

## Research Article

# Occluded Street Objects Perception Algorithm of Intelligent Vehicles Based on 3D Projection Model

Wei Liu , Longsheng Wei, and Yongbo Li

*School of Automation, China University of Geosciences, Wuhan, Hubei 430074, China*

Correspondence should be addressed to Wei Liu; [liuwei@cug.edu.cn](mailto:liuwei@cug.edu.cn)

Received 7 March 2018; Accepted 14 May 2018; Published 14 June 2018

Academic Editor: Yair Wiseman

Copyright © 2018 Wei Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a super perception system of intelligent vehicles for perceiving occluded street objects by collecting images from neighbor front vehicles V2V (vehicle to vehicle) video streams based on 3D projection model. This super power can avoid some serious accidents of driver-assistant systems or automatic driving systems which can only detect visible objects. Our street perception system can “see through” the front vehicles to realize detecting of the occluded street objects only by analyzing the pair images received from front and host (back) vehicles. Upon the 3D projection model based on the pair images, the system uses affine transformation to achieve augmented reality method to increase the visibility perspective of driver system. Experimental results on different datasets are shown to validate our approach. Evaluation method was first introduced into our perception system.

## 1. Introduction

According to the World Health Organization 2016 Global's report on road traffic injuries, about 1.25 million people die each year as a result of road traffic crashes. Between 20 and 50 million more people suffer nonfatal injuries [1]. The statistics from NHTSA show that 31% of traffic accidents are due to rear-end collisions which happens when front vehicles suddenly slow down or stop because of the chain reaction of the road objects ahead of them [2]. This is mainly because the unawareness of the situation ahead of the front vehicles or incapability of perception of the existence of street objects which are occluded by the front vehicles, which are especially serious in urban street. These can also lead to other kinds of accidents. For example, a car or a pedestrian may burst into the street at the head of other vehicles at any time, while the driver cannot have enough reaction time and will definitely cause serious traffic accident. Another example is when cars execute a passing maneuver through trucks or buses. This will cause one of the most severe road crashes when a vehicle shifts into an opposing traffic lane and crashes head on with an oncoming vehicle.

The WHO report also mentioned that improving the safety features of vehicles will be an efficient way to prevent

the road traffic injuries [1]. By analyzing the driving aid technologies of advanced vehicles, it is obvious that an enhanced forward road perception system is an efficient way to decrease the traffic incidents mentioned above. For mostly forward street perception system can only perceive things that vehicles can “see” but failed when street objects are partly or fully occluded. As human beings tend to believe in what they see with their own eyes, the drivers can have enough reaction time if they can see through the front vehicles and can realize the situation happened in front of the front vehicles. In this way, accidents will be decreased sharply in urban street.

The idea of V2V in our paper is similar to the truck platoons [3–5]. The difference is that the truck platoon link trucks or cars in train-like line which can save fuel, fit more cars on road, and potentially improve safety. However, in our systems, vehicles can drive on both the same lanes and different lanes and connect to each other to improve the safety based on sharing the video data. Using V2V communication based on DSRC, a video stream of vehicles' forward-looking cameras can be transmitted between vehicles with low delay [6]. Our perception system of the host vehicle (vehicle A in Figure 1) can be augmented by combining the video information from the front vehicles (vehicle B in Figure 1).

