

Split-and-Match: A Bayesian Framework for Vehicle Re-identification in Road Tunnels

Andrés Frías-Velázquez, Peter Van Hese, Aleksandra Pižurica,
Wilfried Philips

Ghent University

LIBERUM SCIENTIA

DISCLAIMER:

This manuscript has been accepted for publication in Engineering Applications of Artificial Intelligence. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all disclaimers that apply to the journal apply to this manuscript. A definitive version was subsequently published in *Engineering Applications of Artificial Intelligence*, [VOL 45, (October 2015)] DOI <http://dx.doi.org/10.1016/j.engappai.2015.06.024>

Split-and-Match: A Bayesian Framework for Vehicle Re-identification in Road Tunnels

Andrés Frías-Velázquez*, Peter Van Hese, Aleksandra Pižurica,
Wilfried Philips

*Department of Telecommunications and Information Processing (TELIN-IPi-iMinds),
Ghent University, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium.*

Abstract

Vehicle re-identification is key to keep track of vehicles monitored by a multi-camera network with non-overlapping views. In this paper, we propose a probabilistic framework based on a two-step strategy that re-identifies vehicles in road tunnels. The first step consists of splitting the re-identification problem by connecting groups of vehicles observed in different cameras using certain motion and appearance criteria. In the second step, we build a Bayesian model that finds the optimal assignment between vehicles of connected groups. Descriptors like trace transform signatures, lane change, and motion discrepancies are used to derive our probabilistic framework. Experimental tests reveal that connected groups derived from the first step are composed of 4 vehicles on average. This allow us to constrain the number of candidate matches and increase the chances of getting the correct match. In the second step, our Bayesian model succeeds in matching vehicles among candidates with very similar appearance and under uneven illumination conditions. In general, our system reports a re-identification accuracy of 92% using a nearest-neighbor matcher, and 98% using a one-to-one matcher. These results outperform previous works and encourage us to further develop our solution for other re-identification applications.

Keywords: Vehicle matching, Multicamera tracking, Tunnel surveillance, Non-overlapping cameras, Trace transform.

1. Introduction

Tunnel surveillance is of utmost importance for transportation authorities because the risk of being killed in a tunnel accident is twice as high as in open roads (Naussbaumer, 2007). One factor that increases the fatalities in tunnels is the risk of crashing into the tunnel walls. Another factor is the risk of fire because the chances of death by suffocation and burning increases in confined

*Corresponding author: Tel.:+3292644226, Fax: +3292644295.

Email address: Andres.FriasVelazquez@telin.ugent.be (Andrés Frías-Velázquez)

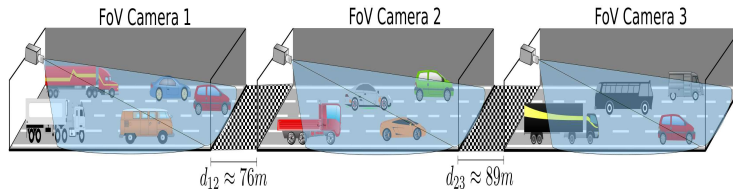


Figure 1: Example of a multicamera setup for tunnel surveillance with non-overlapping views. Vehicle re-identification is needed to correctly associate each of the vehicles across different cameras installed in the tunnel.

spaces. To improve the security in tunnels, multicamera surveillance systems have been extensively deployed to monitor traffic activity. The intention is to provide valuable information to security and emergency corps on hazardous situations in a timely fashion.

A downside of a multicamera surveillance system is the considerable amount of video data generated and its storage. As a tunnel length may reach up to 25 km, hundreds of cameras and human operators are needed to have a full coverage of the tunnel. To minimize the cost of the multicamera network, its maintenance, and data flow, the camera array is placed with non-overlapped Fields of View (FoV), as shown in Fig. 1. In exchange to these benefits, the identity of the vehicles is lost while passing through the blind zone due to the lack of tracking information. Consequently, it is necessary to perform an *identity handover* or re-identification to associate the tracking information from corresponding vehicles. In practice, this re-identification is key to detect incidents in areas not covered by cameras. For example, collisions and broken-down vehicles can be detected when a vehicle is seen by one camera, but not by the next one after a certain time. Re-identification is also crucial to keep track of vehicles that pose a high risk to the tunnel safety, such as trucks carrying dangerous goods. By knowing the position of dangerous vehicles in the tunnel, proper measures can be taken in case of accident.

Even though vehicle matching has been extensively studied in the last decade, dealing with strong illumination and appearance variations remain major challenges. In road tunnels for example, illumination is not isotropically distributed, causing shades and drastic changes on the appearance of the vehicles. Moreover, head and rear lights are normally turned on, which introduce local illumination changes in the scene. All these disturbances turn the vehicle re-identification into a very challenging problem, as illustrated in Fig. 2.

In this work, we propose a Bayesian framework that jointly exploits the appearance and motion information of vehicles to perform a fast and robust re-identification. A simple example illustrating our re-identification approach is presented in Fig. 3. The proposed framework is basically composed of two steps: The first step, called *Matching Problem Decomposition* (MPD), splits the re-identification problem into several matching subproblems based on a statistical hypothesis test. Each subproblem represents the assignment problem of a group of vehicles observed in different cameras. The second step, called

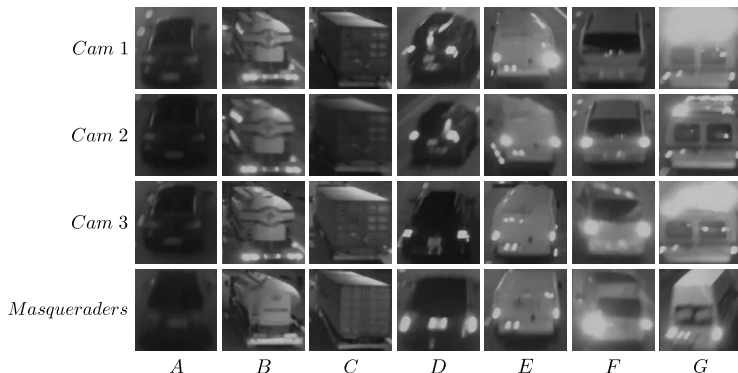


Figure 2: Illustration of the re-identification challenges faced in a tunnel. The first three rows of each column show images of the same vehicle captured in three different cameras. The last row shows vehicles that look quite similar to their counterparts found in every column, but they come from a different vehicle. We name these distractors *masqueraders*. In columns *A*, *B*, and *C* we can see the similarity between corresponding images and masqueraders, which may lead to false alarm matches. Moreover, in columns *D* and *E* we can see that scale and pose changes may also cause mismatches. Finally, in columns *F* and *G* we can observe the impact of vehicle lights on the appearance of the vehicles.

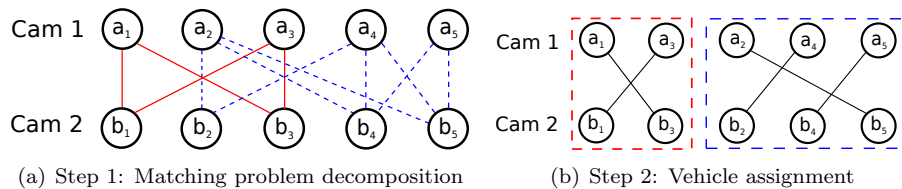


Figure 3: A simple example of our two-step strategy performing the vehicle re-identification between adjacent cameras. In Fig. 3(a) we can see a graph that shows the first step of our strategy, called matching problem decomposition. The vertices labeled as a and b represent the vehicles observed in camera 1 and 2, respectively. The edges of the graph represent the candidate matches resulting from a statistical hypothesis test. With this information, we find the connected components of the graph like those formed with solid and dashed edges that we call matching subproblems. In Fig. 3(b) we can see another graph that represents the second step of our strategy, referred to as vehicle assignment. In this stage, a MAP estimation is used to get the best bijective assignment for each matching subproblem.

Vehicle Assignment (VA), finds the best bijective assignment for each matching subproblem using a Maximum-a-Posteriori (MAP) estimation. The posterior probability is built with descriptors based on appearance and motion cues.

Considering the challenging illumination conditions in tunnels, the fusion of appearance and motion information plays a key role in obtaining an optimal re-identification performance. The proposed spatio-temporal model is described in terms of several motion features related to the entry and exit points of the fields of view like lane change, speed, displacement, and transition time. Moreover, we assume that vehicles move from camera to camera following a linear acceleration model. This assumption holds for the traffic dynamics encountered in road

tunnels, since the flow is normally unidirectional and the road between cameras is usually straight. To evaluate the consistency of the motion data with the kinematic model, we propose a descriptor called *spatial discrepancy*. With this descriptor we expect to discard unlikely matches and reduce the number of possible assignments.

Our appearance model is based on trace transform signatures (Petrou and Kadyrov, 2004). These image descriptors robustly deal with illumination and affine pose variations. The signatures, also called *circus functions*, are compared between images to yield a similarity measure. Also, we propose an appearance descriptor based on corner point detection to measure the *shape complexity* of the vehicles. This feature intends to describe the type of vehicle, either car or truck, to prevent matching vehicles of different kind. Color information is not available in the videos analyzed in this paper, but if available, color descriptors such as the hue histogram and color names (Van de Weijer et al., 2009) can be easily integrated into the appearance model.

The major contributions of this paper are the following: (i) Descriptors based on spatial discrepancies and vehicle type information that verify the motion and appearance consistency of candidate matches. (ii) A preclassification scheme that divides the re-identification problem into several matching subproblems. Thanks to this step, we can process each matching subproblem independently and increase the chances of getting the correct match. (iii) A MAP formulation that finds the optimal vehicle assignment for each matching subproblem. By fusing features like circus functions, spatial discrepancy, and lane change, our re-identification method outperforms previous works based on Haar features (Rios-Cabrera et al., 2012) and Radon-like signatures (Jelača et al., 2013).

