MATCHING 5G CONNECTED VEHICLES TO SENSED VEHICLES FOR SAFE COOPERATIVE AUTONOMOUS DRIVING

Zuoyin Tang, Jianhua He, Kun Yang, and Hsiao-Hwa Chen[†]

[†]Corresponding Author's Address:
Hsiao-Hwa Chen
The Department of Engineering Science
National Cheng Kung University, Taiwan
Tel: +886-6-2757575 ext. 63320
Fax: +886-6-2766549
Email: hshwchen@ieee.org

Zuoyin Tang (email: z.tang1@aston.ac.uk) is with the School of Computer Science and Digital Technologies, Aston University, UK. Jianhua He (email: j.he@essex.ac.uk) and Kun Yang (email: kunyang@essex.ac.uk) are with the School of Computer Science and Electronic Engineering, Essex University, UK. Hsiao-Hwa Chen (email: hshwchen@ieee.org) is with the Department of Engineering Science, National Cheng Kung University, Taiwan.

The paper was submitted on May 3, 2023; revised on June 26, 2023 and July 22, 2023, respectively.

Abstract

5G connected autonomous vehicles (CAVs) help enhance perception of driving environment and cooperation among vehicles by sharing sensing and driving information, which is a promising technology to avoid accidents and improve road-use efficiency. A key issue for cooperation among CAVs is matching communicating vehicles to those captured in sensors such as cameras, LiDAR, etc.. Incorrect vehicle matching may cause serious accidents. While centimeter level positioning is now available for autonomous vehicles, matching connected vehicles to sensed vehicles (MCSV) is still challenging and has rarely been studied. In this paper, we are motivated to investigate the MCSV problem for 5G CAVs, propose and assess solutions for the problem to bridge the research gap. We formulate the MCSV problem and propose two MCSV approaches to support cooperative driving. The first approach is based on vehicle registration number (VRN), which is unique to identify a vehicle and can be shared among CAVs for MCSV. VRN is hashed before sharing to protect privacy, and will be compared to the shared one for vehicle matching. The second MCSV approach is based on visual features of vehicle's external views, which are shared with other CAVs and compared to those obtained from visual sensors to match the vehicles of interest. A new MCSV dataset is developed to assess the effectiveness of the proposed approaches. Experiment results show that both approaches are feasible and useful, and they achieve a very low false positive rate, which is critical for cooperative driving safety.

Index Terms

Connected vehicles; Vehicle to everything; 5G; Autonomous vehicles; Cooperative driving.

I. INTRODUCTION

Road accidents is a global concern in road transport systems. More than one million people died on the roads globally every year [1]. There is an urgent need to develop advanced road and vehicle technologies to improve driving safety. Connected vehicles (CV) and automated vehicles (AV) are promising technologies to reduce road accidents and improve road-use efficiency [2] [3] [4]. AVs have a great potential to avoid road accidents caused by human driving errors, which contribute to more than 90% road accidents. There are also many use cases identified for CVs, such as forward collision warning, blind intersection, queue warning, curve speed warning, traffic signal, emergency vehicles, and adaptive cruise control. However, both AV and CV technologies have inherent shortcomings [5]. For example, AV sensors are limited to line-of-sight sensing, and are not aware of the intents of the surrounding vehicles, which is important for cooperative driving such as lane merging and platooning. On the other hand, CV technology depends on message exchanges to build up mutual awareness.

Due to the limitations on CV and AV, connected autonomous vehicles (CAV) technologies, which combine the features of CV and AV, have attracted a lot of research and development interests [5]. CAVs can share their status, sensing and driving intents for comprehensive perception of driving environments and cooperation on safe driving. Third Generation Partnership Project (3GPP) is developing 5G enhanced cellular vehicle to everything (V2X) standards to support advanced driving applications [6]. The advanced driving includes cooperative sensing, vehicle platooning, remote driving and cooperative driving [6]. In some connected vehicles applications such as forward collision warning or avoidance, sharing accurate 5G CAV status (e.g., position, speed and acceleration) could be effective. However, in the advanced driving applications such as cooperative driving and platooning, they require not only reliable and low latency communications, but also reliable MCSV to identify and match the communicating vehicles from those captured in the sensors such as visual cameras and LiDAR.

The task of MCSV can be best described by an ego vehicle to identify some relevant communicating 5G CVs on cooperative driving among the vehicles detected by the sensors at the ego vehicles. As the ego vehicle will rely on object detection results and the shared information from the communicating CVs for assisted or automated driving, correct MCSV is important for safety-critical cooperative driving applications, such as unprotected turning, lane merging and platooning. It is expected that the AV safety level needs to be two orders higher than human drivers, i.e., about 10^{-8} fatalities per mile. CAV cooperation and especially MCSV need to be very reliable in order not to introduce any safety issue. However, the implementation of reliable MCSV is very challenging. According to [4], for AVs operating on local streets to achieve the safety target, the longitudinal location estimation error should be 0.29 m with an orientation requirement of 0.50 deg. Centimeter level location estimation is now available for autonomous vehicles with advanced sensors and technologies, such as global position system (GPS), inertial measurement unit (IMU), and LiDAR [7].

Note that even if CAVs can have precise location information for themselves that can be exchanged with others, it is not sufficient for solving the MCSV problem. The CAVs will plan their trajectories and make vehicle control decisions based largely on the perception of the driving environment with the sensors such as cameras, LiDAR and radars. The shared locations of communicating CAVs need to be transformed to the sensing domain (such as images for cameras and point clouds for LiDAR) for safety cooperative driving. While advanced deep learning models can boost up the accuracy of object detection and AVs' location estimation, it is still an open issue to map them to the real world precisely. Therefore, there is a huge research gap to fill up in MCSV. The problem is especially prominent with mixed road traffic, where there are human driven vehicles (HDV) and vulnerable road users. There is an urgent need for developing reliable and yet efficient MCSV solutions for cooperative CAVs. To our best knowledge, the MCSV problem has not been studied in the literatures so far.

In this paper we investigate the MCSV problem and propose two approaches for MCSV to support cooperative driving, such as unprotected turning and platooning. The first approach is based on vehicle registration number (VRN), which is unique for a vehicle and can be used to identify a vehicle. With this approach the candidate CAVs to be matched (VBMs) will share VRN for the purpose of cooperative driving. The VRNs are hashed before sharing to protect the privacy. The CAV to perform matching (VPM) will detect and recognize the VRNs of the candidate VBMs in a region of interest (ROI). The VRNs will be hashed and compared to the shared hash values by the candidate VBMs. If a pair of hash values match with each other, then the corresponding CAV is claimed to be matched. The second approach is based on visual features. The traditional feature descriptors such as SURF and deep learning features can be used [9] [10]. With this approach the candidate VBMs share visual features of their external images, and VPM will compare locally measured and shared visual features and determine if the CAV is matched.

