

Vehicle Logo Detection and Classification using Discriminative Pixel-patches Sparse Coding

Yi Ouyang^{1, a*}

¹College of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, China

^{a*}oyy@mail.zjgsu.edu.cn

Keywords: Linear regression, Sparse coding, Classification, SVM.

Abstract. Sparse coding has been successfully applied to many numerical object detection and classification problems in recent years. However, sparse coding has rather slow convergence rates on some optimization problems. In this article, an online framework for fast and classification of different vehicles logo in a natural cluttered background is proposed. Our basic idea is that the better a detection is for vehicle logo by two stages logo detection. Two techniques are applied to achieve it: making Pixel-patches Sparse Coding (PSC) to help suppress noise and enhance translation invariance of sparse representation. PSC describes the local appearance and encodes the each component relative position. The process starts with license plates localization using linear 3-channel pixels regression, followed by vehicle logo classification. This is achieved first computing the structural sparse coding and then looking for highly structured areas which usually correspond to the license plate and log areas. Detected areas are then fed to the classifier operating in the multi-class online structural SVM. The proposed detection and classification algorithm runs in real-time and performs favorably against state-of-the-art algorithms on challenging sequences in terms of efficiency, accuracy and robustness.

Introduction

Vehicles identification is a hot research field in the development of intelligent transport systems (ITS) involving the detection and classification of urban road vehicles. Vision-based vehicle classification system uses low cost camera to capture images of the scene of a vehicle. Therefore, it plays an important role in many applications like traffic congestion avoidance and accidents prevention. Despite the large amount of literature on vehicle detection and tracking, there has been very little work done in the field of vehicle logo detection. This is because vehicle logo detection is an inherently hard problem. In real-world occlusions, illumination, shadows and camera noise are facts of life.

Recently, many methods of vehicle detection have been proposed[1] [2] [3]. In Ma and Grimson [2], the keypoint detection algorithm is adopted and modified to fit in vehicle classification task and return a good result. Nowak and Jurie [3] proposed a high-level data transformation algorithm and a feature selection framework for categorization of different types of vehicles, and adapted to hierarchical SVM classifiers to improve the performance of part-based vehicle models. Chen et al. [4] proposed to segmenting moving road vehicles from the color video data supplied by a stationary roadside camera and classifying those vehicles in terms of type (car, van and Heavy Goods Vehicle) and dominant color using the SVM. However, the existing solutions lack efficacy or are not suitable for reliable operations with poor quality images and the classification relies on motion silhouettes (binary mask) which is noisy and can be affected by camera shake, shadows, object occlusions and so on. Sparse Coding This method is shown in experiment to converge almost surly to a stationary point of the objective function[5]. This paper makes two main contributions: (1)We propose a coarse-to-fine location algorithm for vehicle logo location, which can be composed by coarse-location and fine-location two stages. In coarse-location step, we implemented license plate location using nonlinear 3-channel pixels regression, and assumed the vehicle logo coarse position is above the license plate. (2)We propose a multi-class structural SVM with latent variables for vehicle logo classification. We present a novel algorithm for adapting dictionaries so as to represent vehicle images sparsely.

Fast Vehicle License Plate location using linear 3-channel pixels regression

Locating the vehicle license plate plays an important role in the vehicle license plate automatic recognition system. A novel locating approach based on the 3-channel pixels regression is presented in this paper. Firstly, the license plate color regression model are learned in a color car image; secondly, a line regression model is made for every license plate pixel, and its magnitude is perpendicular to the liner relationship and its center is located on the edge pixel, and then the color pattern of the pixels on both sides of the edge point in the window is investigated and the centric edge point of the window is reserved when the color pattern matches the combination of the background color and text color of the plates. A 3-channel pixels regression filter is applied to the candidate binary image for removing the regions without the structure feature of the license plate. Finally, the license plate is extracted correctly from the candidate regions by analyzing the texture feature of the plate. Consider the linear regression model, we assume the Blue-channel is linear with red and green channel:

