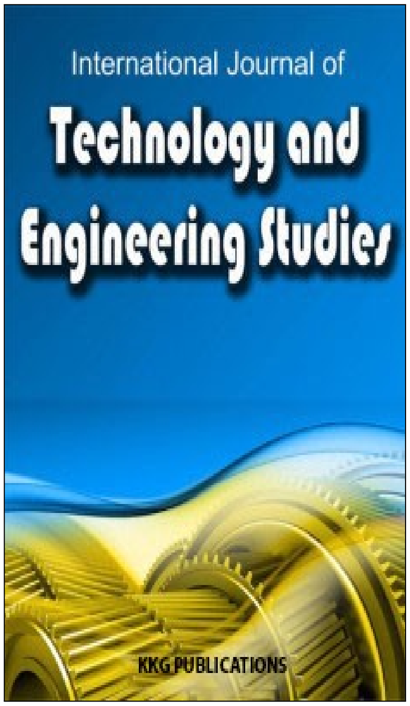

Key Knowledge Generation

Publication details, including instructions for author and subscription information:
http://kkgpublications.com/technology/

A Construction of Vehicle Image and Ground Truth Database for Developing Vehicle Maker and Model Recognitions

JOON WOO JEON [1], DONG-HYUN KIM [2], BUMSUK CHOI [3], GEONWOO KIM [4], YOO-SUNG KIM [5]

[1, 2, 5] Inha University, Incheon, South Korea
[3, 4] Electronics and Telecommunications Research Institute, Korea

Published online: 28 December 2017

To cite this article: J. W. Jeon, D. H. Kim, B. Choi, G. Kim and Y. S. Kim, "A construction of vehicle image and ground truth database for developing vehicle maker and model recognitions," *International Journal of Technology and Engineering Studies,* vol. 3, no. 6, pp. 229-235, 2017.
DOI: https://dx.doi.org/10.20469/ijtes.3.40002-6

To link to this article: http://kkgpublications.com/wp-content/uploads/2017/3/IJTES-40002-6.pdf

KKG PUBLICATIONS

International Journal of Technology and Engineering Studies
vol. 3, no. 6, pp. 229-235, 2017

IJTES

A CONSTRUCTION OF VEHICLE IMAGE AND GROUND TRUTH DATABASE FOR DEVELOPING VEHICLE MAKER AND MODEL RECOGNITIONS

JOON WOO JEON [1], DONG-HYUN KIM [2], BUMSUK CHOI [3], GEONWOO KIM [4], YOO-SUNG KIM [5*]

[1, 2, 5] Inha University, Incheon, South Korea

[3, 4] Electronics and Telecommunications Research Institute, Korea

Keywords:
Vehicle Database
Vehicle Maker and Model Classifier
Machine Learning
Video Annotation System
Ground Truth Data

Received: 05 October 2017
Accepted: 02 December 2017
Published: 28 December 2017

Abstract. In this paper, the construction process of the INHA Vehicle Database, which can be used for developing of a Vehicle Maker & Model (VMM) classifier, is introduced. In order to develop a vehicle detector and/or VMM classifier with the machine learning technology for the social security services or Intelligent Transportation Systems (ITS), a large volume of vehicle images and ground-truth data acquired from various surveillance cameras in real environments should be obtained. For such purpose, fixed CCTV cameras, dash-cams attached to operating vehicles, and smartphones are being utilized for recording vehicle videos. From these vehicle videos, about 11,855 vehicle images and ground-truth data are being created in a month using INHA-VAS (Video Annotation System) for the INHA Vehicle Database. The applications of the current system and recommendations for future research are also discussed.

INTRODUCTION

Some social safety services, which can be utilized by the general public, have been nationally recognized and utilized to voluntarily report and request information on encountered vehicle accidents using images and videos acquired from CCTV cameras, dash-cams, and smartphones [1], [2], [3]. Nowadays, related to these social safety services, news about early resolution of accident-related conflicts through assessment of CCTV camera or dash-cams recorded videos can be commonly encountered. For these social safety services or ITS, automatic car detection and VMM classification functions from input video are essentially required [4], [5], [6].

