Vehicle Make & Model Recognition

by
Zhihao DAI

COP507 Computer Vision & Embedded Systems
Coursework Report

Loughborough University

Jan. 2020

Abstract

In this coursework, I implement a JPEG Image Compression Simulation using MATLAB as frontend GUI and Python as backend JPEG CODEC. There are 2 simulation parameters K and Q' in the application. Several specific design considerations are introduced to the implementation, inclding an end-to-end MATLAB interface, an "Video Compression" functionality and DCT as Matrix Computation. I conclude that both K and Q' can significantly affect the quality of the compressed image.

Contents

List of Figures

List of Tables

List of Listings

Chapter 1

Introduction

Automatic Number Plate Recognition (ANPR) systems are widely used for policing, traffic monitoring and access control. They have proven to be accurate and efficient under most scenarios. However, ANPR systems are vulnerable to plate cloning, forgery or erosion.

A Vehicle Make & Model Recognition (VMMR) system receives an image of a vehicle as input and outputs the make and model of that vehicle. Such system could strengthen the security of existing ANPR systems by providing a matching between vehicle types and number plates. For example, in access control, if the number plate is not registered under the detected vehicle type, a security warning is raised and manual intervention is required.

In this paper, we design and implement a VMMR system. The input to the system is a cropped frontal image of a vehicle and the output is the make and model of the vehicle.

1.1 Related Work

Due to the significance of VMMR systems, many approaches have been proposed for building VMMR systems in recent years. Petrovic and Cootes [10] extracted simple features such as Sobel Gradients, Edge Orientation, Square Mapped Gradients from images in the database. Features are then represented and stored either in full dimension or in low dimension through Principal Component Analysis (PCA). Given a new image, the VMMR system predicts the vehicle type by finding the closest match in dot product distance. Their experiments on a dataset of 1132 frontal images of 77 vehicle classes showed that direct matching by Square Mapped Gradients features achieved the lowest vertification error of 3.5%.

AbdelMaseeh et al. [1] observed that unlike most object recognition tasks, VMMR poses a challenge of distingushing between similar classes under the same category (ie. vehicle). Based on this observation, they proposed the combination of global and local descriptors for VMMR. While global shape descriptors capture differences across categories, local shape and appearance descriptors for segmented regions capture inter-class varieties. An image is matched to the class with the smallest weighted sum of global and local dissimilarity measures.

Pearce and Pears [8] suggested using Harris corner detectors [4] for feature extraction and either k-nearest neighbour (KNN) or Naive Bayes Classifier for classification in VMMR systems. Local Harris strengths are computed through recursively dividing the image into quadrants and computing the sum of Harris corner response for each quadrant. Such features are then normalised through being devided by the sum of higher level strengths. For an input image of 150 by 150, a feature vector of Locally normalised Harris strengths (LNHS) of depth 5 is retrived and only one-twentieth the size of the original response. Their experiments on a dataset of 262 frontal images of 74 vehicle classes showed that LNHS with Naive Bayes Classifier achieved the highest accuracy of 96%. Using LNHS as features speeds up the training of a classifier and does not reduce the accuracy.

Siddiqui et al. [11] proposed using Speeded Up Robust Features (SURF) [2] for feature extraction and Support Vector Machines (SVM) for classification. Following Sivic and Zisserman's work on Bag-of-Features method [12], a dictionary (bag) of SURF features was constructed using K-Means clustering algorithm. An image can be then transformed into a fixed-length vector of visual words occurances and be fed into a SVM classifier for vehicle type recognition. High accuracy score of 94.84% was obtained on a large dataset of 6601 frontal images of 29 vehicle classes.

Zafar et al. [14] observed that dimensionality reduction methods used in many VMMR systems such as Principal Component Analysis (PCA) enhances the inner-class variance and can lead to miss-classification. In their setting, the raw pixel values of the image is directly projected to low-dimension space through Two Dimensional Linear Discriminant Analysis (2D-LDA) [6]. A match is found by minmizing the Euclidean distance to those in the training images set. The usage of 2D-LDA instead of PCA solves the variance problem by maximizing the ratio of intra-class variance to the inter-class variance. An accuracy score of 91% was obtained on a dataset of 271 frontal images of 25 vehicle classes (8 images per class for training and the rest for validation).

