2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

PERFORMANCE EVALUATION OF LOCAL IMAGE FEATURES FOR MULTINATIONAL VEHICLE LICENSE PLATE VERIFICATION

Muhammad Rizwan Asif, Qi Chun, Irfana Bibi, Muhammad Sadiq Fareed, Zhang Zhe, Zhang Zhaoqiang

Abstract— **The verification of vehicle License Plates (LPs) has not been given much importance in existing LP recognition systems as only a handful of methods deal with this problem explicitly. For an efficient system, it is imperative that a detected LP is validated first before the recognition of characters on it. Majority of the existing methods make use of geometrical constraints for the elimination of false LP regions which is not an effective way as multinational LPs have variable geometrical attributes and diversity in styles. To overcome these limitations, in this paper, we evaluate three kinds of representative local descriptors (SURF, HOG and LBP) and their combinations along with AlexNet CNN for the classification of LP and non-LP regions to provide a unique solution for the validation of multinational LPs. Experiments on 13490 LP and non-LP images show that the HOG feature individually gives the best recognition rate of 96.94% while considering collectively, best of 98.35% is achieved for SURF+HOG; whereas, the fine-tuned AlexNet outperform all others in terms of recognition accuracy of 99.27% but requires extensive processing. Furthermore, the proposed model is incorporated in one of the existing LP detection methods to demonstrate improved performance.**

I. INTRODUCTION

Automatic License Plate Recognition (ALPR) is a challenging topic in the field of Intelligent Transportation Systems (ITS) which is a mass surveillance technique used to recognize LPs for vehicle identification as a part of intelligent vehicle infrastructure. It is a technology that allows computer systems to read the LP registration numbers automatically from digital images by transforming digital image pixels in ASCII text of number plate. These systems assume that the identity of each vehicle is already displayed by means of registration plate. Therefore, no additional hardware such as responder or transmitter is required to be installed.

This topic has been the center of attention among researchers due to its vast potential applications. It can be used for the identification of stolen, unregistered and uninsured vehicles. It can also be extremely effective for security control in restricted areas and for border control systems. A parking lot equipped with ALPR system can offer numerous benefits including ticketless parking fee management, automated payment process, car theft prevention and automated parking. Moreover, it can be installed in intelligent vehicles to recognize the identity of neighboring vehicles for communication purposes.

ALPR system is a combination of four processes including image acquisition, LP detection, character extraction and its recognition. True detection of LPs and then its recognition is imperative for a system to work in real-time. During LP detection, several regions could be identified as LPs that should be analyzed using some verification method for true LP region detection before sending all patches with characters to the recognition step which would result in improved performance.

Most of the research conducted does not deal with this problem effectively as one or more geometrical constraints such as area, height, width and aspect ratio of LPs are usually applied along with some other limitations in literature [1-5] to eliminate false LP regions due to its simplicity and effectiveness for known datasets. But in real time environment, the position of the vehicle with respect to the camera might be unknown and the vehicles may move toward or away from the camera which could result in the change of geometrical attributes significantly and specific constraints would fail to provide desired results. Additionally, these geometrical features differ for multinational LPs. As can be seen in Fig. 1, the multinational LPs have different color, size and content. Likewise, the numbers of characters in multinational LPs vary from each other that result in further complication for the characterization of LP and non-LP regions.

Figure 1. Examples of some multinational license plates

Local image features, such as Speed-Up Robust Features (SURF), Histograms of Oriented Gradients (HOG), Local Binary Patterns (LBP), Scale Invariant Feature Transform (SIFT) and Gabor features are employed to achieve object classification, face recognition, image matching, etc. and the performance of these features vary for different conditions and applications. Although there are many feature descriptors, their complementarities have not been fully investigated that results in better performance if a good

*Research supported by the National Natural Science Foundation of China (Grant No. 61572395 and 61133008).

†M. Rizwan Asif is with Xi'an Jiaotong University, Xi'an, 710049, P. R. China. He is also with COMSATS Institute of Information Technology, Lahore, 54000, Pakistan (e-mail: rizwansheikh123@hotmail.com).