Recently, many researchers have been working to develop new technologies for VMM recognition using computer vision and machine learning technologies [5].

A large number of vehicle images are needed to develop VMM recognition schemes using the machine learning technology. Previous researches report i.e., development of vehicle detector and VMM classifier with database containing a large number of vehicle images [5], [6], [7], [8], [9], [10], [11], [12]. In previous researches, database containing vehicle images were either obtained by a research team collecting for vehicle images themselves [4], [5], [7], [8], [9], [12] or utilizing public vehicle image database constructed for vehicle learning researches [6], [10], [11]. A research Gao & lee [5] conducted in Korea developed a VMM classifier utilizing deep neural network with self-collected vehicle image database [5]. In this database, only the front images of 107 vehicle types were collected for a total of 3,210 images. In the previous studies [7], [8], [9] carried out in other foreign countries, a self-collected database with 1,000 vehicle images was utilized to recognize front located vehicle logo [7], [8] and on the other hand, about 200 vehicle images were utilized for 3-dimensional modeling of a vehicle exterior and to classify VMM [9]. A vehicle image database used in [11] included a total of 7,000 front and rear view images of various vehicles and was constructed during the Vision-Based Intelligent Environment (VBIE) project in Taiwan [13]. This database was used not only for a research of [11] published by the authors of [11], but also by another foreign research institute for a publication [6]. In a recent study [10], in order to study a vehicle model recognition technique using the Convolutional Neural Network, a large volume of image data provided by ILSVRC 2012 was first run through pre-training and then actual training process with self-collected 31,148 vehicle images to optimize the parameters of CNN and analyze precision in recognition of 13,333 images. The car image database used in [10] utilized the car related database of [11]. This database contained about 50,000 images taken by a surveillance camera in actual environments among the total

*Corresponding author: Yoo-Sung Kim
†Email: yskim@inha.ac.kr

of 136,727 images. Recent study trend on VMM recognition development indicates using deep learning technique and a large number of vehicle images compared to other studies. For example, in a study conducted in Korea, [5] Deep Neural Network was used with 3,210 car images. In a foreign study [10], Convolutional Neural Network was used with a total of 44,481 car images.

Since the VMM classifier to be used in the Korean social safety services or the intelligent transportation systems should function properly for all vehicles operated in Korea, videos and images of all vehicle models operated in the country should be taken in various environments and various surveillance cameras. Therefore, in order to develop a domestic vehicle-related recognizer which can be used in Korea, a large-scale database must be constructed from images of vehicles actually photographed in actual diverse environments for vehicles widely used in Korea.

In order to fulfill the purpose of this study, various environments, surveillance cameras, shooting angles, and distances should be differentiated during recording of exterior images of various vehicle types operated in Korea. Furthermore, generating their ground truth data should be dealt with for constructing vehicle image and ground truth database. For this, in Section 2, existing vehicle image databases and developing of previous VMM classifiers using such databases will be introduced. Section 3 will describe the vehicle video collection method, ground truth attributes, and the generation method of these ground truth data. In addition, the amount of data in the constructed vehicle image and ground truth database will be analyzed. Finally, conclusion of the study will be included in Section 4.

RELATED WORK

As the vehicle image database constructed for development of vehicle detectors and/or VMM classifiers, there are the vehicle image DB (hereafter, referred to as VBIE Vehicle DB) constructed by the VBIE Project [13], image DB (hereafter, referred to as Comp Cars DB) constructed by [14], and a vehicle image database (hereafter, refer to as ULouisville's VMMR DB) constructed by [15], [16].

