8th IEEE International Conference on Advanced Video and Signal-Based Surveillance, 2011

Automatic Make and Model Recognition from Frontal Images of Cars

Greg Pearce and Nick Pears
Department of Computer Science
University of York, UK
nep@cs.york.ac.uk

Abstract

We investigate a range of solutions in car 'make and model' recognition. Several different feature detection approaches are investigated and applied to the problem including a new approach based on Harris corner strengths. This approach recursively partitions the image into quadrants, the feature strengths in these quadrants are then summed and locally normalised in a recursive, hierarchical fashion. Two different classification approaches are investigated; a k-nearest-neighbour classifier and a Naive Bayes classifier. Our system is able to classify vehicles with 96.0% accuracy, tested using leave-one-out cross-validation on a realistic dataset of 262 frontal images of cars.

1. Introduction

Automatic vehicle surveillance is an increasingly important technology due to a rising trend in license plate cloning. In the UK, for example, tens of thousands of license plates are cloned every year. Reasons for plate cloning range from evading speed camera fines and congestion fees to selling stolen cars and even disguising a vehicle for use in a serious crime. To deal with this, automatic recognition of the actual vehicle itself is required to supplement standard automatic number plate recognition (ANPR). Vehicle recognition systems usually either classify the vehicle into generic classes (car/lorry etc) [3], or they classify the vehicle into specific 'make and model' classes. Here we focus on the latter application of make and model recognition (MMR).

1.1. Related work

Petrovic and Cootes [10] looked for structures such as headlights and grill in car images to use as a basis for MMR. Image samples were position/scale normalised using the size and location of the license plate. A number of different features are then extracted over a region of interest. Classification uses the nearest neighbour to an input sample, by minimising a cosine distance measure. The best feature is found to be *square mapped gradients*, which are gradients formed from vertical and horizontal Sobel edge responses. An identification rate of 97.7% on more than 1000 images and a verification error rate of 3.5% was achieved.

Munroe and Madden [8] use thickened Canny edges as the extracted features, and test 3 different classifiers: k-nearest neighbour (k-NN), feed-forward neural network and a decision tree. The testing dataset is comprised of 5 classes with 30 samples of each class. Each training set is comprised of 134 images with 10 of the class being tested, 30 of each of the other classes and 4 images of unknown class. The k-NN classifier is found to be most effective with 97.46% correct identification rate, but the dataset for this result can be considered less challenging with only 5 fairly distinct looking classes of car tested.

In Clady et al's work [4], Sobel edges are extracted from the license plate based ROI and oriented-contour points are then obtained from the edges using a histogram based thresholding process. For each class an array is formed containing the oriented-contour points that are stable across the class training samples. These are then used to vote on whether or not a sample belongs to that class. The training dataset contains 50 classes and is comprised of 291 high quality frontal view images captured in car parks. The testing dataset is comprised of 830 outdoor images again with variance in lighting, angle, distance and resolution. A correct identification rate of 93.1% is reported.

Many other approaches appear in the literature for example, Huang at al's 2D-LDA approach [6], Psyllos et al. [11] use symmetry measurements and Sarfraz et al. [13] present a local energy based shape histogram to encode vehicle shape. When dealing with vehicle images at a wide variety of scales, SIFT-based approaches are popular [7].

In this paper, we investigate several of the approaches mentioned above, and compare them to our own method, which is based on normalised Harris corner strengths over recursively partitioned image regions. In the next sections, we describe our dataset, followed by our position/scale normalisation and cropping processes. The evaluated methods are then described and their performance compared.

978-1-4577-0845-9/11/$26.00

2. The evaluation dataset

In many image classification applications, a uniform distribution of training/testing image samples across the classes is used. However, this is not what one encounters for car models, both when collecting training data and testing. Some models, such as the *Ford Focus* are much more popular than other models, and this is reflected in our dataset. Our dataset consists of 262 frontal car images with 74 different 'make and model' classes, collected from the car parks around our university. There are 21 'common' vehicle classes that have 5 or more sample images and these constitute 177 images in total. There are another 53 'uncommon' vehicle classes in the dataset, which mostly have one or two samples, and these constitute the remaining 85 samples.