The rest of the paper is organized as follows: in Section 2, we present an overview of the related work. Some insights about matching problem decomposition are presented in Section 3. A detailed description of our two-step strategy is presented in Section 4. In Section 5, an evaluation of our method and a comparison with previous work is presented. Finally, the conclusions of this work are stated in Section 6.

2. Related Work

In general, the re-identification problem is considered part of the more general multicamera tracking problem. A review of early works related to multicamera tracking for overlapping and non-overlapping views can be found in Javed et al. (2008). A more recent and comprehensive review of the state-of-the-art is presented by Wang (2013). This survey analyzes several multicamera tracking methods in terms of their camera overlapping, network topology, motion/appearance model, and target application. We refer the reader to these surveys to get a broad overview of the multicamera tracking literature. In the rest of this section, we will focus on reviewing re-identification methods for traffic surveillance with non-overlapping view cameras.

2.1. Re-identification methods based on appearance models

Under moderate appearance variations, vehicle re-identification is posed as a *same-different classification* problem by many methods based on appearance descriptors (Ferencz et al., 2005; Shan et al., 2005a,b). That is, the discriminatory information is not only extracted from the vehicle image itself, but also from the similarities and differences among other vehicles. For instance, Shan et al. (2005a) proposed a feature vector composed of edge-map distances between a given vehicle and some exemplars within the same camera. Then, feature vectors of same and different vehicles from distinct cameras are used to build the same-different classifier. On the other hand, Ferencz et al. (2005) built a same-different classifier based on image patches of same and different vehicles. Features like position, edge contrast, and patch energy are used to implement the classifier. In a previous paper (Frías-Velázquez et al., 2012), we performed an initial exploration of the vehicle re-identification problem using trace-transform signatures (Petrou and Kadyrov, 2004). In particular, these image descriptors are robust to illumination changes and affine transformations, although some of them yield redundant information. Therefore, we performed a feature selection analysis to identify the set functionals that return the most salient signatures. Then, we verified them with a simple identification test by using an unsupervised matching scheme, which is based on the similarity between signatures. In conclusion, this study served as a baseline to develop the appearance model of the present work.

Re-identification methods based on 3D models (Guo et al., 2008; Hou et al., 2009) better deal with large variations of pose and illumination than approaches based on a single-view image. In a first step, the 3D methods estimate the pose and appearance of the reference and target vehicles using 3D models. In a second step, the vehicles are rendered in a normalized 3D space making the comparison of the vehicles geometrically invariant. In particular, Hou et al. (2009) not only performed the pose normalization, but also estimated the albedo to compare the vehicles under the same illumination conditions. Note that these methods are computationally too demanding for most real-time applications due to the precision constraints imposed by the 3D models and the computation of the pose estimation.

2.2. Re-identification methods based on the fusion of appearance and spatio-temporal models

Statistical models represent a natural framework to combine multiple cues and to pose the re-identification problem in a probabilistic fashion. For instance, Javed et al. (2008) presented a multicamera tracking system based on Maximum Likelihood (ML) estimation using space-time and appearance cues. The appearance cues are derived from a subspace of brightness transfer functions, while the space-time cues depend on physical constraints that the tracked objects may encounter on their way. By using both appearance and motion cues, the identification rate significantly improves in comparison with the performance of each feature. Although this approach was not tested for traffic applications,

it has served as inspiration to methods intended for traffic surveillance (Dixon et al., 2009; Choe et al., 2010; Matei et al., 2011).

Multicamera vehicle tracking in large-scale networks, like those installed in urban areas, has attracted a lot of interest in recent years. An example of this multicamera setup is described by Dixon et al. (2009), which exploits the structure of the urban environment to convert the 2D tracking problem into a 1D problem by generating tracklets. Then, a MAP formulation is used to infer connected tracklets as the vehicles switch fields of view. Similarly, Matei et al. (2011) developed a multicamera tracking method based on Multi Hypothesis Tracking (MHT) (Reid, 1979) by incorporating appearance information into the predictive motion model. In Choe et al. (2010), a method to analyze the traffic flow in an urban area with low-frame rate cameras is presented. The traffic analysis is based on detecting and tracking key vehicles to evaluate the traffic conditions of the network. Significant contextual information such as position, lane, orientation, and speed is fused with local image descriptors to match the vehicles. Kogut and Trivedi (2001) presented a method for tracking groups of vehicles in a metropolitan area. The method relies on color data and the spatial organization of vehicles to find corresponding groups of vehicles.

Multicamera tracking for tunnel surveillance was explored by Rios-Cabrera et al. (2012). This work aimed to detect, track, and re-identify vehicles by sharing the same Haar features on each processing stage. These features are primarily used by an AdaBoost cascade classifier to detect the vehicles. Then, the features are reused to build another classifier for the identification step, but in this case to generate a binary descriptor named *vehicle fingerprint*. This descriptor is compared between vehicles yielding a similarity feature, which is later used by a matching voting system to perform the assignment.

In our earlier work (Niño Castañeda et al., 2011), we also studied multicamera tracking in tunnels. In a first step, we performed the vehicle detection and classification with an Adaboost classifier using Haar features. In a second step, we used the vehicle detections and optical-flow data to track the vehicles with a Kalman filter. Finally, the vehicles were matched between cameras by comparing 1D signatures from every vehicle image. This last step was further developed by Jelača et al. (2013). More concretely, the re-identification step is based on Radon-like signatures to describe the appearance of the vehicles. Four signatures were computed by finding the projections of each vehicle image at four different angle orientations (0° , 45° , 90° , 135°). Then, the signatures of vehicles observed in different cameras were compared to yield a similarity measure. A confusion matrix was filled in with the similarity measure of every matching combination, and the Hungarian algorithm (HA) was used to optimally solve the one-to-one assignment.

The main differences between the re-identification methods proposed in the present and previous paper (Jelača et al., 2013) are the following: In the present paper, the re-identification problem is modeled with a probabilistic framework using a hypothesis test for the split step, and a MAP estimation for the matching step. The statistical model is built with context-rich features that verify the motion and appearance consistency of matching vehicles. These features provide

evidence to reduce the number of candidate matches and increase the chances of getting the correct match. By contrast, in our previous paper, we matched vehicles by simply comparing their image features without any probabilistic support. These image features are composed of Radon-like signatures, which are less robust to illumination and pose changes than the trace-transform signatures used in the present work. Therefore, by relying only on appearance features, the matching chances are lower than combining motion and appearance cues, as many vehicles have similar appearance. This last statement is verified in Section 5.3.2.

3. Insights about Matching Problem Decomposition

To effectively use a re-identification system for multicamera tracking, corresponding vehicles need to be re-identified shortly after being detected in consecutive cameras. This kind of implementation involves additional tasks before performing the vehicle assignment. On the one hand, we have to ensure that corresponding vehicles have been observed in consecutive cameras at the moment of the assignment. We refer to this problem as the *observation delay problem*, which arises because vehicles take some time to pass from one camera to other. On the other hand, we have to split the re-identification problem into smaller subproblems to restrict the possible matches and to quickly perform the assignment. In practice, this decomposition can be performed on the basis of time slots or batches of vehicles. Unfortunately, by splitting the re-identification problem, corresponding vehicles may end up in different groups of vehicles causing mismatches. Therefore, the decomposition method has to minimize this type of mismatches, while allowing a fast re-identification with small groups of vehicles.

In Fig. 4, we show with graphs a high-level representation of the decomposition methods proposed in previous works and in this paper. In particular, the method proposed by Rios-Cabrera et al. (2012) is depicted in Fig. 4(a). This figure shows a fully connected graph where a matching window slides over the graph applying a voting mechanism to perform the vehicle assignment. The window length is carefully chosen to ensure that corresponding vehicles are duly included in the matching window. In practice, the optimal window length is found offline by selecting the one that yields the highest re-identification rate. In conclusion, this window-based re-identification is mostly effective if vehicles move at a constant speed and keep a strict ordering between cameras, which is unlikely to happen in real situations.

In Fig. 4(b), we present a graph that illustrates the method proposed in our earlier paper (Jelača et al., 2013). Unlike Fig. 4(a), this graph is not fully connected because a candidate match selection is applied to discard unlikely matches. The criterion for the candidate selection establishes that every vehicle observed in a given camera can be potentially matched with vehicles observed within certain time frame in the previous camera. This time frame is estimated with the maximum and minimum delay between cameras to deal with the ob-

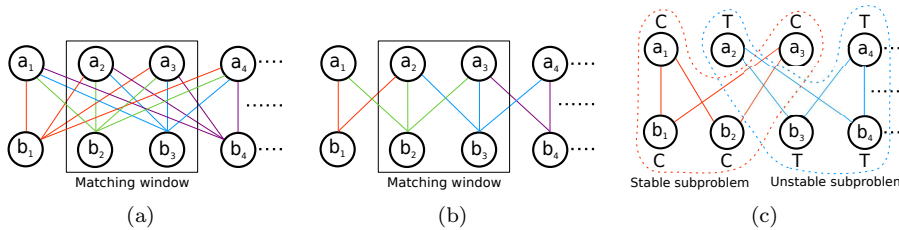


Figure 4: Example graphs representing the re-identification methods proposed in: (a) the paper of Rios-Cabrera et al. (2012), (b) our previous work (Jelača et al., 2013), and (c) this paper. The vertices labeled as a and b represent the vehicles observed in camera 1 and 2, respectively. Every line connecting two vertices represents a candidate match. These links are drawn with different colors to enhance their visualization. In Fig. 4(a), no candidate match selection is used yielding a highly connected graph. In Fig. 4(b), a candidate match selection based on the elapsed time between cameras is applied, thus reducing the number of edges of the graph. In both Fig. 4(a) and Fig. 4(b), a matching window that slides over the graph is used to split the re-identification problem. In Fig. 4(c), the labels C and T next to the vertices stand for car and truck, respectively. A candidate selection using a statistical test generates a sparsely connected graph by linking vehicles that are consistent with the motion model and of the same type. This strategy splits the graph into smaller bipartite graphs, as shown with dashed lines, that we call matching subproblems. Finally, the vehicle assignment is performed once the composition of each subproblem remains stable over time.

servation delay problem. To perform the assignment, this method also uses the sliding window and voting mechanism described in Rios-Cabrera et al. (2012).

