

A survey of advances in vision-based vehicle re-identification

Sultan Daud Khan¹, Habib Ullah^{*,1}

College of Computer Science and Engineering, University of Hail, Hail, 55424, Saudi Arabia

ARTICLE INFO

Communicated by S. Ilic

MSC:

68T05

94A08

03C48

68T10

Keywords:

Re-identification

Hand-crafted methods

Convolutional neural network

Traffic analysis

ABSTRACT

Vehicle re-identification (V-reID) has become significantly popular in the community due to its applications and research significance. In particular, the V-reID is an important problem that still faces numerous open challenges. This paper reviews different V-reID methods including sensor based methods, hybrid methods, and vision based methods which are further categorized into hand-crafted feature based methods and deep feature based methods. The vision based methods make the V-reID problem particularly interesting, and our review systematically addresses and evaluates these methods for the first time. We conduct experiments on four comprehensive benchmark datasets and compare the performances of recent hand-crafted feature based methods and deep feature based methods. We present the detail analysis of these methods in terms of mean average precision (mAP) and cumulative matching curve (CMC). These analyses provide objective insight into the strengths and weaknesses of these methods. We also provide the details of different V-reID datasets and critically discuss the challenges and future trends of V-reID methods.

1. Introduction

In modern public transportation systems, video surveillance for traffic control and security plays a significant role. Therefore, obtaining accurate traffic information is increasingly in demand for many reasons including collection of statistical data (de Dios OrtAozar and Willumsen, 2011), controlling traffic signals and flows (Lämmer and Helbing, 2008), law enforcement (Vishnu et al., 2017; Maaloul et al., 2017; Saghaei et al., 2012; Hsu et al., 2013), smart transportation (Zhang et al., 2011), and urban computing (Zheng et al., 2014). In traffic areas, many surveillance cameras are already installed. It would be advantageous to use these cameras for analysis of traffic scenes with no need of replacing them with some special hardware. The data from these cameras have been used tremendously to handle the problems of vehicle detection (Ren et al., 2017; Zhou et al., 2016; Zha et al., 2007; Sun et al., 2006) and classification (Yang et al., 2015; Chen et al., 2014; Zhang, 2013). However, the problem of V-reID has escalated only over the past few years.

To re-identify a particular object, is to identify it as the same object as one observed on a previous occasion. When being presented with a vehicle of interest, V-reID tells whether this vehicle has been observed in another place by another camera. The problem of V-reID is to identify a target vehicle in different cameras with non-overlapping views as shown in Fig. 1. The emergence of V-reID can be attributed to (1) the increasing demand of public safety and (2) the widespread large camera networks in road networks, university campuses and streets.

These causes make it expensive to rely solely on brute-force human labor to accurately and efficiently spot a vehicle-of-interest or to track a vehicle across multiple cameras.

V-reID research started with sensor based methods. Several important V-reID methods have been developed since then. This development included hybrid methods and computer vision based methods. We briefly depict V-reID history in Fig. 2. Recently, much attention has been focused on the development of V-reID methods. This is evidenced by an increasing number of publications in the different venues. In Fig. 3, the percentages of papers published in both categories: sensor based methods (first row) and vision based methods (second row) are presented.

In this survey, we walk through existing research on V-reID in the hope of shedding some light on what was available in the past, what is available now, and what needs to be done in order to develop better methods that are smartly aware of the traffic environment. To the best of our knowledge, comprehensive survey on V-reID is not available. This paper fills this gap by providing comprehensive summaries and analysis of existing V-reID methods and future trends. In this survey, we begin our discussion from sensor based to deep learning based methods. Therefore, our survey is spanned by comprehensive presentation of different works starting from 1990. It is worth noting that the main focus of our paper is vision based methods. However, we briefly categorize and discuss sensor based methods for the sake of completion. We focus on different V-reID methods currently available or likely to be visible in

* Corresponding author.

E-mail address: h.ullah@uoh.edu.sa (H. Ullah).

¹ Both authors contributed equally.

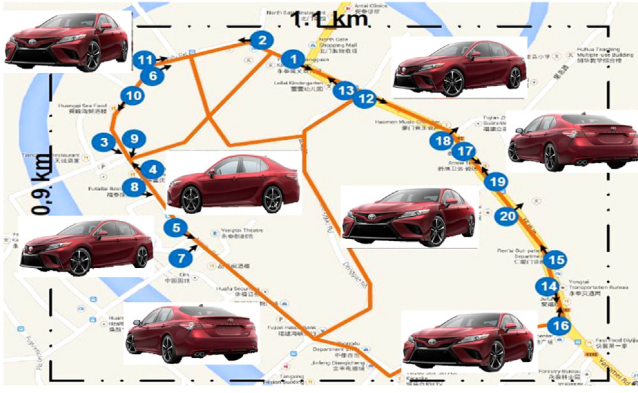


Fig. 1. Multi-view vehicle. Different views of the same vehicle are presented in the map. These views of the vehicle depict non-overlapping views of the cameras installed in different locations.

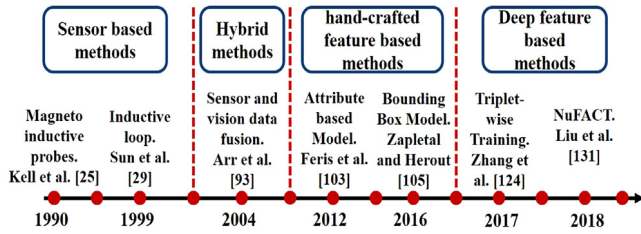


Fig. 2. V-reID methods: History of vehicle re-identification in terms of sensor based methods, hybrid methods, hand-crafted features based methods, and deep feature based methods. The hybrid methods combine image/video processing techniques with the data coming from different sensors.

the future. We have given special emphasis to deep learning methods, which are currently popular topics or reflect future trends.

The rest of the paper is organized as follows. In Section 2, we present sensor based and vision based methods. The details of the datasets are presented in Section 3. Experiments and evaluation on three benchmark datasets considering 20 different methods are shown in Section 4. The challenges and future trends of V-reID are discussed in Section 5 and the conclusion is presented in Section 6.

2. Vehicle re-identification methods

In this section we consider sensor based and vision based methods. We divide sensor based methods into five categories: magnetic sensors, inductive loop detectors, GPS-RFID-cellular phones, multi-sensor, and hybrid methods. Vision based methods are divided into two categories: hand-crafted feature based methods and deep feature based methods. These categories are depicted in Fig. 4.

2.1. Sensor based methods

Sensor based methods use vehicle signatures (Balke et al., 1995; Kuhne and Immes, 1993) for V-reID. There are several hardware detectors to extract the vehicle signatures including inductive loop, infrared, ultrasonic, microwave, magnetic, and piezoelectric sensors (Klein et al., 1997; Davies, 1986). These methods estimate the travel time of individual vehicle by matching vehicle signature detected at one location (the upstream station) with the vehicle signature detected at another location (downstream station). The locations are separated by a few hundred meters. At each location, two sensors are installed in a speed-trap configuration as depicted in Fig. 5. The speed-trap configuration calculates the speed of each detected vehicle using its pair of signatures captured by the lead and lag sensors of the traveled lane. Next we discuss different sensor based methods individually.

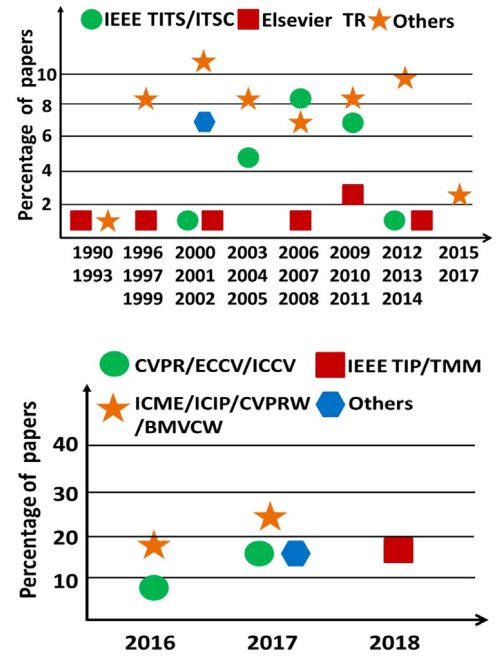


Fig. 3. Published papers: The percentage of papers related to sensor based methods are presented in the first row. For compactness, papers published in different venues in multiple years are shown. IEEE TITS/ITSC, Elsevier TR, and Others represent IEEE transactions on intelligent transportation systems/intelligent transportation systems conference, Elsevier journals of transportation research, and other journals/conferences, respectively. The percentage of papers related to vision based methods are presented in the second row. CVPR/ECCV/BMVCW, IEEE TIP/TMM, ICME/ICIP/CVPRW, and IPTA/ICTIS represent computer vision and pattern recognition conference/European conference on computer vision/British machine vision conference workshop, IEEE transactions on image processing/transactions on multimedia, international conference on multimedia and expo/international conference on image processing/CVPR workshop, and other venues, respectively.

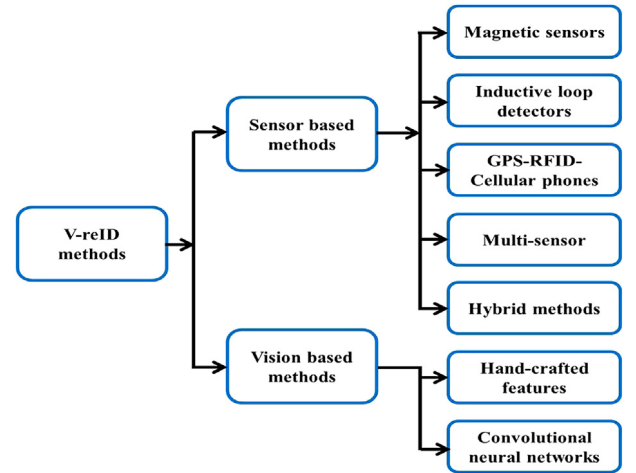


Fig. 4. Literature. We divided sensor based methods into five categories: magnetic sensors, inductive loop sensors, GPS-RFID-Cellular phones, multi-sensor, and hybrid methods. We divided vision based methods into two categories: hand-crafted feature based methods and deep feature based methods.

