# Probabilistic Inference for Occluded and Multiview On-road Vehicle Detection

Chao Wang, *Member, IEEE*, Yongkun Fang, *Member, IEEE*, Huijing Zhao, *Member, IEEE*, Chunzhao Guo, *Member, IEEE*, Seiichi Mita, *Member, IEEE*, and Hongbin Zha, *Member, IEEE*

***Abstract*—Visual-based approaches have been extensively studied for on-road vehicle detection; however, it faces great challenges as the visual appearance of a vehicle may greatly change across different viewpoints and as a partial observation sometimes happens due to occlusions from infrastructure or scene dynamics and/or a limited camera vision field. This paper presents a visual-based on-road vehicle detection algorithm for a multilane traffic scene. A probabilistic inference framework based on part models is proposed to overcome the challenges from a multiview and partial observation. Geometric models are learned for each dominant viewpoint to describe the configuration of vehicle parts and their spatial relations in probabilistic representations. Viewpoint maps are generated based on the knowledge of the road structure and driving patterns, which provide a prediction of the viewpoints of a vehicle whenever it happens at a certain location. Extensive experiments are conducted using an onboard camera on multilane motor ways in Beijing. A large-scale data set that contains more than 30 000 labeled ground truths for both fully and partially observed vehicles in different viewpoints across various traffic density scenes is developed. The data set will be opened to the society together with this publication.**

***Index Terms*—Intelligent vehicle, on-road vehicle detection, occlusion handling.**

## I. Introduction

ROBUST and accurate on-road vehicle detection is a key issue for Advanced Driving Assistant System (ADAS) and autonomous driving systems. Vision-based approach has been extensively studied for such a purpose, while it faces great challenges even on well-structured road environments. Except the difficulties from such as illumination condition, shadow, cluttered background, various size and shape that visual processing may general meet in outdoor applications, on-road vehicle detection has its own challenges: vehicles may come into the view at different orientations yielding largely different visual appearances, and partial observation may happen always due to limited camera's vision field or occlusions from infrastructure and/or scene dynamics. Examples of such challenges are depicted in Fig. 1.

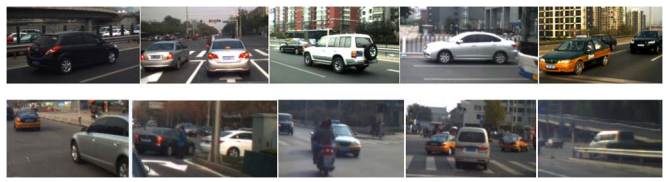


Fig. 1. Challenges in visual-based on-road vehicle detection. Top: varying viewpoints; Bottom: partial observations.

This research studies visual-based on-road vehicle detection with its focus on solving the multi-view and occlusion problems at multi-lane motor way scenes. Inspired by the prior works on part-based detection, especially the success of deformable part model (DPM) [10] and the subsequent researches on on-road vehicle detection [17], [23], this research propose a method of probabilistic inference through part-based detection. To this end, geometric models describing the configuration of vehicle parts as well as their spatial relations in probabilistic representations are learned for each dominant viewpoint, and viewpoint maps are generated based on the knowledge to road structure and driving patterns, which provide a prediction to the viewpoints of a vehicle whenever it happens at a certain location. Extensive experimental studies are conducted using the on-board camera data on the multi-lane motor ways in Beijing, where a large-scale data set is developed containing more than 30 thousands of labeled ground truth of both fully observed and partially occluded vehicles on four distinctive viewpoints across the scenes of various traffic densities. The data set will be opened to the society in accompany with this publication. Proposed work solves multi-view and occlusion problems in on-road vehicle detection, and is demonstrated of efficiency through large-scale experiments using the data at complex multi-lane motor ways.

The paper is structured as below. A review to the literature works is given in Section II. The method of on-road vehicle detection through part model learning and probabilistic inference is presented in Section III, with the data set and experimental study given in Section IV. Conclusion and future works are addressed in Section V.

Manuscript received December 9, 2014; revised May 27, 2015; accepted July 29, 2015. Date of publication October 2, 2015; date of current version December 24, 2015. This work was supported in part by the NSFC under Grants 61161130528 and 91120004 and in part by the Hi-Tech Research and Development Program of China under Grant 2012AA011801. The Associate Editor for this paper was M. Bertozzi.

C. Wang, Y. Fang, H. Zhao, and H. Zha are with the Peking University, with the Key Laboratory of Machine Perception (MOE), and also with the School of Electronics Engineering and Computer Science, Beijing 100871, China (e-mail: zhaohj@cis.pku.edu.cn).

C. Guo is with Toyota Central Research and Development Laboratories, Inc., Nagakute 480-1131, Japan.

S. Mita is with the Research Center for Smart Vehicles, Toyota Technological Institute, Nagoya 468-8511, Japan, and also with the Toyota Technological Institute at Chicago, Chicago, IL 60637 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2015.2466109

## II. Literature Review

### A. On-Road Vehicle Detection

On-road vehicle detection using vision is a very challenging problem which has been extensively studied by many researchers [1]. In outdoor scenes, various challenges such as illumination changes, cluttered background affect detection results greatly. On-road vehicles in image have large variability of appearance in size, viewpoint and color. Partially observation always happens due to the occlusion of static/moving object or limited vision field of camera. A variety features of appearance have been used to detect vehicles, such as symmetry [2], corners [3], edge [4], shadows [5], etc. In recent works, high dimensional feature sets transit from simple image features like edge, orientation have been used widely in object detection and classification, such as Haar like features [19], [20], and histogram of orientation features [6]–[8].

Haar like features can be defined as the difference of the sum in rectangles over an image according different patterns, which are very sensitive to vertical, horizontal and symmetric structures [19]. In [25], Haar features were used to detect the front view of cars in behind for lane change decision assistant. By detecting front and rear wheels, side view vehicles were detected using Haar in [15]. In [20], an active learning framework using Haar features and Adaboost was proposed to detect and track on-road vehicles. HOG features are widely used in many representative work in object detection [6]–[8], [10]. In [6], Dalal proposed HOG feature for pedestrian detection and proved that HOG based classifier has very good performance. In [7], vehicles were detected by using HOG and Gabor features and using SVM and neural network for classification. In [8], HOG feature has been used to detect vehicles and determine vehicle pose. In [9], implemented feature extraction on GPU to speed up HOG based detection algorithm.

Vehicles usually have a variety of poses in image due to orientation changes relative to the ego vehicle. Usually researchers build detector for each dominant viewpoint to adapt the large variability of appearance in different viewpoints [10], [12], [13]. In [12], Zhang used various features including location, color and texture to detect multi-view vehicles in conditional random field. Combine with several geometric constrains of vehicle and road, velocity and pose of other vehicles on the road were obtained in [26]. HOG features have been widely used to distinguishing vehicles orientations. In [13] and [14], HOG feature based SVM classifiers were trained for several orientations. In [10], part-based models with based on HOG features with deformable part position configurations were trained for different orientations. In [14], detection and pose estimation were jointly learned in a multiplicative kernel functions using HOG and SVM.

### B. Part-Based Approaches

Recent researches have been observed of increasing interests on part-based detection [10], [21], [22], where an object is modeled as a combination of parts under geometric constraints. Part-based vehicle detection has been studied a lot in recent years [10], [11], [15]. Detecting side view vehicles by parts has been studied in [15], which presented method to find vehicles in adjacent lane by detecting two wheels. Combine with SURF and edge features, vehicle parts were identified as key points for detection in [24]. In [28], proposed a side view vehicle detection method using spatially constrained part detectors to locate a set of car parts. In [18] and [29], front and rear parts were used to detect vehicles with structural constrains.

The deformable part model (DPM) [10] and the subsequent researches on on-road vehicle detection [17], [23] are among the most representative recent works. DPM was proposed in [10] which implemented multi-view part models using latent Support Vector Machine (SVM), achieved robust and effective detection results, and was used for on-road vehicle detection in [17], [23]. In [16], a cascade classifier using DPM improved detection speed and efficiency. Since vehicles are detected as a deformable configuration of parts, the methods have been demonstrated of more adaptiveness to varying viewpoints [10], [24]. On the other hand, since parts are detected to generate instances, the methods should have more advantages on finding partially occluded objects, while such a potential has not been fully demonstrated in literature for on-road vehicle detection. Such as in DPM work [10], a root filter along with part filters were used to find object, evaluated detections by considering all parts with their detection scores and position deformable cost, detection effect descends evidently for occluded objects.

Occluded vehicle detection using part-based methods has also been addressed in literature. Tehrani [17] combined DPM with Conditional Random Field (CRF) to detect occlusion, but their assumption of occlusions in CRF model is limited in horizontal direction, and experiments on artificially occlusion samples cannot strongly prove effectiveness of their work. Sivaraman [18], [23] introduced a framework using independent parts to detect oncoming, preceding, side view and partially occluded vehicles by active learning algorithm in urban scene, they only consider vehicle's viewpoint in front, rear and side view classes, and only front and rear parts are considered in this work, which is too rough for accurate visible parts detection and occluded parts estimation. In [27], proposed a method to detect both clean and partially occluded front vehicles in static images using SIFT features and hidden CRF, a probabilistic graph model was proposed to represent part spatial structural configurations. However only front vehicles were considered in this paper, the problem of multi-view was not addressed.

Recent researches have been observed of a new tendency on modeling the frequent occluded patterns into subcategories, and making use of them in occluded vehicle detection [33]–[36]. However a trade-off have to be concerned between the number of subcategories and the computation efficiency, and as the number of subcategories increased, the problems of limited training samples are not trivial. For example, although the performance of [36] has been ranked to the top on the KITTI vision benchmark suite [37], more than one hundred models are used, which brings big problems to model training and online computation efficiency in real-world applications.

Prior works have demonstrated that part-based methods have good performance on detecting multi-view vehicles or occluded cases, while few of them address both challenges together. This work proposes a visual-based on-road vehicle detection algorithm for multi-lane traffic scene that address both multi-view and occlusion problem in one probabilistic inference framework.