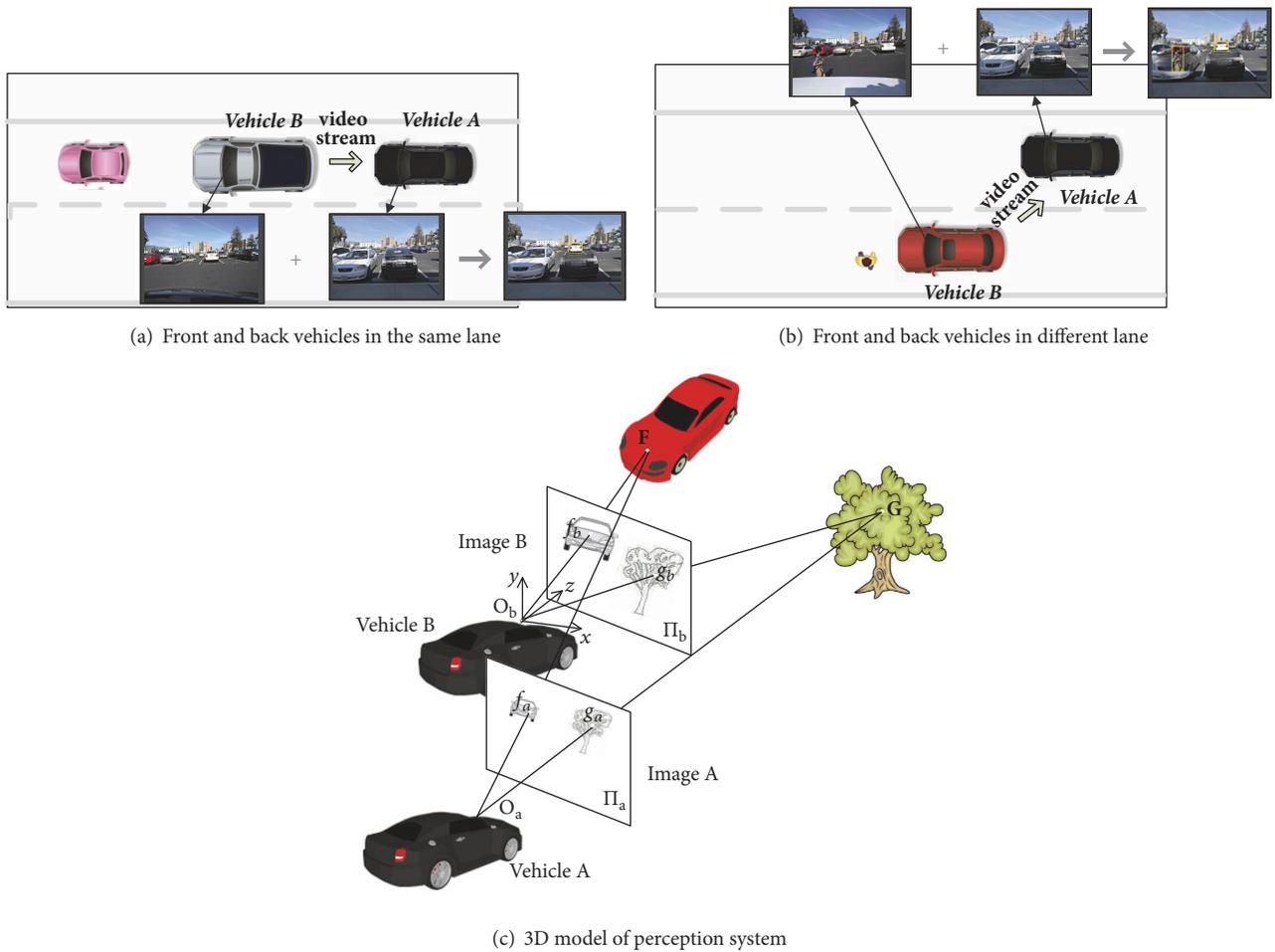


FIGURE 1: Street super perception system.

The same object will project in different location, scale, and shape in image A (vehicle A) and image B (vehicle B) as described in Figure 1(c).  $F$  represents the 3-space point of car, and corresponding image points  $f_a$  and  $f_b$  are the image of  $F$ . We only know car's image points  $f_b$  in image B, and image points  $f_a$  of the same car in image A need to be calculated.  $G$  is the 3-space point of another object (such as tree) and the corresponding image points in image A and image B are  $g_a$  and  $g_b$ . This object can be seen by both vehicles A and B. Our purpose is trying to find the 3D projection model between two images based on matching points in image A and image B (such as  $g_a$  and  $g_b$ ). Then, points  $f_a$  can be computed based on 3D projection model and point  $f_b$ .

Few works have been done on front street perception based on V2V video stream. Most of the forward perception systems cannot be deployed as see-through windows [7–11]. These perception systems can detect their distance to other vehicles and can only perceive visible street objects via sensors, such as laser, radar, and camera. As to the collaborative method, [12, 13] employed location information of vehicles which is periodically exchanged to prevent potential danger in advance. Similarly, [14, 15] adopted the route data coming from a conventional navigation system. But drivers

are not sensitive to digital information. Human beings tend to believe in what they see with their own eyes. The works [6, 16] propose vehicles transparent method based on V2V video streams in order to deal with passing maneuvers. Their method need accurate distance information gathered from radar sensor to realize the object projection between two images. The work [17] proposes a vehicle blind spot elimination system based on videos captured from other vehicles. But the linear projection fusion causes superior deformation when front and back vehicles are in different lanes.

Great progress has been made on modern object detector based on convolution neural network (CNN) recently, such as Fast-RCNN [18], R-FCN [19], Multibox [20], SSD [18], and YOLO [19], which are good enough to detect most of the objects on road but still failed on perception invisible things (including the fully occluded objects or objects more than 80% occluded).

Our street perception system can “see through” the front vehicles to realize detecting of the “invisible” objects (such as vehicles, pedestrians, and motorcycles) by analyzing the images received from front and back vehicles. We name it super perception system due to its super ability to see things through, shown in Figure 1. Only when street objects are

detected in front vehicle's windshield images, the perception system will turn on its "super ability." Then, the system will combine images from front vehicles and project occluded objects on windshield image of the host vehicle. And there is no limitation of the location of front and back vehicles. The main idea is inspired from [6, 17]. The main contributions of this paper are as follows.

(1) In our paper, a novel super perception system for perceiving street object partly or fully occluded by front vehicles is proposed. Our system gives the architecture of the new generation driving aid systems based on V2V stream.

(2) The system also drafts simple communication protocol between two vehicles.

(3) We improved the results of the augmented reality method on projecting objects from front vehicle image to host vehicle image.

(4) We also proposed object-based fusion method based on affine transformation. The fusion will happen only when street objects are detected on the images of front vehicles. And only object's areas were fused to the host vehicle image to augment the perception of the intelligent vehicles.

(5) A performance evaluation method is first proposed in this paper to show the accuracy of the projection.

## 2. Architecture of Super Perception System

The super perception system, shown in Figure 1, of the host vehicle (vehicle A in Figure 1) can be enhanced by combining the video information from the front vehicles (vehicle B in Figure 1). Vehicle B periodically sends beacon signal to surrounding vehicles based on DSRC equipment after detecting street objects based on CNN. If street objects were detected, B sends beacon signal to A and began to receive the video stream from B. The perception system on A began the fusion process based on 3D projection model. The 3D projection model parameters were computed depending on the synchronize images from A and B. The flowchart in Figure 2 describes the processing procedure of the perception system.

The system is constituted of four function blocks boxed in blue dashed line. Function blocks includes communication protocol block, object detection block, 3D projection model block, and object-based fusion block. In the following sections, these blocks will be described in detail.

## 3. DSRC-Based V2V Communication

A lot of advanced driver-assistance system applications are available based on V2V communication standards DSRC [21]. V2V communication plays a decisive role in several of these cooperative approaches. Figure 3 shows a flowchart describing the communication process (in black) of each vehicle and the communication protocol (in blue) between two vehicles.

