2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance

Vehicle Logo Recognition Based on Bag-of-Words

Shuyuan Yu[1] Shibao Zheng[1,2] Hua Yang[1,2] Longfei Liang[3]
(1Institution of Image Communication and Information Processing,
Department of EE, Shanghai Jiaotong University, Shanghai, 200240, China)
(2Shanghai Key Laboratory of Digital Media Processing and Transmission, Shanghai, China)
(3Bocom smart network technologies Inc, Shanghai, China)
yu_shuyuan@sina.com {sbzh, hyang}@sjtu.edu.cn lianglf@bocom.cn

Abstract

The recognition of vehicle manufacturer logo is a crucial and very challenging problem, which is still an area with few published effective methods. This paper proposes a new fast and reliable system for Vehicle Logo Recognition (VLR) based on Bag-of-Words (BoW). In our system, vehicle logo images are represented as histograms of visual words and classified by SVM in three steps: firstly, extract dense-SIFT features; secondly, quantize features into visual words by 'Soft-assignment'; thirdly, build histograms of visual words with spatial information. Compared with traditional VLR methods, experiment results show that our proposed system achieves higher recognition accuracy with less processing time. The proposed system is evaluated on a dataset of 840 low-resolution vehicle logo images with about 30×30 pixels, which verifies that our system is practical and effective.

1. Introduction

Vehicle manufacturer logo is a remarkable characteristic of vehicle type. Recognition of vehicle manufacturer logos has many potential applications such as in traffic monitoring and intelligent vehicle management system etc. However it meets great challenges, including poor illumination conditions, background color interference, distortion from viewpoint variation and small logo size etc. Many vision-based intelligent techniques for detecting, tracking and recognizing vehicles in image sequence have been adequately addressed. While up to now, vehicle manufacturer logo recognition is still an area lacking effective methods [1].

Currently, researches on vehicle manufacturer type recognition mainly contain two categories. In the first category, full frontal-view or rear-view vehicle images are used to simultaneously recognize the vehicle manufacturer and model which are treated as one integrated attribute such as VW Passat, VW Beetle and VW Golf etc. Dlagnekov and Belongie [2] used the rear-view vehicle images to recognize car make and model based on SIFT matching but without real-time performance. Conos[3] proposed new SIFT-based descriptors to recognize vehicle model from frontal-view image, which results extensive computational cost. Petrovic and Cootes [4] solved the problem based on edge gradient and match refinement algorithm, but their solution is sensitive to viewpoint change and plane rotation. Since each manufacturer contains large amounts of models and most vehicle models update frequently, this brings high difficulty and heavy computation to the methods in this category. In the second category, solutions mainly focus on the recognition of vehicle logo. Attempts have been made to recognize vehicle logos using Invariant moments [5], while which are either sensitive to noise or suffered from heavy computation. Wang et al. [6] presented a method using template matching and edge orientation histograms with a result of 90% recognition accuracy, but their approach is limited to complicated environments such as shadows and light reflections. Psyllos et al. [1] proposed a SIFT-based enhanced matching scheme, which seems promising but suffers from the illumination variations, wide range of viewpoints, etc. In all these methods, environment sensitivity is the main obstruction for higher recognition accuracy. Besides, most solutions currently are not suitable for real-time applications due to their heavy computation.

In this paper, we propose an improved system based on Bag-of-Words for VLR which not only reaches high recognition accuracy, but also is time-saving. Our main improvements are the following: (a) We use dense-SIFT to extract strong features which are invariant to scale, rotation, distortion and illumination change in a fast speed. (b) We adopt 'Soft-assignment' quantization method to promote the descriptive ability of visual words. (c) We incorporate the spatial information to Bag-of-Words by spatial pyramid. Our proposed system is evaluated on a dataset containing 840 logo images with about 30×30 pixels which are much smaller compared with the 100×100 ones in [1][17]. According to current traffic monitoring situation, it is impossible to segment vehicle logo images with large size from the monitor screen. Images used in our experiment are more close to actual situation in traffic monitoring video, which promote our system more adequate for practical application.

978-1-4799-0703-8/13/$31.00 ©2013 IEEE

Authorized licensed use limited to: FUDAN UNIVERSITY. Downloaded on May 24,2022 at 07:46:56 UTC from IEEE Xplore. Restrictions apply.

