Vehicle Color Recognition on Urban Road by Feature Context

Article *in* IEEE Transactions on Intelligent Transportation Systems · October 2014

DOI: 10.1109/TITS.2014.2308897

CITATIONS
12

READS
503

3 authors:

Pan Chen

3 PUBLICATIONS **25** CITATIONS

SEE PROFILE

Xiang Bai

Huazhong University of Science and Technol…

145 PUBLICATIONS **3,535** CITATIONS

SEE PROFILE

Wenyu Liu

Huazhong University of Science and Technol…

215 PUBLICATIONS **3,596** CITATIONS

SEE PROFILE

Short Papers

Vehicle Color Recognition on Urban Road by Feature Context

Pan Chen, Xiang Bai, *Member, IEEE*, and Wenyu Liu, *Member, IEEE*

***Abstract*—Vehicle information recognition is a key component of intelligent transportation systems. Color plays an important role in vehicle identification. As a vehicle has its inner structure, the main challenge of vehicle color recognition is to select the region of interest (ROI) for recognizing its dominant color. In this paper, we propose a method to implicitly select the ROI for color recognition. Preprocessing is performed to overcome the influence of image quality degradation. Then, the ROI in vehicle images is selected by assigning the subregions with different weights that are learned by a classifier trained on the vehicle images. We train the classifier by linear support vector machine for its efficiency and high precision. The experiments are extensively validated on both images and videos, which are collected on urban roads. The proposed method outperforms other competing color recognition methods.**

***Index Terms*—Color recognition, region of interest (ROI), vehicle.**

I. Introduction

As an important part of intelligent transportation systems (ITS) or Smart City, vehicle information recognition has received much attention in recent years. The information of a vehicle is very helpful for video surveillance and many applications of city public security. Color is one of the most dominant cues for vehicle identification. Vehicle color recognition in natural scenes can provide useful information in vehicle detection [1], [2], vehicle tracking [3] and automatic driving system [4]–[7]. However, it is a challenging task to recognize the colors of vehicles in images/videos due to the following difficulties.

1) Color can be easily influenced by the change of natural environment. For example, haze, snow, and other illumination changes may cause a significant color variation.
2) The performance of color recognition is also limited by the low quality of images/videos, which are affected by noise, overexposure, and color shift in natural images.
3) As different parts of a vehicle have different colors, it is necessary to choose the appropriate regions for vehicle color recognition. The color of the vehicle body is usually more informative and discriminative than other parts, such as windows, wheels, etc. Thus, in this paper, we focus on the vehicle body for recognizing the dominant color. However, determining the focused region of a vehicle is a critical issue.

Before recognizing the colors of vehicles, localizing their positions in images/videos is an essential step. Several well-known detection approaches [8]–[11] can provide an accurate bounding box for each vehicle. Thus, our color recognition is performed in the detected bounding boxes of vehicles. In this paper, both our training and testing images are collected by a vehicle detector.

In object/image color recognition, the color features are collected from image patches. Then, a classifier is trained for color recognition. Wang *et al.* [12] investigated the effectiveness of multiple support vector machine (SVM) recursive feature elimination for feature selection in the classification of lip color. Zheng *et al.* [13] described an automatic low-cost method based on SVM and color histogram for classifying lip colors. Son *et al.* [14] proposed a novel convolution kernel to extract the color information of vehicle images. Park and Kim [15] proposed a method for color classification of objects with two techniques of dimension reduction: 1) projecting a color histogram generated from a 3-D colorspace into 2-D; and 2) converting the color histograms to class-based features by a naive Bayesian classifier. Wang *et al.* [16] adopted the hue saturation value color space for the color recognition of license plates. However, the methods above cannot be applied to the task of vehicle color recognition due to the lack of region-of-interest (ROI) selection.

In the general framework of object/image color recognition, the color feature plays an important role. Qiu *et al.* [17] investigated the redundancy and performance of histogram based color descriptors in the context of automatic color photo categorization. Kender [18] proposed a descriptor based on a new color space named normalized RGB. Van de Sande *et al.* [19] surveyed popular color descriptors and analyze their invariance to illumination change and color shift. Illumination change and color shift are described as a linear model. Huang *et al.* [20] developed a descriptor named color correlogram for the indexing and comparison of images. Correlogram is a table indexed by color pairs. These features only contain the spatial information of different color types, not the object or the scene.

As the spatial information of object and scene can be described by bag-of-word (BoW)-based methods [21]–[23], we adopt the framework of BoW in our method. The BoW is first used in object and scene retrieval by Sivic and Zisserman [24]. The original feature is encoded by a codebook, which is learned from a set of features by clustering algorithm. Lazebnik *et al.* [21] extended the original BoW method with spatial information. Motivated by the classic shape descriptor shape context [25], Wang *et al.* [23] proposed feature context (FC), which divides a image into several fan-shape-like subregions in the log-polar coordinate system. Combined with radial basis coding (RBC) [23] and reference points, FC outperforms spatial pyramid matching (SPM) [21] in scene categorization. Recently, Bolovinou *et al.* [26] has presented a novel approach to encode the spatial configurations of visual words in order to add context information in the representation. The method introduces a bag of spatio–visual words representation obtained by clustering visual words correlogram ensembles and shows a significant improvement over the state-of-the-art BoW model. However, all these BoW-based methods focus on scene recognition. They cannot be directly applied to vehicle color recognition because the local feature describes the texture, rather than the color information of a patch.