To test the performance of the proposed approaches, a dataset for MCSV was developed from the public dataset for vehicle re-identification of the AI City Challenge 2019 [11]. Preliminary experiment results show that the proposed approaches can be useful for MCSV with a good false negative rate and a very low false positive rate. Note that these two approaches are complementary and thus can work jointly with other existing approaches to improve MCSV performance. The main contributions of this work can be summarized as follows.

1) We defined an MCSV problem, which is the first in the literature to trigger more investigations on such an important issue for connected autonomous driving safety;
2) We proposed two MCSV approaches to support reliable and safe cooperative driving for 5G CAVs;
3) We developed new datasets for the assessment of the proposed approaches, and conducted the experiments to assess the feasibility of the proposed approaches.

The rest of the paper can be outlined as follows. Section II formulates the MCSV problem and presents an overview on the related works, where the differences between MCSV and vehicle re-identification (VERI) are also discussed. Section III introduces the proposed MCSV approaches. Section IV illustrates experiment settings and evaluation results, followed by the conclusions

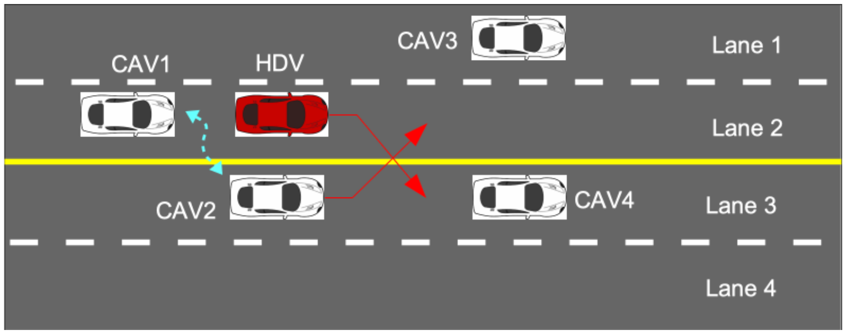


Fig. 1. Example of MCSV problem for lane merging. CAV: connected autonomous vehicle; HDV: human driven vehicle. CAV2 and CAV1 communicate and cooperate on driving. CAV1 gives ways to CAV2 through V2X communication. But for safe cooperative driving CAV2 needs to know which of CAV1 and HDV in the captured image is the communicating CAV1.

made in Section V with some identified open research issues.

II. Research Topics and Related Works

Following the discussions about the research gap on MCVS problem, in this section we will talk about the related works in the areas relevant to MCSV research.

A. MCSV Research Problem

CV uses vehicle to everything (V2X) communications, including vehicle to vehicle (V2V) communications, to exchange information with the other road users and network infrastructure. There are two mainstream V2X communication standards, i.e., IEEE 802.11p and 3GPP cellular V2X, which are dedicated to short range communications [2] [3] [4] [6]. Compared to the DSRC V2X technology, cellular V2X can provide a better communication services with low latency, high data rate and reliability [2], but requires a heavy investment on communication infrastructure.

Advanced driving assistance systems (ADAS) can support autonomous driving and reduce accidents [12]. Equipped with different sensors and advanced data processing algorithms, ADAS can warn drivers of impending danger so that the drivers can take corrective actions or even intervene on the drivers' behalf [5]. It can provide many enhanced safety features, such as blind spot detection and forward collision warning (FCW). ADAS is evolving towards self-driving autonomous vehicle, which has the highest automation level of AVs. Self-driving autonomous vehicles have three essential components, i.e., 360 degree perception, accurate mapping/positioning,

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530

and trajectory planning [5]. With advanced sensors and wireless technologies such as GPS and LiDAR, centimeter level location sensing is now possible. On the other hand, according to the latest KITTI vision benchmark results, the accuracy on detecting pedestrians and cyclists is still not high enough [12]. Thus, there are sensor-related sensing problems, which can not be solved by the AVs alone.

CVs is an important complementary to AVs for autonomous driving. Recently, various issues on CAV were studied [2] [5]. The most research works on CAV were focused on reliable and low latency V2X communications for advanced cooperative autonomous driving, cooperative data fusion, and cooperative autonomous driving applications, such as platooning and cooperative driving [2] [5] [8] [3]. A comprehensive review on the V2X communication research issues and enabling technologies for CAV applications was presented in [2]. [8] established a comprehensive benchmark for cooperative data fusion to evaluate different data fusion strategies (i.e., early, late, and intermediate fusions) and proposed a new fusion pipeline to aggregate data from multiple CAVs. [3] gave a survey on merging control strategies of CAVs at freeway on-ramps. Note that in the existing CAV research works, either the MCSV problem was not considered, or a perfect matching between the communicating and sensed vehicles was assumed.

While there are recent advances on connected autonomous driving, MCSV is a critical and still missing element to ensure safe and effective cooperation for CAVs. Without accurate MCSV, road accidents could happen during cooperative driving.

One example to illustrate MCSV is shown in Fig. 1, where there are one HDV and four CAVs. The CAVs are equipped with V2X devices and vision based sensors. CAV1 and HDV plan to merge to lane 3 and CAV2 plans to merge to lane 2. In this scenario, CAV1 and CAV2 communicate and negotiate on merging. Suppose that CAV1 informed CAV2 via V2X communication network that CAV1 will give way to CAV2 for merging. CAV1 also shared its location with CAV2 through V2X messages. From CAV2's perspective, it is aware of CAV1's locations from V2X messages obtained from GPS or HD-maps. On the other hand, CAV2 can detect surrounding vehicles and get their rough positions through camera/LiDAR sensors and deep learning models. However, both the CAVs positions obtained via V2X messages and via vision sensors have inevitable inaccuracy, e.g., around 5 meters inaccuracy in GPS positioning. The location inaccuracy makes it very challenging to align the vehicles in both V2X communication domain and object sensing domain. In Fig. 1, as CAV1 and HDV are very close, CAV2 may match the detected vehicle HDV wrongly as CAV1. As CAV1 gave way to

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530

CAV2 and CAV2 thought HDV is CAV1, CAV2 presumed that HDV would give way to itself. If HDV also decides to turn to lane 3, CAV2 could collide with HDV, leading to unsafe cooperative driving.

B. Connected Vehicle Matching

With the improved capability of GPS and V2V communications, an intuitive idea for MCSV is to use the reported position of the VBM through V2V communications and the position or distance detected by the sensors of the VPM. However, note that the position information of vehicles obtained from GPS has large errors (especially in urban scenarios), which will affect the matching accuracy. While high-performance GPS technologies may have a much lower location error, they can still have a large impact on vehicle matching. The locations of the VBMs needs to be associated to the detected objects in the sensing domain. However, accurate location sensing of objects in the captured images is an open issue for mono-camera based object detection. Even with RGB-D cameras or LiDAR, it is still challenging to obtain accurate positions of the detected vehicles.