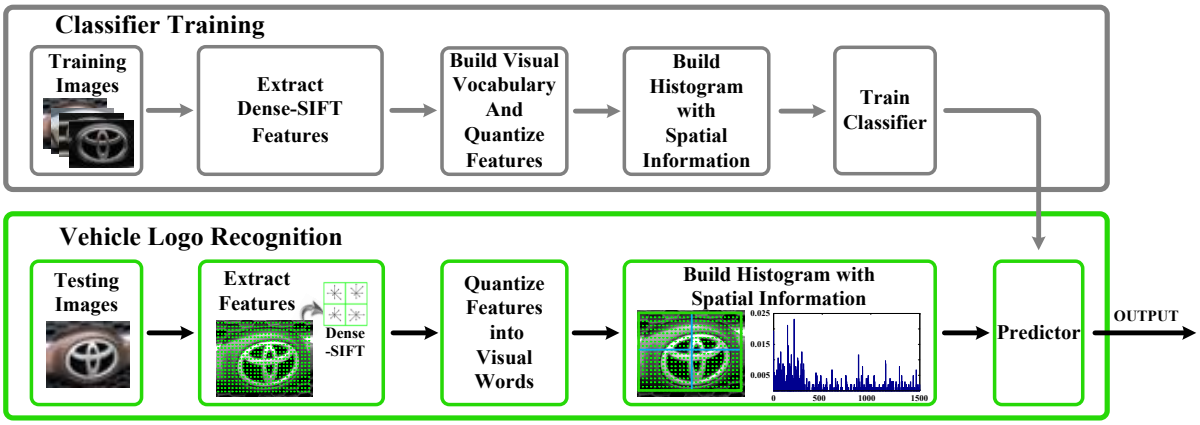


Figure 1: Framework of our vehicle logo recognition system.

The rest of this paper is organized as the following: in section II, the framework of the proposed VLR system is presented. Section III describes the experiments and analyzes the recognition performance. Section IV concludes this paper and refers to future extensions of this research.

2. Framework of Vehicle Logo Recognition System

Currently, the Bag-of-Words method is one of the hotspots in image semantic analysis. For the past few years, systems based on the Bag-of-Words method showed the best performance on several scene classification and large scale content based image retrieval benchmarks, such as Caltech-256 dataset [7] and Pascal VOC challenge [8]. Considering the characteristics of vehicle logos, this paper introduces the outstanding Bag-of-Words method to vehicle logo recognition and makes some promotion to improve the system performance. The framework of our proposed VLR system is illustrated in figure 1. The system mainly consists of the following four steps: (a) Extracting dense-SIFT features from images; (b) Quantizing features into visual words; (c) Building histogram with spatial information; (d) Training classifier for vehicle logo recognition.

2.1. Extracting dense-SIFT features

Due to poorly detected features, recognition of vehicle logo might fail in the cases such as poor illumination conditions, background color interference, viewpoint variation, blot and shelter, etc. In order to increase the recognition accuracy and robustness even in outdoor conditions, we implement dense-SIFT to extract stable local features following the approach of Anna Bosch et al. [9]. In our work, SIFT descriptors of 4×4 pixel patches are computed at points on a regular grid, which are marked as green dots in figure 1 and figure 2. The displacement of each point on the grid to the next is set 5, ie. sampling density.

The idea behind adopting dense-SIFT is that few interest points could be detected from small logo images by SIFT due to lacking abundant information while dense-SIFT does. Additionally, the dense-SIFT features could prevent spurious feature detections from blot and shelter by using the dense regular grid instead of interest points (in SIFT). As a fast dense version of SIFT, the dense-SIFT is also invariant to rotation, scale and illumination difference. Many works [10] [11] have already shown its better work for classification.

2.2. Quantizing features into visual words

After extracting dense-SIFT features, a vehicle logo is represented as large numbers of feature vectors. In the following, we build a more compact representation by quantizing each feature into visual word against a visual vocabulary.

In the offline training phase, a set of local descriptors randomly sampled from our training dataset are used to construct a visual vocabulary by k-means clustering, which partitions the descriptors into k (here $k=300$) disjunct clusters C_i. The visual vocabulary V is shown in equation (1) where μ_i is considered as cluster centers of C_i. Especially, in order to decrease the search time of visual words, we build a KD-Tree[12] to index the visual vocabulary.

(1)

(2)

Where d_j is the dense-SIFT descriptors, $|C_i|$ is the number of descriptors in cluster C_i.

