2015 IEEE Intelligent Vehicles Symposium (IV)
June 28 - July 1, 2015. COEX, Seoul, Korea

Visual-Based On-road Vehicle Detection: A Transnational Experiment and Comparison

Chao Wang, Huijing Zhao, Chunzhao Guo, Seiichi Mita, Hongbin Zha

Abstract— **As a key technique in ADAS (Advanced Driving Assistant System) or autonomous driving systems, visual-based on-road vehicle detection has been studied widely, while it faces still great challenges, among which are the complexity, diversity and unpredictable changes of the real-world environments. In the authors' previous work, an algorithm was developed in a probabilistic inference framework with its focus on solving the multi-view and occlusion problems at multi-lane motor way scenes. In this research, we seek to answer the questions: how efficient is the system during a long-term operation across a large area of changed conditions? To this end, a large scale experiment is conducted, where three testing data sets are developed containing the samples of more than 30,000 on Beijing's ring roads, 800 on Nagoya's fast road, and 3,000 on Nagoya's downtown streets, and the performance of visual-based vehicle detection concerning the multi-view and occlusion problems across extensive regions and at transnational environments are studied. We present our preliminary findings in this paper, which leads to a more extensive study in future work.**

I. Introduction

Visual-based approaches have been studied widely for on-road vehicle detection as a key technique in ADAS (Advanced Driving Assistant System) or autonomous driving systems. During the past decades, this area represents a large body of research efforts in the literature [1], [2], [3], [4], [5], the state-of-the-art algorithms are evaluated within such as the PASCAL Visual Object Classes Challenge [6] and the KITTIs Vision Benchmark Suite [7] that provide challenging real-world data sets and benchmarks, and significant progress are demonstrated on the intelligent vehicle systems [8], [9] and research platforms [10], [11], [12]. However, visual-based approaches face still great challenges, among which are the complexity, diversity and unpredictable changes of the real-world environments. It is difficult to model the real worlds exhaustively and predict all kinds of changes, and a well-established system could experience degradation on its performance when confronts a new environment that is different with the data sets in algorithm training.

In the authors previous work, an algorithm of visual-based on-road vehicle detection was developed with its focus on solving the multi-view and occlusion problems at multi-lane motor way scenes [13] in a probabilistic inference framework. In this research, we seek to answer the questions: how efficient is the system during a long-term operation across a large area of changed conditions of roads and traffic? Can we make the system transregional or transnational, where the visual appearance of roads and vehicles could have much difference due to external environments and cultural background?

To this end, a large scale experiment has been conducted using the on-road data in Beijing and Nagoya to examine the performance of visual-based vehicle detection across extensive regions and at transnational environments. The experiment is based on our initial algorithm [13] that is trained using the multi-view image samples on the motor ways in Beijing [14]. As a reference, the results using deformable part model (DPM) method [15], which has an open source released, are compared. The Beijing's data was captured while driving a total distance of 65 km about 120 min on the ring roads, which are multi-lane motor ways with no signalized intersections, and the traffic speed and density during the course changed dramatically. On the other hand, Nagoya's data was captured covering its fast road and downtown streets, which are also multi-lane road, and had a total distance of 70km about 60 min. However, the data segments across the signalized intersections are removed in this study to make the data sets comparable. Three testing data sets are developed by manually labelling ground truth on the images of both fully observed and partially occluded vehicles on four distinctive viewpoints, which contains the samples of more than 30,000 on Beijing's ring roads, 800 on Nagoya's fast road, and 3,000 on Nagoya's downtown streets, where the road surface, surroundings and illumination conditions present much differences. The detection performance on each site are studied concerning both fully observed and partially occluded vehicles, and at different traffic densities. The results are compared to analyze the performance that could be affected due to different environment and/or underlying culture. We present our preliminary findings in this paper, which leads to a more extensive study in future work.

The paper is structured as below. We briefly outline the vehicle detection method in section II. We describe the experimental setting in Beijing and Nagoya in section III, and present the results and comparative studies in section IV. Conclusion and future works are addressed in section V.

II. Visual-based Vehicle Detection Algorithm

In this section, we outline briefly the visual-based on-road vehicle detection [13], which is in a probabilistic inference framework using part-based models and viewpoint

This work is partially supported by the NSFC Grants (61161130528, 91120004), and the Hi-Tech Research and Development Program of China [2012AA011801].

C. Wang, H. Zhao, and H. Zha are with the Key Lab of Machine Perception (MOE), Peking University, Beijing, China.

C.Guo is with Toyota Central R&D Labs., Inc., Nagakute, Aichi, Japan.

S.Mita is with the Research Center for Smart Vehicles, Toyota Technological Institute, Hisakata, Nagoya, Aichi, Japan.