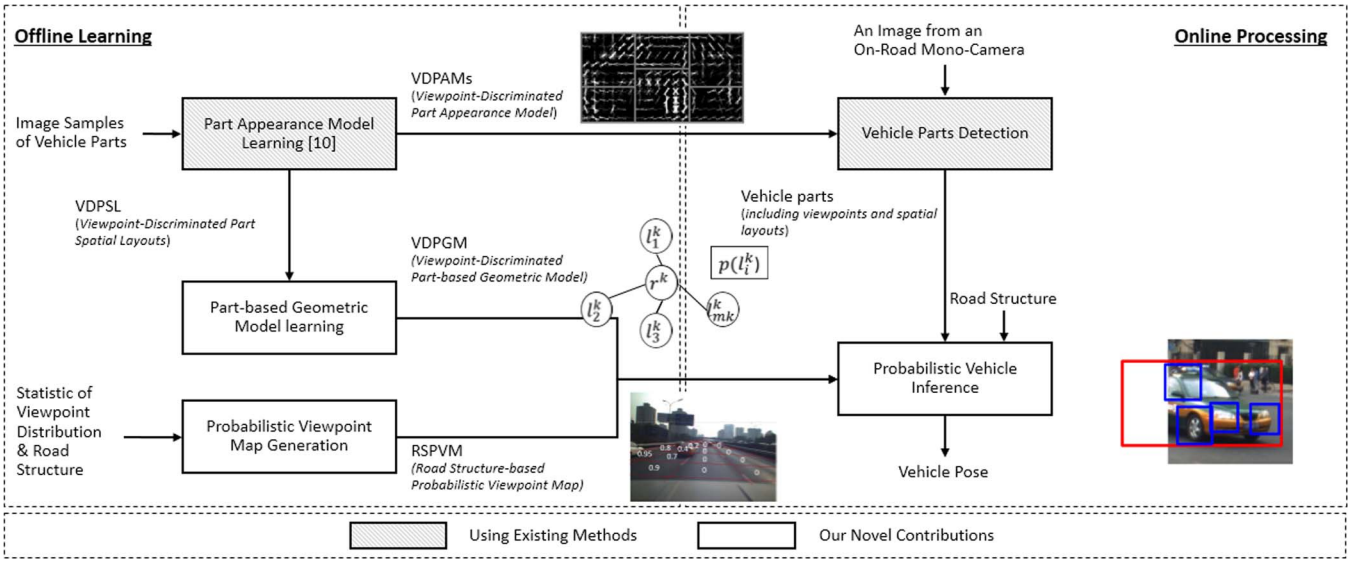



Fig. 2. A probabilistic framework of on-road vehicle detection with part model learning and probabilistic vehicle pose inference.

## III. Algorithm

### A. Outline

In this research, we propose a part-based vehicle detection method with its focus on a probabilistic inference by addressing the issues of partial observation and varying viewpoints in one framework. The system is depicted in Fig. 2, which consists of two parts, i.e., offline learning and online processing. Vehicles are modeled on their major parts, i.e., part-based vehicle models, 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) in this paper respectively. Model learning as well as viewpoint map generation are conducted offline, while in online detection, with the results of vehicle parts detection by exploiting an existing approach [10], a probabilistic inference is conducted based on the geometric models of vehicle parts and viewpoint maps, which are the focus of this research.

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, geometric models are learned for each dominant viewpoint such as front, left-front, right-front views, using part detection results, thus are called VDPGM. On the other hand, viewpoints of potential vehicles could be predicted for any on-road location at an ego or world frame, if traffic direction or lane structure of the present road is known. This is a reasonable assumption, as a road map is nowadays generally available and visual lane detectors have been integrated in many commercial ADAS systems. Viewpoint maps provide probabilistic description to the distribution of viewpoints that could improve the efficiency of vehicle inference, thus are called road structure-based probabilistic viewpoint map (RSPVM).

Below we present the method of probabilistic vehicle inference, part-based vehicle model that consists of the definition and learning of both VDPAM and VDPGM, and the generation of road structure-based probabilistic viewpoint maps.

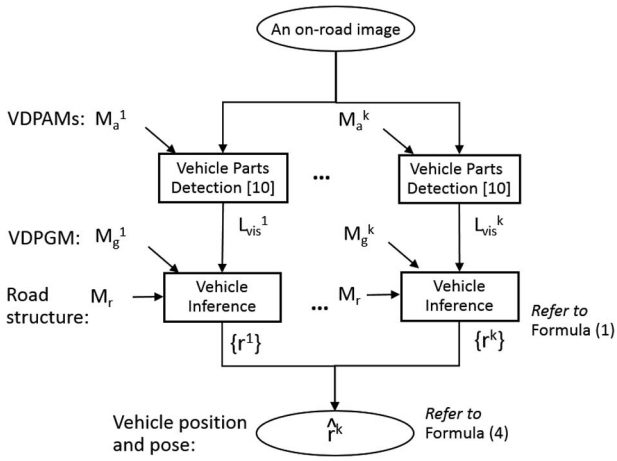



Fig. 3. Online procedure of vehicle inference based on part detections.

### B. Probabilistic Vehicle Inference

Given an image, the vehicle inference procedure in online processing is depicted in Fig. 3. For each dominant viewpoint, $M_a^k$, VDPAM, is learned in offline procedure (refer to Section IV for details). Given the current frame of an on-road image, a set of vehicle parts $L^k = \{l_i^k, \phi(l_i^k)\}$ is detected on each viewpoint $k$, where $l_i^k$ is a part instance recording the corresponding indexes of part and viewpoint, as well as its location on image, $\phi(l_i^k)$ is the detection score. In this research, $k = 1, \ldots, 4$ are defined as the front, left-front, right-front, and side views. Given the sets of detected vehicle parts $L_{vis}^k$ on all dominant viewpoints, the problem of vehicle inference can be represented in a probabilistic way as below.

$$\hat{r^k} = \arg\max_{r^k} p\left(r^k | L^k, M_g^k, M_r\right)_{k=1,\ldots,4} \tag{1}$$

where, $M_r$ is a rough knowledge of road structure (see Section V for details), which could be retrieved from a map or obtained using a visual lane detector. $M_g^k$ is the part-based geometric model (VDPGM) of viewpoint $k$ (see Section IV for details),

which records parts configuration of the specified viewpoint as well as their spatial layouts. $r^k$ 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 on Bayesian rules as below.

$$p\left(r^k|L^k, M^k_g, M_r\right) \propto p\left(L^k|r^k, M^k_g, M_r\right) p\left(r^k|M^k_g, M_r\right) \quad (2)$$

where, $p(r^k|M^k_g, M_r)$ can be further simplified to $p(r^k|M_r)$, which is the probability of a vehicle at location $r$ on viewpoint $k$, given the knowledge of road structure. $p(L^k|r^k, M^k_g, M_r)$ can be further simplified to $p(L^k|r^k, M^k_g)$, meaning for the probability that the set of part instances $L^k_{vis}$ is observed, given a model $M^k_g$ of the part configuration and spatial layouts, and that a vehicle is at location $r$ on viewpoint $k$.

Concerning that $L^k_{vis}$ might be a subset of the full configuration of $M^k_g$, as only visible parts can be detected under partial occlusions, let $L^k_{occ} = \neg L^k_{vis} = \{l^k_j\}$ be the rest set of the full configuration, where $l^k_j \in M^k_g \wedge l^k_j \notin L^k_{vis}$ meaning for the occluded parts, the estimation of $p(L^k|r^k, M^k_g)$ is converted to as below.

$$\begin{aligned} p\left(L^k|r^k, M^k_g\right) &= p\left(L^k_{vis}, L^k_{occ}|r^k, M^k_g\right) \\ &= p\left(L^k_{vis}|r^k, M^k_g\right) p\left(L^k_{occ}|r^k, M^k_g\right). \end{aligned} \quad (3)$$

In summarizing the above derivations, the $p(r^k|L^k_{vis}, M^k_g, M_r)$ in formula (1) can be estimated as in formula (4).

$$\begin{aligned} &p\left(r^k|L^k, M^k_g, M_r\right) \\ &\quad \propto p(r^k|M_r) p\left(L^k_{vis}|r^k, M^k_g\right) p\left(L^k_{occ}|r^k, M^k_g\right) \\ &\quad = p(r^k|M_r) \prod_{l^k_i \in L^k_{vis}} \phi\left(l^k_i\right) p\left(l^k_i|r^k, M^k_g\right) \prod_{l^k_j \in L^k_{occ}} \phi_{occ}. \end{aligned} \quad (4)$$

A set of $r^k$ is sampled, from which, the $r^k$ that maximize the posterior $p(r^k|L^k, M^k_g, M_r)$ is selected as the vehicle pose. In this research, sampling of $r^k$s is conducted as below. Giving $l^k_i$, which has part detector response $\phi(l^k_i)$ greater than $\phi_m in$, 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 $\dot{r}^k$. In this research, a sample of $r^k$ is generated on a subset of the visible parts ($L' \subseteq L$), with the missed parts of the subset as the occluded ones. It could happen that a $r^k$ on fewer visible parts has higher posterior than those on more. This property ensures robustness to noisy/partially incorrect part detection results.

Below, we define the part-based vehicle models including VDPAM and VDPGM, address their learning procedures, and give an analytical estimation to the $p(l^k|r^k, M^k_g)$, $\phi(l^k_i)$ and $L^k_{occ}$ of formula (4) in Section III-C; we define the primary types of road structures, address the method of generating RSPVM, and give an analytical estimation to the $p(r^k|M_r)$ of formula (4) in Section III-D.

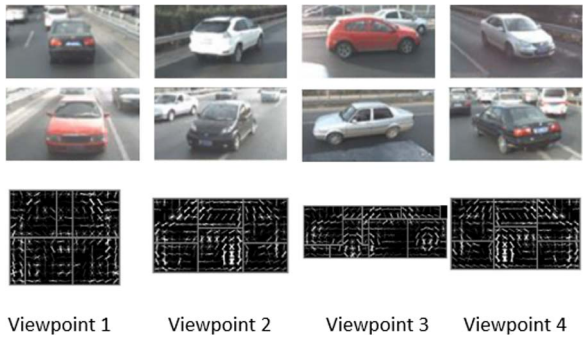



Fig. 4. A result of learned VDPAMs of different viewpoints.

