Analysis of Features for Rigid Structure Vehicle Type Recognition

V.S.Petrović and T.F.Cootes
Imaging Science Bimedical Engineering
University of Manchester, M13 9PT, UK
v.petrovic@man.ac.uk and t.cootes@man.ac.uk

Abstract

We describe an investigation into feature representations for rigid structure recognition framework for recognition of objects with a multitude of classes. The intended application is automatic recognition of vehicle type for secure access and traffic monitoring applications, a problem not hitherto considered at such a level of accuracy. We demonstrate that a relatively simple set of features extracted from sections of car front images can be used to obtain high performance verification and recognition of vehicle type (both car model and class). We describe the approach and resulting system in full, and the results of experiments comparing a wide variety of different features. The final system is capable of recognition rates of over 93% and verification equal error rates of fewer than 5.6% when tested on over 1000 images containing 77 different classes. The system is shown to be robust for a wide range of weather and lighting conditions.

1 Introduction

Increased security awareness in recent years has made the need for various authentication technologies increasingly pertinent. Various access control systems that use techniques such as biometrics and smart cards are increasingly applied to authenticate and restrict access to users and intruders respectively. Recently, vehicle based access control systems for buildings, outdoor sites and even housing estates have become commonplace. Additionally, various traffic monitoring and control systems that depend on user (man+vehicle) identification, such as congestion charging would also benefit by augmenting existing number-plate recognition with an additional authentication mechanism. Given an image containing a frontal view of a vehicle (car), a system is proposed here that determines its exact class (make and model). The aim is to obtain reliable classification of a vehicle in the image from a multitude of possible classes (vehicle types) using a limited number of prior examples (only one per class).

Although classification of road going vehicles has been a subject of interest in the past, e.g. traffic control systems and toll levy automation, vehicle type recognition has not hitherto been considered at this level of accuracy. Instead, most systems either detect (clasify vehicle or background) or classify vehicles in broad categories such as cars,

BMVC 2004 doi:10.5244/C.18.61

buses, heavy goods vehicles (HGVs) etc. [5, 2, 6, 7, 11, 1]. Kato et. al. [5] propose a vehicle detection and classification method based on the multi-clustered modified quadratic discriminant function (MC-MQDF) that is reported to exhibit high levels of detection of a range of vehicle types against road environment backgrounds. For the same application, Matthews et. al. [7] propose a region of interest designator based on simple, horizontal and vertical edge responses and shadow detection followed by a Principal Component Analysis (PCA) feature extractor. Features are input into a multi-layer perceptron (MLP) network that discriminates between vehicle and background [7].

Parameterised 3D deformable models of vehicle structure are another approach used in classification between broad categories of vehicles [2, 6, 11]. Ferryman et. al. [2] use a PCA description of the manually sampled geometric data to define a deformable model of a vehicles 3D structure. By fitting this model to an image, both the pose and structure of the vehicle can be recovered and used to discriminate between the different vehicle categories [2]. An extension of this approach uses MLP networks to perform the classification based on the model parameters [11]. A similar approach is adopted in Lai et. al. [6] where a deformable model is fit to the image in order to obtain vehicle dimensions, which are the basis for discrimination between vehicle categories. A related approach using deformable templates for vehicle segmentation and recognition is employed in [1].

The recognition system proposed in this paper is based on recognising rigid structure samples obtained using specific feature extraction techniques from an image of the object (vehicle). Recognition is initiated through an algorithm that locates a reference segment on the object, in this case the front number plate. The location and scale of this segment is used as reference to define a region of interest in the image from which the structure is sampled. A number of feature extraction algorithms that perform this task, including direct and statistical mapping methods are investigated. Feature vectors are finally classified using simple nearest neighbour classification. Different system configurations are tested on a realistic data set of over 1000 images and show that this approach achieves very high levels of both identification and verification performance. In the following the method is described in more detail and the results of verification and recognition performance on a large dataset of realistic vehicle images are presented.

2 Vehicle Type Recognition

The scope and complexity of the recognition problem considered in this paper is exemplified by the extensive database of car images, illustrated in Figure 1, with 1132 images ordered into 77 distinct classes, such as Mercedes A class, Ford Puma, Peugeot 406 etc. Many classes describe different versions of the same car, e.g. Fiat Punto and Fiat Punto New. In our experiments, 105 images are used for registration (training) and 1027 for evaluation of the proposed system. Subtle varieties of the same model, e.g. different bottom grill between two Puntos in the bottom right of Figure 1, are represented by additional examples in the otherwise one example-per-class registration set. This approach is considered as more testing for the recognition process compared to using several examples of each class displaying more variation in lighting and pose.