Contact: chao.wang@pku.edu.cn

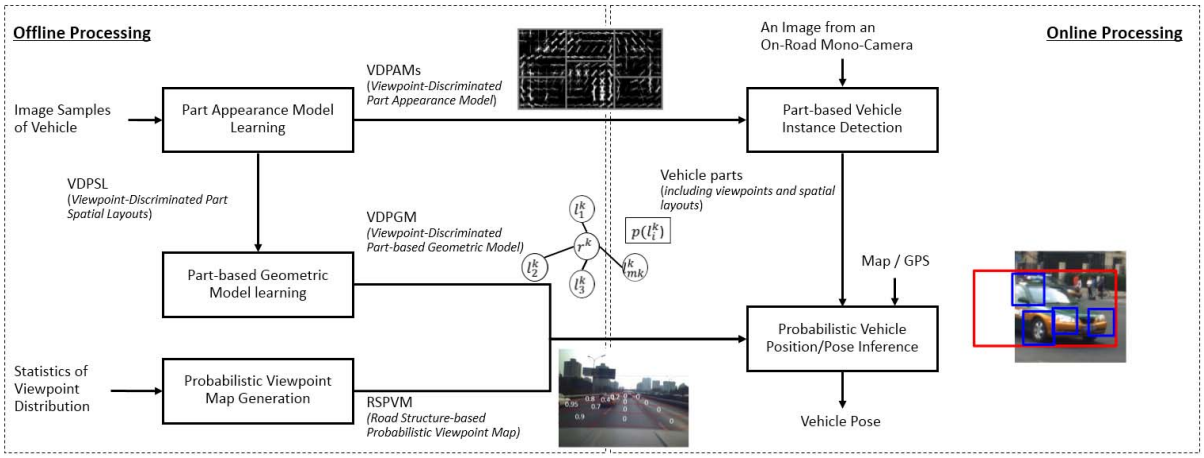

Fig. 1. A probabilistic framework of on-road vehicle detection with part model learning and probabilistic vehicle pose inference.

map to address the multi-view and occlusion problems at multi-lane on-road conditions. As depicted in Fig.1, the system consists of two parts: offline learning and online processing. Part-based vehicle models are used to represent vehicle's major parts on each discriminative viewpoint, and on both appearance and geometry, which are named viewpoint-discriminative part appearance models (VDPAM) and viewpoint-discriminative part-based geometric model (VDPGM) respectively. In offline procedure, part-based models are learned and viewpoint map is generated. While in online detection, a novel probabilistic inference procedure is conducted based on the geometric models which describe vehicle parts spatial configurations and viewpoint maps which provide vehicle viewpoint prediction.

A. Part-based Models

Part-based vehicle models are generated on each discriminative viewpoint k on both appearance and geometry. VDPAMs M_g^k are trained by the existing approach DPM [15], and sample data of part configuration and spatial layout is generated by applying appearance models on training samples to train VDPGM M_g^k. Fig.2 shows examples of learned VDPAMs in different viewpoints. Geometric models describe the configuration of vehicle parts as well as their spatial relations in probabilistic representations. As the configuration of vehicle parts varies greatly in different viewpoints, VDPGMs are learned for each dominant viewpoint using part detection results. For each part its relative location with respect to a vehicle center is described in a 2D Gaussian model which is learned using the statistics of the training data. An example of geometric model on viewpoint 2, as well as its learning details are depicted in Fig.3.

B. Probabilistic Viewpoint Map

On-road vehicle motion follows certain rules which has strong correlation with road structure that is defined by road geometry, lanes and traffic rules. Given a type of road structure, viewpoints of vehicles could be predicted for any on-road location at an ego or world frame. A probabilistic

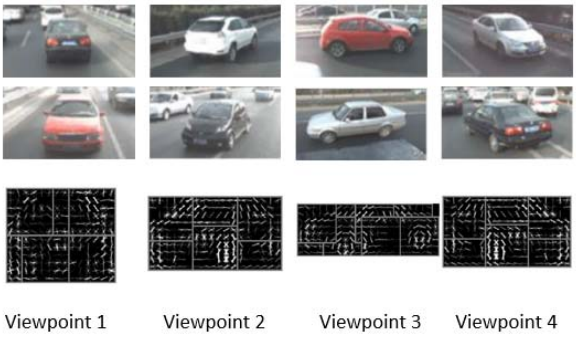

Fig. 2. A result of learned VDPAMs of different viewpoints.

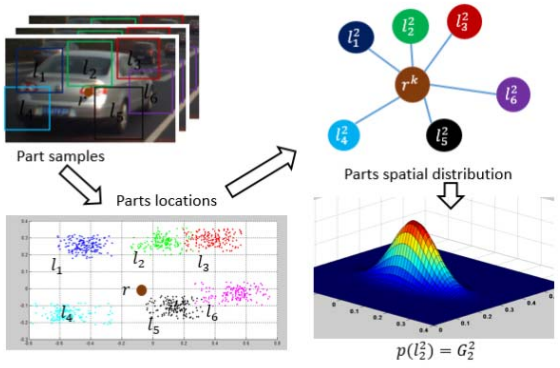

Fig. 3. A result of learned VDPGM of viewpoint 2.

representation M_r is used to estimate the predictions of a subject vehicles viewpoint. Viewpoint maps provide probabilistic description to the distribution of viewpoints that could improve the efficiency of vehicle inference for certain road strucutre, thus are called road structure-based probabilistic viewpoint maps (RSPVMs). A viewpoint prediction map as shown in Fig.4, for each viewpoint(viewpoint 2 in Fig.4), a grid map at ego vehicle frame is generated in an offline

learning procedure, values in grids are generated to represent the probability of a vehicle to be observed on each viewpoint at the grid's area.