Manuscript received August 6, 2013; revised November 29, 2013 and January 15, 2014; accepted February 23, 2014. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 61222308, 61173120, and 61300028; by the Program for New Century Excellent Talents in University in China under Grant NCET-12-0217; by the Fundamental Research Funds for the Central Universities under Grant HUST 2013TS115; and by the National 863 Project under Grant 2012AA011504. The Associate Editor for this paper was U. Nunes. *(Corresponding author: X. Bai.)*

P. Chen and W. Liu are with the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: chenpan.male@gmail.com; liuwy@hust.edu.cn).

X. Bai is with the Department of Electronics and Information Engineering, Huazhong University of Science and Technology (HUST), Wuhan 430074, China, and also with the National Center of Anti-Counterfeiting Technology, HUST, Wuhan 430074, China (e-mail: xbai@hust.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2014.2308897

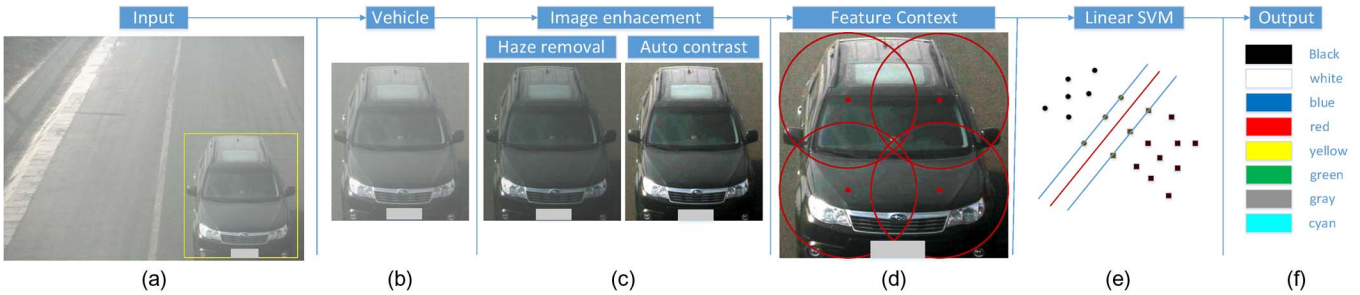

Fig. 1. Framework of our method. (a) Original image. (b) Output of vehicle detection in (a). (c) Results of haze removal and color contrast stretch. (d) Description of image by FC. (e) Training/testing by linear SVM. (f) Output of vehicle color recognition.

Since the vehicle has its specific inner structure and different parts may be in different colors, it requires us to recognize the dominant color of a vehicle. For example, we are more interested in the color of vehicle body than other parts such as wheels or windows. Thus, it is very important to select the region of interest (ROI) of a vehicle for color recognition. We propose a BoW-based method to solve the problem. The BoW representation can map the color features into a higher dimensional subspace by quantizing them using a large codebook, in which different color types are separated more easily by a classifier such as SVM. The flowchart is shown in Fig. 1. The vehicle image is cropped via a detector first. FC, which utilizes fan-like subregions, shows much better performance in scene recognition than that of the SPM method [21], which uses rectangular blocks. In our application, by setting multiple reference points, FC can generate many subregions with the irregular shapes on a vehicle. We train a linear SVM as our classifier.

The contributions of our method are threefold: First, we solve the problem of vehicle color recognition using the BoW model, which has not been applied to this topic yet. The BoW model is very common in object detection, scene recognition, image retrieval, etc. To our knowledge, this is the first study of vehicle color recognition based on BoW. Since the representation of BoW is more discriminative after mapping the original features to a higher dimension using a codebook, the BoW-based color feature achieves the better performance in vehicle color recognition. Second, the ROI is implicitly selected by our method. To select the ROI of a vehicle, we do not explicitly segment it into subregions of different colors. Instead, FC divides the image into fan-like subregions from which the local color features are collected. The ROI can be selected by assigning the weights of a SVM classifier to subregions. Third, we generate a vehicle data set with images and videos. Each vehicle in our data set is labeled with its color, brand, bounding box (for training a vehicle detector), and category. The data set can be helpful for others in the various applications of ITS, such as vehicle detection, vehicle color recognition, and vehicle brand recognition.

The rest of this paper is organized as follows. In Section II, we introduce the details of our method, which includes the preprocessing and representation of images. The experiment results on images and videos are presented and discussed in Section III. Section IV is the conclusion of this paper.

II. Approach

Here, the details of our method are introduced as follows.

A. Preprocessing

The quality of the images/videos taken by the cameras on urban roads is usually poor due to the impact of haze, strong light, and color shift caused by bad weather conditions or inappropriate configuration of equipment. The poor quality is a challenge for color recognition. In order to overcome the influences, we adopt the haze removal method [27] and color contrast method [28] as the preprocessing in our method.

The results of the preprocessing are shown in Fig. 1(b) and (c). In Fig. 1(b), the original image is under thick haze that makes the color of the vehicle biased to gray. After the haze removal, the image is much clearer. The quality of the image can be improved further using a color contrast stretch. We apply haze removal first in the preprocessing, since the color contrast stretch cannot significantly improve the quality of the images under thick haze. The dominant color of the vehicle image is more obvious after the haze removal algorithm and the color contrast stretch method.

