Bayesian Prior Models for Vehicle Make and Model Recognition

[1]M.S. Sarfraz, A. Saeed, M.H. Khan, Z. Riaz*

[1]Computer Vision Research Group (COMVis)

Department of Electrical Engineering, COMSATS Institute of Information Technology, M.A.Jinnah Campus, Defense Road off Raiwind road, Lahore, Pakistan

{ssarfraz,ahmadsaeed, mhkhan}@ciitlahore.edu.pk

*Institut für Informatik,

Technische Universität München,

D-85748 Germany.

riaz@in.tum.de

ABSTRACT
Automatic vehicle type recognition (make and model) is very useful in secure access and traffic monitoring applications. It compliments the number plate recognition systems by providing a higher level of security against fraudulent use of number plates in traffic crimes. In this paper we present a simple but powerful probabilistic framework for vehicle type recognition that requires just a single representative car image in the database to recognize any incoming test image exhibiting strong appearance variations, as expected in outdoor image capture e.g. illumination, scale etc. We propose to use a new feature description, local energy based shape histogram 'LESH', in this problem that encodes the underlying shape and is invariant to illumination and other appearance variations such as scale, perspective distortions and color. Our method achieves high accuracy (above 94 %) as compared to the state of the art previous approaches on a standard benchmark car dataset. It provides a posterior over possible vehicle type matches which is especially attractive and very useful in practical traffic monitoring and/or surveillance video search (for a specific vehicle type) applications.

Keywords
Vehicle MMR, car type classification, feature extraction, LESH.

1. INTRODUCTION
The need for vehicle identification and classification has become relevant in recent years as a result of increased security awareness for access control systems in parking lots, buildings and restricted areas. Vehicle recognition also plays an important role in road traffic monitoring and management. It may also prevent the fraudulent use of bogus registration plates by providing an additional security mechanism to automatic number plate recognition systems. Vehicle classification have been limited mostly to algorithms distinguishing between broad categories of vehicles i.e. car, bus, truck etc. In contrast, an effective vehicle recognition system requires correctly identifying the make and model of vehicles within a given category. Recent research in this direction aims at identifying the correct make and model from a frontal or rear image of a vehicle. Methods dealing with vehicle make and model recognition 'MMR' typically extract relevant features from the vehicle image and either directly matches them in a template matching fashion or use a classifier to pose it as a multiclass learning problem. A relatively limited number of techniques that directly relate to vehicle MMR have been proposed in literature. Petrovic and Cootes [9] proposed techniques for the recognition of cars by extracting different gradient features from images. A number of feature extraction algorithms including direct and statistical mapping methods were applied to frontal views of cars. These feature vectors were then extracted and classified using simple nearest neighbor classification methods. Munroe and Madden [8] investigated the use of machine learning classification techniques in vehicle MMR. Initially a Canny edge detector followed by a dilation process was used to extract feature vectors. Subsequently different machine learning classifiers were used to determine vehicle make and model associated with each feature vector. Dlagnekov [2] explored the problem of MMR by using Scale Invariant Feature Transforms (SIFT) features [5]. Kazemi et.al in [3] investigated the use of Fast Fourier Transforms, Discrete Wavelet Transforms and Discrete Curvelet Transforms based image features in vehicle MMR. Rahati et.al in [10] proposed the direct replacement of Curvelet transforms in [3] with Contourlet transforms for vehicle MMR. Zafar et al.[13] extended the work of [10] by restricting the contourlet features in different subbands and reported superior performance. Clady et.al [1] proposed an oriented-contour point based voting algorithm for multiclass vehicle type recognition, which is particularly proven to be effective under occlusion.

The methods proposed in the literature so far may be categorized largely in two domains. After a feature extraction stage, The first one matches an incoming unknown vehicle image features directly with that of the registered vehicle images in the database by using some distance measure [2][3]. The test image is assigned the label for which the similarity is highest. The problem with this approach, however, is that it can not cater for the high in-class appearance variability present in the vehicle images. Its performance therefore depends on the effectiveness of the extracted features. The second one tries to learn a classifier directly in the multidimensional feature space. It therefore pose the problem as a multi-class classification problem [1][13]. Such an approach however, requires sufficiently large training examples for each class (vehicle type in this case). Furthermore,

FIT'09, December 16–18, 2009, CIIT, Abbottabad, Pakistan.

such a multi-class learning approach is meaningful only when the feature dimensionality is low as opposed to the current feature extractions (used in MMR) that results in very high dimensional features. Another problem with such an approach is the requirement of re-training the models whenever a new class (vehicle) is added in the database.

