applied sciences

MDPI

Article

Autonomous Driving Assistance with Dynamic Objects Using Traffic Surveillance Cameras

Kuk Cho [1] and Dooyong Cho [2,*]

[1] Korea Land and Geospatial Informatix Corporation (LXSIRI), Wanju-Gun 55365, Korea; kcho@lx.or.kr
[2] Department of Convergence System Engineering, Chungnam National University, Daejoen 34134, Korea
* Correspondence: dooyongcho@cnu.ac.kr; Tel.: +82-42-821-5693

Abstract: This paper describes a method that precisely estimates the position of images of traffic surveillance camera objects. We suggest a projection method with multiple traffic surveillance cameras through a local coordinate system into a global coordinate system. The transformation of coordinates uses detected objects, parameters of the camera and the geometric information of high- definition (HD) maps. Traffic surveillance cameras that pursue traffic safety and convenience use various sensors to generate traffic information. We suggest a transformation method with images of the camera and HD maps and an evaluation method. Therefore, it is necessary to improve the sensor-related technology to increase the efficiency and reliability of the traffic information. Recently, the role of the camera in collecting video information has become more important due to advances in artificial intelligence (AI) technology. The objects projected from the traffic surveillance camera domain to the HD domain are helpful to identify imperceptible zones, such as blind spots, on roads for autonomous driving assistance. In this study, we proposed to identify and track dynamic objects (vehicles, pedestrian, etc.) with traffic surveillance cameras, and to analyze and provide information about them in various environments. To this end, we conducted the identification of dynamic objects using the Yolov4 and DeepSort algorithms, established real-time multi-user support servers based on Kafka, defined transformation matrices between images and spatial coordinate systems, and implemented map-based dynamic object visualization. In addition, a positional consistency evaluation was performed to confirm its usefulness. Through the proposed scheme, we confirmed that multiple traffic surveillance cameras can serve as important sensors to provide relevant information by analyzing road conditions in real-time in terms of road infrastructure beyond a simple monitoring role.

Keywords: autonomous driving; road infrastructure; traffic surveillance camera; camera calibration

check for updates

Citation: Cho, K.; Cho, D. Autonomous Driving Assistance with Dynamic Objects Using Traffic Surveillance Cameras. *Appl. Sci.* **2022**, *12*, 6247. https://doi.org/10.3390/app12126247

Academic Editors: Wen-Hsiang Hsieh, Jia-Shing Sheu and Minvydas Ragulskis

Received: 15 April 2022
Accepted: 13 June 2022
Published: 20 June 2022

1. Introduction

The commercialization of autonomous vehicles is expected to have a positive impact in terms of mobility and convenience. Recently, a considerable amount of research and development is being conducted by automobile manufacturers and the IT industry focusing on control algorithms and related technologies using vehicle sensor information. However, there is a limitation in recognizing the surrounding environment of a vehicle with only the vehicle's sensors. It is hard to guarantee safe driving. In particular, there are imperceptible zones such as buildings, sensor blind spots for cars, and bad weather.

Therefore, an efficient and useful road infrastructure is helpful for safe driving [1–3]. The autonomous driving infrastructure provides information on weather, construction, surrounding vehicles, and pedestrians to autonomous vehicles through communication. It also supports the driving of autonomous vehicles, such as precise positioning. Road infrastructure can be classified into road facilities, roadside sensors, traffic sensors, and communications.

Road facilities refer to facilities applied to roads to improve the recognition performance of a car driving and reduce the risk of accidents. The car driving is included whether it is performed autonomously or not. A roadside sensor refers to a technology that detects

objects and the environment around the road through various sensors. A transportation center refers to a technology that comprehensively manages and analyzes data collected from vehicles, road facilities, roadside sensors, etc. Communication technology that transmits and receives data between vehicles and vehicle infrastructure that is collected and analyzed for autonomous driving is also an important technology belonging to road infrastructure. In addition, road infrastructure includes research on planning/strategy for construction.

Intelligent Transport Systems (ITS) have an information chain, including data acquisition, data processing, information distribution, and so on [1,2]. It aims to provide continuous communications between a vehicle side and an infrastructure side [2]. For example, this has the advantage that it is possible to respond quickly to driving-oriented changes in traffic conditions such as accidents and construction [3]. Generally, a digital map is an ITS enabling technology. A high-definition map (HD map) is a very important advanced form of ITS technology for autonomous driving because it holds a large amount of road information, as shown in Table 1. It consists of a kind of node-link (A and C categories) and additional road information (B category) [4,5]. This HD map technology uses information from the infrastructure, surrounding vehicles, and dynamic objects (vehicles, pedestrians, etc.) and can support reliable autonomous driving. R&D and pilot projects using the technology are being actively promoted not only overseas [6–8] but also in Korea [9].

Table 1. Data model of high-definition map.

Category	Table Name	Type	Type
A	Lane marking	Line	Left and right lanes, lane color, centerline, bus-only lane
	Road facilities	Line	Signal stop line, cross walk stop line
	Centerline of lane	Line	Highway, national highway, local road, city road, country road
B	Sign facilities	Point	Attention, regulation, indication subsidiary sign
	Signal	Point	Signal
	Road direction	Point	Road marking, direction marking
	Guide line	Line	Guide line for left turn
	Crosswalk	Polygon	No stopping zone, guide zone, uphill road, crosswalk
	No-autonomous driving zone	Polygon	Protection zones (children, disabled elderly, and others)
	Roadside facilities	Point	Roadside facilities such as streetlamps and road signs
C	Node	Point	Node

A large number of traffic surveillance cameras are built and operated for certain regions with a main control center. Using the cameras, moving objects such as vehicles and pedestrians can be recognized and location information can be derived from a map. It can be used in a variety of approaches for traffic surveillance camera services such as the detection of unexpected blind spots and pedestrian care [10,11]. In this regard, research on object recognition and image analysis based on deep learning for road traffic information analysis is steadily progressing. Until now, studies have mainly been conducted to improve the recognition rate of a specific object, such as a vehicle or a person, or to analyze a motion. Research for utilization and application such as visualization in a new form through the transformation of the coordinate system of the recognized object is still insufficient. The public cloud uses the Google Cloud platform for large-scale data processing to recognize vehicles from images or to estimate the traffic volume in an area by estimating the location information of the vehicle, which is recognized by a stereo traffic surveillance camera [12].