†Qi Chun is the corresponding author and is with Xi'an Jiaotong University, Xi'an, 710049, P. R. China (e-mail: qichun@mail.xjtu.edu.cn).

†Irfana Bibi is with the Xidian University, Xi'an, 710071, P. R. China.

M. Sadiq Fareed, Zhang Zhe and Zhang Zhaoqiang are with Xi'an Jiaotong University, Xi'an, 710049, P. R. China.

†The authors contributed equally to this work.

combination of features is chosen. Therefore, in this paper, we evaluate three local image features (SURF, HOG and LBP) and their combinations for the categorization of LP and non-LP regions. However, the concept of using these features is somewhat old fashioned now by the introduction of deep learning methods. Deep architectures like AlexNet, GoogleNet, and Residual Networks learn a hierarchy of discriminate features automatically that richly describe image content. Hence, we also investigate the performance of AlexNet CNN by modifying it for our task.

The selected (SURF, HOG and LBP) feature methods are chosen as they describe edginess, orientation and/or texture information around pixels that are the main characteristics of LPs while the AlexNet is chosen due to its popularity for many application-specific methods. The extracted features are evaluated one-by-one with linear SVM using cross-validation on the available dataset to provide a unique solution that is best suited for vehicle LP classification. Such models can easily be incorporated in existing ALPR methods to increase the overall efficiency of the system in real-time.

The rest of the paper is organized in the following order. Section II discusses the related work and Section III explains the used local image feature models in detail. The results are discussed in Section IV and Section V provide the conclusions remarks.

II. Related Work

Only a handful of methods in literature [5,6] explicitly deals with the problem of LP verification. In [5], false regions are removed using height and width ratio initially and then orientation algorithm [7] is applied to decide if the candidate region is a valid LP as the characters in LPs have similar orientation. Connected Component Analysis (CCA) is used in [6] to extract the character information and four weak verification approaches were employed to create a strong classifier for the verification of LPs.

Some other methods discuss this problem implicitly as a part of LP detection process. As LPs usually have abundance of edges, the average transition at three different positions in the candidate regions is used in [1] for the validation of LP regions. Moreover, a uniformity test for edge density is conducted by partitioning the candidate region in equal subparts as the distribution of edges in the LP region is usually uniform. In [2], orientation angle and geometrical information is used to filter out non-LP regions. Geometrical attributes and a scan-line approach is used in [3] that counts the number of black or white runs in the middle of the candidate region for the exclusion of false LPs. A two layered approach is presented in [4] to discard the false LP regions in which energy map of binary candidate image and the vertical edge information extracted using scan-line approach is used. A dedicated method for the verification of Chinese LPs is presented in [8] that make use of stroke-width relation of Chinese characters, number of connected components and aspect ratio for the determination real or fake LP region. For Iranian plates, projection method is used in [9] to detect the peaks in the candidate region. If the region under consideration detects at least eight peaks using semi-local maximum and minimum, it is considered to be a true plate as standard Iranian LP includes seven numbers and a letter.

Most of the existing schemes employed for the verification of LPs make use of the geometrical attributes or CCA information which gives effective results for known datasets but may not provide reasonable results if the dataset is unknown. In real-time, it is not an effective way to hard-code the geometrical information as the geometrical information such as height, width and other parameters might be unknown and may change with the vehicle depth in the image. Furthermore, fixed parameter for the number of connected components needs to be chosen for the verification of plates if CCA is to be used. As seen in Fig. 1, multinational vehicles may have different numbers of characters ranging from three to nine characters. Some LPs may exceed this limit. Therefore, use of this information might not give effective results.

To get rid of parameter selection and to form a method applicable for the verification of multinational LPs, local image features are used to train the SVM classifier to categorize LP and non-LP regions. In this paper, SURF, HOG and LBP features are assessed individually and collectively for performance evaluation. Moreover, the AlexNet CNN is also investigated for LP verification task. It is validated that a good combination of local image features results in superior performance as compared to individual features. Furthermore, the proposed model has been merged in one of the existing LP detection methods to verify improved performance.