For the uncommon vehicle classes, we consider that we do not have enough training data to test for these particular classes. In the experiments where we test 3-nearest neighbour schemes, this is self-evident when we have less than 3 samples (2 needed for a majority vote, one for testing). Thus our testing consists of using the 177 images within the 21 'common' classes in a 'leave-one-out' scheme, yet the other 85 samples are included when matching in order to create a more difficult, realistic scenario where the feature space is populated by a large number of different classes (i.e. 74).

Images are collected so that there are small pose variations: the distance from the car ranges from 1.5m-3m and the camera pan angle varies from 10 degrees. Also images are collected in a variety of lighting conditions: midday, evening, in bright sunshine and in cloudy conditions.

3. Normalisation and ROI selection

In all images, we manually mark up the three corners of the numberplate, the fourth is computed from the other three such that the four corners form a parallelogram. Firstly, this allows us to normalise the scale, rotation and skew of the imaged license plate and hence vehicle front. Secondly it allows us to mask the plate in the database for security reasons. Finally, features from the number plate are not directly associated with the class and so should be masked in the classification process. Although this amounts to manual intervention, license plate localisation is a mature, high performance technology particularly when using active (LED) light projection onto the retro-reflective plate surface. Additionally, there are many highly successful passive plate localisation techniques [1].

Given we have the four corners of the numberplate, we can map them to canonical positions using a planar projectivity (homography). Since we restrict the corners of the number plate to be on a parallelogram, this is a 6-DOF affine mapping, and this appears to give a more stable mapping than when we specify all four corners independently and allow a full 8-DOF projectivity, see figure 1. Thus we effectively normalise the position, scale, rotation and skew of the plane containing the license plate.

Figure 1. Undesirable image warping when normalising with four corners of license plate. Notice how the vehicle's right headlamp is enlarged.

Figure 2. Example of an extracted region of interest, h is the license plate height and w is the license plate width.

The region of interest (ROI) over which the image is processed and a feature vector extracted influences both signal to noise ratio (and hence classification rate) and the speed of classification. In order to find a suitable size for the ROI, an optimal box shaped ROI was manually plotted on 60 different vehicles (after normalisation), such that the full width of the vehicle and the lights and grill were included within the image. The mean of these box coordinates was then taken as an estimate for the optimal ROI for the dataset. The position and size of the ROI relative to the license plate is illustrated in figure 2.

4. Methods evaluated

Our approach was to examine some approaches in the literature, understand their performance limitations and then try and use this information to inform a new approach. To try and keep the comparison of methods fair, we ensured that the feature vectors extracted for any one method were not excessively large and, where possible, we tried to make them approximately the same size across different methods.

4.1. Canny edges

Munroe and Madden [8] outline a MMR system using thickened Canny edge features [2]. Since, in this implementation, a feature is considered to be one pixel, this restricts the size of the image. A resolution of 150 by 150 pixels gives 22,500 pixels in total. Petrovic and Cootes [10] observed that most of the information in a vehicles front is in the vertical direction, so compressing along the horizontal and making the image square may be a better use for extra pixels than maintaining the original aspect ratio. Histogram based methods were used to set the threshold parameters automatically on a per-image basis. The edges were thickened by 1 pixel in each direction, as this was found to increase correct classification rates more than thickening by 2 pixels. Thickening improves classification rates because edge pixel values are being compared directly across the dataset. This means that if an edge in one sample is just one pixel off from the location of the same edge in another sample it will not match across the samples. Thickening the edges increases the chance that the edge pixels will line up at least partially. We evaluated the 1 pixel-thickened system on our 177 test images using a Euclidean metric in a k-nearest neighbour (k-NN) scheme. For the 1-NN scheme we obtained a correct classification rate of 79.1% and for a 3-NN scheme we obtained 81.4%. We view this very simple solution as a baseline performance which more sophisticated systems can be measured against.