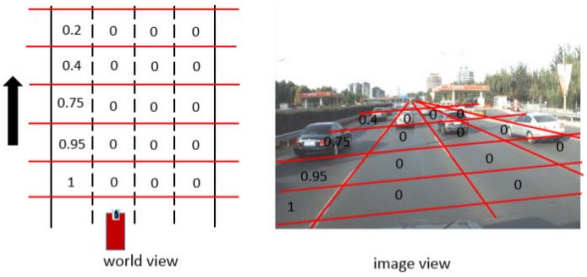

Fig. 4. A result of learned RSPVM of viewpoint 2 on straight road.

C. Probabilistic Inference

Given an image, a set of vehicle part $L^k_{vis} = \{l^k_i, \phi(l^k_i)\}$ is detected on each viewpoint k with learned VDPAMs, where l^k_i is a part instance recording the corresponding index of part and viewpoint, as well as its location on image, $\phi(l^k_i)$ is the detection score. With learned VDPGMs and RSPVMs, vehicle's position, viewpoint and occlusion status can be inferred by evaluating part instances' position and spatial relations. The problem of vehicle inference can be represented in a probabilistic way as below.

$$\hat{r^k} = \arg\max_{r^k} p(r^k | L^k_{vis}, M^k_g, M_r)_{k=1,..,4} \tag{1}$$

where r^k is the estimated location of vehicle center on viewpoint k. The objective is defined to find the r^k that maximize the posterior $p(r^k|L^k_{vis}, M^k_g, M_r)$, which can be further extended as below.

$$\begin{aligned} & p(r^k|L^k_{vis}, M^k_g, M_r) \\ & \propto p(L^k_{vis}|r^k, M^k_g, M_r) p(r^k|M^k_g, M_r) \\ & = \prod_{l^k_i \in L^k_{vis}} \phi(l^k_i) p(l^k_i|r^k, M^k_g) \cdot \\ & \prod_{l^k_j \in L^k_{occ}} \phi(l^k_j) p(l^k_j|r^k, M^k_g) p(r^k|M_r) \end{aligned} \tag{2}$$

where $L^k_{occ} = \neg L^k_{vis} = \{l^k_j\}$ is the rest set of the full configuration of M^k_g. A set of r^k is sampled, from which, the r^k that maximize the posterior $p(r^k|L^k_{vis}, M^k_g, M_r)$ is selected as the vehicle pose. In this research, sampling of r^ks is conducted as below. For each detected part l^k_i, it is compared with the geometric model M^k_g of the viewpoint to propose for a potential vehicle pose $\dot{r}^k$, a number of samples $\{r^k_s\}$ are thereupon drawn regularly nearby $\dot{r}^k$.

III. Experimental Settings

A large scale experiment has been conducted using the on-road data in Beijing and Nagoya to examine the performance of visual-based vehicle detection across extensive regions and at transnational environments. The Beijing's data was captured while driving a total distance of 65 km about 120 min on the ring roads as shown in Fig.6 (a), which are bi-directional motorways with four- or five-lane's on each side that has no signalized intersections, and the traffic speed and density changed dramatically during the course. On the other hand, Nagoya's data was captured covering its fast road and downtown streets, which are multi-lane roads too, and had a total distance of 70km about 60 min as shown in Fig.6 (b). However, there had been a number of signalized intersections during the course, the data across these road segments are not studied in this research. In order to examine the performance of vehicle detection at all around viewpoints, an omni-vision system, the PGR Ladybug3 is used to capture on-road images in the experiments at both Beijing and Nagoya (see Fig.5 (a)). However as an omni-image is different with those of normal cameras, it is divided equally into five pieces as shown in Fig5 (b), and corrected from distortions, so as to simulate the images that are captured by a mono-vision camera to the direction.

Testing data sets are developed by manually labelling ground truth vehicles on the omni-images at a rate of 1Hz, where the vehicles are marked by bounding boxes with a label for either a fully or partially observed one, and a viewpoint in four distinctive directions as defined in Fig.2. As shown in Fig.5 (b), a fully observed vehicle is marked using a red bounding box, while an occluded one is bounded in twofold, i.e. one for observed part and one for a guessed full body, which are drawn in Fig.5 (b) in green in blue respectively. In addition, the vehicles on the opposite roadway occluded by high road barriers are not considered in this research, which are marked using black bounding boxes as background. For occluded vehicles, sub-category labels are assigned according to the reason of occlusion, i.e. due to road infrastructure, other vehicles or limited camera's vision field which can be used in further examine the performance on different kinds of occlusions. The corresponding detection results are also shown at Fig.5 (c) for an example. The observable vehicle parts are detected and marked by blue bounding boxes, while those occluded parts are inferred as marked in green, and locations of the whole vehicle body are shown in red as the final detection results. Note that vehicles are labeled as the ground truth if they have larger than 40 pixels on their heights and have more than one-third of the body parts observable. And vehicle's viewpoints are labeled by subjective judgement. As the data sets were labeled by more than ten volunteers, there exists differences and inaccuracies in manual operations.

