A Coarse-to-Fine Strategy for Vehicle Logo Recognition from Frontal-View Car Images[1]

S. Sotheeswaran[a],* and A. Ramanan[b],**

[a] *Department of Mathematics, Eastern University, Sri Lanka*
[b] *Department of Computer Science, University of Jaffna, Sri Lanka*
** e-mail: sotheeswarans@esn.ac.lk*
*** e-mail: a.ramanan@jfn.ac.lk*

This paper proposes a vehicle logo recognition (VLR) system centered on front-view cars, which has been largely neglected by vision community in comparison to other object recognition tasks. The study focuses on local features that describe structural characteristics by locating the logo of a car using a coarse-to-fine (CTF) strategy that first detects the bounding box of a car then the grille and at last, the logo. The detected logo is then used to recognize the make of a car in a reduced time. Our system starts to progress in detecting the bounding box of a car by means of a vocabulary voting and scale-adaptive mean-shift searching strategy. The system continues to process in locating the bounding box of an air-intake grille using a scale-adaptive sliding window searching technique. In the next level, the bounding box of a logo is located by means of cascaded classifiers and circular region detection techniques. The classification of vehicle logos is carried out on the patch-level as occurrences of similar visual words from a visual vocabulary, instead of representing the patch-based descriptors as bag-of-features and classifying them using a standard classifier. The proposed system was tested on 25 distinctive elliptical shapes of vehicle logos with 10 images per class. The system offers the advantage of accurate logo recognition of 86.3% in the presence of significant background clutter. The proposed scheme could be independently used for part recognition of grille detection and logo detection.

Keywords: car detection, coarse-to-fine strategy, logo classification, vehicle logo recognition

DOI: 10.1134/S1054661818010170

1. INTRODUCTION

Vehicle logo recognition (VLR) is vital, yet challenging task in the area of intelligent transportation surveillance systems. The enormous number of car designs and car model makes car to be a rich object class, which can potentially foster more sophisticated and robust computer vision models and algorithms. Cars present several unique properties that other objects cannot offer, which provides more challenges and enables more challenging fine-grained tasks in object categorization. Existing VLR systems solely rely on a preprocessing stage where the car images are tailored and close-fitting region-of-interest selection is performed to reduce the complexity of background. Though existing systems significantly rely on license plate recognition they are generally misled by fake number plates. In this regard, the recognition of vehicles make adds another level of security through verification. VLR is a challenging process due to the presence of extensive background, clutter, different degree of illumination, varying sizes of vehicles in motion, and change of weather conditions (see Fig. 1).

This paper focus to detect front-view cars and recognize their make from static images. To detect moving cars on road in video sequences, the most general approach is to extract motion features through the background subtraction. However, this technique is impossible when the background includes different environmental variations such as lighting changes or camera vibrations. Usually, recognizing a specific instance of an object may be easier due to particular distinctive features. When considering the detection of front-view cars some basic features such as logo, air-intake grille, side mirrors, and bumpers are prominent to discriminate one from the other.

A vehicle detection system for cars can be viewed with different perspectives such as front, back, side, and oblique. For the front, back or side-view cars, the knowledge of horizontal structures can be used to generate the hypothesis area, and for oblique car, a template matching technique can be applied to generate the vehicle hypothesis. In general the detection of vehicles from dynamic scenes consist of three main approaches: (i) Background subtraction methods, (ii) frame differencing and motion based methods, and (iii) feature based methods. In background subtraction methods the non-adaptation is a drawback which is due to the variation in lighting and weather conditions. The frame differencing and motion based meth-

[1] The article is published in the original.

Received December 12, 2016

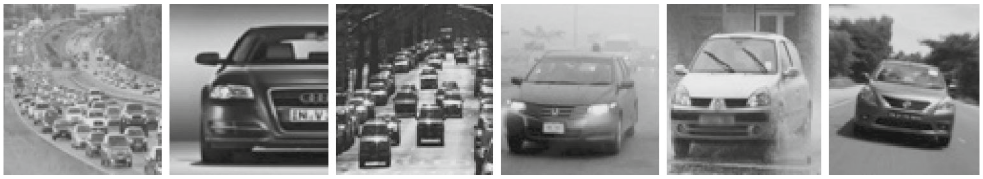

Fig. 1. Challenges of car detection.

ods are very sensitive to the changes in the external environment and has poor anti-interference ability. The feature based method supports the occlusion handling between overlapping vehicles and shows less computational difficulty in comparison with the background subtraction method. Most of the works in the literature on VLR have mainly been addressed by means of license plate (LP) detection [1–5].

This paper contributes in the context of car logo detection using a coarse-to-fine strategy that first detects the bounding box of a car from the given input image then the grille and at last, the logo. For car detection, we demonstrate a novel vocabulary voting and mean-shift searching technique [6]. For localizing the car logo, we demonstrate a scale-adaptive sliding window searching technique to first detect the air-intake grille using histogram of oriented gradient (HOG) features [7] and support vector machine (SVM) classifier [8] and then the detection of logo by means of Viola-Jones method [9] with Circular hough transform (CHT) [10] technique. The detected logo will be then recognized for the make of the car. For logo recognition, we demonstrate a novel classifier-free vocabulary-based classification technique that not only speeds up the classification process but also improves the classification rate when compared with the traditional bag-of-features (BoF) approach [11] and [12].

The rest of this paper is organized as follows. Section 2 reviews various techniques that have been used in the literature for vehicle detection and logo recognition. Section 3 explains in detail the proposed methodology for VLR using a coarse-to-fine strategy. Section 4 provides details of exploratory work and testing results. Finally, in Section 5 we draw conclusions about this work.

2. PREVIOUS WORK

Pan and Zhang [13] have proposed an approach for accurate vehicle logo detection using a coarse-to-fine strategy. The authors have used two steps to localize a logo: (i) detection of vehicle region and (ii) segmentation of a small region of interest (ROI) which rapidly localize the logo. The detection of vehicle region is achieved by following three steps: (i) Images are resized to 128 × 128, 256 × 256 and 512 × 512, respectively; (ii) Haar and HOG features are extracted from each scaled image using a sliding window approach. These features are passed to an Adaboost classifier to generate hypothesis images delineated by rectangular boxes, which indicate potential vehicle objects; and (iii) the final bounding box is detected from the hypothesis images that will undergo scaling up operation followed by a thresholding operation, which disregards the small hypothesis images of size less than that of the threshold. Segmentation of logo RoI from vehicle's low frontal part is performed by localizing the number plate as the reference. Thereafter, HOG features are extracted from these ROIs. Finally, logo candidate regions are classified using a two-stage classifier, which is mainly composed of gentle Adaboost and SVM. Authors' dataset consists of 1225 vehicle images together with 521 non-vehicle images for training and 274 vehicle images for testing purpose each with 320 × 320 pixels corresponding to 11 distinctive vehicle manufactures obtained from the local Police Department of Dushu Lake Higher Education Town in Suzhou traffic surveillance cameras. The vehicle detection accuracy is reported to be 95.18%. The choice of an appropriate initial size of the sliding window over the segmented RoI has influence over the detection rate, and the hybrid approach of the two-stage cascaded classification strategy takes more processing time.