III. Local Image Features

In this section, we explain the selected local features in detail. The local image feature descriptors are extracted for the classification of LP and non-LP regions. The classification performance of SURF using BOW, HOG and LBP features have been evaluated separately and collectively for the verification of LP regions. Moreover, the performance of AlexNet CNN is also investigated.

A. SURF descriptor using Bag of Words (BOW)

SURF [10] is a feature detector that determines the scale, position, orientation, sign of Laplacian and descriptors using second order Hessian matrix. The scale, position, orientation and sign of Laplacian are called as interest points. For each such point, a feature vector is formed by the descriptor. Several computation steps are required to extract the interest points and descriptors.

$$H\left(f\left(x,y\right)\right)=\begin{bmatrix}\dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y}\\ \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2}\end{bmatrix} \quad (1)$$

For improved computational efficiency, Hessian matrix as in (1) is computed on integral image for the matrix *f(x,y)*. The number of interest points required for the representation of image is determined by the determinant of this Hessian detector. Haar wavelet transform is used on integral image for the computation of orientation on 4x4 sub-regions around

the position of interest point. Each sub-region has the following descriptors.

$$\left\{\sum dx, \sum |dx|, \sum dy, \sum |dy|\right\}$$

Therefore, in total, the number of descriptor becomes 4×4×4=64 descriptors. The locations of interest points using SURF for a sample LP and non-LP is given in Fig. 2.

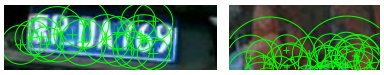


Figure 2. SURF interest points for LP and Non-LP region

Usage of raw SURF descriptors at the input of SVM might not be adequate enough to effectively identify the decision boundaries. Therefore, SURF descriptors are represented using BOW model. The BOW model is typically used for the classification of documents in which the frequency of occurrence of each word is used as a feature. To represent SURF using BOW model, the descriptors are considered as words or codes and are converted into a histogram. A set of codes and their sparse histograms is the output of BOW for each feature descriptor to group LPs uniquely.

The descriptors of LP and non-LP images meant for training are extracted using SURF and provided at the input of K-means clustering to find the centroid of clusters to group the data into K categories with a centroid for each group. For LP validation, K=500 gives the optimal classification performance. Minimum Euclidean Distance Measure (MEDM) is then used to determine the closest centroid of each training image for the descriptors and a code is assigned to it. The count of all the codes for each image creates a histogram which is used as feature vector and assigned a unique label. The visual word occurrence for a sample LP is shown in Fig. 3.

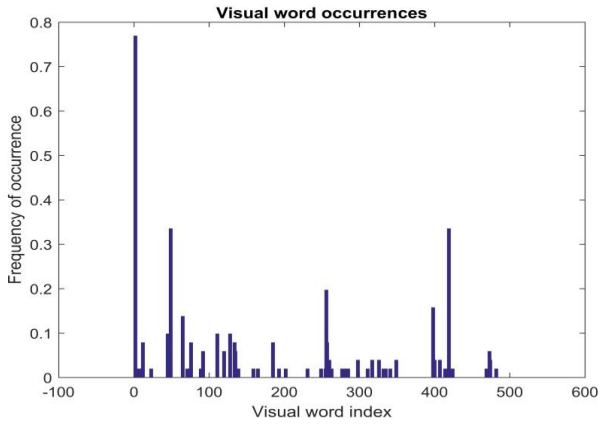


Figure 3. Visual word occurrences for a test LP image

The centroids and their associated codes extracted during the training are used as SVM decision boundaries. The descriptor of each test image is extracted and a code is assigned to it after calculating MEDM. Then the histogram of the codes is formed and is provided at the input of SVM for classification by predicting the group by using decision boundaries of SVM projected in training.