In [13], the problem of matching connected vehicles was initially discussed for the formation of platooning. Location and distance based approaches were proposed to identify the preceding vehicles for cooperative adaptive cruise control platoon. In their approaches, radar was used to measure the distance and angle by the VPM. In the location based approaches, the position of the preceding vehicle was obtained with GPS by the VBM and passed to the VPM. The VPM compares the distance and angle measured by radar with those computed from the locations of the VBM and VPM. Note that the position error due to GPS and the inaccurate association of objects with real vehicles by radar can introduce significant reliability and safety problems in cooperative driving. In the distance based approach, ultra wideband technique was used to send the distance information measured by the VPM to the target vehicles. It was reported that a relative positioning accuracy below 1.1 m is required to ensure an acceptable identification time. However, the distance based approach has the same problem as that in the location based approach. For a vehicle matching problem where two vehicle patches are given, a projection profiles based approach was proposed in [14]. The image rows and columns were projected to generate row and column vectors, which were used as image profiles for vehicle matching. It is noted that this projection profiles based approach was used for smart camera network applications instead of MCSV, where there are many design variables which may cause reliability and effectiveness issues [14].

C. Vehicle Re-identification

Accurate traffic information plays an important role in public transport systems. There are increasing research interests in vehicle re-identification from surveillance videos [15] [16]. The task of vehicle re-identification is to identify a given vehicle from observations made by different surveillance cameras. It can be used in many applications such as collecting traffic data for statistics, law enforcement and traffic control, accident detection, etc.. Traditional sensors, such as inductive loop detection, magnetic sensors, and cellular phones, have been used for vehicle re-identification.

Computer vision based approaches become increasingly popular. There are many deep learning based methods proposed for vehicle re-identification. [16] gave an overview on VERI with sensors and computer vision based features. There are also vehicle re-identification competitions organized by NVidia since 2017 [11]. It is noted that MCSV problem was not considered in the aforementioned VERI research works. While the vehicle re-identification problem looks similar to MCSV, there are major differences between them as discussed below, and innovative solutions for MCSV task are needed.

1) In the VERI task, there is a relatively loose delay requirement, while in the MCSV task fast matching with a high accuracy is required.
2) In the VERI task, the vehicle to be re-identified does not participate in the process; while in the MCSV task VBM is aware of the process and actively cooperates with the VPM.
3) There is no communication between the vehicles to be re-identified and the executor of the VERI process. In the MCSV task, the VBM and the VPM have direct communications through V2V networks.
4) In the VERI process, images of the vehicles taken by different cameras are available for identification purpose. In the MCSV task, usually only VPMs will have images taken by their own cameras, in which the VBMs may be included. The images of the VBMs may not be available to the VPMs for identification and matching, which makes the problem even more difficult to deal with.

III. VRN AND VISUAL FEATURE BASED MCSV APPROACHES

Having discussed about the MCSV problem and introduced the related works in the previous section, we formulate the MCSV problem and present the VRN and visual feature based MCSV approaches in this section.

Assume that there are N_c CAVs and N_h HDVs in a CAV system. The CAVs are equipped with positioning devices such as GPS, a number of sensors (including cameras, radar and/or LiDAR), and V2X communication devices. The CAVs use V2X communications to broadcast their position and driving information to establish mutual awareness and to report hazard or emergency events. Some of them may also work in cooperative driving. Suppose that among the CAVs, there are one VBM and one VPM cooperating on driving (such as joining platoon and negotiating turning in unprotected intersections). At a given time t, the estimated location of VPM is $L_p(t)$ and the shared location of VBM is $L_b(t)$. Note that there are certain position estimation errors for the VPM and VBM, and there is also a latency between the position estimation for the VBM and the reception of the position information at the VPM.

We assume that VPM detected N_o objects from the sensors at time t with estimated location $L_o(n,t)$ and class $C_o(n,t)$ for the nth object, where $n \in [1, N_o]$. Again, there can be estimation errors on the location and class for the detected objects. The VPM also has auxiliary information (denoted as Aux) about the VBM, such as detected VRNs and shared VRNs of VBM. The VPM attempts to match the VBM to one of the N_o detected objects. Let I_n be an optimization variable taking value of 1 if the nth detected object is chosen to be the VBM, and 0 otherwise. Let R denote the reward of an object being successfully matched to the VBM (taking value of 1), Y denote the penalty if an object is wrongly matched to the VBM, and U denote the overall utility for the MCSV process. The value of Y is configurable and is larger than 1 (such as 50), manifesting the serious consequence of wrong vehicle matching on driving safety. Furthermore, let P_n denote the estimated probability that the nth detected object is matched to the VBM, which will be computed according to sensor's object detection performance and the Aux information.

With the above notations, we can formulate the MCSV problem as

$$\begin{aligned} &\underset{I_n, n=1,\ldots,N_o}{\text{Max}} \quad U = \sum_{n=1}^{N_o} [P_n R - (1-P_n)Y], \\ &\text{Subject to} \quad \sum_{n=1}^{N_o} I_n \leq 1, \\ &\text{Given} \qquad L_p(t), L_b(t), L_o(n,t), C_o(n,t), \text{and Aux.} \end{aligned} \tag{1}$$

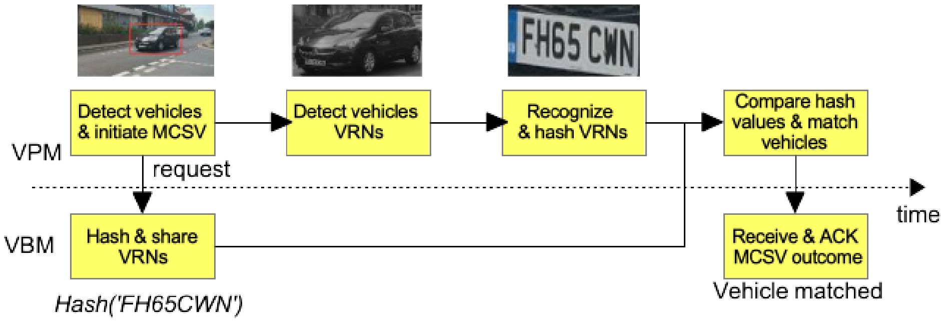
Next we will propose two MCSV approaches, in which the VRN and visual features are used as the identification information for vehicle matching at the VPMs. In this paper we assume that the VPMs will mainly use the camera sensors with optional radar and LiDAR sensors for perception. It is noted that the formulated MCSV problem takes a general form. In the VRN and visual feature-based approaches, we first use the shared location information from the VBM to modify the constraints, by reducing the set size of candidate objects to include only a very small number of objects. Those detected objects not in the reduced set of candidate objects will not be chosen as the VBM. In addition, we use the shared VRN (or visual feature) and the detected VRN (or visual feature) as the Aux auxiliary information to compute the probability that a candidate object may be the VBM to determine the VBM. The computed probabilities and the utility are used to determine if a candidate object is the VBM.