4.2. Square mapped gradients

Square mapped gradients [10] are calculated using the vertical and horizontal edge responses, s_x and s_y, returned by a Sobel edge detector such that:

$$g_x = \frac{s_x^2 - s_y^2}{s_x^2 + s_y^2}, \qquad g_y = \frac{2 s_x s_y}{s_x^2 + s_y^2} \tag{1}$$

When both edge responses are zero, these undefined numbers were replaced with zeros. The values of g_x and g_y were concatenated together to form a feature vector with two values for each pixel in the input image. If a resolution of 150 by 150 was used , as with our original Canny edge based system, this would have been 45,000 features. Clearly the resolution needs to be reduced to make a fairer comparison. Petrovic and Cootes recommended a resolution of 50 by 120 which would produce a feature vector of 12,000 features which is slightly too low. Using their proportions and scaling upwards, a resolution of 68 by 162 produces feature vectors of size 22,032. An example of the feature vector created by square mapped gradients is given in figure 3. Testing on our test set of 177 vehicles gives performances of 91.0% and 89.8% correct classification for 1-NN and 3-NN classifiers respectively.

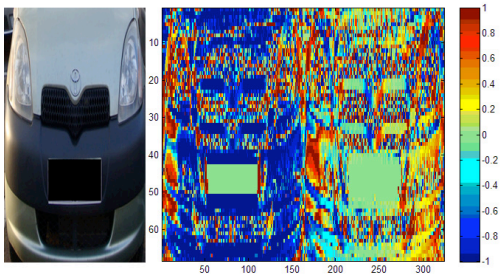

Figure 3. Square mapped gradient features. Left: the original vehicle image. Right: the square mapped gradients feature vector. The left side of the feature vector shows the g_x responses, the right side shows the g_y responses. These have been concatenated to form the full feature vector.

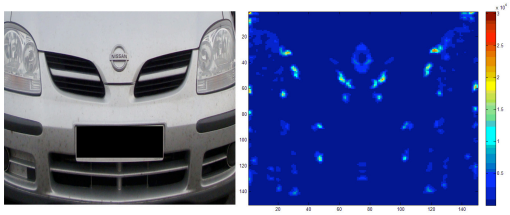

Figure 4. Harris corner strength output. Left: the original vehicle image. Right: the corner strength image.

4.3. Harris corners

One of the most well-known and often used interest point detectors is the Harris detector [5]. We used this detector along with Noble's suggested corner strength measure [9], given by:

$$C_N = \frac{I_x^2 I_y^2 - (I_{xy})^2}{I_x^2 + I_y^2} \tag{2}$$

where I_x, I_y, I_{xy} are smoothed image derivatives. An example of this Harris corner strength output is given in figure 4 and it can be seen that the output is rather uniform in many areas of the image. Clearly, as expected, responses are only evident where there are local brightness variations in orthogonal directions. As in the Canny feature test, a feature vector was comprised of 22,500 elements representing pixels in a 150 by 150 corner strength image. Testing on our test set of 177 vehicles gives a performances of 78.0% using a 1-NN classifier, a performance similar to the Canny feature (79.1%).

4.4. Recursive partitioning and local normalisation

One of the issues faced by the techniques experimented with so far is that once the image sample has been normalised with respect to position and scale, the pixels of individual structures in the image are assumed to line up across

samples. Of course these structures do not line up perfectly across all samples but can be slightly off, mainly due to variances in the normalisation stage. There are two possible approaches with dealing with this. The first is to employ a metric that has a concept of local neighbourhood, in other words replace the Euclidean metric with an 'earth mover' metric [12]. The second is to rethink the feature vector so that it is less sensitive its alignment due to localisation and scaling at the normalisation stage. One way to do this is to recursively divide the image into quadrants and construct summed feature outputs within the divisions at each level of the recursion. In this way, feature strength matching at a high level in the image has an effect on the matching as well as feature strength matching at the lower level, smaller partitions. Thus we have a recursive algorithm that gradually partitions up the image into smaller partitions, and sums the feature response at each level into a new feature value.

An advantage of this recursive structure is that feature strength sums can be normalised in a localised way, such as dividing by the sum of the feature strength in the associated higher level region. This ensures that even if a feature is detected with a much lower strength in one image sample, it will still match as long as that strength is the same in proportion to the rest of the feature strengths within its 'one-level-up' region of the image. Our feature vector is formed in such a way that the feature sums at the lowest level of recursion will form roughly 2/3s of the feature vector and will have the most effect on classification. This is preferable since feature matching in smaller and more precise regions is likely to be more discriminating than matching them at higher level regions. (In larger regions, many response distributions can sum to the same overall response, which reduces discrimination.) However, it is important that feature matching is not too precise since this degrades back into matching individual pixels; clearly an optimal value for recursion depth needs to be found. Our algorithm partitions the image into 2 columns by 2 rows each recursion, this should be suited well to classification using Harris corner strength since Harris corners are localised to small, compact areas in the image rather than stretching across large segments. Harris corners were retested using the same conditions as previously, only now using locally normalised feature strengths to a recursion depth of 5. This gives a correct classification rate of 94.9% using a 1-NN classifier, a large improvement over the 78.0% rate found when using pixel-level matching. It should also be noted that the feature vector generated using this method is only 1364 features in size, roughly 5% of the size it was beforehand, thus speeding up classification.