When recognizing a new vehicle logo image in the online phase, dense-SIFT descriptors are first extracted as shown in sec 2.1 and then quantized into visual words against

Authorized licensed use limited to: FUDAN UNIVERSITY. Downloaded on May 24,2022 at 07:46:56 UTC from IEEE Xplore. Restrictions apply.

vocabulary. As for quantizing, we adopt 'Soft-assignment' method from researches [13][14] against the baseline of 'Hard-assignment'. 'Hard-assignment' is that each descriptor is assigned to the closest visual word. A feature vector may be equally close to multiple visual words, but this information cannot be captured in the case of 'Hard-assignment', which makes this method sensitive to image noise and interference(e.g. illumination). However 'Soft-assignment' solves the problem by assigning weights to all visual words depending on the distance between the descriptor and cluster centers. Obviously, "Soft-assignment" could improve the descriptive ability of features to a great extent. We use a kernelized distance between descriptors and visual words in 'Soft-assignment' quantization to construct histograms. The Soft-assignment histogram *(h-soft)* is:

— (3)

Where standing for the image, $|D|$ is the number of descriptors in the image D, and is a vector-valued quantization function that quantizes a local descriptor against the visual vocabulary V:

(4)

K is a Gaussian kernel to measure distance between samples.

——— (5)

2.3. Building histogram with spatial information

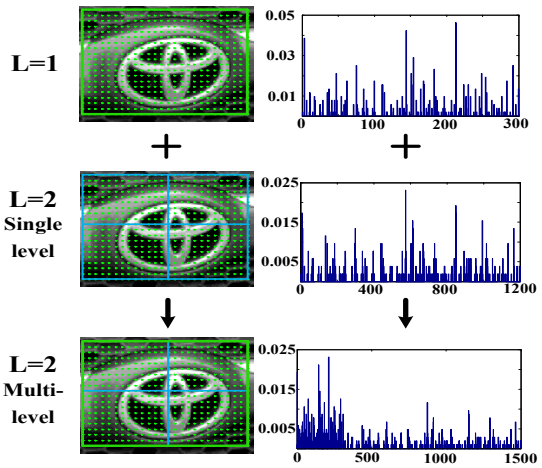


Figure 2: Vehicle logo histogram with spatial pyramid scheme. The green dots are sample points for feature extraction by dense-SIFT

According to the traditional Bag-of-Words, the histogram of vehicle logo image is represented as an orderless collection of visual words. The notorious disadvantage of BoW is that it ignores the spatial relationships among the features, which is very important in image representation. Thus we apply spatial pyramid scheme [15] into our system. The basic idea to form spatial pyramid is partitioning the image into increasingly fine sub-regions, computing histograms of local features inside each sub-region and concatenate all histograms. In spatial pyramid, grid at level l has 2^{l-1}cells along each direction. When l =1, i.e. without geometry information, the image is represented by k visual words. For an image with l level subdividing, the highest-level pyramid descriptor of the image is a histogram with dimensionality $k \times 4^{l-1}$. Consequently, the multi-level pyramid descriptor is a vector with dimensionality :

– (6)

Our experiments use the setting of k =300 and multi level of L=[1 2] pyramid, resulting in a 1500-dimensional histogram. The representation is illustrated in Fig. 2

2.4. Vehicle logo classifier

We apply multi-class support vector machines (SVM), a popular and effective supervised learning technique, to our vehicle logo classifier. For comparation, we also employ standard nearest-neighbor (KNN) as a baseline, which is fast, easy to implement and does not require an explicit training step. Histograms of visual words described in Sec.2.3 serve as the inputs to the classification training process.

Given a histogram h, the algorithm KNN will find its k nearest neighbors in the feature space of training dataset based on a similarity measure (e.g. distance functions). We classify h as the category which owns the majority of k nearest neighbors. For the optimized parameters, we perform a grid search using $k \in \{1, ..5\}$ and distance measures $d \in$ {code book, cosine}.

For each class of vehicle logo, one model is trained by the multi-class SVM classifier. When training the model for a specific class, the vehicle logos in this class are marked as positive, and the remaining negative. We train 14 models to distinguish 14 vehicle logo classes. Each model is able to discern its category from the union of the remaining (one-vs-all approach). When recognized, the input logo image is classified according to the max classification predict value. The classification predict value $P^i(h)$ for class i is

(7)

(8)

Where weight w_j, support vectors s_j and bias β are generated during SVM training phrase, h is the histogram

Authorized licensed use limited to: FUDAN UNIVERSITY. Downloaded on May 24,2022 at 07:46:56 UTC from IEEE Xplore. Restrictions apply.

and $K(\cdot,\cdot)$ is SVM RBF kernel.