In Fig. 4(c), we present a graph that represents the method proposed in this paper. First, as vehicles are detected in consecutive cameras, a candidate match selection based on a statistical test is applied. This test verifies that the candidate matches are consistent with the motion model, and of the same type of vehicle. The aim is to discard numerous unlikely matches while minimizing the risk of rejecting true matches. Then, vehicles with common candidate matches are dynamically clustered to form matching subproblems. Since these subproblems may change their composition as new vehicles are detected, the vehicle assignment is performed once a subproblem remains stable over time using a MAP estimation. In comparison with previous work, our method evaluates each subproblem independently, and infers the correct match more robustly than with a voting mechanism.

4. Proposed Method

In this section, we derive the two steps of our Bayesian framework to solve the vehicle re-identification problem. The first step, called *matching problem decomposition*, is described in section 4.1. The second step, called *vehicle assignment*, is explained in section 4.2.

4.1. Matching Problem Decomposition

The matching problem decomposition consists of first selecting candidate matches for every vehicle based on features like spatial discrepancy and vehicle

type information. This procedure will be described in section 4.1.1. Then, based on the candidate selection, we group vehicles with common candidate matches to compose the matching subproblems. This step will be detailed in section 4.1.2.

4.1.1. Candidate Match Selection

A candidate match is established when a pair of *similar type* of vehicles provide motion data that agree with a linear acceleration model. To evaluate whether a given pair of vehicles is a candidate match or not, we propose a statistical hypothesis test. The null H_0 and alternative H_1 hypotheses are defined as follows:

H_0 : The pair of vehicles is a candidate match.

H_1 : The pair of vehicles is *not* a candidate match.

The decision rule of this test is based on a candidate match probability, denoted as P_c . This probability is derived from two probability functions that evaluate certain motion and appearance criteria. The probability function P_κ verifies the consistency of the motion data of a pair of vehicles with a kinematic model. On the other hand, the probability P_a verifies that the paired vehicles are of similar type. The subscripts κ and a refer to the kinematic and appearance evidence on which the probabilities are based. As a result, we state that $P_c = P_\kappa P_a$ by assuming that motion and appearance features are independent to each other. Finally, the decision rule that determines the best hypothesis H is stated as follows

$$H = P_c \underset{H_1}{\overset{H_0}{\gtrless}} \Upsilon, \quad (1)$$

where Υ is a probability threshold. In the following paragraphs we will derive the density functions P_κ and P_a to estimate P_c .

Motion Gating Function P_κ

In order to derive P_κ , we propose a feature called spatial discrepancy. Let us introduce the spatial discrepancy by considering Fig. 5. This figure shows the position-time graph that relates any vehicle i observed at the exit point in camera r with any other vehicle j observed at the entry point in camera $r + 1$. The motion variables involved are the displacement s_{ij} , elapsed time t_{ij} , exit velocity v_i , and entrance velocity v_j . According to the constant acceleration model, these variables are related to each other as follows

$$s_{ij} = \frac{1}{2}(v_i + v_j)t_{ij}, \quad (2)$$

and measured using the intracamera tracking information. The *spatial discrepancy* ε evaluates the fitness between the displacement estimate s_{ij} and the actual distance between cameras d as stated in (3). According to the error

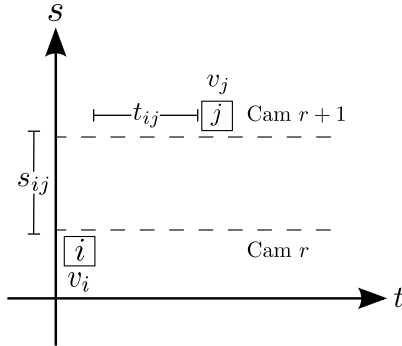


Figure 5: Position-time (s - t) graph showing the motion variables that relate vehicle i and j observed in camera r and $r + 1$, respectively. The variables v_i and v_j denote the speed of the vehicles i and j . Meanwhile s_{ij} and t_{ij} represent the displacement and elapsed time, respectively.

analysis (Taylor, 1996), the displacement estimation can be also expressed as $s_{ij} = \hat{d}_{ij} + \delta s_{ij}$, where \hat{d}_{ij} is the best estimate of d , and δs_{ij} is the margin of error. By substituting this equation into (3), we rewrite the spatial discrepancy as shown in (4).

$$\varepsilon_{ij} = s_{ij} - d \quad (3) \quad \varepsilon_{ij} = \hat{d}_{ij} - d + \delta s_{ij} \quad (4)$$

Note that the spatial discrepancy essentially depends on the margin error δs_{ij} when the displacement is computed for corresponding vehicles (true matches). Distinct vehicles (false matches), on the other hand, are expected to have discrepancy values of larger deviation. These statements can be verified in Fig. 6, where the conditional densities of the spatial discrepancy given the occurrence of a true match ($m = T$) and false match ($m = F$) are presented. In both instances, the probability density is modeled with a normal distribution $\mathcal{N}(\varepsilon; \mu, \sigma^2)$ of mean μ and standard deviation σ . Note that the false match density resembles to a uniform distribution due to its large standard deviation. The parameters of the distributions are estimated with data from our training set, which is described in Section 5.1. Prior probabilities of true and false matches are also derived from our training set. The rate of true and false matches is obtained from all matching combinations with an observation delay between 0 and 6 seconds. The lower bound verifies that the paired vehicles are chronologically consistent. The upper bound is the maximum delay between cameras by traveling at the minimum speed of 50 km/h established for the monitored tunnel.

Having established the parametric distributions of the spatial discrepancy for true and false matches, we use the *Likelihood Ratio* (LR) (5) to derive the motion gating function P_κ (6).

$$LR^\kappa = \frac{P_\kappa}{1 - P_\kappa} \triangleq \frac{P(\varepsilon | m = T)P(m = T)}{P(\varepsilon | m = F)P(m = F)} \quad (5)$$

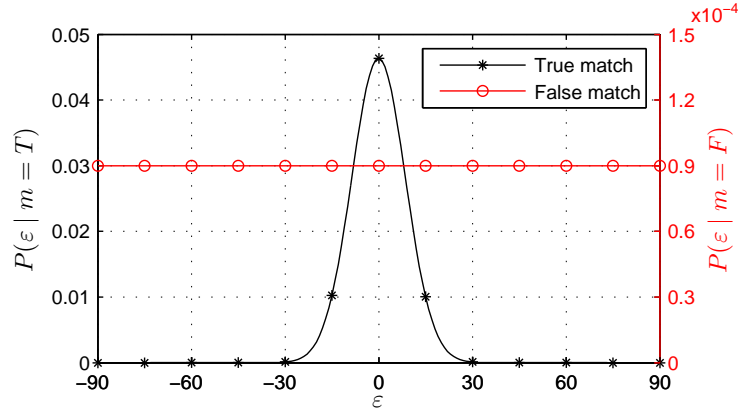


Figure 6: Conditional probability functions of the spatial discrepancy given the occurrence of a true and false match. The left and right y-axes show the range of the conditional probability of the true and false match instances, respectively.

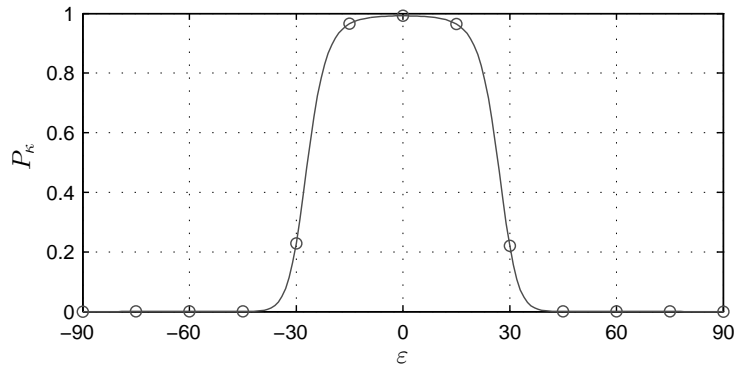


Figure 7: Motion gating function P_κ .

$$P_\kappa = \frac{LR^\kappa}{1 + LR^\kappa} \quad (6)$$

The distribution of P_κ can be observed in Fig. 7. This graph shows that the probability of having a true match increases as the spatial discrepancy tends to zero. Based on this probability, we can define a criterion of motion consistency to establish candidate matches. The criterion of consistency should at least verify that the candidate matches are chronologically consistent, that is $t_{ij} > 0$, which implies that $\varepsilon_{ij} > -d$. In section 5.2, we will evaluate the performance of the motion gating function and its impact on the re-identification performance.

Appearance Gating Function P_a

The probability P_a aims to evaluate the vehicle type similarity between vehicles observed in consecutive cameras based on descriptors like the size and

the bodywork complexity of the vehicles. The size of every vehicle was estimated with the area of its bounding box detected at the entrance of the FoV. Therefore, we expected the area of a truck to be much larger than the one of a car. On the other hand, the complexity descriptor C_x intends to quantify the texture and structure of the bodywork of a vehicle by using corner detection information. Before estimating the complexity, we applied to every image an illumination normalization procedure (Zuiderveld, 1994) to cope with the lighting changes. Then, the complexity was computed as the weighted mean of the N_c distances between the position \mathbf{q} of each corner detection and the centroid \mathbf{g} of the vehicle detection. In mathematical terms, it follows that

$$C_x = \sum_{n=1}^{N_c} w_n \|\mathbf{q}_n - \mathbf{g}\|, \quad (7)$$

where w is the normalized corner score returned by the corner detector (Rosten and Drummond, 2006). Note that as the size of a vehicle increases, the bodywork tends to be more textured and structurally more complex. Therefore, the complexity descriptor increases with the vehicle size because the number of corners and their dispersion around the centroid increases.