2.1.1. Magnetic sensors

Vehicles are composed of metallic masses that disrupt the Earth's magnetic field. Therefore, magnetic sensors generate vehicles signatures that are different from one vehicle to the other. Magnetic sensors can provide detail temporal data related to the vehicle. This temporal data can be used in travel time estimation and V-reID process. Sanchez et al. (2011) proposed a wireless magnetic sensor to

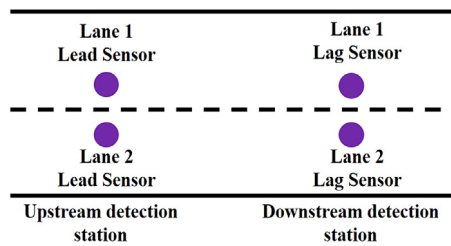


Fig. 5. Sensors configuration. Lead sensors and lag sensors are installed at the upstream and downstream detection stations, respectively.

detect the vehicle signature. Charbonnier et al. (2012) proposed a single three-axis magnetic sensor to detect a tri-dimensional magnetic signature. They extract temporal features from the sensor and then train a Gaussian maximum likelihood classifier. Kwong et al. (2009) propose a wireless magnetic sensors to extract vehicle signatures for real-time estimation of travel time across multiple intersections. Yokota et al. (1996) used ultrasonic technology and Gimeno et al. (2013) used anisotropic magneto-resistive (AMR) sensors for the V-reID. Kell et al. (1990) and Caruso and Withanawasam (1999) proposed other sensors including magnetoinductive probes also known as microloop and anisotropic magnetoresistive sensors, respectively. Kreeger and McConnell (1996) and Christiansen and Hauer (1996) proposed laser profile and weigh-in-motion axle profile (WIM) for the V-reID, respectively.

2.1.2. Inductive loop detectors

Inductive loop detectors (ILDs) are the most conventional ways for obtaining traffic data. They are widely deployed on major freeway networks. ILDs can provide various measurements like speed, length, volume, and occupancy information of a vehicle. Inductive loop signatures from vehicles are utilized by researchers Sun et al. (1999), Kuhne and Immes (1993), JENG and Chu (2013), Guilbert et al. (2013), Ndoye et al. (2008), Lunt (0000), Jeng (2007), Manual (2000), Dailey (1993) and Kwon et al. (2000) to solve the problem of V-reID. Sun et al. (1999) formulates and solves the V-reID problem as a lexicographic optimization problem. JENG and Chu (2013) proposed a real-time inductive loop signature based V-reID method also called RTREID-2M, which is based on their previous method (Jeng, 2007) also called RTREID-2. Guilbert et al. (2013) track vehicle using inductive loop detector to obtain origin–destination matrix for the vehicle. Ndoye et al. (2008) and Lunt (0000) considered inductive loop data for extracting vehicle signature and estimated travel time for V-reID. Geroliminis and Skabardonis (2006) proposed an analytical model which estimates the travel time on signalized arterials based data. They utilized flow and occupancy information provided by the ILDs. Ndoye et al. (2009) and Oh et al. (2005) proposed a signal processing framework to improve travel time estimation for V-reID. Ali et al. (2013) proposed multiple inductive loop detector system for V-reID and lane change monitoring. V-reID using the ILDs are most sensitive to speed changes in between vehicle detection stations. These systems are based on the unrealistic assumption of constant speed.

2.1.3. GPS, RFID, and cellular phones

For V-reID, some methods explore beacon based vehicle tracking and monitoring systems for locating the position of the vehicles. These systems include global positioning systems (GPS), cell phones (Smith et al., 2003), automatic vehicle identification (AVI) tags (Zhou and Mahmassani, 2006), radio frequency identification (RFID) tags (Roberts, 2006), and medium access control (MAC) addresses of blue tooth-enable devices (CONTAIN, 2008). The methods based on these systems are also called vehicle probe based methods (Turner et al., 1998). Several methods related to cell phone based measurements for traffic monitoring are reported in the literature

(Alessandri et al., 2003; Astarita and Florian, 2001; Bar-Gera, 2007; Cheng et al., 2006; Liu et al., 2008; Smith et al., 2003; Ygnace et al., 2000; Yim and Cayford, 2001; Byeong-Seo et al., 2005; Wunnava et al., 2007). Prinsloo and Malekian (2016) proposed V-reID method based on RFID tags for electronic toll collection. Mazloumi et al. (2009) proposed GPS based method for vehicle travel time estimation. The aforementioned methods considered one sensor node (GPS, cell phones, AVI, and RFID) for measurement. A single sensor node is susceptible to speed variations and environmental disturbances (Golob et al., 2004). Moreover, when the traffic flow is congested, two cars close to each other will likely be identified as a single, merged vehicle. Therefore, to determine precise vehicle status, some of researchers used multiple sensor nodes to deal with V-reID.

2.1.4. Multi-sensor

The literature on multi-sensor fusion is extensive (Waltz et al., 1990; Hall and McMullen, 2004; Yang et al., 2014; Cho et al., 2014). There are many applications of multi-sensor data fusion including image fusion (Hellwich and Wiedemann, 1999; Sharma, 1999; Hellwich and Wiedemann, 2000), medical image analysis (Leonard, 1998), and intelligent transportation systems (Klein, 2000; Dailey et al., 1996). For V-reID, it was expected to get more accurate, reliable, and comprehensive information from multi-sensor fusion. Therefore, Ndoye et al. (2011) combine signal processing technique with multi-sensor data for V-reID. Tian et al. (2014) proposed a multi-sensor spatio-temporal correlation method for V-reID. The method relied on magnetic wireless sensor network. Several signature matching methods have been proposed by Zhang and Heidemann (2011) that uniquely identify vehicle signature by matching signatures coming from different sensors. Oh et al. (2007) used sensors in conjunction with inductive loop detectors. El Faouzi (2004) applied multi-sensor fusion to the problem of road travel time estimation. Cheu et al. (2001) proposed a neural network model (NNM) (Rowley et al., 1998) for information fusion of traffic data. However, NNM requires a huge amount of data for training which is hardly available in engineering practices. Choi and Chung (2002) proposed a fusion algorithm based on fuzzy regression for travel time estimation. However, it is difficult to generalize fuzzy membership function fitting all the links of the road. To manage, analyze, and unify traffic data, linear estimation and weighted least square methods for information fusion are proposed by Zhang et al. (2005). To deal with incomplete and inaccurate traffic information coming from multi-sensor fusion, El Faouzi and Lefevre (2006) used evidence theory (Yong et al., 2004). Kong et al. (2009) combine ILDs data with GPS data. They used federated Kalman filter (Carlson and Berarducci, 1994) and evidence theory to provide a robust and flexible fusion multi sensors data. However, El Faouzi and Lefevre (2006) and Kong et al. (2009) did not consider traffic signals which affect link travel time of vehicles. To address this problem, Bhaskar et al. (2008, 2009) estimated the cumulative number of vehicles plots on the upstream and downstream of a link. Hage et al. (2011) used unscented Kalman filter (Wan and Van Der Merwe, 2001) to estimate travel time in urban areas by fusing data from ILDs with GPS sensor. Kerekes et al. (2017) proposed a multi-modal sensing method for vehicle classification and identification. They used electromagnetic emanations (Vuagnoux and Pasini, 2010), acoustic signatures (Atalar, 1979), and kernel regression (Takeda et al., 2007) to classify and identify the vehicles. Multi-sensor methods are not capable of monitoring multiple lanes. They also depend on the assumption of constant vehicle speed. Therefore, hybrid methods were investigated that combine sensor techniques with computer vision techniques.

2.1.5. Hybrid methods

Computer vision based methods are conveniently used in intelligent transportation systems. For example, Mallikarjuna et al. (2009) proposed video processing method for traffic data collection. Therefore,

researchers proposed methods (Wang et al., 2014; Ramachandran et al., 2002; Arr et al., 2004; Sun et al., 2004) for V-reID by combining image/video processing techniques with sensor data. Deng et al. (2017) proposed a V-reID method based on dynamic time warping (Müller, 2007) and magnetic signature. Wang et al. (2014) proposed a vehicle re-identification system with self-adaptive time windows to estimate the mean travel time for each time period on the freeway under traffic demand and supply uncertainty. Unlike inductive loop detectors, this method provide speed independent signature. Ramachandran et al. (2002) and Arr et al. (2004) proposed hybrid methods based on the signature captured from inductive loop detectors, vehicle velocity, traversal time, and color information based on images acquired from video cameras. Sun et al. (2004) fused data collected from inductive loop detector with color information captured from video camera.

Vehicle re-identification by matching vehicle signatures captured from inductive/ magnetic sensors are the most prominent, efficient and cost effective approaches. These approaches have several advantages over other methods. Firstly, these methods protect the privacy of the traveling public since vehicle signatures cannot be traced to individual vehicles. Secondly, probe penetration is 100% since no equipment inside the vehicle is needed. Finally, inductive/magnetic sensors are cost-effective. Despite these advantages, sensors based V-reID methods have several limitations. For example, these methods are not capable of monitoring multiple lanes. They are also having the limitation of constant speed constraint since they cannot provide speed independent signatures. The waveforms of vehicle signatures extracted from different sensors are cumbersome. Therefore, signatures matching algorithms are complex and depend on extensive calibrations. Moreover, sensor based methods cannot provide information about various features of the vehicles including color, length, and type.

Table 1 summarizes sensor based methods covered in this section in terms of their strengths and weaknesses.

2.2. Vision based methods

In computer vision, the goal of V-reID is to identify a target vehicle in multiple cameras with non-overlapping views. Due to the increase in traffic volumes on the roads and high demand in public safety, large camera network are mounted in different areas of parks, universities, streets, and other public areas. It is expensive and impractical to use traditional loop detectors or other sensors for the V-reID in such diverse environments. It is also a laborious job for security personnel to manually identify a vehicle of interest and track it across multiple cameras. Computer vision can automate the task of V-reID that can be broken down into two major modules. (1) Vehicle detection and (2) vehicle tracking through multiple cameras. Generally the first module is independent task and lots of energies have been poured to detect the vehicle in challenging and diverse environments. The challenging problem is V-reID, i.e., how to correctly match multiple images of the same vehicle under intense variations in appearance, illumination, pose, and viewpoint. The V-reID is considered as a multi camera tracking problem (Javed et al., 2008; Wang, 2013). We divided vision based methods into two categories: hand-crafted feature based methods and deep feature based methods.

2.2.1. Hand-crafted feature based methods

Hand-crafted features refer to properties derived using different methods that consider the information present in the image itself. For example, two simple features that can be extracted from images are edges and corners.