To investigate whether it is actually necessary to include the feature strengths from higher level regions, a new version of the algorithm was coded that only included the normalised feature strengths at the lowest level of iteration. This resulted in two extra misclassifications compared to the original algorithm, with an overall correct classification rate of 93.9%, although we need a larger dataset to verify that the higher level features are of significance. Another modification to the algorithm investigated whether the local normalisation step was actually more powerful than just normalising the feature strengths across the strength exhibited by the entire image. This gave very poor results, reducing the overall correct classification rate to just 75.7%. It would appear that it is the local normalisation step that gives the algorithm its biggest advantage. A final modification used square mapped gradient strengths rather than harris corner strengths and we obtained a 90.4% classification rate. Since this is still below that attained by Harris features this method was abandoned in favour of using Harris strengths.

4.5. Naive Bayes Classifier

We were motivated to evaluate a Naive Bayes classifier in order to exploit prior information concerning the relative frequencies of cars in each class. This classifier uses Bayes' theorem in order to evaluate the probability of each class (C), given the observed feature vector (X). The class with the largest *posterior* probability, $P(C/X)$, is selected as the predicted class. Bayes' theorem gives us

$$p(C/X) = \frac{P(X/C)P(C)}{P(X)} \tag{3}$$

where $P(.)$ is a probability. $P(C)$ is derived from the relative frequencies of the different classes in the training data. Clearly, this should closely correspond to the relative frequencies seen when the system is live. (The integration of an MMR system with an ANPR system would allow these *prior* probabilities to be adaptively learnt over time.) In the Naive Bayes classifier, the term *naive* comes from the naive assumption of independence between the observations associated with the different dimensions of the feature vector. Often, as in our case, this is an oversimplification, and yet the classifier can often give good results. This independence assumption allows us to fit separate probability density functions (PDFs) for each dimension, x_i, of the feature vector X, for each class. Then, when evaluating the likelihood, $P(X/C)$, for some class, we can simply form the product of $P(x_i/C)$ over all dimensions. This is scaled by the prior $P(C)$ in equation 3 and the maximum value is selected over all classes. Note that the denominator $P(X)$ is the same for all classes and hence does not need to be evaluated. In our implementation, we used the Naive Bayes function in MATLAB's statistics toolbox.

Using this classifier brings locally normalised Harris strengths up to a 96.0% correct classification rate from 94.9%. Square mapped gradients also improve to 96.0% from a 93.8% rate. A word of caution is required here

though. In our 'leave one out' testing scenario, the relative frequencies in the test set correspond closely to the relative frequencies in the training set and hence we have a very accurate estimate of $P(C)$ in equation 3. In practise, the level of accuracy in the Bayes Naive classifier may be harder to replicate, when compared to the k-NN classifier figures.

5. Comparative evaluation

In summary, we tested 10 different classification systems, by varying the combination of feature extraction method and classification technique. Our feature extraction methods include (i) Canny edge detection, (ii) Square mapped gradients (SMG), (iii) SMG 'improved' (by prefiltering with a 9x9 Gaussian mask), (iv) Harris corner strength and (v) Locally normalised Harris strengths (LNHS). Our classification methods include (i) 1-nearest neighbour, (ii) 3-nearest neighbour and (iii) Naive Bayes. Figure 5 shows the comparative performance of each of these MMR systems.

The Naive Bayes classifier marginally outperforms the k-NN classifier for both of the feature detectors tested. Improved SMG and LNHS significantly outperform Canny edge features. When using the k-NN classifier locally normalised Harris corner strengths slightly outperform improved square mapped gradients but when using the Naive Bayes classifier they achieve the same classification rate on our data set. Both 'improved' SMG and LNHS using the Naive Bayes classifier merit further investigation in terms of their misclassifications. The seven misclassifications in the 177 classification tests for these two systems are shown in table 1.