In this contribution, we attempt to solve the aforementioned problems by proposing to use a new feature extraction method that is based on local energy model of feature perception and is proven to be invariant with in-class variability due to illumination, scale and other appearance variations. The vehicle MMR problem, then, is formulated in a Bayesian framework. In particular instead of modeling the extracted features directly in a typical multi-class learning strategy, we rather model the computed similarities between the examples when the class (vehicle) is same and when it is different. Such a strategy is beneficial in that it turns a multi-class problem in to two-class one and does not depend on the dimensionality of the feature space. An offline training set is used to learn the prior 'same' and 'different' models from the distributions of the computed similarities. These prior models are then used to classify an incoming unknown test image by using its similarity with each of the representative reference vehicle image in the database (single per class) in a pure Bayesian setting. This provides us with a posterior over possible matches that is more meaningful in such applications since it gives the probabilities and the confidence for each class label. Note that because the models are learnt once, offline, for the same and different classes, adding a new vehicle image in the database does not require any retraining. In section 2 we explain the feature extraction stage. Section 3 provides details on the proposed method. In section 4 we demonstrate the effectiveness of our method by providing experimental results followed by discussion and conclusion in section 5.

2. Feature Extraction

Feature extraction is the first stage in any vehicle MMR application. Most of the existing MMR systems focus on minute portions of the image and use interest point based feature detectors such as SIFT. While such a bag-of-feature based approach has proven useful in general object recognition tasks, the vehicle MMR relies more on the global configuration of local parts, since it is the only discriminating feature between images of vehicle as an object category. We, therefore advocate the use of a global feature description (preserving the configural relationships of the local parts) summarizing the local content of the image parts in a scale, color and illumination invariant manner. For this we propose to use a new feature description, local energy based shape histogram "LESH', introduced in our earlier work in the context of head pose estimation [13] [14]. The rest of the section details LESH extraction slightly adapted for the purpose of vehicle MMR in this work.

2.1 Local Energy Model

The local energy model developed by [7] postulates that features are perceived at points in an image where the local frequency components are maximally in phase.

$$E(x) = max_{\bar{\phi}(x)\in[0\ ,2\pi]} \frac{\sum_n A_n cos(\phi_n(x) - \bar{\phi}(x))}{\sum_n A_n} \qquad (1)$$

where A_{n} and ϕ_n are the magnitude and phase of the nth Fourier component. This frequency information must be obtained in a way such that underlying phase information is preserved. This is achieved by convolving the image with a bank of Gabor wavelets kernels tuned to 5 spatial frequencies and 8 orientations. At each image location, for each scale and orientation, it produces a complex value comprising the output of even symmetric and odd symmetric filter, which gives the associated magnitude and phase of that pixel, $G(e_{n,v}, o_{n,v}) = I(x,y) * \psi_{n,v}(z)$, where $\psi_{n,v}$ is the bank of Gabor kernel and n,v is the scale and orientation, $G(\cdot)$ is the response at image position (x,y) having a real and imaginary part comprising output of even symmetric and odd symmetric filter at scale n and orientation v. The amplitude A_n and phase ϕ_n in equation 1, thus can be written in terms of these responses at a given scale n as,

$$A_n = \sqrt{e_n^2 + o_n^2}\,, \qquad \phi_n = \tan^{-1}\frac{e_n}{o_n} \qquad (2)$$

Originally [11] has proposed to use cosine of the deviation of each phase component from the mean phase as a measure of the symmetry of phase, however, this measure results in poor localization and is sensitive to noise. [4] extended this framework and developed a modified measure, as given in equation 3, consisting of sine of the phase deviation, including a proper weighing of the frequency spread 'W' and also a noise cancelation factor 'T'. The normalization by summation of all component amplitudes makes it independent of the overall magnitude of the signal, making it invariant to illumination variations in images. For details of this measure see [4].