Many researchers considered appearance descriptors (Woesler, 2003; Shan et al., 2005; Ferencz et al., 2005; Shan et al., 2008; Feris et al., 2012; Zheng et al., 2015; Zapletal and Herout, 2016; Liu et al., 2016b) for V-reID. In these methods, discriminatory information is extracted from the query vehicle image. Woesler (2003) extracted 3d vehicle models and color information from the top plane of vehicle for

V-reID. Shan et al. (2005) proposed a feature vector which is composed of edge-map distances between a query vehicle image and images of other vehicles within the same camera view. A classifier was trained on the feature vectors extracted from the same and different vehicles. Ferencz et al. (2005) trained a classifier using image patches of the same and different vehicles. These image patches consisted of different features including position, edge contrast, and patch energy. Shan et al. (2008) proposed an unsupervised algorithm for matching road vehicles between two non-overlapping cameras. The matching problem is formulated as a same-different classification problem, which aims to compute the probability of vehicle images from two distinct cameras being from the same vehicle or different vehicle(s). Guo et al. (2008) and Hou et al. (2009) proposed 3D models for V-reID. They deal with large variations of pose and illumination in a better way. In the first step, pose and appearance of reference and target vehicles are estimated by 3D methods and in the second step vehicles are rendered in a normalized 3D space by making geometrically invariant comparisons of vehicles. Feris et al. (2012) proposed large-scale features using a feature pool containing hundreds of feature descriptors. Their method explicitly modeled occlusions and multiple vehicle types. Zheng et al. (2015) introduced multi-poses matching and re-ranking method for searching a car from a large-scale database. Watchar (2017) proposed V-reID method based on license] number plate (Anagnostopoulos et al., 2008). They match vehicles through their license plate using cameras with automatic number plate recognition. Zapletal and Herout (2016) also extracted 3D bounding boxes for V-reID. They used color histogram and histogram of oriented gradients followed by linear regressors. These methods (Hou et al., 2009; Guo et al., 2008; Zapletal and Herout, 2016) are computationally expensive due to constraints of 3D models. Moreover, the performance of appearance based approaches is limited due to different colors and shapes of vehicles. Other cues such as license plates and special decorations might be hardly available due to camera view, low resolution, and poor illumination of the vehicles images. To cope with these limitations, deep feature based methods for V-reID are proposed.

2.2.2. Deep feature based methods

In recent years, the success of convolutional neural networks (CNNs) (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; Szegedy et al., 2015) in different computer vision problems has inspired the research community to develop CNN based methods for vehicle recognition (Hu et al., 2015; Ramnath et al., 2014), vehicle categorization and verification (Yang et al., 2015), object categorization (Krause et al., 2013) and recognition (Xiang et al., 2015).

Liu et al. (2016b) considered large-scale bounding boxes for V-reID. They combine color, texture, and high level semantic information extracted by deep neural network. For V-reID, Liu et al. (2016c) proposed two networks: a convolutional neural network (CNN) for learning appearance attributes and a siamese neural network (SNN) for the verification of license number plates of vehicles. For learning appearance attributes, they adopted a fusion model of low-level and high-level features to find similar vehicles. The SNN was used to verify whether two license plates images belong to the same vehicle. This network is trained with large number of license number plate images for the verification. The SNN is based on deep distance metric learning (Song et al., 2016) and it simultaneously minimizes the distances of similar object pairs and maximizes the distances among different object pairs. The SNN was initially proposed by Bromley et al. (1994) for the verification of hand written signatures. It was also adopted by Chopra et al. (2005) for face verification and Zhang et al. (2016) for gait recognition for human identification. Liu et al. (2016a) proposed deep relative distance learning (DRDL) method for V-reID. They aimed to train a deep convolutional neural network considering a triplet loss function to speed up the training convergence. Their model takes raw images of vehicles and project them into a Euclidean space where L2 distance can be used directly to measure the similarity between

Table 1

Summary of sensor based methods for V-reID. We listed the strengths and weaknesses of magnetic sensors, inductive loop detectors, GPS, RFID, cellular phones, multi-sensor, and hybrid methods in the table.

Sensor based methods	Strengths	Weaknesses
Magnetic sensors	Can be used where loops are not viable (for example bridge decks). Less sensitive to stresses of traffic. Insensitive to inclement weather such as snow, rain, and fog.	Requires pavement cut or tunneling under roadway. Cannot detect stopped vehicles unless special sensor layouts or signal processing software are considered.
Inductive loop detectors	Understood and well-known technology. Provides different traffic parameters including volume, presence, occupancy, speed, headway, and gap. Resilient design to address large variety of applications.	Poor installation reduces pavement life. Multiple detectors usually required to monitor a location. Installation and maintenance require lane closure.
GPS, RFID and cellular phones	The GPS signal is available anywhere on the globe. The system gets calibrated by its own, and therefore it is easy to be used. Multiple lane operation available.	Susceptible to speed variations and environmental disturbances. In congested, two cars close to each other will likely be identified as a single car.
Multi-sensor	Combine the strengths of many sensors. Reliable even if traffic information is incomplete. Models with small detection zones do not require multiple units for full lane detection.	Not capable of monitoring multiple lanes. Depend on constant vehicle speed.
Hybrid methods	Generally cost-effective in term of hardware requirements. Easy to add and modify due to flexibility. Monitor multiple lanes.	Performance affected by inclement weather such as fog, rain, and snow. Performance also affected by vehicle shadows and vehicle projection.

two or more arbitrary vehicle images. The key idea of the DRDL was to minimize the distance between the arbitrary views of the same vehicle and maximize those of other vehicles. [Zhang et al. \(2017\)](#) proposed triplet-wise training of convolutional neural network (CNN). This training adopts triplets of query, positive example, and negative example to capture the relative similarity between vehicle images to learn representative features. They improved the triplet-wise training at two ways: first, a stronger constraint namely classification-oriented loss is augmented with the original triplet loss; second, a new triplet sampling method based on pairwise images is modeled. [Li et al. \(2017b\)](#) proposed a deep joint discriminative learning (DJDL) model, which extracts discriminative representations for vehicle images. To exploit properties and relationship among samples in different views, they modeled a unified framework to combine several different tasks efficiently, including identification, attribute recognition, verification and triplet tasks. The DJDL model was optimized jointly via a specific batch composition design. [Tang et al. \(2017\)](#) investigated that deep features and hand-crafted features are in different feature space, and if they are fused directly together, their complementary correlation is not able to be fully explored. Therefore, they proposed a multi-modal metric learning architecture to fuse deep features and hand-crafted features in an end-to-end optimization network, which achieves a more robust and discriminative feature representation for V-reID. [Cui et al. \(2017\)](#) proposed a deep neural network to fuse the classification outputs of color, model, and pasted marks on the windshield. They map them into a Euclidean space where distance can be directly used to measure the similarity of arbitrary two vehicles. [Kanaci et al. \(2017\)](#) proposed a CNN based method that exploits the potentials of vehicle model information for V-reID. The method avoids expensive and time consuming cross-camera identity pairwise labeling and relies on cheaper vehicle model. [Shen et al. \(2017\)](#) proposed a two stage framework that takes into account the complex spatio-temporal information. The method takes a pair of vehicle images with their spatio-temporal information. Each image is associated with three types of information, i.e., visual appearance, time stamp, and geo-location of the camera. By employing MRF model, a candidate visual-spatio-temporal path is generated, where each visual-spatio-temporal state corresponds to actual image with its visual-spatio-temporal information. A siamese-CNN + Path-LSTM is employed that takes candidate path and image

pair and compute their similarity score. [Wang et al. \(2017\)](#) proposed a method that incorporates orientation invariant feature with spatial-temporal information for V-reID. The method consists of two main components: the orientation invariant feature and spatial-temporal regularization component. The method starts by feeding vehicle images into a region proposal module, which computes response maps of 20 key points. The key points are then clustered into four orientation base region proposal masks. A global feature vector and four region vectors are then generated through a learning module. These features are fused together through an aggregation module that gives final orientation invariant feature vector. The second component models the spatio-temporal relations between the query and gallery images. [Liu et al. \(2018\)](#) introduced a Null space based fusion of color and attribute feature model. They adopted a Null Foley-Sammon transform (NFST) ([Guo et al., 2006](#)) based metric learning method for the fusion of features. Their model learns discriminative appearance features from multiple and arbitrary view points and reduces dimensionality of the feature space. This method utilized time and geo-location information for each query image and spatio-temporal information are exploited for every pair of images. The method works well for the pair of images that are spatially and temporally close to each other. [Zhou et al. \(2018\)](#) addressed multi-view V-reID problem by generating multi-view features for each query image which can be considered as a descriptive representation containing all information from the multiple views. The method extracts features from one image that belong to one view. Transformation models are then learned to infer features of other viewpoints. Finally, all the features from multiple viewpoints are fused together and distant metric learning is used to train a network. For the purpose of inferring features from hidden viewpoint, two end-to-end networks, spatially concatenated convolutional network and CNN-LSTM bi-directional loop (CLBL) are proposed. [Bai et al. \(2018\)](#) proposed a deep metric learning method, group sensitive triplet embedding (GSTE), to recognize and retrieve vehicles, in which intra-class variance is magnificently modeled by using an intermediate representation “group” between samples and each individual vehicle in the triplet network learning. To acquire the intra-class variance attributes of each individual vehicle, they utilized online grouping method to partition samples within each vehicle ID into a few groups, and build up the

triplet samples at multiple granularities across different vehicle IDs as well as different groups within the same vehicle ID to learn fine-grained features.

Table 2 summarizes vision based methods covered in this section in terms of feature/model, dataset, performance metric, venue, and publication year.

3. Datasets

In order to explore V-reID problem, several methods and datasets have been introduced during past couple of years. The V-reID problem not only faces challenges of enormous intra-class and least inter-class differences of vehicles but also suffers from complicated environmental factors including changes in illumination, viewpoints, scales, and camera resolutions. Therefore, in order to develop robust V-reID methods, it is important to acquire data that captures these factors effectively. A dataset should consist of sufficient amount of data so that a V-reID model can learn intra-class variability. It should also consist of huge amount of annotated data collected from a large camera network. In order to address these challenges, attempts have been made in this direction and a few datasets are collected. We discuss them individually in detail.

3.1. CompCars

Yang et al. (2015) collected CompCars dataset. It is large scale and comprehensive vehicle dataset with 214,345 images of 1716 car models. The dataset is labeled with five viewpoints including front, rear, side, front-side, and rear side. This dataset also consists of car parts as well as other attributes such as seat capacity, door number, and type of car. The images are collected from web and urban surveillance camera. Most of the images are collected from the web which cover different viewpoints and 50,000 images are captured from the surveillance camera that cover only front view. Each image is annotated with bounding box, model, and color of the car. This dataset offers four unique and important features. (1) Car hierarchy where car models are organized into a hierarchical tree of three layers, namely, car make, car model, and year of manufacturing. (2) Car attributes where each car is labeled with five attributes, namely, maximum speed, displacement, number of doors, number of seats, and type of car. (3) Viewpoints where each car is labeled with five view points. (4) Car parts where each car is labeled with eight car parts, namely, headlight, taillight, fog light, air intake, console, steering wheel, dashboard, and gear lever. This dataset was originally collected for car classification, attributes prediction, and car verification. In Fig. 6, sample images from CompCars dataset are depicted.

3.2. VehicleID

Liu et al. (2016a) collected VehicleID dataset which consists of 2,21,763 images of 26,267 cars captured from multiple non-overlapping surveillance cameras. The dataset contains car images captured from only two viewpoints (front and back) and the information about the other viewpoints is not provided. All car images are annotated with ID numbers indicating correct identities according to the car's license plate. Additionally, 90,000 images of 10,319 vehicles are also annotated with the vehicle model information. This dataset can be used for vehicle model verification, vehicle retrieval, and V-reID problems. In Fig. 7, sample images from VehicleID dataset are presented.