B. HOG Feature Descriptor

HOG is a feature descriptor in which the frequency of occurrences of gradient orientation is computed on a dense grid of uniformly spaced cells in localized portions of image. The main theme behind HOG descriptor is that the edge directions or intensity gradients' distribution can describe the shape and appearance of local object within an image. This distribution is evaluated by dividing the image in small connected regions termed as cells. Then, a histogram of gradient directions is compiled within each cell. By concatenating these histograms, a descriptor is formed.

The gradient values are computed by applying one dimensional derivative mask in vertical and horizontal directions. The filtering of intensity is required with the following filter kernels:

$$[-1,0,1] \text{ and } [-1,0,1]^T$$

Once the gradient is computed, histogram for each cell is created. A weighted vote for pixels within each cell based on the gradient computation values is used to create a histogram channel based on orientation. The histogram channels are uniformly spread over 0° to 360° or 0° to 180°. Gradient magnitude or a function of gradient magnitude can be used in pixel contribution for weighted vote. Usually, the gradient magnitude itself is used which results in optimal performance. The direction of the gradient for each pixel is calculated by using (2).

$$\theta(x,y) = \arctan\left(\frac{I(x,y+1) - I(x,y-1)}{I(x+1,y) - I(x-1,y)}\right) \qquad (2)$$

where I(x,y) is the grayscale value at (x,y).

We consider using the approach presented in [11] to model the appearance and shape of LP and non-LP regions. Histogram for local gradients is formed according to the orientation and gradient magnitude within a grid of cells. A feature vector is acquired within each overlapping block of cells by sampling the histograms from contributing cells. These feature vectors are then concatenated to result in final feature vector that is used at the input of classifier. Therefore, it is essential that the HOG feature vector encodes reasonable amount of information about the object for good classification accuracy.

(a) Sample license plate

(a) For 4x4 cell Features = 16236 (b) For 8x8 cell Features = 3600 (c) For 16x16 cell Features = 648

Figure 4. Visualization of shape information for different cell sizes

The visualization of HOG features shown in Fig. 4 depicts that a cell size of 16×16 lacks insufficient shape information whereas a 4×4 cell size encodes excess of shape information, consequently results in a significant increase in dimensionality of HOG feature vector. A good compromise is a cell size of 8×8 that encodes reasonable amount of shape

and appearance information to identify a LP shape visually while limiting the number of HOG feature vector dimension which helps in speeding up the training process.

To account for resolution variations and for facilitating the process, images are scaled to 50×170 pixels and blocks are formed by grouping 2×2 neighboring cells while each cell size is chosen to be 8×8. The 0° to 180° degree angles are divided into 9 bins and $\theta(x,y)$ is projected onto the closest one of nine bins.

C. LBP Feature Descriptor

Finally, we will test the use of LBP [12] to assess if the texture of LPs contains appropriate information for the classification task of LP and non-LP regions. It is an algorithm that represents local texture features in images. LBP was initially introduced in [13] and is discussed in detail in [14]. The original LBP operator considers a 3×3 kernel and the difference between the center pixel and eight neighbors is performed to assign a binary number on the basis of the sign of difference for each pixel. If the difference is greater than zero, binary number 1 is assigned otherwise 0 is assigned.

Different neighborhood sizes were considered in [15] by using a circular neighborhood and the values were interpolated at non-integer pixel coordinates bi-linearly. A generic LBP operator is shown in (3) that uses P points on a circle of an arbitrary radius R to result a LBP code for a pixel (x_c,y_c).

$$LBP_{P,R}(x_c,y_c)=\sum_{p=0}^{P-1}s\left(f(x,y)-f(x_p,y_p)\right)2^p \quad (3)$$

$$s(z)=\begin{cases}1, & z\geq 0\\ 0, & z<0\end{cases} \quad (4)$$

where s(z) in (4) is a thresholding function.

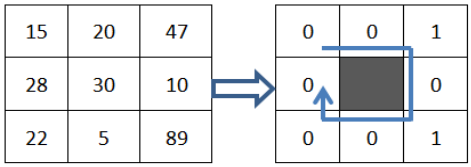


Figure 5. The principle of LBP

In our experiment, LBP code is computed considering 8 neighbors with radius R = 1. An example for the calculation of an LBP code is shown in Fig. 5. The LBP code for each cell is computed and a histogram is formed. The histograms of all the cells are combined together to result in a single feature vector that is used for the training of classifier. For the processing of LBP, the images are also resized to 50×170 pixels.