$$Y = XW + \epsilon \quad (1)$$

where X is a vector of RG-channel color value, Y is $n \times 1$ vector of Blue color value, ϵ is identically independent distributed error vector, n is the number observations. After the least squares estimates of the parameters W have been computed, the predicted value of the response variable can be written in the form as follows:

$$\hat{y} = XW \quad (2)$$

We use \hat{W} to denote the least squares estimate of W . The formula as:

$$\hat{W} = (X^T X)^{-1} X^T Y \quad (3)$$

The elements of W are shown by $w_{i,j}$. It can be seen from (2) that the influence of the response values on the prediction, depends on the values of $w_{i,j}$. The coarse vehicle logo region results of color of license plate of vehicle can be seen in Figure 1.



Figure 1 Coarse vehicle logo region (The red rectangle of second row denote the coarse logo region by vehicle license plate location)

The proposed method focuses on matching background color and character color in a license plate and combines its structure feature and texture feature. The experiments on 163 car images that were taken under various conditions show the extraction rate of 98.2%. Integrating color edge detection, edge color pair, mathematical morphology and neural network into our method, the approach offers robustness when dealing with noisy car images, car images in variant lighting conditions and car images with skew number plate.

Learning Discriminative Pixel-patches Sparse Coding Features

Because of its advantage and promising performance for many computer vision applications, sparse coding has been received growing attention. Researchers have developed a lot of algorithms to make sparse coding process more efficient [6]. The coefficient X can be computed by optimizing the L_1 regularized least square problem, which typically provides a sparse solution [7]. However, linear representations might not be adequate for representing the nonlinear data. By kernel trick, the data can be mapped into a high dimensional feature space. Thus, the kernel sparse coding method which extends sparse coding into high dimensional space has been studied by many researchers [8],[9]. As the vehicle

logo image exist remarkably different on difference image scene, neither pixel template detector nor the HOG template from the image is able to capture its variations. To build an effective prior for classification, we focus on describing the deformable vehicle logo combined features set S , which is composed of the Pixel sparse coding features from different vehicle logo images.

We describe the pixel patches template learned from the sampled sets. In contrast to conventional sparse coding approaches [10] that use only positive or negative template to represent objects, we incorporate Pixel-level and Gradient-level information into the visual dictionary to acquire extra robustness. Object Pixel-patches Sparse Coding (PSC) is constructed by standard vehicle logo object directly. PSC describes the local appearance and encodes the each component relative position, so they can handle local appearance variation effectively. Given the hypersphere's center $w = \phi(y_\theta)$, the descriptive loss term is:

$$L_P(w) = \min_R R^2 + \lambda \sum_i [|\phi_P(Y_\Theta) - w|^2 - R^2]_+ \quad (4)$$

where R is the radius of the hypersphere, and $[z]_+ = \max(0, z)$ is the hinge loss. Let $\phi(\cdot)$ denote the be constrained to have a constant form τ , i.e., $|\phi(y_\theta)|^2 = \tau$.

Let $\xi^2 = \tau - R^2 + w^2$, then $\xi^2 \geq \tau + w^2$. Thus $L_P(w)$ can be rewritten as:

$$L_P(w) = \min_\xi w^2 + \tau - \xi^2 + \lambda \sum_i [\xi^2 - 2w\phi]_+ \quad (5)$$

$L_P(w)$ is related to 1-class SVM. The descriptive component captures the global density and modality of the samples in S , hence it can be less affected by the influence of outlier target samples but capture the major characteristics of target samples. The learning of descriptive component needs low learning cost and less prior knowledge, which can adapt to the complex changes of the object and is suitable for model-free object tracking..