Every vehicle periodically sends beacon signals to nearby vehicles after street objects are detected in its forward images. In Figure 3 B detected cars or pedestrian in its image and then sent beacon signals to vehicle A. A receives beacon signal from B and the perception system is activated. The

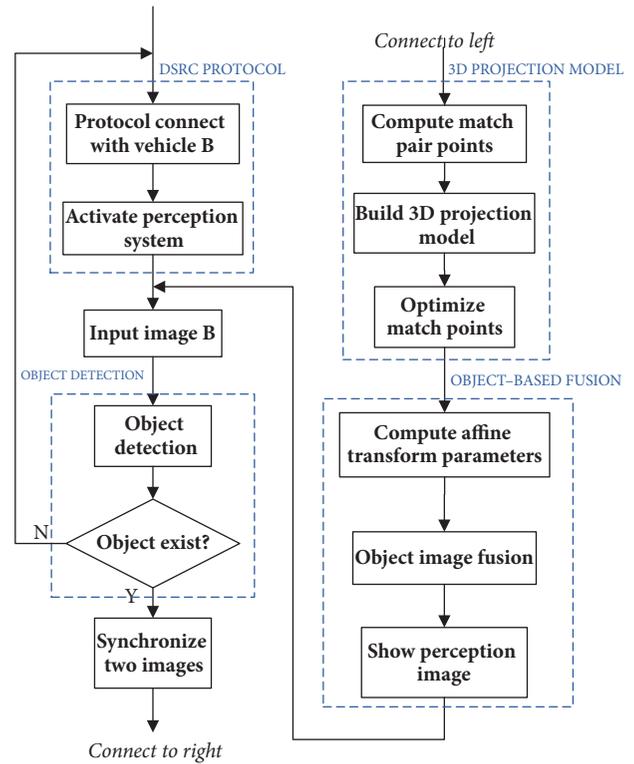


FIGURE 2: Flowchart describing the processing procedure of the perception system.

cooperative protocol between two vehicles is initiated, with vehicle A request for video stream and camera intrinsic parameter from vehicle B. Then, B sends those data to A. The delay caused by transmission will be lower than 100ms for per frame image. And the effect under 100ms can be ignored in perception systems, which is proved in experiments in [3]. If no objects were detected, vehicle B will stop sending videos and will send stop signal to A, which will terminate the communication between A and B.

## 4. SSD Based Street Objects Detection

Before the perception system is activated, the image of vehicle B needs to be analyzed first. Detection must be applied before pursuing projection, as the projection only happens when there exist some occluded street objects (which mean there are some street objects in front of B). Here we adopt an end-to-end single deep neural network to realize detection algorithm, named SSD [22]. The architecture of SSD network is shown in Figure 4. VGG16 form the early network layers (truncated before any classification layer). The layers added to the truncated VGG16 are 5 convolutional feature layers which progressively decrease in size and can produce feature maps in different scale. A small set of default boxes slide on several feature maps. For each default box, the shape offset and the confidence score were calculated for all street object categories (here we have 3 kinds of street objects; they are vehicle, pedestrian, and bicycle). The objective loss function

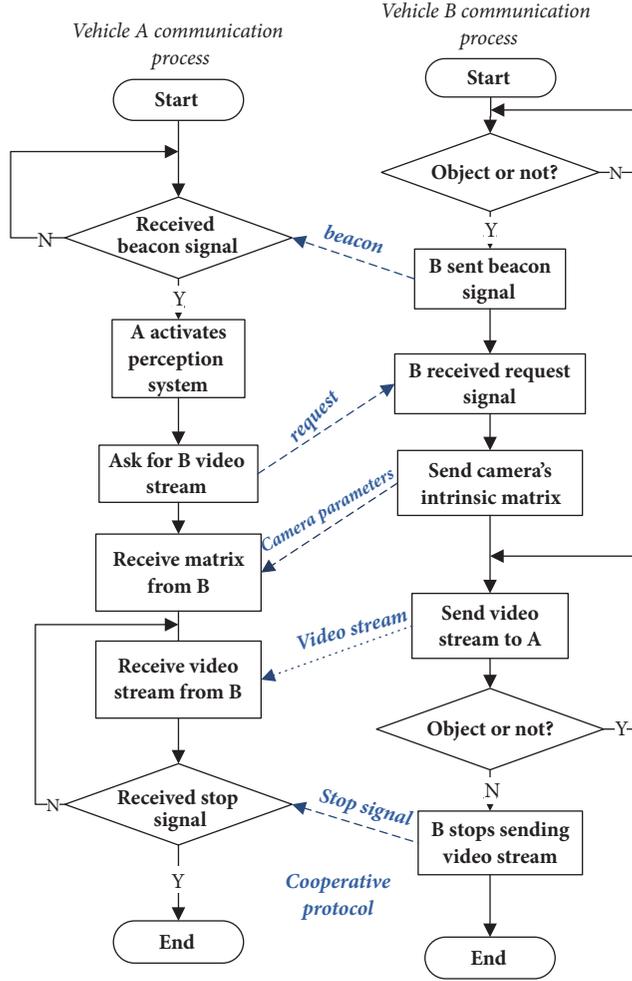


FIGURE 3: Flowchart showing the communication protocol between two vehicles and processing flow of each vehicle.

consists of the confidence loss (conf) and the localization loss (loc):

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)). \quad (1)$$

$N$  is the number of matched default boxes.  $l$  means the location of predicted box and  $g$  presents the ground-truth box.  $x = \{1, 0\}$  indicate the matching of default box and the ground-truth box. Then, a set of default boxes with each feature map cell are associated at the top of the network. Finally, model supplies nonmaximum suppression algorithm to find the most confidence level boxes depending on the score and region. This algorithm can achieve 90.27% mAP on our testing street dataset at 49 FPS on a Nvidia Titan GTX 1080i GPU.

SSD achieves good performance both on accuracy and on speed which can meet our request. Though SSD performs poor on small object detection, small object means far distance from vehicle's perception system which will not cause accident in this situation.

## 5. 3D Front-and-Back Projection Model

If street objects are detected in image of vehicle B, two images from both front and back vehicles are used to construct the 3D projection model, shown in Figure 5. Different from the projection model proposed by Yair [23–25], our 3D projection model is based on the epipolar geometry which only depends on the cameras' internal parameters and their relative pose, but independent of scene structure. As the locations of the two cameras are different, the same object will have different size, location, and deformation after projection into two cameras. Finding the transformation function between two images is the key to projection. Feature pairs in both images achieve the realization of 3D projection model. The processing can be divided into 3 parts: (1) feature pair selection; (2) camera epipolar geometry estimation; (3) object-based fusion. More details are discussed below.