A. VRN based MCSV Approach

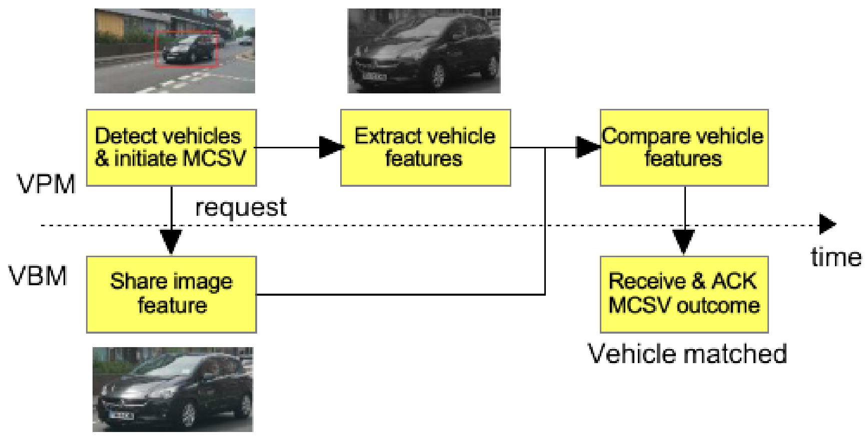
The process of the VRN based MCSV approach is shown in Fig. 2(a), which works in six steps to be explained as follows.

1) Detect vehicles and initiate MCSV (by VPM): The VPM will first detect vehicles using sensors including cameras, optional radar and LiDAR sensors. Popular deep learning models such as Faster-RCNN, YOLO-V3 and CenterNet can be used for vehicle detection. In the cooperative driving applications (such as turning and platooning), the VPM will identify the needs of cooperation with one CAV (i.e., VBM) according to the shared information such as position and pose of the CAV. The VPM knows the communication identifier (ID) and rough position estimation of the VBM, but does not know which vehicle it corresponds to among the vehicles captured in the camera image. The VPM initiates the MCSV process and requests for the VRN from the VBM.
2) Hash and share VRN (by VBM): As the VBM's communication ID is included in the message sent by the VPM, the VBM knows it is contacted for MCSV. It will respond to

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530



(a) VRN based MCSV.



(b) Visual features based MCSV.

Fig. 2. Framework of MCSV approaches. a) VRN based MCSV; and b) visual features based MCSV.

the request by hashing its VRN and sending the hash value to the VPM. The widely used SHA-512 algorithm can be used for hashing, which can protect the privacy and produce a 512-bits hash value.

3) Detect vehicle number plate (by VPM): After receiving the shared hash value from VBM, the VPM will detect vehicle number plates from the detected vehicles in the region of interests (ROI). For example, in the platooning and car-following applications, the vehicles of interest will be the preceding ones in the same lane. Using image segmentation and lane detection technologies, the VPM can locate the candidate vehicles to be processed. Then, deep learning models for object detection such as Faster-RCNN and YOLO can be used to detect the number plates in the vehicles.
4) Recognize and hash VRNs (by VPM): For those detected number plates, the VPM will use deep learning models for character recognition such as CRNN to recognize the VRNs. As shown in the example given in Fig. 2(a), a VRN of ”FH65CWN” is detected and

recognized. Then, the recognized VRNs are hashed.

5) Compare the hash values and match vehicles (by VPM): The shared VRN hash value is compared to those obtained from the detected VRNs by the VPM. If the shared hash value from the VBM matches to any computed one by the VPM, it is said that the VBM is matched to the CAV in the image with the matched hash value.
6) Receive and acknowledge MCSV outcome: The VBM will be notified of the MCSV outcome (either matched or not) by the VPM. It will acknowledge the notification of the outcome to the VPM. In the above process, it is assumed that the VPM knows the ROI and the communication ID of the candidate VBMs located in the ROI, so that it can request for the VRNs from the candidate VBMs to check if the vehicles are matched. Alternatively, if the VPM finds a need of MCSV and identifies a candidate VBM (such as the preceding vehicle in the same lane), it can recognize the VRN of the candidate VBM. Then, it broadcasts the hashed value of the VRN of the candidate VBM to surrounding CAVs and requests for a MCSV process. The CAVs receiving the message will compare the received hash value and the hash value of their own VRNs. If two hash values match, then the CAV will respond to the VPM and confirm the match of vehicles. The VPM will know the communication ID of the VBM.

B. Visual Feature based MCSV Approach

The process of the VRN based MCSV approach is shown in Fig. 2(b). Its main steps are explained as follows.

1) Detect vehicles and initiate MCSV (by VPM): This process is identical to the one in the VRN based MCSV approach. The VPM will identify the need of MCSV in the cooperative driving applications, and request MCSV for candidate CAVs located in a ROI determined by the driving applications. Visual features of the vehicle external image extracted from traditional feature descriptors such as SURF and ORB and deep learning models will be used. A specific example of the features is the raw external image. Optional information such as the vehicle pose can be included in the request message sent by the VPM to narrow down the scope of vehicles to respond.
2) Share image feature (by VBM): CAVs receiving the VPM request message will check if they are located in the ROI specified in the message. Those located in the ROI will respond it with the corresponding raw image patch of the CAVs or their image features.

Sending image features such as SURF can be a way of protecting privacy (just like using hashing) and can reduce bandwidth requirement. A possible negative side-effect is that some information about the image patches may be lost. There should be a trade-off to strike according to the available bandwidth and MCSV accuracy requirement.

3) Extract vehicle features (by VPM): Using some supporting technologies such as image segmentation and lane detection, the VPM can locate the candidate vehicles in the ROI for MCSV. Then, deep learning models for object detection such as the Faster-RCNN and YOLO can be used to detect the candidate vehicles. Either traditional handcrafted feature such as SURF and deep learning features will be extracted from the detected vehicles, depending on which image features are requested by the candidate VBMs.
4) Compare the vehicle features and match vehicles (by VPM): The shared vehicle features will be compared to those obtained from the detected candidate VBMs by the VPM. If the shared features from a VBM is close to those obtained by the VPM, it is claimed that the VBM sharing the features is matched to the candidate VBM determined by the VPM.
5) Receive and acknowledge the MCSV outcome: The VBM will be notified of the MCSV outcome (either being matched or not) by the VPM. It will acknowledge the notification of the outcome to the VPM. In the above MCSV process, it is assumed that the VPM knows the ROI of the candidate VBMs and request image features from the candidate VBMs located in the ROI. Alternatively, the VPM can identify the candidate VBM (such as the preceding vehicle in the same lane), and extract image features for the image patch of the identified candidate VBM. Then, it can share the image patch with the vehicles in the ROI and request the candidate vehicles to check if their image features match to the shared image features. If the image features of a candidate CAV match to the shared ones, then the candidate CAV will respond to the VPM and confirm the match of the vehicle. The VPM will know the communication ID of the matched vehicle.