Fig. 6. CompCars dataset. These sample images of different vehicles are captured from the front, rear, rear side, and front viewpoints.



Fig. 7. VehicleID dataset. Each row depicts four view points for two vehicles.

3.3. BoxCars

Sochor et al. (2016) collected BoxCars dataset from 137 surveillance cameras. This dataset consists of two variants: BoxCars21K and BoxCars116K. The first variant BoxCars21K contains 63,750 images of 21,250 vehicles of 27 different models. The second variant BoxCars116K contains 1,16,826 images of 27,496 vehicles of 45 different models. The dataset contains vehicle images captured from arbitrary viewpoints, i.e., front, back, side, and roof. All vehicle images are annotated with 3D bounding box, make, model, and type. Vehicle type is annotated by tracking each vehicle across multiple cameras. Each correctly detected vehicle has three images in BoxCars21K which has been extended to 4 images per track. BoxCars dataset is designed for fine-grained vehicle model, make classification, and recognition. The images in the dataset are arranged according to the vehicle identities. Therefore, it can be also be used for V-reID problem. In Fig. 8, sample images from BoxCars dataset are shown.

3.4. VeRi-776

Liu et al. (2016c) collected VeRi-776 dataset which is the extension of VeRi dataset collected by Liu et al. (2016b). All the vehicle images in this dataset are captured in natural and unconstrained traffic environment. The dataset is collected from 20 surveillance cameras with arbitrary orientations and tilt angles. Most of the scenes include two lane roads, four lane roads, and cross roads. The dataset consists of 50,000 images of 776 cars, where each image of a vehicle is captured from 2~18 viewpoints with different illuminations and resolutions. The vehicles are labeled with bounding boxes over whole vehicle body,

Table 2

Summary of vision-based methods for V-reID. Hand-crafted feature based methods and deep feature based methods published in different venues are listed. In the table, ‘Authors collected’ means the dataset is collected by the same authors of the paper.

Category	Reference	Features/Model	Dataset	Performance metric	Venue	Year
Hand-crafted features	Woesler (2003)	3D Model Color	Wegedomstreet Author collected	Standard deviation	IEEE ITSC	2003
	Shan et al. (2005)	Edge based model	Author collected	Hit rate	IEEE ICCV	2005
	Guo et al. (2008)	3D model appearance features	Author collected	Probability of correct association	IEEE CVPR	2008
	Hou et al. (2009)	Pose and illumination model	Author collected	Hit rate	IEEE CVPR	2009
	Feris et al. (2012)	Motion shape	Authors collected	Hit rate false positive	IEEE TMM	2012
	Zheng et al. (2015)	Individual paintings matching	Car dataset	Cumulative match curve	IEEE MMSP	2015
	Zapletal and Herout (2016)	3D box color histogram	Authors collected	False positive	IEEE CVPRW	2016
Deep features	Watchar (2017)	String matching	Authors collected	Hit rate	IEEE AVSS	2017
	Liu et al. (2016c)	CNN SNN	VeRi-776 Author collected	Mean average precision (mAP)	Springer ECCV	2016
	Liu et al. (2016a)	Deep relative distance model	VehicleID CompCars	Hit rate	IEEE CVPR	
	Liu et al. (2016b)	Texture color	VeRi Author collected	Hit rate false positive	IEEE ICME	
	Wang et al. (2017)	Siamese-CNN+Path-LSTM network	VeRi-776 CompCars	mAP	IEEE CVPR	2017
	Shen et al. (2017)	Siamese-CNN+Path-LSTM network	VeRi-776	mAP	IEEE ICCV	
	Kanaci et al. (2017)	CNN	VehicleID	CMC measure	BMVC	
	Zhang et al. (2017)	CNN	VeRi	mAP	IEEE ICME	
	Tang et al. (2017)	CNN	VeRi	mAP	IEEE ICIP	
	Li et al. (2017b)	Deep joint discriminative model	VehicleID	mAP	IEEE ICIP	
	Liu et al. (2018)	Deep neural network	VeRi	mAP	IEEE TMM	2018
	Zhou et al. (2018)	CNN-LSTM	BoxCars Author collected	mAP	IEEE TIP	
	Bai et al. (2018)	GSTE	PKU-Vehicle (Author) VehicleID, VeRi, CompCars	mAP	IEEE TMM	



Fig. 8. BoxCars. Sample images of different vehicles from BoxCars dataset are shown.

type, color, and cross camera vehicle correlation. Moreover, vehicle license plate and spatial-temporal relations are also annotated for tracks of all vehicles. Due to high recurrence rate and with large number of vehicles images captured with diverse attributes, the dataset is widely used in the V-reID problem. In Fig. 9, different vehicles, from VeRi-776 dataset, in terms of color, type, model, and different viewpoints are shown.

3.5. Toy Car ReID

Zhou et al. (2018) collected Toy Car ReID dataset. This is the first synthetic vehicle dataset collected in an indoor environment using multiple cameras. The dataset contains 200 metal toy cars of common vehicle types including sedan, SUV, hatchback, van, and pickup. The

dataset is constructed with only those metal toy cars which have close resemblance with their real counter parts. In addition, changes in lighting are also incorporated to simulate the illumination changes by sun. In order to obtain densely sampled viewpoints, each vehicle is rotated by 360 degrees. The cameras are set at three different angles: 30, 60, and 90 degrees to capture images with different altitudes. In each angle, vehicle images are sampled from 50 viewpoints and then cropped to generate dataset comprising of 30,000 images. Sample images from Toy Car ReID dataset are shown in Fig. 10.

3.6. VRID-1

Li et al. (2017a) collected vehicle re-identification dataset-1 (VRID-1). This dataset consists of 10,000 images captured in daytime. There are 1000 individual vehicles. For each vehicle model, there are 100 individual vehicles, and for each of these, there are ten images captured at different locations. The images in VRID-1 are captured using 326 surveillance cameras. Therefore, there are various vehicles poses and levels of illumination. This dataset provides images of good enough quality for the evaluation of V-reID methods.

3.7. PKU-VD

The PKU-VD dataset collected by Yan et al. (2017) consists of two subdatasets VD1 and VD2 based on real-world unconstrained scenes from two cities, respectively. The images in VD1 (total 1,097,649 images) are captured from high resolution traffic cameras, and images in VD2 (total 807,260 images) are obtained from surveillance videos. For each image in both two datasets, the authors provide different annotations including identity number, precise vehicle model, and vehicle color. In the dataset, identity number is unique and all images belong to the same vehicle have the same ID. Additionally, the authors annotated 11 common colors in the dataset. Sample images from PKU-VD dataset are depicted in Fig. 11.



Fig. 9. VeRi-776 dataset. In column (a), sample images show the variations of vehicles in terms of color, type and model. In column (b), sample images show same vehicles from different viewpoints, illuminations, resolutions, and occlusions in different cameras.



Fig. 10. Toy Car ReID. Sample images of different synthetic vehicles are presented from Toy Car ReID dataset.



Fig. 11. PKU-VD. Sample images of different vehicles are presented from PKU-VD dataset.

Table 3 summarizes the datasets covered in this section in terms of venue, publication year, number of images, number of vehicles, make, number of viewpoints, V-reID Anno, and availability.

4. Experiments and evaluation

For experimental analysis, we consider 8 hand-crafted feature based methods and 12 deep feature based methods. These are recently published methods in both categories. The 8 hand-crafted feature based methods are: the 3d and color information (3DCI) Woesler (2003), the edge-map distances (EMD) (Shan et al., 2005), the 3d and piecewise model (3DPM) (Guo et al., 2008), the 3d pose and illumination model (3DPIM) (Hou et al., 2009), the attribute based model (ABM) (Feris et al., 2012), the multi-pose model (MPM) (Zheng et al., 2015), the bounding box model (BBM) (Zapletal and Herout, 2016), and the

license number plate (LNP) (Watchar, 2017). The 12 deep feature methods are: the progressive vehicle re-identification (PROVID) (Liu et al., 2016c), the deep relative distance learning (DRDL) (Liu et al., 2016a), the deep color and texture (DCT) (Liu et al., 2016b), the orientation invariant model (OIM) (Wang et al., 2017), the visual spatio-temporal model (VSTM) (Shen et al., 2017), the cross-level vehicle recognition (CLVR) (Kanaci et al., 2017), the triplet-wise training (TWT) (Zhang et al., 2017), the feature fusing model (FFM) (Tang et al., 2017), the deep joint discriminative learning (DJDL) (Li et al., 2017b), the Null space based Fusion of Color and Attribute feature (NuFACT) (Liu et al., 2018), the multi-view feature (MVF) (Zhou et al., 2018), and the group sensitive triplet embedding (GSTe) (Bai et al., 2018).

To assess the performance of the aforementioned V-reID methods, we select four comprehensive datasets including VeRi-776, VehicleID, CompCars, and PKU-VD. This selection is based on the available attributes of these datasets. These datasets consist of large scale vehicle models with multiple views. Therefore, they present reasonable information for proper comparison of different V-reID methods. Furthermore, we consider two widely used V-reID performance metrics: the mean average precision (mAP) and the cumulative matching curve (CMC) for quantitative evaluation. The mAP metric evaluates the overall performance for V-reID. The average precision is computed for each query image as formulated in Eq. (1),

$$A = \frac{\sum_{l=1}^n S(l) \times c(l)}{M_c} \quad (1)$$

where l is the rank in the order of retrieved vehicles, n is the number of fetched vehicles, M_c is the number of relevant vehicles. $S(l)$ is the precision at cut-off l in the recall list and $c(l)$ shows whether the l th recall image is correct. Therefore, the mAP is formulated in Eq. (2),

$$mAP = \frac{\sum_{u=1}^V A(u)}{V} \quad (2)$$

where V is the number of total query images. The CMC indicates the probability that a query identity occurs in different-sized candidate list. The CMC at rank k can be formulated as in Eq. (3),

$$CMC = \frac{\sum_{u=1}^V c(u, k)}{V} \quad (3)$$

where $c(u, k)$ equals 1 when the ground truth of u image occurs before rank k . The CMC evaluation is valid only if there is only one ground truth match for a given query image.

For proper comparison, we divide the training and testing samples of each dataset according to the standard protocol. VeRi-776 dataset consists of 50,000 images of 776 vehicles, in which each vehicle is captured by 2–18 cameras in different viewpoints, illuminations, resolutions and occlusions. The training set has 576 vehicles with 37,781 images and the testing set has 200 vehicles with 11,579 images. VehicleID dataset consists of 221,763 images of 26,267 vehicles. There are 110,178 images of 13,134 vehicles for training and 111,585 images of 13,133 vehicles for testing. CompCar dataset contains 136,727 vehicle images of 1687 different vehicle models. We choose the Part-I subset for training that contains 16,016 images of 431 vehicle models and

Table 3

Summary of datasets. We listed 9 datasets for V-reID problem in terms of venue, publication year, number of images, number of vehicles, make, number of viewpoints, V-reID Anno, and availability. V-reID_Anno is the annotation of ids for V-reID task.