There are two key parameters for an RBF kernel: C (penalty parameter) and γ (kernel parameter). To identify good (C, γ) so that the classier can accurately predict unknown data, we adopt "Cross-validation and grid-search" and scan various pairs of (C, γ) values, where $C \in \{2^{-5} \ldots 215\}$ and $\gamma \in \{2^{-24} \ldots 22\}$. Table 1 lists the model parameters in best performance, by using which we generate our best classifiers in the following vehicle logo classification experiments.

	KNN		SVM	
	k	d	C	
Hard	5	cosine	512	0.0078
Soft	4	11	32	2

Table 1 Best parameters for KNN/SVM model using histograms generated with soft/hard quantization

3. Experimental results and analysis

3.1. Dataset

Our system is evaluated on a dataset of 840 segmented vehicle logo images. The dataset is composed of 14 classes and each class contains 60 vehicle logo images. The major sources of our dataset include images captured by ourselves, Medialab LPR Dataset [16], web resource from Google image search and actual road surveillance video. Images in our dataset are captured under various conditions such as: dusk, dawn, cloudy & sunny day and from various viewpoints. Figure 3 illustrates some images in our dataset.

We rescale all the vehicle logo images to a fixed height of 30 pixels. Given an image with $W \times H$, the size of the

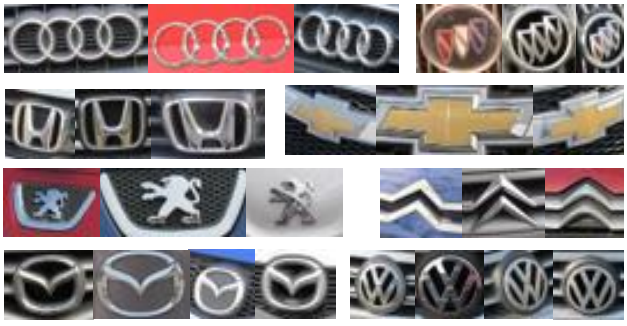


Figure 3: Some images in our vehicle logo dataset.

resized image thus is — , here . Compared with the images of 100×100 pixels in the dataset of Psyllos' work [1], ours are with much smaller size. It should be noted that smaller images meet greater challenge in recognition due to the lack of discriminative features. Images in our dataset are much similar to the actual monitoring situation which proves our system to be more practical.

3.2. Experiment

In this section, we present and discuss the experimental results of the vehicle logo recognition. In our experiments, a part of vehicle logos in dataset are randomly chosen as a training subset, while the rest are evaluated as testing images in recognition. To get a relative stable result, average 6-fold-cross-validation accuracy is measured. The accuracy we used is given in the following.

———————— (9)

Where TP is the number of correct classified vehicle logo images.

We compare our proposed solution with the traditional Bag-of-Words method, mainly focusing on feature extraction, quantization methods and spatial pyramid etc. Thus five methods for contrasts are evaluated. 'DSS' as our proposed system: using dense-SIFT to extract features, 'Soft-assignment' for quantization and SVM classifier; 'SSS': using SIFT descriptor, 'Soft-assignment' and SVM; 'DHS': using dense-SIFT, 'Hard-assignment' and SVM; 'SHS': using SIFT descriptor, 'Hard-assignment' and SVM; 'DSK': using dense-SIFT, 'Soft-assignment' and KNN. In our experiments, the spatial bin of dense-SIFT is set to 4 and sampling density to 5. The vocabulary size is 300.

In order to evaluate the performance of our system more specifically, we split the whole dataset into increasing larger sub-datasets (6, 12,...,60 logo images per category). For each sub-dataset, training and testing images are chosen randomly and the average 6-fold-cross-validation accuracy of the different methods are measured. Accuracy results are shown in figure 4.