IV. EXPERIMENTS SETTINGS AND RESULTS

MCSV is an important issue in safety critical cooperative driving applications. In this section the feasibility of the proposed MCSV approaches is assessed by a desktop computer with Intel i7 CPU. The key performance metrics include false negative rate, false positive rate, and computation time. Note that the computation times for the MCSV approaches are measured by

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530

the computer CPU and out of box software tools. The false negative rate (denoted by P_{fnr}) is defined for the detected VBMs present in the detected CAVs at the VPM, as the ratio of the number of these detected VBMs being not matched to any detected vehicles to the total number of detected VBMs. False positive rate (denoted by P_{fpr}) is defined for these detected VBMs as well, as the rate of the detected VBMs being matched to wrong CAVs detected by VPM to the total number of these detected. Let N_{all} denote the number of detected VBMs by the VPM from the cameras. Let N_{neg} denote the number of detected VBMs that are not matched to any detected CAVs and N_{pos} denote the number of detected VBMs that are matched to wrong CAVs. Then, false negative and false positive rates can be computed by $P_{fnr} = \frac{Nneg}{N_{all}}$ and $P_{fpr} = \frac{N_{pos}}{N_{all}}$. It should be noted that, in the cooperative driving applications, false positive has a much stronger impact than false negative. The false positive rate should be kept very low.

A. Experiment Settings and MCSV Dataset

In the experiments, we considered cooperative driving in two challenging scenarios, i.e., an uncontrolled intersection where there is no traffic light, and a complex urban road with T-junctions. A mixed traffic is assumed where some of the vehicles in the scenarios are CAVs with high-performance positioning and ranging devices, and some vehicles are HDVs. The CAVs use LTE V2X radios for communications and camera sensors for detecting road objects. The LTE V2X communication settings follow those used in [17], where the carrier frequency is 5.9 GHz with a bandwidth of 10 MHz. Five schedule assignment resources and two data resources are configured in each subframe. Transmit power is 23 dBm. We fix the number of CAVs in the considered scenarios to eight. A VPM is fixed at a location of the road in the scenarios. It is assumed that a ROI can be effectively determined by the VPM with the positioning and ranging devices. Information such as VRN or visual features from the CAVs in the ROI is requested and then compared with the corresponding information measured by the VPM.

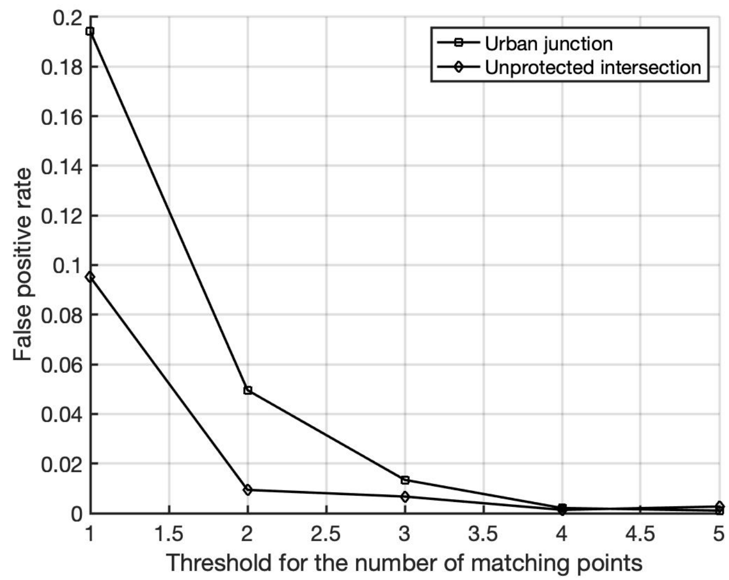
As there is no publicly available MCSV dataset, we developed a new MCSV dataset for performance evaluation of MCSV. The new dataset was developed on top of the dataset for VERI track from the AI City Challenge 2019 [11]. The AI City Challenge series are very popular for research and development of advanced AI based solutions for city challenges including vehicle re-identification and multiple object motion tracking. Video clips were provided for the VERI competition track, which were collected by surveillance cameras. The frames in the video clips are labelled with vehicle bounding boxes and ID. A given vehicle may appear in a number of frames in a specific video. A number of vehicles are chosen as CAVs to be matched from the videos. An image patch of a given CAV is extracted as the profile image for a VBM, which is denoted as VBM patch. From each frame of the tested video clips, a ROI is chosen for a given CAV. If the CAV is present in a frame, a ROI centered on the CAV with doubled size of the CAV is selected. The image patch corresponding to the ROI will be extracted for MCSV operation by the VPM. If the CAV is not present in a frame, a ROI centered on another randomly selected vehicle with doubled size of that vehicle will be selected. The image patch corresponding to the ROI will be extracted from the frame for MCSV operation by the VPM, which is denoted as VPM patch. Two videos from the AI City Challenge scenarios for VERI Track of train_S04_c029 and train_S04_c028 were selected to develop MCSV dataset, which correspond to urban junction and uncontrolled intersection scenarios, respectively.

B. Experiment Results for Visual Features based MCSV

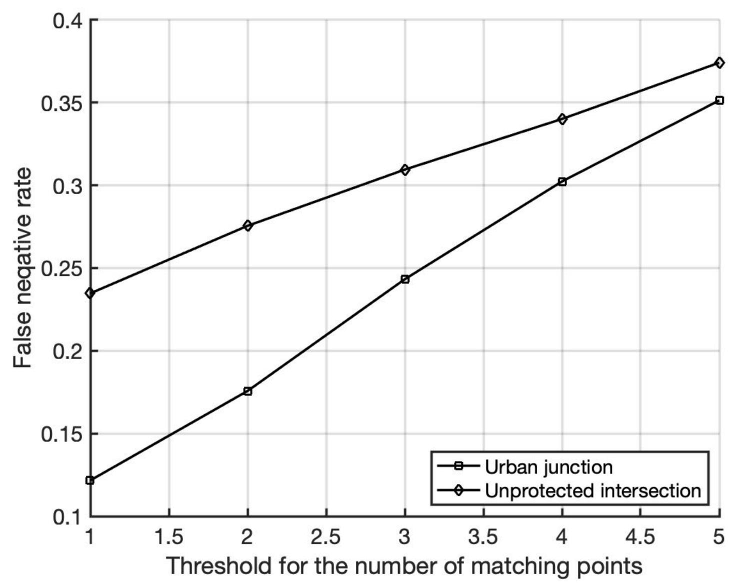
In the two selected video clips corresponding to urban junction and uncontrolled intersection scenarios, three and four CAVs are chosen as VBMs, respectively. The visual features of VBMs patches shared by VBMs via V2X are compared to the patches captured by VPM sensors. According to the visual feature comparison results, a VBM will be claimed as being matched or not by the VPM. The VPM's claimed matching results are then compared to the ground truth. Then, the MCSV performance of false negative and false positive rates can be computed. In this paper, we use the traditional SURF algorithm to extract visual features of the image patches [9], and then match the feature points of the VBM and VPM patches. A threshold of matched feature points (N_{threh}) is set, which means that only when there are no less than N_{threh} matched feature points between the VBM and VPM patches, a VBM is claimed to be matched by the VPM in the corresponding frame.