Datasets	Venue	Year	No. of images	No. of vehicles	Images per vehicle	Make	No. of viewpoints	V-reID_Anno	Availability
CompCars	CVPR	2015	214,345	1,687	127.05	161	~5	–	On request
VehicleID	IEEE CVPR	2016	221,763	26 267	8.44	250	~2	✓	On request
VeRi	IEEE ICME	2016	40,000	619	64.62	–	~20	✓	On request
VeRi-776	Springer ECCV	2016	50,000	776	64.43	–	~20	✓	On request
BoxCars21k	IEEE CVPRW	2016	63,750	21,250	3.0	27	~4	✓	✓
BoxCars116k	IEEE ITS	2017	116,286	27 496	4.22	45	~4	✓	✓
Toy Car ReID	IEEE TIP	2018	30,000	150	200	–	~8	✓	On request
VRID-1	IEEE ITS	2017	10,000	1 000	10	10	~10	✓	On request
PKU-VD	IEEE ICCV	2017	1,904,909	2 344	2	–	~2	–	On request

the remaining 14,939 images for test. For PKU-VD dataset, we split it into training and testing sets according to Yan et al. (2017) scheme. To build the training set, we randomly choose nearly half of vehicles from each vehicle model to construct the training set. The rest of the vehicles are used to build the testing set. The numbers of intra-class groups in CompCar, VeRi-776 and VehicleID are empirically set to be 5, 5 and 2, respectively. Learning rate is divided by 10 every 15 epochs and the models are trained for 50 epochs.

In Fig. 12, we present the mAP results for the four datasets for all the methods. The hand-crafted feature based methods are annotated in red and the deep feature based methods are annotated in blue. The first, second, third, and last row list the mAP results on VeRi-776, VehicleID, CompCars, and PKU-VD datasets, respectively. The experimental results show that the GSTE (Bai et al., 2018), NuFACT (Liu et al., 2018), OIM (Wang et al., 2017), DRDL (Liu et al., 2016a), and VSTM (Shen et al., 2017) outperform the rest of the methods. They are the latest available state-of-the-art methods for V-reID. The gain of the performance of the GSTE (Bai et al., 2018) originates from the GSTE loss function, such that GSTE should be able to generalize to other network structures. In fact, the deeper networks of GSTE learn better feature representation. From the comparison, the GSTE loss based network outperforms the hand-crafted feature based methods significantly. The improvements across networks suggest that GSTE is generic work with the deep network structure to present consistently better performance in V-reID. The NuFACT (Liu et al., 2018) utilizes the multimodality features of the data. The method also considers coarse-to-fine search in the feature domain and near-to-distant search in the physical space. The orientation invariant feature embedding module of the OIM (Wang et al., 2017) handle multiple views of a vehicle efficiently. The DRDL (Liu et al., 2016a) exploits a two-branch deep convolutional network to map vehicle images into a Euclidean space. Their coupled clusters loss function and the mixed difference network structure perform key role in achieving a high predict accuracy. The VSTM (Shen et al., 2017) method based on the visual-spatio-temporal path proposal does provide vital priors for robustly estimating the vehicle similarities.

In Fig. 13, we present the CMC curves for the four datasets for all the methods. The first, second, third, and last row depict the CMC curves of VeRi-776, VehicleID, CompCars, and PKU-VD datasets, respectively. All the four rows show that the GSTE (Bai et al., 2018), NuFACT (Liu et al., 2018), OIM (Wang et al., 2017), DRDL (Liu et al., 2016a), and VSTM (Shen et al., 2017) outperform the rest of the methods. The best performance of GSTE (Bai et al., 2018) in the first and last rows shows that the introduction of the intra-class variance structure and its relevant loss function to triplet embedding bring significant improvement over the other methods. Moreover, we can observe that considering large data and deep network structure can ensure more effective network training, and generate more discriminative feature representation for fine-grained recognition. The OIM (Wang et al., 2017) outperforms all the other methods in the second row showing the results on VehicleID dataset. The OIM extracts local region features of different orientations based on 20 key point locations. They retrieve

Table 4

Computational complexity. Speed represents the complexity of each method in term of frames per seconds (fps). The complexity is presented for both deep feature (DF) based methods and hand-crafted feature (HCF) based methods.

DF based methods	Speed (fps)	HCF based methods	Speed (fps)
PROVID	08	3DCI	Real-time
DRDL	10	EMD	19
DCT	11	3DPM	13
OIM	09	3DPIM	18
VSTM	06	ABM	21
CLVR	15	MPM	Real-time
TWT	17	BBM	Real-time
FFM	14	LNP	Real-time
DJDL	12		
NuFACT	09		
MVF	13		
GSTE	10		

refined result using the log-normal distribution to model the spatial-temporal constraints. In the third row, the NuFACT (Liu et al., 2018) surpasses all the other methods considering the CompCars dataset. Thus NuFACT achieves much better improvement on CompCars than VeRi-776 and VehicleID. In fact, the fusion of color feature with semantic attributes can work better on CompCars. Moreover, during training of the null space, more information can be learned on CompCars. Thus, the NuFACT achieves greater improvement.

Both the mAP results in Fig. 12 and the CMC curves in Fig. 13 show large differences in the overall performances of hand-crafted feature based methods and deep feature based methods. The deep learning empowers the creation of complex networks, where deep layers act as a set of feature extractors that are often quite generic and, to some extent, independent of any specific task at hand. This means that deep learning extracts a set of features learned directly from observations of the input images. The clue is to discover multiple levels of representation so that higher level features can represent the semantics of the data, which in turn can provide greater robustness to intra-class variability.

To calculate the computational complexity, a 16 GB RAM computer with a 4.20 GHz CPU and a powerful NVIDIA GPU are used to perform the experiments. Further reduction in the computational complexities is possible since these implementations are not optimized. In Table 4, we provide the computational complexities for both deep feature (DF) based methods and hand-crafted feature (HCF) based methods. These complexities are presented in term of average frames per second (fps) calculated over all the datasets. Deep feature based methods are computationally more complex than hand-crafted feature based methods. Therefore, they execute less frames per seconds.

5. Challenges and future trends of V-reID

V-reID is a broad and difficult problem with numerous open issues (Gong et al., 2014). In this section we discussed the general problems of V-reID and its broader challenges.

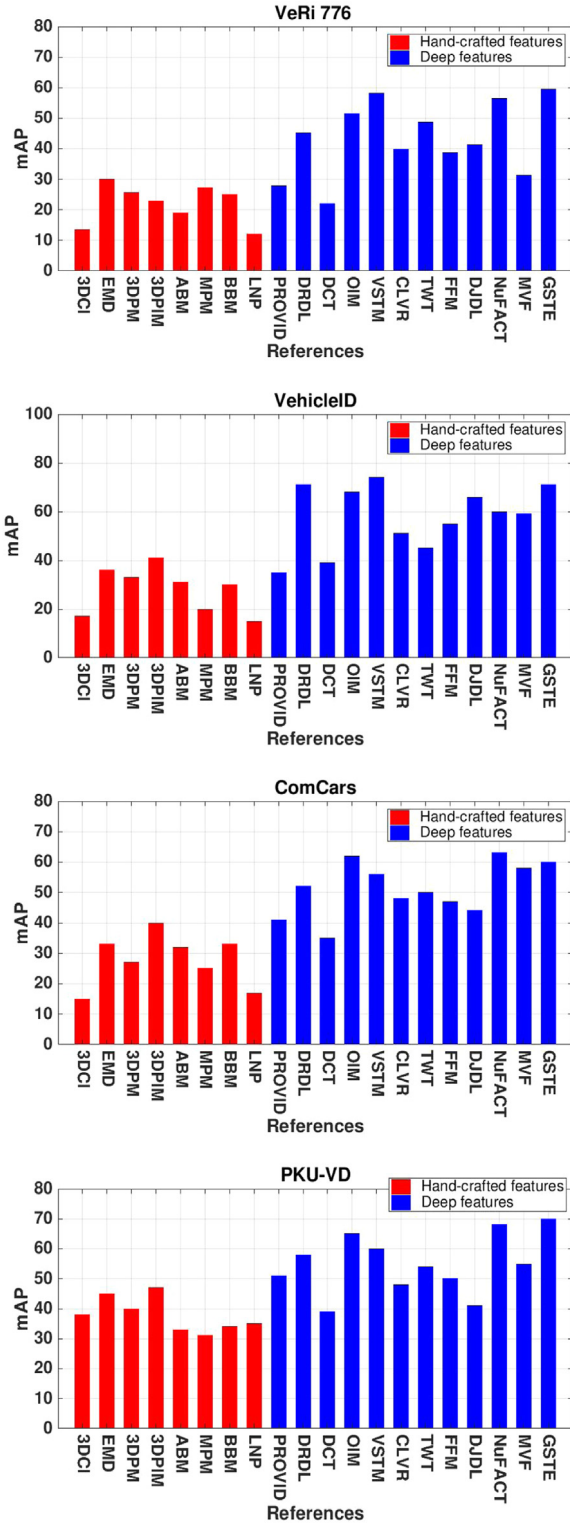


Fig. 12. Results based on mAP. The first, second, third, and last row show the results in term of mAP on VeRi-776, VehicleID, CompCars, and PKU-VD datasets, respectively.