**5.1. Feature Pair Selection.** In order to provide a representative description of the object, we can extract characteristic feature points on the object in an image. We make use of these descriptions to find the correspondence points of the same object in two images. To perform trustful points matching, it is very important that the description of the points should be invariant to the changes in scale, noise, illumination, and deformation. Feature Pairs Selection includes feature detection and feature matching.

(1) *Feature Detection.* Here, we adopt Lowe's SIFT method [26] as feature selection and description method. SIFT method uses a 128-element-long feature vector descriptor to characterize the gradient pattern in a properly oriented neighborhood surrounding a SIFT feature. These features are (semi-)invariant to incidental environmental changes in lighting, viewpoint, and scale. Here, we used a public-domain SIFT implementation (<http://www.vlfeat.org/>).

(2) *Feature Matching.* Emphatically, matching SIFT features in front and back images is trying to search the similarity of the those descriptors, as lacking of the relative pose of two cameras. Brute-force algorithm is adopted here to match feature pairs in the front and back images. The Euclidean distance was computed between feature vectors as the matching score. The selected matching pair needs to meet this equation (2) [26].

$$ratio = \frac{\max(dis(f_a, f_b))}{\max sec(dis(f_a, f_b))} > 0.8. \quad (2)$$

$\max(dis(f_a, f_b))$  means the best matching pair and  $\max sec(dis(f_a, f_b))$  is the second best one.  $f_a, f_b$  represent points from feature map of vehicle images A and B. Figure 7(a) depicts the matching results between front and back images. The results reveal that there exists error pair matching only based on similarity. In the following section, geometrical constraints will be introduced in filter error matches.

**5.2. Camera Epipolar Geometry Estimation.** No matter what relative position two vehicles are at, if two images acquired by

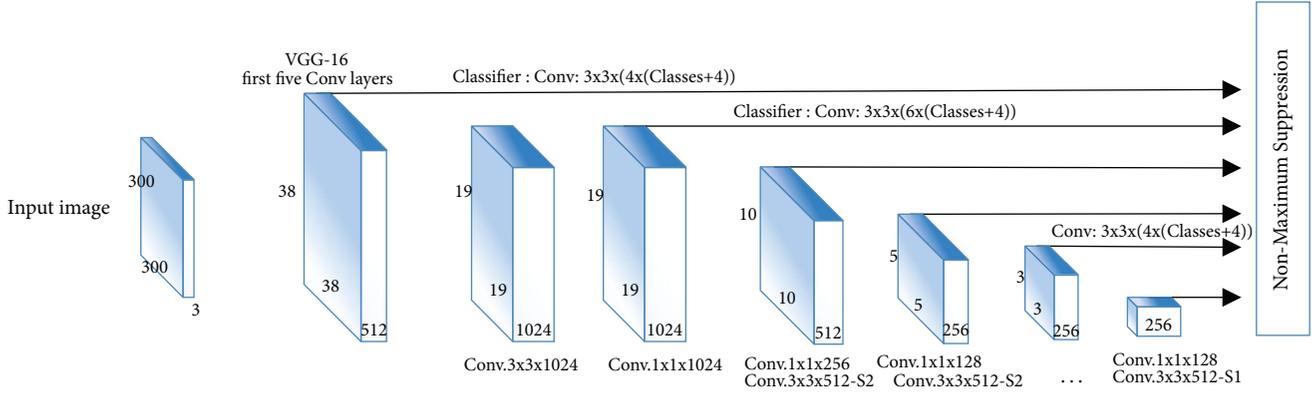


FIGURE 4: Network architecture of SSD.

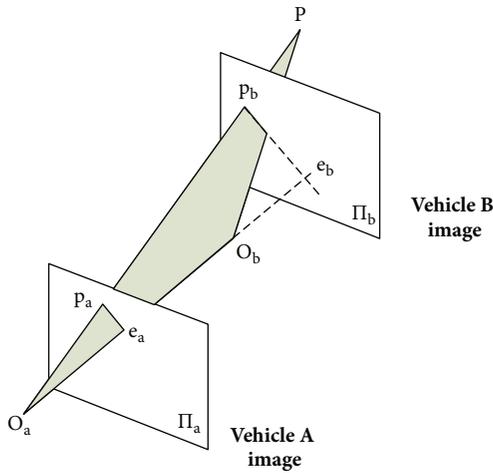


FIGURE 5: Front and back 3D projection model.

two vehicles have noncoincident centers, then the fundamental matrix  $F$  is the unique  $3 \times 3$  rank 2 homogeneous matrix which satisfies [27]

$$p_a^T F p_b = 0 \quad (3)$$

where  $(p_a, p_b)$  is any pair of the corresponding points in two images from vehicles A and B. This enables  $F$  to be computed from image correspondences alone. Here, we used the five-point algorithm [28]. The Normalized Eight-Point Algorithm [29] can also be used to improve the performance while lowering the efficiency.

Based on the fundamental matrix  $F$ , we can calculate the parameters rotation  $R$  and direction of translation  $T$ , which shows the relative pose between two cameras. As shown in Figure 5, a point 3-space  $P$  is imaged as  $p_a$  in the view of vehicle A and  $p_b$  in the view of vehicle B.  $O_a$  and  $O_b$  denote two cameras' optical centers, and  $\Pi_a$  and  $\Pi_b$  are the corresponding image A and image B. Geometrically, points  $O_a, O_b, p_a, p_b,$  and  $P$  all lie on the same epipolar plane, which gives the expression

$$O_b p_b \cdot (O_b O_a \times O_a p_a) = 0. \quad (4)$$

Suppose  $K_a$  and  $K_b$  give the intrinsic matrix for the two cameras.  $R$  and  $T$  denote the movement between two cameras.

$$T = O_b O_a \quad (5)$$

$$O p = R K^{-1} p + T. \quad (6)$$

Substitute (5) and (6) into formula (4):

$$(K_b^{-1} p_b)^T (T \times (R K_a^{-1} p_a + T)) = 0 \quad (7)$$

$$p_b^T (K_b^{-T} T \times (R K_a^{-1})) p_a = 0.$$

Comparing formulas (7) and (3), we have

$$(K_b^{-T} T \times (R K_a^{-1})) = F. \quad (8)$$

As  $F$  is calculated, we can compute  $R$  and  $T$  with formulas (5) and (8).