There are several considerations on using SURF features instead of more advanced deep learning features. First, SURF features can be extracted with much less computing and memory resources than deep learning features, which is important for the embedded computing devices in CAVs. Second, open software resources for computing SURF features are widely available and no training is needed. The SURF features can be shared and are compatible across CAVs. It is difficult to share and run deep learning features or models across CAVs. Third, deep learning models usually take a longer time to compute the results, which is not desirable in the real time safety critical cooperative driving applications. Based on the above considerations, SURF

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530



(a) FPR results



(b) FNR results

Fig. 3. Experiment results for visual feature based MCSV approach. (a) False positive rate; (b) False negative rate.

features are used in this paper. In our future works, deep learning will be investigated for the visual feature based MCSV approach.

The false positive and false negative rate performances of the visual feature based MCSV approach are presented in Fig. 3. The false matching performance is plotted versus the threshold

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530

of the required number of feature matching points for both urban junction and unprotected intersection scenarios. It is noted that under a light traffic load condition, V2X communication is highly reliable with an average packet successful ratio larger than 0.97. The impact of V2X communication packet losses on MCSV is very small.

It can be observed from the above results that with the visual feature based MCSV, the false positive rate decreases fast with the threshold for the number of required matching points. It can be kept very low (e.g., smaller than 0.1% with a threshold of 4), which is desirable in the safety critical cooperative driving applications. On the other hand, the false negative rate increases with the threshold. With a threshold of 4, the false negative rate in urban junction and unprotected intersection scenarios is 0.3 and 0.34, respectively. While the false negative rate is high, it is found that this is due partly to the small size of the CAVs and possible occlusions in the captured VPM patches. As false negative matching will not lead to cooperation failure on driving between the involved VBM and VPM, it will not introduce extra safety issues. Further more approaches such as using the VRN based MCSV and location based approach can be combined with the visual feature based approach to improve false negative rate performance.

For the VRN based approach, we used YOLOv3 to detect vehicles and VRNs in the images, and used the open source optical character recognition (OCR) software Tesseract to recognize the numbers and letters in the detected VRNs. As most VRNs in the video clips from the AI City Challenge were masked, we collected a dataset of driving images with cameras from a test vehicle and extracted VRNs from the collected images to develop another dataset of VRNs. For the OCR task, the original trained Tesseract model was used without fine tuning, which may not produce the optimal results.

Next, the results of false positive rate, false negative rate and latency are presented for the VRN based approach.

- As matching VRNs is relatively easy and reliable, the VRN based approach has an average false positive rate of zero, which is highly desirable in safety critical cooperative driving applications.
- On the other hand, the false negative rate performance is largely affected by the relative distance, lighting and background views. Within a relative distance of 10 meters between VPM and VBM, the average false negative rate is 38.4% with full VRN match over the collected VRN dataset. With a relative distance of about 20 meters between VPM and VBM, the average false negative rate results is 26% with full VRN match.

The latency analysis for the two proposed approaches is presented as follows, which includes communication and computation latencies.

- Communication latency: In the experiments, the MCSV data packets were broadcast at a rate of 50 packets per second. There is a communication latency of 20 ms for both visual feature and VRN based approaches, which can be reduced with a higher data transmission rate.
- Computation latency: In the object detection task, which is also needed by other autonomous driving applications, the computing time with YOLOv3 deep learning model and the given CPU is 65 ms, which can be largely reduced with GPU devices. There is some additional computing time of 45 ms for the detection and recognition of VRN with Tesseract in the VRN based MCSV. The time of computing SURF features of a vehicle patch with Matlab is 40 ms.
- Overall latency: The overall latencies (calculated as the sum of communication and computation latencies) for the VRN and visual feature based MCSV approaches are 130 ms and 125 ms, which can meet the real time requirement in the cooperative driving applications.

It is noted that the experiments were for demonstration purpose only. In practical applications, more advanced and customized software tools and GPU devices can be used to speed up the MCSV computing process and reduce the overall latency.

V. CONCLUSION

In this paper we presented and investigated an important issue on matching 5G connected vehicle to sensed vehicles for safe cooperative driving in mixed traffic conditions, which has been rarely studied. The MCSV problem was formulated and two novel MCSV approaches were proposed based on vehicle registration numbers and visual features, respectively. The proposed approaches were assessed in terms of false positive and false negative rates. A new MCSV dataset was developed from the AI City Challenge VERI dataset. In the studied urban junction and unprotected intersection scenarios, both approaches can achieve a very low false positive rate, which is desirable for safe cooperative driving. While the false negative rate is relatively high for the proposed approaches, it is not a major safety concern and can be improved by combining the proposed and additional approaches. In the future, advanced deep learning models will be used for visual feature based approach, and new MCSV datasets and performance metrics will be introduced for performance evaluation in cooperative driving applications. In addition, identifying the relevant vehicles for MCSV and improving the robustness of the MCSV approaches will be our future works.

ACKNOWLEDGEMENT

This work was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 824019 and No 101022280, Horizon Europe MSCA programme under grant agreement number 101086228, and EPSRC with RC Grant reference EP/Y027787/1.

REFERENCES

[1] Pedestrian safety: A road safety manual for decision-makers and practitioners, 2nd edition, World Health Organization, 2023.

[2] M. Noor-A-Rahim, *et al*., "6G for vehicle-to-everything (V2X) communications: enabling technologies, challenges, and opportunities", Proc of IEEE, Vol. 10, No. 6, pp. 712-734, June 2022.

[3] J. Zhu, S. Easa, and K. Gao, "Merging control strategies of connected and autonomous vehicles at freeway on-ramps: a comprehensive review", Journal of Intelligent and Connected Vehicles, Vol. 5, No.2, pp. 99-111, Dec. 2022.

[4] T. Reid *et al*., "Localization requirements for autonomous vehicles," SAE Int. J. Connected Automated Vehicles, Vol.2, No.3, pp.173-190, Sep. 2019.

[5] J. He, K. Yang, and H. Chen, "6G cellular networks and connected autonomous vehicles", IEEE Network, Vol.35, No.4, pp.255-261, July 2020.

[6] 3GPP TS 22.186 V16.2.0, Enhancement of 3GPP support for V2V scenario, Nov. 2022.

[7] Y. Gao, *et al*., "Performance analysis of robust cooperative positioning based on GPS/UWB integration for connected autonomous vehicles", IEEE Trans. on Intelligent Transport Systems, Vol.8, No.1, pp.790-802, Jan. 2023.

[8] R. Xu, *et al*., "OPV2V: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication", International Conference on Robotics and Automation (ICRA), May 2022.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Gool, "Speeded-Up Robust Features (SURF)," Computer Vision and Image Understanding, Vol. 110, No. 3, June 2008.

[10] Z. Lu, R. Lin, and H. Hu, "MART: mask-aware reasoning transformer for vehicle re-Identification", IEEE Trans. on Intelligent Transport Systems, Vol. 24, No. 2, pp.1994-2009, Feb. 2023.

[11] M. Naphade *et al*., "The 2019 AI City Challenge," IEEE CVPR Workshop, June 2019.

[12] J. Wei, *et al*., "Enhanced object detection with deep convolutional neural networks for advanced driving assistance," IEEE Trans. on Intelligent Transport Systems, Vol. 21, No. 4, pp.1572-1583, April 2019.