2.2 Local Energy based Shape Histogram-LESH

The local energy analysis in the preceding section is intended to detect interest points in images with a high reliability in presence of illumination and noise. We use this raw energy information and attempt to encode the underlying shape. This is done in a way that makes it invariant to scale variations. Motivated by the fact that this local orientation energy response varies with respect to the underlying shape and since local energy signifies the underlying corners, edges or contours, we generate a local histogram accumulating the local energy along each filter orientation on different sub-regions of the image. The local histograms are

$$E = \frac{\sum_n W(x) \lfloor A_n(x)(cos(\phi_n(x) - \bar{\phi}(x)) - |sin(\phi_n(x) - \bar{\phi}(x))|) - T \rfloor}{\sum_n A_n(x) + \varepsilon} \qquad (3)$$

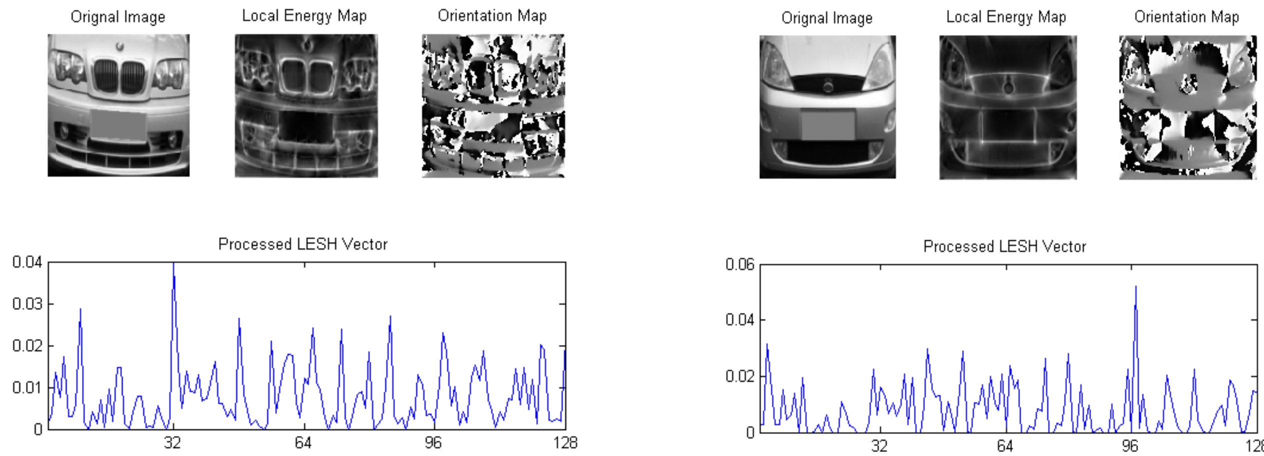

Figure 1. Example LESH feature extraction on two of the vehicle images from our database

extracted from different sub-regions of the image, and then concatenated together, to keep the spatial relationship between vehicle parts. We proceed by obtaining an orientation label map where each pixel is assigned the label of the orientation at which it has largest energy across all scales. The local histogram *h* is extracted according to the following,

$$h_{r,b} = \sum w_r \times E \times \delta_{Lb} \tag{4}$$

where subscript *b* represents the current bin, *L* is the orientation label map, *E* is the local energy, as computed in equation 3, δ_{Lb} is the Kronecker delta and *w* is a Gaussian weighing function centered at region *r*.

$$w_r = \frac{1}{\sqrt{2\pi\sigma}} e^{[(x-r_{xo})^2+(y-r_{yo})^2]/\sigma^2} \tag{5}$$

This weight is used to provide soft margins across bins by small weighted overlap among neighboring sub-regions to overcome the problems induced due to scale variations.

As mentioned earlier, in order to keep the spatial relation between vehicle parts, we extract 8 bins local histogram corresponding to 8 filter orientations on 16 image partitions, which makes it a 128-dimensional feature vector. Example feature extraction and associated energy and orientation maps on two different vehicle images, from our database, are shown in figure 1.