As a result, three testing data sets are developed on Beijing's ring roads, Nagoya's fast road, and Nagoya's downtown streets, where the number of ground truth vehicles are listed in Tab.I. Fig.6 (a) and (b) show the number of ground truth vehicles in each frame of omni-image along the driving route, which depicts traffic density nearby the ego-vehicle, the color represents vehicle number in each frame. It can be found that the testing data covers a broad range of traffic condition, from smooth to dense traffic. We divide

the testing data into two groups for low and high traffic density. The group of low contains the image frames with less than 10 labeled vehicles, while others are contained in the high one. On the other hand, Fig.6 (c), (d) and (e) are the two-dimensional histograms on three testing data sets by cross-correlating the numbers of all vehicles v.s. occluded ones in a single omni-image frame. It can be found that the number of occluded vehicles increases along with traffic density, which varies in comparable wide ranges in both Beijing ring road and Nagoya downtown, while the traffic density on Nagoya fast road was low during the experiment. The vehicle detection performance are studied on three testing data sets, with the algorithm trained using the multi-view image samples on the motor ways in Beijing [14]. As a reference, the results using deformable part model (DPM) method [15], which has an open source released, are compared.

IV. Results and Comparative Study

A. Vehicle Detection Results

Experimental evaluations are conducted on testing data sets from Beijing and Nagoya to examine the proposed method's performance under different road and environment conditions. Fig.7 shows sequences of detection results on the mono-vision images in Beijing ring road. (a) is a front view sequence under low traffic density condition, which presents the results while a white car entered image from the ego's rear left to the front, and from partially to fully observed. At first white car had only its front observed which was too small a part for detection. Later the white car had larger parts observable in the image and it was succeeded in detection that the red bounding box shows the estimation of its whole body, while green for the inference of occluded parts. Detection was succeed in the later frames. We would stress that in the experimental results of this paper, detection was conducted independently on each image without consideration to inter-frame relations, i.e. no tracking technique was used. The other sequences in Fig.7 can be discussed similarly, where (b) is the results while a car on a farther lane was occluded by a front one; (c) is a rear view one under high traffic density, where front vehicles blocked greatly the vision to farther ones; (d) is a side view one under high dense traffic.

Fig.8 (a) and (b) show sequences of detection results in Nagoya fast road. Most of time there are only 1 or 2 vehicles in camera's sight, thus the detection work is easier than that in Beijing ring road even with similar road conditions. In Beijing ring road and Nagoya fast road, it can be found that background environmental conditions are quite simple at most of the time. Usually there is only closed road guardrail next to the road and buildings are very far. While in Nagoya downtown, as shown in Fig.8 (c) and (d), background environment conditions are much more complex with buildings, road signs, trees and pedestrians just next to the road, which bring more challenges for detection. Besides, as shown in (d), isolation facility of the bi-directional road is too low to prevent the vehicle on opposite roadway in an emergency, thus vehicles on both roadways are considered in Nagoya downtown experiment. In (d) a white car's bottom part is occluded by the green belt and vehicles are occluded by the tree while passing it.

Fig.9 shows several images of false or miss detections in experiment. (a), (b) and (c) are in Beijing ring road. (a) is a front view scene, where a sedan drove next to a van, and two bounding boxes were labelled as the ground truth with one as fully and the other as partially observed. However, the image region of the two vehicles are largely overlapped. The proposed algorithm detected them as one entirety as the red bounding boxes on them, while a miss detection was counted on each frame, where the black bounding box is the ground truth one that has no matching in detection results. The miss detection in (b) depicts also a typical case, where the vision to a farther car was heavily blocked by a front one, reliable detection on such heavily occluded vehicle is still a challenge of the proposed algorithm. False alarms as shown in (c) happened on a overhead bridge that has similar feature with a side view vehicle. These false/miss detections are very common for on-road scene which also happened in Nagoya experiments. While in Nagoya downtown, challenges from complex background bring new problems. As shown in (d), a white car was heavily blocked by a motorcycle which is hard to detect parts to infer the vehicle. (e) and (f) are side view scene with road signs, trees and pedestrians in camera's sight. A road sign and a box were recognized as vehicles because their shape is similar with the vehicle and they located just next to the road where is actually a reliable area in viewpoint map.

B. Quantitative Evaluations

The datasets with labeled ground truth as described in Tab.I are used in quantitative evaluations of detection accuracy for comparative experimental test, where detection results are compared with the labeled ground truth by following the classical object detection protocol of Pascal VOC [1]. Evaluations are conducted on 3 levels, i.e. detection results in Beijing and Nagoya compare with the results of DPM method [1], results on occluded vehicles and different kinds of occluded types.

Fig.10 shows precision/recall curves of detection results on all vehicles in test datasets of Beijing and Nagoya, which compare performance of the proposed method and DPM. It can be found that the proposed method performs better than DPM on the detection results. For Beijing ring road we evaluated the performance on overall frames and then on different traffic density conditions. With less vehicles in camera's sight, detection performance of low traffic density is better than as more severe occlusions and challenges happen under higher traffic density. Detections in Nagoya fast road has best performance as the traffic density is much lower in Nagoya fast road than Beijing ring road even in low traffic density condition. While performance of detections in Nagoya downtown is worse than that in Beijing ring road and comparable with the high traffic density condition in