Hsieh et al. [14] have proposed a system to detect vehicle and to recognize the make and model of the vehicle using a novel symmetrical SURF descriptor [15]. The symmetrical SURF enriches the power of SURF descriptor to detect all possible symmetrical matching pairs through a mirroring transformation. ROI detection is done using symmetric property of vehicles without using any motion features. Afterward, ROI is divided into 4 × 6 regions and the extracted features via HOG and SURF from each region are used to train various weak vehicle classifiers by using an SVM learning algorithm. With a Bayesian averaging technique, these weak classifiers are then integrated to form an ensemble classifier to recognize vehicle types with better accuracy. Authors dataset consists of 2846 vehicle images for training and 4090 vehicle images for testing purpose corresponding to 29 distinctive vehicle make and models. The vehicle detection accuracy is reported to be 98% and the vehicle type recognition rate to be 97.95%. The method proposed by these authors highly depends on prior knowledge about symmetric property of vehicles, thus license plate recognition, in general, does not generate satisfactory results in countries or regions where vehicles have diverse license plates which might be issued by various authorities.

Zhou et al. [16] have proposed an algorithm to detect the location of vehicle logo based on the combination of feature appearance and symmetric property of the front image of a vehicle. The contributions of the proposed system are adaptable and independent, i.e., does not rely on the LPR process of the system. Authors have used a coarse-to-fine strategy to detect the coarse location of the vehicle logo by-product of air-intake grille detection of the vehicle. The region of the air-intake grille is detected using edge detection, bilateral symmetry of the image and Hough transformation. The authors have used SIFT-match based symmetry detection algorithm for detecting bilateral symmetry in the image. A coarse location of the vehicle logo is extracted from the image by detecting bilateral symmetry of the air-intake grille image. The overall logo detection accuracy is reported to be 98.3% with an average detection time of 332 ms per image. The authors have used their own dataset consisting of 1798 digital images captured from traffic surveillance videos in areas, such as highways, roads and parking lots to evaluate their proposed method.

Ou et al. [17] have proposed a system for VLR based on a weighted spatial pyramid framework. Firstly, HOG features are extracted from the given image and ROIs are detected using a 12 staged AdaBoost-based detector. Secondly, dense SIFT descriptors [18] are extracted from the union of detected ROIs for robust description of the image. A weighted image has been generated by means of accumulating the weights from all the ROIs in the image. A spatial pyramid framework is then introduced from these ROIs. Thereafter the system is implemented using locality-constrained linear coding to generate the code vectors for extracted local descriptors. Furthermore, the weighted local code vectors are calculated and pooled in each region separately under the spatial pyramid by using max pooling. Finally, the concatenated feature vector of vehicle logos are classified using one-versus-one linear SVM classifier. The authors have evaluated their proposed system on images collected from surveillance cameras at toll-gates on the highways. An image consists the region above the LP which is obtained by a LP detection technique. The dataset consists of 1791 images with 15 distinctive vehicle manufacturers. The vehicle logos have an average size of 45 × 45 pixels. Their overall logo recognition accuracy is reported to be 98% with a recognition time of 300 ms per image.

Yu et al. [19] have proposed a system to VLR based on bag-of-features approach. In their approach three steps have been followed to recognize vehicle logo images: (i) dense-SIFT features were extracted from the logo images; (ii) visual vocabulary was constructed by soft-assignment; (iii) histograms of visual features were constructed in multi-level with spatial information. They have investigated two kinds of quantitative features into visual features and classification techniques such as hard-assignment and soft-assignment, with k-nearest-neighbour and one-vs-all SVMs, respectively. The proposed system is tested and compared using six-fold cross-validation technique. Their system is evaluated on a vehicle logo image database which consists of 14 distinctive vehicle classes with 60 images per class with about 30 × 30 pixels. The authors have used their own dataset that consists of well-segmented vehicle logo images. The framework that is made of dense-SIFT, soft-assignment and SVM achieves better recognition rate of 97.3% with 23 ms recognition time per image.

Llorca et al. [20] have proposed a system to VLR using HOG features and SVM classifier. The authors have used a sliding window technique combined with a majority voting scheme to recognize the vehicle logo images. The authors have compared the performance of HOG and linear-SVM against HOG and RBF-SVM. Their system is evaluated on a vehicle image database which consists of 3579 images belonging to 27 distinctive vehicle classes with resolution of 640 × 480 pixels. They have created 162 samples from each training images by horizontal mirroring, geometric jittering and varying the size. It has reported that, HOG with linear-SVM achieves an accuracy of 92.6%. However, the sliding window technique would cost much more time.

Huang et al. [21] have proposed a VLR scheme that first captures images from a monitoring kit and the vehicle license plate is then identified using an embedded module that is available in the kit. The vehicle logo is detected from the region of interest and the background around the logo is removed. The segmented logo is then categorized by a back-propagation neural network classifier using the last layer of the CNN structure. In addition, PCA-based pretraining strategy is introduced to improve the CNN-based system by reducing the computational cost of the training stages and enhancing the classification performance. The coarse location of the logo is segmented and the ROI is normalized to 70 × 70 pixels. The authors have created their own image set consisting of 11,500 car logo images belonging to 10 distinct manufacturers with 10,000 images used for training and 1,500 for testing. An average logo recognition rate is reported to be 99.07%. The authors have considered only 10 distinctive logos in testing with the assumption that the vehicle logo is always located within a certain ROI above the license plate. The recognition accuracy of the entire system strongly depends on the success of license plate detection.

3. METHODOLOGY

This section provides detailed information of the proposed method for VLR in the following four steps:

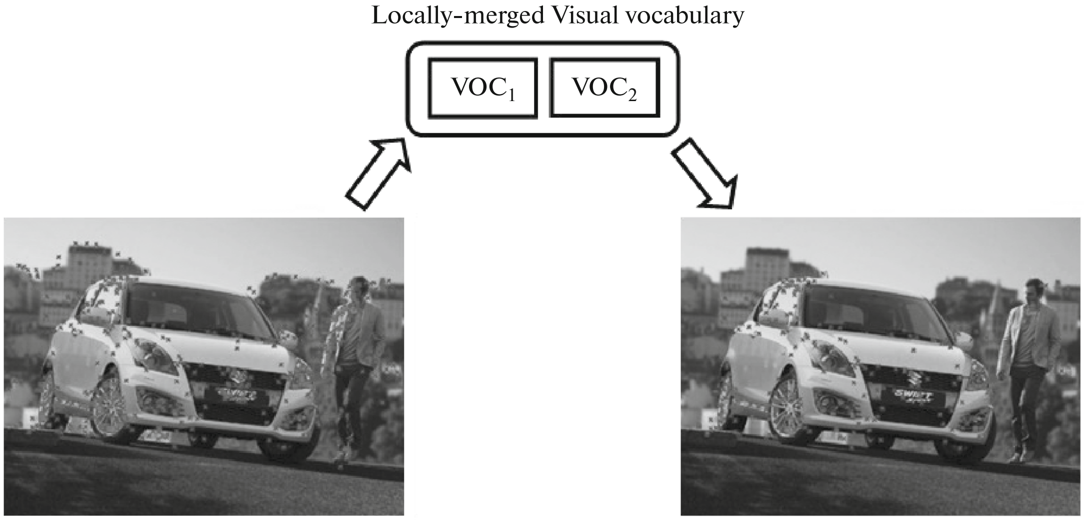

Fig. 2. Retaining foreground keypoints using vocabulary voting strategy. (a) Initially detected e-SURF keypoints (b) Retained foreground keypoints after eliminating possible background keypoints.