[13] Z. Chen and B. Park, "Preceding vehicle identification for cooperative adaptive cruise control platoon forming," IEEE Trans. Intelligent Transport Systems, Vol. 21, No. 1, pp. 308-320, Jan. 2020.

[14] V. Jelaca, *et al*., "Vehicle matching in smart camera networks using image projection profiles at multiple instances," Image and Vision Computing, Vol. 31, No. 9, pp. 673-685, Sep. 2013.

[15] R. Wei, J. Gu, S. He, and W. Jiang, "Transformer-based domain-specific representation for unsupervised domain adaptive vehicle re-identification", IEEE Trans. on Intelligent Transport Systems, Vol. 23, No. 3, pp. 2935-2946, Mar. 2023.

[16] S. Khana and H. Ullaha, "A survey of advances in vision-based vehicle re-identification," Computer Vision and Image Understanding, Vol. 182, pp. 50-63, May 2019.

[17] J. He, Z. Tang, Z. Fan, and J. Zhang, "Enhanced collision avoidance for distributed LTE vehicle to vehicle broadcast communications," IEEE Communications Letters, Vol. 22, No. 3, pp. 630-633, March 2018.

PLACE PHOTO HERE

Zuoyin Tang is currently a Lecturer at Aston University, UK. She received her PhD degree from the University of Bath, UK. She has authored or co-authored over 70 technical papers in major international journals and conferences. Her current research interests include connected and autonomous vehicles, mobile networks and computing, AI and machine learning, and intelligent transport. She is a TPC chair of IEEE Workshop on Connected Autonomous System and Applications, TPC member of international conferences including IEEE GLOBECOM, IEEE VTC, IEEE ICNP and ACM MobiArch.

PLACE PHOTO HERE

Jianhua He is a Professor at University of Essex, UK. He received his PhD degree from Nanyang Technological University. His research interests include mobile networking and computing, 5G/6G networks, Internet of Things, edge computing and intelligence, CAV, machine learning, large language models and intelligent document understanding. Dr He has published over 150 scientific papers and filed 20 patents. He served as editor for IEEE Wireless Communication Letters and Computer Journals, IEEE Access. Dr He is TPC co-chair of MobiArch2020, ICAV201 and ICASM2022.

PLACE PHOTO HERE

Kun Yang (F'2023) received his PhD from University College London (UCL), UK. He is currently a Chair Professor at University of Essex, leading the Network Convergence Laboratory (NCL), UK. His main research interests include wireless networks and communications, IoT networking, data and energy integrated networks and mobile computing. He manages research projects funded by various sources such as UK EPSRC, EU FP7/H2020 and industries. He has published 400+ papers and filed 30 patents. He serves on the editorial boards of both IEEE (e.g., IEEE TNSE, IEEE ComMag, IEEE WCL) and non-IEEE journals (e.g., Deputy EiC of IET Smart Cities). He was an IEEE ComSoc Distinguished Lecturer (2020-2021). He is a Member of Academia Europaea (MAE), a Fellow of IEEE, a Fellow of IET and a Distinguished Member of ACM.

PLACE PHOTO HERE

Hsiao-Hwa Chen (S'89-M'91-SM'00-F'10) is currently a Distinguished Professor in the Department of Engineering Science, National Cheng Kung University, Taiwan. He obtained his BSc and MSc degrees from Zhejiang University, China, and a PhD degree from the University of Oulu, Finland, in 1982, 1985, and 1991, respectively. He has authored or co-authored over 400 technical papers in major international journals and conferences, six books, and more than ten book chapters in the areas of communications. He served as the TPC Chair for IEEE Globecom 2019. He is the founding Editor-in-Chief of Wiley's Security and Communication Networks Journal. He is the recipient of the best paper award in 2021 IEEE Systems Journal and the IEEE 2016 Jack Neubauer Memorial Award. He served as the Editor-in-Chief for IEEE Wireless Communications from 2012 to 2015. He was an elected Member-at-Large of IEEE ComSoc from 2015 to 2016. He is a Fellow of IEEE, and a Fellow of IET.

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530

Figure 1.

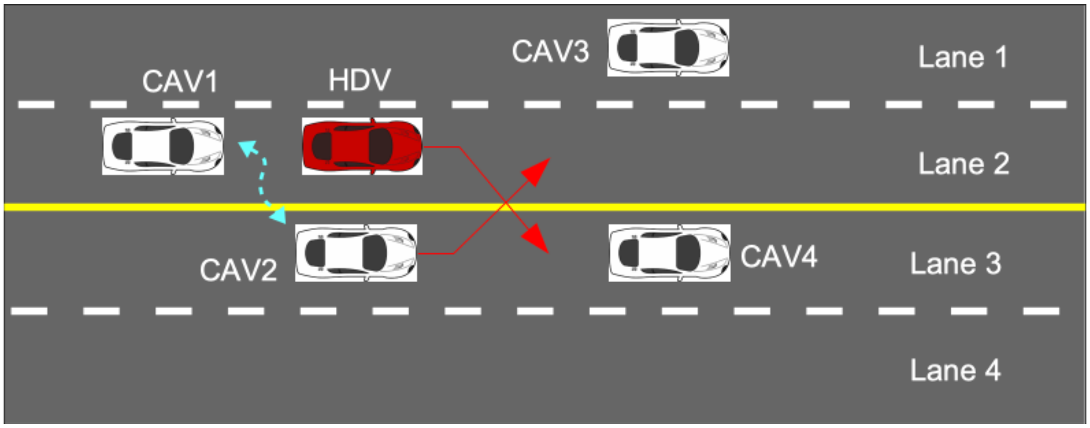


Fig. 4. Example of MCSV problem for lane merging.

Figure 2.a

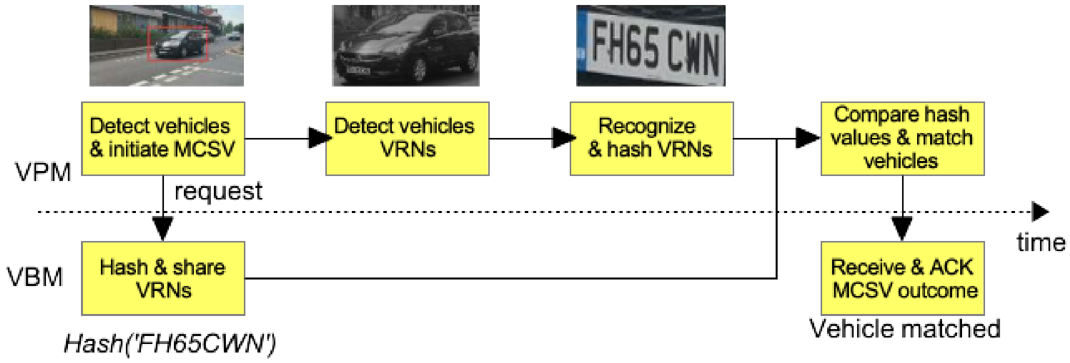


Fig. 5. VRN based MCSV.

This article has been accepted for publication in IEEE Network. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MNET.2023.3321530

Figure 2.b.