3. Proposed Recognition Framework

As mentioned earlier, instead of modeling the extracted features directly in a multi-class learning manner, we rather formulate the whole problem in a similarity feature space. This is achieved by computing similarities between images of the vehicle, when the class (make and model) is same and when it is different, from a offline training set. This effectively turns the multiclass problem into a two class problem. Motivated by the success of similar approach [6][12] ,our approach briefly is to learn probabilistic models describing the approximated joint probability of registered (database) and test vehicle image. Since we assume that only one representative vehicle image in each class is available in our database, we learn such models by explicitly modeling the vehicle appearance change when the class is similar and when it is dissimilar. This is done by using the distributions of the similarities from the offline generic training set. These distributions are used to obtain the likelihood functions of the form

$$P(I_r, I_t \mid C) \quad where \quad C \in \{same, diff\} \tag{6}$$

C *refers* to classes when the registered database image I_r and test image I_t are similar (*same*) and dissimilar (*diff*) in terms of vehicle identity (make and model).

3.1 Prior Appearance Models

We approximate the joint probability of test and database image as,

$$P(I_r, I_t \mid C) \approx P(\chi_{rt} \mid C, \Omega) \tag{7}$$

where χ_{rt} is the similarity between test and registered images and $\Omega = (\theta_{same}, \theta_{diff})$ are the model parameters for the similar and dissimilar class. The distributions of the extracted similarities from the offline training set are used to learn these model parameters. We employ a Gaussian model to approximate these similarities distributions.

Any distance metric may be used to compute the similarity. We however use the cosine metric,

$$\chi_{rt} = \frac{I_r \cdot I_t}{\|I_r\|_2 \cdot \|I_t\|_2} \tag{8}$$

where $\|\cdot\|$ denotes the Euclidean norm of the vectors.

The likelihoods in equations 7 are then obtained from the similar (*same)* and dissimilar (*diff)* distributions of similarities. Figure 2a depicts the histogram of these prior *same* and *diff* distributions obtained from the offline training set. These distributions are approximated by using a Gaussian density. Figure 2b shows the