Related information by tracing a vehicle with a traffic surveillance camera image and visualizing an event information map by linking the tracking results with Google Map through a transformation relationship is provided [13]. In addition, a method of object visualization from a top view was proposed with object tracking and a coordinate transformation in a specific area through the transformation relationship between six cameras employing the stereo vision method [14]. A fusion method using an image sensor, GPS, and LiDAR was proposed to estimate the location of a target object by tracking the location of the object [15]. In the case of related existing studies, various approaches for the use of traffic surveillance cameras have been presented from a monitoring point of view [16]. However, since the purpose of providing information is focused only on semi-real-time image analysis, it has disadvantages such as poor usability other than monitoring or requiring multiple sensors. There are only a few approaches that can be used to match coordinates between 2D images and 3D real roads because they are difficult to measure.

In this study, we suggest a transformation method with multiple traffic surveillance camera images and HD maps and an evaluation method as a way to increase the utilization of established traffic surveillance cameras. The main contribution of this paper can be summarised by the following points. First, we proposed a fusion method to evaluate accurate vehicle positions on the road. Additionally, a high-precision localization approach was used with a combination of traffic surveillance camera equipment and HD maps. Secondly, we proposed a simple method to unify heterogeneous coordinates. Furthermore, moving objects (vehicles, people, etc.) were identified and tracked in traffic surveillance camera images and information related to their location was provided as spatial information on the map. We proposed a method that can be used in various environments. To this end, the moving object location information of the image was converted to the left position on the map, and the object was visualized as spatial information using precision, aerial, and general maps through a web-based visualization module.

This paper is organized as follows. The identification and tracking of dynamics objects are introduced in Section 2. In Section 3, an overview of the proposed localization method and a detailed introduction of the basic theoretical method are provided. In Section 4, the factors that affect the accuracy of estimation based on a real environment are analyzed. In Section 5, an evaluation and discussion are presented, along with the conclusion and indications of future research efforts.

2. Identification of Dynamics Objects

Recently, with the development of artificial intelligence (AI) technology, technologies that can identify and track target objects based on images in various devices have been developed. For this reason, the traffic surveillance camera, an image sensor incorporating AI technology, can be used for a variety of purposes in many fields. This goes beyond a simple monitoring role that can be checked with the naked eye, and can serve as an important sensor that can provide relevant information by analyzing road conditions in real time in terms of road infrastructure. In addition, for the scalability of monitoring system connection, which is the purpose of traffic surveillance camera utilization, related functions were implemented along with modular-based systems, and Figure 1 shows multiple traffic surveillance cameras in a certain area. It is helpful in understanding traffic flow on a road as well as for obtaining information on blind spots at an intersection.

Traffic surveillance camera images are used as multiple inputs, dynamic objects (vehicles, people, bicycles, etc.) included in the image frame are identified using AI, and the location is tracked by creating a minimum bounding box of the object. For the identified and tracked object, the position defined in the image coordinate system is transformed into the map coordinate system through a transformation matrix in which the coordinate system transformation relationship between the camera coordinate system and the map is established and also configured for multi-user support. It performs the process of uploading object-related information to the server in real time, and uses the JSON format for the standardization of information delivery [17]. The object information delivered in this way

is visualized using precision, aerial, and general maps through a web-based visualization module designed to support various devices, as shown as Figure 2.

Figure 1. Example images of multiple traffic surveillance cameras.

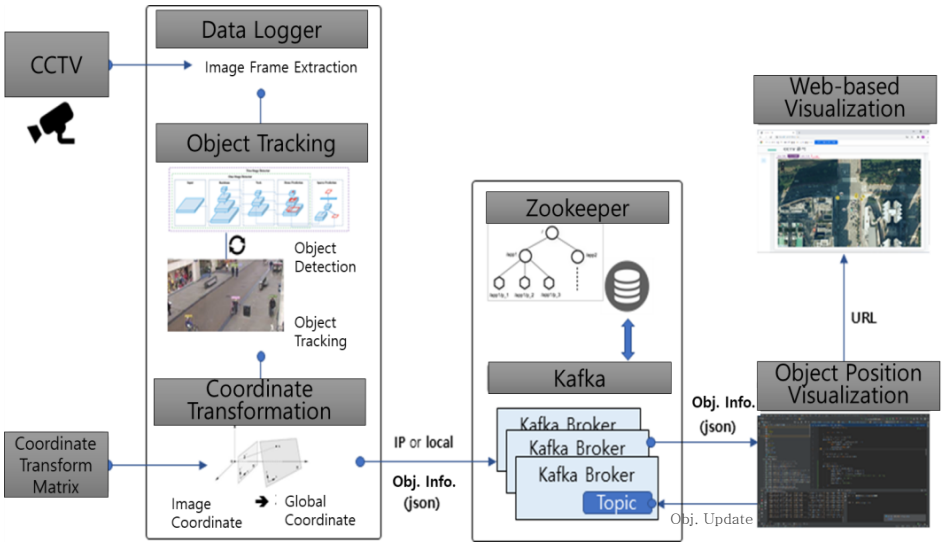


Figure 2. Overview of the implemented system.

In order to accurately identify and track the target object in an image, information such as the shape and color of the object can be used. Therefore, the process of recognizing one object in one image requires a large amount of effort and time before AI technology can support high-level functions. However, with the advent of artificial intelligence technology, a technology for recognizing multiple objects in one image at a near real-time speed is in development. As these technologies are used in various fields, they are gradually becoming more advanced.

Image processing technology, including object detection, has rapidly improved recently [18]. It is utilized in autonomous driving with various algorithms such as Faster Region-based Convolutional Neural Networks (R-CNN) [19], You Only Look Once (YOLO) [20], and Single Shot Detector (SSD) [21]. Faster R-CNN is an improved model with the aim to achieve a real-time processing speed using Fast R-CNN from the initial R-CNN. Although the structure is similar to that of Fast R-CNN, it uses a regression network called a very small region proposal network for selective search, showing a processing speed performance of 250 times that of R-CNN and 25 times that of Fast R-CNN. YOLO is a method that tracks objects by dividing them into grid units corresponding to n boxes rather than pixel units of an image, and is a model suitable for real-time detection systems with a near real-time processing performance. SSD is a model that has a balanced accuracy and processing speed by tracking many target objects in one image based on feature maps of various sizes.

Therefore, we adopted the YOLO algorithm, which is known for its fast object identification time. The YOLO model for training used the YOLOv4 architecture. In addition, the recently announced YOLOv4 shows a considerable improvement in performance [22,23]. In order to use such a model, learning of the target object to be identified and tracked should be performed. A total of five types of target objects for learning were defined, including four types of objects defined as vehicles, people, two-wheeled vehicles (including bicycles), strollers, and others corresponding to moving objects on and off the road. It was constructed using the COCO data set [24], which has various learning sets, and the KODAS data set [25] from LX (Korea Land Information Corporation), which provides domestic autonomous driving data.