### C. Part-Based Vehicle Model

Part-based vehicle is modeled on each discriminative viewpoint $k$ on both appearance and geometry. The existing approach [10] is used in this research in training the viewpoint-discriminative part appearance model $M^k_a$ (VDPAM), and provide sample data of part configuration and spatial layout in training the viewpoint-discriminative part-based geometric model $M^k_g$ (VDPGM). Given a set of image samples with bounding boxes on the subject vehicles, which covers all possible views, the approach in [10] discriminates viewpoints of the vehicle samples in an unsupervised manner, generates HOG-based appearance models on the most significant vehicle parts of a specific viewpoint $M^k_a = \{M^k_{a_i}|i = 1, \ldots, n\}$, where $n$ is a predefined constant number of parts to be modeled, and outputs relative locations as well as viewpoints of all detected vehicle parts $\{l^k_i|i = 1, \ldots, n^k; k = 1, \ldots, \kappa\}$. Although in the implementation of this research, $n^k$ is the same across different viewpoints, we would like to formulate it as a variable that addresses a viewpoint discriminative property, the implementation of which can be extended in future work.

Given a data set $\{l^k_i|i = 1, \ldots, n\}$ on viewpoint $k = 1, \ldots, \kappa$, a geometric model $M^k_g$ is trained, which can be formulated as below.

$$M^k_g = \left\{\zeta^k_i|i = 1, \ldots, n^k, \zeta^k_i \sim N\left(\mu^k_i, \Sigma^k_i\right)\right\} \quad (5)$$

where, the geometric model $M^k_g$ consists of $n^k$ parts. For each part $\zeta^k_i$, its relative location with respect to a vehicle center $r^k$ is described in a 2D Gaussian with its mean at $\mu^k_i$ and covariance $\Sigma^k_i$, which are valued using the statistics of the training data $\{l^k_i|i = 0, \ldots, n^k\}$.

In this research, four dominant viewpoints are defined, corresponding to the front, front-left, front-right and side views. Some image samples and learned appearance models on each viewpoint are demonstrated in Fig. 4. On the other hand, examples of geometric model on viewpoint 1 and 2, as well as their learning details are depicted in Fig. 5. It can be found that the spatial distribution of the sample set of each part $\{l^k_i\}$ can be reasonably fitted using a 2D Gaussian, while its mean and variance varies across different parts and viewpoints.

Given a geometric model $M^k_g$ and a vehicle pose $r^k$, for any vehicle parts $l^k_i \in L^k_{vis}$, the likelihood that part $i$ be measured on viewpoint $k$ at location $l^k_i$ is estimated as below.

$$p\left(l^k_i|r^k, M^k_g\right) = N\left(l^k_i - r^k; \mu^k_i, \Sigma^k_i\right). \quad (6)$$

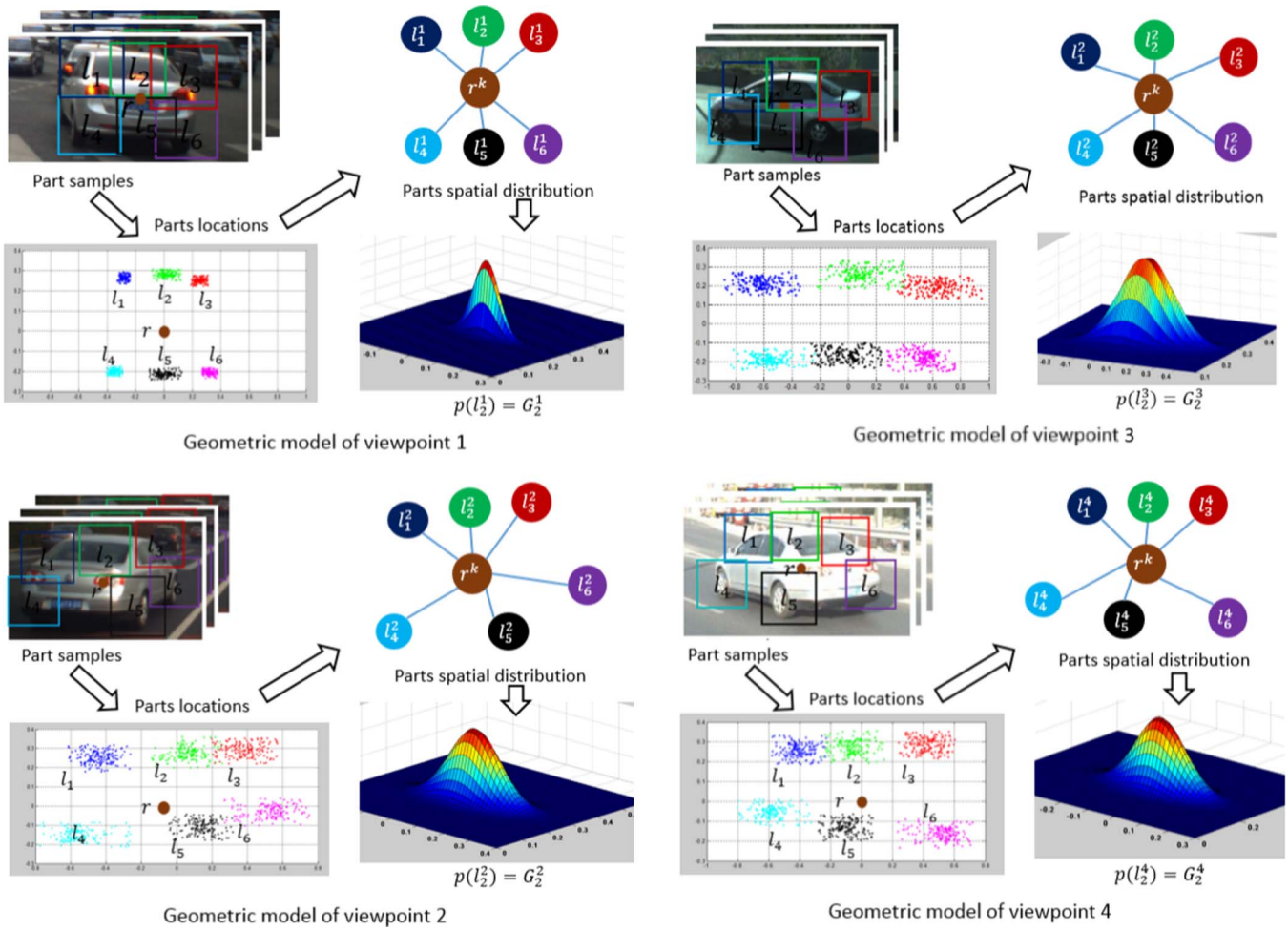



Fig. 5. A result of learned VDPGMs on different viewpoints.

On the other hand, for any vehicle parts $l_j^k \in L_{occ}^k$, as no observation of the part is obtained, it is hard to define an explicit estimation to $\phi(l_j^k)p(l_j^k|r^k, M_g^k)$. In this research, it is treated as a punish factor, which is assigned a constant value $\phi_{occ} = \eta\phi_{\min}$, which is the detector threshold $\phi_{\min}$ multiply by a punishing item $\eta(\eta = 0.7)$, to simulate the occluded part and filter the low quality detected part instance.

### *D. Road Structure-Based Probabilistic Viewpoint Map*

Normally, on-road vehicle motion follows certain rules, which has strong correlation with road structure that is defined by road geometry, lanes, and traffic rules. Regular vehicle motion patterns at a certain road structure can be simulated on the traffic rules or a prior knowledge implied by road geometry and lane, and they can be modeled with more accuracy using real-world measurements too. As a road map is generally available and visual lane detectors have been integrated in many commercial ADAS systems, it is reasonable to assume that the type of the present road structure can be an online knowledge, and is used in this work as an online input. Given a type of road structure, the viewpoint of a subject vehicle at any specific location can be predicted by registering regular vehicle motion patterns, e.g., trajectories, to the ego vehicle frame, and subsequently projected onto image frame using the calibration parameters between ego vehicle and on-board camera. However such predictions are not exclusive nor accurate enough as an individual vehicle motion may vary from the regular patterns; road parameters such as lane width differ slightly at different road sections, which are not addressed in nowadays map; and furthermore, the model of regular vehicle motion patterns is always lack of accuracy, the procedure in registering it to the ego vehicle frame using a commercial GPS could yield large error too. A probabilistic representation is thus needed to address the predictions of a subject vehicle's viewpoint.

A viewpoint prediction map as shown in Fig. 6 is proposed in this research. For each viewpoint $k$, $k = 1, \ldots, 4$ in this research, a grid map at ego vehicle frame is generated in an offline learning procedure (see Fig. 6 world grid view), for each grid values are generated to represent the probability of a vehicle to be observed on viewpoint $k$ at the grid's area. Thus, an estimation to $p(r^k|M_r)$ in formula (4) can be achieved by retrieving the corresponding grid values after projecting the grid map onto the image frame of an on-board camera (see Fig. 6 right). In this research, grid maps are generated at a pixel size of 2 meter, and grid values are assigned by taking the statistics of the data from a Lidar-based vehicle detection system [31] (see next section for details).

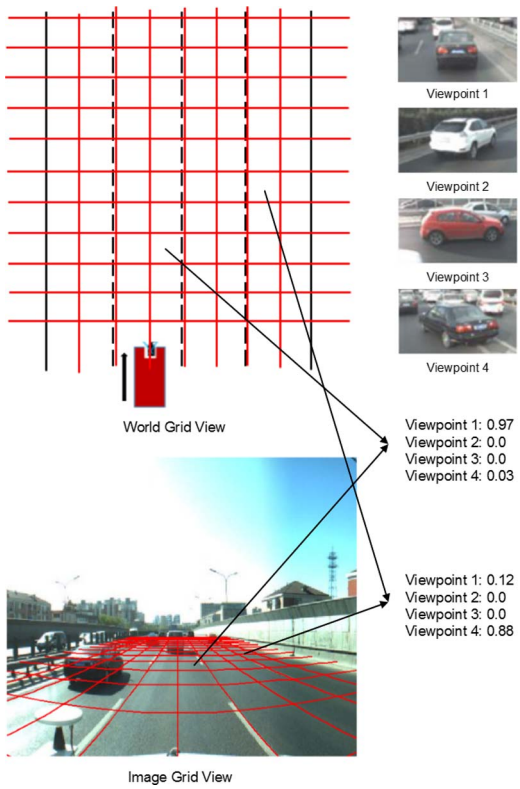



Fig. 6. A result of learned RSPVMs of the straight road. Numbers in each grid represent the probability of vehicle's viewpoint.