The VBIE Vehicle DB used in [11] is a database constructed during the VBIE [13] of the 4-year national project plan conducted by Taiwan's Ministry of Economic Affairs. This DB includes front and rear vehicle images of a total of 39 models photographed during actual day-time and night-time periods with various weather environments such as sunny, cloudy, rain, and so on [11]. This VBIE vehicle DB was used in other similar study [11] by the same authors as well as in [11] because it contains vehicle images photographed in different weather and time conditions. It also was used for other foreign study [6], [17]. However, since the VBIE Vehicle DB consists only of the front and rear images of the vehicle, it cannot be used to develop vehicle detectors or VMM classifiers for vehicles photographed from the side, front and rear sides.

The Comp Cars DB constructed in [16] is a large-scale vehicle image database built for the purpose of developing automobile-related applications with computer vision technology. It includes not only exterior images but also images of vehicle parts. The Comp Cars DB contains 136,727 car exterior images for a total of 1,687 car models, and only 50,000 car real images taken by surveillance camera in real environments. The rest of the images are automobile advertisement photographs collected from the web pages. In addition, 50,000 real images taken by surveillance cameras in real life contain only the front view images.

Since the Comp Cars DB images captured with surveillance cameras in real environment are relatively small and most of the images are front view images despite its large sized database, Comp Cars DB is not appropriate for detecting car images of different view angles for developing VMM classifier.

The ULouisville's VMMR DB built in [15] contains a large number of images taken in multiple view angles using different surveillance cameras and environments. The ULouisville's VMMR DB contains a total of 291,752 images of 9,170 vehicle models produced between 1950 and 2016. However, these car images were collected through crawling of the internet homepages related to vehicle sales.

Since the existing vehicle image databases were constructed mostly in foreign countries without the images domestically operated vehicles in Korea, these databases are not appropriate for developing a VMM classifier for the public participated social safety service or ITS which will be used in Korea. In addition, since most of the vehicle images were of groomed and fabricated for the sales purpose unlike the images taken by surveillance cameras in real environments and most of the pictures taken by the surveillance cameras were limited to the front or rear view of the vehicles, image DB of previous studies are not suitable for developing a system to recognize vehicles situated in real environment and classify VMM. Therefore, in order to develop a vehicle recognizer suitable for widely used vehicles situated in real environments in Korea, a large-scaled database with real car images captured with various surveillance cameras such as CCTV cameras, dash-cams, and smartphones in different view angles was considered.

CONSTRUCTION OF VEHICLE IMAGE AND GROUND TRUTH DATABASE

In this section, we describe the process of constructing a vehicle image and ground truth database named as INHA Vehicle DB for developing a domestically suitable VMM classifier to be used by the social safety services or ITS in Korea.

Overall Procedure

CCTV cameras installed in the campus of Inha University, dash-cams attached to operating vehicles, and portable smartphones are used to capture vehicle videos for the construct of INHA Vehicle DB. CCTVs were installed around the campus of Inha University to take videos of vehicles at 8 different angles and dash-cams were installed in vehicles operating around the university campus and nearby roads to collect videos of vehicles in motion. Videos of cars parked around the campus were taken by smartphones at 360 degree view angle. Manufacturers and models of vehicles widely sold and operated in Korea were analyzed and included in the INHA Vehicle DB. Among 77 models of 6 domestic manufacturers and 50 models of 6 overseas manufacturers were targeted and a total of 127 models of 12 manufacturers were selected for the DB.

As shown in Figure 1, vehicle images were collected from collected vehicle videos described in step 1 and the video annotation system, INHA-VAS, developed and used by our research team was utilized to generate annotation information related to collected vehicle images. As shown in Step 2, when a vehicle video is inputted with the modified INHA-VAS, the annotation creator selects a frame for which ground truth data should be generated, and Region of Interest (ROI) marked vehicle images from that frame, and Ground-Truth data for the ROI marked vehicles are generated semi-automatically and saved as the XML format. In order to add and manage vehicle images and their ground truth data generated in such ways, the INHA Vehicle DB is generated which can be used to develop a vehicle detector or VMM classifier using the Database Management & Retrieval System (Step 3).

