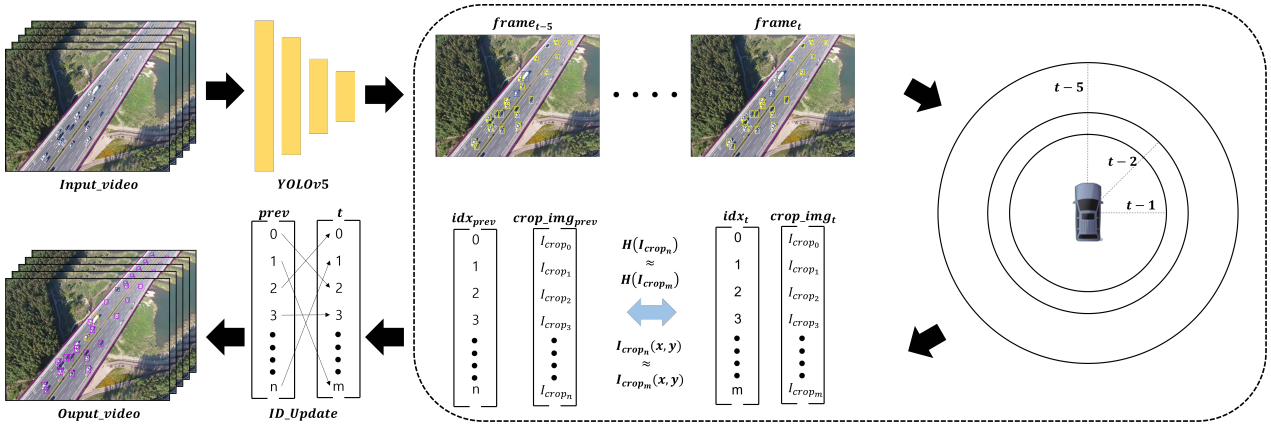

Fig. 2: Overall framework of VTS using detection results and histogram of cropped images

not constant, the results are compared between consecutive frames. In this paper, we maintain the detection results of the most recent 5 frames. To minimize calculations, we consider the vehicle's speed and only add vehicles within a certain range in the recent 5 frames to the histogram calculation list. As the probability of a wider movement range increases for older frames, we gradually expand the search range. Vehicles captured within the range are determined to be the same object or not through HSV histogram calculations. Once the histogram and coordinate comparison process is finished, ID updates occur. ID updates are applied to all objects detected in each frame, and when the input image displays the object detection results and ID, the final result video is completed.

III. Experiment

A. Dataset

The data used for the experiment are videos from the autonomous flight drone dataset [2] built by the University of Ulsan in 2020. A total of four videos were used for the experiment, and information about altitude and angle can be found in Table 1.

TABLE I: The information of drone data.

Region	Altitude(m)	Angle(°)	Time(s)
Ulsan_Samhogyo	90	60	120
	50	50	
Ulsan_Taehwagyo	60	45	
	40	30	

B. Evaluation Metric

To evaluate the performance of a VTS, we utilize evaluation metrics such as multiple object tracking accuracy (MOTA), false negatives (FN), false positives (FP), ID switches(IDs), and GT (Ground Truth). MOTA is a comprehensive metric for evaluating object tracking performance, and its formula is as follows:

$$\text{MOTA} = 1 - \frac{\text{FP} + \text{FN} + \text{IDs}}{\text{GT}} \quad (1)$$

FN occur when the system fails to detect an existing object. FP occur when the system incorrectly detects a non-existing object as existing. ID inconsistencies occur when the same ID is assigned to different objects or when different IDs are assigned to the same object.

IV. Result

In this study, the proposed Vehicle Tracking System (VTS) was tested using an autonomous drone dataset without ground truth (GT) for tracking. The results are shown in Figure 3 using data from the Ulsan_Samhogyo area's 50m 50° imagery. The accuracy was found to be higher than other models, but it's important to note that the testing was conducted in a limited situation. The selection of the radius for histogram comparison significantly impacts VTS accuracy, and the evaluation metrics are detailed in Table 2.

TABLE II: The results of proposed VTS

GT	MOTA	FP	FN	IDs
4,306	90%	1	401	28

V. Conclusion

This study proposes a vehicle tracking algorithm that can be used for traffic analysis and crime prevention in urban areas. The prominent object detection model YOLOv5 is used to detect objects, and the results are compared using histograms for each object. A unique vehicle ID is assigned in every frame, and as a result, a 90% MOTA (Multi Object Tracking Accuracy) performance was achieved.

References

[1] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, NanoCode012, Y. Kwon, K. Michael, TaoXie, J. Fang, imyhxy, Lorna, Yifu), C. Wong, A. V, D. Montes, Z. Wang, C. Fati, J. Nadar, Laughing, UnglvKitDe, V. Sonck, tkianai, yxNONG, P. Skalski, A. Hogan, D. Nair, M. Strobel, and M. Jain, "ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022. [Online]. Available: https://doi.org/10.5281/zenodo.7347926

[2] K. Jo. (2020) Autonomous drone dataset. [Online]. Available: https://aihub.or.kr/aihubdata/data/view.do?currMenu=115topMenu=100