**5.3. Feature Pair Optimization.** Generally, five pairs of points or eight pairs of points are enough to compute fundamental matrix  $F$ . In fact, we often match more features than that. Hence, we can iteratively use these features to gain an optimistic result. Here, RANSAC algorithm is employed to improve the robustness in camera motion inference. Randomly selected  $n$  small subsets "seed" ( $n$  pairs of matching points), fundamental matrix  $F$  is calculated in  $n$  times. The value of  $|p_a^T F p_b|$  calls the residual error, which is, ideally, supposed to be zero. An  $F$  will be computed by those outlier-free seeds and will produce small residual errors in  $|p_a^T F p_b|$  for mostly inlier matching pairs. We preserve those seeds that produce the minimum median  $|p_a^T F p_b|$  residual errors.

After filtering the error pair points, five-point algorithm is then performed to compute the precise value of  $F$ ,  $R$ , and  $T$  by using all remaining pairs in a least-square way. Also, epipolar  $e_a$  and  $e_b$  can be solved by calculating the standard Singular Value Decomposition (SVD) of  $F$  [27].

#### 5.4. Object-Based Data Fusion

(1) *Estimation of Transformation Parameters.* In order to realize fusing objects on image B to image A, we need to

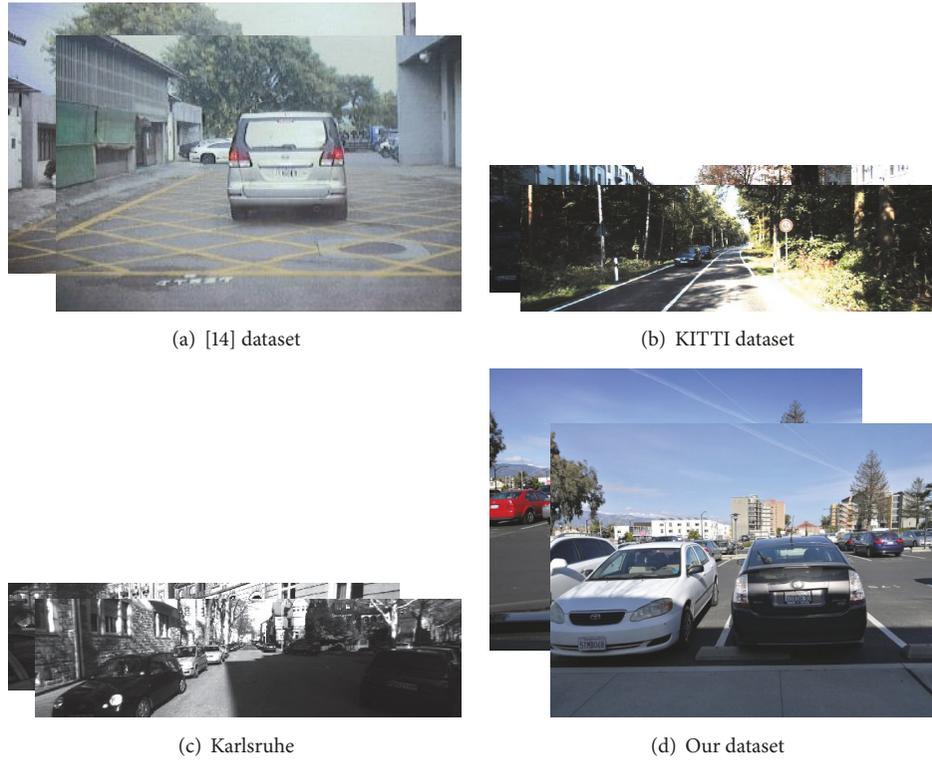


FIGURE 6: Image dataset used in this paper.

figure out some information related to detected objects. The information includes size, shape, and location of the fusion region. Hence, the mapping parameters between two images need to be estimated. The work [17] used polar coordinates to approximately simulate the mapping relation between two images in a global, linear form to get visually appealing color fusion result. This method can achieve good performance in vehicles in the same lane (shown in Figure 1(a)) but failed in vehicles in different lane (shown in Figure 1(b)). As an affine transformation is a nonsingular linear transformation followed by a translation, we regard the mapping relationship between two images as an affine transformation. It has the matrix representation

$$P_B = HP_A. \quad (9)$$

$P_B$  and  $P_A$ , respectively, represent matching pair point matrix in two images.  $H$  is the parameter matrix of affine transformation. The homogeneous formula is as follows:

$$\begin{pmatrix} x_b \\ y_b \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & t_1 \\ a_{21} & a_{22} & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_a \\ y_a \\ 1 \end{pmatrix}. \quad (10)$$

$a_{11}$ ,  $a_{12}$ ,  $a_{21}$ ,  $a_{22}$ ,  $t_1$ , and  $t_2$  are six parameters in  $H$  matrix. Five pairs of point are enough to calculate the parameters of affine transformation. However, we have more feature pairs than that. Hence, RANSAC is also used to optimize the parameter of the transformation.