(i) Detection of front-view car using vocabulary voting and mean-shift search techniques. (ii) Detection of air-intake grille using scale-adaptive search technique. (iii) Localizing the logo using circular region detection techniques, and (iv) recognizing the make of a logo using a classifier-free vocabulary-based classification technique.

3.1. Front-View Vehicle Detection

The detection of cars in front-view static images is described here. The images can be gray-level or color. The proposed method focuses on local patch-based descriptors that describe the structural characteristics of a particular object. The assumption of such patch-based descriptors is that in different images the statistical distribution of the patches is different. If a vehicle is partially occluded in a given scene, it can be detected with the aid of patch-based visual descriptors such as SIFT and SURF. SURF features can be extracted faster than SIFT using the gain of integral images and yield, a lower dimensional feature descriptor resulting in faster matching and less storage space. Moreover, SURF slightly outperforms SIFT in illumination changes, which is a desired property to the VLR process as vehicles are highly exposed to external weather conditions such as fog, rain and sun. Therefore, extended speeded up robust features (e-SURF) descriptors were selected in the experiments of front-view vehicle detection. Limited experiments conducted in this work favors the choice of e-SURF descriptors than SIFT in front-view car detection. The proposed method [22] is based on foreground-background separation technique by means of constructing visual vocabularies for the car (i.e., foreground) and non-car objects (i.e., background) using K-means algorithm. The keypoints for the foreground vocabulary are extracted from manually segmented bounding boxes that consist of car only.

When an unseen test image of a car is produced to the system, e-SURF descriptors are extracted from it and will cast votes to the foreground and background vocabularies in order to retain the candidate keypoints of the car. We measure the L2 norm between each descriptor of a test image with each of the cluster center in the visual vocabulary. This process is continued until all the descriptors of a test image are explored with the visual vocabularies (see Fig. 2).

The vocabulary voting scheme can suffer from keypoints that are detected in background when casting votes to the foreground vocabulary. This will result in retaining some keypoints from the background image. It can be assumed that such keypoints do not belong to densest regions in the image. Thus we run a scale-adaptive mean-shift search technique on a test image of a car by down-sampling the image as shown in Fig. 3 using Eq. (1) to find set of possible bounding boxes of the car. A test input image is rescaled using bilinear interpolation with a start scale (S_s) to an end scale (S_e) with scale stride (S_r) [23]. The number of scale steps is computed as follows:

$$\text{Number of scale steps} = \left\lfloor \frac{\log\left(\frac{\mathrm{S_e}}{\mathrm{S_s}}\right)}{\log(\mathrm{S_r})} + 1 \right\rfloor. \qquad (1)$$

In every scale, we find the dense region of the candidate keypoints by means of mean-shift searching algorithm which is controlled by its bandwidth. The bandwidth is calculated as $\max(W_s, H_s)/3$, where W_s and H_s are the width and height of the test image at sth

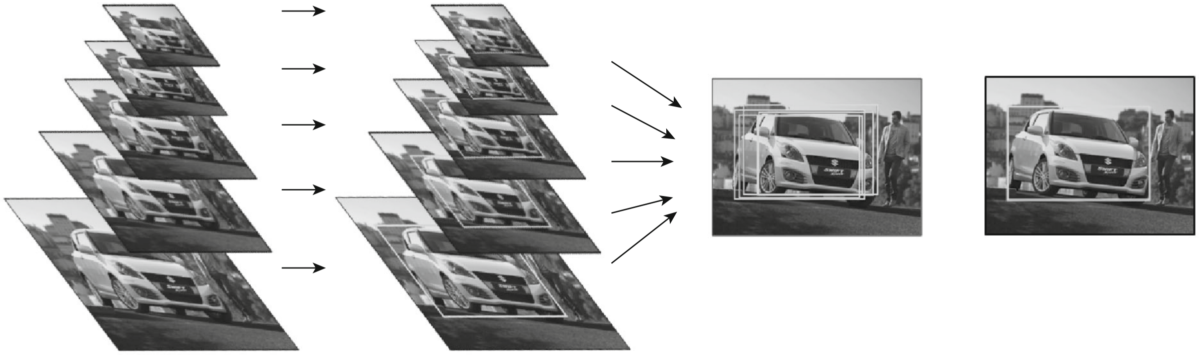

Fig. 3. An example test image: Detecting the bounding box of a car using vocabulary voting and scale-adaptive searching methods. Each test image is downscaled from S_s to S_e. Voting of e-SURF descriptors is performed with the constructed vocabularies using a nearest neighbour approach to retain a set of possible keypoints at each scale that belongs to the car. A set of bounding boxes is found using mean-shift algorithm applied in different scales and the final bounding box of the car is detected by fusing the set of bounding boxes using non-maximum suppression technique.

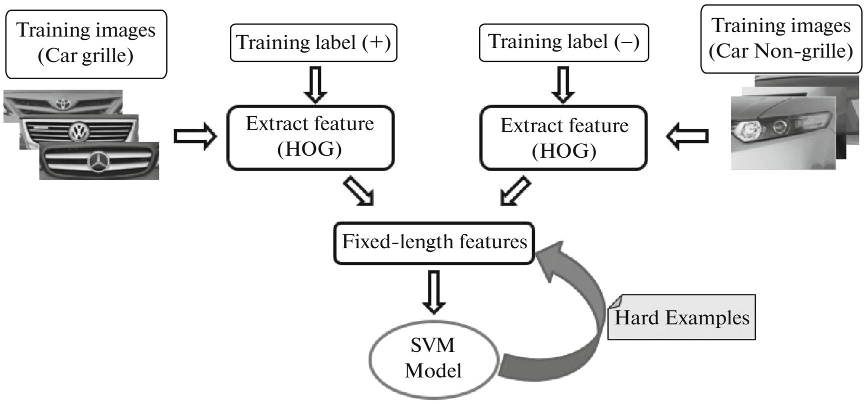

Fig. 4. Creating model to detect the car-grille.

scale, respectively. At every scale, the bounding box is drawn by means of finding the minimum and maximum values of the X-axis and Y-axis in the densest region.

The set of bounding boxes found at every scale are mapped in to the original test image and a non-maximum suppression technique is then used to predict the final bounding box of the car.

3.2. Grille Detection

In most of the published works, the position of the license plate was used as an aid to locate the area of the logo. To overcome the problem of existing logo localization techniques, in this section, vehicle logo localization has also been taken into consideration along with air-intake grille detection. The air-intake grille is the most recognizable visual shape, when comparing for the logo in the front-view of the vehicle. Air-intake grille is detected by using a scale-adaptive searching technique which combines HOG features and SVM classifier. The proposed algorithm of the air-intake grille detection is as follows:

(1) HOG features are extracted from positive samples (i.e., training images of air-intake car grille).

(2) HOG features are extracted from negative samples that do not contain any of the air-intake grille images.

(3) A linear SVM classifier model that detects bounding box of the air-intake grille is trained.

(4) Apply hard-negative mining: The sliding window technique is applied to every image in the negative training set at each possible scale. At each window, HOG descriptors are computed and fed to the SVM classifier (see Fig. 4). The false-positive samples are collected during the hard-negative mining stage.

(5) The SVM classifier model is then retrained by using the hard-negative samples.