In order to derive a compact appearance descriptor, we reduced the feature space composed by the area and complexity descriptor into a 1D feature ξ by applying *Non-parametric Discriminant Analysis* (NDA) (Fukunaga, 1990). In Fig. 8, we can see the conditional distribution of ξ given the occurrence of a car ($\nu = 0$) and truck ($\nu = 1$). The distribution of the car type was modeled with a *Log-Logistic* (*LL*) density function of median α and shape parameter β . On the other hand, the distribution of the truck type was modeled with a *Generalized Gamma Distribution* (*GGD*) of scale a_g , shape b_g , and free parameter c_g . The kernels *LL*($\xi; \alpha, \beta$) and *GGD*($\xi; a_g, b_g, c_g$) are described in Walck (2007), and their parameters were estimated with data from our training set. In Fig. 8, we can see the good separability between these distributions, which suggests a low risk of misclassification.

To derive P_a , let ξ_i^r and ξ_j^{r+1} be the appearance descriptors of vehicle i and j in camera r and $r + 1$, respectively. Moreover, let $h_a = \{0, 1\}$ be the random variable that states whether vehicles i and j are of the same type ($h_a = 1$) or not ($h_a = 0$). Based on all these variables, we defined the likelihood ratio as in (8) to later express P_a as in (9).

$$LR^a = \frac{P_a}{1 - P_a} \triangleq \frac{P(\xi_i^r, \xi_j^{r+1} | h_a = 1)P(h_a = 1)}{P(\xi_i^r, \xi_j^{r+1} | h_a = 0)P(h_a = 0)} \quad (8)$$

$$P_a = \frac{LR^a}{1 + LR^a} \quad (9)$$

The conditional distributions required by the likelihood ratio (8) can be expressed in terms of the car and truck distributions referred above by assuming that $h_a = (\nu_i \wedge \nu_j) \vee (\neg\nu_i \wedge \neg\nu_j)$. Consequently, the conditional distribution

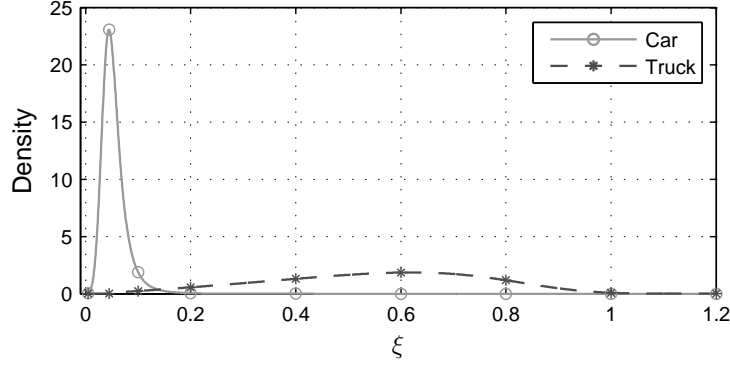


Figure 8: Conditional distributions of ξ given the occurrence of a car and truck.

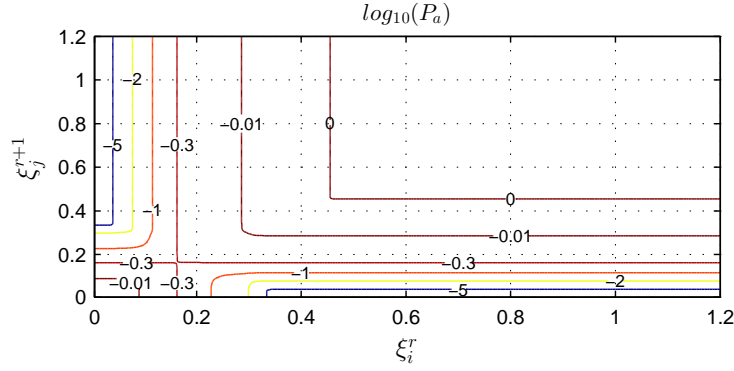


Figure 9: Contour plot of the appearance gating function P_a . The probability that two vehicles are of the same type increases as their appearance features lie within the reddish isolines.

$P(\xi_i^r, \xi_j^{r+1} | h_a = 1)$ when the vehicles i and j are of the same type is defined as follows

$$P(\xi_i^r, \xi_j^{r+1} | h_a = 1) = \lambda P(\xi_i^r | \nu_i = 0) P(\xi_j^{r+1} | \nu_j = 0) + (1 - \lambda) P(\xi_i^r | \nu_i = 1) P(\xi_j^{r+1} | \nu_j = 1), \quad (10)$$

where λ is defined as

$$\lambda = \frac{P^2(\nu = 0)}{P^2(\nu = 0) + P^2(\nu = 1)}. \quad (11)$$

Furthermore, when $h_a = 0$, the conditional distribution is determined as follows

$$P(\xi_i^r, \xi_j^{r+1} | h_a = 0) = \frac{1}{2} P(\xi_i^r | \nu_i = 0) P(\xi_j^{r+1} | \nu_j = 1) + \frac{1}{2} P(\xi_i^r | \nu_i = 1) P(\xi_j^{r+1} | \nu_j = 0). \quad (12)$$

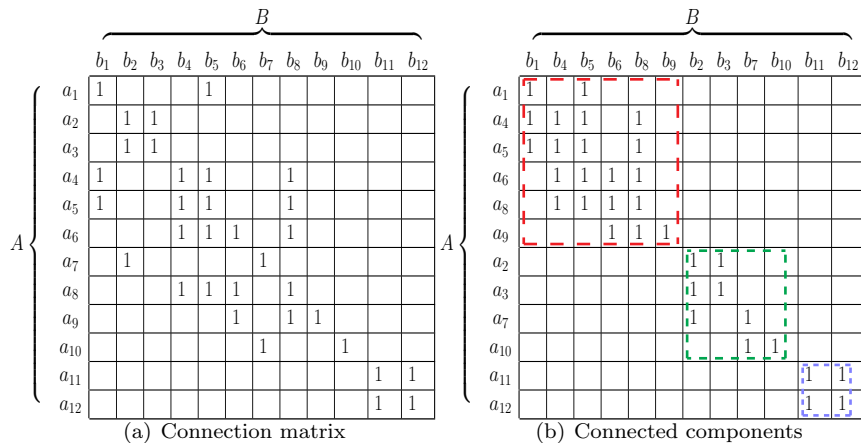


Figure 10: Example of a connection matrix that relates two sets of vehicles observed in different cameras (Fig. 10(a)), and its bundled form after finding the connected components (Fig. 10(b)). The cells filled with ones denote the candidate matches after applying the criterion defined in (1) to all possible matching combinations. The empty cells, on the other hand, denote the unlikely matches. Note that in Fig. 10(a), the candidate matches are scrambled in the matrix, making it difficult to recognize the connected components. In Fig. 10(b), we show in different colors the connected components obtained with the algorithm described in Tarjan (1972).

The prior probabilities referred in (8) are defined as $P(h_a = 1) = P^2(\nu = 0) + P^2(\nu = 1)$ and $P(h_a = 0) = 2P(\nu = 0)P(\nu = 1)$. In Fig. 9, we present the contour plot that shows the isoprobability lines of the appearance gating function (9). Note that a high probability is returned when the features ξ_i^r and ξ_j^{r+1} most likely come from vehicles of the same type. This probability progressively decreases as the features approach to the intersection of the distributions depicted in Fig. 8. Finally, the chances practically vanish when the features most likely come from different type of vehicle.

4.1.2. Matching Subproblems

Based on the candidate match selection, we can now group vehicles with common candidate matches to compose the matching subproblems.

In Fig. 10(a), we can see the connection matrix that represents the bipartite graph of a given set of vehicles A and B detected in camera r and $r + 1$, respectively. Specifically, this matrix denotes the pairwise relationship between vehicles after applying the hypothesis test defined in (1). Note that many unlikely matches have been discarded thanks to the hypothesis test, yielding a sparse connection matrix. Based on this matrix, we find the connected components, as shown in Fig. 10(b), by using the algorithm proposed by Tarjan (1972). Note that each connected component represents a matching subproblem.

To dynamically determine the matching subproblems, a connection matrix keeps record of the candidate matches as the vehicles are detected. If we compute the connected components of that matrix at a given time, some components

Algorithm 4.1: STABLECOMPONENTS(G, cc)

comment: Determine the Stable Components SC from the connection matrix G with connected components cc .

initialization: Get the initial connected components cc_{old} from the connection matrix G_{old}

loop

$$\left\{ \begin{array}{l} \text{wait till detect a new vehicle in cam } r + 1 \\ \text{assign } G_{old} \text{ to } G_{new} \text{ and include the new vehicles} \\ cc_{new} = \text{CONNECTEDCOMPONENTS}(G_{new}) \\ \text{if exist common components between } cc_{old} \text{ and } cc_{new} \\ \quad \text{then} \left\{ \begin{array}{l} \text{compute the buffering time } \tau_b \\ \text{for each common component} \\ \quad \text{do} \left\{ \begin{array}{l} \text{compute the stability time } \tau_s \\ \text{if } \tau_s / \tau_b \geq \zeta \\ \quad \text{then} \left\{ \begin{array}{l} \text{remove } SC \text{ from } cc_{new}, G_{new} \\ \text{return } (SC) \end{array} \right. \end{array} \right. \\ G_{old} = G_{new} \\ cc_{old} = cc_{new} \end{array} \right.$$

might not be complete due to the observation delay between cameras. These components are called unstable, since their composition changes with the arrival of new detected vehicles. Consequently, after finding the connected components, it is necessary to resolve which of them remained stable over time. To identify a stable component, we measure its stability period τ_s . Also, we measure the elapsed time between the first and last vehicle included in the connection matrix, called buffering time τ_b . Then, we consider a stable component if $\tau_s / \tau_b \geq \zeta$. For our tunnel setting, we have experimentally found that a threshold ζ of 0.7 is a suitable point of operation. This procedure to dynamically split the re-identification problem is detailed in Algorithm 4.1.

