

Vehicle Classification for Road Tunnel Surveillance

Andrés Frías-Velázquez, Peter Van Hese, Aleksandra Pižurica, and Wilfried Philips
Ghent University-TELIN-IPI-iMinds, Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium

ABSTRACT

Vehicle classification for tunnel surveillance aims to not only retrieve vehicle class statistics, but also prevent accidents by recognizing vehicles carrying dangerous goods. In this paper, we describe a method to classify vehicle images that experience different geometrical variations and challenging photometrical conditions such as those found in road tunnels. Unlike previous approaches, we propose a classification method that does not rely on the length and height estimation of the vehicles. Alternatively, we propose a novel descriptor based on trace transform signatures to extract salient and non-correlated information of the vehicle images. Also, we propose a metric that measures the complexity of the vehicles' shape based on corner point detection. As a result, these features describe the vehicle's appearance and shape complexity independently of the scale, pose, and illumination conditions. Experiments with vehicles captured from three different cameras confirm the saliency and robustness of the features proposed, achieving an overall accuracy of 97.5% for the classification of four different vehicle classes. For vehicles transporting dangerous goods, our classification scheme achieves an average recall of 97.6% at a precision of 98.6% for the combination of lorries and tankers, which is a very good result considering the scene conditions.

Keywords: Vehicle classification, tunnel surveillance, trace transform, shape complexity

1. INTRODUCTION

In the last decade, several tragic accidents occurred in road tunnels, drawing the attention of authorities to adopt measures to reinforce their security. Unlike to highway or urban settings, in road tunnels, the traffic flow takes place in a confined space reducing the driver's maneuverability space to react against unexpected incidents. In the event of fire, enclosed environments also boost the chances of being killed due to the inhalation of toxic gases and the extreme temperatures.

Determining the type of vehicles passing through a tunnel is an important source of information in order to prevent potential risks. Particularly, tracking vehicles carrying dangerous goods is of paramount importance in road tunnel surveillance. For instance, a lorry transporting margarine and flour triggered a fire at the Mont Blanc* tunnel claiming 39 lives. This sort of events prompted the authorities to revise the security protocols and improve the infrastructure in tunnels. An extensive deployment of closed-circuit television systems (CCTV) was one of the outcomes that is now part of the current security directives. As a result, boundless information received from the cameras has surpassed the capabilities of tunnel operators. Hence, computer-vision schemes are needed to automatically extract and categorize information in an efficient and robust way with a minimum human intervention.

Vehicle classification has been extensively studied for highway and urban traffic settings as reviewed by Buch et al.¹ Most of the related works deal with the vehicle classification problem under good and steady illumination conditions. More challenging scenarios for urban settings have been studied by Buch et al.² by comparing the vehicle silhouettes against projected 3D models of several vehicle classes. On the other hand, vehicle classification at nighttime has been studied by Chen³ and Robert⁴ employing PCA-based eigenspaces as appearance feature. Particularly, in road tunnels, vehicle detection and tracking has been explored by Wu⁵ and Niño.⁶ However, no information about vehicle classification is reported.

Edge maps, SIFT descriptors, and region-based features are the most common methods employed in the literature to describe the vehicle appearance. Unlike previous works, in this paper we propose a classification

Corresponding author: Andrés Frías-Velázquez E-mail: Andres.FriasVelazquez@telin.ugent.be

*<http://news.bbc.co.uk/2/hi/europe/304946.stm>

method that does not rely on the length and height estimation of the vehicle detection. Instead, we extract salient signatures from the trace transform to model the vehicle appearance. Also, we suggest a feature that expresses the structural shape complexity of the vehicles based on interest point detection. Then, these features serve to train a multi-class SVM classifier to discriminate between car, van, lorry and road tanker. In overall, our approach achieves a correct classification rate (CCR) of 97.5%.

The rest of the paper is organized as follows: In section 2, the trace transform signatures and the shape complexity feature are described. Section 3 shows the method to train the features proposed. The experimental results are provided in section 4. Finally, the conclusions of this work are given in section 5.

2. VEHICLE IMAGE DESCRIPTORS

As stated in section 1, a novel set of features are proposed in order to extract salient information that allows to describe the vehicle image in a robust way against geometrical transformations and photometrical changes experienced in tunnels.

2.1 Orthonormal Trace Transform Signatures

The *Trace transform* is a radial projection transformation that allows one to construct image features that are invariant to a chosen group of image transformations.⁷ From the trace transform we can apply the so called *P-invariant functionals* in order to extract several signatures from an image, also called circus functions, to describe the appearance of an object. According to Petrou and Kadyrov,⁷ these signatures have proved to be invariant to affine transformations, illumination variations, and even partial occlusion.