B. Image Color Representation by FC

The color features of patches are extracted from the enhanced images. Following the framework of FC [23], the color features are encoded as histograms based on the visual words in a codebook. The encoded features are aggregated into one vector by pooling function to describe a region. The details are introduced as follows.

The patch features are important as they describe local color information, which affects the performance of our method significantly. Van de Sande *et al.* [19] surveyed the current color histograms and analyze their robustness to illumination change and color shift. The histograms include transformed color histogram [19], hue histogram [29], opponent histogram [19], normalized RG histogram [30], and RGB histogram. Their abilities to overcome the illumination change and color shift are different. However, the robustness analysis in [19] is not reliable for our task, since the changes in color shift and the illumination in the natural scene are sophisticated and cannot be described by a simple model in [19]. Thus, in order to keep the merit of each local color feature, we combine transformed color histogram, hue histogram, opponent histogram, normalized RG histogram, and color moment [19] together as our patch feature. The total dimension is 211 ($= 48 + 62 + 48 + 32 + 21$).

In our method, the patches are densely sampled. In a given image I, the features of patches are denoted by $Z = \{z_1, \ldots, z_L\}$. Each patch feature $z \in Z$ is encoded as a vector $C(z) = (w_1^z, \ldots, w_K^z)$. The vector has the same size as the codebook S. The codebook S is learned from the patch features in the training data set by clustering method. We use K-means in this paper. The visual words of the codebook are the centers of K clusters $\{S_1, S_2, \ldots, S_K\}$. Recent coding approaches [22], [31], [32] have been applied in our method. We choose RBC [23] as our coding method. The RBC obtains the state-of-the-art performance in scene recognition. The activation strength is measured with a Gaussian function

$$w_i = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{-\|x-\mu_i\|^2}{2\sigma_i^2}\right), & \mu_i \in \mathrm{NN}(x) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

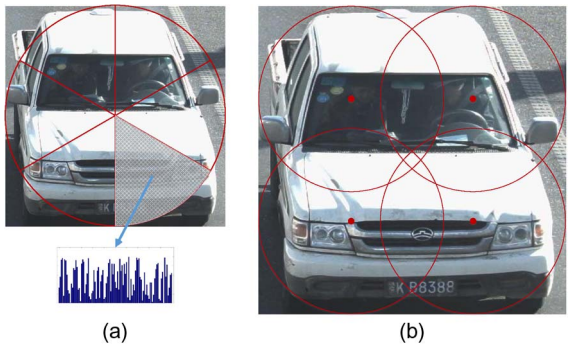

Fig. 2. Illustration of FC. (a) Description of FC of fan-shape-like subregions. (b) Configuration of four reference points.

where μ_i is the center of the cluster S_i, and σ_i is the standard deviation of the distances between S_i and μ_i. $\mathrm{NN}(x)$ denotes a set of nearest neighbors of x. We restrict the activation to $\mathrm{NN}(x)$ for efficiency reasons. The number of nearest neighbors is set to 5 in Section III.

We take the representation of FC [23] to describe the image. FC divides the image into fan-like subregions. The value of each histogram bin of FC is a histogram of the color features. The partitions of subregions of FC are illustrated in Fig. 2. Let p be the location of a reference point in I. The area around p is divided into subregions $Region_r^\theta$ in log-polar coordinate system for $r = 1, \ldots, R$, $\theta = \pi/N, \ldots, \pi$, where R and N are the number of the bins in radius and angle, respectively. Our FC representation for the reference point p is defined as follows:

$$\mathrm{FC}(p, r, \theta, i) = M\left\{w_i^z | (z - p) \in Region_\theta^r\right\} \tag{2}$$

where i indexes the ith codeword $(i = 1, \ldots, K)$. The feature of fan-like subregions are described by the pooling function M, which extracts the most relevant codewords in the region. The function M can be $\max$, sum, mean, or some other functions. We use the $\max$ pooling in our method since it outperforms the others according to [33]. In order to get more precise representation and the spatial information of a vehicle, we set multiple reference points $P = \{p_1, p_2, \ldots, p_L\}$. As an example of the configuration of the reference points in Fig. 2(b), the image is divided into four subregions. Each subregion that belongs to a reference point is described by FC illustrated in Fig. 2(a). Therefore, the feature of an image I is a tensor of $L \times R \times N \times K$ dimension given by

$$\mathrm{FC}(I) = \mathrm{FC}(p, r, \theta, i)_{p,r,\theta,i} \in \mathbb{R}^{L \times R \times N \times K}. \tag{3}$$

Linear SVM is trained for color recognition of vehicle images. Although nonlinear SVM generally outperforms linear SVM, it takes more time to train a model. We chose the implementation of LibLinear [34] to solve the multiclass classification problem.

III. Experimental Evaluation

To fully evaluate the performance of the proposed method, extensive experiments are conducted on the images and the videos collected on urban roads. The proposed algorithm is implemented in C++ on a Windows 7 $\times$ 64-based system. The experiments are carried out on a desktop machine with an Intel(R) Core(TM) i7-2600 K CPU (3.40 GHz) and 8-GB memory. We divide the colors of vehicles into eight classes, including black, white, blue, yellow, green, red, gray, and cyan. Linear SVM is used in our method for color classification.