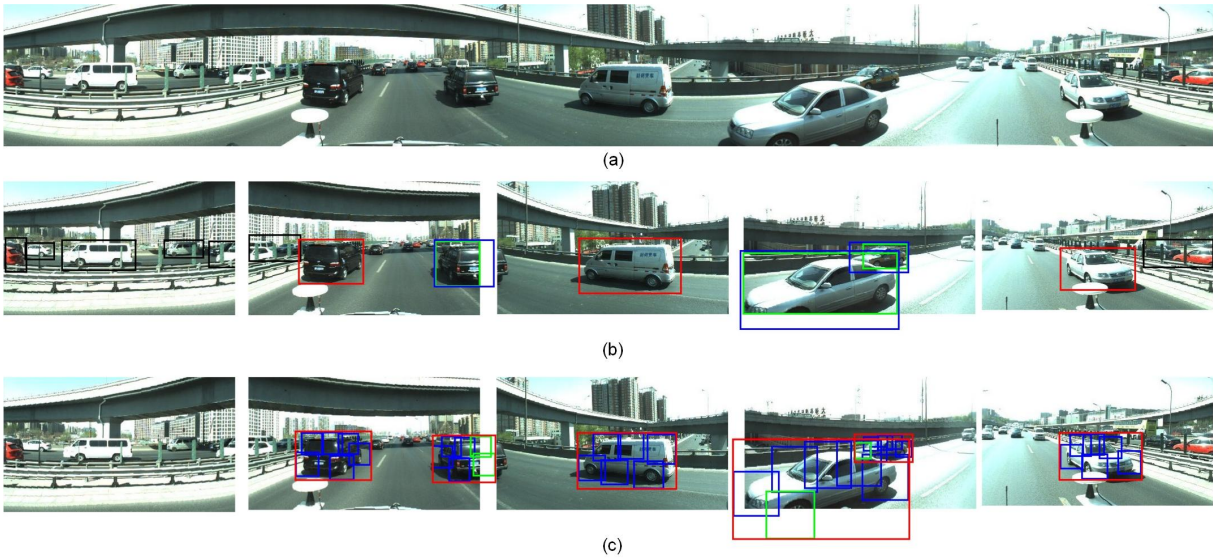

Fig. 5. (a) Omni image; (b) Ground truth bounding boxes on processed images; (c) Detection results.

TABLE I
TEST DATA SET DETAILS

			Fully Observed	Partially Observed		
				Occ. Road.	Occ. Vehi.	Occ. Cam.
Beijing Ringroad	Viewpoint	1	2829	0	958	683
		2	5792	0	113	631
		3	6057	56	2016	3877
		4	1302	89	764	6279
	Traffic Density	Low	11337	97	2283	7381
		High	4643	49	1568	4089
	Total		15980	145	3851	11470
				15466		
Nagoya Fast Road	Viewpoint	1	258	0	2	31
		2	122	0	0	3
		3	141	0	0	93
		4	60	2	6	138
	Total		581	2	8	265
				275		
Nagoya Downtown	Viewpoint	1	354	10	84	288
		2	539	12	54	49
		3	505	44	87	229
		4	172	28	43	560
	Total		1570	94	268	1126
				1488		

Beijing, this is because the complex background conditions bring more false alarms.

Fig.11 shows evaluation of detection results only for occluded vehicles compared proposed method with DPM in Beijing and Nagoya. It can be found that the proposed system performs much better on occluded vehicles. The result demonstrates importance of part-based inference in occluded scene. And for Nagoya downtown, as more false alarms appear due to the complex background, the performance is worse than that in Nagoya fast road and Beijing ring road.

Performance on detecting different kinds of occluded vehicles is also evaluated. For each ground truth that labeled as a occluded vehicle, a sub-category label is assigned denoting the reason of occlusion. In the test data set as shown in Tab.I, there are few samples that are occluded by road infrastructures compared with the other two occluded types, we only evaluate the performance on types of occlusions due to surrounding vehicles or limited vision field. For Nagoya test data sets we evaluate the performance in downtown and ignore that in fast road as there are nearly all occluded vehicle by camera's limited vision field. It can be found in Fig.12 that detection of the vehicles partially occluded due to limited vision field has better performance on those due to surround vehicles, which could be explained that the occlusion pattern of the former one is rather stable, the model of which could be learned with higher reliability, however the later one has much more variety of patterns and suffers more uncertainties from the dynamic environment. Especial-