Affine transformation can offer better performance than linear method in [17], but the projection of the objects from

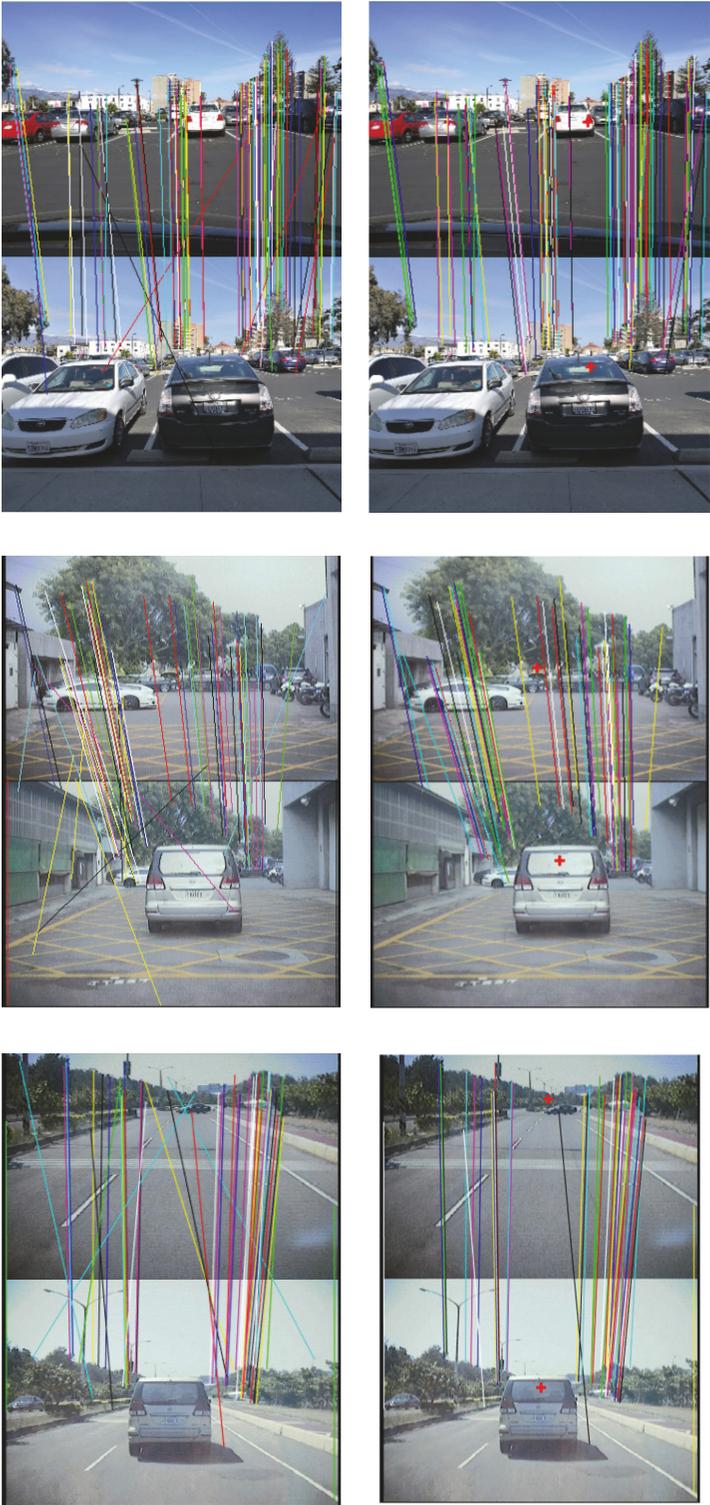
image B to image A is still not precisely correct. Because we use matching points  $g_a$  and  $g_b$  of object G to estimate the parameters of affine transformation of points  $f_a$  and  $f_b$  of object F, this rough method can only be applied to realize visualization, which helps drivers to “see” objects occluded and perceive the approximate situation ahead of the front vehicles. In order to get precise results, more information should be adopted, such as deep information of every pixel. Or we can use other transformation to replace affine transformation.

(2) *Images Fusion*. The fusing region, where the fusing process is applied to, is a circle area in the image of vehicle A. The center and radius of the circle depend on the detected object region location and size. Epipolar  $e_a$  and  $e_b$  can be used to eliminate those objects that are not occluded by vehicle B. The blending method is similar to [14]. The blending weight is adjusted to use more color from the front image B close to fusion center and more color from the back image A away from center toward the edge of the circle. The transparency parameter controls the mixture of two images.

## 6. Experiment Results

Our proposed system runs on a server with a Nvidia Titan GTX 1080i GPU.

6.1. *Datasets*. Experiments were performed on dataset from [17], Karlsruhe dataset, KITTI dataset, and our own dataset. Four Datasets were shown in Figure 6.



(a) Matching pairs based on similarity (b) Optimized matching pairs based on geometry constraint

FIGURE 7: Matching pairs features between front and back vehicles.



FIGURE 8: Affine transformation results based on front and back vehicle images.

In [17], two minivans were equipped with forward-looking cameras and video capturing system which capture on-board video and analyze the store videos. Two vehicles were driven in a front-and-back style following configuration on many different roads. Two vans were driven in a front-and-back style in the same lane. The data were captured on many different streets.

Our perception system was applied in two situations: front and back vehicles in the same lane and vehicles in the difference lanes. The [17] dataset includes images in which vehicles are in the same lane. However, our dataset contains both situations. Our database images were acquired from videos captured by mobile phones camera. We equipped three cars with forward-looking cameras. Most of our data now were captured in parking lot area.

Karlsruhe dataset [30] contains high-quality stereosequences recorded from a moving vehicle in Karlsruhe. The sequences, which are captured by Pointgrey Flea2 firewire cameras, are saved as rectified images in \*.png format. Ground-truth odometry from an OXTS RT 3000 GPS/IMU system is provided in a separate text file. Here, we use two

frames ( $\Delta t$ ) in the video to simulate the front and back vehicle images. And we used Karlsruhe dataset and KITTI dataset [31] to evaluate the accuracy of the projection results. However, it is hard to evaluate the results by using dataset in [17] and our dataset because the objects in back images are occluded by the front vehicles.

**6.2. Feature Matching and Optimization Results.** Feature matching and optimized matching results are shown in Figure 7. Results in column (a) are the matching results after pursuing Brute-force algorithm and results in column (b) show the matching results after being optimized by adopting geometrical constraint of projection model. The experiment results reveal the optimized results, in which error matches were deleted.

**6.3. Affine Transformation Results.** In our system, the objects in two images are supposed to meet the affine transformation. The affine transformation results are shown in Figure 8 and the quantitative evaluation is performed on KITTI and



FIGURE 9: IoU of two methods compared with ground-truth.