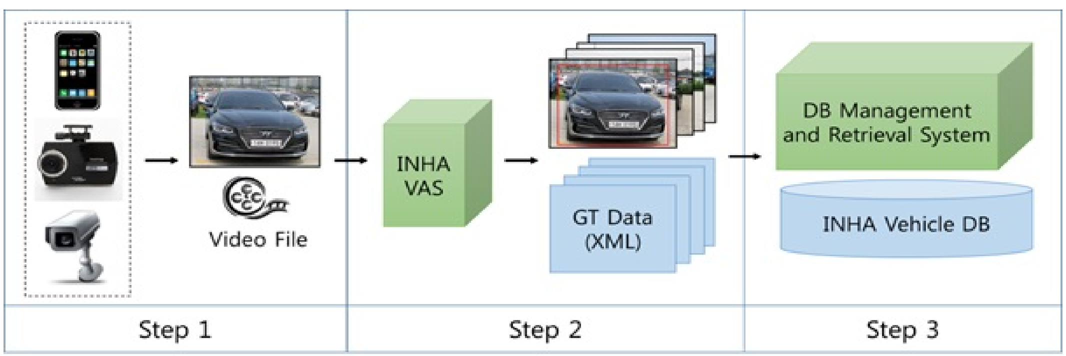

Fig. 1. Overall procedure for constructing INHA vehicle DB

Annotation for Vehicle Videos with INHA-VAS

The video annotation generation system, INHA-VAS, [18], [19] which is developed by our research team, mainly supports the manual annotation and automatic annotation functions to efficiently generate annotation data for various objects and events in videos. The manual annotation method is the most basic method for generating video annotation data. This method, which allows direct confirmation and set-up of the annotation area at the same time, also allows generating annotation data for objects as well as the events in video frames. As for the object annotation, annotations can be efficiently generated through the Propagation and Interpolation algorithms. The automatic annotation method is a method which allows automatically generating initial annotation and allows manual correction against the incorrect initial annotation created by the recognizer. For automatic annotation function, however, the previous other systems could not support automatic generation of annotation data for various objects and events since they were limited to only using the built-in recognizer. Therefore, we developed an open API [19] that can be easily linked to a third-party detector developed by other researchers independently. The third-party detector linked with Open API has the automatic generating function to annotate the specific objects and events after simple registration to the INHA-VAS as a module. The above INHA-VAS (Video Annotation System) was modified to have a structure as shown in Figure 2. It was modified to obtain vehicle images and annotation information from collected vehicle videos. Using the modified INHA-VAS, when a vehicle video is entered, third-party detector can detect vehicles in every video frames and then, human user can manually reviewed the initial annotation data generated by using a third-party detector. Thereafter, finally selected frame, vehicle image included in the selected frame, and annotation data are generated in the XML format per video frame.