In many cases, both features struggled with classifying the same classes, such as in the cases of the shared *Peugeot207*, *VWGolfMk3*, *MiniCooper*, *HondaJazzMk1* and *FordFiestaMk4* failures. The *HondaJazzMk1* class accounts for 3 failures in both approaches. This is surprising since it is a fairly well populated class in the feature space, with 8 samples in total and 7 samples to match to in the training set. It may be that the samples in this class do not display enough stable features across the dataset, and the class is spread across the feature space. The Peugeot 207 class may be difficult to classify against the Peugeot 407 class. Problems with classifying Peugeot vehicles were expected in the system since they have the license plate located in a different location to most other car manufacturers. This causes a poor region of interest to be extracted. The Peugeot 207 and 407 are already fairly similar vehicles, but once the classifier is unable to use the features from the bonnet and headlights it makes for a particularly difficult classification, especially considering the variation in this region of the vehicle. Two other reoccurring misclassifications are sample 121; a *VWGolfMk3* sample misclassified both times as a *VWGolfMk2* and sample 202; a *MiniCooper* sample that misclassifies differently each time. Upon inspection, nothing of note could be seen in the misclassification of sample 121 except that the *VWGolfMk2* and *VWGolfMk3* classes are very similar. In real applications, it may be better to group models that are highly similar into larger superclasses that contain several similar models. For example, this could prevent too many false alarms where the model class does not match that associated with the ANPR read, whilst still providing a highly useful automatic surveillance system.

The single image classification speeds in a standard MATLAB implementation (Intel Core 2 Duo E6600 PC with 2GB of RAM) for various feature/classifier combinations were (i) LNHS+k-NN, 0.01s, (ii) SMG + k-NN, 0.14s, (iii) LNHS + Naive Bayes, 0.1s, (iv) SMG + Naive Bayes, 1.48s. Each classification time was calculated as an average over 5 independent experiments, using the same timing code located at the same point in the system, and 261 comparisons per experiment. A marked difference in speed can be seen between a k-NN classifier a Naive Bayes classifier. Another marked difference in speed can be seen between locally normalised Harris strengths and square mapped gradients. This is due to the smaller feature vector size of locally normalised Harris strengths, which is about one twentieth the size of the square mapped gradients feature vector.

6. Conclusions

Locally normalised Harris strengths (LNHS) classify faster due to a smaller feature vector size around one twentieth of the size of the square mapped gradients (SMG) feature vector. LNHS features also slightly outperform SMG features when using a k-NN classifier on our data set, although a larger dataset would help statistical significance arguments. Primarily as a result of their compactness and speed, we propose LNHS as our advocated feature extraction approach from those evaluated. Although Naive Bayes improves on our k-NN systems, the improvement is marginal and relies on a good estimate of the priors. More extensive testing on a larger body of training and test data is required to further investigate these systems. Our future work will be to incorporate a passive automatic number plate localisation system and we will test our system on a significantly larger dataset. We will also include evaluations based on *verification* in addition to our current *identification* tests, as this may be more appropriate for systems that integrate MMR and ANPR.

References

[1] C.-N. Anagnostopoulos, I. Anagnostopoulos, I. Psoroulas, V. Loumos, and E. Kayafas. License plate recognition from still images and video sequences: A survey. *Intelligent Transportation Systems, IEEE Transactions on*, 9(3):377 – 391, 2008.