4.2. Vehicle Assignment

The optimal vehicle assignment for each matching subproblem is performed with a *Maximum-a-Posteriori* (MAP) estimation. The features on which the posterior probability is based are presented in section 4.2.1, while the model itself is described in section 4.2.2.

4.2.1. Features

A fusion of appearance and motion features builds the MAP estimation that finds the best bijective assignment. These features are described as follows:

Table 1: Lane change distribution of true and false matches

Matching case (m)	Number of lane changes (ℓ)			
	0	1	2	3
% True match	94.96	4.91	0.13	0.00
% False match	21.74	43.25	24.6	10.41

Appearance feature

According to [Petrou and Kadyrov \(2004\)](#), the *Trace transform* is a generalization of the Radon transform that allows one to construct image features that are invariant to a chosen group of image transformations. Signature descriptors based on the trace transform, also called *circus functions*, are global features invariant to affine transformations and global illumination changes. These features have been successfully applied in different recognition tasks ([Zarpalas et al., 2007](#); [Kadyrov and Petrou, 2001](#); [Goudelis et al., 2011](#)). By comparing the signatures S^i and S^j computed from the vehicle i and j , we can evaluate their appearance distance θ_{ij} as follows

$$\theta_{ij} = \frac{2}{\pi\sqrt{N_s}} \sqrt{\sum_{n=1}^{N_s} \left(\arccos \left(\rho_{max}(S_n^i, S_n^j) \right) \right)^2}, \quad (13)$$

where ρ_{max} indicates the maximum correlation coefficient, and N_s represents the number of signatures. The distribution of θ is modeled with a *Correlation Angle Distribution CAD*($\theta; \vartheta, \eta$), where ϑ and η represent the population and sample parameters, respectively. This distribution is the arccosine transformation of the correlation coefficient distribution described by [Hotelling \(1953\)](#).

The appearance feature ξ was not considered in the vehicle assignment step because its discriminatory capabilities are more effective to distinguish between cars and trucks rather than to re-identify vehicles directly. The fast computation of this feature allows us to perform a quick candidate match selection, and leave the heavier computation of the trace transform for the vehicle assignment step. Moreover, experimental tests revealed that the re-identification results did not improve by including the feature ξ in the assignment step.

Spatio-temporal features

Let ℓ_{ij} be the discrete random variable that describes the number of lane changes between the vehicle i and j observed in different cameras. The distribution of this feature is presented in [Table 1](#), which is estimated with data from our training set. In this table, we can see that most vehicles do not change lane between cameras. This behavior is expected, since most drivers remain in the same lane for safety reasons, and few vehicles can change lane in a blind zone not greater than 100 m. As a result, 95% of the true matches stay in the same lane, while 78.26% of the false matches change at least one lane. In conclusion, the lane change provides valuable evidence to discriminate among potential matches.

In addition to the lane information, we have also included the spatial discrepancy descriptor ε into the MAP estimation. Recall that this feature conveys important information about the motion consistency of a pair of vehicles with the kinematic model. In section 4.2.2, we will show that the combination of features like the spatial discrepancy and the number of lane changes creates a strong synergic effect in solving the correspondence problem.

4.2.2. Model

In general, an assignment problem can be mathematically described by a permutation function. Therefore, a matching subproblem with K vehicles can be described by the permutation function ϕ as follows

$$\phi = \begin{pmatrix} 1 & 2 & \cdots & K \\ p_1 & p_2 & \cdots & p_K \end{pmatrix}, \quad (14)$$

where every vehicle i in camera r is assigned to vehicle p_i in camera $r + 1$. In order to find the best assignment, we derive a MAP estimation based on the multivariate random variable $\mathcal{O}_\phi = (O_{1p_1}, O_{2p_2}, \dots, O_{Kp_K})$ that assembles the observation vectors $O_{ij} = (\theta_{ij}, \varepsilon_{ij}, \ell_{ij})$ of every correspondence in the assignment ϕ . Similarly, $\mathcal{M}_\phi = (m_{1p_1}, m_{2p_2}, \dots, m_{Kp_K})$ gathers the binary states $m_{ij} = \{T, F\}$ that denote whether vehicle i matches (T) or not (F) to vehicle j . Meanwhile, \mathcal{H}_ϕ collects the candidate match hypotheses of every correspondence in the assignment ϕ .

Having defined the random variables involved in our problem, we estimate the optimal assignment $\hat{\phi}$ with the following MAP formulation

$$\hat{\phi}_{\text{MAP}} = \arg \max_{\phi \in \mathcal{S}} \left(P(\mathcal{M}_\phi = \vec{T} \mid \mathcal{O}_\phi, \mathcal{H}_\phi = \vec{H}_0) \right), \quad (15)$$

where the statements $\mathcal{M}_\phi = \vec{T}$ and $\mathcal{H}_\phi = \vec{H}_0$ imply that the K correspondences in ϕ are true matches and candidate matches, respectively. The solution space \mathcal{S} represents the set of possible permutations. By assuming that the correspondences are conditionally independent to each other, we can express the posterior probability of (15) as follows

$$P(\mathcal{M}_\phi = \vec{T} \mid \mathcal{O}_\phi, \mathcal{H}_\phi = \vec{H}_0) = \prod_{i=1}^K P(m_{ip_i} = T \mid O_{ip_i}, H = H_0). \quad (16)$$

Then, by applying the logarithm to (16), we can rewrite (15) in its equivalent *log-posterior* form as follows

$$\hat{\phi}_{\text{MAP}} = \arg \max_{\phi \in \mathcal{S}} \sum_{i=1}^K \mathcal{P}_{i\phi(i)}, \quad (17)$$

where $\mathcal{P}_{i\phi(i)} = \log(P(m_{ip_i} = T \mid O_{ip_i}, H = H_0))$. Note that the MAP estimation in (17) is posed as a *Linear Sum Assignment Problem* (LSAP) (Burkard et al., 2009), which can be solved with the LAPJV algorithm proposed by Jonker and Volgenant (1987).

Data Distribution

In Fig. 11(a)-(d), we present the scatter plots of true and false match samples in terms of the appearance distance, spatial discrepancy, and number of lane changes. Note that these graphics were obtained with data from our training set. By analyzing the scatter plots, we can see that most true matches have a spatial discrepancy between -15 and 15 . By contrast, false matches are distributed in a larger range of discrepancy values. The distribution of samples along θ shows that the distance between true and false matches decreases as ℓ increases. This means that the differences between corresponding vehicles increase due to the pose variations when changing lane. On the other hand, the distribution of samples for different number of lane changes shows that many true matches do not change lane, whereas most false matches change at least one lane, as previously noted in section 4.2.1. Finally, it is worth to remark that in Fig. 11(a) our three descriptors show a synergic effect on the separability of true and false matches. That is, the population of false matches significantly decreases in the range of $-15 \leq \varepsilon \leq 15$ while most true matches are found in the same interval. Moreover, the appearance feature enhances the separability of the distributions along θ .

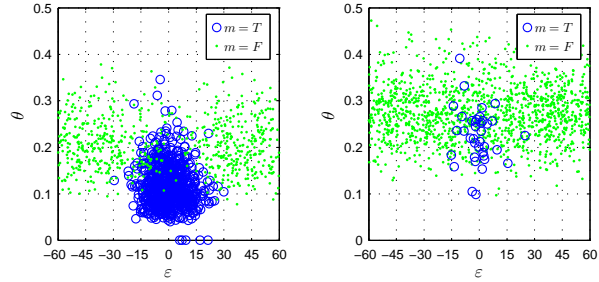
Probability Map

The probability maps sketched in Fig. 11(e)-(h) show the posterior probability for different number of lane changes. Note that the probability maps for $\ell = 2$ and $\ell = 3$ were estimated with density functions derived from an imputation technique (Little and Rubin, 1986) due to the scarce number of true matches. The imputation consists in applying a regression method to estimate the parameters of missing densities using all the observed data.

By analyzing the probability maps along ε , we can see that the largest probabilities occur when $-15 \leq \varepsilon \leq 15$. Therefore, our statistical inference gives a higher matching probability to candidates with a strong motion consistence. Regarding the inference along θ , the matching probability generally increases as the appearance distance tends to zero. However, note that in Fig. 11(e)-(f) there are regions with a large appearance distance that yield a high matching probability. In these specific cases, the evidence provided by the spatial discrepancy dominates over the appearance distance. Finally, the statistical inference shows that the matching probability quickly decreases as the number of lane changes increases.

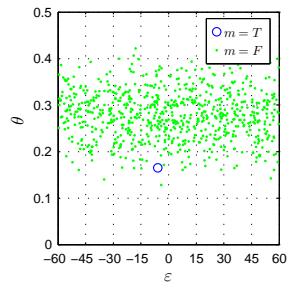
5. Experimental Results

In this section we present an evaluation of our vehicle re-identification system. First, a description of the experimental setup is presented in section 5.1. Then, an assessment of the matching problem decomposition is presented in section 5.2. Finally, an overall analysis of the re-identification performance is conducted in section 5.3.

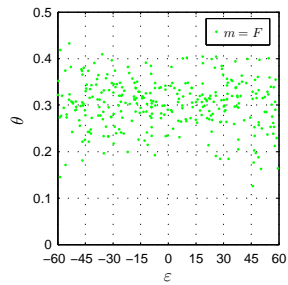


(a) Scatter plot for $\ell = 0$.

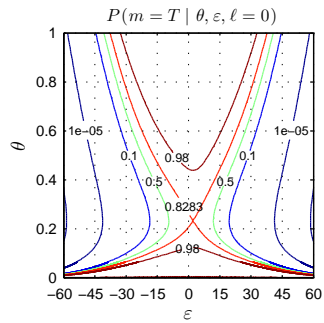
(b) Scatter plot for $\ell = 1$.