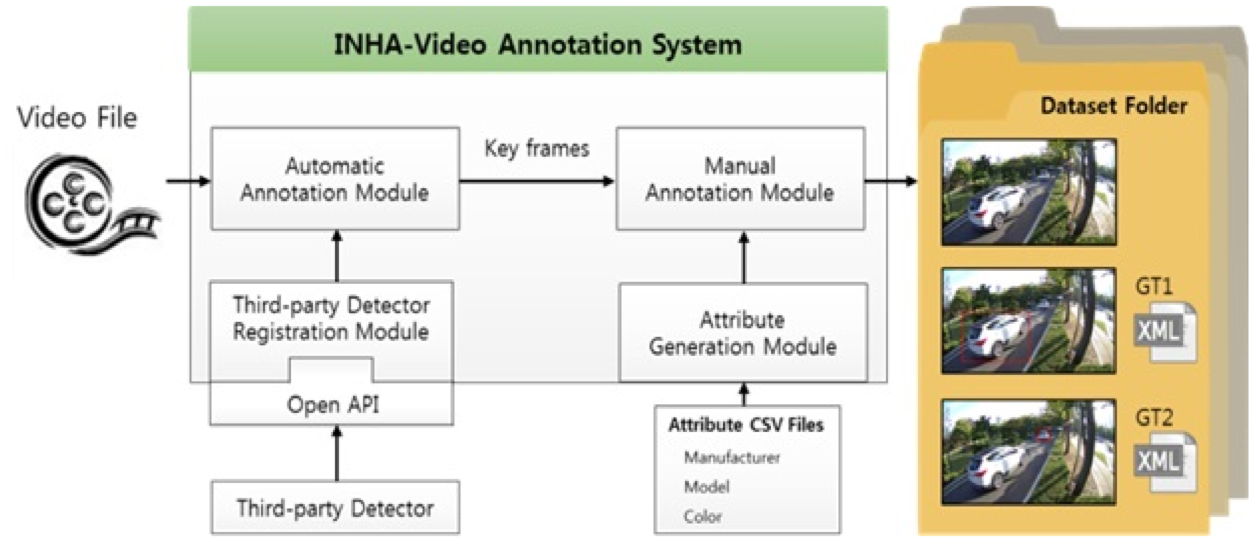

Fig. 2. Architecture of modified INHA-VAS

TABLE 1
ATTRIBUTES FOR VEHICLE ANNOTATION

Attribute	Description	Example Values
Manufacturer	Name of manufacturer	Hyundai, KIA, BMW, Audi etc (12)
Model	Name of vehicle model	Sonata, Santafe, BMW 5, etc (127)
Vehicle Class	Vehicle classification	Economy, Compact, Full-size etc
Color	Color of vehicle	Black, White, Silver etc
View Direction	Direction of camera view to vehicle Front	Rear, Front-side, Rear-side etc (8)
Distance	Distance between camera and vehicle	1∼30 m
Image Quality	Clearness of vehicle image	5 levels of 1-5
Shadow	Amount of the shadow covering vehicle	4 levels of 0-3
Time	Time for video recording	2017-09-14-09-59
Resolution	Resolution of input video	720x640, 1080x780 etc
Weather	Weather of video recording	Sunny, Overcast, Rainy, Snow
Device	Types of recording surveillance cameras	CCTV, Dashcam, Smartphone, etc.

The ground truth attributes for developing vehicle detector and/or VMM classifier are shown in the Table 1.

In order to generate an annotation of a vehicle with the INHA-VAS, the car object should be registered and the vehicle video (for example, Video ID: 090_20171022_08404) should be entered prior to viewing the first frame of the vehicle video in the Annotation window' of the INHA-VAS. First, select a frame (for example, Frame number 226) with a target car to generate an annotation using the 'Video control buttons' and 'Sliding bar' at the bottom, and draw ROI rectangle of the desired car and press the 'Annotation button' for a pop-up of an attribution entry window as shown as in the right image of the Figure 3.

In the following property input window, the contents previously entered are presented as default. Alteration to the contents can be made by selectively changing information and pressing the 'Done button'. Once the 'Done button' is clicked, annotation is created within the folder with a name as the entered video ID. Image files and ground truth data in the XML format will be separately saved based on the corresponding frame number and object ID of selected vehicle.