The process of continuously detecting the target object being identified must also be carried out. Therefore, the DeepSort algorithm [26], which predicts the actual value by applying past information to the present, is attached to the back of YOLOv4 so that a specific object can be tracked without missing the specific object in a continuous image. This is a method of matching the same object by storing the characteristic information of the object identified in the previous frame and applying it to the subsequent frame. First, a loop that recognizes the type and location of an object in a frame based on YOLOv4 and applies the results to DeepSort to track a specific object in successive frames is repeatedly performed.

3. Problem Formulation

The traffic surveillance camera image used as input information in this study has a high resolution in the horizontal (u) and vertical (v) directions to detect long distance objects, whereas in visualization precision, aerial and general maps are used. Therefore, in order to map the exact position of an object tracked in an input image, a transformation matrix capable of explaining the differences (translate, rotation, and scale) between coordinate systems is required.

We assumed that our camera follows the general pinhole camera model [27]. In this model, the perspective projection of a 3D point into the 2D image point can be represented as follows:

$$\lambda\ P_c = [K|C][R|T]P_w \tag{1}$$

where,

$$[K|C] = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$[R|T] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}$$

$$R = R_z(\phi)\ R_x(\theta) = \begin{bmatrix} cos\phi & -cos\theta sin\phi & sin\theta sin\phi \\ sin\phi & cos\theta cos\phi & -sin\theta cos\phi \\ 0 & sin\theta & cos\phi \end{bmatrix}$$

$[K|C]$ is the matrix of the camera's intrinsic parameters, i.e., focal length, skew, and the optical center of the camera. $[R|T]$ is the matrix of the camera's extrinsic parameters, the 3×3 rotation matrix and the 3×1 translation matrix of the camera, respectively. R is referred to as a pan-tilt camera model. The camera calibration problem is the problem of estimating $[K|C]$ and $[R|T]$. $[K|C]$ depends only on the intrinsic parameters of the camera. $[R|T]$ is the extrinsic parameters of the position of the cameras (camera translation positions $t_x, t_y,\ t_z$ and rotational angles θ and ϕ), and projective depth (λ). We assume that the traffic surveillance cameras have zero pixel skew and an equal aspect ratio [28,29] and that the image center is the same as the optical center. The road section was also assumed to be flat and straight. It is a pan-tilt camera model. We rewrote $[K|C] = diag(f,\ f,\ 1)$ and $R = R_z(\phi),\ R_x(\theta)$. Figure 3 shows the calibration procedure.

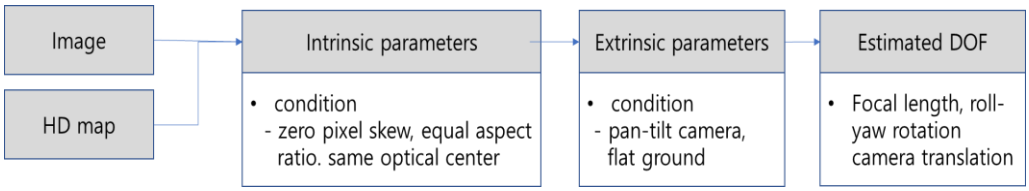


Figure 3. Data processing flowchart with estimated parameters.

For this purpose, as shown in Figures 4 and 5, invariable feature points (lanes, road markings, crosswalks, etc.) were defined in the traffic surveillance camera image of the target location for object location information visualization, and road information was measured and produced in the world coordinates with precision on the map. The map was built in high definition for automated vehicles. Corresponding points were required to match the coordinates. The corresponding points on the map with image feature points and precision were performed manually to maintain consistency. Here, when a feature point for a partial region of an image is used, the feature point defined in the image should be defined to cover as wide a range as possible because the region without the feature point may be distorted and the location of the object may be incorrectly mapped. In this study, after defining 25 invariant feature points for each image to calculate the transformation relationship between coordinate systems, the position of the object was transformed and visualized using the transformation equation.

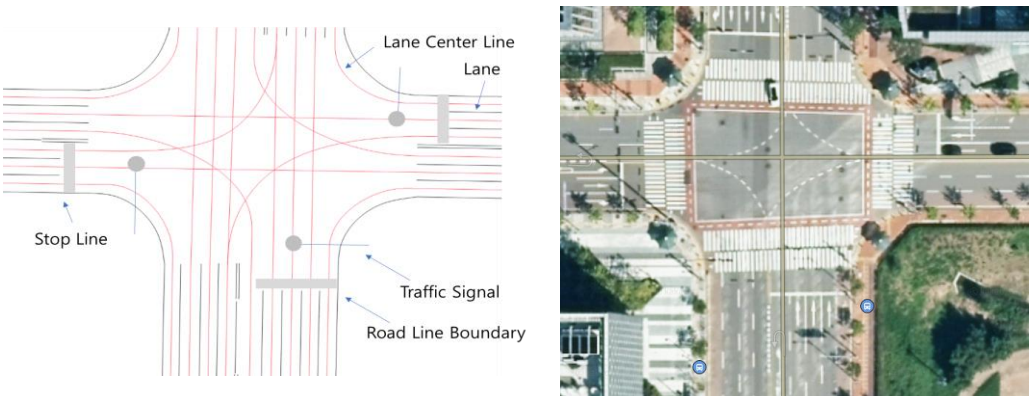


Figure 4. Road line and critical path in the HD map and aerial image.

To check the transformation relationship between the two coordinate systems with different dimensions in the traffic surveillance camera image and visualization, a transformation matrix was constructed using a point on the world coordinate system that matches a specific point on the traffic surveillance camera image together with the intrinsic and extrinsic parameters. A method of calculating and extracting a large number of invariant feature points from an image and mapping the corresponding feature points in the world coordinate system to calculate a transformation matrix such as a homography matrix

(H) in which the transformation relationship between the two coordinate systems was established [30]. In the former case, the location matching accuracy may be high, but there is a disadvantage in that it is not easy to obtain internal and external parameters through actual measurement of a traffic surveillance camera. In the latter case, the accuracy may be lower than that of the former, but it is advantageous in that the establishment of the transformation relationship is relatively simple and the positioning accuracy can be improved according to the precision of the corresponding feature point.