The images in the dataset were obtained over a period of two months and exhibit a variety of weather and lighting conditions. Most are outdoor, although a proportion (10%) were captured in a multi-storey car park and exhibit extreme lighting conditions.

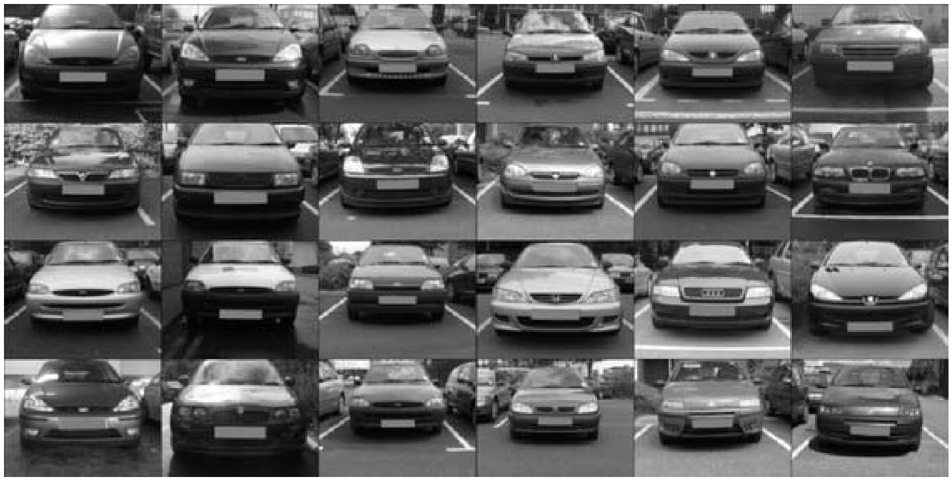


Figure 1: Examples from the data set of frontal car images

All contain frontal views of a single vehicle (with no occlusion) captured from varying distance and a height of approximately 1.2m. However, the camera was not fixed and there is significant variation in both the scale and in-plane rotation of the vehicle in the images. In particular, the horizontal axis is within a range of ± 5 degrees of the image horizontal. The images are 640x480 colour pixels, although the proposed system consideres only grey level intensities produced using a weighted sum of the colour channels.

The recognition system proposed in this paper is based on the principle of locating, extracting and recognising normalised structure samples taken from a reference image patch on the front of the vehicle, structure Figure 2. The process initialled by locating a reference segment on the object (in this case number-plate) and defining a region of interest (RoI) relative to it. The RoI is processed by the feature extraction element to define a normalised sample of the structure within it. The structure is expressed in a feature vector of pre-defined length that is representative of the vehicle identity. Finally, simple nearest neighbour classification is used to determine the vehicle type associated with each vector.

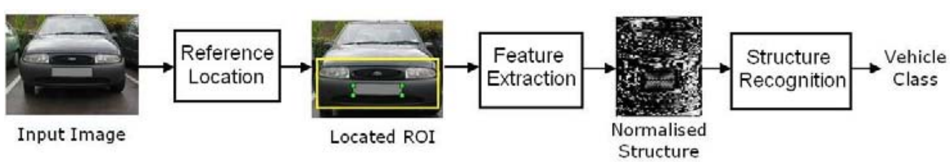


Figure 2: Basic structure of the automatic vehicle type recognition system

3 Recognition Approach

3.1 Reference Detection and Location

The location and scale of a reference structure on the object defines a reference frame for the region of interest to be sampled. An RoI defined relative to the number-plate is thus independent on the actual location and scale of the vehicle in the image.

Number-plates are highly regular rectangles. To locate them, our system finds all possible righ-angle corners using suitably tuned, separable gradient filters. A hierarchical algorithm for aggregation of corner points into valid rectangular constallations is used to generate hypotheses for the plate location in the image. A number of scale and aspect constraints are used to remove unsuitable candidates, many caused by the characters on the plate and regular vehicle structure features, and of the remaining candidates the one with best corner structure fit to each of its corners is chosen. RoI is defined in terms of number plate width w_p relative to its center, as shown on Figure 3.