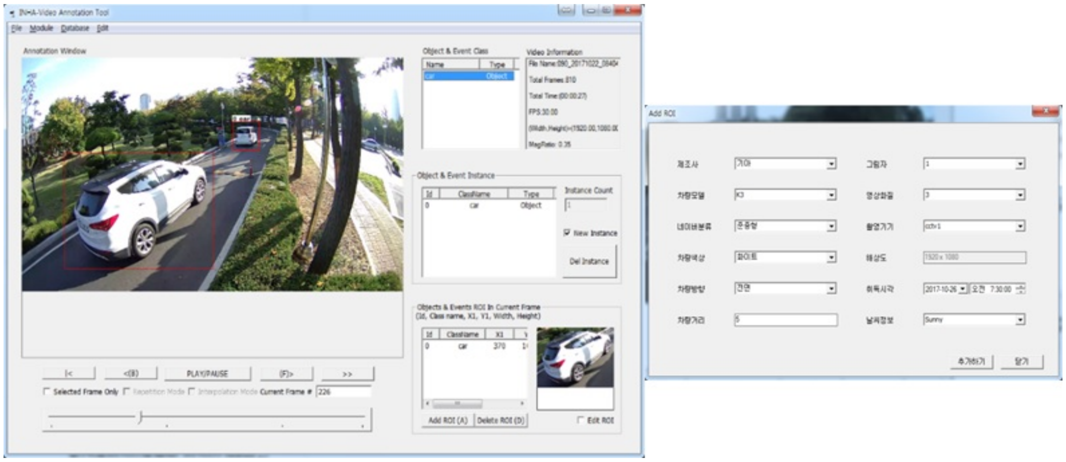

Fig. 3. Creation of annotations for cars with INHA-VAS

The image in Figure 4 shows vehicle images of two vehicles within Frame number 226 of Video ID 090_20171022_08404 and the folder contains created annotations and file names based on the frame number and car object ID. Moreover, the image shown in Figure 4 illustrates the ground truth data of a vehicle with object ID-1 in the frame 226 generated in the XML format.

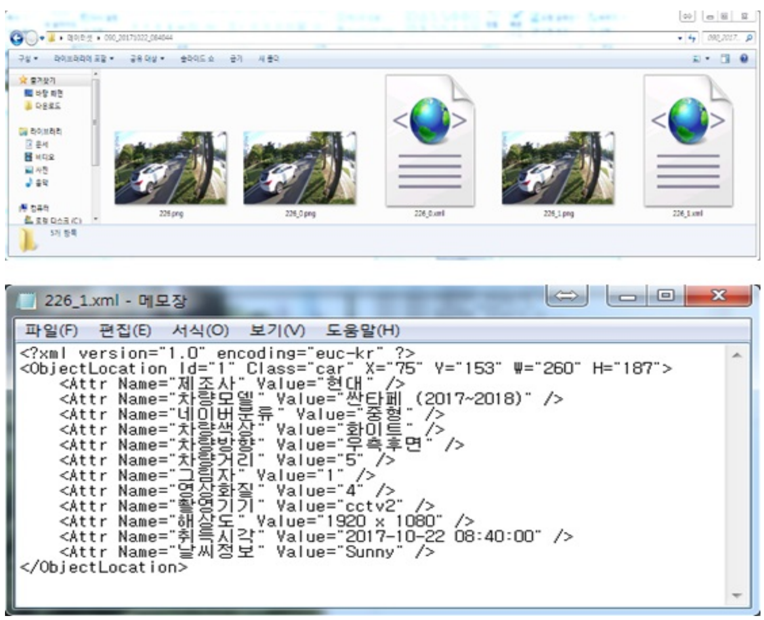

Fig. 4. Annotation Results in the video folder and ground-truth data in XML

Construction of INHA Vehicle DB

This project, which is currently in progress, has been collecting vehicle videos using various surveillance cameras such as CCTV cameras, dash-cams, and smartphones. Using 8 CCTV cameras installed within the campus of Inha University, more than 1,000 registered vehicles to the university have been videoed per day compounding to an average of 8,000 vehicles per day. In addition, more than 100 vehicle images have been acquired to an average daily basis using dash-cams attached to operating vehicles and portable smartphones.

Up till now, a total 11,855 vehicle images and their ground truth data are generated and included into INHA Vehicle Database. The database schema for INHA Vehicle Database is as shown in Figure 5.