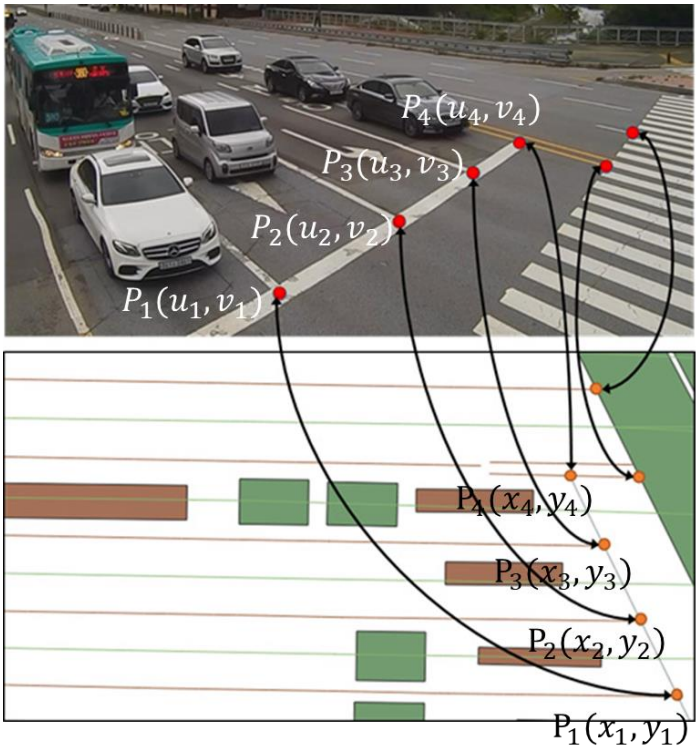


Figure 5. Corresponding points of two heterogeneous coordinate systems.

In this study, the H_{ij} calculation method was adopted and used for usefulness verification from the coordinate of the *i*-th camera to the coordinate of the *j*-th camera. H is a method mainly used for transforming the coordinate system from 2D to 2D and 2D to 3D, and this is an essential method for establishing the relationship between the marker and the camera in augmented reality (AR) technology [31]. We were able to merge both coordinates from 1 to the *n*-th the coordinate of traffic cameras. As shown in Figure 6, a 3×4 matrix was produced in which the translation and rotation between the points formed on the images of each camera were defined when looking at the same point from the cameras in two different directions. In the coordinate system transformation problem of this study, the image from 1 to *n* in Figure 7 is defined as a traffic surveillance camera. Additionally, by defining the image as a world coordinate system, the relationship of the transformation between the two coordinate systems can be evaluated. That is, the transformation matrix was calculated using the world coordinate system's corresponding points for the invariant feature points of the traffic surveillance camera image. Unfortunately, traffic surveillance cameras rarely film overlapping regions.

We used a virtual primary trajectory on the HD map, which is the center line of the lane. As a result of identifying and detecting the target object in Section 2, the object type, location, ID, etc., were output as monitoring information. These are the properties of one object, and for easy transfer and utilization, a process of tying them into one data set is required, and there are many applicable formats for this purpose. In this study, we defined the attribute information output as a result of object identification and detection by adopting the JSON format, which is used in many fields and provides fast and stable

encoding and decoding. The dataset for each object defined in this way is then delivered to the server. During visualization, the relevant information is received from the server, decoded, and then the object information is mapped to the background map.

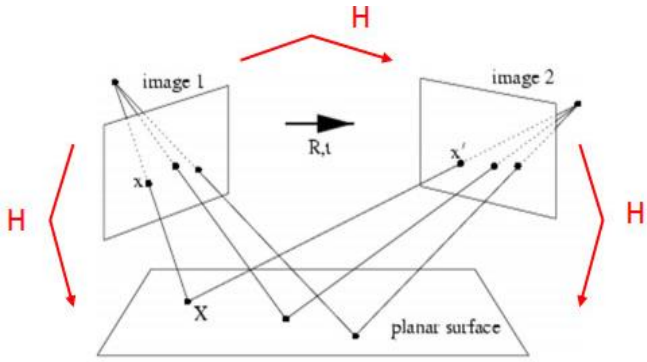


Figure 6. Definition of a homography matrix for projection (not real world).

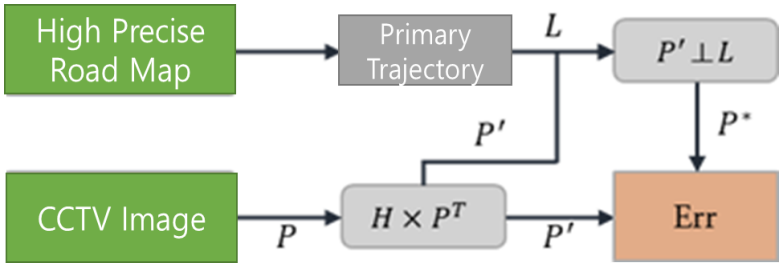


Figure 7. Process for error estimation in the global coordinate system.

4. Results Evaluation

In the proposed method, the mapping of moving objects such as vehicles and people to exact positions on the spatial coordinate system in the traffic surveillance camera image is a very important factor. Figure 8 shows the target object identification results in successive frames. For the object location detection method, a minimum square shape was adopted, and the square for each tracked object was used as location information of the object for visualization and registration. Therefore, the verification of the degree of position matching is required, and for this purpose, the coordinate system conversion accuracy was evaluated based on the process to examine the degree of coincidence between any point on the main path in the traffic surveillance camera image and the converted world conversion point. So, we proposed that it uses a virtual primary trajectory on the HD map, which is the center line of the road or the lane. The error is measured between a virtual primary trajectory from the HD map and the sample points of the estimated trajectory from the image. The virtual primary trajectory is a reference line used as the critical path.

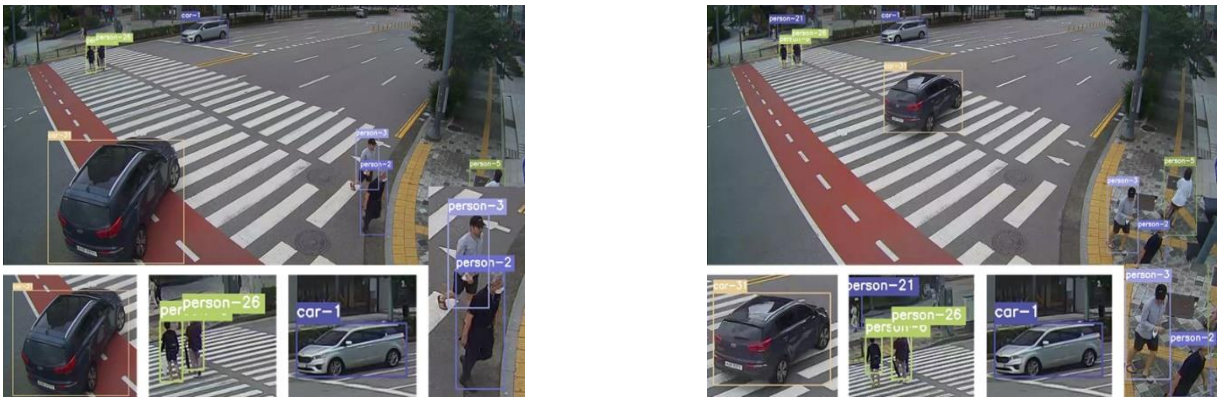
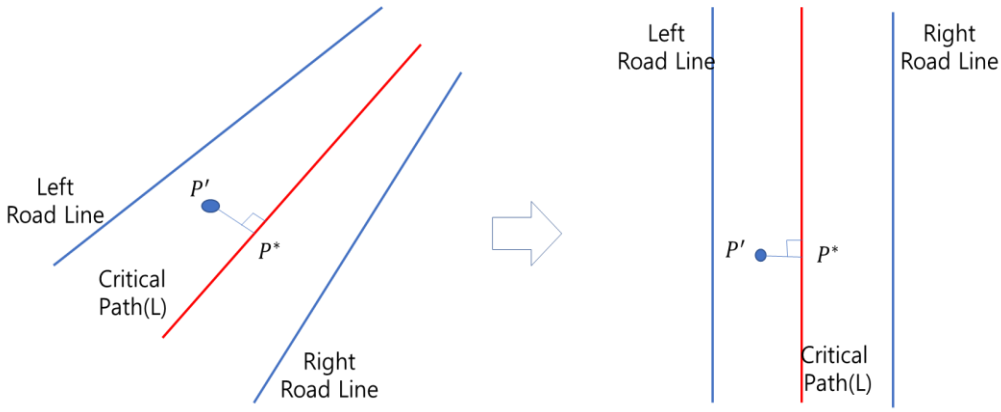