### E. Implementation Details

The methods provided by [10] are used in training visual appearance models. 500 positive samples, each consists of a vehicle image $I$ and a bounding box of the subject vehicle, are picked up equally from all categories covering all possible viewpoints. Negative samples are generated from the 100 images without vehicle instance. According to [30], the vehicle images on front and rear viewpoints are similar in their visual feature space, same observations are gained on the data of other three pairs of viewpoints, i.e., front-right with rear-left, front-left with rear-right, right with left. So that, in this research, we merge the training samples in the original eight categories to four, and models are learned on four viewpoints accordingly, i.e., $\kappa = 4$. For each viewpoint $k$, a root filter $F_0^k$ for the whole vehicle body is learned using the image samples of the category. The most dominant regions are then extracted from the root $F_0^k$ to generate part filters $F_i^k, i = 1, \ldots, n$, i.e., the viewpoint discriminative part appearance models (VDPAM) which are used later in vehicle part detection. Here $n$ is a pre-assigned constant value, and $n = 6$ in this research.

For each image sample $j$ in the category of viewpoint $k$, each part filters $F_i^k, i = 1, \ldots, n$ is applied to detect the dominant part instance, which is denoted by $l_{j_i}^k$, meaning that a vehicle part $i$ on viewpoint $k$ is detected at the $j$th sample at relative image location $l$ with respect to the vehicle center. Given the set of detected parts on all image samples $\{\{l_j\}_i^k | i = 1, \ldots, n^k\}$, the $\{l_j\}_i^k$ are fitted on a 2D Gaussian distribution, a 2D mean location $\mu_i^k$ and a covariance matrix $\Sigma_i^k$ is estimated for each part $i$ and viewpoint $k$, composing the viewpoint-discriminative part-based geometric model (VDPGM) $M_g^k$ in formula (5).

Four viewpoint maps are generated at ego-vehicle frame with a pixel size of 2 meter. As described previously, the Lidar-based detection and tracking results (i.e., trajectories) are used to take the statistics of viewpoints at each grid location of the viewpoint maps. For each grid location, the vehicle trajectories that crossing the grid cell are used to estimate votes for viewpoints. The votes at each grid location are then normalized across $k = 1, \ldots, 4$ viewpoint maps. The pixel value $M_r^k(i)$ meaning for a probabilistic prediction of a vehicle be observed at the grid location $i$ on viewpoint $k$.

### F. Computational Complexity Analysis

Comparing with the original DPM[10] that uses a root and part filters, the proposed approach makes inference based on part filters only. As formulated in (4), $\phi(l_i^k)$ is the score of detecting part $l_i^k$. It is the most time-consuming portion of this method, while could be accelerated through GPU as demonstrated in [9]. On the other hand, the $M_g^k$ and $M_r$ are generated by taking statistics in an off-line procedure, and online estimation of $p(l_i^k | r^k, M_g^k)$ and $p(r^k | M_r)$ can both be done in $O(1)$ time complexity. Thus we can summarize that the proposed approach has the time complexity at the same level with that of the original DPM.

## IV. Experimental Study

### A. The Datasets for Training and Testing

Experiments have been conducted on the ring roads of Beijing, which are bi-directional motorways with four- or five-lane's on each side that has no signalized intersections. In order to examine the performance of vehicle detection at different viewpoints, an omni-vision system, the Ladybug3 by the Point Grey Research Inc., is mounted on the roof of a vehicle platform, and omni-images as shown in Fig. 8(a) that are captured through the driving experiments at different days are used in developing the datasets for training and testing.

In this research, the image samples that were developed in the authors' previous work [32] in an automatic manner are used in the training of multi-view detectors, which were developed originally from the omni-images after distortion removal and viewpoint discrimination, and consist of the image clips of un-occluded vehicles in eight sub-categories that correspond to each view direction (see Fig. 7).

Testing data are developed on the sequence of omni-images, which were captured while driving a total distance of 65 km about 120 min, however additional consideration is needed on correcting their distortions. In order to make use of the data to study the performance of vehicle detection on normal cameras, an omni-image is divided equally into five pieces as shown in Fig. 8(f), and distortions are corrected to simulating the images that are captured from a mono-vision camera to the direction. Ground truth of the vehicles are labeled through manual operation. As listed in Table I, five thousand frames of omni-images (i.e., 20 thousand mono-vision images) are labeled at a rate of 1 Hz, consisting of 16 thousand fully observed vehicles and 15 thousand partially observed ones. As shown in Fig. 8(f), a fully observed vehicle is marked using a red bounding box, while a partially observed one is bounded in twofold, i.e., one


Fig. 7. Training samples.

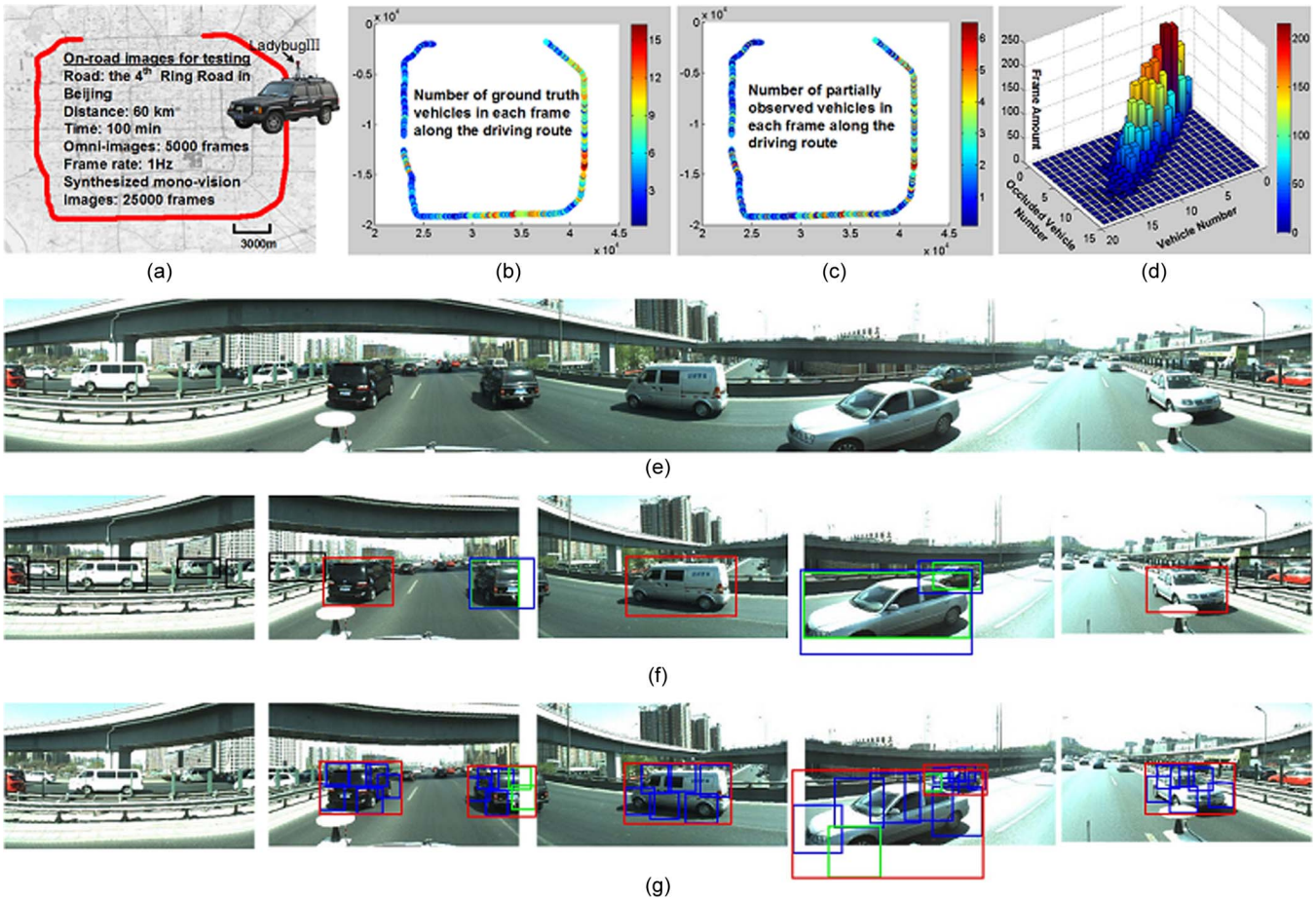



Fig. 8. Experiment data set details. (a) Experiment environment. (b) Number of ground truth vehicles in each frame along the driving route. (c) Number of partially observed vehicles in each frame along the driving route. (d) Statistic of vehicle number. (e) Omni image. (f) Ground truth bounding boxes on processed images. (g) Detection results.

for observed part and one for a guessed full body, which are drawn in Fig. 8(f) in green in blue respectively. The corresponding detection results are also shown at Fig. 8(g) 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. As for partially observed vehicles, sub-category labels are assigned according

TABLE I
TEST DATASET DETAILS

| | | Fully Observed | Partially Observed | | |
|---|---|---|---|---|---|
| | | | By Road Infrastructure | By Other Vehicles | Camera Sight Limitation |
| 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 | | |