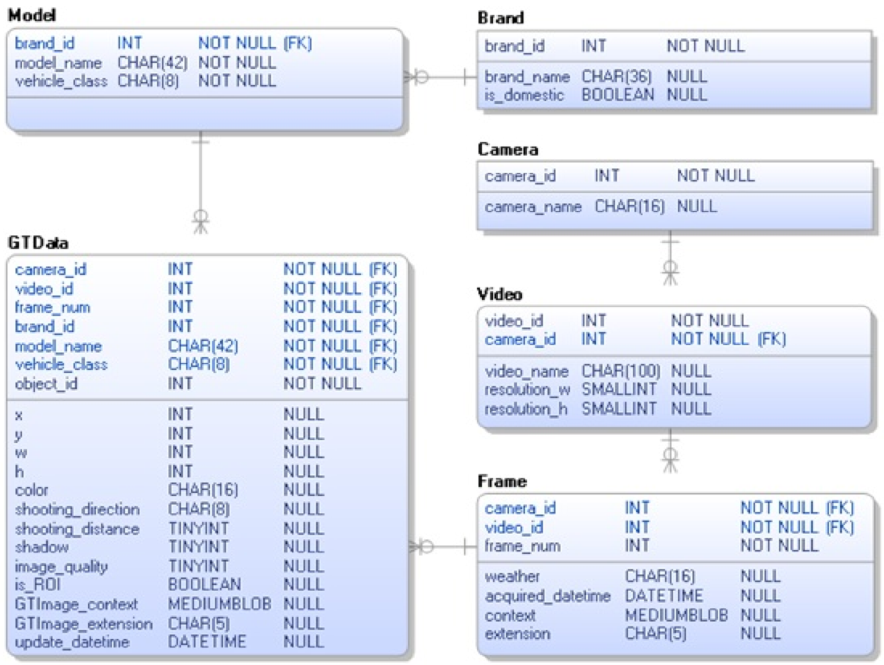

Fig. 5. Database scheme for INHA vehicle database

According to the analysis of the current INHA Vehicle Database, we have 9 major vehicle manufactures such as Hyundai, KIA, Reault Samsung, Chevrolet, Genesis, Ssangyong, Benz, Volkswagen, and BMW by which more than 200 vehicle images are collected and included into the database. The vehicle images recorded by CCTV cameras are 11,617 while the number of vehicle images recorded by dash-cams is 147 and 91 vehicle images are recorded by smartphone cameras. And the vehicle models, for which more than 200 vehicle images are included in the database are 21. The numbers of images for recording directions are as shown in Table 2.

TABLE 2
NUMBER OF IMAGES BASED ON RECORDING DIRECTIONS

Recording Directions	Number of Images
Front	2,986
Right rear	2,220
Right front	2,135
Left rear	2,067
Left front	1,529
Right side	438
Rear	376
Left side	104
Total	11,855

CONCLUSION AND RECOMMENDATIONS

This paper have described the construction process of a large-scale INHA Vehicle DB for development of a vehicle detector and VMM classifier using the machine learning technology essential for public participated social safety service or ITS. In order to collect various video images of various vehicle models operated in Korea, CCTV cameras installed around the Inha University campus, dash-cams attached to vehicles in operation, and portable smartphones have been utilized to collect vehicle videos.

The ground truth attributes are defined which are to be used for developing vehicle detector and VMM classifier and the modified INHA-VAS was utilized to produce 11,855 vehicle images at total up to now in the INHA Vehicle DB.

In the future, we plan to develop a vehicle detector and VMM classifier using the INHA Vehicle DB constructed through the abovementioned process and the latest machine learning technique such as the Convolution Neural Network.

Declaration of Conflicting Interests

No conflicts of interest are present in this study.

Acknowledgement

This work was supported by Institute for Information & Communications Technology Promotion(IITP) grant funded by the Korea government (MSIT) (No.B0717-16-0107 , Development of Video Crowd Sourcing Technology for Citizen Participating-Social Safety Services)

REFERENCES

[1] H. J. Kang, "A study on analysis of intelligent video surveillance systems for social security," *Journal of Digital Contents Society,* vol. 17, no. 4, pp. 273-278, 2016.