Figure 8. Results of object detection from the traffic surveillance camera.

At the beginning, in the map structure with precision built with actual measurement, the critical path (L) was obtained as shown in Figure 9. It is between lanes and guides the vehicle in the direction it should drive at the junction, branch, intersection, etc., of roads, and the traffic surveillance camera. A lane is found in the image, and n-random points are sampled by calculating the image-based main path located in the center between the lanes. For sampling, lane extraction and main route calculation were performed with an evaluation approach and an image processing library to ensure accuracy [32].



(**a**) View point of traffic surveillance camera (**b**) View point of HD map

Figure 9. Definition of error estimation for the world coordinate.

In this way, by using the coordinate system transformation method of Section 2 for the points sampled from the traffic surveillance camera image, it was converted to a point $P'_i = (x'_i,\ y'_i)$ of the spatial coordinate system of the map with precision, and the vertical intersection $P^*_i = (x^*_i,\ y^*_i)$ that is closest to and located at the correspondence position was found. The estimated error is as follows:

$$Err = \left|P' - P^*\right| \tag{2}$$

If the coordinate system transformation is performed well without error, the calculated point is located on the map's main path with precision. In other words, it can be said that the closer it is to zero, the higher the degree of position matching. Figure 9 shows the structure of the map with precision defined by lanes (black line) and a middle line (red line). Figure 9 shows random points sampled for evaluation in the main path of the traffic surveillance camera image corresponding to the map with precision. In this study, after sampling 26 random points, the degree of localization of the proposed method was evaluated with this process. As shown Figure 10, we used virtual points to verify the proposed method from the HD map.

The results of comparing the transformed and vertical intersections, and the distance between the two points are provided in Appendix A. As a result of the distance comparison, errors occurred from as small as 0.02 m to as large as 0.48 m, and the average (Avg) and root mean square (RMS) showed errors of 0.15 m and 0.19 m, respectively. Points sampled from a region close to the traffic surveillance camera showed relatively few errors, which is why the feature points obtained in the region with low distortion close to the camera field of view were well reflected when calculating the transformation matrix for the two coordinate systems, while the region far from the camera field of view was judged to be a phenomenon caused by not accurately reflecting the characteristic points due to severe distortion.

Therefore, it is necessary to define the feature points for calculation of the transformation matrix in a direction in which more detailed information can be obtained (increase in the number of feature points as the distance from the traffic surveillance camera increases, etc.) in a region with severe distortion. This way, a higher degree of position matching can be expected.

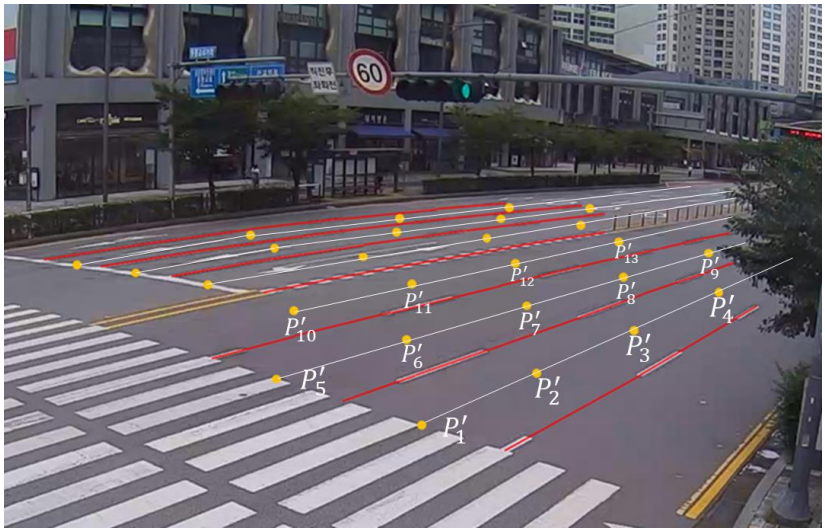


Figure 10. Sample points of the traffic surveillance camera.

4.1. Object Recognition of the Traffic Surveillance Camera

Table 2 shows the average precision (AP) test results according to the intersection over union (IoU) for objects with more than 30 pixels in the image. AP_{50} indicates an AP when an object with an IoU of 0.5 or more is set as a true positive, and AP_{75} indicates an AP when an IoU is 0.75. $AP_{0.5:0.05:0.95}$ represents the average of AP results for up to 0.95 while increasing the IoU by 0.5 to 0.05 in our proposed method as shown in Figure 11. In this test, vehicle objects and people were considered based on their frequency of appearance by class, but later, the performance evaluation was performed using traffic surveillance camera images of various regions with various classes.

Table 2. Results of detection using our model.