(6) The sliding window technique is applied to every image in the testing set at each possible scale. At

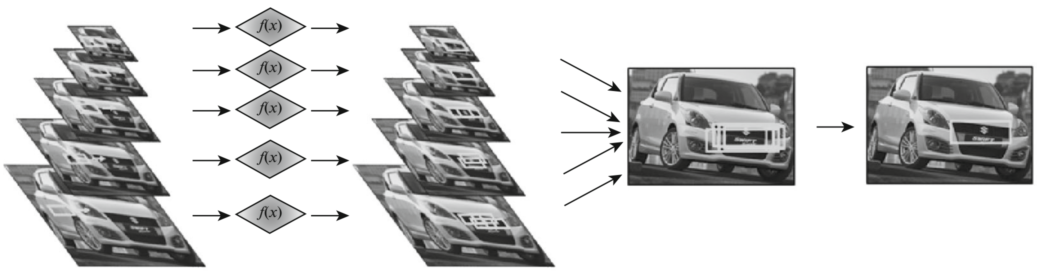

Fig. 5. Scale-adaptive sliding window search.

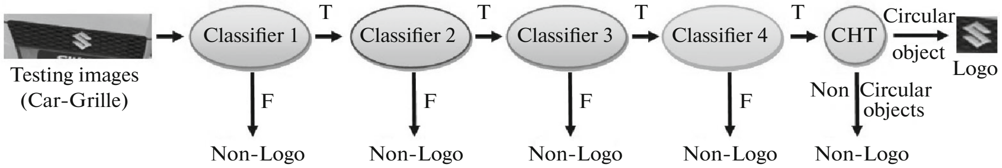

Fig. 6. Localization of the logo using cascade of classifiers and CHT.

each window, HOG descriptors are computed and fed to the retrained SVM classifier.

If the retrained SVM detects an air-intake grille with sufficiently large probability, then the bounding box of the window is recorded. After completion of scanning the image, a non-maximum suppression is applied to remove redundant and overlapping bounding boxes (see Fig. 5).

3.3. Logo Localization

Mostly, the vehicle logo lies on the middle of the air-intake grille in the front-view of the vehicle. In our approach the logo is localized from the detected grille image by using cascade of classifiers and circular hough transform technique. The algorithm of logo localization is as follows:

(1) Images of vehicle logos and air-intake grille without logos were used as positive and negative training examples, respectively.

(2) The cascade classification model is trained by extracting Haar-like features from the training examples.

(3) Viola-Jones algorithm is then applied to select region of interest (i.e., logo) by discarding the background regions.

(4) The circular hough transform technique is then applied to the ROIs to localize the logo.

A Cascade of classifiers is used to detect the logo sub-window with less computational time. Through this technique, boosted classifiers were constructed which reject many of the negative sub-windows (non-logo) while detecting almost all positive instances (logo). A series of four classifiers were applied to every sub-window of the detected air-intake grille image (see Fig. 6). Cascade classifiers eliminate non-logo sub-windows and pass the possible logo sub-windows. After four stages of processing, the number of sub-windows and false prediction rate have been reduced radically. At last, it will be classified as a possible logo sub-window.

3.4. Logo Classification

The detected logos in the previous stage are then classified for recognition. In this regard, a novel approach is proposed [24] to classify the vehicle logo to predict the particular class of the vehicle. We propose a classifier-free vocabulary-based image classification technique which employs a nearest neighbour max-win voting strategy on a learnt class-wise vocabularies of logos to predict the class label of a given test image. The proposed approach is free from the histogram representation of patch-based descriptors as fixed-length feature vectors and then feeding those vectors to a standard classifier as required in the traditional BoF method. Our proposed approach is evaluated on 25 distinctive vehicle classes and the schema of the technique is illustrated in Fig. 7 in rounded rectangle with dotted line. In our approach we measure the L2 norm between each descriptor of a test image with each of the cluster center in the locally merged global vocabulary.

The relative position of the visual word in the global vocabulary that results in minimum distance is recorded in a *list*. This process is continued until all the descriptors of a test image is explored with the global vocabulary. Finally a max-win voting strategy is applied on the *list* to predict the class label of a test

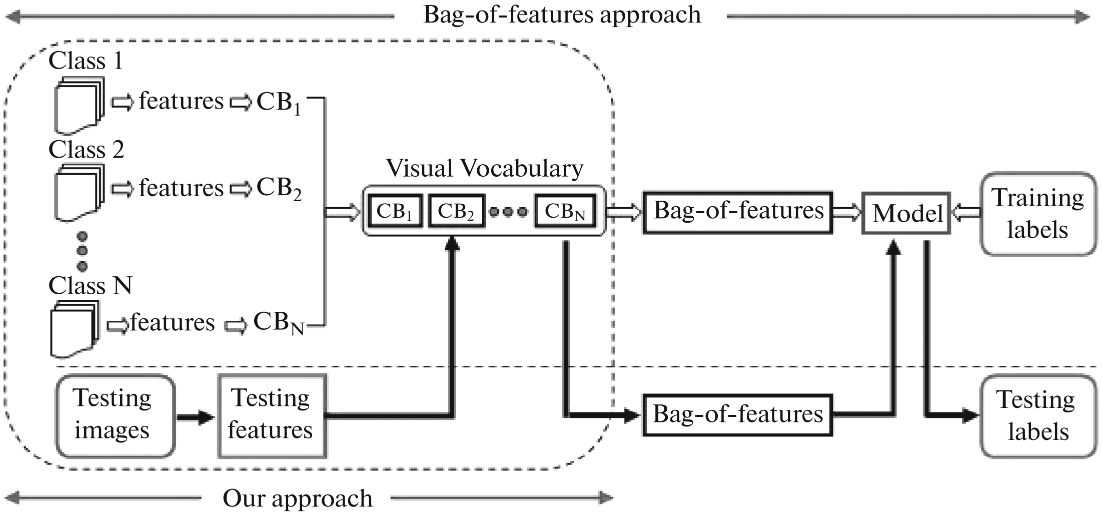

Fig. 7. Bag-of-features vs. our approach. (a) BoF approach: Training stage is applied to the training set to obtain a vocabulary and a classifier. Testing stage is applied to the test set using the previously computed vocabulary and classifier to obtain a label for every test image. (b) Our approach: In contrast the training stage of the proposed method is applied to the training set to obtain a locally merged global vocabulary. Testing stage is applied to the test set using the learnt global vocabulary to obtain a label for every test image by means of a nearest neighbour max-win voting strategy applied on the visual descriptors.