A. Evaluated Data Sets

Since there is no existing public benchmark for vehicle color recognition, we built two data sets for our experiments: a) a vehicle image data set; and b) a vehicle video data set. Both data sets are collected on urban roads, where the images and videos are taken in the frontal view captured by a high-definition camera with the resolution of 1920 $\times$ 1080. The collected data set is very challenging due to the noise caused by illumination variation, haze, and overexposure.

The image data set contains 15 601 vehicle images of various categories, such as sedan, truck, and bus. The recognition rate is used to evaluate the performance of image based color recognition and defined as the ratio between the correctly predicted numbers and the totals.

The video data set contains ten pieces of surveillance videos. The frame rate of all videos is 10 frames per second, and the average video length is about 5 min. We use an average recognition rate to evaluate the performance in video sequences. A frame is successfully recognized if all the vehicles appearing in the frame are assigned to the correct colors. Only the frames with vehicles are taken into account in the evaluation process.

B. Evaluation on Image Data set

Implementation Details: In our experiments, the data set is randomly divided into two groups, i.e., half for training and half for testing. The procedure is repeated for five times, and the final recognition rate is the average of all runs. All the vehicle images are resized to 300 $\times$ 300 for both training and testing. Our color histogram is a combination of transformed color histogram [19], hue histogram [29], opponent histogram [19], normalized RG histogram [30], and color moment [19] with a dimension of 211 $(= 48 + 62 + 48 + 32 + 21)$. Local features are computed on an image patch with a grid of size 24 $\times$ 24 and a stride of 8 pixels. The codebook size is fixed to 512 in all the experiments. The parameters for FC are $R = 1$ and $N = 6$. The number of the reference points for FC is set as 4 for each vehicle image. Thus, each image is described by a 12 288-dimension feature $(= 512 \times 1 \times 6 \times 4)$.

Performance Evaluation of Our Method and Comparisons: We adopt several conventional color features as the input for the classifier for comparison, including the global features such as color correlogram [20], layered color indexing [35], and the local patch features such as transformed color histogram [19], hue histogram [29], opponent histogram [19], normalized RG histogram [30], and RGB histogram. To demonstrate the advantage of BoW representation over other methods, these features and our color histogram are directly fed to the linear SVM, respectively, and the corresponding results are listed in the second and third rows of Table I. Note that, to provide a fair comparison, in all the experiments we use the linear SVM as the classifier. As shown in Table I, global features work much better than local patch features, including our color histogram. This is reasonable, since our color histogram is a combination of the local patch features without spatial information, whereas color correlogram [20] and layered color indexing [35] are facilitated by the spatial information of different color types.

In addition, when we apply the BoW paradigm to the local patch features, the recognition rates are significantly improved, as shown in the fourth row of Table I. In the BoW paradigm, the local patch features are encoded by the RBC method at first. Then, the encoded local features of an image are aggregated by $\max$ pooling, establishing a holistic representation of the corresponding image. These results demonstrate that BoW representation brings more discriminative power after transforming the original color feature to a feature space of a higher dimension. The best results are obtained based on the BoW

TABLE I
AVERAGE RECOGNITION PERFORMANCES OF DIFFERENT COLOR DESCRIPTORS ON IMAGE DATASET. EACH ROW LISTS THE PERFORMANCE OF CERTAIN COLOR DESCRIPTORS. COLUMNS SHOW A SPECIFIC RECOGNITION RATE OF DIFFERENT COLOR TYPES, AND THE AVERAGE OF THEM IS DISPLAYED IN THE LAST COLUMN