Class	$AP_{0.5:0.05:0.95}$	AP_{50}	AP_{75}
car	0.54	0.66	0.56
person	0.72	0.93	0.78
mAP	0.63	0.79	0.67

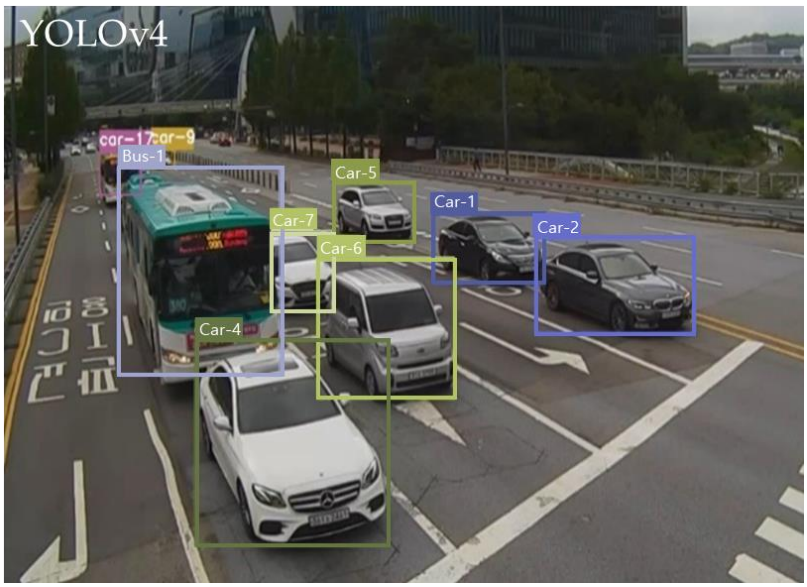


Figure 11. Results of detection using our model.

As hyperparameters for training the YOLOv4 model, the training step was set to 15,000 times, the batch size was set to 64, and the learning rate was set to 0.1. Additionally, momentum and weight changes were defined as 0.9 and 0.0005, respectively. To verify the performance of the model designed in this study, a test was performed using traffic surveillance camera image event data of the Gyeonggi Autonomous Driving Center from a public data portal (data.go.kr, accessed on 24 January 2022). Open traffic surveillance camera image data were used, and the object of the image was mainly divided into vehicle and person classes due to the nature of the traffic surveillance camera location for road-condition monitoring.

4.2. Data Synchronization and Communication

In the integrated center for autonomous driving or the road traffic control center, several traffic surveillance cameras are installed and operated in the management area to monitor road conditions. The process of analyzing multiple traffic surveillance camera information with a shooting speed of 30 frames per second (FPS) in real time and delivering it to multiple users requires a large amount of resources in terms of hardware and software, and specifically, it can be said that fast data loading and provision are necessary conditions. To this end, Kafka specialization for distributed processing environments was adopted and used for loading object information and delivering visualization information in this study. Kafka is a distributed data-streaming platform that can store and process data in real time. In other words, it is an open-source-based solution that processes data streams from multiple sources and delivers them to multiple users.

4.3. Web-Based Visualization

Recently, due to the improvement of the specifications of terminals such as smartphones and changes in information delivery methods, the visualization method for information delivery has also changed from a single software method to a web method that can be used on various platforms. In this study, a dynamic web-based visualization function was implemented based on the Spring framework [33,34], which facilitates web development. It supports a variety of services compared to general web development languages, and it provides an environment for integrating and operating existing libraries because of its excellent compatibility with database processing libraries. The visualization environment of objects in traffic surveillance camera images in this study was configured as shown in Figure 12.

Figure 12. Web-based object visualization.

The web module for visualization has a simple structure. It supports precision maps, aerial maps, and general maps as base maps, and users can use functions such as base map selection. The web module is a client corresponding to the consumer in Section 2, and provides a function to access the related topic while the Kafka server is running and to monitor the location of objects and related information. Figure 13 shows the results of applying the proposed method for one, two, and three images of traffic surveillance cameras, respectively. In terms of usability, the target object was identified and tracked based on one traffic surveillance camera image taken during the day. As a result, the performance for object visualization in a main CPU with 3.6 Ghz, memory 32 Gb RAM, and graphic card GPU 1080 Ti environments had a processing speed of about 18 FPS. The entire infrastructure system was installed in a data center in the Pangyo Zero City autonomous driving pilot city in Korea.

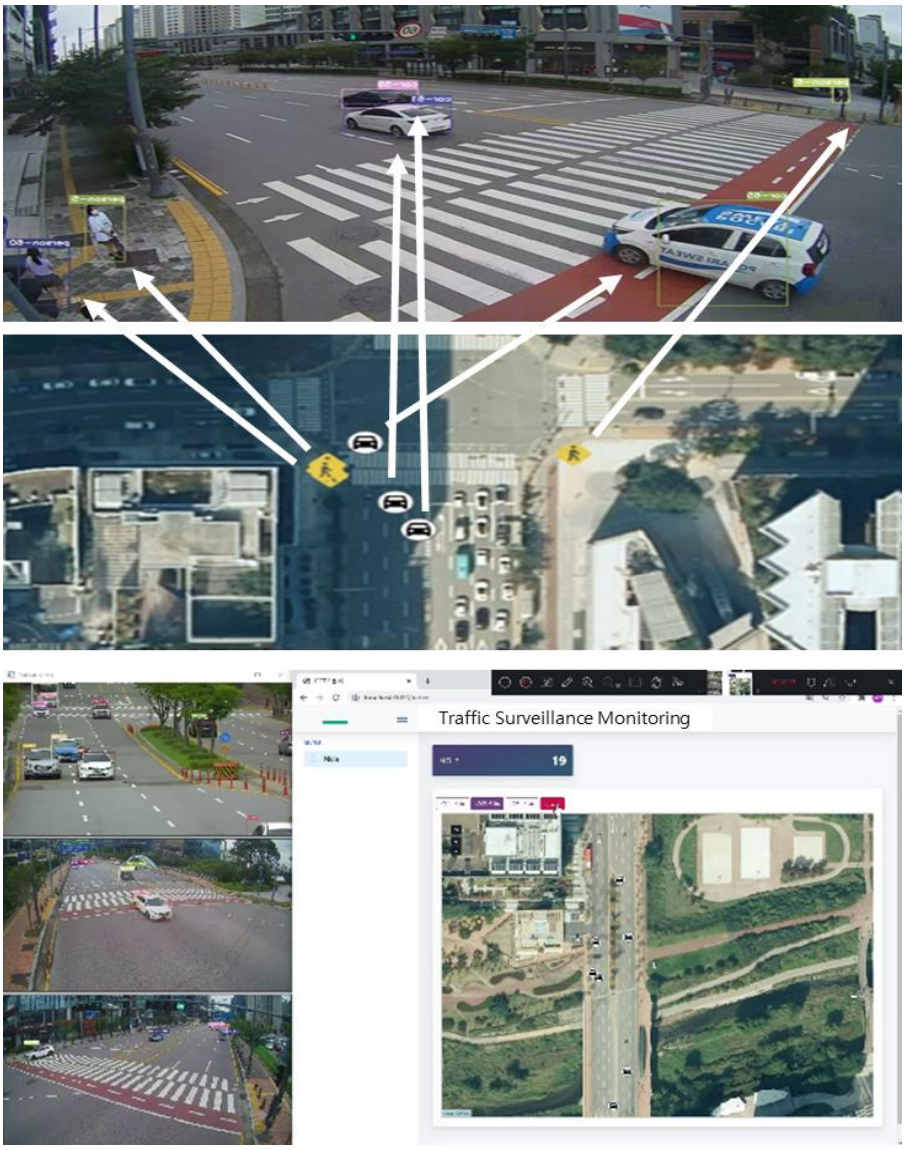


Figure 13. Results of the proposed method.