Inter- and Intra-class variations: A basic challenge of any V-reID method is to cope with the inter-class and intra-class variations. Considering inter-class variation, different vehicles can look alike across camera views. Two vehicles from the same viewpoint always look more alike than two different viewpoints of one vehicle by machine vision. Considering intra-class variation, the same vehicle from different sides

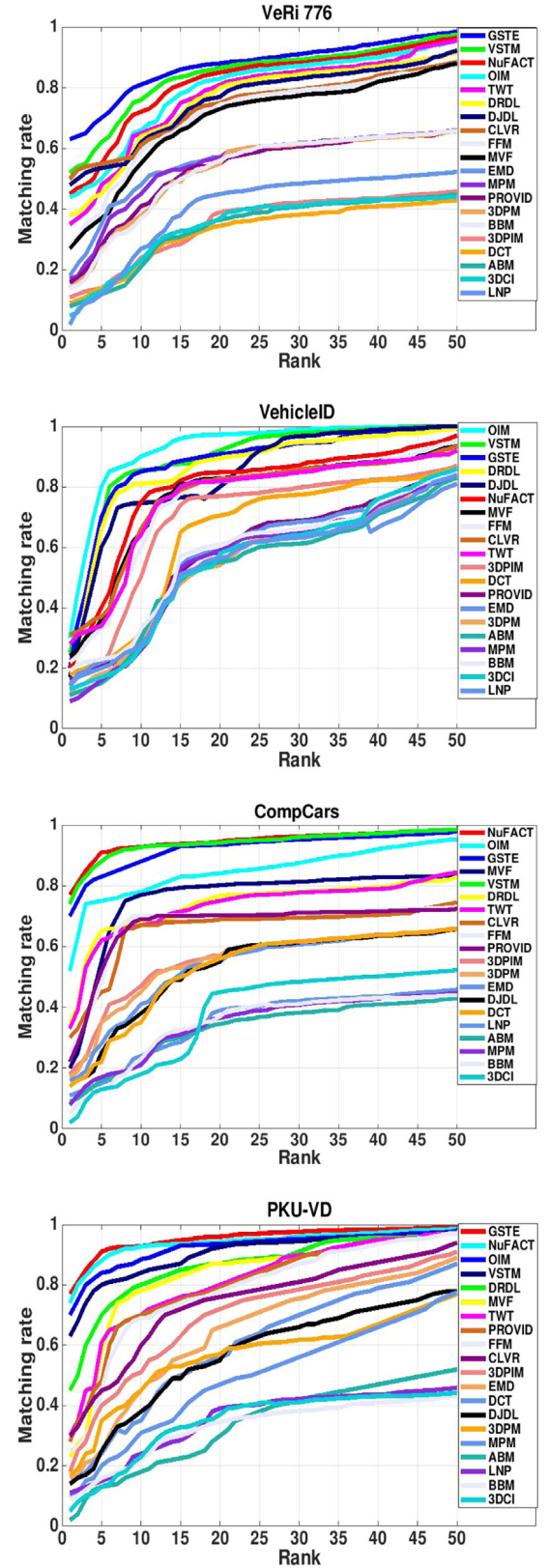


Fig. 13. CMC curves. The first, second, third, and last row show the results on VeRi-776, VehicleID, CompCars, and PKU-VD datasets, respectively.

looks different when observed under different camera views. Such variations between camera view pairs are in general complex and multi-modal. Visual features of vehicles with varying viewpoints from

different cameras have significant differences. There is no overlap of the visual features for a vehicle across the front, side, rear, and top viewpoints. The chassis of a vehicle is not upright like a person, thus the texture or color will change severely in different views. Therefore, the appearance variation across different viewpoints of a vehicle is much larger. Additionally, different views of a vehicle from multiple cameras leads to deformable shapes of silhouettes and different geometries.

Data requirements: In general a V-reID method may be required to match single vehicle image to a gallery of images. This means there is likely to be insufficient data to learn a good model of each vehicle's intra-class variability. The datasets we discussed hardly reflect the challenges of real world surveillance. For example, vehicle has to be tracked through a large number of cameras with overlapping and non-overlapping views in real world surveillance. The currently available datasets are collected from non-overlapping views with limited number of cameras, which capture less variations in resolution and viewpoints. Due to unavailability of unconstrained data, the impact of integrating and exploiting temporal information cannot be considered. However, this type of information is important for learning inter-camera relations which can increase the efficiency of V-reID methods by suppressing the false positives. Moreover, multiple vehicles re-identifications are not considered in the current datasets.

Generalization capability: This is the other side of the training data scalability. If a model is trained for a specific pair of camera, it does not generalize well to another pair of cameras with different viewing conditions. Good generalization ability of a model is generally desirable that can be trained once and then applied to a variety of different camera configurations from different locations. This would avoid the issue of training data scalability.

Long-term V-reID: The longer the time and space separation between views, the greater the chance that vehicles may appear with some changes. In fact, the nature of separation between the cameras determines the level of difficulty in V-reID system. For example, the V-reID modeling could be easy if the two images are taken from similar views only a few minutes apart. However, if the images/video are taken from different views hours apart, the modeling could not be easy due to variations in illumination and view angles. This highlights the sensitive nature of the V-reID modeling. The temporal segregation between the images is a key factor in the complexity of the V-reID system. Therefore, a V-reID method should have some robustness. The current datasets cannot address long-term V-reID problem due to the unavailability of long duration videos, recorded over several days using the same or different set of cameras.

Other challenges: To develop a V-reID model, features are extracted from an image input or a video input. In the former case, a V-reID system detects and localize a vehicle in an image. In the latter case, the system establishes correspondence between the detected vehicles across multiple frames to ensure that the features belong to the same vehicle of interest. This procedure is called tracking that caters a consistent label to each vehicle in multiple frames. The V-reID system uses multiple instances of a vehicle for feature extraction and subsequent descriptor generation for V-reID. However, vehicle detection and multiple vehicle tracking are different problems with their own challenges. Moreover, it is difficult to model discriminative visual descriptors since vehicles can be partially or completely occluded due to crowded traffic or clutter. It is also difficult to control factors such as resolution, frame rate, imaging conditions, and imaging angles. Therefore, extracting a unique and discriminative descriptor is dependent upon availability of good quality observations. Additionally, the quality of the V-reID system is affected due to incorrect detection and trajectory estimation.

Despite these challenges, the performance of deep feature based methods is noticeable. One of the main benefits of deep learning over other methods is its ability to generate new features from limited series of features located in the training dataset. These methods use more complex sets of features in comparison with hand-crafted feature

based methods. In fact, deep learning generates actionable results when solving the V-reID problem. The performance of deep feature based methods can be further improved if large amount of training data is provided.

6. Conclusions

In this paper we presented the problem of vehicle re-identification and an overview of current research in the computer vision community. We discussed sensor-based and vision based methods in detail. We categorized vision based methods into hand-crafted feature based methods and deep feature based methods. The details of different datasets are presented and summarized. We also presented results on three benchmark datasets considering 20 different V-reID methods including both hand-crafted feature based and deep feature based methods. We highlighted open issues and challenges of the V-reID problem with the discussion on potential directions for further research.

In our future work, we would collect large scale real surveillance multi-view vehicle datasets to improve the training of the state-of-the-art models for performance enhancement.

Acknowledgments

This work is supported by the University of Hail, Saudi Arabia. Portions of the research in this paper use the PKU-VD dataset collected under the sponsor of the National Basic Research Program of China and the National Natural Science Foundation of China.

References

- Alessandri, A., Bolla, R., Repetto, M., 2003. Estimation of freeway traffic variables using information from mobile phones. In: American Control Conference, 2003. Proceedings of the 2003, Vol. 5. IEEE, pp. 4089–4094.
- Ali, S.S.M., George, B., Vanajakshi, L., 2013. Multiple inductive loop detectors for intelligent transportation systems applications: Ramp metering, vehicle re-identification and lane change monitoring systems. In: Computers & Informatics (ISCI), 2013 IEEE Symposium on. IEEE, pp. 176–180.
- Anagnostopoulos, C.N.E., Anagnostopoulos, I.E., Psoroulas, I.D., Loumos, V., Kayafas, E., 2008. License plate recognition from still images and video sequences: A survey. IEEE Trans. Intell. Transp. Syst. 9 (3), 377–391.
- Arr, G., Sun, C., Ramachandran, R., 2004. Fusion of wavelet transform and color information features for automatic vehicle reidentification in intelligent transportation systems. In: Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, Vol. 5. IEEE, pp. V–285.
- Astarita, V., Florian, M., 2001. The use of mobile phones in traffic management and control. In: Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE, pp. 10–15.
- Atalar, A., 1979. A physical model for acoustic signatures. J. Appl. Phys. 50 (12), 8237–8239.
- Bai, Y., Lou, Y., Gao, F., Wang, S., Wu, Y., Duan, L., 2018. Group sensitive triplet embedding for vehicle re-identification. IEEE Trans. Multimed..
- Balke, K.N., Ullman, G., McCasland, W., Mountain, C., Dudek, C., 1995. Benefits of real-time travel information in Houston, Texas. Technical Report.
- Bar-Gera, H., 2007. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from israel. Transp. Res. C 15 (6), 380–391.
- Bhaskar, A., Chung, E., Dumont, A.G., 2009. Estimation of travel time on urban networks with midlink sources and sinks. Transp. Res. Rec. J. Transp. Res. Board (2121), 41–54.
- Bhaskar, A., Chung, E., de Mouzon, O., Dumont, A.G., 2008. Analytical modeling and sensitivity analysis for travel time estimation on signalized urban networks.
- Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R., 1994. Signature verification using a “siamese” time delay neural network. In: Advances in Neural Information Processing Systems. pp. 737–744.
- Byeong-Seo, Y., Kang, S.P., Park, C.H., 2005. Travel time estimation using mobile data. In: Proceedings of Eastern Asia Society for Transportation Studies, Vol. 5. Citeseer, pp. 1533–1547.
- Carlson, N.A., Berarducci, M.P., 1994. Federated kalman filter simulation results. Navigation 41 (3), 297–322.
- Caruso, M.J., Withanawasam, L.S., 1999. Vehicle detection and compass applications using amr magnetic sensors. In: Sensors Expo Proceedings, Vol. 477. p. 39.