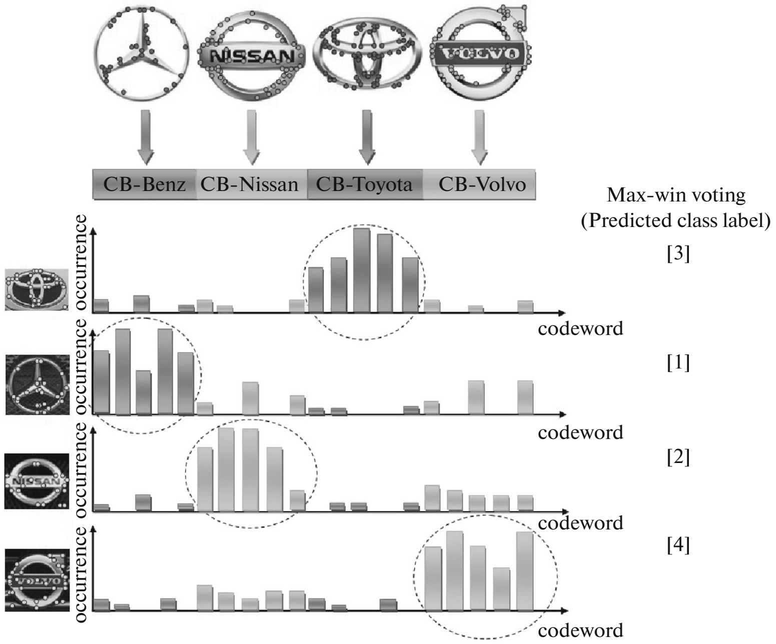

Fig. 8. Vocabulary-based classification.

image. This approach is free from the histogram representation and classification steps which not only speeds up the classification process but also improves the classification rate. Figure 8 depicts the nearest neighbour max-win voting strategy that is used in our approach to predict the class label of a test image.

4. EXPERIMENTAL SETUP AND TESTING RESULTS

This section provides a brief description of the dataset and the experimental setup of front-view car detection, air-intake grille detection, logo localization, and logo recognition along with the obtained testing results.

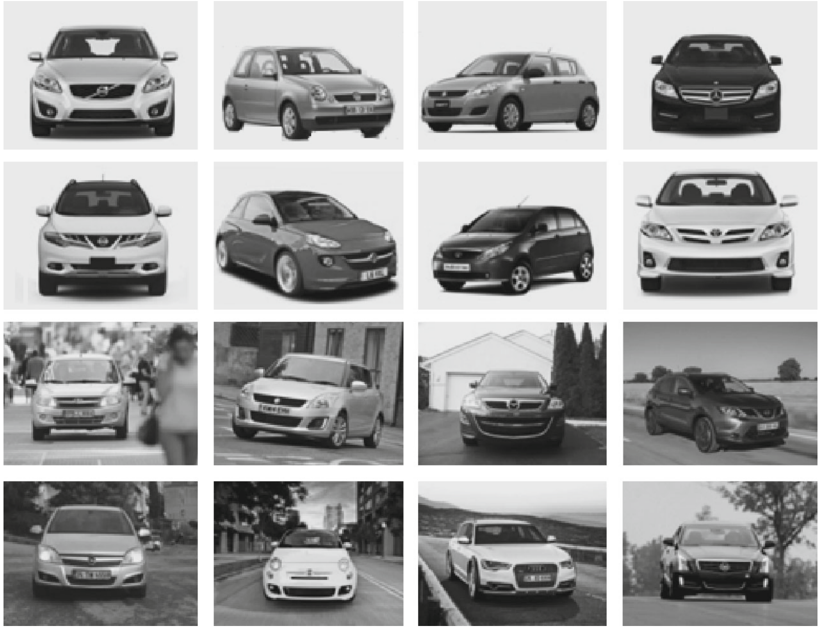

Fig. 9. Example car images of the dataset.

4.1. Dataset

We obtained images of 25 distinct classes: Acura, Audi, Buick, Cadillac, Daihatsu, Fiat, Ford, Honda, Hyundai, Infiniti, Kia, Lada, Lexus, Mazda, Mercedes Benz, Nissan, Opel, Renault, Subaru, Suzuki, Tata, Toyota, Vauxhall, Volkswagen and Volvo, of front-view cars with 20 images per class from Google Images with crucial variations such as scale, rotation, background, and lighting. Images were used without any preprocessing and each image consists only one car. Figure 9 shows some of the example car images of the dataset[2]. Each of the images is of size 800 × 600 pixels. All images within a class is numbered 1 to 20 where odd numbered images are of cars with different backgrounds, whereas even numbered images are with plain background. 120 images consisting no cars were obtained from Google Images as training images for background category.

4.2. Experimental Setup

Experimental setup is divided into the following four parts: Front-view car detection, grille detection, logo localization and logo classification. Experimental results reported are the mean together with the standard deviation evaluated against the following four different training-testing subsets:

Subset 1: All even numbered images as training and the rest as testing.

Subset 2: All odd numbered images as training and the rest as testing.

Subset 3: First ten images as training and the rest as testing.

Subset 4: First ten images as testing and the rest as training.

Through this four different training-testing subsets, all images of the dataset are tested against different set of training images. The system was tested with each subset having 25 distinctive vehicle classes with 10 images per class resulting in a total of 250 images.

4.2.1. Front-view car detection. The vocabulary for the foreground object was constructed by using 250 training images of cars. The vocabulary for the background object was constructed by using 120 training images of non-cars. e-SURF descriptors extracted from car images are clustered using K-means algorithm with $K = 1000$ to constructing the foreground vocabulary. Similarly the background vocabulary was constructed with $K = 1000$ resulting in a locally-merged visual vocabulary of size 2000. Each test image is downscaled by setting $S_s = 1$, $S_e = 3$ and $S_r = 1.05$ with the scale steps mentioned in Eq. (1). For each scale of a test image, e-SURF descriptors are extracted and the vocabulary voting strategy is performed using the nearest neighbour approach to retain a set of possible keypoints that belongs to the foreground object. Each image scale is then densely scanned using mean-shift algorithm with a Gaussian kernel to detect all bounding boxes. Thereafter a threshold θ is employed to reject false prediction bounding boxes, where θ is estimated in every possible bounding box using Eq. (2).

[2] www.csc.jfn.ac.lk

$$\theta = \frac{\#\text{Vehicle keypoints detected inside the bounding box}}{\#\text{Keypoints detected inside the bounding box}}. \quad (2)$$

In this work, bounding boxes having $\theta \le 0.52$ are rejected. We follow the criteria of true object detection from PASCAL VOC Challenges [25]. The object detection is considered according to the overlap ratio between the predicted bounding box of the object B_{po} and the groundtruth bounding box of the object B_{gto}.

$$\text{True detection} = \frac{\text{area}(\mathrm{B}_{\mathrm{po}} \cap \mathrm{B}_{\mathrm{gto}})}{\text{area}(\mathrm{B}_{\mathrm{po}} \cup \mathrm{B}_{\mathrm{gto}})} > 0.5. \quad (3)$$

Where $(B_{po} \cap B_{gto})$ denotes the intersection of the predicted and groundtruth bounding boxes and $(B_{po} \cup B_{gto})$ their union. The object detection rate is then computed as follows.

$$\text{Object detction rate} = \frac{\text{Number of true object detections}}{\text{Total number of testing images}} \times 100\%. \quad (4)$$

Here, front-view car is referred as an object. The above evaluation criteria was used in grille detection and logo localization using air-intake grille and logo as objects, respectively.