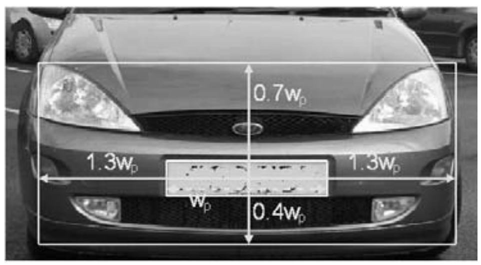


Figure 3: Region of Interest defined relative to vehicle number plate

3.2 Feature Extraction

Feature extraction from the Region of Interest provides a structure representation used to recognise the object. The approach used in vehicle type recognition is illustrated in Figure 4. Initially, the RoI is down-sampled to a desired fixed resolution NxM (preceded by suitable smoothing). For vehicle images, horizontal resolution is more severely reduced as the structure is significantly more redundant in that direction.

From the scale normalised image I', structure was extracted into a feature vector using a broad range of both direct and statistical structure mapping methods. Direct mapping with relatively simple, local structure operators such as gradients, produces feature samples same size as the original signal that preserve both the global structure and finer localised details. A subset of the most significant feature extraction techniques investigated is summarised in Table 3.2 [8, 9, 3, 10].

Raw image values concatenated into a feature vector $\mathbf{f}_{raw}$, form the most direct representation of vehicle appearance and serve as a general reference for other, more complex and robust representations. Sobel gradients s_x and s_y, as derivatives, are less sensitive to lighting conditions and can be even more robust by making them independent of contrast; Direct Normalised (DN) gradients are vectors on the unit circle with orientation in the range $[-\pi$ to $\pi)$, $\mathbf{f}_{DN}$. Locally Normalised (LN) gradients g^{LN} also take into account local structure by normalising each Sobel gradient (s_x, s_y) with the average edge strength

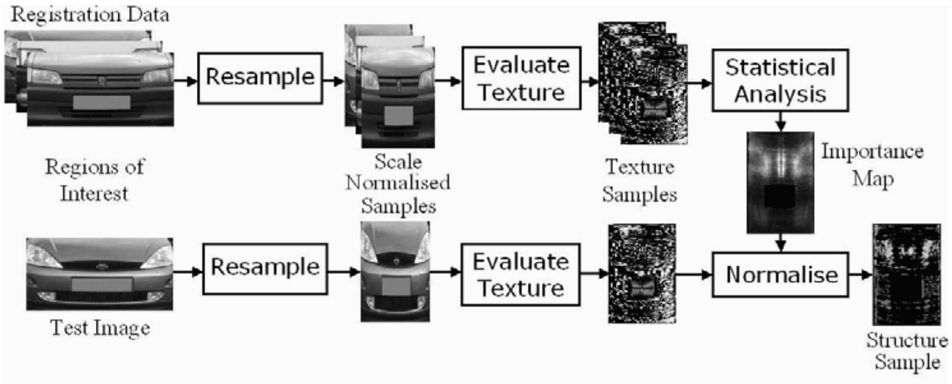


Figure 4: Feature extraction process in the proposed vehicle recognition system

Approach	Structure Representation
Raw Image	I'
Sobel Edge Response	(s_x, s_y)
Edge Orientation	$s_\alpha = arctan(s_x/s_y)$
Direct Normalised Gradients	$(g_x^{DN}, g_y^{DN}) = (\frac{s_x}{\sqrt{s_x^2+s_y^2}}, \frac{s_y}{\sqrt{s_x^2+s_y^2}})$
Locally Normalised Gradients	$(g_x^{LN}, g_y^{LN}) = (\frac{s_x}{g_L}, \frac{s_y}{g_L})$
Square Mapped Gradients	$(g_x^{SM}, g_y^{SM}) = (\frac{s_x^2-s_y^2}{s_x^2+s_y^2}, \frac{2s_xs_y}{s_x^2+s_y^2})$
Harris Corner Response	R_{harris}[3]
Spectrum Phase	$\phi = arg(FFT(I'))$

Table 1: Feature extraction approaches investigated

$\sqrt{s_x^2+s_y^2}$, g_L over an LxL neighbourhood around the pixel. Pixels lying on prominent edges thus exhibit higher values while less significant gradients tend towards zero. Further robustness can be added by restricting the range of orientation to 0 to π. The Square Mapped (SM) gradients (g_x^{SM}, g_y^{SM}) lie on the unit circle and represent axis parallel (g_x^{SM}) and diagonal (g_y^{SM}) components of signal change. This representation is robust to noise and small changes in edge direction, which makes it particularly suitable for local structure representation. Harris corner detector [3], which measures the similarity of local structure to a corner was also considered since vehicle fronts contain regular features and numerous, tightly localised and well-defined corners. Finally, spectrum phase $\mathbf{f}_\phi$, of the scale normalised image sample was also considered as a representation of global structure.