	Method	Black	Blue	Gray	Green	Red	Cyan	White	Yellow	Average
Original Global Features	Color Correlogram [20]	0.8083	0.7422	0.5568	0.6598	0.9475	0.9078	0.8170	0.8694	0.7886
	Layered Color Indexing [35]	0.8937	0.6998	0.6776	0.6058	0.9640	0.9291	0.9009	0.9038	0.8218
Original Local Features	Hue Hist [29]	0.5969	0.5299	0.0086	0.6370	0.9550	0.8414	0.6585	0.8062	0.6292
	Normalized RG Hist [30]	0.5722	0.2129	0	0.5824	0.8581	0.2851	0.7996	0.4428	0.4691
	Opponent Hist [19]	0.7735	0.4955	0.1501	0.5217	0.8910	0.7163	0.8352	0.7412	0.6406
	RGB Hist	0.6446	0.2276	0.1216	0.6788	0.8110	0.7464	0.7819	0.6597	0.5839
	Transformed Color Hist [19]	0.7685	0	0	0.3272	0.7397	0	0.9080	0.0035	0.3437
	Our Color Hist	0.8684	0.7558	0.4700	0.6100	0.9755	0.3121	0.9125	0.8192	0.7154
BoW based Features	Hue Hist [29]	0.6284	0.5549	0.2929	0.6538	0.9440	0.6843	0.7215	0.6600	0.6425
	Normalized RG Hist [30]	0.5161	0.6998	0.1727	0.6393	0.9271	0.7872	0.7508	0.7848	0.6597
	Opponent Hist [19]	0.8193	0.7037	0.5708	0.5667	0.9450	0.8681	0.8738	0.8595	0.7759
	RGB Hist	0.7737	0.7209	0.5450	0.6149	0.9371	0.7623	0.7994	0.8160	0.7462
	Transformed Color Hist [19]	0.7790	0.7269	0.4965	0.5766	0.9245	0.7511	0.8466	0.8125	0.7392
	Our Color Hist	0.8999	0.8475	0.5835	0.6963	0.9683	0.8142	0.8620	0.8089	0.8101
BoW based Features (SPM [21])	Hue Hist [29]	0.8161	0.7504	0.4846	0.6933	0.9667	0.7957	0.8353	0.8290	0.7714
	Normalized RG Hist [30]	0.6874	0.7394	0.3706	0.4946	0.9483	0.7986	0.8103	0.7869	0.7045
	Opponent Hist [19]	0.9419	0.7448	0.7014	0.6197	0.9633	0.8895	0.9232	0.8806	0.8330
	RGB Hist	0.9397	0.7742	0.7267	0.5707	0.9646	0.9139	0.9168	0.8612	0.8335
	Transformed Color Hist [19]	0.9391	0.8323	0.7400	0.5013	0.9638	0.7971	0.9323	0.8609	0.8209
	Our Color Hist	0.9700	0.9297	0.8037	0.7593	0.9870	0.9560	0.9367	0.9141	0.9071
BoW based Features (FC [23])	Hue Hist [29]	0.8654	0.8437	0.6032	0.7630	0.9729	0.9014	0.8550	0.8869	0.8364
	Normalized RG Hist [30]	0.7731	0.8229	0.5896	0.6803	0.9610	0.8766	0.8270	0.8386	0.7962
	Opponent Hist [19]	0.9312	0.8172	0.7856	0.6350	0.9753	0.9518	0.9250	0.8796	0.8626
	RGB Hist	0.9213	0.8414	0.7837	0.6473	0.9601	0.9594	0.9097	0.8646	0.8610
	Transformed Color Hist [19]	0.9206	0.8479	0.7570	0.6461	0.9528	0.8676	0.9192	**0.9634**	0.8518
	Our Color Hist	**0.9714**	**0.9363**	**0.8218**	**0.7859**	**0.9885**	**0.9660**	**0.9415**	0.9395	**0.9189**

TABLE II
AVERAGE PERFORMANCE COMPARISONS USING DIFFERENT STRATEGIES. THE STRATEGIES TAKEN LIE IN DIFFERENT STAGES OF OUR METHOD SUCH AS IMAGE PREPROCESSING, FEATURE POSTPROCESSING, AND THE EMPLOYED CLASSIFIERS

Method	Black	Blue	Gray	Green	Red	Cyan	White	Yellow	Average
Our Method (Without Preprocessing)	**0.9730**	0.9076	0.8198	0.7668	0.9854	0.9589	**0.9423**	0.9010	0.9068
Our Method (PCA)	0.9694	**0.9535**	0.8395	0.7709	0.9874	0.9687	0.9361	**0.9553**	0.9227
Our Method (CAC [36])	0.9718	0.9477	**0.8466**	**0.7884**	**0.9876**	0.9688	0.9392	0.9368	0.9234
Our Method (PCA + CAC)	0.9713	0.9451	0.8461	0.7834	**0.9876**	**0.9787**	0.9414	0.9457	**0.9249**

representation of our color histogram, since our combined feature can keep the merits of different features to overcome the influence caused by overexposure and color shift. In addition, different features can be weighted by learning with SVM.

In addition to the experiments on the BoW representation, we also examine the impact of two different ways of involving spatial information, i.e., SPM and FC. Their results are shown in the fifth and the last row of Table I, respectively. For SPM, the images are partitioned into 1 $(= 1 \times 1)$, 4 $(= 2 \times 2)$, 16 $(= 4 \times 4)$ rectangles, following the configuration in [21]. For each subregion, the encoded local features inside a certain rectangle are aggregated into a vector by $\max$ pooling as the representation of the corresponding region. Then the vector representations of all the subregions are concatenated into one to describe the color of a vehicle. As shown in Table I, both SPM and FC improve the performance of BoW, and FC slightly outperforms SPM. The best result in Table I is obtained by our method when integrating BoW representation of our color histogram with FC. Although color correlogram and layered color indexing also contain the spatial information, they cannot robustly capture the structure of a vehicle.

In our data set, the variances in green and gray are much larger than the others. As these two colors are less unified, their features are more scattered in the feature space and difficult to be classified. Thus, the recognition rates for these two colors are lower than that of the other color types in all the evaluated methods.

When the preprocessing stage is removed in our method, the designed color descriptor still achieves promising results, as shown in Table II. This further demonstrates the robustness of our method against impacts such as color shift and haze. In addition, some other strategies such as principle component analysis (PCA) and coordinate augmented codebook (CAC) [36] can be also employed to enhance the performance of our system. PCA is often used for dimension reduction and helps eliminate the noise embedded in the original feature space. CAC is used to combine the appearance information and coordinate of a patch into a feature vector. Due to the constraint of the location, the patches with the same coordinates are assigned in one cluster center. Thus, the features of the subregions of different locations are encoded by the different codewords. The dimension of the proposed color descriptor is reduced from 211 to 100 by PCA in our practice. In Table II, we observe that both PCA and CAC are able to improve the performance of our descriptor. As the structure of our vehicle images is fixed to some extent, various codewords for different subregions make the features more discriminative. The combination of PCA and CAC leads to the highest accuracy, as shown in the last row of Table II.