[2] Ministry of the Interior and Safety. (2017). *Life inconvenience report service* [Online]. Available: https://goo.gl/HA25fL

[3] National Police Agency. (2017). *Smart people reporting* [Online]. Available: https://goo.gl/h595YY

[4] H. Emami, M. Fathi and K. Raahemifar, "Real time vehicle make and model recognition based on hierarchical classification," *International Journal of Machine Learning and Computing,* vol. 4, no. 2, pp. 142-145, 2014.

[5] Y. Gao and H. J. Lee, "Local tiled deep networks for recognition of vehicle make and model," *Sensors,* vol. 16, no. 2, pp. 226-231, 2016.

[6] A. J. Siddiqui, "A robust vehicle make and model recognition system for its applications," Doctoral dissertation, School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada, 2015.

[7] A. Psyllos, C. Anagnostopoulos and E. Kayafas, "Vehicle model recognition form frontal-view image measurements," *Computer Standard and Interfaces,* vol. 33, no. 2, pp. 142-151, 2012.

[8] L. Zhai, L. Li, Z. Liu, Y. Luo, Y. Wang, H. Lai and M. Guan, "An efficient vehicle model recognition method," *Journal of Software,* vol. 8, no. 8, pp. 1952-1959, 2013.

[9] K. Ramnath, S. N. Sinha, R. Szeliski and E. Hsiao, "Car make and model recognition using 3D curve alignment," in *Proceedings of IEEE Winter Conference on Applications of Computer Visions,* Steamboat Springs, CO, pp. 285-292, 2014.

[10] J. Fang, Y. Zhou, Y. Yu and S. Du, "Fine grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture" *IEEE Transactions on Intelligent Transportation Systems (ITS),* vol. 18, no. 7, pp. 1782-1792, 2017.

[11] L. C. Chen, J. W. Hsieh, Y. Yan and D. Y. Chen, "Vehicle make and model recognition using sparse representation and symmetrical SURFs" *Pattern Recognition,* vol. 48, no. 6, pp. 1979-1998, 2015.

[12] T. J. Ha, "A recognition system for car images based on the HTM network," Master thesis, Department of Computer and Information Science, Kyung Sang University, Korea, 2010.

[13] The Industrial Liaison Program of VBIE. (n.d.). *Vision-Based Intelligent Environment (VBIE) project in Taiwan* [Online]. Avaiable: https://goo.gl/4T7NQZ

[14] L. Yang, P. Luo, C. C. Loy and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Boston, MA, pp. 3973-3981, 2015.

[15] F. Tafazzoli, H. Frigui and K. Nishiyama, "A large and diverse dataset for improved vehicle make and model recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop,* Honolulu, HI, pp. 874-881,2017.

[16] C. H. Huang, M. H. Chang and I. H. Lin, "Research on single-board computers clustering the computing performance," *International Journal of Technology and Engineering Studies,* vol. 2, no. 5, pp. 125-133, 2016.

[17] S. Wittayapiyanon and S. Chuepeng, "Sports utility vehicle seat structure simulation of frontal oblique crash," *Journal of Advances in Technology and Engineering Research,* vol. 3, no. 3, pp. 69-78, 2017.

[18] J. S., Kim, H. I. Kim and Y. S. Kim, "A video annotation system with automatic human detection from video surveillance data," in *Proceedings of the Korean Information Science Society Conference,* Korean Institute of Information Scientists and Engineers, Korea, 2012.

[19] D. Kim, Y. Kim, K. Kim and Y. Kim, "An open API for third-party detectorsto automatically generate ground truth datawith INHA video annotation system," in *Proceedings of 3rd International Conference on Big Data Application and Services,* pp. 129-138, 2016.

— This article does not have any appendix. —