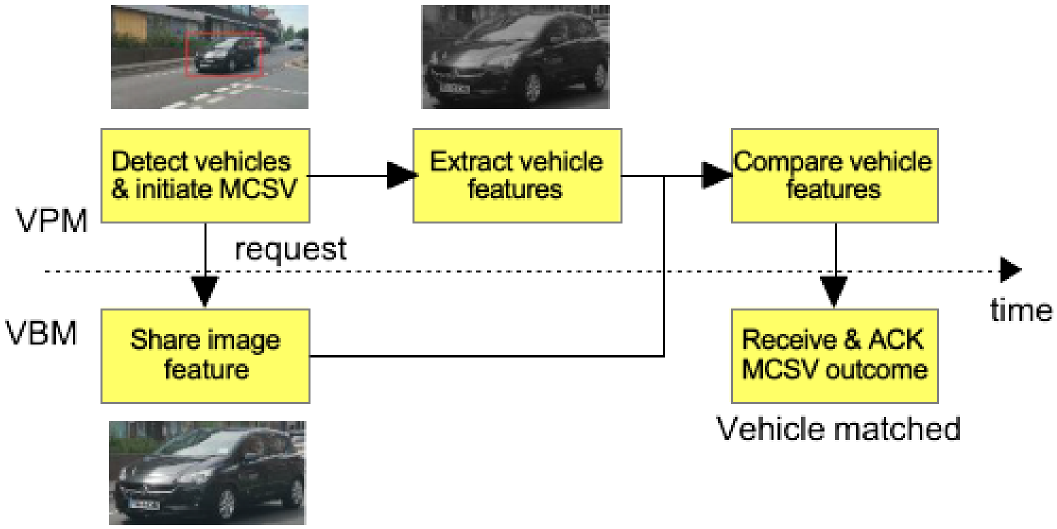


Fig. 6. Visual features based MCSV.

Figure 3.a

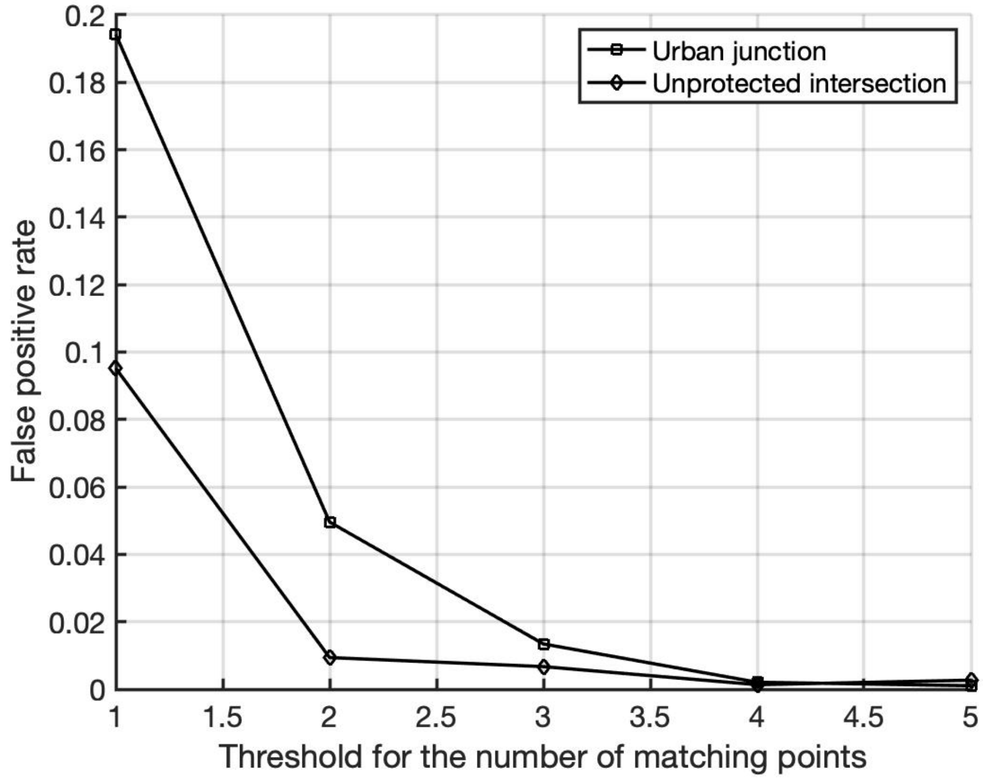


Fig. 7. FPR results.

Figure 3.b

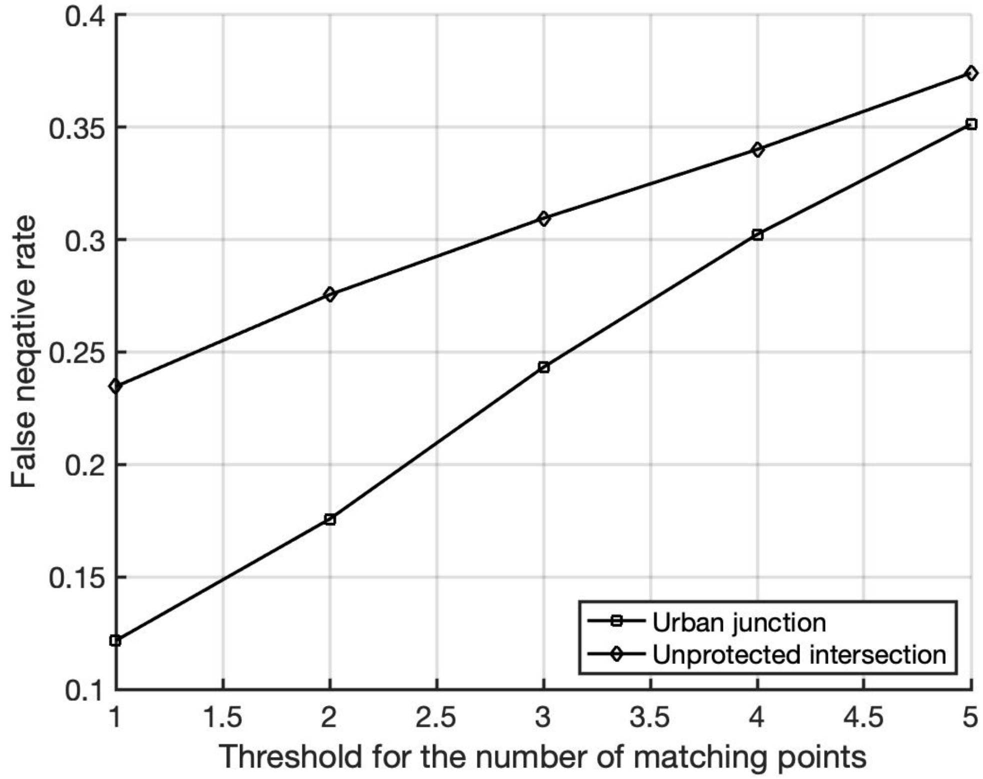


Fig. 8. FNR results.