4.2.2. Grille detection. Vehicle grille images are manually segmented from training images. The segmented grille images are resized to 40 × 100 pixels. HOG features are extracted from the vehicle grille images. One grille image has 1584 (i.e., 11 × 4 × 4 × 9) dimensional fixed-length feature vector. Thereafter HOG features are extracted from non-grille images of vehicle using sliding window search. The size of the sliding window is 40 × 100 pixels. One sliding window has 1584 dimension fixed-length feature vector. In order to avoid the imbalance problem in the feature set, vehicle grille features are oversampled to the size of the non-grille features. These two feature sets are fed to binary SVM classifier. The SVM training model is then generated with a linear kernel by setting C to 0.01. To detect the hard negatives, training images of non-grille vehicle are down sampled using Eq. (1). HOG features are extracted from each scale using sliding window technique. Extracted HOG features from each scale are then fed to the SVM training model. If there is any sliding window that predicts it as the grille then the class label of the windows feature vector is changed from positive to negative. SVM training model is retrained by using new fixed-length feature vectors of the hard negatives. Testing images are also down sampled using Eq. (1). HOG features are extracted from each scale using sliding window technique. Possible bounding boxes of the grille are predicted at each scales by means of the retrained SVM training model. All predicted bounding boxes at each scales are then mapped to the original size of the image. Finally, non-maximum suppression technique is used to predict the final bounding box of the grille.

4.2.3. Logo localization. The vehicle logo images and non-logo images are manually segmented from grille area in the training images. Haar-like features are extracted from segmented logo and non-logo training images. The following formula are used to calculate the number of positive samples and negative samples at each stage of the cascaded classifier:

$$\#\text{positive samples} = \left\lfloor \frac{\text{total positive samples}}{1 + (\#\text{CascadeStages} - 1) \times (1 - \text{TPR})} \right\rfloor. \quad (5)$$

$$\#\text{negative samples} = 2 \times \#\text{positive samples}. \quad (6)$$

where TPR is true positive rate.

The number of available positive samples are used to train the cascaded classifier at each stage depends on the true positive rate and number of cascade stages. We have used four cascade stages and true positive rate with 0.9. Hence, 192 positive samples and 384 negative samples have been used at each stage. A sliding window over the test image is applied as a detector. The detector then uses a cascade classifier to decide whether the window contains the ROI of the logo or not. The size of the window varies to localize the ROI of logos at different scales, but its aspect ratio remains fixed. CHT is applied to ROI of logos to localize the final logo of a vehicle. For this experiment, we used the range of circle radii from 8 to 20 pixels.

4.2.4. Logo classification. Vehicle logo is classified by using a novel classifier-free vocabulary based approach which is compared to the standard BoF approach. The logos of front-view vehicles are manually segmented from training images. Figure 10 shows some of the examples of manually segmented logo images of Volkswagen logo from our dataset. SIFT features were extracted from regular patches on the segmented logos.

SIFT features were then clustered using the K-means algorithm. We also studied the effect of cluster size (i.e., the size of the visual vocabulary) on the classification rate by setting the K of K-means to 30, 40 and 50. We constructed separate vocabularies of size K for each logo and concatenated the vocabularies to form a global vocabulary of size 25 × K. The following two different approaches are used to classify the make of a vehicle logo.

Bag-of-features approach: Following the construction of a global vocabulary, the extracted image descriptors were then mapped into a feature vector by computing the frequency histograms with the learnt clusters. This mapping produces a BoF representation. The size of the BoF representation is equal to the size of the locally merged global vocabulary. Each feature vector of size 25 × K was fed separately into multiclass

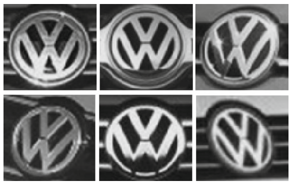

Fig. 10. Some example images of volkswagen logo in our dataset.

SVM classifiers. We used one-versus-all SVMs with RBF kernel for the classification of logos. The implementation of multiclass classifiers was performed using the SVM^{light} package [8]. We estimate the generalized accuracy using different kernel parameters γ and cost parameters C such that $\gamma = [2^{-10},...,2^3,2^4]$ and $C = [2^{-2},...,2^{12}]$.

Classifier-free vocabulary-based approach: In our approach, SIFT features were extracted from regular patches on the testing images of vehicle logos. The Euclidean distance was measured between each descriptor of a test image with each of the cluster center in the locally merged global vocabulary. The relative position of the visual word in the global vocabulary that results in minimum distance is recorded in a list. This process is continued until all the descriptors of a test image is explored with the global vocabulary. Finally a max-win voting strategy has been applied on the list to predict the class label of a test image. The classification rate was computed as follows:

$$\text{Classification rate} = \frac{\text{Number of correctly classified images}}{\text{Total number of testing images}} \times 100\%. \tag{7}$$

4.3. Testing Results

Figure 11 depicts the overall framework and the corresponding testing results obtained in VLR at each stage of the CTF strategy.

4.3.1. Front-view car detection. A true bounding box of the front-view car detection achieved an average rate of **97.2 ± 0.6%** which yields low rate of false positives on all four different testing subsets. It has been observed that varying the bandwidth of the mean-shift algorithm gives an increased detection rate by **1.2%**. Figure 12 shows the empirical results obtained using the proposed approach to detect the car from its background.

4.3.2. Grille detection. The output of the detected car images from the previous method becomes the input to the grille detector. On the other hand, manually segmented car images were also tested with the grille detector. A true bounding box of the air-intake grille detection using the proposed method achieved a rate of **94.0 ± 2.4%**, whereas the manually segmented images yielded a rate of 95.5 ± 1.8%.

Figure 13 shows the empirical results obtained using the scale-adaptive sliding window approach to detect the air-intake grille in its detected car image.

4.3.3. Logo localization. The output of the detected air-intake grille images from our method becomes the input to the logo localizer. On the other hand, manually segmented air-intake grille images were also tested with the logo localizer. A true bounding box of the logo localization using our method achieved a rate of **90.4 ± 2.9%**, whereas the manually segmented images yielded a rate of **94.0 ± 2.5%**.

Figure 14 shows the empirical results obtained using Viola-Jones and CHT techniques to detect the logo from the detected grille images.

4.3.4. Logo classification. The testing results of the logo classification compares the BoF approach and the proposed classifier-free vocabulary-based approach. Each of the approaches were tested with manually segmented logo images and automatically detected logos from the proposed CTF strategy. The testing results obtained from BoF approach and the classifier-free vocabulary-based classification approach are reported in Table 1 as the mean classification rate with standard deviation for all subsets of the datasets. Testing results compare different sizes of vocabularies of BoF and classifier-free vocabulary-based approaches.

In all categories, the testing results show that our approach is comparable or slightly better than the traditional BoF approach.

The CTF mechanism proposed in our VLR system aims on restricting the intensive processing to a very small portion of the original input image, i.e., those regions containing actual logo or logo-like clutter. In addition, CTF strategy sufficiently discriminates the region-of-interests against background to limit the number of false detections. We consider detecting and localizing shapes of front-view air-intake grille and logo of a car in cluttered gray-level images under a wide range of geometric presentations and imaging conditions (resolution, lighting, background, etc.).

The proposed method for detecting a vehicle from its background is flexible and can be easily extended to other views such as side-or rear-view by modifying the visual vocabularies. The proposed classifier-free vehicle logo classification technique not only speeds up the classification process but also reduces the storage requirement when comparing with the traditional BoF approach while maintaining comparable performance. The computational time saving of our proposed approach is summarized in Table 2.