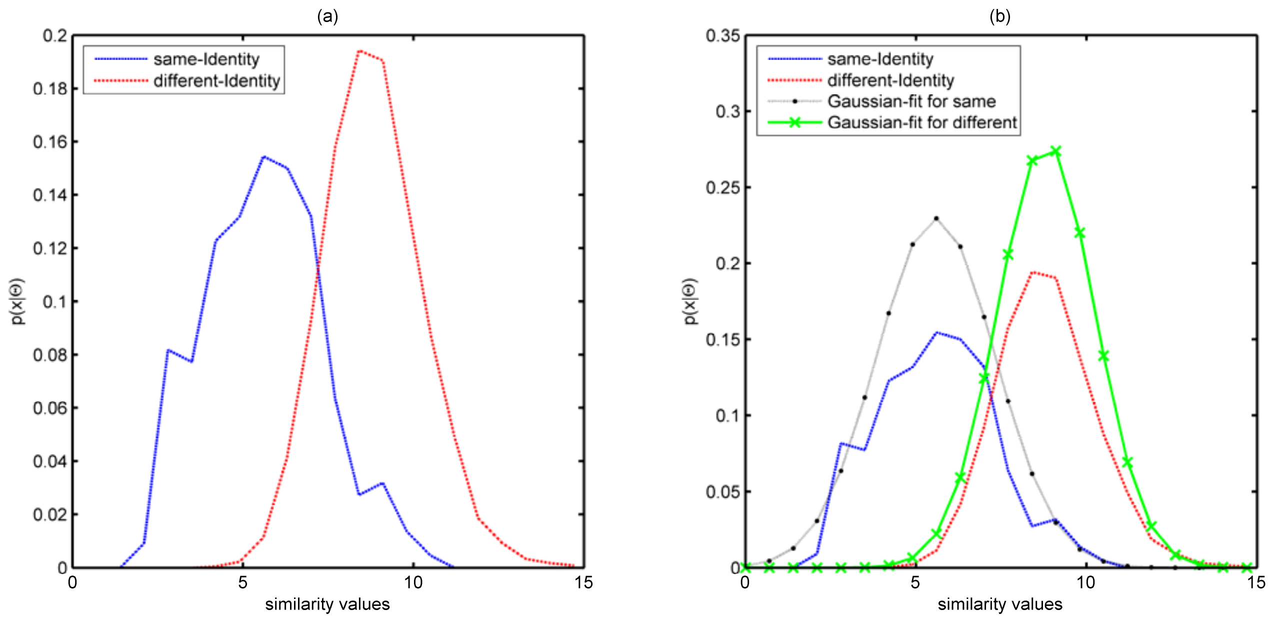

Figure 2. Similarity distributions for similar and dissimilar classes. x-axis denotes the similarity χ and y-axis the density approximation. (a) Histograms for prior same and different distributions. (b) Gaussian density fits to obtain prior models.

fitted Gaussian model on these distributions. The Gaussian model parameters $\Omega = (\theta_{same}, \theta_{diff})$ are estimated directly from these distributions. The parameters are the means and variance of the estimated density i.e. $\theta_{same} = (\mu_{same}, \sigma_{same})$ and $\theta_{diff} = (\mu_{diff}, \sigma_{diff})$.

Note that the more separated the two distributions are the more power it has to discriminate between the same and different classes. It is clear that the similarities distributions obtained, as depicted, provides a good basis to derive the prior models.

3.2 Vehicle Type Recognition

Prior models obtained in the previous section can now directly be used in computing the posterior probability of two images to be similar or dissimilar. By using this posterior as a match score we can now decide for a test image I_t of unknown make and model, if it is coming from the same vehicle type as database I_r, with each of the registered database images. Using Bayes rule the posterior probability can be written as in equation 9. The conditional probabilities $P(\chi_{rt} \mid same, \theta_{same})$ and $P(\chi_{rt} \mid diff, \theta_{diff})$ are found from the offline training set as detailed in the previous section. $P(same)$ *and* $P(diff)$ are the class priors. The class priors are set to $P(same) \ll 1$ and $P(diff) = 1 - P(same)$ in all of our experiments.

We compute this posterior for an unknown test vehicle image with all of the registered database images and choose the make and model of the database image with the highest score as recognition result.

4. Experimental Setup and Results

Experiments are performed on a database of cars [9][13] comprising 300 frontal view images of 25 different car types such as Honda civic, Toyota corolla, Audi, Fiat etc. Each car type (class) has eleven to fourteen different images. Several of the car types exhibit some hard classes such as same make but different model e.g. Vauxhall Astra and Vauxhall Vectra, Fiat Punto and Fiat Punto new, where the only difference lies in the shape of the front grill. Frontal area of the cars are cropped from the images using the method described in [9], the area is segmented based on the number plate coordinates such that it includes the important components of the frontal views of cars such as lights, grill, logo, bumper area, etc. Images are then resized to 128x128 pixels. Figure 3 depicts some of the cropped frontal views of the cars from the database.

We use four images from each class (a total of 100 images) as offline training set for training of the prior models as described in section 3.1. One image from each class (a total of 25 images corresponding to 25 classes) is taken as registered database (for recognition). Whereas remaining are used as test images. Note that all the three sets are disjoint.

After extracting features, as described in section 2, we proceed to compute the posterior of a test image (equation 9) by calculating

$$P(same \mid \chi_{rt}, \Omega) = \frac{P(\chi_{rt} \mid same, \theta_{same}) P(same)}{P(\chi_{rt} \mid same, \theta_{same}) P(same) + P(\chi_{rt} \mid diff, \theta_{diff}) P(diff)} \qquad (9)$$

Figure 3. Example segmented images of three different car types from the database

its similarity (using equation 8) with each of the registered database car image feature vectors. The database car type, for which this posterior is maximum, is assigned to the test image. Our average recognition accuracy on all of the test images is 94% (rank-1). Figure 4 plots the performance across each class (vehicle type) as a confusion matrix.