To illustrate our method more intuitively, the selected ROI is presented in Fig. 3. More discriminative subregions are attached with a higher opacity (see Fig. 3). The property is determined by the product of the SVM weight and the feature vector of the region. As shown in Fig. 3(a), the regions of window and background are less considered in the color prediction process, whereas the engine hood tends to be more dominant in color decision. For different categories of vehicles, the structures in the frontal view are similar. Our method can perform successful color recognitions of the different categories under mild structure variances.

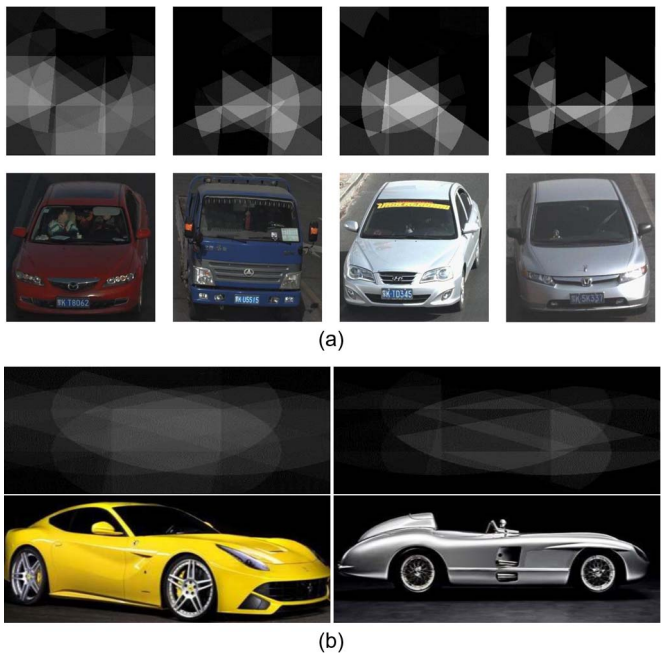

Fig. 3. ROI selected implicitly by our method. The higher opacity means the corresponding subregion is more discriminative. (a) ROI of the frontal view vehicle images. (b) The ROI of the side-view vehicle images.

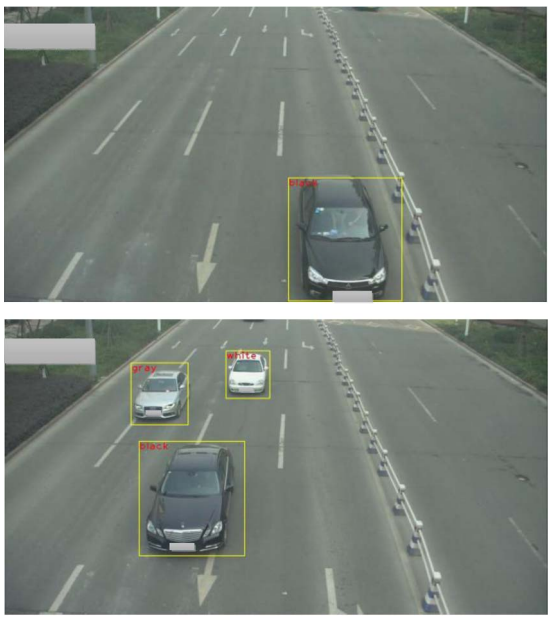

Fig. 4. Some samples of vehicle color recognition on videos.

For side-view vehicle images, the model trained from the frontal view may not make the correct predictions due to the different structure of the ROI. To verify the proposed method on side-view vehicles, we collected 240 side-view vehicle images (30 per class) by Google from the Internet and use half for training and half for testing. With the same experimental settings of the frontal view images, we achieve the recognition rate of 0.8700, which demonstrates the potential of the proposed method on side-view vehicles. Two examples about the ROI selection of side-view vehicle images are shown in Fig. 3(b).

Discussion on System Configurations: Here, we analyze the performance of FC in different system configurations (see Fig. 4). The experiments inspect the influences of the number of partitions in radial and angle, i.e., R and N, the number of reference point Nrp, and the overlapped ratios between rectangular subregions. The efficiency of coding methods, namely, vector quantization (VQ) and RBC, are also presented. The codebook size is set as 512 according to the description in Section III-B1.

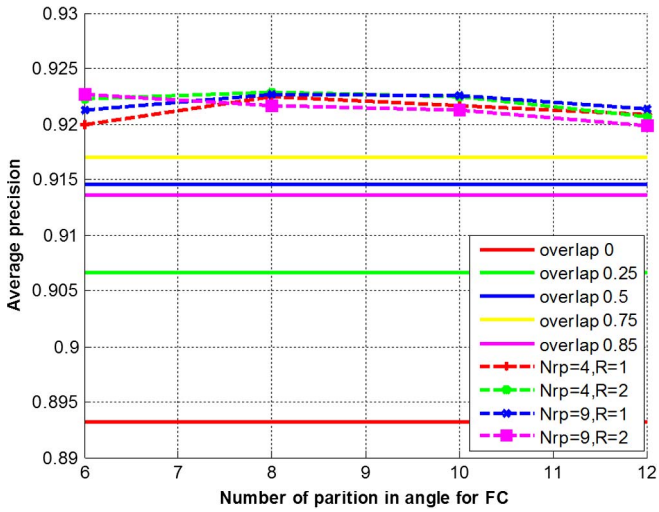

Fig. 5. Average recognition rates of the different settings for our methods. The parameters in the settings contains the number of reference points Nrp, the number of partition in radial R, and the number of partition in angle N. The *overlap* is the ratio between the sizes of the overlapping area and the subregion.