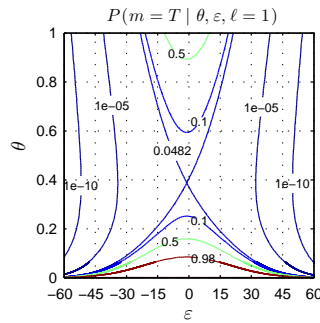
(c) Scatter plot for $\ell = 2$.



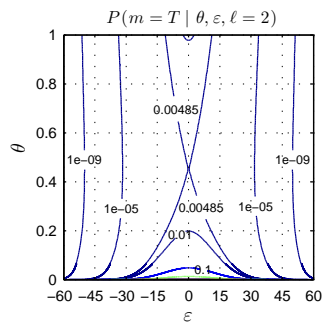
(d) Scatter plot for $\ell = 3$.



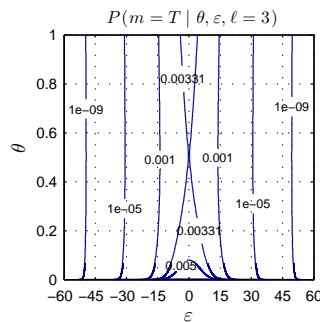
(e) Posterior probability for $\ell = 0$.



(f) Posterior probability for $\ell = 1$.



(g) Posterior probability for $\ell = 2$.



(h) Posterior probability for $\ell = 3$.

Figure 11: Scatter plots and probability maps for different number of lane changes $\ell = \{0, 1, 2, 3\}$. The scatter plots are depicted in the top figures, showing samples of true matches (circles) and false matches (dots) for each instance of ℓ . In the bottom figures, we can see the corresponding probability maps.

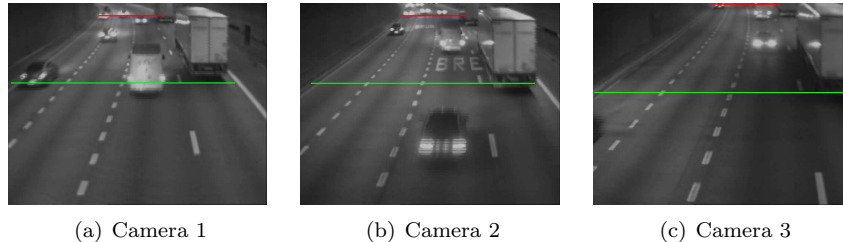


Figure 12: Non-overlapping views of three consecutive cameras installed in a tunnel. The green and red lines represent the entry and exit checkpoints of the field of view, respectively.

5.1. Experimental Setup

The video collection is composed of three sequences of 11 minutes each. These videos are retrieved by three non-overlapping view cameras separated 83 m on average, as shown in Fig. 12. A ground plane camera calibration method (Cathey and Dailey, 2006) based on road marks is used to translate speed and displacement measurements from image coordinates to physical units. In relation to the video properties, the decoded frames have a spatial and temporal resolution of 720×576 pixels and 25 fps, respectively. The color depth is 8-bit gray-scale. Regarding the traffic statistics, the number of vehicles amounts to 564, each of them observed in the three cameras. Moreover, the traffic flow in the tunnel was estimated at 52.18 vehicles/min.

In order to evaluate our method, we partitioned the set of 564 manually annotated vehicles into training and testing set. The training set is composed of 377 vehicles with their corresponding observations in each camera. This group of vehicles was used for the estimation of the gating functions and the posterior probability of our MAP estimation, as described in sections 4.1.1 and 4.2.2. Consequently, the remaining 187 vehicles were used for testing. In order to also evaluate our method using automatic detections, we generate them with the vehicle detector proposed in Frías-Velázquez et al. (2011).

Before extracting the appearance descriptors, every image was preprocessed to compensate lighting changes between cameras with an illumination normalization method (Zuiderveld, 1994). The circus functions were computed using the functionals proposed in Petrou and Kadyrov (2004). As a result, 7 T -functionals and 3 P -functionals returned in total $N_s = 21$ different signatures per vehicle. On the other hand, the complexity descriptor, based on corner detections, was computed with a 9-point FAST detector (Rosten and Drummond, 2006). This detector was chosen due to its outstanding corner repeatability between warped images.

5.2. Assessment of the Matching Problem Decomposition

In this section we evaluate the hypothesis test that rules the candidate match selection. Moreover, we present a statistical analysis of the matching subproblems generated.

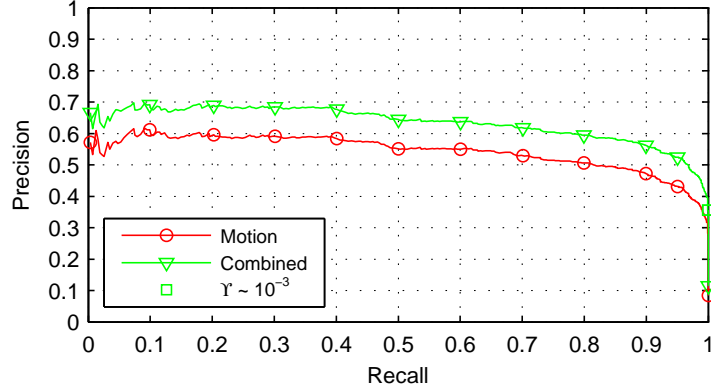


Figure 13: Empirical PR curves of the candidate match probability and motion gating function.

5.2.1. Hypothesis Test Evaluation

The performance of the hypothesis test defined in (1) is evaluated in terms of its precision and recall rates. These scores are based on the number of true positives t_p , false positives f_p , and false negatives f_n as follows

$$Recall = \frac{t_p}{t_p + f_n}, \quad (18) \quad Precision = \frac{t_p}{t_p + f_p}. \quad (19)$$

In Fig. 13, we can see the *Precision vs Recall* (PR) curve that evaluates the performance of the motion gating function P_κ . Moreover, we present the PR curve that evaluates the hypothesis test based on the candidate match probability P_c , which combines the gating functions P_κ and P_a .

The ultimate goal of the hypothesis test is to establish candidate matches without losing any true match ($recall = 1$) while discarding the largest number of unlikely correspondences ($precision \rightarrow 1$). By analyzing the motion PR curve, we can see that the recall was maintained in 1 up to a precision of approximately 0.3. Loosely speaking, this implies that for every vehicle in camera r , approximately 4 candidate matches are selected in camera $r + 1$. This is the case if only the motion gating function P_κ were used. On the other hand, the PR curve of the candidate match probability P_c shows the result of combining the gating functions based on motion and appearance cues. In this case, the recall was kept to 1 up to a precision of approximately 0.4. As a result, for every vehicle observed in camera r , an average of 3 candidate matches are selected in camera $r + 1$. In conclusion, the hypothesis test allows us to considerably reduce the number of unlikely matches without losing any true match.

In Fig. 13, we can also see the operational point that we set for the hypothesis test. This point corresponds to a probability threshold of $\Upsilon = 10^{-3}$, which yields a precision of 0.36 and a recall rate of 1.

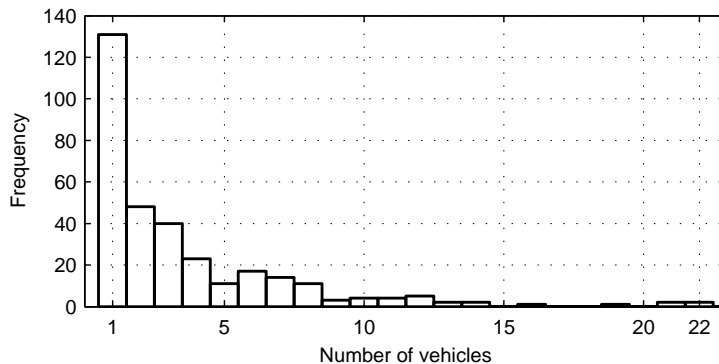


Figure 14: Distribution of the matching subproblems in terms of their number of vehicles. Note that most subproblems are composed of a few vehicles, thus reducing the number of possible assignments. The first bin shows that some vehicles were already re-identified, suggesting that the vehicle-type descriptor and the spatial discrepancy provided enough evidence to get the correct match. On average, the matching subproblems are composed of 4 vehicles, which is also an approximate number of candidate matches per vehicle.

5.2.2. Analysis of the matching subproblems

After selecting the candidate matches, the matching subproblems were found as described in section 4.1.2. In Fig. 14, we present the distribution of the matching subproblems in terms of their number of vehicles. This figure shows that the re-identification problem is split in different matching subproblems of small size. By contrast, previous methods based on a sliding window used a fixed window size of 10 vehicles. This implies that each vehicle has 10 candidate matches, whereas with our approach we have on average 4 candidate matches per vehicle.

An important aspect to analyze of a matching subproblem is its connection matrix, which describes the matching combinations. In this matrix, some matching combinations may be restricted, as shown in Fig. 10(b). As a result, the number of possible assignments can be reduced, and the chances of getting the correct match increase. In order to evaluate the chances of randomly pick the correct assignment in a matching subproblem, we analyzed the entropy of this event. If we assume that $P_R = 1/N_R$ is the probability of getting the correct match out of N_R possible assignments, we can define the entropy E as follows

$$E = \log_2 \left(\frac{1}{P_R} \right) = \log_2(N_R), \quad (20)$$

where the subscript R denotes the connection matrix that is being evaluated. In order to enumerate the N_R permutations of a connection matrix with forbidden combinations is necessary to derive its *rook polynomial* (Tucker, 2006). This polynomial can be automatically generated by using Fielder’s algorithm (Fielder, 2004).