Fig. 6(a-d) shows images with different textures in which the first two are the LPs and other two are non-LP regions. The similarity between LBP features is estimated by computing the squared error between them. Fig. 6(e) shows the squared error to compare LP and non-LP regions. The error is small when images have similar texture.

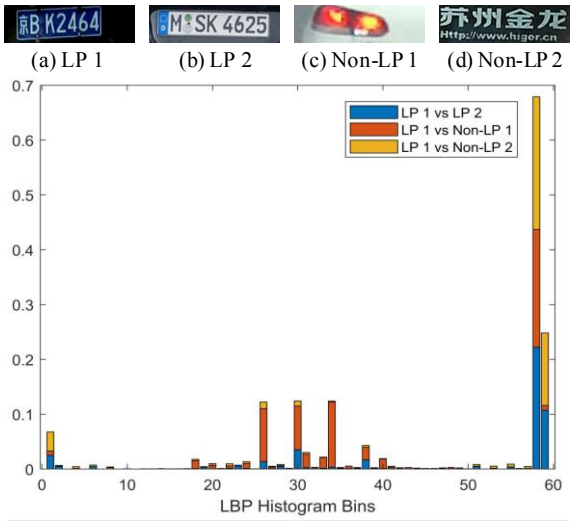


Figure 6. Comparison of difference of texture information using squared error of LBP histogram

D. AlexNet Convolutional Neural Network (CNN)

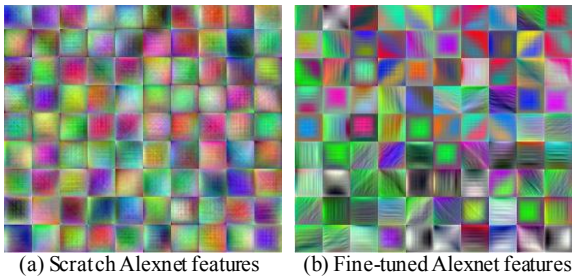
CNNs have become immensely popular in recent years mainly due to the success of the AlexNet [16] on the ILSVRC-2012 image classification benchmark. It consists of five convolution layers, three pooling layers and three fully connected layers having nearly 60 million trainable parameters. Many application-specific methods including [17-20] having their unique datasets make use of transfer learning on AlexNet CNN to good effect. Furthermore, a survey on several state-of-the-art CNNs indicate that AlexNet is flexible to different datasets while having least number of layers [21] that reduces the computational complexity. For these reasons, we investigate the scratch and fine-tuned AlexNet model on our LP dataset. The stochastic gradient descent update rule with momentum is applied in our experiment which is similar to that of [16].

1) Scratch AlexNet

Initially, we train the AlexNet architecture from scratch on our training data to obtain a model that can be used for evaluation on our test set. To fit the input images according to the AlexNet architecture, they are resized to 227×227 pixels. Other parameters are chosen as below: momentum 0.9, weight decay 1×10−4, initial learn rate 1×10−4 and mini batch size of 128 with a maximum number of iterations of 25000 (25 epochs). Some features learned by the scratch AlexNet by the first convolution layer are shown in Fig. 7a.

2) Fine-tuned AlexNet

This version of the architecture relies on the weights that are initialized by a pre-trained network. The pre-trained network is initially trained on a subset of ILSVRC ImageNet dataset. As the ImageNet dataset does not contain the image categories present in our dataset, we have fine-tuned the network according to our data. The same experimental settings are used as discussed earlier. It can be seen from the visualization of Fig. 7b that the features learnt by the fine-tuned network are more enriched and enhanced as compared with the scratch network which is shown in Fig. 7a.