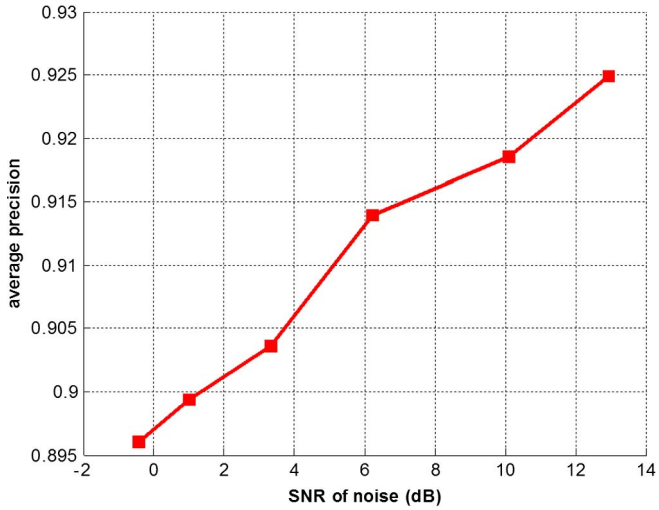

Fig. 6. The average recognition rate of our method under the different levels of noise.

The impacts of parameters R, N, and Nrp are shown in Fig. 5. The number of subregions is the product of R, N, and Nrp (see Fig. 5). Our method achieves a recognition rate of 0.9229, with eight angle partitions, two radial partitions, and four reference points. With the number of subregions increasing, the precision rises at first and then declines. The smaller subregions make the representation of the image too trivial and sensitive to intraclass variance. The larger regions have good generalization property but poor capability in discrimination. Additionally, the length of the feature vector becomes larger when the number of the subregions increases. The mentioned configuration of our method is set to make a tradeoff between the performance and efficiency.

We run the experiments on different subregions overlap ratios of 0, 0.25, 0.5, 0.75, and 0.85. The images are resized into 300 $\times$ 300 and divided into 1 $\times$ 1, 2 $\times$ 2, and 4 $\times$ 4 rectangle subregions in spatial pyramids. The size of subregions in level 1 $\times$ 1 is the same

with the whole image. In the levels of 2×2 and 4×4, the sizes of the subregions are 150, 188, 225, 263, and 278, and 75, 94, 113, 131, and 139, respectively. Fig. 5 shows that the best result achieved by an SPM-based method is still inferior to that of our method. It suggests that the fan-like subregion is more appropriate for vehicle color recognition. The precision increases when the overlap ratio increases from 0 to 0.75. However, the recognition rate decreases when the overlap ratio is higher than 0.75. This can be addressed by the fact that the feature extracted from large partitions is not able to capture the local information, thus losing the discriminative power.

As the number of subregions increases, the image becomes over-partitioned, and the discriminative power of the descriptor decreases. In addition, the size of image representation becomes larger, which increases the computation complexity. In our experiment, the image is partitioned into 1×1, 2×2, 4×4, and 8×8 rectangular subblocks without overlap. The sizes of subregions in four levels are 300, 150, 75, and 38. The proposed setting of subregions achieves a recognition rate of 0.8916.

To test the computational efficiency of the coding step in our method, we use a codebook of size 512 and 1156 local color features. The running time is the average of ten loops. The procedure of VQ and RBC take 0.0271 and 0.0531 s, respectively. The RBC is slower because it involves the operation of applying Gaussian kernel. We show that the step of coding has limited impact on the total color recognition time in the following video experiment.

Robustness Against Noises: In order to verify the robustness of our method, five levels of Gaussian noise are added to the training and test images in the conducted experiments. The average SNR of the noise on the images are 12.94, 10.09, 6.22, 3.33, 1.02, and -0.43 dB, respectively. In this experiment, all the parameter settings are consistent with Section III-B1. The recognition rate increases as SNR rises. The lowest rate is 0.8960 under the noise of SNR -0.43 dB. However, our method still outperforms several methods listed in Table I. This demonstrates our method can overcome the influence of the noise to a certain degree.

C. Evaluation on Video Sequences

We collect 10 081 vehicle images from six pieces of videos as the training samples. The other four pieces of videos are for testing. For a vehicle image, the features are extracted from the images following the settings of Section III-B1. In the video testing, the image patches of the vehicles in the videos are cropped by the detection and tracking system.

The number of the frames in which vehicles of the frontal view are detected is 6547. The average per-frame recognition rate of eight color types is 0.9491. Thus, the performance of our method on video sequences is also encouraging. Fig. 4 shows some samples of the vehicle color recognition results on video sequences. In our experiment, the performance of vehicle color recognition is affected by the vehicle detection results, particularly when the vehicle image is incomplete or contains a large proportion of background. By setting a high threshold in the vehicle detection process, the precision of detection can be increased, thus boosting the performance of color recognition. On average, it costs 0.5 seconds per frame for detection and recognition, suggesting that our method can be applied on videos efficiently.

IV. Conclusion

In this paper, an effective method for vehicle color recognition has been proposed. We demonstrate that the BoW representation of local patch features is powerful to describe object colors. The interesting regions of dominant color of a vehicle can be implicitly selected by assigning the weights to each subregion using a classifier. The extensive experiments on both image and video data demonstrate the potential of the proposed method in real applications. Our future work might be integrating the proposed method and mature image segmentation techniques for localizing the interesting color region more accurately.