The confusion matrix depicts performance for all the test images in each class (vehicle type). A white in the diagonal indicates perfect classification i.e. all the test image in that class are correctly classified. As can be seen the recognition results are 100% for most of the classes, while only three to four classes shows some degradation in performance e.g. some of the test images in class 15 (Fiat Punto) are assigned to class 20 (Fiat Punto new) ,see figure 4. The confusion in these classes is well expected since only subtle differences exist as being the same make. A closer look at the corresponding posterior probabilities

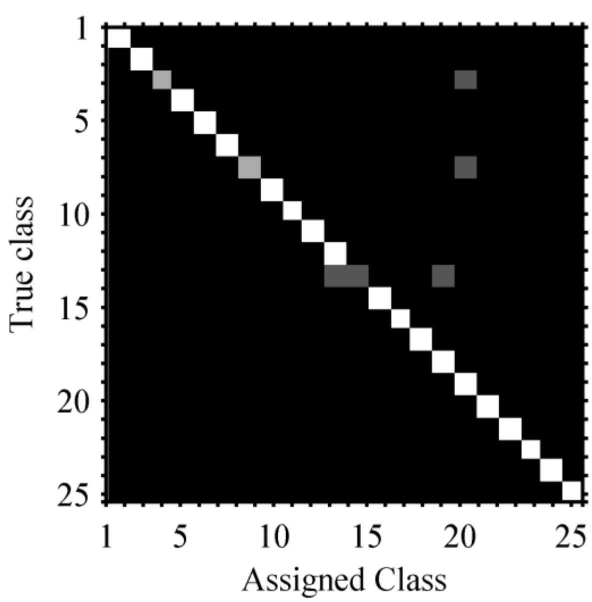

Figure 4. Confusion matrix, white diagonal indicates perfect recognition

reveals that the difference in computed probabilities is very small, in most cases less that 0.05 (5%). Because of the fact that we use

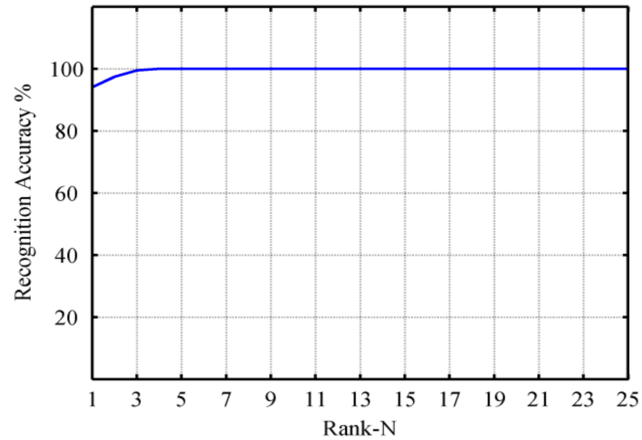

Figure 5. Rank wise score depicting accuracy for top N matches

the maximum probability to assign the class label, differences even this small may result in the wrong classification. We, therefore, note that in a probabilistic framework it is more meaningful to look at the notion of top probabilities or top matches. Figure 5 plots the rank curve for this purpose. As can be seen even considering the first 2 matches (rank-2), according to the assigned probabilities, the overall recognition accuracy improves from 94% to 97.4% and up to 100% for first 3 matches (rank-3).

5. Discussion and Conclusion

Our results indicate the effectiveness of the proposed method on a standard vehicle dataset. We can compare our method with some of the contemporary approaches in vehicle MMR on same or similar databases. Petrovic and Cootes [9] reports a 87.3% accuracy on the similar access control car database by using different gradient features from images. Dlagnekov [2] using SIFT features reported an accuracy of 89%. Kazemi et.al in [3] investigated the use of Fast Fourier Transforms, Discrete Wavelet Transforms and Discrete Curvelet Transforms based image features in vehicle MMR, using a relatively small car database comprising only five different vehicle types, they reported a accuracy of above 92%. Rahati et.al in [10] proposed Contourlet transforms for vehicle MMR, the performance of their method on the same database, as reported in [13], is 52%. Zafar et al.[13] extended the work of Rahati et.al by restricting the contourlet features in different subbands and reported a peak performance of 94%, on the same database. The accuracy figure of 94%, however, is achieved when using ten images in each class for training and only one to two images per class for testing, as demonstrated by their results the accuracy drops to 85% when using four images per class for training (similar to our offline training set size). Clady et.al in [1], using oriented-contour point based voting algorithm reports a maximum performance of 93.1% by using fusion of different classifiers.