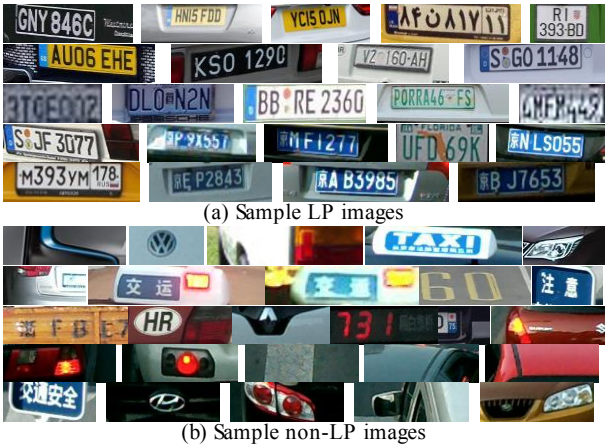
(a) Scratch Alexnet features (b) Fine-tuned Alexnet features

Figure 7. Feature visualization of first convolution layer of retrained AlexNet

IV. RESULTS

We have used MATLAB 2017a on Intel® Xeon® CPU at 2.40GHz for the implementation of proposed models. As we have used SVM for classification which is a supervised learning model that requires training to get a decision boundary, a database of LP and non-LP regions is mandatory for training and validation. The database selected comprises of 6561 LP and 6929 non-LP regions. 30% of images from each category are used for training.

The database consist of LPs of 39 countries including Australia, Canada, China, Hong Kong, USA, Mexico, India, Pakistan, Sri Lanka, Malaysia, UAE, Germany, Russia, Serbia, Denmark, Switzerland, Austria, etc., whereas the non-LP regions are extracted from real-time traffic images that mainly consist of traffic sign boards, vehicle logos, headlights and backlights of vehicles. The chosen dataset focuses on diversity in LP styles to evaluate the performance of feature descriptors under different conditions. Some of the LP and non-LP regions are shown in Fig. 8.



(a) Sample LP images

(b) Sample non-LP images

Figure 8. Some images from the dataset

We test three feature extraction methods and their combinations for the remaining 70% of images that consist of 4593 positive samples and 4850 negative samples. We also extract features from the re-trained CNNs using the fully connected FC7 layer.

The output of the SVM classifier is the prediction score for each category, larger the difference D is between the scores; the more confident the classifier is about the prediction. The final classification result is obtained by using (5) for each image at different acceptance threshold value t.

$$P = \begin{cases} LP, & D \geq t \\ Not\ LP, & D < t \end{cases} \tag{5}$$

The recognition rate for the combination of feature descriptors is given in such a way that, e.g. for SURF+HOG, if any of the feature descriptors classify correctly, it is assumed to be a correct decision. If both of them fail to identify correctly, the region under consideration is then presumed as false detection; while the performance of CNN has been evaluated individually.

Table 1 gives a detailed analysis of each model for performance evaluation. True Positive (TP) is the correctly identified LPs while True Negative (TN) gives correctly recognized non-LP regions. Feature Extracting Time (FE.T) is the time taken by each model for feature extraction from training images. For the case of CNN, it represents the time taken from training of the network till obtaining its activations. Training Time (T.T) is the time consumed for training SVM using the extracted features while Evaluation Time (E.T) is the time taken to extract features from the test image and its evaluation by the classifier. The overall accuracy (Acc%) is obtained by dividing sum of TP and TN with total number of validation images used for validation.

TABLE I. STATISTICAL ANALYSIS OF FEATURE DESCRIPTORS

Descriptor Type	*TP%*	*TN%*	*Acc%*	*FE.T (s)*	*T.T (s)*	*E.T (ms)*
SURF at t = 0.2	96.8	94.7	95.7	1632	218	86.0
HOG at t = 0.4	95.7	98.1	96.9	26.99	2.58	4.9
LBP at t = 0	93.8	86.5	90.04	29.40	0.35	4.6
SURF+HOG at t = 0.2	98.9	97.7	98.3	1659	221	89.6
SURF+LBP at t = 0.6	97.6	97.6	97.5	1661	219	89.1
HOG+LBP at t = 0.6	97.2	94.8	95.9	56.39	2.93	8.8
Scratch AlexNet at t = 0.6	97.1	97.3	97.2	110e4	2.47	990
Finetune AlexNet at t = 0.3	99.6	99	99.3	109e4	4.1	996

The HOG feature individually gives the best recognition rate i.e. 96.94% at t=0.4 while considering collectively, a best of 98.35% recognition rate is achieved at t=0.2 for SURF+HOG. This combination is found to be effective as the SURF features have scale and rotation invariance property while the HOG features are insensitive to geometric and photometric transformations which are essential for effective classification of LPs in real-time.