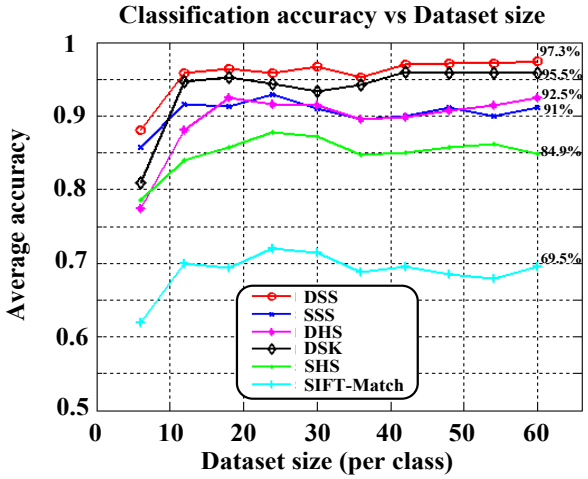


Figure 4: Classification accuracy of vehicle logo images. SIFT-Match is SIFT matching method [1] [2]. All results are measured at a multi-level spatial pyramid of L=[1, 2]

Authorized licensed use limited to: FUDAN UNIVERSITY. Downloaded on May 24,2022 at 07:46:56 UTC from IEEE Xplore. Restrictions apply.

We could infer the performance of dense-SIFT and SIFT on feature extraction from DSS vs. SSS and DHS vs. SHS shown in figure 4. DSS has the highest accuracy about 97.3% and reaches a stable performance of about 95% accuracy even when the sub-datasets size is only 18 images per category. The recognition accuracy of SSS is 3% ~12% lower than that of DSS. Similar results are shown in the performance of DHS vs. SHS. As expected in sec 2.1, dense-SIFT outperforms SIFT for dense-SIFT extracting more stable features by dense grid sampling.

As for quantization methods, both 'Soft-assignment' and 'Hard-assignment' are evaluated, whose results are illustrated as DSS and DHS, SSS and SHS in figure 4. The average recognition accuracies of DHS and SHS are all about 5% lower than that of DSS and SSS respectively, which verifies the better performance when using 'Soft-assignment' for quantization. This mainly because that 'Soft-assignment' expresses features more accurately by using all visual words with different weights.

The classification results of DSS and DSK indicate that multi-class SVM used in our system outperforms KNN which serves as a baseline in this paper.

To analyze the influence of spatial information, we evaluate the BoW methods with different pyramid settings. As shown in table 2, results improve as we go from L=1 to a higher level setup. Almost all methods achieve the highest recognition accuracy at spatial pyramid in multi-level. The maximum growth from L=1 to L=[1,2] is about 5%, and the minimum growth is larger than 1%. Although the growth is not obvious enough, the improvement is much difficult and valuable since our recognition accuracy is quite high already. In our experiments shown in figure 4, we choose the multi-level of L=[1,2] pyramid which achieves the best performance as shown in Table2.

L	DSS	SSS	DHS	SHS	DSK
1	0.960	0.894	0.910	0.809	0.943
2	0.967	0.913	0.915	0.854	0.950
[1,2]	**0.973**	0.911	0.925	0.849	0.955

Table 2 Influence of spatial information on classification accuracy. The highest results are shown in bold.

We also evaluate the SIFT matching method [1] [2] which reaches an accuracy of highest 72.0% shown in figure 4. While based on the method of SIFT matching, Ref.[2] and Ref.[1] achieved an accuracy of 89.5% and 91% on their dataset respectively. This mainly because that small logo images in our dataset meet greater difficulty in recognition due to the low resolution.

As the analyses, our proposed system (DSS) obviously outperforms all the methods. The system acquires the best performance with recognition accuracy of 97.3% at the subset 60 images per class which also exceeds most VLR systems previously published in [1] [5] [17] [18] as listed in table 3.

Method	Tche-bichef SVM[5]	MFM [1]	Merge-SIFT [17]	Ada-boost [18]	DSS Proposed
Accuracy	100%	94%	94.6%	92%	97.3%
Time	1.0 s	1.02s	0.8s	2.2s	0.022s
Dataset	200	400	1500	200	840
Classes	unknown	10	10	10	14

Table 3 Comparation with other VLR methods. Noted that the results listed in this table is from the published paper [1] [5] [17] [18] based on their own dataset which is not access to us, and the time in [18] includes detection phrase. DSS is the proposed system evaluated on our dataset.

3.3. Computational time

In this paper, experiments are performed with a Matlab implementation on a 2GHz Pentium R (Dual CPU-T2410). We train the classifier in the offline phase and recognize vehicle logo images on-line. For our system, the average recognition speed is about 0.0226s per image which is much faster than many other researches [1, 5,17,18] whose recognition time is about 1s per image as shown in table 3. We have contributed to the time-saving system in four aspects: (a) using BoW to prevent from heavy computation. (b) using dense-SIFT to assure a fast feature extraction. (c) building KD-tree to accelerate visual-words searching. (d) low-resolution images with less computation.