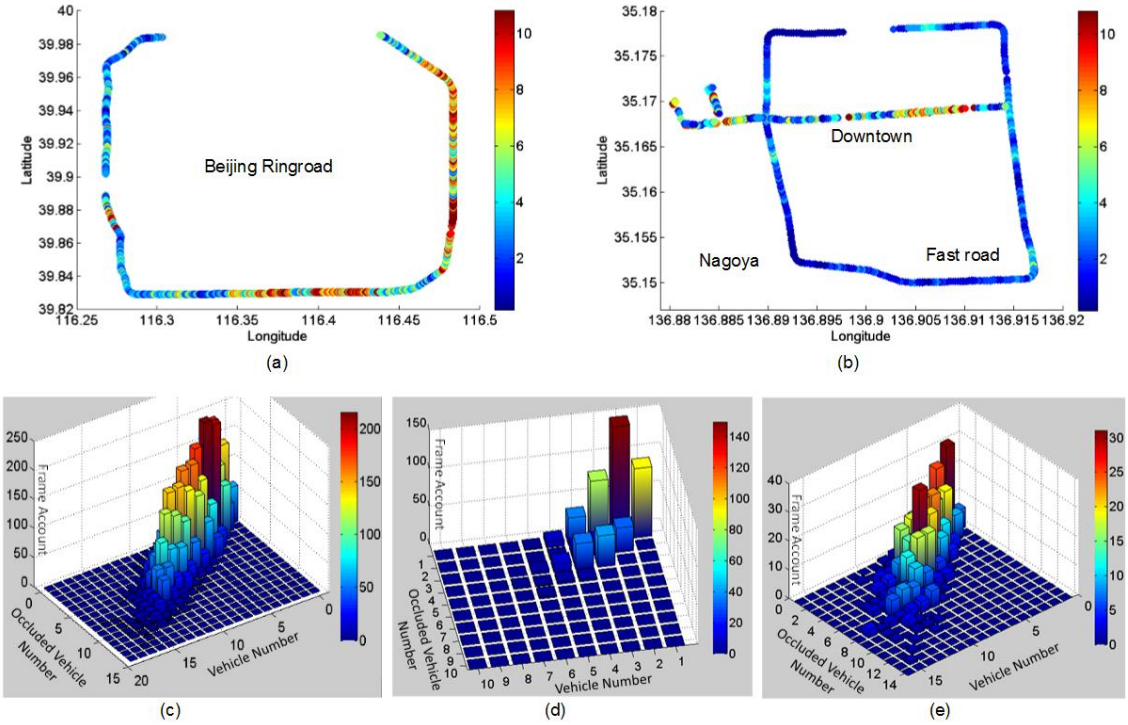

Fig. 6. Traffic density in comparative experiments between Beijing and Nagoya.

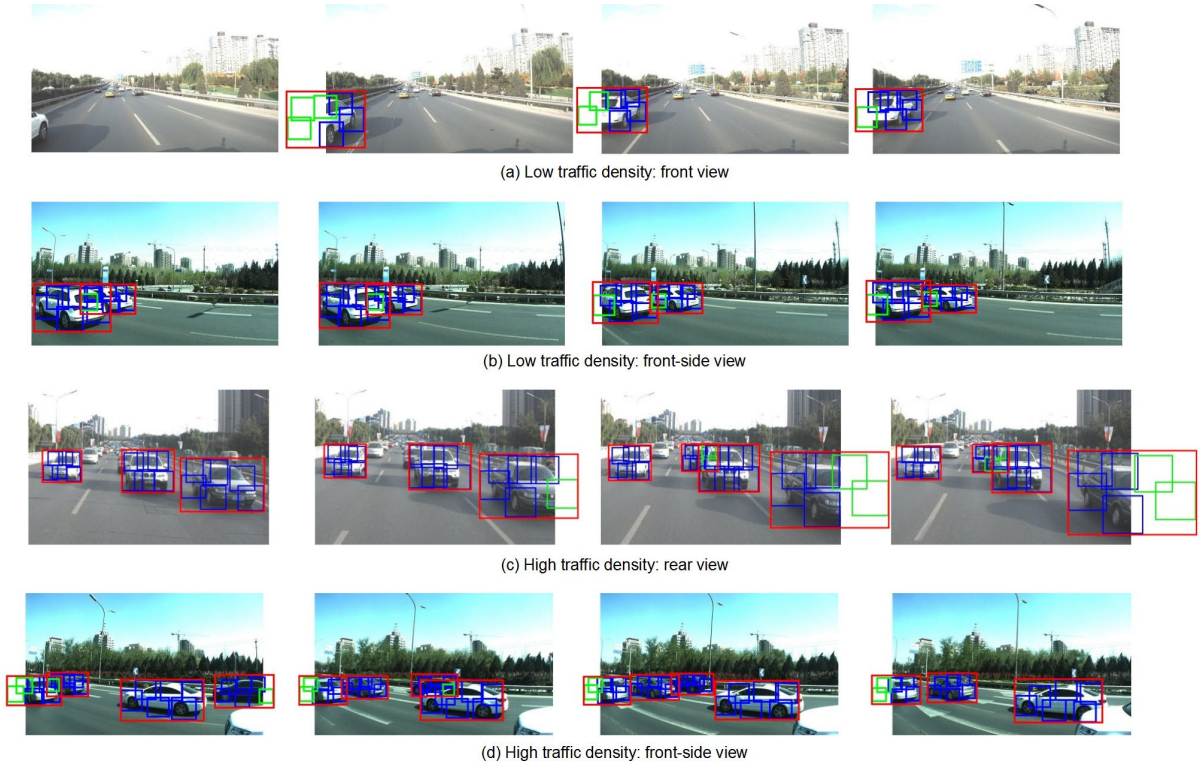

Fig. 7. Detection results in Beijing ring road under different traffic density.

ly when occlusion is caused by a vehicle with similar color. While detection of vehicles occluded by road infrastructures has similar performance compared with that occluded by limited vision field, which is because the occlusions due to