(a) Fusion based on polar linear method [14]



(b) Fusion based on affine transformation

FIGURE 10: The results of two projection methods.

Karlsruhe dataset. The images in these two datasets are captured by one vehicle; here we use two frames with interval  $\Delta t$  in the data frames to simulate the front and back vehicle images.  $\Delta t$  is a random value within 3~10.

The (a) images represent the front vehicle images and the (b) pictures are supposed to be the back vehicle images. The (c) pictures show the results that transform the front images to the images of the back vehicles based on affine transforming.

The (b) images were taken as the ground-truth images and the affine results (c) are compared to the ground-truth images. The evaluation results are represented in Figure 9 and Table 1 shows the improving results of our method compared to [17] especially if the locations of the front and back vehicles do not follow the linear relationship. Method in [17] supposes that the front vehicle and back vehicle meet the linear model, but, in fact, most of the situations do not meet this hypothesis.

IoU (Intersection over Union) is an evaluation metric used to measure the accuracy of an object detector.

$$IoU = \frac{ObjectBox \cap GroundTruth}{ObjectBox \cup GroundTruth}. \quad (11)$$

We use IoU here to evaluate affine results which is key to the accuracy of perception. In Figure 9, red box is the ground-truth bounding box, yellow box is the result of [14] method, and green box is the result of our method.

Figure 10 reveals the projection results based on different fusion methods. Result of (a) adopts the method used in [17] which supposes the projection between front and back vehicle images satisfied the polar linear relationship, whereas result in (b) adopts the affine transform to hypothesis of the relationship between two images. Obvious improvement

TABLE 1: Results on the KITTI and Karlsruhe dataset for different IoU thresholds.

Methods	0.6	0.7	0.8	0.9
[14]	80.43%	79.16%	65.22%	56.12%
Our method	82.52%	80.32%	67.01%	62.82%

can be seen from the compared results. However, the results of affine transformation still were not accurate according to the facts, because the pair features choose from images that belong to different objects (most of them belong to background) in different depth. And these pair features were used to calculate the affine parameters of one object, which results in the deviation results. So we need more complicated model to imitate the relationship between the objects.

**6.4. Object Fusion Results.** Affine transformation is supposed to satisfy the relationship between front and back vehicles, so the augmented fusion can be done based on the above calculated affine parameters. Figure 11 reveals the final fusion results. If the front vehicle (vehicle B) detected the object on street, it will send its image B to back vehicle (vehicle A) and the detected objects are fused in image A. The yellow rectangular shows the detected objects which are occluded by front vehicle B. The fusion process is to blend the pixels color in image A with the corresponding pixels of objects' area in image B. The fusion region is a circle where the center is the center of the rectangular and we set the transparency parameter to be 1. The blending weight is adjusted to use more color from the front image close to the center and more color from the back image away from the center.

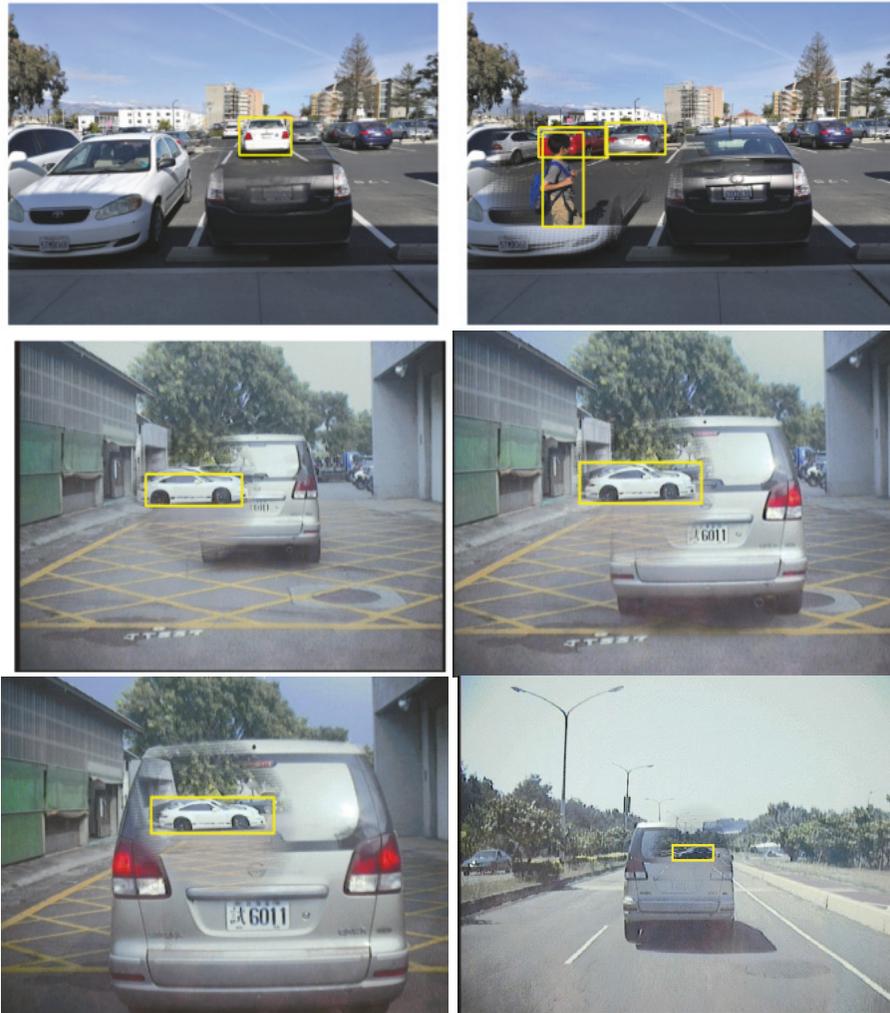


FIGURE II: The object-based fusion results.