Unfortunately, circus functions yield high recognition rates at a expense of also high computational power. In order to reduce the computational requirements while keeping a good recognition performance, we propose a method to derive uncorrelated circus functions. An orthonormality condition is enforced between circus functions in order to maximize the information provided by each signature. We have found that such condition does not hold for most of the circus functions derived with the traditional functionals.⁸ As a result, several of those signatures convey redundant information that do not contribute with new discriminatory evidence, and thus representing a computational burden. In order to ensure orthonormality, the circus functions have to fulfill the following condition

$$\rho = \frac{\langle h_n(\theta), h_m(\theta) \rangle}{\|h_n(\theta)\| \|h_m(\theta)\|} = \delta(n - m) \quad (1)$$

where $h_n(\theta) = \langle \psi_n(p), G(p, \theta) \rangle$ is the circus function resulting of the inner product between the weighting function $\psi_n(p)$ and the trace transform sinogram $G(p, \theta)$. Note that p and θ are the radial and angular variables in the Hough domain, respectively. If we assume that the sinogram is orthonormal, that is $\langle G(p, \theta), G(p', \theta) \rangle = \delta(p - p')$, the correlation coefficient between signatures lies exclusively on the weighting functional as follows

$$\rho = \frac{\langle \psi_n(p), \psi_m(p) \rangle}{\|\psi_n(p)\| \|\psi_m(p)\|} = \delta(n - m) \quad (2)$$

Hermite functions comply with the orthonormal requirements stated in (2). Moreover, multiple weighting kernels can be easily derived by using the following recurrence formula $\psi_{n+1}(p) = \sqrt{2/n+1} p \psi_n(p) - \sqrt{n/n+1} \psi_{n-1}(p)$, where $\psi_n(p) = (-1)^n (2^n n! \sqrt{\pi})^{-1/2} e^{p^2/2} \frac{d^n}{dp^n} e^{-p^2}$. Consequently, the novel group of diametrical functionals based on Hermite functions are defined as follows

$$\mathcal{H}_n(g(p)) = \int_{\mathbb{R}} \psi_n(z) \gamma(z) dz \quad (3)$$

where $z = p - c$ and $c = \text{median}(\{p_k\}_k, \{|g(p_k)|\}_k)$, while $g(p)$ represents a diametric trace line of the sinogram $G(p, \theta)$. Note that $\gamma(z)$ is a diametric trace line of the sinogram orthonormalization, which is obtained by solving the *orthonormal Procrustes problem*⁹

$$\Gamma = \underset{\Omega}{\text{argmin}} \|\Omega - G\|_F \quad \text{subject to} \quad \Omega^T \Omega = I \quad (4)$$

where Ω is any orthonormal matrix, and Γ is the nearest orthonormal matrix.

In summary, the orthonormal circus functions are generated following the next steps:

1. Generate the trace transform sinogram $G(p, \theta)$, as described by Petrou and Kadyrov.⁷
2. Change the domain of the sinogram from p to z using the following transformation $z = p - c$, given that $c = \text{median}(\{p_k\}_k, \{|g(p_k)|\}_k)$.
3. Orthonormalize the sinogram by solving the Procrustes problem as described by Schönemann.⁹
4. Apply the Hermite functionals (3) over the orthonormal sinogram Γ .

This procedure is repeated for the different sinograms derived from the image using different T -functionals.

2.2 Shape Complexity

Object similarity is strongly determined by the shape of the compared objects. Thus, embedding the shape of an object into quantitative features that describe its human perception is the ultimate goal of shape complexity. For instance, Chen and Sundaram¹⁰ have proposed a method to estimate the complexity of 2D shapes based on distance entropies of contour points. For our application, we aim to derive a feature that describes the complexity of vehicle shapes. Hence, we propose a metric that assumes the complexity of a given vehicle shape proportional to its vehicle type. For instance, the shape of a car can be considered less complex than a lorry and even much less than a road tank. Therefore, in order to derive such metric, we use the FAST¹¹ corner detector to extract some structural information of the vehicles. Then, we employ a global distance measure to reflect such structural shape information into a quantitative manner. The complexity metric is formally defined as follows