- Charbonnier, S., Pitton, A.C., Vassilev, A., 2012. Vehicle re-identification with a single magnetic sensor. In: *Instrumentation and Measurement Technology Conference (I2MTC)*, 2012 IEEE International. IEEE, pp. 380–385.
- Chen, P., Bai, X., Liu, W., 2014. Vehicle color recognition on urban road by feature context. *IEEE Trans. Intell. Transp. Syst.* 15 (5), 2340–2346.
- Cheng, P., Qiu, Z., Ran, B., 2006. Particle filter based traffic state estimation using cell phone network data. In: *Intelligent Transportation Systems Conference*, 2006. ITSC'06. IEEE. IEEE, pp. 1047–1052.
- Cheu, R.L., Lee, D.H., Xie, C., 2001. An arterial speed estimation model fusing data from stationary and mobile sensors. In: *Intelligent Transportation Systems*, 2001. Proceedings. 2001 IEEE. IEEE, pp. 573–578.
- Cho, H., Seo, Y.W., Kumar, B.V., Rajkumar, R.R., 2014. A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In: *Robotics and Automation (ICRA)*, 2014 IEEE International Conference on. IEEE, pp. 1836–1843.
- Choi, K., Chung, Y., 2002. A data fusion algorithm for estimating link travel time. *ITS J.* 7 (3–4), 235–260.
- Chopra, S., Hadsell, R., LeCun, Y., 2005. Learning a similarity metric discriminatively, with application to face verification. In: *Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE Computer Society Conference on, Vol. 1. IEEE, pp. 539–546.
- Christiansen, I., Hauer, L.E., 1996. Probing for travel time: Norway applies avi and wim technologies for section probe data. *Traffic Technol. Int.*
- CONTAIN, T.A., 2008. Real-time travel time estimates using media access control address matching. *ITE J.*
- Cui, C., Sang, N., Gao, C., Zou, L., 2017. Vehicle re-identification by fusing multiple deep neural networks. In: *Image Processing Theory, Tools and Applications (IPTA)*, 2017 Seventh International Conference on. IEEE, pp. 1–6.
- Dailey, D.J., 1993. Travel-time estimation using cross-correlation techniques. *Transp. Res. B* 27 (2), 97–107.
- Dailey, D.J., Harn, P., Lin, P.J., 1996. ITS data fusion. Technical Report. Washington State Department of Transportation.
- Davies, P., 1986. Vehicle detection and classification. *Inf. Technol. Appl. Transp.* 11–40.
- Deng, G., Zhou, H., Yu, G., Yan, Z., Hu, Y., Xu, X., 2017. Scalable and parameterized dynamic time warping architecture for efficient vehicle re-identification. In: *Transportation Information and Safety (ICTIS)*, 2017 4th International Conference on. IEEE, pp. 48–53.
- de Dios OrtAozar, J., Willumsen, L.G., 2011. *Modelling Transport*. John Wiley Sons.
- El Faouzi, N.E., 2004. Data-driven aggregative schemes for multisource estimation fusion: A road travel time application. In: *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2004*, Vol. 5434. International Society for Optics and Photonics, pp. 351–360.
- El Faouzi, N.E., Lefevre, E., 2006. Classifiers and distance-based evidential fusion for road travel time estimation. In: *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006*, Vol. 6242. International Society for Optics and Photonics, p. 62420A.
- Ferencz, A., Learned-Miller, E.G., Malik, J., 2005. Building a classification cascade for visual identification from one example. In: *Computer Vision*, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 1. IEEE, pp. 286–293.
- Feris, R.S., Siddiquie, B., Petterson, J., Zhai, Y., Datta, A., Brown, L.M., Pankanti, S., 2012. Large-scale vehicle detection, indexing, and search in urban surveillance videos. *IEEE Trans. Multimed.* 14 (1), 28–42.
- Geroliminis, N., Skabardonis, A., 2006. Real time vehicle reidentification and performance measures on signalized arterials. In: *Intelligent Transportation Systems Conference*, 2006. ITSC'06. IEEE. IEEE, pp. 188–193.
- Gimeno, R.V.C., Celda, A.G., Pla-Castells, M., Plume, J.M., 2013. Improving similarity measures for re-identification of vehicles using amr sensors. In: *Information, Communications and Signal Processing (ICICS)* 2013 9th International Conference on. IEEE, pp. 1–5.
- Golob, T.F., Recker, W.W., Alvarez, V.M., 2004. Freeway safety as a function of traffic flow. *Accid. Anal. Prev.* 36 (6), 933–946.
- Gong, S., Cristani, M., Loy, C.C., Hospedales, T.M., 2014. The re-identification challenge. In: *Person Re-Identification*. Springer, pp. 1–20.
- Guilbert, D., Le Bastard, C., Ieng, S.S., Wang, Y., 2013. Re-identification by inductive loop detector: Experimentation on target origindestination matrix. In: *Intelligent Vehicles Symposium (IV)*, 2013 IEEE. IEEE, pp. 1421–1427.
- Guo, Y., Rao, C., Samarasekera, S., Kim, J., Kumar, R., Sawhney, H., 2008. Matching vehicles under large pose transformations using approximate 3d models and piecewise mrf model. In: *Computer Vision and Pattern Recognition*, 2008. CVPR 2008. IEEE Conference on. IEEE, pp. 1–8.
- Guo, Y.F., Wu, L., Lu, H., Feng, Z., Xue, X., 2006. Null foley–sammon transform. *Pattern recognition* 39 (11), 2248–2251.
- Hage, R.M., Le, E., Peyret, F., Meizel, D., et al., 2011. Link travel time estimation in urban areas by detectors and probe vehicles fusion. In: *2nd International Conference on Models and Technologies for Intelligent Transportation Systems*.
- Hall, D.L., McMullen, S.A., 2004. *Mathematical Techniques in Multisensor Data Fusion*. Artech House.
- Hellwich, O., Wiedemann, C., 1999. Multisensor data fusion for automated scene interpretation. In: *Image and Signal Processing for Remote Sensing V*, Vol. 3871. International Society for Optics and Photonics, pp. 284–296.
- Hellwich, O., Wiedemann, C., 2000. Object extraction from high-resolution multisensor image data. In: *Third International Conference Fusion of Earth Data*, Sophia Antipolis, Vol. 115.
- Hou, T., Wang, S., Qin, H., 2009. Vehicle matching and recognition under large variations of pose and illumination. In: *Computer Vision and Pattern Recognition Workshops*, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. IEEE, pp. 24–29.
- Hsu, C.Y., Kang, L.W., Liao, H.Y.M., 2013. Cross-camera vehicle tracking via affine invariant object matching for video forensics applications. In: *Multimedia and Expo (ICME)*, 2013 IEEE International Conference on. IEEE, pp. 1–6.
- Hu, C., Bai, X., Qi, L., Chen, P., Xue, G., Mei, L., 2015. Vehicle color recognition with spatial pyramid deep learning. *IEEE Trans. Intell. Transp. Syst.* 16 (5), 2925–2934.
- Javed, O., Shafique, K., Rasheed, Z., Shah, M., 2008. Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. *Comput. Vis. Image Underst.* 109 (2), 146–162.
- Jeng, S.T., 2007. Real-time vehicle reidentification system for freeway performance measurements.
- JENG, S.T.C., Chu, L., 2013. Vehicle reidentification with the inductive loop signature technology. *J. East. Asia Soc. Transp. Stud.* 10, 1896–1915.
- Kanaci, A., Zhu, X., Gong, S., 2017. Vehicle re-identification by fine-grained cross-level deep learning. In: *BMVC AMMDS Workshop*, Vol. 2.
- Kell, J.H., Fullerton, I.J., Mills, M.K., 1990. *Traffic detector handbook*. Technical Report.
- Kerekes, R.A., Karnowski, T.P., Kuhn, M., Moore, M.R., Stinson, B., Tokola, R., Anderson, A., Vann, J.M., 2017. Vehicle classification and identification using multi-modal sensing and signal learning. In: *Vehicular Technology Conference (VTC Spring)*, 2017 IEEE 85th. IEEE, pp. 1–5.
- Klein, L.A., 2000. Dempster-shafer data fusion at the traffic management center. In: *Transportation Research Board 79th Annual Meeting*.
- Klein, L.A., Kelley, M.R., Mills, M.K., 1997. Evaluation of overhead and in-ground vehicle detector technologies for traffic flow measurement. *J. Test. Eval.* 25 (2), 205–214.
- Kong, Q.-J., Li, Z., Chen, Y., Liu, Y., 2009. An approach to urban traffic state estimation by fusing multisource information. *IEEE Trans. Intell. Transp. Syst.* 10 (3), 499–511.
- Krause, J., Stark, M., Deng, J., Fei-Fei, L., 2013. 3d object representations for fine-grained categorization. In: *Computer Vision Workshops (ICCVW)*, 2013 IEEE International Conference on. IEEE, pp. 554–561.
- Kreeger, K., McConnell, R., 1996. Structural range image target matching for automated link travel time computation. In: *Intelligent Transportation: Realizing the Future. Abstracts of the Third World Congress on Intelligent Transport Systems ITS America*.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. pp. 1097–1105.
- Kuhne, R.D., Immes, S., 1993. Freeway control systems for using section-related traffic variable detection. In: *Pacific Rim TransTech Conference (1993: Seattle, Wash.)*. Proceedings Pacific Rim TransTech Conference. Vol. 1.
- Kwon, J., Coifman, B., Bickel, P., 2000. Day-to-day travel-time trends and travel-time prediction from loop-detector data. *Transp. Res. Rec. J. Transp. Res. Board* (1717), 120–129.
- Kwong, K., Kavalier, R., Rajagopal, R., Varaiya, P., 2009. Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors. *Transp. Res. C* 17 (6), 586–606.
- Lämmer, S., Helbing, D., 2008. Self-control of traffic lights and vehicle flows in urban road networks. *J. Stat. Mech. Theory Exp.* 2008 (04), P04019.
- Leonard, J., 1998. Sensor fusion for surgical applications. In: *Dayton Section Symposium*, 1998. The 15th Annual AESS/IEEE. IEEE, pp. 37–44.
- Li, Y., Li, Y., Yan, H., Liu, J., 2017b. Deep joint discriminative learning for vehicle re-identification and retrieval.
- Li, X., Yuan, M., Jiang, Q., Li, G., 2017a. Vrid-1: A basic vehicle re-identification dataset for similar vehicles. In: *Intelligent Transportation Systems (ITSC)*, 2017 IEEE 20th International Conference on. IEEE, pp. 1–8.
- Liu, H., Danczyk, A., Brewer, R., Starr, R., 2008. Evaluation of cell phone traffic data in minnesota. *Transp. Res. Rec. J. Transp. Res. Board* (2086), 1–7.
- Liu, X., Liu, W., Ma, H., Fu, H., 2016b. Large-scale vehicle re-identification in urban surveillance videos. In: *Multimedia and Expo (ICME)*, 2016 IEEE International Conference on. IEEE, pp. 1–6.
- Liu, X., Liu, W., Mei, T., Ma, H., 2016c. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In: *European Conference on Computer Vision*. Springer, pp. 869–884.
- Liu, X., Liu, W., Mei, T., Ma, H., 2018. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Trans. Multimed.* 20 (3), 645–658.