- Total time required by the BoF approach is $t_1 + t_2 + t_3 + t_4$,
- Total time required by our approach is $t_1 + t_2 + t$,

where $t << (t_3 + t_4)$. From our testing results, the average time of t is reduced by tenfold as compared to $(t_3 + t_4)$. Furthermore, the storage saving of our approach is summarized in Table 3,

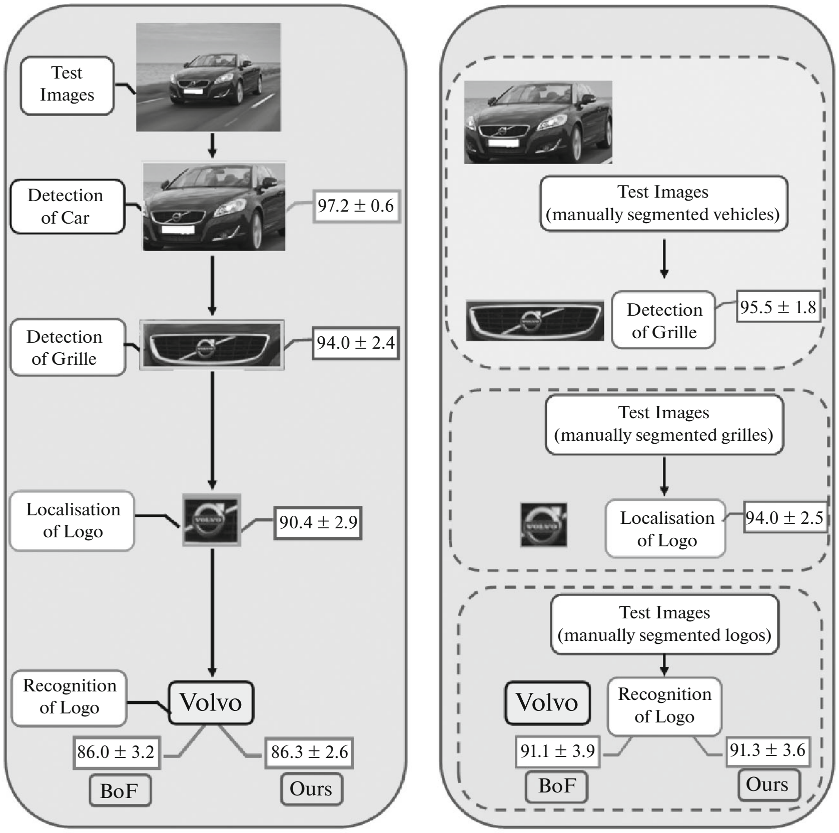

Fig. 11. Testing results of Region-of-interest detection (i.e., bounding box of a car, air-intake grille, and logo) of the proposed CTF strategy. (a) Series of detection stages by passing output of one detector as an input to another detector and the detection rate at each stage. (b) Manually cropped ROIs as input to every detector and the detection rate.

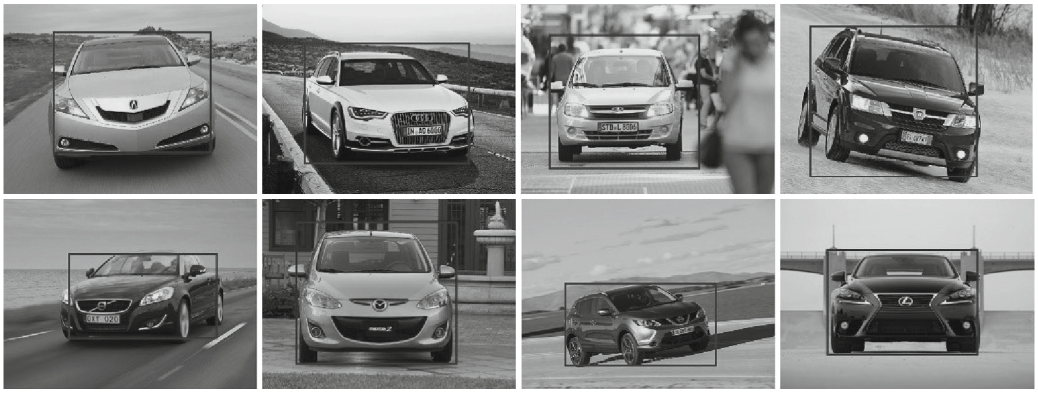

Fig. 12. Some of the cars detected results on the testing dataset.

where, K is the size of a local vocabulary, L is the number of distinctive logos, M is the number of training images, N is the number of testing images, and P is the total number of SIFT descriptors of $M + N$ images.

- Total storage required by BoF approach is $(P \times 128) + (L \times K \times 128) + (M + N) \times (L \times K)$,

- Total storage required by our approach is $(P \times 128) + (L \times K \times 128)$.

In our approach there is no need to store the training and testing feature vectors in the form of histograms as needed by the BoF approach. Thus there is a storage savings of $(M + N) \times (L \times K)$ matrices in our approach.

All of our experiments were implemented in MATLAB and executed on a desktop computer with an Intel Core i3 running at 3.1GHz and 4GB of RAM.

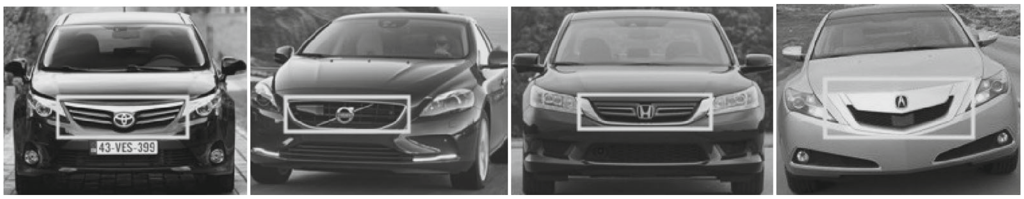

Fig. 13. Some of the grille detection results on the testing dataset.

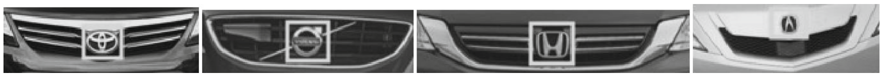

Fig. 14. Some example images of localized logos through CTF strategy.

5. CONCLUSIONS

This paper addressed the front-view VLR system in static images and focused on local features that describe structural characteristics by locating the logo of a car using a coarse-to-fine strategy that first detects the bounding box of a car then the grille and at last, the logo. The detected logo is then used to recognize the make of a car by a novel approach in reduced time. Our system captures the inter-class variability and intra-class similarity to recognize the make of the vehicle in a coarse-to-fine setup without any manual segmentation of a precise region of interest. An inaccurate ROI detection of a bounding box of car and/or air-intake grille in the CTF strategy results in clipping of important information associated with the logo in some cases or in other case it takes in unnecessary information from background region that affects the final performance. The classification of vehicle logos is performed on patch level as occurrences of similar visual words from a visual vocabulary, instead of representing the patch-based descriptors as bag-of-features and classifying them using a standard classifier. The proposed system was tested on 25 distinctive elliptical shapes of vehicle logos with 10 images per class resulting in a total of 250 images obtained from Google Images without any preprocessing. The system offers the advantage of accurate logo recognition of 86.3% in the presence of significant background clutter.