## 7. Conclusion

In this paper, we introduce a super perception system which can “see through” vehicle and detect the fully occluded street objects. Our perception system is a good example of advanced driver-assistance system (ADAS) that can collect information from sensors in neighbor vehicles. Our future research will focus on the algorithm on how to improve the accuracy of the projection model, which means correct location and correct size of the occluded objects after projecting from front image to the back image. We also will invent a performance evaluation method to evaluate the projection results of the system.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported in part by the National Nature Science Foundation of China under Grants 61503349 and 61603357. The authors would like to thank Professor Shi Jinlong for helpful discussions. They also thank Professor Yuanfang Wang for supplying dataset of [17] and helpful comments.

## References

- [1] “Global Status Report on Road Safety,” <http://www.who.int/mediacentre/factsheets/fs358/en/>, 2016.
- [2] “Traffic Safety Facts,” <https://crashstats.nhtsa.dot.gov/#/>, 2015.
- [3] K. Liang, J. Mårtensson, and K. H. Johansson, “When is it fuel efficient for a heavy duty vehicle to catch up with a platoon?” *IFAC Proceedings Volumes*, vol. 46, no. 21, pp. 738–743, 2013.
- [4] K.-Y. Liang, J. Martensson, and K. H. Johansson, “Fuel-saving potentials of platooning evaluated through sparse heavy-duty vehicle position data,” in *Proceedings of the 25th IEEE Intelligent Vehicles Symposium, IV 2014*, pp. 1061–1068, Dearborn, MI, USA, June 2014.

- [5] K.-Y. Liang, J. Mårtensson, and K. H. Johansson, "Heavy-duty vehicle platoon formation for fuel efficiency," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1051–1061, 2016.
- [6] P. Gomes, C. Olaverri-Monreal, and M. Ferreira, "Making vehicles transparent through V2V video streaming," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 930–938, 2012.
- [7] G. Burnett, "A road-based evaluation of a head-up display for presenting navigation information," in *Proceedings of the HCI International Conference*, vol. 3, pp. 180–184.
- [8] S. Ammoun and F. Nashashibi, "Design and efficiency measurement of cooperative driver assistance system based on wireless communication devices," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 3, pp. 408–428, 2010.
- [9] M. A. Akhlaq, *Smart-Dashboard Augmenting Safe & Smooth Driving*, Blekinge Institute of Technology, Karlskrona, Sweden, 2010.
- [10] A. Amditis, M. Bimpas, G. Thomaidis et al., "A situation-adaptive lane-keeping support system: overview of the SAFELANE approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 617–629, 2010.
- [11] C. Lin, C. Lin, D. Huang, and Y. Chen, "Design a support vector machine-based intelligent system for vehicle driving safety warning," in *Proceedings of the 2008 11th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 938–943, Beijing, China, October 2008.
- [12] Q. Xu, T. Mak, J. Ko, and R. Sengupta, "Vehicle-to-vehicle safety messaging in DSRC," in *Proceedings of the the first ACM workshop*, p. 19, Philadelphia, PA, USA, October 2004.
- [13] T. ElBatt, S. K. Goel, G. Holland, H. Krishnan, and J. Parikh, "Cooperative collision warning using dedicated short range wireless communications," in *Proceedings of the 3rd ACM International Workshop on Vehicular Ad Hoc Networks (VANET'06)*, pp. 1–9, September 2006.
- [14] M. Ferreira, R. Fernandes, and H. Conceicao, "Self-organized traffic control," in *Proceedings of the 7th ACM International Workshop on Vehicular Inter-NETworking*, pp. 85–90, 2010.
- [15] M. Ferreira and P. M. D'Orey, "On the impact of virtual traffic lights on carbon emissions mitigation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 284–295, 2012.
- [16] "Transparent Trucks Show the Road Ahead," [https://www.thecarconnection.com/news/1040575\\_transparent-trucks-show-the-road-ahead](https://www.thecarconnection.com/news/1040575_transparent-trucks-show-the-road-ahead), 2009.
- [17] C.-T. Lin, Y.-C. Lin, L.-T. Chen, and Y.-F. Wang, "Front vehicle blind spot translucitization based on augmented reality," in *Proceedings of the 2013 IEEE 78th Vehicular Technology Conference, VTC Fall 2013*, September 2013.
- [18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.
- [19] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: object detection via region-based fully convolutional networks," <https://arxiv.org/abs/1605.06409>, 2016.
- [20] C. Szegedy, S. Reed, D. Erhan, and D. Anguelov, "Scalable, high-quality object detection," *Computer Science > Computer Vision and Pattern Recognition*, pp. 1412–1441, 2014.
- [21] R. Penazzi, P. Capozio, M. Duncan, A. Scuderi, M. Siti, and E. Merli, "Cooperative safety: a combination of multiple technologies," in *Proceedings of the Design, Automation and Test in Europe, DATE 2008*, pp. 959–961, March 2008.
- [22] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *European Conference on Computer Vision*, pp. 21–37, Springer, 2016.
- [23] Y. Wiseman and I. Grinberg, "Circumspectly crash of autonomous vehicles," in *Proceedings of the 2016 IEEE International Conference on Electro Information Technology (EIT)*, pp. 0387–0392, Grand Forks, ND, USA, May 2016.
- [24] Y. Wiseman and I. Grinberg, "When an inescapable accident of autonomous vehicles is looming," *International Journal of Control and Automation*, vol. 9, no. 6, pp. 297–308, 2016.
- [25] Y. Wiseman and I. Grinberg, "Autonomous vehicles should not collide carelessly," *Advanced Science and Technology Letters*, vol. 133, pp. 223–228, 2016.
- [26] D. Lowe, "Distinctive image feature from scale-invariant key points," *International Journal Computer Vision*, vol. 60, pp. 91–100, 1997.
- [27] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2nd edition, 2003.
- [28] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, 2004.
- [29] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 6, pp. 580–593, 1997.
- [30] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: dense 3D reconstruction in real-time," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '11)*, pp. 963–968, Baden-Baden, Germany, June 2011.
- [31] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: the KITTI dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