The preceding comparison shows that our method is able to achieve better results by using a relatively simple but effective recognition strategy. Most of the previous approaches use gradient based features which in general are very much susceptible to outdoor realistic image conditions, such as illumination, casting shadows, surface reflections etc. We on the other hand proposed to use a feature description that is based on the multi-resolution and multi-scale analysis of the underlying frequency content. The proposed feature description attempts to encode the pure shape information such as contours, corners etc and is invariant to such appearance variations. Our method requires just one reference image of the respective vehicles in the database as opposed to the requirement of multiple example images per vehicle type in most of the existing approaches.

On concluding remarks, we have presented a novel approach to automatic vehicle make and model recognition. We have proposed to employ a new feature description LESH in vehicle MMR. The features are then modeled in a similarity feature space by using a probabilistic Bayesian framework. Our method provides the posterior over possible matches. The assigned probabilities for possible vehicle type are more useful in a typical vehicle MMR application, such as traffic management system, access control or security video archive search for a specific vehicle type.

REFERENCES

[1] Clady, X., Negri, P., Milgram, M., and Poulenard, R. Multi-class Vehicle Type Recognition System. L. Prevost, .Marinai, and F. Schwenker, Eds. Lecture Notes In Artificial Intelligence, vol. 5064. Springer-Verlag. 2008.

[2] Dlagnekov, L., "Video-based car surveillance: License plate make and model recognition" Masters thesis, University of California at San Diego. 2005.

[3] Kazemi, F., Samadi, S., Pourreza, H., Akbarzadeh, M. Vehicle Recognition Based on Fourier, Wavelet and Curvelet Transforms - a Comparative Study, Int. Journal Of Computer Science and Network Security 7(2), 2007.

[4] P.D. Kovesi, Phase congruency: A low-level image invariant, Psychological Research, 64, pp 136-148, 2000.

[5] Lowe, D.G., Distinctive Image Features from Scale Invariant Keypoints, International Journal of Computer Vision, 2004.

[6] S. Lucey and T. Chen, Learning Patch Dependencies for Improved Pose Mismatched Face Verication. In Proceeding IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 17-22, 2006.

[7] M.C. Morrone and R.A. Owens. Feature detection from local energy, Pattern Recognition Letters vol.6, pp 303-313. 1987.

[8] Munroe, D.T., Madden, M.G., Multi –class and single-class classification approaches to vehicle Model recognition from images, In Proceedings of IEEE AICS, 2005.

[9] Petrović, V., Cootes, T. Analysis of features for rigid structure vehicle type recognition, In Proc. of the British Machine Vision Conference. 2004.

[10] Rahati, S., Moravejian, R., Kazemi, E.M., Kazemi, F.M., Vehicle recognition using contourlet transform and SVM", Proc. of the 5th IEEE Int. Conference on Information technology: New Generations, 894-898, 2008.

[11] B. Robbins and R. Owens, 2D feature detection via local energy, Image and Vision Computing 15, pp 353-368, 1997.

[12] M.S.Sarfraz and O. Hellwich. "Statistical Appearance Models for Automatic Pose Invariant Face Recognition." 8th IEEE Int. conference on Face and Gesture recognition, FG pp 1-6, 2008.

[13] M.Saquib Sarfraz and Olaf Hellwich, "On Head Pose Estimation in Face Recognition", J. Braz, A. Ranchordas, H. Araújo and J. Jorge (Eds), Advances in Computer Graphics and Computer Vision. Lecture notes CCIS 24, pp 162 -175 Springer, 2009.

[14] M.S.Sarfraz, O.Hellwich," An efficient front-end facial pose estimation system for face recognition", Pattern Recognition and Image Analysis, distributed by Springer, Vol.18 (3) pp. 434–441, 2008.

[15] Zafar,I., Edirisinghe,E.A., Acar,B.S. Localized contourlet features in vehicle make and model recognition, In Proceedings SPIE, Electronic Imaging, Vol. 7251, 725105. 2009.

,