4. Conclusions

This paper has presented a new reliable system for vehicle logo recognition based on Bag-of-Words. Specifically, we use dense-SIFT to extract stable features, quantize features by 'Soft assignment', and compute histogram with spatial information to improve our performance. Our vehicle logo dataset has 14 classes and 840 images with small size about 30×30 pixels. Our approach is demonstrated with promising result, yielding 97.3% recognition accuracy rate. Besides the recognition speed is rather fast, about 0.0226s per image. From the above, the approach proposed in our paper is obviously more adequate in practical application such as traffic monitoring etc. To further improve the practicality of vehicle intelligent recognition system in video stream, we will do more research on the combination of vehicle detection, logo detection and recognition in the future.

5. Acknowledgement

This research is supported by NSFC (No.61102099, No. 61171172), STCSM (No.10231204002, No.11231203102), National Key Technology R&D Program (No. 2011BAK14B02) and The High Technology Research and

Authorized licensed use limited to: FUDAN UNIVERSITY. Downloaded on May 24,2022 at 07:46:56 UTC from IEEE Xplore. Restrictions apply.

Development Program of China (2011AA01A107, 2012AA011702)

References

[1] Psyllos A P, Anagnostopoulos C N E and Kayafas E. Vehicle logo recognition using a SIFT-based enhanced matching scheme. Intelligent Transportation Systems, IEEE Transactions on, 2010, 11(2): 322-328.

[2] Dlagnekov L and Belongie S. Recognizing cars. University of California San Diego, Tech. Rep. CS2005-0833, 2005.

[3] M.˘Conos. Recognition of vehicle make from a frontal view. M.S. thesis, Fac. Elect. Eng., Czech Tech. Univ., Prague, Czech Republic, 2006.

[4] Petrovic V S and Cootes T F. Analysis of features for rigid structure vehicle type recognition. British Machine Vision Conference. 2004, 2: 587-596.

[5] Dai S, Huang H, Gao Z, et al. Vehicle-logo recognition method based on Tchebichef moment invariants and SVM. Software Engineering, 2009. WCSE'09. WRI World Congress on. IEEE, 2009, 3: 18-21.

[6] Yunqiong W, Zhifang L and Fei X. A fast coarse-to-fine vehicle logo detection and recognition method. Robotics and Biomimetics, 2007. ROBIO 2007. IEEE International Conference on. IEEE, 2007: 691-696.

[7] G. Griffin, A. Holub, and P. Perona. Caltech 256 object category dataset. Technical Report UCB/CSD-04-1366, California Institute of Technology, 2007.

[8] M. Everingham, L. V. Gool, C. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge 2008 (voc2008). In ECCV Workshop, 2008.

[9] Bosch, Anna, Andrew Zisserman and Xavier Munoz. Scene classification via pLSA. Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006:517-530.

[10] Bosch A, Zisserman A and Muoz X. Image classification using random forests and ferns. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007: 1-8.

[11] Uijlings J R R, Smeulders A W M and Scha R J H. Real-time bag of words, approximately. Proceedings of the ACM international Conference on Image and Video Retrieval. ACM, 2009: 6.

[12] J. L. Bentley. Multidimensional binary search trees used for associative searching. Communications of the ACM, 18(9):509-517, 1975

[13] Eitz M, Hays J and Alexa M. How do humans sketch objects?. ACM Transactions on Graphics (TOG), 2012, 31(4): 44.

[14] Fares Alnajar, Caifeng Shan, Theo Gevers, and Jan-Mark Geusebroek. Learning-based Encoding with Soft Assignment for Age Estimation under Unconstrained Imaging Conditions. Image and Vision Computing, Vol. 30, Issue 12, December 2012, Pages 946-953.

[15] Lazebnik S, Schmid C and Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, 2: 2169-2178.

[16] Medialab LPR dataset, March. 2013[online]. Available: http://www.medialab.ntua.gr/research/LPRdataset.html

[17] Psyllos A, Anagnostopoulos C N and Kayafas E. M-SIFT: A new method for Vehicle Logo Recognition. IEEE International Conference on Vehicular Electronics and Safety (ICVES), 2012:261-266

[18] Sam, Kam-Tong, and Xiao-Lin Tian. Vehicle Logo Recognition Using Modest AdaBoost and Radial Tchebichef Moments. International Proceedings of Computer Science & Information Tech, 2012, 25: 91.

Authorized licensed use limited to: FUDAN UNIVERSITY. Downloaded on May 24,2022 at 07:46:56 UTC from IEEE Xplore. Restrictions apply.