Figure 5 visualises some of these structure representations. The top row contains the 50x120 pixel normalised raw image sample, Harris corner detector response, spectrum phase, and the *x* and *y* Sobel edge responses. The second row in turn contains **w** (see below) and the *x* and *y* components of the LN and SM gradients. Vehicle structure is easily discernable only in gradient images.

As an alternative to direct mapping methods, a statistical approach using Principal Component Analysis [4] was considered to determine a low dimensional, optimal structure representation that contains the majority of between class variation. Initially, feature

Figure 5: Feature extraction examples, see text for explanation

extraction is performed on the registration set using the methods outlined above. PCA is then applied to the registration set, obtaining at most N_{reg} eigenvectors of the data covariance matrix, where N_{reg} is the number of registration set examples. Eigenvectors corresponding to the largest eigenvalues and pre-defined proportion of the total variation in the data are ordered into a transformation matrix $\mathbf{P}$, to project original feature vectors into a lower dimensional subspace, $\mathbf{f}_{PCA} = \mathbf{P}^T\mathbf{f}$. Thus, with 105 registration examples, section 2, the transformed PCA space can have at most 105 dimensions.

3.3 Statistical Sample Normalisation

In order to improve recognition performance of rigid structure feature extraction, additional normalisation of structure samples is proposed. The aim is to ephasise areas of the rigid structure that exhibit the greatest variation over the registration set (between different classes). Initially, the average confidence of each structure element (pixel) is found as the average, normalised, edge strength at the pixel across the registration set, $\mathbf{s} = E[\mathbf{s}_g^i]$ where $\mathbf{s}_g^i$ is edge strength accross the i-th registration image segment. This elementwise confidence score modulates the elementwise variance of the registration structure vectors, $\mathbf{v}$, into a weighting vector $\mathbf{w}$, equation 1. Accordingly, $\mathbf{w}$ should exhibit larger values in locations where feature samples of various classes exhibit higher energy and differ significantly. Finally, $\mathbf{w}$ is used to weight and normalise feature vectors extracted from image data, equation 2.

$$\mathbf{w} = \frac{diag(\mathbf{v})\mathbf{s}}{|diag(\mathbf{v})\mathbf{s}|} \tag{1}$$

$$\mathbf{f}' = \frac{diag(\mathbf{w})^{\frac{1}{2}}\mathbf{f}}{|diag(\mathbf{w})^{\frac{1}{2}}\mathbf{f}|} \tag{2}$$

For vehicle type recognition the elements of $\mathbf{w}$ corresponding to the number plate region are set to 0, as the licence characters do not correlate with vehicle class and merely confound classification. The $\mathbf{w}$ obtained for the x component of $\mathbf{f}_{SM}$, shown in the bottom left of Figure 5 clearly highlights the structures important for discriminating between various vehicle classes: shape of the lights and the badge.

3.4 Classification

We investigated two distance measures to compare test and registration samples, the dot product $d = 1 - \mathbf{f}_1^T\mathbf{f}_2$ and the euclidean distance $d = |\mathbf{f}_1 - \mathbf{f}_2|$. The identity of the test sample was then determined using the nearest neighbour rule. We found the two measures gave similar results, with the dot product slightly outperforming the euclidean distance. Results presented below are thus given for the dot product measure.

4 Results

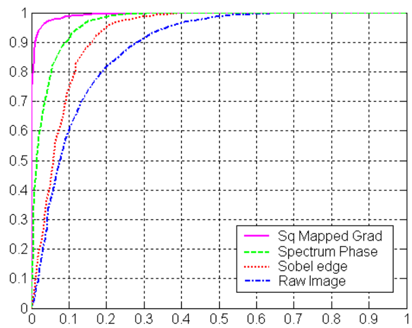
The various feature extraction methods described in the previous section were tested within the proposed recognition system using the vehicle data described in section 2. Both their direct identification (*what type of vehicle is it?*) and verification (*is it this type of vehicle?*) performance is given in Table 2, where P_{id} is the probability of correct identification and Equal Error Rate, EER the point at which probabilities of false acceptance and false rejection are equivalent. The results were obtained using manually marked number-plate locations, i.e. ideally defined RoIs for both registration and test images, so that only feature extraction strategy influences the performance. In each case, the scale-normalised sample was 50x120 pixels and classification used the dot product distance metric. For the PCA representation, 95% of the registration set variation was considered as it produces the optimal results for most of the feature extraction techniques considered. The resulting number of principal components (dimensions) for each feature representation is given in the rightmost column of Table 2.