Table 1. Classification rate of testing logo images using BoF and our proposed approach

Vocabulary Size	Manually segmented logo		Automatically detected logo	
	BoF	Ours	BoF	Ours
30	89.4 ± 3.8	89.2 ± 3.3	84.1 ± 3.2	84.2 ± 2.9
40	91.1 ± 3.9	**91.3 ± 3.6**	86.0 ± 3.2	**86.3 ± 2.6**
50	89.8 ± 2.6	90.8 ± 2.9	84.8 ± 2.9	85.5 ± 3.1

Table 2. The computational time saving of our proposed approach

Process	Time
Extracting SIFT descriptors from the image set	t_1
Constructing visual codebooks	t_2
Representing features of image set	t_3
Classifying testing images using a standard classifier	t_4
Classifying testing images using our approach	t

Table 3. The storage saving of our proposed approach

Storage	Dimension
SIFT descriptors of the image set	$P \times 128$
Logo-specific visual vocabularies	$(L \times K) \times 128$
Feature vectors (histograms)	$(M + N) \times (L \times K)$

Car enthusiasts are able to distinguish car models by examining the fine-grained parts. Our method for detecting a vehicle from its background is flexible and can be easily extended to other views such as side- or rear-view by modifying the visual vocabularies. The proposed VLR framework using a CTF strategy could be independently used for part recognition of air-intake grille detection and logo detection. Our noval approach for vehicle logo classification technique not only speeds up the classification process but also reduces the storage requirement when comparing with the traditional bag-of-features approach while maintaining comparable performance. Apart from classifying vehicle logos, several applications such as fine-grained style analysis, part recognition and model analysis could be focused in future research.

A limitation in vehicle logo recognition research is the absence of a comprehensive benchmark dataset. Authors who contributed in this research field have compiled their own datasets to report the superiority of their techniques and those datasets are not publicly available. We believe that the lack of high quality benchmark datasets greatly limits the exploration of the above mentioned applications.

REFERENCES

1. W. Li and L. Li, “A novel approach for vehicle-logo location based on edge detection and morphological filter,” in *Proc. 2nd IEEE Int. Symp. on Electronic Commerce and Security* (Nanchang, 2009), pp. 343–345.
2. Y. Li and S. Li, “A vehicle-logo location approach based on edge detection and projection,” in *Proc. IEEE Int. Conf. on Vehicular Electronics and Safety* (Beijing, 2011), pp. 165–168.
3. S. Mao, M. Ye, X. Li, F. Pang, and J. Zhou, “Rapid vehicle logo region detection based on information theory,” Comput. Electr. Eng. **39** (3), 863–872 (2013).
4. Y. Wang, Z. Liu, and F. Xiao, “A fast coarse-to-fine vehicle logo detection and recognition method,” in *Proc. IEEE Int. Conf. on Robotics and Biomimetics* (Sanya, 2007), pp. 691–696.
5. A. Psyllos, C. Anagnostopoulos, and E. Kayafas, “Vehicle logo recognition using a SIFT-based enhanced matching scheme,” IEEE Trans. Intellig. Transport. Syst. **11** (2), 322–328 (2010).
6. D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis,” IEEE Trans. Pattern Anal. Mach. Intellig. **24**, 603–619 (2002).
7. N. Dalal and M. Triggs, “Histograms of oriented gradients for human detection,” in *Proc. Computer Vision and Pattern Recognition (CVPR)* (San Diego, 2005), pp. 886–893.
8. T. Joachims, B. Scholkopf, C. Burges, and A. Smola, *Making Large-Scale SVM Learning Practical* (MIT Press, Cambridge, MA, 1999).
9. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proc. Conf. on Computer Vision and Pattern Recognition (CVPR)* (Kauai, HI, 2001), pp. 511–518.
10. J. T. Atherton and D. J. Kerbyson, “Size invariant circle detection,” Image Vision Comput. **79**, 795–803 (1999).
11. C. Csurka, R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints,” in *Proc. Workshop on Statistical Learning in Computer Vision (ECCV)* (Prague, 2004), pp. 1–22.
12. A. Ramanan and M. Niranjan, “A review of codebook models in patch-based visual object recognition,” J. Signal Processing Syst. **68** (3), 333–352 (2012).
13. H. Pan and B. Zhang, “An integrative approach to accurate vehicle logo detection,” Comput. Vision Image Understand. **2013**, 12 (2013).
14. J. Hsieh, L. Chen, D. Chen, and S. Cheng, “Vehicle make and model recognition using symmetrical SURF,” in *Proc. 10th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance (AVSS)* (Krakow, 2013), pp. 472–477.
15. H. Bay, A. Ess, T. Tuytelaars, and V. L. Gool, “SURF: speeded up robust features,” Comput. Vision Image Understand. **10** (3), 346–359 (2008).
16. K. Zhou, M. Varadarajan, M. Vincze, and F. Liu, “Hybridization of appearance and symmetry for vehicle-logo localization,” in *Proc. 15th Int. IEEE Conf. in Intelligent Transportation Systems (ITSC)* (Anchorage, 2012), pp. 1396–1401.
17. Y. Ou, H. Zheng, S. Chen, and J. A. Chen, “Vehicle logo recognition based on a weighted spatial pyramid framework,” in *Proc. 17th IEEE Int. Conf. on Intelligent Transportation Systems (ITSC)* (Qingdao, 2014), pp. 1238–1244.
18. D. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vision **60**, 91–110 (2004).
19. S. Yu, S. Zheng, H. Yang, and L. Liang, “Vehicle logo recognition based on bag-of-words,” in *Proc. 10th IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance* (Krakow, 2013), pp. 353–358.
20. D. Llorca, R. Arroyo, and M. Sotelo, “Vehicle logo recognition in traffic images using HOG features and SVM,” in *Proc. 16th Int. IEEE Annu. Conf. on Intelligent Transportation Systems (ITSC)* (Hague, 2013), pp. 2229–2234.
21. Y. Huang, R. Wu, Y. Sun, W. Wang, and X. Ding, “Vehicle logo recognition system based on convolutional neural networks with a pretraining strategy,” IEEE Trans. Intellig. Transport. Syst. **16** (4), 1951–1960 (2015).
22. S. Sotheeswaran and A. Ramanan, “Front-view car detection using vocabulary voting and mean-shift search,” in *Proc. 15th IEEE Int. Conf. on Advances in ICT for Emerging Regions (ICTer’15)* (Colombo, 2015), pp. 16–20.
23. N. Dalal, “Finding People in images and videos,” *Thesis* (Grenoble, 2006).
24. S. Sotheeswaran and A. Ramanan, “A classifier-free codebook-based image classification of vehicle logos,” in *Proc. 9th IEEE Int. Conf. on Industrial and Information Systems (ICIIS’14)* (Gwalior, 2014), pp. 87–91.
25. M. Everingham, L. Van-Gool, C. Williams, J. Winn, and A. Zisserman, “The PASCAL visual object classes challenge: a retrospective,” Int. J. Comput. Vision **111** (1), 98–136 (2015).

Sittampalam Sotheeswaran is a Senior Lecturer at the Department of Mathematics at Eastern University, Sri Lanka. He received his B.Sc. Honours in Computer Science (2008) and MPhil in Computer Science (2016) from the University of Jaffna, Sri Lanka. His research interests are in the field of Image Processing and Machine Learning.

Amirthalingam Ramanan is a Senior Lecturer at the Department of Computer Science at University of Jaffna, Sri Lanka. He received his B.Sc. Honours in Computer Science (2002) from the University of Jaffna, Sri Lanka and his PhD in Computer Science (2010) from the University of Southampton, United Kingdom. His research interests are in the algorithmic and applied aspects of Machine Learning and Computer Vision.