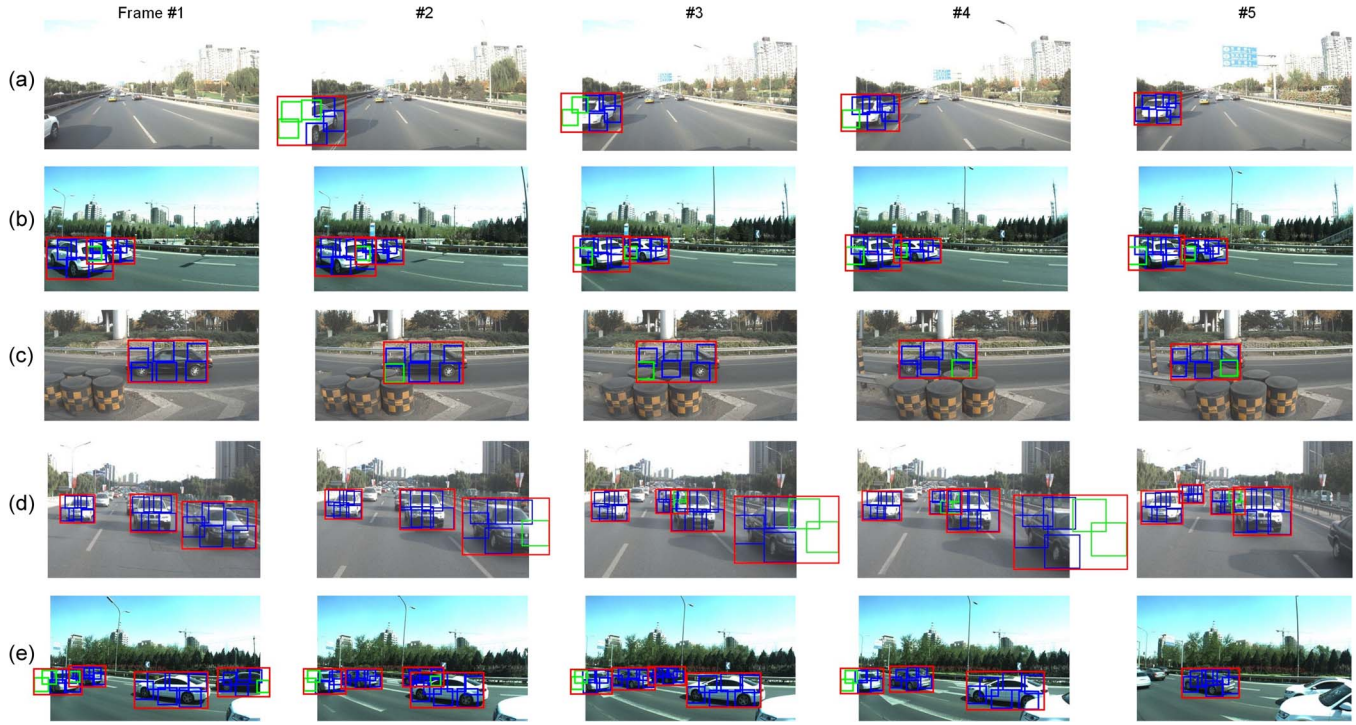



Fig. 9. Multi-view vehicle detection sequences. (a) Front view of low density. (b) Front-side view of low density. (c) Side view of low density. (d) Rear view of high density. (e) Side view of high density.

to the reason of occlusion, i.e., due to road infrastructure, other vehicles or limited camera's vision field, which have 146, 3851, and 11 470 images samples respectively. These data are used in further examine the performance on different kinds of occlusions. Note that the 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. However, as the judgments are taken based on the size of bounding box, which were drawn by more than ten volunteers, there exists differences and inaccuracies in manual operations. In addition, as the experimental sites are bi-directional motorways, the vehicles on the opposite roadway are sometimes observable with occlusion from road barriers. The detection performance on these vehicles are not studied in this research, which are marked using black bounding boxes as shown in Fig. 8(f).

The testing data covers a broad range of traffic condition, from smooth to dense traffic. Fig. 8(b) shows 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 detection performance at different traffic density is also studied.

Below we first study some cases at typical traffic situations to examine the advantage and major challenges of the proposed algorithm, then analyze the performance by taking statistics on the whole testing data sets. Part-based examination to find the significance of different parts in vehicle detection is presented too.

### B. Vehicle Detection Results

Fig. 9 shows sequences of detection results on the mono-vision images on each view direction. For the front, front-side and rear views, each has one sequence to present results, while as side view images may suffer more from occlusion, two sequences are given. Seq.1 is a front view one, 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 (a), the white car had only its front observed, which was too small a part to pass the threshold of validation, so that was failed in detection. Later at (b), 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

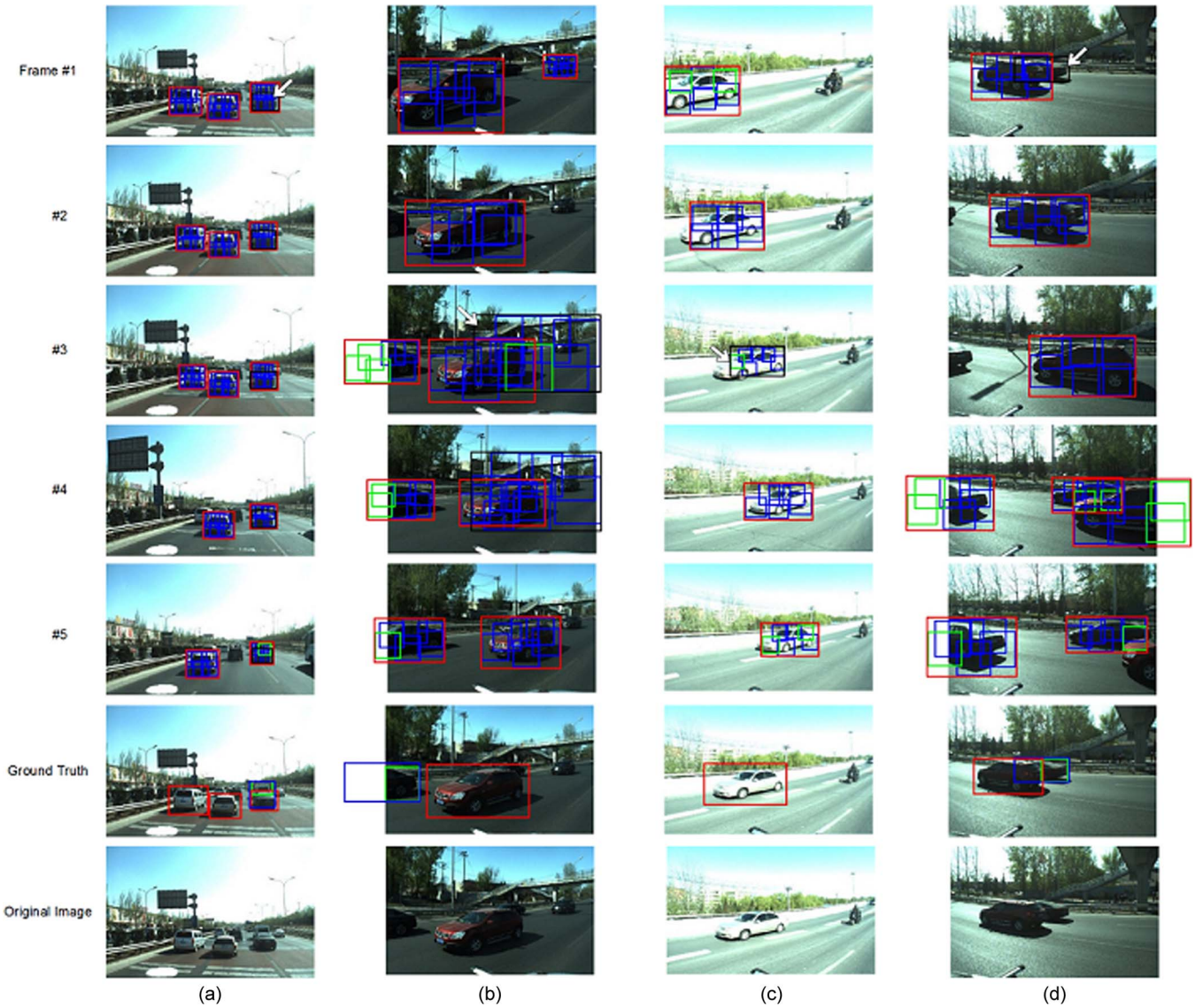



Fig. 10. False detections. (a) Miss detection because of heavily occluded. (b) False detection because of similar features. (c) Miss detection because of an incorrect bounding box. (d) Another typical miss detection because of heavily occluded.

body, while green for the inference of occluded parts. Detection was succeeded 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. 9 can be discussed similarly, where seq.2 is the results while a car on a farther lane was occluded by a front one; seq.3 is that a black car was occluded by road barriers; seq.4 is a rear view one, where front vehicles blocked greatly the vision to farther ones; seq.5 is a scene of dense traffic, and occlusion is even more severe on side view images. An interesting point is that in seq.5(h), a white car on the right-bottom corner of the image was failed in detection, which had even larger area be observed than that in seq.1(b), but the results are contrary. This prompt us that parts of a vehicle may have different significance in detection, which will be studied later.

The detection results are compared with the labeled ground truth. For those labeled as partial observed ones, the bounding box of a whole vehicle is used. For a detection result, if a ground truth is found that has a ratio of overlapped area more than 0.7 in between of the two bounding boxes, it is recognized as a *true positive*, otherwise a *false positive* one. On the other hands, if a ground truth is failed in obtaining a matched detection result, a miss detection (i.e., *false positive*) is counted. The negative results and challenging situations have some typical patterns that are depicted in Fig. 10.

Each sequence in Fig. 10 has an original image that give a clear view to the scene and one that shows the ground truth at the last. The bounding boxes in black are those of either false of miss detections. Seq.a is a front view scene, where a sedan drove next to a van, and two bounding boxes were labeled 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

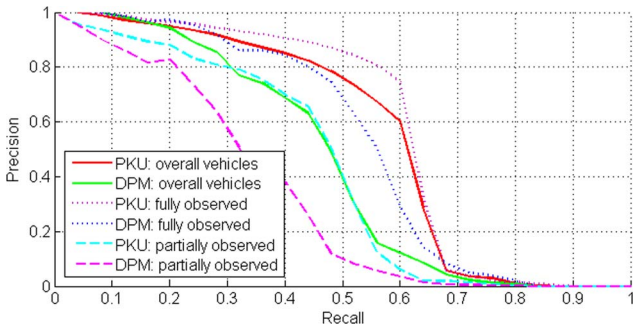



Fig. 11. Detection results compared with DPM.

box is the ground truth one that has no matching in detection results. The miss detection in seq.d #1 depicts also a typical case, where the vision to a farther car was heavily blocked by a front one. A ground truth was labeled, however, reliable detection on such heavily occluded vehicle is still a challenge of the proposed algorithm. There are also false alarms as shown in seq.b, which happened on a overhead bridge that has similar feature with a side view vehicle. Seq.c is a scene that a white car has a background of white road barrier. Part detection was hard especially on the front end of the car. Although the vehicle was detected at the final results, the estimated bounding box was not as correct, and a false detection was counted at (c) as the ratio of overlapped area with the ground truth one was less than the predefined threshold.