$$C = \sum_{i=1}^N \|\mathbf{q}_i - \mathbf{g}\| \quad (5)$$

where \mathbf{q}_i represents the vector coordinates of one of the N corners found in the foreground image of the vehicle, while \mathbf{g} is the centroid of the vehicle silhouette. The proposed metric measures the number of corners detected and their distribution around the centroid of the vehicle. We expect that as the vehicle size increases, the number of corners detected also increases, but with a larger dispersion around the centroid. In Fig. 1 we can see the distribution of the corner detections around the vehicle’s centroid for each vehicle class. Note that the bodywork of a car (*e.g.* Fig. 1a) is perceived as a flat surface generating small number of corner detections located close to the vehicle centroid. In the case of vans (Fig. 1b), the bodywork is very similar to a car; however, most of the corner detections are farther from the centroid. Cargo vehicles such as lorries and road tankers normally have a more “textured” bodywork, providing a larger number of corner detections than small vehicles. Fig. 1c shows that the corner detections found in lorries are distributed around the vehicle silhouette, while in road tanks (Fig. 1d) the distribution is more uniform through the bodywork. Hence, the complexity metric provides a compact and discriminative source of information to support the vehicle classification scheme.

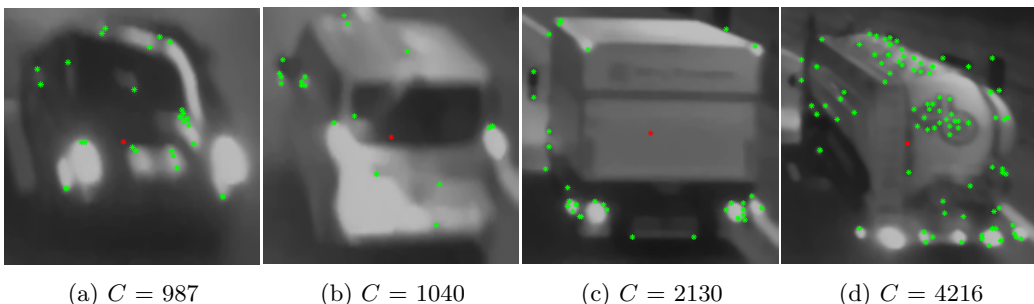


Figure 1: Shape complexity for each vehicle class, green asterisks correspond to corner detections while the red one indicates the centroid of the vehicle silhouette.

3. FEATURE TRAINING

The proposed features based on the trace transform describe the image vehicles in terms of signatures. These signatures could be used by the classifier *as-is* by collecting them in a feature vector. However, a higher-dimensional space may lead to an inefficient classification scheme in terms of information compactness. Hence, it is necessary to summarize the signature information in a concise fashion. Note that we expect similar signatures from same-class vehicles, and dissimilar features for vehicles from distinct classes. As a result, we can describe a signature by measuring its similarity to a representative group of signatures of each vehicle class.

In order to have a compact representation of the training signatures, some reference signatures \mathbf{r}_k are generated for each vehicle category k . These reference signatures are derived by clustering similar training signatures, and then averaging the signatures within the same cluster. This procedure is achieved by applying the Independent Component Analysis (ICA) to the training signatures of each class group, as stated in (6). The column vectors $\mathbf{h}_k = (h_k^1, \dots, h_k^{M-1}, h_k^M)$ represent the training signatures of the class k , while \mathbf{u}_k are the independent components. The matrix \mathbf{A} is known as the mixing matrix. Then, 95% of the variance of \mathbf{A} is regarded as a subset of column vectors stored in \mathbf{P} . This last matrix is eventually used by the *k-means* clustering to return the indices of the clustered signatures, that is $\boldsymbol{\lambda} = kmeans(\mathbf{P})$. Finally, the reference signatures are composed by averaging the signatures with the same cluster label retrieved by $\boldsymbol{\lambda}$, as shown in (7).

$$\mathbf{u}_k = \mathbf{A}^{-1}\mathbf{h}_k \quad (6) \quad \mathbf{r}_k = AVG(\boldsymbol{\lambda}, \mathbf{h}_k) \quad (7)$$

The set of reference signatures is denoted as follows $\mathbf{r}_k = (r_k^1, \dots, r_k^{L-1}, r_k^L)$, where $L \ll M$ allowing to reduce the number of signatures to compare with. Consequently, any signature s can be described in terms of its similarity to the reference signatures as follows

$$f_{s,r_k} = \max\{\rho_{s,r_k^1}, \dots, \rho_{s,r_k^{L-1}}, \rho_{s,r_k^L}\} \quad (8)$$

where ρ_{s,r_k^l} is the correlation coefficient between the signature s and the reference signature l of the class k . As a result, f_{s,r_k} returns the maximum similarity between s and the reference signatures for a given vehicle class. In this way, the signature information is collected in the feature vector $\phi_{s,r} = (f_{s,r_1}, \dots, f_{s,r_{K-1}}, f_{s,r_K})$, reducing the feature space into K dimensions, which is the number of classes.