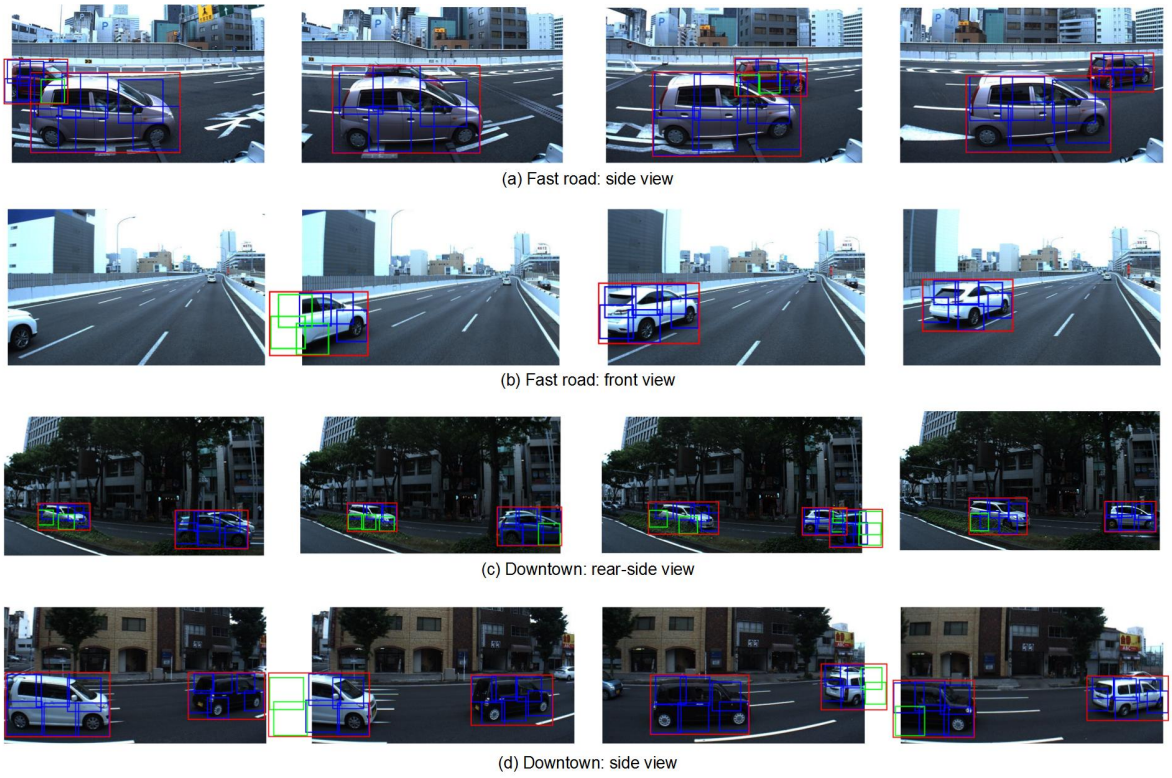

Fig. 8. Detection results in Nagoya under different road conditions.

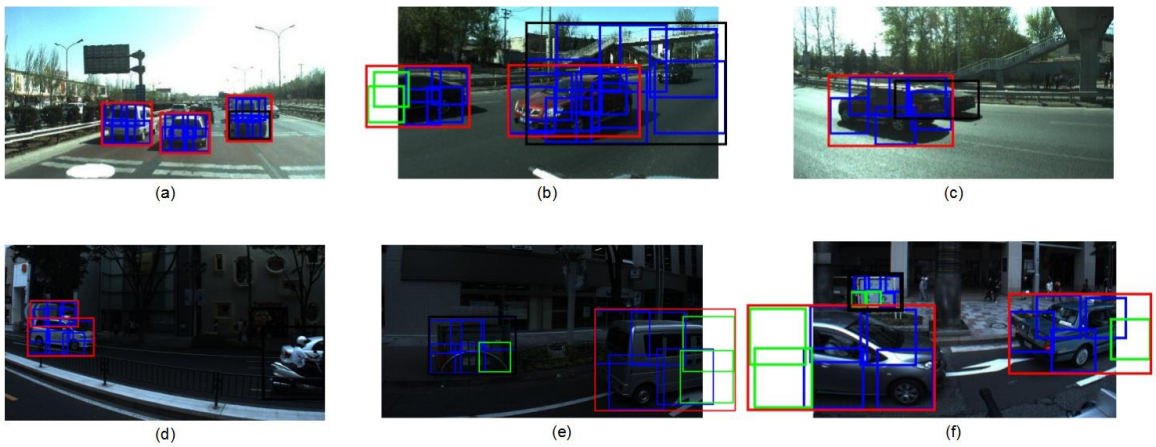

Fig. 9. False detections in Beijing and Nagoya.

green belt or low road barriers are usually small parts.

V. Conclusion

Although visual-based on-road vehicle detection has been studied widely for ADAS or autonomous driving systems, it faces still great challenges from the complexity, diversity and unpredictable changes of the real-world environments. It has always been the question, how efficient is the system during a long-term operation across a large area of changed conditions? This research present an experimental and comparative study based on the authors' previous development, which is a visual-based on-road vehicle detection that solves the multi-view and occlusion problems at multi-lane motor way scenes in a probabilistic inference framework. The experiment is conducted using the on-road data in Beijing and Nagoya, where three testing data sets are developed containing the samples of more than 30,000 on Beijing's ring roads, 800 on Nagoya's fast road, and 3,000 on Nagoya's downtown streets, and the performance of visual-based vehicle detection concerning the multi-view and occlusion problems across extensive regions and at transnational environments are studied. We present our preliminary findings in this paper, while more

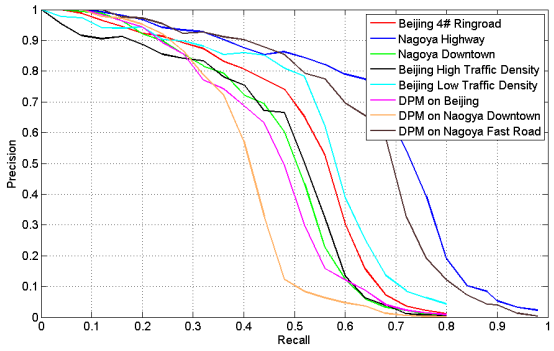

Fig. 10. Quantitative evaluation of detection accuracy.