### C. Quantitative and Comparative Evaluation

The data set with the labeled ground truth as described in Table I is used in quantitative evaluations of detection accuracy for both fully and partially observed vehicles, where the results are compared with the labeled ground truth by following the classical object detection protocol of Pascal VOC [3]. Evaluation are conducted at three levels, i.e., comparison with the results of DPM method [10], the results on detecting different kinds of occluded vehicles and the system performance at different traffic density.

Precision/recall curves of the detection results on the overall testing data set are plotted in Fig. 11(a), which compares performance of the proposed and the DPM methods by following the classical object detection protocol of Pascal VOC [3]. Here, the precision and recall are defined below,

$$\text{Precision}_{total} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall}_{total} = \frac{TP}{GT} \tag{8}$$

where $TP, FP$ are the numbers of *true* and *false positive* detections, and $GT$ is the number of labeled ground truth vehicles.

On the other hand, as each ground truth vehicle has a label of being fully or partially observed one, the set of *true positives* can be divided into two groups accordingly, with their numbers of $TP_{full}$ and $TP_{part}$ respectively. Subsequently, the detection performance on fully or partially observed vehicles can be evaluated by using the set of ground truths of the corresponding tag, and precision/recall (P/R) curves are plotted in Fig. 11(b) and (c) with the following definition.

TABLE II
AVERAGE PRECISION OF DETECTION RESULTS ON DIFFERENT METHODS

| AP: Method / Status | DPM | Our method |
|---|---|---|
| fully observed | 54.6 % | 61.4 % |
| partially occluded | 39.6 % | 48.1 % |

$$\text{Precision}_{full} = \frac{TP_{full}}{TP_{full} + FP} \tag{9}$$

$$\text{Recall}_{full} = \frac{TP_{full}}{GT_{full}} \tag{10}$$

$$\text{Precision}_{part} = \frac{TP_{part}}{TP_{part} + FP} \tag{11}$$

$$\text{Recall}_{part} = \frac{TP_{part}}{GT_{part}}. \tag{12}$$

It can be found in Fig. 11(a) that the proposed system performs better than DPM on the detection results of overall vehicles. We study the reason in Fig. 11(b) and (c). The proposed method and DPM perform better on detecting fully observed vehicles, which have average precision (AP) of 54.6% and 61.4% respectively for ours is more flexible on multi-view problem, as shown in Table II. And the proposed one performs much better on partially observed vehicles with average precision of 39.6% AP comparing to 48.1% by DPM. The result demonstrates importance of part-based inference in occluded scene.

Performance on detecting the different kinds of partially observed vehicles are also evaluated. For each ground truth that labeled as a partially observed vehicle, a sub-category label is assigned denoting the reason of occlusion. Accordingly, the true positives of partially observed ones are divided into three subsets for those occluded by road infrastructure, other vehicles or limited vision field. We denote $TP^{(i)}_{part}$ for the numbers of the true positives of the $i$th subset, where $TP_{part} = \sum_{n=3}^{i} TP^{(i)}_{part}$.

$$\text{Precision}^{(i)}_{part} = \frac{TP^{(i)}_{part}}{TP^{(i)}_{part} + FP} \tag{13}$$

$$\text{Recall}^{(i)}_{part} = \frac{TP^{(i)}_{part}}{GT^{(i)}_{part}}. \tag{14}$$

However in the testing data set as shown in Table I, there are few samples that are occluded by road infrastructures, we compare the performance on the other two types of occlusions due to surrounding vehicles or limited vision field as shown in Fig. 12 and Table II. Detection of the vehicles partially occluded due to limited vision field $(\text{ap} = 46.5\%)$ has better performance on those due to surround vehicles $(\text{ap} = 43.0\%)$, 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. Especially when occlusion is caused

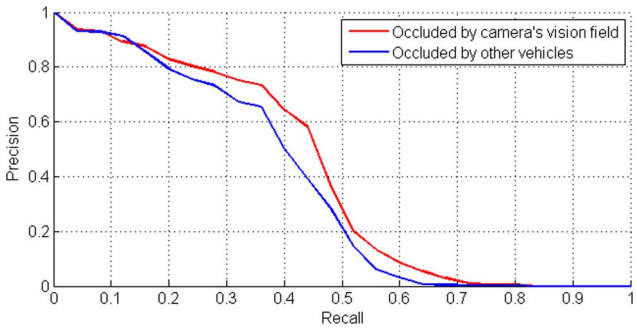



Fig. 12. Detection results on different occluded vehicles.

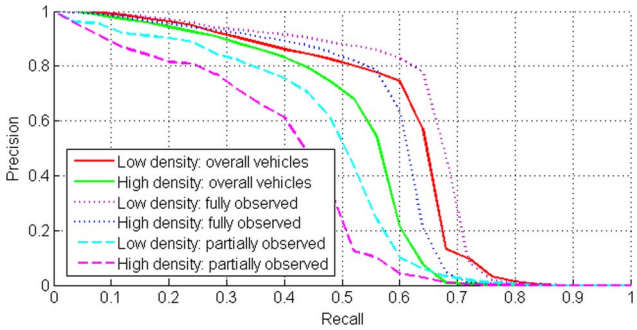



Fig. 13. Detection results under different traffic density.

by a vehicle with similar color as shown in Figs. 9 and 10, detection is extremely difficult, which has not been solved in the proposed method.

Detection performance at different traffic conditions are also studied, where traffic density is concerned as one of the most crucial factors in evaluation. Fig. 8(b) visualizes the number of ground truth vehicles at each omni-image frame along the driving route, which equals to the number of vehicles at the ego's nearby neighbor, and reflects traffic density of the scene. Color represents vehicle number per frame (denoted by $TVN$) with blue for 0 and red for 15. On the other hand, within the ground truth vehicles, the number of partially observed ones at each omni-image frame (denoted by $PVN$) are also counted as visualized in Fig. 8(c) with blue for 0 and red for 6, which reflects severeness of occlusion at the scene. Fig. 8(d) correlates $TVN$ and $PVN$, which shows that $PVN$ increases nearly linearly along with $TVN$, i.e., higher traffic density, more severe occlusion, and more challenges to detection algorithm.

We divide the testing data into two groups for low and high traffic density. The group of low contains the image frames with $TVN < 10$, while others are contained in the high one. Detection performance in low and high traffic density are compared as shown in Fig. 13 and Table III, which are estimated in the same way with those defined in formula (1)–(6). The performance are compared on the sets of overall vehicles, fully and partially observed ones respectively. It can be found that for fully observed vehicles, no matter they are in the scene of low or high traffic density, the detection performance are at similar level. However for partially observed ones, the performance at low traffic density is much better than that at high one. It can be explained that comparing with the scene of low traffic density, there are more occluded cases that are caused by surround vehicles, which are more difficult in detection comparing to other types of occlusions as discussed previously.

TABLE III
AVERAGE PRECISION OF OUR DETECTION RESULTS ON DIFFERENT CONDITIONS

| AP: Density / Status | Traffic Density | |
|---|---|---|
| | Low | High |
| fully observed | 65.8 % | 60.5 % |
| partially occluded | 50.5 % | 44.9 % |

### D. Performance Comparison

The KITTI's Vision Benchmark Suite [37] is a widely acknowledged open resource that contains a large set of labeled vehicles on visual images. Many state-of-the-art algorithms have tested their performance on the dataset, where [33]–[36] are at the top places according to the latest ranking. As the proposed approach makes use of viewpoint maps, which are generated on the knowledge of road structure and driving patterns that are not provided by KITTI, we can not directly compare the performance on the dataset. While on the other hand, a subset of PKU dataset is built to be comparable with KITTI, and a performance test is conducted so as to compare with the other state-of-the-art algorithms. The subset of PKU dataset is generated as below.

1) Only the front and the front-right views are selected. As the front-left views (e.g., the left most one of Fig. 8(f)) face mainly to the opposite road that brings additional challenges, they are not selected.
2) As the vehicle's route was on a looped road (refer to Fig. 8(a)), some images are in extreme illumination conditions such as facing to sun lights. Such images are removed manually.
3) The labeled vehicles are divided into 3 groups, i.e., "Easy" for full observation; "Moderate" for those with an occlusion less than 30%; and "Hard" for those with more than 30% but less than 80% occluded.
4) The vehicles that have their bounding box no less than 40 pixels' height are labels, with the smaller ones marked as the background. However in the original PKU dataset, the small vehicles were neither labeled nor marked as the background, yielding the corrected detections on them be counted as the false positives.

Detection results on the subset of PKU dataset is shown in Fig. 14. Comparing with the previous graphs of Figs. 12 and 13, the figures on recall have been greatly improved, where marking the small vehicles as the background is one of the major reasons. On the other hand, the average precisions are listed in Table IV. As a comparison, the figures that are published on the KITTI's website of the top ranked methods are referred. Since the methods are tested on different datasets, the figures can not be compared directly. However, we can summarize that the proposed approach is competitive with the other state-of-the-art methods.

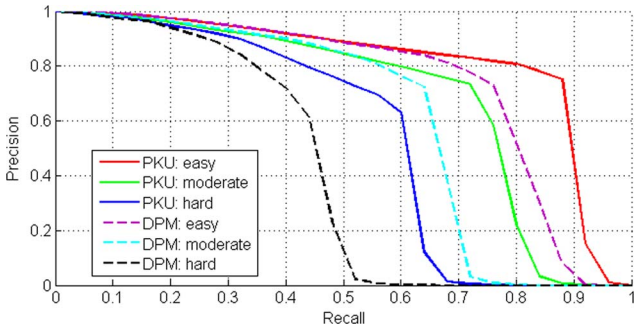



Fig. 14. Detection results on PKU subset.