An entropy analysis of the matching subproblems is addressed in Fig. 15. In this figure, the asterisk markers show the entropy of matching subproblems

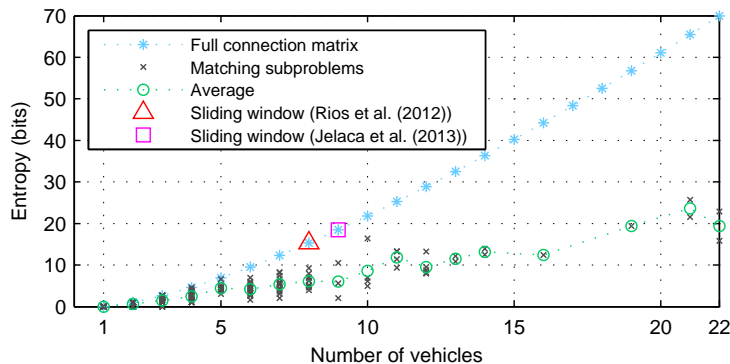


Figure 15: Distribution of the matching subproblems in terms of their number of vehicles and entropy. The asterisk markers show the entropy of matching subproblems of different sizes with a full connection matrix. The cross markers show the entropy of the subproblems returned by the decomposition step, while the circle markers denote their average entropy. For comparative purposes, we include with a triangle and a square marker the entropy of previous work using a sliding-window approach.

with a full connection matrix. In this case, getting the correct assignment is very challenging because the number of permutations increases in a factorial way with the number of vehicles. On the other hand, the cross markers show the entropy of the matching subproblems returned by the decomposition step. In this case, most of the subproblems reported a lower entropy than their equivalents with a full connection matrix. This means that by restricting some matching combinations we can gain some “knowledge” about the correct assignment. Interestingly, matching subproblems of big size lose more entropy because their connection matrix is more sparse.

In comparison with previous methods, we found that our approach yields an average entropy of 2.24 bits, whereas sliding window methods like in [Rios-Cabrera et al. \(2012\)](#) and [Jelača et al. \(2013\)](#), returned an uncertainty of 15.30 bits and 18.47 bits, respectively. In summary, the entropy study reveals that our decomposition step may significantly improve the chances of getting the correct assignment, and thus the re-identification rate.

5.3. Re-identification Assessment

In this section, we evaluate our performance in terms of the ranking capabilities of our Bayesian model to discriminate between true and false matches. Moreover, we present a comparative analysis of our re-identification performance in relation to previous work.

5.3.1. Match Ranking Performance

After a matching subproblem has been found, the log-posterior probability is computed for all valid matching combinations. Then, based on this evidence, we use an assignment method to finally solve the vehicle re-identification.

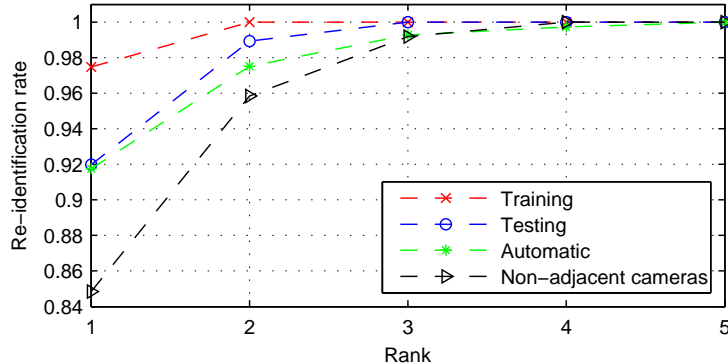


Figure 16: Cumulative matching curves using a nearest neighbor matcher. The CMCs obtained with the training and testing set are shown with cross and circle markers, respectively. The curve with asterisk markers represents the CMC obtained with automatic detections. Finally, the curve with triangle markers shows the performance of matching vehicles from non-adjacent cameras.

For comparative purposes, we tested our performance with a Nearest Neighbor (NN) matcher and a bijective matcher like the LAPJV algorithm (Jonker and Volgenant, 1987). The intention is to evaluate the ranking capabilities of our inference model using these two assignment methods. According to Bolle et al. (2005), the ranking capabilities of a re-identification system can be assessed with a Cumulative Matching Curve (CMC). The CMC estimates the cumulative accuracy of correct identifications found within certain rank of ordered solutions.

An assessment of the re-identification performance using the nearest neighbor matcher is presented in Fig. 16. In general, the performance of a CMC is evaluated in terms of its first ranked identification rate, and the number of ranks need it to reach the 100% of the cumulate accuracy. In these terms, the re-identification rate of our system is 97.5% with the training set and 92.0% with the testing set. The performance with automatic detections reaches to 91.7%. Meanwhile, when we match vehicles from non-adjacent cameras, the performance drops to 84.8%. In this case, note that the distance between cameras is larger than the case of adjacent cameras, which increases the spatial discrepancy and the chances that vehicles change lane and pose. In relation to the convergence of the CMCs, we can see that all of them converge at most in the fifth rank. This implies that all the correct matches were found within the first five ranked solutions. The fast convergence of these curves can be explained by two factors: 1. Most matching subproblems are composed of less than 15 vehicles, which constrains the possible assignments. 2. Our inference model evaluates the probability of true and false matches in a consistent way.

The ranking capabilities of our inference model can be also evaluated with the LAPJV algorithm (Jonker and Volgenant, 1987). Unlike the NN matcher, the LAPJV algorithm performs a one-to-one assignment, and finds the optimal

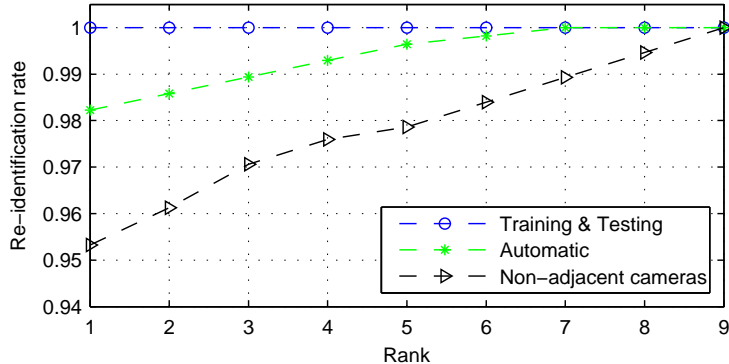


Figure 17: Cumulative matching curves using a bijective matcher. The curve with circle markers denote the CMCs obtained with the training and testing set. The curve with asterisk markers represents the CMC obtained with automatic detections. Finally, the curve with triangle markers shows the performance of matching vehicles from non-adjacent cameras.

solution of the MAP formulation described in (17). In order to determine the rank of the bijective assignments we used Murty’s algorithm (Murty, 1968).

The re-identification performance using the LAPJV algorithm can be analyzed in Fig. 17. Note that the re-identification rate increased in relation to the NN matcher for all cases. This improvement is thanks to the MAP formulation, which jointly optimizes the bijective assignment, and not each vehicle independently. For the training and testing set, the CMCs reach a re-identification performance of 100%. With automatic detections the performance is 98%. Interestingly, the re-identification performance between non-adjacent cameras is 95%. This is an encouraging result considering the larger distance between cameras and the larger dissimilarity between vehicles. Regarding the convergence of the CMCs, we can see that all of them converge at most in the ninth rank. Note that the convergence is slower than with the NN matcher because a bijective matcher has a larger solution space. Nevertheless, thanks to the small size of the subproblems and their sparse connection matrix, our method reaches the convergence within a few ranked solutions.

5.3.2. Comparative Analysis

Having investigated the ranking capabilities of our inference model, we compared our re-identification performance with previous work. Recall that these approaches use a sliding window to group the vehicles between cameras and to perform the assignment. Therefore, their re-identification performance is reported in terms of the window size chosen. Although our method does not require a sliding window, we evaluated it using this scheme for the sake of comparison.

In Fig. 18, we can compare the performance of our approach with earlier work using manual detections. Note that the matching curves reported by previous work decrease as the window size increases. This behavior is caused by two

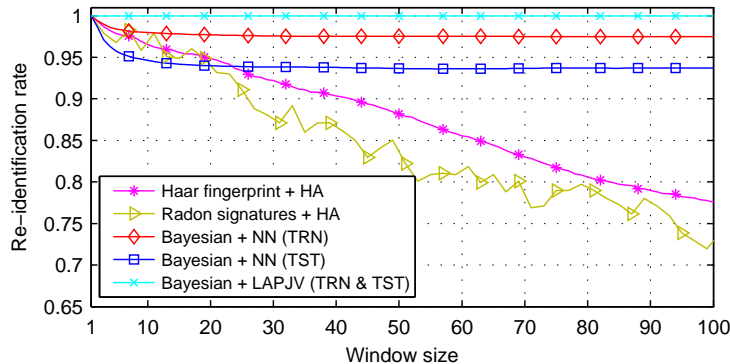


Figure 18: Re-identification performance for different matching windows using manual detections. The curve with asterisk markers shows the performance reported by [Rios-Cabrera et al. \(2012\)](#). The performance presented in our earlier work ([Jelača et al., 2013](#)) is depicted with triangle markers. On the other hand, the performance curves of our Bayesian framework with the training set (TRN) and testing set (TST) are shown with diamond and square markers, respectively. A NN matcher is used in these last cases. Finally, the curve with cross markers shows our performance with the training and testing set using a bijective matcher.

factors: 1. The factorial increase of the number of possible assignments with the window size. 2. The difficulty of matching vehicles of similar appearance within large windows. This last factor becomes more relevant for methods based exclusively on appearance models. On the contrary, the matching curves obtained with our Bayesian framework exhibit a transient and a steady state. These curves reach a steady state thanks to our candidate match selection, which significantly constrains the number of possible assignments. By using a nearest neighbor matcher, the steady state is reached when the re-identification is 97.5% with the training set, and 94% with the testing set. These percentages improve to 100% by using the LAPJV matcher. It is worth to remark that all the performance curves in Fig. 18 were estimated assuming that the vehicles keep a strict ordering between cameras. Therefore, corresponding vehicles were always included within the matching window, thus avoiding the observation delay problem discussed in section 3. In a later evaluation with automatic detections this assumption is not longer considered.