Note that if we use all the training samples to compute the reference signatures and then we generate the training feature vectors with the same signatures, the training pool may experience over-fitting. That is, the training feature vectors lack of independence due to the correlation between the reference and training signatures. Therefore, the training process is carried out using *cross-validation with active pattern selection*, which consists of the following steps:

1. *Pattern Selection*: Find the vehicles whose signatures have the largest deviation from the reference signatures of their respective class.
2. *Cross-validation*: After finding the “worst” vehicle samples that may contribute to the classification error, we apply the leave-one-out cross-validation method to generate the training pool.

The training process described above is repeated for the J signatures that describe every vehicle image. Therefore, J feature vectors like $\phi_{s,r}$ are derived from all the signatures. Finally, these feature vectors are averaged as follows $\Phi = \frac{1}{J} \sum_{j=1}^J \phi_{s,r}^j$ and the shape complexity feature described in section 2.2 is integrated into the final feature space defined by the tuple (Φ, C) in \mathbb{R}^{K+1} .

After the training process is performed, the testing samples are derived by comparing the signatures of the test vehicle images against the reference signatures, and then composing their respective feature vector. Note that in this case the reference and testing signatures are independent to each other, so they can be compared straightforwardly without further processing.

It is worth to remark that the information provided by the feature vector Φ not only comes from the intrinsic information of the signature itself, but also from its similarities and differences to the reference signatures of each vehicle class. This sort of information embedding allows to transfer the signature information into a more discriminative and compact feature descriptor.

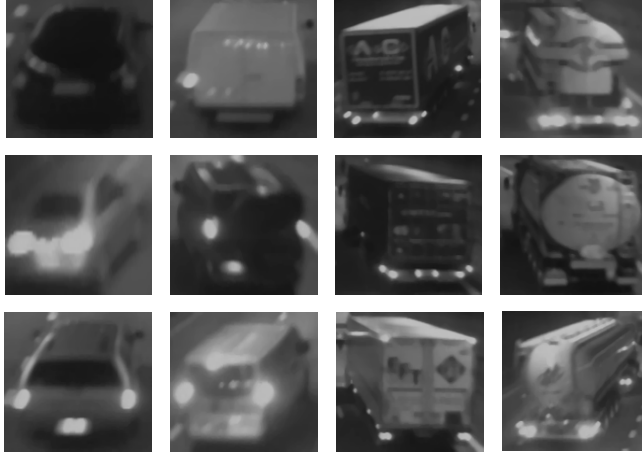


Figure 2: From left to right, each column shows some images of cars, vans, lorries and tankers captured in a road tunnel.

4. EXPERIMENTAL RESULTS

The proposed classification scheme intends to discriminate among 4 different vehicle classes like car, van, lorry and tanker. The classification experiments were performed with a database of 1131 vehicle images captured in a road tunnel. The database is composed of 756 cars, 84 vans, 264 lorries, and 27 tankers. In Fig. 2, we present an image array with some examples of those vehicles. The vehicle images are 8-bit gray-scale and their size is 85 x 85 pixels. Note that the vehicles were captured in the near field of view due to the limited tunnel height. Consequently, the vehicle images may experience translation, scale, and projective view variations. Furthermore, the vehicle appearance is strongly affected by the vehicle lights and uneven scene illumination. As a result, the combination of these variations turns the vehicle classification into a challenging problem. Conversely, most of previous works intended for highway and urban settings deal with more indulgent conditions, since their images are captured in the mid field of view, with a fixed view angle and under daylight.

For each image we computed 4- T and 5- P functionals resulting 20 orthonormal circus functions. The T -functionals used correspond to G14-G17 described by Petrou and Kadyrov,⁷ while the P -functionals correspond to the first five Hermite functionals proposed in (3). The shape complexity feature was computed using a 9-point FAST corner detector and using non-maximal suppression to control the number of corners detected.

The evaluation of the classifier was made with a 10-fold cross-validation scheme. Every training fold was processed as denoted in section 3. The classification performance can be assessed in the confusion matrix depicted in Table 1. The overall accuracy of the system is 98.6% for the training set and 97.5% for the testing set. From Table 1, we can see that most of the classification errors come from the confusion between car and van, as well as between lorry and tanker. This result is expected due to the large similarity between these classes. However, despite the significant similarities between car and van, the orthonormal circus functions were able to well discriminate between these two classes.