TABLE IV
AVERAGE PRECISION OF DETECTION RESULTS

| | Easy | Moderate | Hard |
|---|---|---|---|
| Our method (PKU) | 80.3 % | 72.1 % | 60.3 % |
| DPM (PKU) | 74.9 % | 65.2 % | 45.1 % |
| DPM (KITTI) | 71.2 % | 62.2 % | 48.4 % |
| OC_DPM [33] (KITTI) | 75.0 % | 66.0 % | 53.9 % |
| AOG [34] (KITTI) | 84.4 % | 71.9 % | 59.3 % |
| SubCat [35] (KITTI) | 84.1 % | 75.5 % | 59.7 % |
| 3DVP [36] (KITTI) | 87.5 % | 75.8 % | 65.4 % |

### E. Discussion

Vehicle parts may have different significance in detection through the above experiment. For example comparing with other parts, a wheel gives stronger message with less uncertainty in suggesting that a vehicle may exists at the location. This inspired us to examine the difference of parts in the detection results of the vehicles at each viewpoint.

In each detection result, an index $k$ representing the vehicle's viewpoint is estimated simultaneously with a location (e.g., a bounding box). Thus both the *true positive*s and *false positive*s are divided into subsets on each viewpoint, and recurrence ratios of each part on the particular viewpoint are estimated as defined below.

$$TPr_i^k = \frac{TP_i^k}{TP^k} \tag{15}$$

$$FPr_i^k = \frac{FP_i^k}{FP^k}. \tag{16}$$

$TPr_i^k$ evaluates the recurrence ratio of part $i$ on viewpoint $k$ of *true positive*s, where $TP_i^k$ is the numbers of part $i$ at the subsets on viewpoint $k$, while $TP^k$ is the total number of detection results of the subset. $FPr_i^k$ is defined similarly evaluating the recurrence ratio of part $i$ on viewpoint $k$ of *false positive*s. In this research, we have $k = 1, \ldots, 4$ and $i = 1, \ldots, 6$.

The results are plotted in Fig. 17(a) and (b) for *true positive*s and *false positive*s respectively, which are estimated at the point when average precision is 50%. In the result of each viewpoint, the recurrence ratios of different parts are compared with the meaning of each part be illustrated in (c). The recurrence of different parts are correlated. For example at viewpoint 3 that are illustrated in Fig. 15, parts are always observed in pairs as $l_3$ and $l_6$ in (a), and $l_1$ and $l_4$ in (c). In addition, as most of

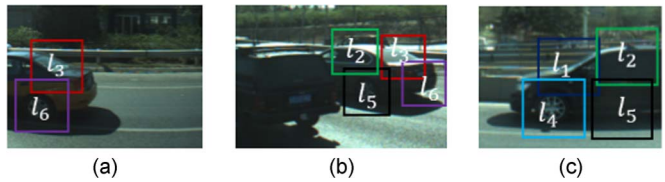



Fig. 15. Typical occlusion types.

TABLE V
WEIGHT FACTORS

| Part \ View. | | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Overall | 1 | 1.04 | 1.0 | 1.0 | 1.0 |
| | 2 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 3 | 1.06 | 1.0 | 1.0 | 1.0 |
| | 4 | 1.0 | 1.03 | 1.4 | 1.29 |
| | 5 | 1.0 | 1.45 | 1.0 | 1.5 |
| | 6 | 1.12 | 1.31 | 1.36 | 1.0 |
| Low density | 1 | 1.07 | 1.0 | 1.0 | 1.0 |
| | 2 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 3 | 1.03 | 1.01 | 1.0 | 1.0 |
| | 4 | 1.0 | 1.05 | 1.38 | 1.32 |
| | 5 | 1.0 | 1.42 | 1.0 | 1.48 |
| | 6 | 1.1 | 1.28 | 1.39 | 1.0 |
| High density | 1 | 1.04 | 1.0 | 1.0 | 1.0 |
| | 2 | 1.0 | 1.0 | 1.0 | 1.0 |
| | 3 | 1.05 | 1.0 | 1.01 | 1.0 |
| | 4 | 1.0 | 1.08 | 1.43 | 1.29 |
| | 5 | 1.0 | 1.44 | 1.0 | 1.51 |
| | 6 | 1.13 | 1.29 | 1.37 | 1.02 |

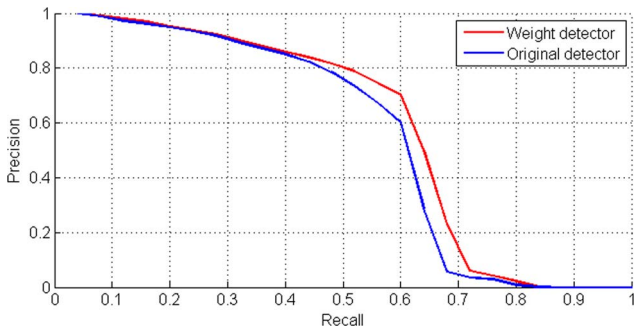



Fig. 16. Comparative results between weighted detector and original detector.

the occluded vehicles have more than 50% of their parts be observable, $l_2$ and $l_5$ in (b) have always higher recurrence ratio. Thus the recurrence ratios of *true positive*s of Fig. 17 represent parts' correlation, where their difference at each viewpoint are caused due to the major patterns in occlusions.

On the other hand, the recurrence ratios of *false positive*s as shown in Fig. 17 reflect to some extent the significance of each parts in vehicle detection. For example, it can be found that the parts corresponding to a wheel have lower recurrence ratios, such as $l_5$ at viewpoint 2, $l_4$ and $l_6$ at viewpoint 3, and $l_4$ and $l_5$ at viewpoint 4, which means that *false positive*s happen always in the image clips where vehicle wheels are not observable, and a wheel is a more unique feature comparing with other vehicle parts, which provides more certain a message and is harder to be confused.

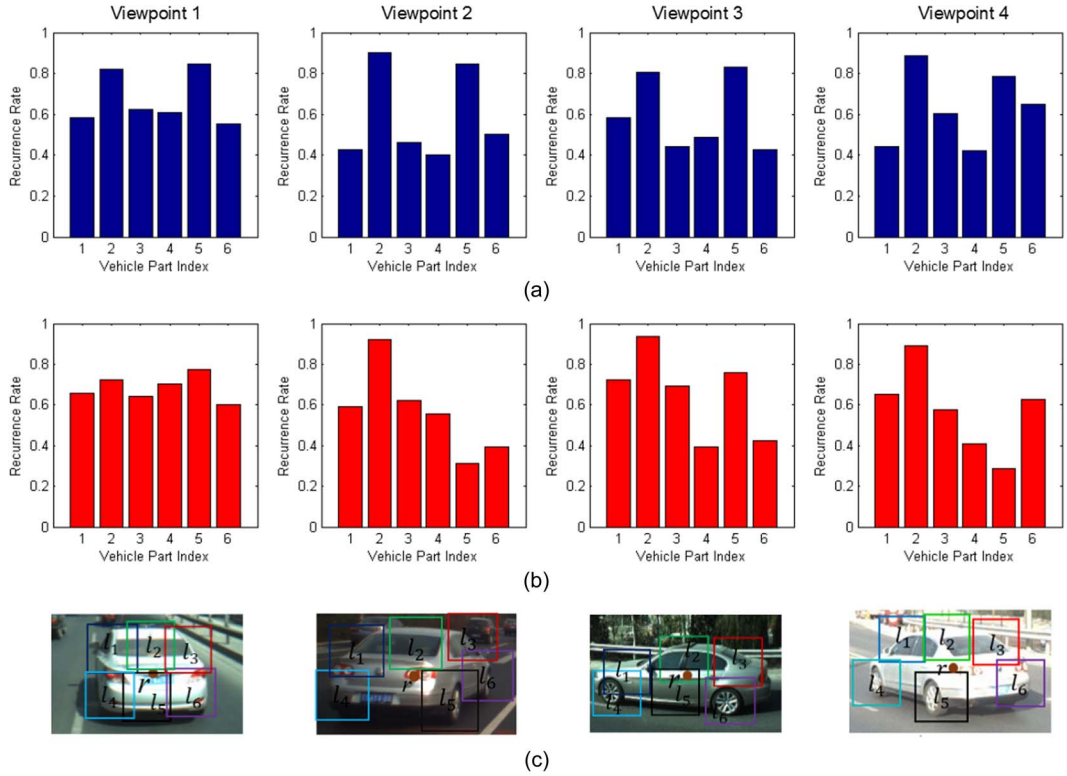



Fig. 17. Recurrence rate of vehicle part in each viewpoint class. (a) Recurrence rate of vehicle part in true positive detections. (b) Recurrence rate of vehicle part in false positive detections. (c) Vehicle parts in each viewpoint class.

Prompted by the above examination that parts may have different weights in inferring for a vehicle, we reshape formula (6) as below.

$$p\left(l_i^k | r^k, M_g^k\right) = w_i^k \cdot N\left(l_i^k - r^k; \mu_i^k, \Sigma_i^k\right) \tag{17}$$

where $w_i^k$ is a weight factor for part $i$ at viewpoint $k$, the value of which is assigned as below.

$$FPr_{avg}^k = \sum_i^{n=6} FPr_i^k / n \tag{18}$$

$$w_i^k = 1 + \max\left(0, 1 - FPr_i^k / FPr_{avg}^k\right). \tag{19}$$

Take the viewpoint 4 of Fig. 17(b) as an example, where $l_4$ and $l_5$ are the only parts having lower $FPr_i^k$s than $FPr_{avg}^k$, meaning that false positives happen more frequently without observations on the two parts, and on the other hand, prompting that the two parts could have higher significance in detection. Subsequently the weights of $w_4^4$ and $w_5^4$ are assigned more than 1.0 on formula (19), with the rest parts be 1.0. Accordingly, weights have been estimated as listed in Table V on overall data, low and high traffic densities respectively. A result is shown in Fig. 16, demonstrating an improvement in detection results after weighting parts in formula (17), where the overall values were used. On the other hand, it can be found in Table V that although minor vibrations exist, the weight values are not sensitive with traffic conditions. Learning a more accurate model in weighting parts according to their significance in detection will be studied in our future work.