	Identification P_{id}(%)		Verification EER(%)		PCA(95% var)
	Direct	PCA	Direct	PCA	N_{dims}
Locally normed grads	47.3	79.3	23.8	9.8	91
Edge orientation	54.9	57.9	19.8	18.9	96
Raw pixel values	62.8	4.5	19.8	46.7	68
Harris corner detector	72.4	25.0	9.6	29.6	70
Sobel edge responses	75.0	74.7	13.8	13.3	85
Spectrum phase	82.5	48.4	9.4	23.8	97
Direct normed grads	84.5	83.9	8.1	8.7	93
Square mapped grads	97.7	95.1	3.5	4.9	95

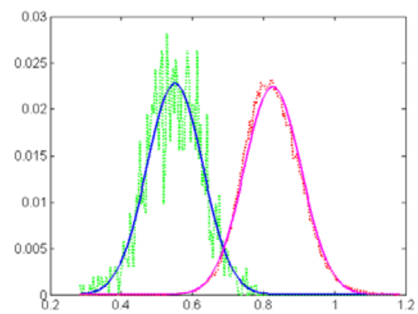
Table 2: Recognition performance of various feature representations, Parking Lot data set

For all feature types, except LN gradients, rigid, direct feature mapping provides for correct identification of a majority of vehicles in the test image set. At the same time, PCA produces slightly worse results for most structure representations. Square mapped gradients exhibit the best performance overall with P_{id} as high as 97.7% and EER is as low as 3.5% for direct and $P_{id} = 95.1$%, EER=4.9% for PCA structure representation. Gradient features generally perform best, with spectrum phase also achieving EER$<$ 10%. This result suggests that independence from vehicle colour and contrast is an advantage in representing vehicle structure, which is expected considering the wide range of conditions under which the test images were obtained. The square normalised gradients with limited orientation range are additionally robust to small noise induced changes in gradient orientation, key recognition information. Contrast information, incorporated into locally normalised and Sobel gradients, varies unpredictably with the conditions and more importantly with the colour of the vehicle. Finally, results close to ideal also justify the relatively simple rigid structure approach.

The effect of structure sample normalisation on performance is easily demonstrated by performing recognition with **w** set to 1 (apart from number plate). Such a system, using SM gradients, achieves lower P_{id} =95.8% and higher EER=6.5%.



a) ROC curves for different features



b) PDFs (left true, right false matches)

Figure 6: Verification ROC curves and PDFs of the true and false distances for the Square Mapped Gradient case

Verification performance is also illustrated graphically for a subset of direct feature representations on the ROC curve in Figure 6a. For an 80% correct acceptance rate, false acceptance rate of the SM gradients system is less that 0.4% (solid line). Distributions of classification distances between examples of the same and different classes are shown in dotted lines in Figure 6b. Almost perfect Gaussians, Gaussian fit shown as solid lines, indicates that an analytic confidence score associated with the recognition performed by the system can easily be derived from these distributions.

The proposed vehicle recognition system was also tested on an additional, realistic access control application data set. Frontal images were captured from a distance of 10 and a height of 3m, approximately 20 degrees of the main vehicle axis. 21 different vehicle types were enrolled with one image each and the system was tested with 204 test images. The test images were captured under a variety of athmospheric conditions (sunny, rain and overcast) and various times of day (daylight, dusk). In general, this data set contains fewer different vehicles (registered vehicles repeat in the test set), however the vehicles exhibit a higher range of out-of-plane rotation and location and the images are of lower quality than the ”Parking Lot” data set. Recognition performance for this data set, given for a chosen subset of structure representations in Table 3, correspond well to the results in Table 2 with gradient features giving the best results. This time however, both Harris corner detector and raw pixel values perform well as the number of vehicles in the test set is reduced and colour correspondency exists with the registered examples.