Acknowledgment

The authors would like to thank the three anonymous reviewers for their valuable comments. The authors also thank the Third Research Institute, Ministry of Public Security, for nicely providing part of the experimental data.

References

[1] R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 453–462, Jun. 2010.

[2] L. Unzueta, M. Nieto, A. Cortes, J. Barandiaran, O. Otaegui, and P. Sanchez, "Adaptive multicue background subtraction for robust vehicle counting and classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 527–540, Jun. 2012.

[3] T. Xiong and C. Debrunner, "Stochastic car tracking with line- and color-based features," in *Proc. IEEE Trans. Intell. Transp. Syst.*, Oct. 2003, vol. 2, pp. 999–1003.

[4] N. Wu, F. Chu, S. Mammar, and M. Zhou, "Petri net modeling of the cooperation behavior of a driver and a copilot in an advanced driving assistance system," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 977–989, Dec. 2011.

[5] Q. Li, N. Zheng, and H. Cheng, "Springrobot: A prototype autonomous vehicle and its algorithms for lane detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 300–308, Dec. 2004.

[6] Y. Kang, K. Yamaguchi, T. Naito, and Y. Ninomiya, "Multiband image segmentation and object recognition for understanding road scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1423–1433, Dec. 2011.

[7] J. W. Choi, J. Y. Lee, D. W. Kim, G. Soprani, P. Cerri, A. Broggi, and K. Yi, "Environment-detection-and-mapping algorithm for autonomous driving in rural or off-road environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 974–982, Jun. 2012.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2001, pp. 511–518.

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1627–1645, Sep. 2010.

[10] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, San Francisco, CA, USA, Jun. 2010, pp. 2241–2248.

[11] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. Conf. Brit. Mach. Vis. Conf.*, 2009, pp. 91.1–91.11.

[12] J. Wang, X. Li, H. Fan, and F. Li, "Classification of lip color based on multiple SVM-RFE," in *Proc. IEEE Int. Conf. BIBMW*, Nov. 2011, pp. 769–772.

[13] L. Zheng, X. Li, X. Yan, F. Li, X. Zheng, and W. Li, "Lip color classification based on support vector machine and histogram," in *Proc. 3rd Int. CISP*, Oct. 2010, pp. 1883–1886.

[14] J. W. Son, S. B. Park, and K. J. Kim, "A convolution kernel method for color recognition," in *Proc. 6th Int. Conf. Adv. Lang. Process. Web Inf. Technol.*, Aug. 2007, pp. 242–247.

[15] S. M. Park and K. J. Kim, "Color recognition with compact color features," *Int. J. Commun. Syst.*, vol. 25, no. 6, pp. 749–762, Jun. 2012.

[16] F. Wang, L. Man, B. Wang, Y. Xiao, W. Pan, and X. Lu, "Fuzzy-based algorithm for color recognitioncolor recognition of license plates," *Pattern Recognit. Lett.*, vol. 29, no. 7, pp. 1007–1020, May 2008.

[17] G. Qiu, X. Feng, and J. Fang, "Compressing histogram representations for automatic colour photo categorization," *Pattern Recognit.*, vol. 37, no. 11, pp. 2177–2193, Nov. 2004.

[18] J. R. Kender, "Saturation, hue, and normalized color: Calculation, digitization effects, and use," Carnegie-Mellon Univ., Pittsburgh, PA, USA, Tech. Rep., Nov. 1976.

[19] K. E. Van De Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.

[20] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 1997, pp. 762–768.

[21] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2006, vol. 2, pp. 2169–2178.

[22] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 3360–3367.

[23] X. Wang, X. Bai, W. Liu, and L. J. Latecki, "Feature context for image classification and object detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 961–968.

[24] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nice, France, Oct. 2003, pp. 1470–1477.

[25] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[26] A. Bolovinou, I. Pratikakis, and S. Perantonis, "Bag of spatio-visual words for context inference in scene classification," *Pattern Recog.*, vol. 46, no. 3, pp. 1039–1053, Mar. 2013.

[27] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proc. IEEE Conf. CVPR*, 2009, pp. 1956–1963.

[28] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing using MATLAB*. Knoxville, TN, USA: Gatesmark Publishing, 2009.

[29] J. Van De Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 150–156, Jan. 2006.

[30] T. Gevers, J. Van De Weijer, and H. Stokman, "Color feature detection," in *Color Image Processing: Methods and Applications*, R. Lukac and K. N. Plataniotis, Eds. Boca Raton, FL, USA: CRC Press, 2007, pp. 203–226.

[31] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 1794–1801.

[32] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, vol. 22, pp. 2223–2231.

[33] L. Liu, L. Wang, and X. Liu, "In defense of soft-assignment coding," in *Proc. IEEE ICCV*, Nov. 2011, pp. 2486–2493.

[34] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.

[35] G. Qiu and K. Lam, "Spectrally layered color indexing," in *Image and Video Retrieval*. Berlin, Germany: Springer-Verlag, 2002, pp. 100–107.

[36] S. McCann and D. G. Lowe, "Spatially local coding for object recognition," in *Proc. Conf. Asian Conf. Comput. Vis.*, 2013, pp. 204–217.