The information presented in Table 1 can be also summarized in terms of recall and precision percentages, as shown in Table 2. For cars, the precision and recall were very good, and as mentioned above, most of the errors are due to the confusion with vans. Note that the number of false positives ($FP = 10$) and negatives ($FN = 10$) between cars and vans are much smaller than the sample size ($N_{car} = 756$) of cars. However, these errors have a greater impact on the precision and recall percentages of vans, since its sample size is much smaller ($N_{van} = 84$). A similar phenomenon occurs between lorries and tankers. The precision and recall percentages for lorries were very good, but the number of misclassified samples between lorries and tankers have a greater influence on the tankers' classification due to the large sample-size difference.

For the recognition of vehicles transporting dangerous goods is of utmost importance not to misclassify any lorry or tanker. Consequently, recall is more relevant than precision. In this domain, our method achieves a recall greater than 96% for lorries and tankers, showing its reliability to recognize hazardous vehicles.

Table 1: Confusion Matrix

		Ground Truth			
		Car	Van	Lorry	Tanker
Classified	Car	745	10	2	0
	Van	10	74	1	1
	Lorry	1	0	258	0
	Tanker	0	0	3	26

Table 2: Precision and recall performance

	Car	Van	Lorry	Tanker
Recall	98.5%	88.1%	97.7%	96.3%
Precision	98.4%	86.0%	99.6%	89.6%

We can conclude that the orthonormal circus functions convey the vehicle appearance information in a salient and compact way. In fact, the overall performance is strongly supported by the information provided by these signatures, since the accuracy of the system is 96.3% by using the signatures alone. Furthermore, the signatures were able to deal with illumination variations as well as the different vehicle views.

5. CONCLUSIONS

A novel set of features for vehicle classification in tunnels were proposed in this paper. The orthonormal circus functions were able to successfully describe the vehicle’s appearance despite the challenging scene conditions that prevail in tunnels. Also, the classification results suggest that the discriminatory information among classes was preserved despite their similar appearance and the feature scale invariance. For vehicles transporting dangerous goods, our classification scheme achieves a recall of at least 96% for lorries and tankers, which is a very good result considering the scene conditions. However, further efforts have to be made to minimize the risk of misclassifying these two classes.

ACKNOWLEDGMENTS

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

REFERENCES

- [1] Buch, N., Velastin, S., and Orwell, J., “A review of computer vision techniques for the analysis of urban traffic,” *IEEE Trans. on ITS* **12**, 920–939 (sept. 2011).
- [2] Buch, N., Orwell, J., and Velastin, S., “Detection and classification of vehicles for urban traffic scenes,” in [*5th Int. Conf. on Visual Inf. Eng.*], 182–187 (aug 2008).
- [3] Chen, Y., Wu, B., Huang, H., and Fan, C., “A real-time vision system for nighttime vehicle detection and traffic surveillance,” *IEEE Trans. on IE* **58**, 2030–2044 (may 2011).
- [4] Robert, K., “Night-time traffic surveillance: A robust framework for multi-vehicle detection, classification and tracking,” in [*6th IEEE AVSS’10*], 1–6 (sept. 2009).
- [5] Wu, B., Kao, C., Liu, C., Fan, C., and Chen, C., “The vision-based vehicle detection and incident detection system in hsueh-shan tunnel,” in [*IEEE Int. Symp. on Ind. Electronics*], 1394–1399 (30 2008-july 2 2008).
- [6] Niño, J., Jelača, V., Frías, A., Pižurica, A., Philips, W., Rios, R., and Tuytelaars, T., “Non-overlapping multi-camera detection and tracking of vehicles in tunnel surveillance,” in [*DICTA ’11*], 591–596 (dec. 2011).
- [7] Petrou, M. and Kadyrov, A., “Affine invariant features from the trace transform,” *IEEE Trans. PAMI* **26**, 30–44 (2004).
- [8] Frias-Velazquez, A., Ortiz, C. A., Pizurica, A., Philips, W., and Cerda Villafaña, G., “Object identification by using orthonormal circus functions from the trace transform,” in [*ICIP*], 2153–2156, IEEE (2012).
- [9] Schönemann, P., “A generalized solution of the orthogonal procrustes problem,” *Psychometrika* **31**, 1–10 (1966).
- [10] Chen, Y. and Sundaram, H., “Estimating complexity of 2d shapes,” in [*IEEE 7th Workshop on Multimedia Signal Processing*], 1–4 (nov 2005).
- [11] Rosten, E. and Drummond, T., “Fusing points and lines for high performance tracking,” in [*IEEE ICCV*], **2**, 1508–1511 (October 2005).