On the other hand, the fine-tuned AlexNet outperform all others in terms of recognition accuracy of 99.27% at t=0.3 but requires more processing time, both in training and testing, as compared to non-deep learning approaches.

The precision-recall graph for different values of acceptance threshold t for given models is shown in Fig. 9. It can be seen that the accuracy of SURF+HOG and fine-tuned AlexNet slightly differ but training and testing a CNN requires extensive processing as compared to other models. Considering the trade-off between time and accuracy, the

method with considerably higher accuracy can be incorporated in the existing or new developing ALPR algorithms to increase system's efficiency.

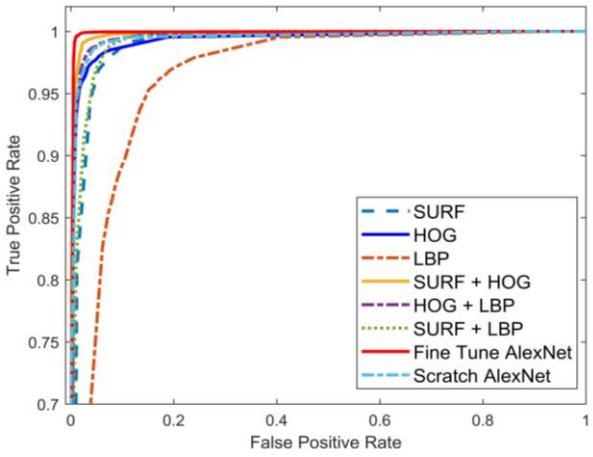


Figure 9. ROC curve for LP and non-LP region classification

The LP verification method employed in one of the existing LP detection methods [4] has been replaced by the proposed HOG+SURF combination method to verify performance improvement. The results show that the recall rate decreases marginally from 93.86% to 93.85% as some of the correctly detected LPs were considered as non-LPs at verification phase while the precision rate rises considerably from 87.15% to 90.73% as many fake regions were rejected at the cost of 0.16s in average processing time that rises from 0.25s to 0.41s on average per image. There is a slight difference between the validation and testing results as several non-LPs had similar pattern as LP regions.

The use of local image features is an effective way for the verification of LPs regardless of color, size, content and style. Moreover, such methods restrict the use of geometrical constraints to allow the system to be more dynamic to multinational LPs. In future, advanced feature combination techniques and low level features extracted from CNN architectures can be examined with different types of SVMs for improved performance. Furthermore, these models can be merged in existing systems for superior performance.

V. CONCLUSION

For an effective ALPR system, it is essential that a detected LP is validated first before segmentation and recognition of characters on it. This paper presents comparison experiments on three local image feature descriptors (SURF, HOG and LBP) and their complementarities along with the modified AlexNet CNN for LP classification. Cross validation on 9443 images show that HOG descriptor is most appropriate for the validation of LPs individually. When feature combination is used, SURF+HOG report the best performance while the fine-tuned AlexNet outperform others in terms of recognition accuracy. It is verified that the complementarities of features result in superior performance as compared to individual features. The proposed method is also incorporated in an existing LP detection method to verify improved performance. The use of local image features gives a dynamic solution for the verification of multinational LPs without the use of geometrical attributes.

REFERENCES

[1] J. L. Tan, S. A. Abu-Bakar, M. M. Mokji, "License plate localization based on edge-geometrical features using morphological approach," 20th IEEE Intl. Conf. Image Processing (ICIP), 2013, pp. 4549-4553.