## V. Conclusion

Partial observation and varying viewpoints are the key challenges in visual-based on-road vehicle detection. Inspired by the works on part-based detection, this research proposed a framework for on-road vehicle detection with its focus on vehicle pose inference on the set of detected part instances by addressing both partial observation and varying viewpoints in one probabilistic framework. To this end, geometric models describing the configuration of vehicle parts as well as their spatial relations are learned for each dominant viewpoint, and viewpoint maps are generated on the knowledge to road structure, which provide probabilistic prediction to the viewpoints of a vehicle at each location of an ego frame. Extensive experimental studies are conducted to examine the performance, where a large-scale data set is developed using the data from an on-board vision system on the motor ways in Beijing, which contains more than 30 thousands of labeled ground truth of both fully observed and partially occluded vehicles on four viewpoints at the scenes of various traffic densities. The data set will be opened to the society in accompany with this publication at www.poss.pku.edu.cn. Experimental study demonstrates efficiency in detecting occluded vehicles through part-based inference, which shows that although the proposed work has a similar performance on fully observable vehicles with other state-of-the-art approaches, it is much outstanding on detecting those occluded ones. In addition, the experimental results reveal that different vehicle parts may have different significance in vehicle detection, proper weighting could improve the performance, which needs more intensive study in future.

Future works will be addressed on the scene of more complex structure such as junction and bypass. Reducing computation cost towards on-board processing will also be addressed.

## REFERENCES


[1] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1–23, Dec. 2013.

[2] S. S. Teoh and T. Braunl, "Symmetry-based monocular vehicle detection system," *Mach. Vis. Appl.*, vol. 23, no. 5, pp. 831–842, Sep. 2012.

[3] M. Bertozzi, A. Broggi, and S. Castelluccio, "A real-time oriented system for vehicle detection," *J. Syst. Architect.*, vol. 43, no. 1–5, pp. 317–325, Mar. 1997.

[4] P. Parodi and G. Piccioli, "A feature-based recognition scheme for traffic scenes," in *Proc. Intell. Veh. Symp.*, 1995, pp. 229–234.

[5] W. Seelen and C. Tzomakas, "Vehicle detection in traffic scenes using shadows," IR-INI, Institut Nueroinformatik, Ruhruniversitat, Bochum, Germany, Tech. Rep. 98-06, 1998.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 886–893.

[7] Z. Sun, G. Bebis, and R. Miller, "Monocular precrash vehicle detection: Features and classifiers," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 2019–2034, Jul. 2006.

[8] R. Wijnhoven and P. de With, "Unsupervised sub-categorization for object detection: Finding cars from a driving vehicle," in *Proc. IEEE ICCV Workshops*, 2011, pp. 2077–2083.

[9] T. Machida and T. Naito, "GPU and CPU cooperative accelerated pedestrian and vehicle detection," in *Proc. IEEE ICCV Workshops*, 2011, pp. 506–513.

[10] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2009.

[11] The PASCAL Object Recognition Database Collection. [Online]. Available: http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html

[12] X. Zhang and N. Zheng, "Vehicle detection under varying poses using conditional random fields," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, 2010, pp. 875–880.

[13] P. Rybski, D. Huber, D. Morris, and R. Hoffman, "Visual classification of coarse vehicle orientation using histogram of oriented gradients features," in *Proc. IEEE Intell. Veh. Symp.*, 2010, pp. 921–928.

[14] Q. Yuan, A. Thangali, V. Ablavsky, and S. Sclaroff, "Learning a family of detectors via multiplicative kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 514–530, Mar. 2011.

[15] W.-C. Chang and C.-W. Cho, "Real-time side vehicle tracking using parts-based boosting," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2008, pp. 3370–3375.

[16] P. Felzenszwalb, R. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 2241–2248.

[17] H. T. Niknejad, T. Kawano, Y. Oishi, and S. Mita, "Occlusion handling using discriminative model of trained part templates and conditional random field," in *Proc. IEEE Intell. Veh. Symp.*, 2013, pp. 750–755.

[18] S. Sivaraman and M. M. Trivedi, "Vehicle detection by independent parts for urban driver assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1597–1608, Dec. 2013.

[19] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, pp. 511–518.

[20] S. Sivaraman and M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 267–276, Jun. 2010.

[21] C. Huang and R. Nevatia, "High performance object detection by collaborative learning of joint ranking of granules features," *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 41–48.

[22] B. Wu and R. Nevatia, "Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 185–204, Apr. 2009.

[23] H. T. Niknejad, A. Takeuchi, S. Mita, and D. McAllester, "On-road multivehicle tracking using deformable object model and particle filter with improved likelihood estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 748–758, Jun. 2012.

[24] B.-F. Lin *et al.*, "Integrating appearance and edge features for sedan vehicle detection in the blind-spot area," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 737–747, Jun. 2012.

[25] W. Liu, X. Wen, B. Duan, H. Yuan, and N. Wang, "Rear vehicle detection and tracking for lane change assist," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 252–257.

[26] J. Nuevo, I. Parra, J. Sjoberg, and L. Bergasa, "Estimating surrounding vehicles pose using computer vision," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2010, pp. 1863–1868.

[27] X. Zhang, N. Zheng, Y. He, and F. Wang, "Vehicle detection using an extended hidden random field model," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2011, pp. 1555–1559.

[28] A. Chavez-Aragon, R. Laganiere, and P. Payeur, "Vision-based detection and labeling of multiple vehicle parts," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2011, pp. 1273–1278.

[29] S. Sivaraman and M. M. Trivedi, "Real-time vehicle detection using parts at intersections," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2012, pp. 1519–1524.

[30] C. Kuo and R. Nevatia, "Robust multi-view car detection using unsupervised sub-categorization," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2009, pp. 1–8.

[31] H. Zhao *et al.*, "Omni-directional detection and tracking of on-road vehicles using multiple horizontal laser scanners," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 57–62.

[32] C. Wang, H. Zhao, F. Davoine, and H. Zha, "A system of automated training sample generation for visual-based car detection," in *Proc. IEEE Conf. Intell. Robots Syst.*, 2012, pp. 4169–4176.

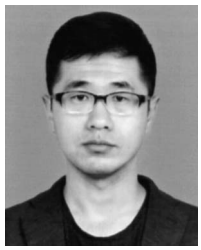
[33] B. Pepik, S. Michael, P. Gehler, and B. Schiele, "Occlusion patterns for object class detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 3286–3293.

[34] B. Li, T. Wu, and S. Zhu, "Integrating context and occlusion for car detection by hierarchical and-or model," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 652–657.

[35] E. Ohn-Bar and M. M. Trivedi, "Learning to detect vehicles by clustering appearance patterns," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2511–2521, Oct. 2015.

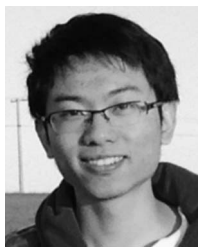
[36] Y. Xiang, W. Choi, Y. Lin, and S. Savarese, "Data-driven 3D voxel patterns for object category recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Boston, MA, USA, 2015, pp. 1903–1911.

[37] The KITTI Vision Benchmark Suite. [Online]. Available: http://www.cvlibs.net/datasets/kitti/

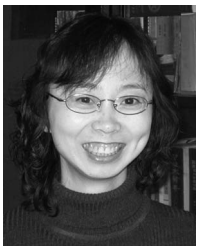
**Chao Wang** received the B.S. degree in automation from Tsinghua University, Beijing, China, in 2010 and the Ph.D. degree in computer science (intelligent science and technology) from Peking University, Beijing, China, in 2015. He is currently working toward the Ph.D. degree in intelligent robots with Key Laboratory of Machine Perception (MOE), and also the School of Electronics Engineering and Computer Science, Peking University.

His research interests include intelligent vehicles, intelligent transportation systems, computer vision, and machine learning.

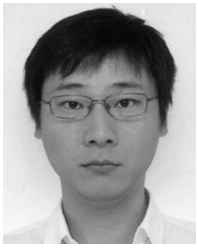


**Yongkun Fang** received the B.S. degree in computer science (intelligent science and technology) from Peking University, Beijing, China, in 2013. He is currently working toward the Ph.D. degree in intelligent robots with Key Laboratory of Machine Perception (MOE), and also the School of Electronics Engineering and Computer Science, Peking University.

His research interests include computer vision, machine learning, and intelligent vehicles.



**Huijing Zhao** received B.S. degree in computer science from Peking University, Beijing, China, in 1991 and the M.E. and Ph.D. degrees in civil engineering from The University of Tokyo, Tokyo, Japan, in 1996 and 1999, respectively. From 1991 to 1994, she was with Peking University, where she was recruited for a project of developing a GIS platform. In 2003, after postdoctoral research with The University of Tokyo, she was promoted to be a Visiting Associate Professor with the Center for Spatial Information Science. Since 2007, she has been with Peking University as an Associate Professor with the Key Laboratory of Machine Perception (MOE), and also with the School of Electronics Engineering and Computer Science. Her research interests include intelligent vehicles, machine perception, and mobile robots.



**Chunzhao Guo** received the B.S. degree in control science and engineering and the Ph.D. degree in pattern recognition and intelligent system from University of Science and Technology of China, Hefei, China, in 2002 and 2007, respectively.

He is currently a Researcher with Toyota Central Research and Development Laboratories, Inc., Nagakute, Japan. Previously, he was an Assistant Professor with Toyota Technological Institute, Nagoya, Japan. His research interests include machine learning, computer vision, intelligent vehicles, and biomimetic robots.

**Seiichi Mita** received the B.S., M.S., and Ph.D. degrees in electrical engineering from Kyoto University, Kyoto, Japan, in 1969, 1971, and 1989, respectively.

He is currently a Distinguished Professor with and the Director of the Research Center for Smart Vehicles, Toyota Technological Institute, Nagoya, Japan. He is also a Joint Professor with Toyota Technological Institute at Chicago, Chicago, IL, USA. His research interests include digital signal processing for recording and communication channels, and machine learning applications for vehicle environment recognition.

**Hongbin Zha** received the B.E. degree in electrical engineering from Hefei University of Technology, Hefei, China, in 1983 and the M.S. and Ph.D. degrees in electrical engineering from Kyushu University, Fukuoka, Japan, in 1987 and 1990, respectively. After working as a Research Associate with Kyushu Institute of Technology, he joined Kyushu University as an Associate Professor in 1991. In 1999, he was also a Visiting Professor with the Centre for Vision, Speech, and Signal Processing, University of Surrey, Surrey, U.K. Since 2000, he has been with Peking University, Beijing, China as a Professor with the Key Laboratory of Machine Perception (MOE), and also with the School of Electronics Engineering and Computer Science. His research interests include computer vision, digital geometry processing, and robotics.