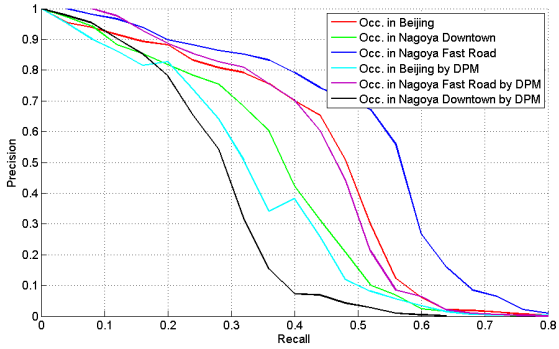

Fig. 11. Quantitative evaluation of detection accuracy on occluded vehicles.

extensive studies will be addressed in future work.

REFERENCES

[1] S. Sivaraman, M. M. Trivedi, *Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis*, IEEE Transactions on Intelligent Transportation Systems, 14(4):1-23, 2013.

[2] Q. Yuan, A. Thangali, V. Ablavsky, S. Sclaroff, *Learning a Family of Detectors via Multiplicative Kernels*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(3):514C530, 2011.

[3] Z. Sun, G. Bebis, R. Miller, *Monocular Precrash Vehicle Detection: Features and Classifiers*, IEEE Transactions on Image Processing, 15(7):2019-2034, 2006.

[4] C. Huang, R. Nevatia, *High Performance Object Detection by Collaborative Learning of Joint Ranking of Granules Features*, IEEE Conference on Computer Vision and Pattern Recognition, 41-48, 2010.

[5] S. Sivaraman, M. Trivedi, *A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking*, IEEE Transactions on Intelligent Transportation Systems, 11(2):267C276, 2010.

[6] PASCAL Visual Object Classes Challenge, *http://pascallin.ecs.soton.ac.uk/challenges/VOC/*

[7] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, *Vision meets Robotics: The KITTI Dataset*, International Journal of Robotics Research, 1-7, 2013.

[8] A Broggi, P Medici, P Zani, A Coati, M. Panciroli, *Autonomous Vehicles Control in the VisLab Intercontinental Autonomous Challenge*, Annual Reviews in Control, 161-171, 2012.

[9] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, *Towards Fully Autonomous Driving: Systems and Algorithms*, In Proceedings of the Intelligent Vehicles Symposium, 163-168, 2011.

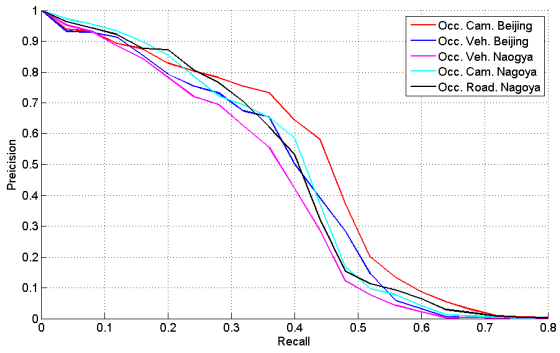

Fig. 12. Quantitative evaluation of detection accuracy on occluded vehicles in different occluded type categories.

[10] C. Caraffi, T. Vojii, J. Trefny, J. Sochman, and J. Matas, *A system for real-time detection and tracking of vehicles from a single car-mounted camera*, International IEEE Conference on Intelligent Transportation Systems, 975-982, 2012.

[11] P. Lenz, J. Ziegler, A. Geiger, M. Roser *Sparse Scene Flow Segmentation for Moving Object Detection in Urban Environments*, Intelligent Vehicles Symposium, 2011.

[12] H. T. Niknejad, A. Takeuchi, S. Mita, D. McAllester, *On-Road Multivehicle Tracking using Deformable Object Model and Particle Filter with Improved Likelihood Estimation*, IEEE Transactions on Intelligent Transportation Systems, 13(2):748C758, 2012.

[13] C. Wang, H. Zhao, C. Guo, S. Mita, H. Zha, *On-road Vehicle Detection through Part Model Learning and Probabilistic Inference*, IEEE/RSJ International Conference on Intelligent Robots and Systems, 4965-4972, 2014.

[14] C. Wang, H. Zhao, F. Davoine, H. Zha, *A System of Automated Training Sample Generation for Visual-Based Car Detection*, IEEE Conference on Intelligent Robots and Systems, 4169-4176, 2012.

[15] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, D. Ramanan, *Object Detection with Discriminatively Trained Part Based Models*, IEEE Transactions on Pattern Analysis and Machine Interlligence, vol. 32, pp. 1627-1645, 2010.