Multi-Class Structural Linear SVM with Latent Variables for Vehicle Logo Classification

The standard structural SVM problem assumes that the structure of h is given and fixed. We solve the following convex optimization problem for training structural SVMs:

$$\min_w \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N [\max_{y,h} (w \cdot \Phi(x_i, y, h) + L(y_i, y, h)) - \max_h (w \cdot \Phi(x_i, y_i, h))] \quad (6)$$

where C is a fixed penalty factors, $\Phi(x_i, y, h)$ is feature mapping function, and $L_i(y) = L(y_i, y, h)$ is the loss function. For object detection problem, $L_i(y) = 1$ if $y_i = y$, otherwise 0. This optimization problem is non-convex. In [11] use a local optimum method CCCP concave-convex procedure to solve this problem. In our Experiment, we follow [11] but modify CCCP to an online incremental CCCP. The feature vector is defined as follows:

$$\Phi(x, y, h) = \begin{cases} (\Phi_A(x_i, h), \Phi_S(h)) & \text{if } y = +1 \\ \Phi_S(x_i) & \text{if } y = -1 \end{cases} \quad (7)$$

The $\Phi_A(x, h)$ is a joint feature vector that describes the relationship between input x and unobserved latent variables $h \in H$, with w being the parameter vector. denote the sparse coding appearance features which contain PSC for all nodes. More precisely, the appearance features is composed by PSC descriptors. The $\Phi_S(h)$ is spatial features which encode the parent-child pairwise spatial relationship. Let $(\Delta u, \Delta v)$ denote the displacement of node b relative to its reference position which is determined by the position of the parent node. $\Phi_{BG}(x_i)$ is the feature count for the background model. For the positive patches, we randomly initialized the object position h uniformly over the whole patches of the image. In this paper, we pursue is to learn a multiclass classifier function $F : X \rightarrow Z$ that maps an instance x to an element z of Z . Given the training data $S = \{(x_1, z_1), \dots, (x_m, z_m)\}$, we assume that each instance x_i is drawn from a domain $X = \{x_i \in \mathbb{R}^n\}$ and that each label $z_i \in Z = \{1, \dots, L\}$ is an integer from the label set. Instead of training a binary classifier over $\phi(X) = [\phi_\Theta, \phi_H]$, we learn a score function $F : X \times Y \mapsto \mathbb{R}$ that measures the compatibility between (ϕ_H, ϕ_Θ) pairs. Considering

its efficiency at predictive stage, F is a linear model and can yield higher scores to those more similar to the target. A linear SVM aims to learn L linear functions $\{w_c^T \phi(X) | c \in Z\}$, for a test data x , its class labels is predicted by .

$$z = \max_{c \in Z} w_c^T \phi(X) \quad (8)$$

In our algorithm, a local PSC sparse representation is used to model the appearance of standard templates of vehicle logo, and the sparse coding represents the basis distribution of the object. For any image patch represented by a set of descriptors, we can compute a single feature vector based on some statistics of the descriptors' codes. We take a one-to many strategy to train L binary linear SVMs, each solving the following unconstrained convex optimization problem

$$\min_{w_c} \{J(w_c) = \|w_c \odot s_p\|_2 + C \sum_{i=1}^n \ell(w_c; y_i, f_i)\} \quad (9)$$

w is the parameter representing the appearance. It is collaboratively learned from two components through a single objective function:

$$\min_w \|w\|_2^2 + \beta \cdot L_P(w) + (1 - \beta) \cdot L_H(w) \quad (10)$$

where L_P, L_H represent the loss terms on the PSC component and LHS component, $\|w\|_2^2$ is the regularized term, $\beta \in [0, 1]$ is scalar parameter to trade-off the impact between the two components. $\ell(w_c; y_i^c, f_i)$ is a hinge loss function. L is the 2^{nd} differential operator and $\|L\|^2$ penalizes the curvature of the approximation object function.