In addition, in a test conducted by increasing traffic surveillance cameras and visualization modules by up to eight units each for real-time multi-user support inspection, the processing speed from object detection to server loading and visualization was about 18 FPS, which is the performance when using one of each. It was confirmed to maintain FPS, and based on this, it was found that most resources for data processing in the image

analysis were used. Although it was not possible to apply up to dozens of traffic surveillance camera images due to the limitation of the amount of data possessed, the Kafka system constructed in this study can be used as a server system for real-time multi-user support if the proposed data delivery format and object properties are followed.

5. Discussion

Camera calibration using a reference checkerboard is a general approach used to estimate intrinsic parameters. It is a good method to extract the parameters. However, this approach is impractical in applications to traffic surveillance cameras in the real world because it requires a reference checkerboard to cover the entire monitored area. It is rarely possible to stop and restrict traffic for a camera calibration process where surveillance cameras are installed. In the field of traffic surveillance camera calibration, there is a vanishing point-based method that is available with man-made structures such as a lane, a road, a building, etc. The parameter of the lens distortion is required to estimate object positions. The vanishing point-based method sometimes assumes that an error from the lens distortion is ignored. The positions of multiple cameras are essential for fusion applications such as the object detection of vehicles and walkers in the system of urban surveillance cameras. To overcome the above issues, we proposed a camera calibration method suitable for traffic surveillance cameras. We assumed that the ground plane is the same height, and then we used only geometric information of the road where the traffic surveillance camera was installed because the HD map provides accurate geometric information. Compared with other methods, the proposed method has the advantage that the intrinsic and extrinsic parameters are directly derived without any other information. We demonstrated the performance of our method by conducting real data experiments. These experiments showed that the proposed method could successfully estimate camera parameters of the pan-tilt traffic surveillance cameras using HD map information as the reference points.

6. Conclusions

As a way to more actively utilize traffic surveillance cameras among the sensors introduced and used in road infrastructure, this paper described a method that precisely estimates the position of a traffic surveillance camera image's objects. It is a transformation method that uses two heterogeneous applications between multiple traffic surveillance cameras and HD maps. It was built in many sections, but has little use other than for simple monitoring, and it is a method used to track a target object in an image with AI technology and visualize it. The projection and detection of objects from the traffic surveillance camera domain to the HD domain is helpful to identify imperceptible zones, such as blind spots on roads for autonomous driving assistance. In the proposed method, target objects such as vehicles and people were defined for detection, and an AI model that can identify and track them, a server that provides real-time data for multi-user support, and precision, aerial, and general road maps were used as base maps. A visualization function was implemented. In addition, to verify the validity of the proposed method, an evaluation procedure was established, and the applicability and effectiveness of our method were confirmed. With the proposed method, the basis for using traffic surveillance cameras in various fields was established, and the process for this is expected to be diversified in products and research using many imaging devices. The system was installed in the data center and inter-operates the infrastructure of the Pangyo Zero City autonomous driving pilot city in Korea.

However, since it is greatly affected by the establishment of the coordinate system transformation relationship, a more accurate establishment method is required, and as a basic study for the advancement of traffic surveillance camera utilization, conditions such as occlusion between objects and the environment (weather) were not considered. An HD map provides precise and accurate reference information. This paper contributed a fusion method to accurately evaluate a vehicle's position on the road, as well as a simple method to unify heterogeneous equipment. Moving objects (vehicles, pedestrians, etc.) are

identified and tracked in traffic surveillance camera images and information related to their location is provided as spatial information on the map.

In this research, traffic surveillance camera shooting speeds were of less than 30 FPS, but considering the current GPU specifications, if a high-end GPU is used, processing speeds of 30 FPS or more can be achieved. Due to the characteristics of image analysis technology, adjusting the image size within the threshold does not have much of an effect on the result, so adjusting the image size to increase the processing speed may be another method.

In future research, we intend to secure a wider variety of types of traffic surveillance camera data in consultation with local governments to improve the visualization processing speed and location matching, and to improve object identification, tracking models and coordinate system transformation methods based on this.

Author Contributions: Conceptualization, K.C. and D.C.; Methodology, K.C. and D.C.; Experiment, K.C.; Validation, D.C.; Formal Analysis, K.C.; Investigation, K.C. and D.C.; Data Curation, K.C.; Writing—Original Draft Preparation, K.C.; Writing—Review and Editing, D.C.; Visualization, K.C.; Supervision, D.C.; Project Administration, K.C. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant titled "Industrial Technology Innovation Project—Establishing a Demonstration Infrastructure of Autonomous Cargo Transportation Service for Commercial Vehicles in Saemangeum (P001187493)" from the Korea Evaluation Institute of Industrial Technology and the National Research Foundation of Korea (NRF) [2018R1D1A3B07049698]. The authors gratefully acknowledge this support.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Results of the Evaluation of Projection Points on World Coordinate.

No.	$P^{'}$ (Transformed Point).	P^{*} (Reference Point).	Error (m)
1.	37.39810236995146, 127.11284371695545.	37.39810236995146, 127.11284458460997.	0.27
2.	37.398056022983745, 127.11284453412071.	37.398056022983745, 127.11284459860855.	0.06
3.	37.398014448299726, 127.11284392441756.	37.39801444225397, 127.1128446110664.	0.41
4.	37.3981003578181, 127.11287948619685.	37.39810034847469, 127.11288081438913.	0.48
5.	37.398053366157214, 127.11288048901872.	37.39805336615721, 127.11288084370179.	0.31
	⋮		
22.	37.39805187414329, 127.11291959905732.	37.398051865349494, 127.11291840016989.	0.36
23.	37.39799629735647, 127.11291879882195.	37.39799629149394, 127.11291837671978.	0.37
24.	37.39804721105009, 127.11296406345524.	37.39804722277515, 127.11296141945408.	0.24
25.	37.39790503296632, 127.11296219917062.	37.397905021241264, 127.11296124944073.	0.44
26.	37.39794793496272, 127.11291813053069.	37.397947934962716, 127.11291835477248.	0.02
	Avg.		0.15
	RMS.		0.19