In Fig. 19 we compare the performance of our approach with previous work using automatic detections. By analyzing the matching curves of previous methods, we can observe a faster decay than their corresponding curves in Fig. 18. On average, the performance of these curves dropped 16%. This drop is not only caused by the misalignments of the automatic detections, but also due to the observation delay problem. Therefore, corresponding vehicles may not be found within the matching window leading to mismatches. The performance of our matching curves using the NN and LAPJV matcher are also presented in Fig. 19. Note that these curves maintain a transient and steady state. The steady state is reached when the re-identification rate is 92% using the NN matcher and 98% using the LAPJV matcher. In Fig. 19, we also evaluate the

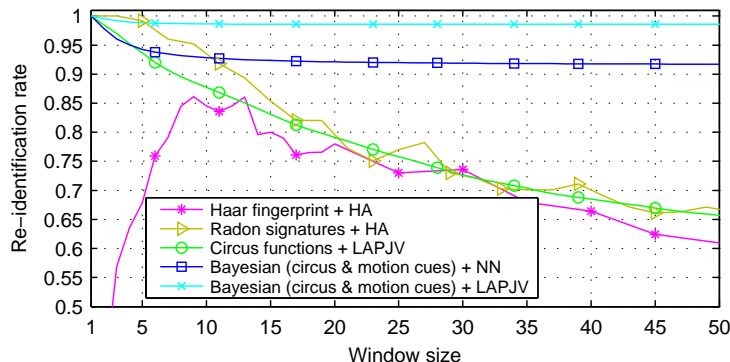


Figure 19: Re-identification performance for different matching windows using automatic detections. The curve with asterisk markers shows the performance reported by Rios-Cabrera et al. (2012). The performance presented in our earlier work (Jelača et al., 2013) is depicted with triangle markers. On the other hand, the curve with circle markers shows the performance of the circus functions using the LAPJV matcher without any statistical support. The matching curves with square and cross markers denote the performance of our Bayesian framework using a NN and LAPJV matcher, respectively.

re-identification performance by only using the circus functions to match the vehicles. The intention is to show the differences in performance when the circus functions are used alone without statistical support, and in combination with the motion cues using our Bayesian framework. As a result, by only using the circus functions, the performance is similar to that reported in previous papers, and lower than the one obtained with our Bayesian framework. In conclusion, the performance gain of our method can be due to two factors: 1. The candidate match selection, which restricts the possible assignments making it easier for the appearance descriptor to discriminate among a few vehicles rather than numerous vehicles of similar appearance. 2. The Bayesian formulation, which effectively exploits the good separability between true and false matches by combining motion and appearance features to infer the best assignment.

6. Conclusion

In this paper, we presented a study of the vehicle re-identification problem for a tunnel surveillance network. The proposed Bayesian framework efficiently combined motion and appearance features to successfully solve the re-identification problem. Descriptors like the spatial discrepancy and the vehicle complexity built a statistical test that dynamically splits the re-identification problem into small matching subproblems. By partitioning the re-identification problem, we showed that the probability of getting the correct match significantly increased in comparison with sliding window methods. On the other hand, descriptors like the circus functions, spatial discrepancy, and lane change were used to derive a MAP formulation to find the optimal assignment for each matching subproblem. An evaluation of this inference model showed that the

correct assignment of the matching subproblems was found within the first 7 out of 120 ranked solutions. Also, the evaluation yielded an overall re-identification rate of 100% and 98% using manual and automatic detections, respectively. This represents an increase of 5% in comparison with our earlier work. Finally, we conclude that the proposed method should be validated in a larger multicamera network and tested for other re-identification applications.

Acknowledgment

This work has been supported by iMinds and by the Flemish Fund for Scientific Research (FWO).

References

- Bolle, R., Connell, J., Pankanti, S., Ratha, N., Senior, A., 2005. The relation between the roc curve and the cmc. In: IEEE Workshop on Autom. Identif. Adv. Technol. pp. 15–20.
- Burkard, R., Dell’Amico, M., Martello, S., 2009. Assignment Problems. SIAM, Philadelphia, PA, USA.
- Cathey, F., Dailey, D., 2006. Mathematical theory of image straightening with applications to camera calibration. In: Proc. IEEE Int. Conf. on Intell. Transp. Syst. pp. 1364–1369.
- Choe, T., Lee, M. W., Haering, N., 2010. Traffic analysis with low frame rate camera networks. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops. pp. 9–16.
- Dixon, M., Jacobs, N., Pless, R., 2009. An efficient system for vehicle tracking in multi-camera networks. In: ACM/IEEE Int. Conf. on Distrib. Smart Cameras. pp. 1–8.
- Ferencz, A., Learned-Miller, E., Malik, J., 2005. Building a classification cascade for visual identification from one example. In: Proc. IEEE Int. Conf. on Computer Vision. pp. 286–293 Vol. 1.
- Fielder, D., 2004. A generator of rook polynomials. The Mathematica Journal 9 (2), 371–375.
- Frías-Velázquez, A., Niño Castañeda, J., Jelača, V., Pižurica, A., Philips, W., 2011. A mathematical morphology based approach for vehicle detection in road tunnels. In: Proc. SPIE 8135. Vol. 8135. p. 8.
- Frías-Velázquez, A., Ortiz Lopez, C. A., Pižurica, A., Philips, W., Cerda Villafañá, G., 2012. Object identification by using orthonormal circs functions from the trace transform. In: IEEE International Conference on Image Processing. IEEE, pp. 2153–2156.

- Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition (2nd Ed.). Academic Press Professional, Inc., San Diego, CA, USA.
- Goudelis, G., Karpouzis, K., Kollias, S., 2011. Robust human action recognition using history trace templates. In: Proc. Int. Workshop on Image Analysis for Multimedia Interactive Services.
- 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: Proc. IEEE Conf. on Computer Vision and Pattern Recognition. pp. 1–8.
- Hotelling, H., 1953. New light on the correlation coefficient and its transforms. Journal of the Royal Statistical Society 15 (2), 193–232.
- Hou, T., Wang, S., Qin, H., 2009. Vehicle matching and recognition under large variations of pose and illumination. In: IEEE Conf. on Computer Vision and Pattern Recognition Workshops. pp. 24–29.
- Javed, O., Shafique, K., Rasheed, Z., Shah, M., 2008. Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. Computer Vision and Image Understanding 109, 146–162.
- Jelača, V., Pižurica, A., Niño Castañeda, J., Frías-Velázquez, A., Philips, W., 2013. Vehicle matching in smart camera networks using image projection profiles at multiple instances. Image and Vision Computing 31, 673–685.
- Jonker, R., Volgenant, A., 1987. A shortest augmenting path algorithm for dense and sparse linear assignment problems. Computing 38, 325–340.
- Kadyrov, A., Petrou, M., August 2001. The trace transform and its applications. IEEE Trans. on Pattern Analysis and Machine Intelligence 23, 811–828.
- Kogut, G. T., Trivedi, M. M., 2001. Maintaining the identity of multiple vehicles as they travel through a video network. In: IEEE Workshop on Multi-Object Tracking. p. 29.
- Little, R. J. A., Rubin, D. B., 1986. Statistical analysis with missing data. John Wiley & Sons, Inc., New York, NY, USA.
- Matei, B. C., Sawhney, H. S., Samarasekera, S., 2011. Vehicle tracking across nonoverlapping cameras using joint kinematic and appearance features. In: IEEE Conf. on Computer Vision and Pattern Recognition. pp. 3465–3472.
- Murty, K. G., 1968. An Algorithm for Ranking all the Assignments in Order of Increasing Cost. Operations Research 16 (3), 682–687.
- Naussbaumer, C., 2007. Comparative analysis of safety in tunnels. In: European Conference of Transport Research Institutes: Young Reseachers Seminar. pp. 1–9.

- Niño Castañeda, J., Jelača, V., Frías-Velázquez, A., Pižurica, A., Philips, W., Rios Cabrera, R., Tuytelaars, T., 2011. Non-overlapping multi-camera detection and tracking of vehicles in tunnel surveillance. In: Proc. Digital Image Computing Techniques and Applications. pp. 591–596.
- Petrou, M., Kadyrov, A., 2004. Affine invariant features from the trace transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 26, 30–44.
- Reid, D., 1979. An algorithm for tracking multiple targets. *IEEE Trans. Automatic Control* 24 (6), 843 – 854.
- Rios-Cabrera, R., Tuytelaars, T., Gool, L. V., 2012. Efficient multi-camera vehicle detection, tracking, and identification in a tunnel surveillance application. *Computer Vision and Image Understanding* 116 (6), 742 – 753.
- Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. In: European Conference in Computer Vision. pp. 430–443.
- Shan, Y., Sawhney, H., Kumar, R., 2005a. Vehicle identification between non-overlapping cameras without direct feature matching. In: IEEE Int. Conference on Computer Vision. pp. 378 – 385 Vol. 1.
- Shan, Y., Sawhney, H. S., Kumar, R., 2005b. Unsupervised learning of discriminative edge measures for vehicle matching between non-overlapping cameras. In: IEEE Conf. on Computer Vision and Pattern Recognition. pp. 894–901.
- Tarjan, R., 1972. Depth-first search and linear graph algorithms. *SIAM Journal on Computing* 1 (2), 146–160.
- Taylor, J. R., 1996. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2nd Edition. Univ. Science Books.
- Tucker, A., 2006. Applied Combinatorics. John Wiley & Sons, Inc., New York, NY, USA.
- Van de Weijer, J., Schmid, C., Verbeek, J., Larlus, D., 2009. Learning color names for real-world applications. *IEEE Transactions on Image Processing* 18 (7), 1512–1523.
- Walck, C., 2007. Hand-book on statistical distributions for experimentalists. Tech. rep., University of Stockholm.
- Wang, X., 2013. Intelligent multi-camera video surveillance: A review. *Pattern Recognition Letters* 34 (1), 3–19.
- Zarpalas, D., Daras, P., Axenopoulos, A., Tzovaras, D., Strintzis, M., 2007. 3d model search and retrieval using the spherical trace transform. *EURASIP J. Appl. Signal Process.* 2007, 023912.
- Zuiderveld, K., 1994. Contrast limited adaptive histogram equalization. In: Graphics Gems IV. Academic Press Professional, pp. 474–485.