A number of incorrectly recognised vehicles (SM grads) from both datasets are shown in Figure 7. The main modes of failure in the ”Parking Lot” data, top row, include severe in-plane rotation and poor lighting conditions. The latter is also a source of errors with ”Access Control” data, bottom row, along with severe out-of-plane rotation.

	Identification P_{id}(%)		Verification EER(%)	
	Direct	PCA	Direct	PCA
Raw pixel values	91.2	6.4	10.5	47.2
Harris corner detector	96.6	40.7	3.5	24.1
Direct normed grads	98.0	93.1	3.4	5.8
Square mapped grads	98.0	93.6	5.9	16.5

Table 3: Recognition performance for Access Control data set

Figure 7: Recognition failures with square mapped gradient feature extraction

4.1 Automatic Recognition System Performance

While the results presented above are based on manually located RoIs and depend on the feature extraction approach only, realistic vehicle recognition systems must locate the vehicle front automatically. Our number-plate location algorithm, tested independently on 1206 ”Parking Lot” car images, correctly located the number-plate (to within 5 pixels) in 93.3% of the cases with an average error of 2.74 pixels per corner point. Integrated with Square Mapped gradient feature extraction, the whole system correctly identifies 93.3% vehicles with a verification Equal Error Rate of 5.6% for the ”Parking Lot” data. For ”Access Control” data performance is worse at $P_{id} = 87.7\%$ and $EER = 10.1\%$. We can thus conclude that reference location limits performance as by en large all the images in which the number plate is incorrectly located result in a recognition error. For both data sets, the low dimensional PCA representation gives worse performace, at best $P_{id} = 90.6\%$ and $EER = 6.8\%$ and $P_{id} = 72.1\%$ and $EER = 16.3\%$. Such high levels of performance for a highly complex recognition task (77 classes) clearly justify the adopted rigid structure approach that preserves well the global structure of object apearance and proves its robustness.

5 Conclusions

This paper presents an investigation into various feature extraction techniques in a rigid structure approach to automatic recognition of vehicle types. It was shown that certain gradient representations, such as the square mapped gradients, are capable of accurate and reliable recognition of vehicles from frontal views under a variety of conditions. Their inherent insensitivity to small changes in gradient orientation and noise coupled with a rigid sampling approach preserves both global and local object structure in a robust manner. Additional statistical normalisation of structure samples provides further recognition confidence by emphasising structural differences between various object classes. Such direct structure recognition systems are capable of Correct identification performance in excess of 93% and a verification Equal Error Rate of less than 5.7% demonstrated on a realistic data sets of over 1000 frontal images of cars taken under a range of varying conditions fully justify this relatively simple approach. At the same time, statistical structure representations, using PCA, showed less robustness in both recognition and verification. Further work on this approach involves extending the system to deal with a wider range of viewpoints and in plane rotations as well recognition of more general classes of objects.

References

[1] MP. Dubuisson-Jolly, S. Lakshmanan, and A. Jain. Vehicle segmentation and classification using deformable templates. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 18(3):293–308, 1996.

[2] J. Ferryman, A. Worral, G. Sulliva, and K. Baker. A generic deformable model for vehicle recognition. In *British Machine Vision Conference*, pages 127–136. British Machine Vision Association, 1995.

[3] C. Harris and M. Stephens. A combined corner and edge detector. In *4th ALVEY vision conference*, pages 147–151, 1988.

[4] I. Jollife. *Principal Component Analysis*. Springer-Verlag, 1986.

[5] T. Kato, Y. Ninomiya, and I. Masaki. Preceding vehicle recognition based on learning from sample images. *IEEE Transactions on Intelligent Transportation Systems*, 3(4):252–260, 2002.

[6] A. Lai, G. Fung, and N Yung. Vehicle type classification from visual-based dimension estimation. In *IEEE Intelligent Transportation Systems Conference*, pages 201–206. IEEE, 2001.

[7] N. Matthews, P. An, D. Charnley, and C. Harris. Vehicle detection and recognition in greyscale imagery. *Control Engineering Practice*, 4(4):472–479, 1996.

[8] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis and Machine Vision*. PWS Publishinh, Pacific Grove, 1998.

[9] T.F.Cootes and C.J.Taylor. On representing edge structure for model matching. volume 1, pages 1114–1119, 2001.

[10] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 1999.

[11] W. Wei, Q. Zhang, and M. Wang. A method of vehicle classification using models and neural networks. In *IEEE Vehicular Technology Conference*. IEEE, 2001.