[2] A. Azam, M. M. Islam, "Automatic license plate detection in hazardous condition," Journal of Visual Communication and Image Representation, 2016, vol. 36, pp. 172-186.

[3] H. J. Lee, S. Y. Chen, S. Z. Wang, "Extraction and recognition of license plates of motorcycles and vehicles on highways," 17th IEEE Intl. Conf. Pattern Recognition (ICPR), 2004, pp. 356-359.

[4] M. R. Asif, C. Qi, S. Hussain, M. S. Fareed, "Multiple license plate detection for Chinese vehicles in dense traffic scenarios," IET Intelligent Transport Systems, Oct 2016, vol. 10 (8), pp. 535-544.

[5] K. Kamel, A. A. Ouamri, M. Keche, "A new method for license plate validation using the orientations algorithm," 3rd European Workshop on Visual Information Processing (EUVIP), 2011, pp. 88-92.

[6] B. Y. Amirgaliyev, C. A. Kenshimov, K. K. Kuatov, M. Z. Kairanbay, Z. Y. Baibatyr, A. K. Jantassov, "License plate verification method for automatic license plate recognition systems," 12th Intl. Conf. Electronics Computer and Computation (ICECCO), 2015, pp. 1-3.

[7] N. Ratha, K. Karu, S. Chen, A. K. Jain, "A Realtime Matching System for Large Fingerprint Database," IEEE Transaction on Pattern Analysis and Machine Intelligence, 1996, vol. 18(8), pp. 799-813.

[8] D. Jingyu, Z. Sanyuan, Y. Xiuzi, Z. Yin, "Chinese License Plate Localization in Multi-Lane with Complex Background Based on Concomitant Colors," IEEE Intelligent Transportation Systems Magazine, 2015, vol. 7(3), pp. 51-61.

[9] A. H. Ashtari, M. J. Nordin, M. Fathyn, "Iranian License Plate Recognition System Based on Color Features," IEEE Transations on Intelligent Transportation Systems, 2014, vol. 15(4), pp. 1690-1705.

[10] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, "Speeded-up robust features (SURF)," Computer Vision and Image Understanding, 2008, vol. 110(3), pp. 346-359.

[11] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection," IEEE Intl. Conf. Computer Vision and Pattern Recognition, 2005, pp. 886–893.

[12] T. Ojala, M. Pietikainen, D. A. Harwood, "A comparative study of texture measures with classification based on featured distributions," Pattern Recognition, 1996, vol. 29, pp. 51-59.

[13] M. Pietikainen, T. Ojala, "Texture Analysis in Industrial Applications," Springer Berlin Heidelberg, 1996.

[14] M. Pietikainen, "Local Binary Patterns," in Scholarpedia, 2010.

[15] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, vol. 24(7), pp. 971-987.

[16] K. Alex, S. Ilya, E. Geoffrey, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, 2017, vol. 60 (6), pp. 84-90.

[17] W. Qian, Z. Cailan, X. Ning, "Street view image classification based on convolutional neural network," 2nd IEEE International Conference on Advanced Information Technology, Electronic and Automation Control (IAEAC), 2017, pp. 1439-1443.

[18] D. Halil, O.G. Ece, K. Mürvet, "Disease detection on the leaves of the tomato plants by using deep learning," 6th IEEE International Conference on Agro-Geoinformatics, 2017, pp. 1-5.

[19] W. Yeqing, L. Yi, P. Fatih, "Finetuning Convolutional Neural Networks for visual aesthetics," 23rd IEEE International Conference on Pattern Recognition, 2016, pp. 3554-3559.

[20] S.E. Mohamed, S.A. Sara, A.Y. Inas, "Low quality dermal image classification using transfer learning," IEEE International Conference on Biomedical & Health Informatics (BHI), 2017, pp. 373-376.

[21] S.M.S. Islam, R. Shanto, R. Mostafijur, K.D. Emon, M. Shoyaib, "Application of deep learning to computer vision: A comprehensive study," 5th Intl. Conf. Informatics, Electronics and Vision (ICIEV), 2016, pp. 592-597.