References

1. Guerna, A.; Bitam, S.; Calafate, C.T. Roadside Unit Deployment in Internet of Vehicles Systems: A Survey. *Sensors* **2022**, *22*, 3190. [CrossRef]
2. Jarašūniene, A. Research into intelligent transport systems (ITS) technologies and efficiency. *Transport* **2007**, *22*, 61–67. [CrossRef]
3. Kiela, K.; Barzdenas, V.; Jurgo, M.; Macaitis, V.; Rafanavicius, J.; Vasjanov, A.; Kladovscikov, L.; Navickas, R. Review of V2X-IoT Standards and Frameworks for ITS Applications. *Appl. Sci.* **2020**, *10*, 4314. [CrossRef]
4. Bezzina, D.; Sayer, J. Safety Pilot Model Deployment: Test Conductor Team Report. NHTSA. Available online: http://www.nhtsa.gov/ (accessed on 24 January 2022).
5. Ham, S.; Im, J.; Kim, M.; Cho, K. Construction and Verification of a High-Precision Base Map for an Autonomous Vehicle Monitoring System. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 501. [CrossRef]
6. National Geographical Institute's Precision Map. Available online: http://map.ngii.go.kr/ms/pblictn/preciseRoadMap.do (accessed on 10 October 2019).
7. CV Pilot. Connected Vehicle Pilot Deployment Program: United States Department of Transportation(ITS). Available online: https://www.its.dot.gov/pilots/ (accessed on 24 January 2022).
8. Kotsi, A.; Mitsakis, E.; Tzanis, D. Overview of C-ITS Deployment Projects in Europe and USA. In Proceedings of the 23rd IEEE International Conference on Intelligent Transportation Systems, Rhodes, Greece, 20–23 September 2020.
9. Kim, J. *A Study on the R&D of the Operating System and Transportation Infrastructure for Road Driving of Self-Driving Cars*; The Road Traffic Authority: Wonju-si, Korea, 2018.
10. Jung, J.; Yoon, I.; Lee, S.; Paik, J. Object Detection and Tracking-Based Camera Calibration for Normalized Human Height Estimation. *J. Sens.* **2016**, *2016*, 1–9. [CrossRef]
11. Yang, I.; Jeon, W.; Lee, J.; Park, J. Development of an Integrated Traffic Object Detection Framework for Traffic Data Collection. *J. Korea Inst. Intell. Transp. Syst.* **2019**, *18*, 191–201. [CrossRef]
12. Seo, H.; Kim, E. Estimation of Traffic Volume Using Deep Learning in Stereo CCTV Image. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2020**, *38*, 269–279.
13. Mehboob, F.; Abbas, M.; Rehman, S.; Khan, S.; Jiang, R.; Bouridane, A. Glyhp-based video visualization on Google Map for surveillance in smart cities. *J. Image Video Process.* **2017**, *28*, 1–16.
14. Sankaranarayanan, A.; Veeraraghavan, A.; Chellapp, R. Object Detection, Tracking and Recognition for Multiple Smart Cameras. *Proc. IEEE* **2008**, *96*, 1606–1624. [CrossRef]
15. Kim, B. Design of Image Tracking System Using Location Determination Technology. *J. Digit. Converg.* **2016**, *14*, 43–148. [CrossRef]
16. Kumar, D.; Raut, S.; Shimasak, K.; Senoo, T.; Ishii, I. Projection-mapping-based object pointing using a high-frame-rate camera-projector system. *Robomech. J.* **2021**, *8*, 1–21. [CrossRef]
17. Schiopu, I.; Cornelis, B.; Munteanu, A. Real-Time Instance Segmentation of Traffic Videos for Embedded Devices. *Sensors* **2021**, *21*, 275. [CrossRef] [PubMed]
18. Fernandes, S.; Duseja, D.; Muthalagu, R. Application of Image Processing Techniques for Autonomous Cars. In Proceedings of the Engineering and Technology Innovation, Online, 17 December 2020; Volume 17, pp. 1–12.
19. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems(NIPS), Montreal, QC, Canada, 7–12 December 2015.
20. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
21. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A. SSD: Single Shot Multibox Detector. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; Las Vegas, NV, USA, 27–30 June 2016.
22. Seung, T.; Kwon, G.; Moon, K.; Lee, S.; Kwon, K. An Estimation Method for Location Coordinate of Object in Image Using Single Camera and GPS. *J. Korea Multimed. Soc.* **2016**, *10*, 112–121. [CrossRef]
23. Bochkovskiy, A.; Wang, C.; Liao, H. Yolov4: Optimal speed and accuracy of object detection. In Proceedings of the The International IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Seattle, WA, USA, 13–19 June 2020.
24. Lin, T.-Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.
25. Cho, K.; Im, J.; Kim, M.; Jin, Y.; Kang, S. Feasibility Assessment of KODAS through Autonomous Driving Recognition Challenge. *Korean Soc. Automot. Eng.* **2021**, *29*, 233–241. [CrossRef]
26. Wojke, N.; Bewley, A.; Paulus, D. Simple Online and Realtime Tracking with a Deep Association Metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
27. Zhang, Z. Camera calibration with one-dimensional objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 892–899. [CrossRef] [PubMed]
28. Guillou, E.; Meneveaux, D.; Maisel, E.; Bouatouch, K. Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. *Vis. Comput.* **2000**, *16*, 396–410. [CrossRef]
29. Sochor, J.; Juránek, R.; Herout, A. Traffic Surveillance Camera Calibration by 3D Model Bounding Box Alignment for Accurate Vehicle Speed Measurement. *Comput. Vis. Image Underst.* **2017**, *161*, 87–98. [CrossRef]

30. Bhardwaj, R.; Tummala, G.K.; Ramalingam, G.; Ramjee, R.; Sinha, P. Autocalib: Automatic traffic camera calibration at scale. *ACM Trans. Sens. Netw.* **2018**, *14*, 1–27. [CrossRef]
31. Prince, S.; Xu, K.; Cheok, A. Augmented reality camera tracking with homographies. In Proceedings of the IEEE Computer Graphics and Application, Tsinghua University. Beijing, China, 9–11 October 2002; Volume 22, pp. 39–45.
32. Farag, W.; Saleh, Z. Road lane-lines detection in real-time for advanced driving assistance systems. In Proceedings of the Conference on Innovation and Intelligence for Informatics, Computing, and Technologies(3ICT), Sakhier, Bahrain, 18–20 November 2018.
33. Arthur, J.; Azadegan, S. Spring framework for rapid open source J2EE Web application development: A case study. In Proceedings of the Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing and First ACIS International Workshop on Self-Assembling Wireless Network, Towson, MD, USA, 23–25 May 2005; pp. 90–95.
34. Gajewski, M.; Zabierowski, W. Analysis and Comparison of the Spring Framework and Play Framework Performance, Used to Create Web Applications in Java. In Proceedings of the Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH), Polyana, Ukraine, 22–26 May 2019; pp. 170–173.