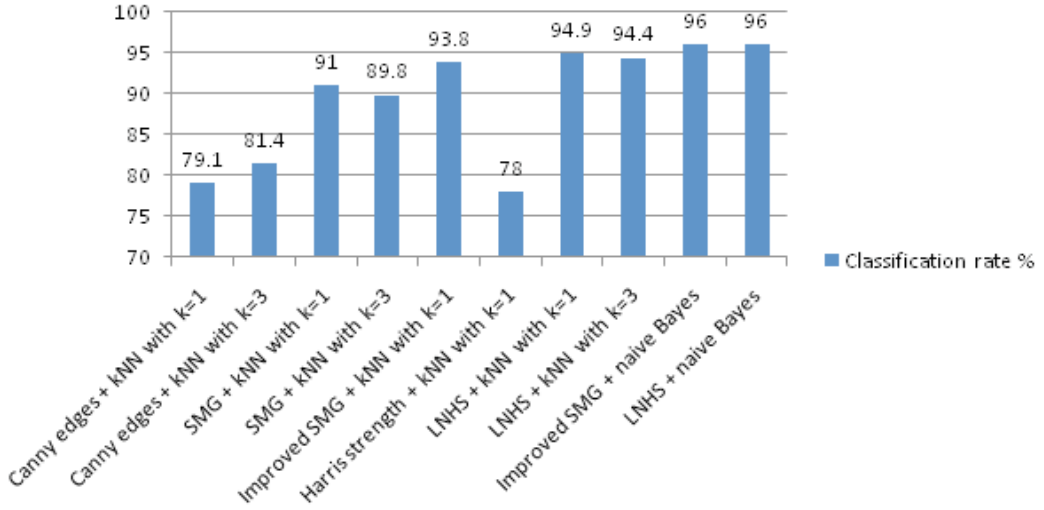

Figure 5. Classification rates for various MMR systems evaluated. SMG indicates 'square mapped gradient' features. LNHS indicates 'locally normalised Harris strength'. 'Improved' square mapped gradients uses a 9x9 Gaussian pre-filter mask.

LNHS + Naive Bayes			Improved SMG + Naive Bayes		
Vehicle class	Sample	Misclassified to	Vehicle class	Sample	Misclassified to
Peugeot207	226	Peugeot407	Peugeot207	134	Peugeot407
VWGolfMk3	121	VWGolfMk2	VWGolfMk3	121	VWGolfMk2
MiniCooper	202	FordMondeoMk1	MiniCooper	202	FordFiestaMk3
HondaJazzMk1	111	VauxhallAstraMk2	HondaJazzMk1	111	VauxhallVectraMk3
HondaJazzMk1	112	VauxhallAstraMk2	HondaJazzMk1	112	VauxhallVectraMk3
HondaJazzMk1	165	FordMondeoMk1	HondaJazzMk1	165	Peugeot607
FordFiestaMk4	138	VauxhallCorsaMk1	FordFiestaMk4	199	FordFiestaMk3

Table 1. Vehicles misclassified in our two best systems. Left: LNHS misclassifications. Right: SMG misclassifications

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[3] Z. Chen, N. E. Pears, M. Freeman, and J. Austin. Road vehicle classification using support vector machines. In *IEEE Int. Conf. Intelligent Computing and Intelligent Systems*, pages 214–218, 2009.

[4] X. Clady, P. Negri, M. Milgram, and R. Poulenard. Multi-class vehicle type recognition system. In *Artificial Neural Networks in Pattern Recognition*, volume 5064 of *Lecture Notes in Computer Science (LNCS)*, pages 228–239. Springer, 2008.

[5] C. J. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference Manchester*, pages 147–151, 1988.

[6] H. Huang, Q. Zhao, Y. Jia, and S. Tang. A 2dlda based algorithm for real time vehicle type recognition. In *11th IEEE Int. Conf. Intelligent Transportation Systems (ITSC 2008)*, pages 298 –303, 2008.

[7] X. Ma and W. E. L. Grimson. Edge-based rich representation for vehicle classification. *IEEE International Conference on Computer Vision*, 2:1185–1192, 2005.

[8] D. Munroe and M. G. Madden. Multi-class and single-class classification approaches to vehicle model recognition from images. In *Proc. Irish Conf. on Artificial Intelligence and Cognitive Science (AICS'05*, 2005.

[9] A. Noble. *Descriptions of Image Surfaces*. PhD Thesis, Department of Engineering Science, Oxford University, 1989.

[10] V. S. Petrovic and T. F. Cootes. Analysis of features for rigid structure vehicle type recognition. In *British Machine Vision Conference (BMVC'04)*, pages 587–596, 2004.

[11] A. Psyllos, C. N. Anagnostopoulos, and E. Kayafas. Vehicle authentication from digital image measurements. In *16th IMEKO TC4 Symposium*, 2008.

[12] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40:99–121, 2000.

[13] M. S. Sarfraz, A. Saeed, M. H. Khan, and Z. Riaz. Bayesian prior models for vehicle make and model recognition. In *Proc. 7th ACM Int. Conf. on Frontiers of Information Technology*, 2009.