- Liu, H., Tian, Y., Yang, Y., Pang, L., Huang, T., 2016a. Deep relative distance learning: Tell the difference between similar vehicles. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2167–2175.
- Lunt, G., 0000. Vehicle re-identification using induction loop data.
- Maaloul, B., Taleb-Ahmed, A., Niar, S., Harb, N., Valderrama, C., 2017. Adaptive video-based algorithm for accident detection on highways. In: *Industrial Embedded Systems (SIES)*, 2017 12th IEEE International Symposium on. IEEE, pp. 1–6.
- Mallikarjuna, C., Phanindra, A., Rao, K.R., 2009. Traffic data collection under mixed traffic conditions using video image processing. *J. Transp. Eng.* 135 (4), 174–182.
- Manual, H.C., 2000. Highway Capacity Manual. Washington, DC, p. 11.
- Mazloumi, E., Currie, G., Rose, G., 2009. Using gps data to gain insight into public transport travel time variability. *J. Transp. Eng.* 136 (7), 623–631.
- Müller, M., 2007. Dynamic time warping. In: *Information Retrieval for Music and Motion*. Springer, pp. 69–84.
- Ndoye, M., Totten, V., Carter, B., Bullock, D.M., Krogmeier, J.V., 2008. Vehicle detector signature processing and vehicle reidentification for travel time estimation. Technical Report.
- Ndoye, M., Totten, V., Krogmeier, J.V., Bullock, D.M., 2009. A signal processing framework for vehicle reidentification and travel time estimation. In: *Intelligent Transportation Systems*, 2009. ITSC'09. 12th International IEEE Conference on. IEEE, pp. 1–6.
- Ndoye, M., Totten, V.F., Krogmeier, J.V., Bullock, D.M., 2011. Sensing and signal processing for vehicle reidentification and travel time estimation. *IEEE Trans. Intell. Transp. Syst.* 12 (1), 119–131.
- Oh, C., Ritchie, S.G., Jeng, S.T., 2007. Anonymous vehicle reidentification using heterogeneous detection systems. *IEEE Trans. Intell. Transp. Syst.* 8 (3), 460–469.
- Oh, C., Tok, A., Ritchie, S.G., 2005. Real-time freeway level of service using inductive-signature-based vehicle reidentification system. *IEEE Trans. Intell. Transp. Syst.* 6 (2), 138–146.
- Prinsloo, J., Malekian, R., 2016. Accurate vehicle location system using rfid, an internet of things approach. *Sensors* 16 (6), 825.
- Ramachandran, R.P., Arr, G., Sun, C., Ritchie, S.G., 2002. A pattern recognition and feature fusion formulation for vehicle reidentification in intelligent transportation systems. In: *Acoustics, Speech, and Signal Processing (ICASSP)*, 2002 IEEE International Conference on, Vol. 4. IEEE, pp. IV–3840.
- Ramnath, K., Sinha, S.N., Szeliski, R., Hsiao, E., 2014. Car make and model recognition using 3d curve alignment. In: *Applications of Computer Vision (WACV)*, 2014 IEEE Winter Conference on. IEEE, pp. 285–292.
- Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6), 1137–1149.
- Roberts, C.M., 2006. Radio frequency identification (rfid). *Comput. Secur.* 25 (1), 18–26.
- Rowley, H.A., Baluja, S., Kanade, T., 1998. Neural network-based face detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (1), 23–38.
- Saghafi, M.A., Hussain, A., Saad, M.H.M., Tahir, N.M., Zaman, H.B., Hannan, M., 2012. Appearance-based methods in re-identification: a brief review. In: *Signal Processing and its Applications (CSPA)*, 2012 IEEE 8th International Colloquium on. IEEE, pp. 404–408.
- Sanchez, R.O., Flores, C., Horowitz, R., Rajagopal, R., Varaiya, P., 2011. Vehicle re-identification using wireless magnetic sensors: Algorithm revision, modifications and performance analysis. In: *Vehicular Electronics and Safety (ICVES)*, 2011 IEEE International Conference on. IEEE, pp. 226–231.
- Shan, Y., Sawhney, H.S., Kumar, R., 2005. Vehicle identification between non-overlapping cameras without direct feature matching. In: *Computer Vision*, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 1. IEEE, pp. 378–385.
- Shan, Y., Sawhney, H.S., Kumar, R., 2008. Unsupervised learning of discriminative edge measures for vehicle matching between nonoverlapping cameras. *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (4), 700–711.
- Sharma, R.K., 1999. Probabilistic model-based multisensor image fusion.
- Shen, Y., Xiao, T., Li, H., Yi, S., Wang, X., 2017. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In: *Proceedings of the IEEE Conference on Computer Vision*.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Smith, B.L., Zhang, H., Fontaine, M., Green, M., 2003. Cellphone probes as an ATMS tool. Technical Report. The Center.
- Sochor, J., Herout, A., Havel, J., 2016. Boxcars: 3d boxes as cnn input for improved fine-grained vehicle recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3006–3015.
- Song, H.O., Xiang, Y., Jegelka, S., Savarese, S., 2016. Deep metric learning via lifted structured feature embedding. In: *Computer Vision and Pattern Recognition (CVPR)*, 2016 IEEE Conference on. IEEE, pp. 4004–4012.
- Sun, C.C., Arr, G.S., Ramachandran, R.P., Ritchie, S.G., 2004. Vehicle reidentification using multidetector fusion. *IEEE Trans. Intell. Transp. Syst.* 5 (3), 155–164.
- Sun, Z., Bebis, G., Miller, R., 2006. On-road vehicle detection: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (5), 694–711.
- Sun, C., Ritchie, S.G., Tsai, K., Jayakrishnan, R., 1999. Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways. *Transp. Res. C* 7 (4), 167–185.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., et al., 2015. Going Deeper with Convolutions. *IEEE*.
- Takeda, H., Farsiu, S., Milanfar, P., 2007. Kernel regression for image processing and reconstruction. *IEEE Trans. Image Process.* 16 (2), 349–366.
- Tang, Y., Wu, D., Jin, Z., Zou, W., Li, X., 2017. Multi-modal metric learning for vehicle re-identification in traffic surveillance environment. In: *Image Processing (ICIP)*, 2017 IEEE International Conference on. IEEE, pp. 2254–2258.
- Tian, Y., Dong, H.h., Jia, L.m., Li, S.y., 2014. A vehicle re-identification algorithm based on multi-sensor correlation. *J. Zhejiang Univ. Sci. C* 15 (5), 372–382.
- Turner, S.M., Eisele, W.L., Benz, R.J., Holdener, D.J., 1998. Travel time data collection handbook. Technical Report.
- Vishnu, V.M., Rajalakshmi, M., Nedunchezian, R., 2017. Intelligent traffic video surveillance and accident detection system with dynamic traffic signal control. *Cluster Comput.* 1–13.
- Vuagnoux, M., Pasini, S., 2010. An improved technique to discover compromising electromagnetic emanations. In: *Electromagnetic Compatibility (EMC)*, 2010 IEEE International Symposium on. IEEE, pp. 121–126.
- Waltz, E., Llinas, J., et al., 1990. Multisensor Data Fusion, Vol. 685. Artech house Boston.
- Wan, E.A., Van Der Merwe, R., 2001. The unscented kalman filter. In: *Kalman Filtering and Neural Networks*, Vol. 5. pp. 221–280.
- Wang, X., 2013. Intelligent multi-camera video surveillance: A review. *Pattern Recognit. Lett.* 34 (1), 3–19.
- Wang, J., Indra-Payong, N., Sumalee, A., Panwai, S., 2014. Vehicle reidentification with self-adaptive time windows for real-time travel time estimation. *IEEE Trans. Intell. Transp. Syst.* 15 (2), 540–552.
- Wang, Z., Tang, L., Liu, X., Yao, Z., Yi, S., Shao, J., Yan, J., Wang, S., Li, H., Wang, X., 2017. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 379–387.
- Watchar, N., 2017. Approximate license plate string matching for vehicle re-identification. In: *Advanced Video and Signal Based Surveillance (AVSS)*, 2017 14th IEEE International Conference on. IEEE, pp. 1–6.
- Woesler, R., 2003. Fast extraction of traffic parameters and reidentification of vehicles from video data. In: *Intelligent Transportation Systems*, 2003. Proceedings. 2003 IEEE, Vol. 1. IEEE, pp. 774–778.
- Wunnavu, S.V., Yen, K., Babij, T., Zavaleta, R., Romero, R., Archilla, C., 2007. Travel time estimation using cell phones (ttec) for highways and roadways.
- Xiang, Y., Choi, W., Lin, Y., Savarese, S., 2015. Data-driven 3d voxel patterns for object category recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1903–1911.
- Yan, K., Tian, Y., Wang, Y., Zeng, W., Huang, T., 2017. Exploiting multi-grain ranking constraints for precisely searching visually-similar vehicles. In: *The IEEE International Conference on Computer Vision (ICCV)*.
- Yang, G.Z., Andreu-Perez, J., Hu, X., Thiemjarus, S., 2014. Multi-sensor fusion. In: *Body sensor networks*. Springer, pp. 301–354.
- Yang, L., Luo, P., Change Loy, C., Tang, X., 2015. A large-scale car dataset for fine-grained categorization and verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3973–3981.
- Ygnace, J.L., Drane, C., Yim, Y., De Lacvivier, R., 2000. Travel time estimation on the san francisco bay area network using cellular phones as probes.
- Yim, Y., Cayford, R., 2001. Investigation of vehicles as probes using global positioning system and cellular phone tracking: field operational test.
- Yokota, T., Inoue, T., Nagai, T., Kobayashi, Y., Takagi, K., 1996. Travel time measuring system based on platoon matching: a field study. In: *Intelligent Transportation: Realizing the Benefits*. Proceedings of the 1996 Annual Meeting of ITS America. ITS America.
- Yong, D., WenKang, S., ZhenFu, Z., Qi, L., 2004. Combining belief functions based on distance of evidence. *Decis. Support Syst.* 38 (3), 489–493.
- Zapletal, D., Herout, A., 2016. Vehicle re-identification for automatic video traffic surveillance. In: *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016 IEEE Conference on. IEEE, pp. 1568–1574.
- Zha, Z.J., Mei, T., Wang, Z., Hua, X.S., 2007. Building a comprehensive ontology to refine video concept detection. In: *Proceedings of the international workshop on Workshop on multimedia information retrieval*. ACM, pp. 227–236.
- Zhang, B., 2013. Reliable classification of vehicle types based on cascade classifier ensembles. *IEEE Trans. Intell. Transp. Syst.* 14 (1), 322–332.
- Zhang, C., Heidemann, J., 2011. Evaluating signature matching in a multisensor vehicle classification system (extended). Technical Report. Citeseer.
- Zhang, C., Liu, W., Ma, H., Fu, H., 2016. Siamese neural network based gait recognition for human identification. In: *Acoustics, Speech and Signal Processing (ICASSP)*, 2016 IEEE International Conference on. IEEE, pp. 2832–2836.

- Zhang, Y., Liu, D., Zha, Z.J., 2017. Improving triplet-wise training of convolutional neural network for vehicle re-identification. In: *Multimedia and Expo (ICME), 2017 IEEE International Conference on*. IEEE, pp. 1386–1391.
- Zhang, J., Wang, F.Y., Wang, K., Lin, W.H., Xu, X., Chen, C., 2011. Data-driven intelligent transportation systems: a survey. *IEEE Trans. Intell. Transp. Syst.* 12 (4), 1624–1639.
- Zhang, H.s., Zhang, Y., Yao, D.y., Hu, D.c., 2005. An architecture of traffic state analysis based on multi-sensor fusion. In: *Intelligent Vehicles Symposium, 2005. Proceedings*. IEEE. IEEE, pp. 855–860.
- Zheng, Y., Capra, L., Wolfson, O., Yang, H., 2014. Urban computing: concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* 5 (3), 38.
- Zheng, Q., Liang, C., Fang, W., Xiang, D., Zhao, X., Ren, C., Chen, J., 2015. Car re-identification from large scale images using semantic attributes. In: *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*. IEEE, pp. 1–5.
- Zhou, Y., Liu, L., Shao, L., 2018. Vehicle re-identification by deep hidden multi-view inference. *IEEE Trans. Image Process.*
- Zhou, Y., Liu, L., Shao, L., Mellor, M., 2016. Dave: a unified framework for fast vehicle detection and annotation. In: *European Conference on Computer Vision*. Springer, pp. 278–293.
- Zhou, X., Mahmassani, H.S., 2006. Dynamic origin-destination demand estimation using automatic vehicle identification data. *IEEE Trans. Intell. Transp. Syst.* 7 (1), 105–114.