Experiments and results

We evaluate the proposed tracking algorithm with our method ,using SVM-based on pixel-patches and SVM-based HOG features. The performances of logo recognition are measured by using accuracy (recognition rate). In this experiment, there are 21 class labels (20 class labels for manufacturer brand and one class label for unknown class) for classification of 32×32 windows of images in 50 and 100 test images. The accuracy is calculated by the number of image windows that are recognized correctly and the number of all image windows in 50 or 100 test images. The result in Table-1 demonstrates that the proposed scheme can provide the higher accuracy in comparison with SVM-based on pixel-patches and SVM-based HOG features.

Table-1 Vehicle Logo Classification Accuracy

Vehicle Logo Type	Training Samples #	Test Samples #	Our Method%	SVM-based on pixel-patches %	SVM –based on HOG%
FORD	100	50	81	78	83
NISSAN	100	50	83	75	81
LEXUS	200	50	88	72	83
MAZDA	200	50	90	85	85
BENZ	200	100	87	80	75
HONDA	200	100	92	87	92
SUZUKI	200	100	82	81	80
AUDI	200	100	85	80	83

We have considered the problem of vehicle logo recognition for sparse coding approach and the trade-off between the accuracy and the number of sparse coding used in the model which will directly affect the computational speed. We proposed a feature selection scheme adapted to hierarchical classifiers, which outperforms feature selection strategies applied to the structural SVM classifier. We achieved good results both for infrared vehicle categorization problem and for more generic visible object categorization problem. Our results show that our approach can improve both the robustness and accuracy of object classification.

Acknowledgments

This research was supported by the grand Project of the Science and Technology Department of Zhejiang Province in China No. (2015C33071) , Zhejiang Province Intelligent Transport Systems Center of Engineering and Technology under Grant No. (2015ERCITZJ- KF1) and Zhejiang Provincial Natural Science Foundation of China No.(LY14F020004) The authors would like to thanks the anonymous reviewers for their constructive and useful comments, which helped in improving the presentation of our work.

References

- [1]. Chen, Li-Chih, Jun-Wei Hsieh, Yilin Yan, and Duan-Yu Chen, *Vehicle make and model recognition using sparse representation and symmetrical SURFs*. Pattern Recognition, 2015. **48**(6): p. 1979-1998.
- [2]. Xiaoxu, Ma and W. E. L. Grimson. *Edge-based rich representation for vehicle classification*. in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*. 2005. p. 1185-1192 Vol. 2.
- [3]. Nowak, E. and F. Jurie. *Vehicle categorization: parts for speed and accuracy*. in *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. 2005. p. 277-283.
- [4]. Chen, L. C., J. W. Hsieh, Y. Yan, and D. Y. Chen. *Vehicle make and model recognition using sparse representation and symmetrical SURFs*. in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. 2013. p. 1143-1148.
- [5]. Mairal, Julien, Francis Bach, Jean Ponce, and Guillermo Sapiro, *Online dictionary learning for sparse coding*, in *Proceedings of the 26th Annual International Conference on Machine Learning*. 2009, ACM: Montreal, Quebec, Canada. p. 689-696.
- [6]. Agarwal, S., A. Awan, and D. Roth, *Learning to detect objects in images via a sparse, part-based representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004. **26**(11): p. 1475-1490.
- [7]. Mei, Xue and Haibin Ling. *Robust Visual Tracking using $l(1)$ Minimization*. in *IEEE International Conference on Computer Vision*. 2009. p. 1436-1443.
- [8]. Aharon, M., M. Elad, and A. Bruckstein, *K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation*. IEEE Transactions on Signal Processing, 2006. **54**(11): p. 4311-4322.
- [9]. Jiang, Z., Z. Lin, and L. S. Davis, *Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. **35**(11): p. 2651-2664.
- [10]. Wang, Q., Chen Feng, Xu Wenli, and M. H. Yang. *Online discriminative object tracking with local sparse representation*. in *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*. 2012. p. 425-432.
- [11]. Yuille, A. L. and Anand Rangarajan, *The Concave-Convex Procedure*. Neural Computation, 2003. **15**(4): p. 915-936.