Zafar et al. [13] later proposed using localized Contourlet transform for feature extraction, 2D-LDA for dimensionality reduction, and SVM for classification. They reported a boosted accuracy of 96% on the same frontal car images dataset in [14].

Fraz et al. [3] introduced an innovative framework of Mid-Level-Representation of densely sampled features into VMMR. The framework starts by extracting patches around key-points detected by Difference of Gaussians (DoG) detector. For each extracted patch, A set of Scale-Invariant Feature Transform (SIFT) [7] feature descriptors are computed and reduced dimensionality by PCA. Fisher Vector [5], a Mid-Level-Representation (MLR), for the patch is then generated based on Gaussian Mixture Model (GMM), following Perronnin et al.'s work [9]. Fisher Vector for patches in images within the same class are visual words and collectively form a sub-lexicon. A lexicon of the training set images is essentially a collection sub-lexicons of all classes. Given a new image, the VMMR system extracts patches from the image, assigns each patch to a visual word by Euclidean distance within each sub-lexicon, classifies the image to the class (sub-lexicon) with the highest sum of similarity score of the word-patch matches. Fraz et al. reported an accuracy of 97.60% on the dataset used in [13] and 84.31% on a new dataset. The new dataset, coined 'Loughborough Cars (LC) Dataset', is composed of 1537 frontal images of 75 vehicle classes.

Our proposed system is trained and evaluated on a superset of the dataset in [14, 13, 3] of 530 frontal images from 27 vehicle classes. Despite having a smaller number of 4 images per vehicle class compared to 8 in [14] and 10 in [14], our method achieves a higher accuracy score of 98% on the validation set.

1.2 Dataset

Chapter 2

System Design

2.1 Assumptions

2.2 Feature Extraction

2.2.1 Raw Image

2.2.2 Sobel Edge Response

2.2.3 Square Mapped Gradients

2.2.4 Recursive Harris Corner

2.2.5 SURF + K-Means

2.3 Dimensionality Reduction

2.4 Classification

Chapter 3

Experiments and Results

Environment, etc.

3.1 Pre-processing

3.2 Cross-Validation

3.3 Merits of Performance

3.4 Effects of Feature Extraction Methods

3.5 Effects of Dimensionality Reduction Methods

3.6 Effects of Classification Methods

Chapter 4

Convolution Neural Network Model

4.1 Architecture

4.2 Overfitting Issues

4.3 Data Augmentation

Chapter 5

Discussion

5.1 Conclusions

5.2 Future Work

References

[1] AbdelMaseeh, M., Badreldin, I., Abdelkader, M. F., and El Saban, M. Car make and model recognition combining global and local cues. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (2012), IEEE, pp. 910–913.

[2] Bay, H., Tuytelaars, T., and Van Gool, L. Surf: Speeded up robust features. In *European Conference on Computer Vision (ECCV)* (2006), Springer, pp. 404–417.

[3] Fraz, M., Edirisinghe, E. A., and Sarfraz, M. S. Mid-level-representation based lexicon for vehicle make and model recognition. In *2014 22nd International Conference on Pattern Recognition (ICPR)* (2014), IEEE, pp. 393–398.

[4] Harris, C. G., Stephens, M., et al. A combined corner and edge detector. In *Alvey Vision Conference* (1988), vol. 15, Citeseer, pp. 10–5244.

[5] Jaakkola, T., and Haussler, D. Exploiting generative models in discriminative classifiers. In *Advances in Neural Information Processing Systems (NIPS)* (1999), pp. 487–493.

[6] Li, M., and Yuan, B. 2d-lda: A statistical linear discriminant analysis for image matrix. *Pattern Recognition Letters 26*, 5 (2005), 527–532.

[7] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV) 60*, 2 (2004), 91–110.

[8] Pearce, G., and Pears, N. Automatic make and model recognition from frontal images of cars. In *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (2011), IEEE, pp. 373–378.

[9] Perronnin, F., Sánchez, J., and Mensink, T. Improving the fisher kernel for large-scale image classification. In *European Conference on Computer Vision (ECCV)* (2010), Springer, pp. 143–156.

[10] Petrovic, V. S., and Cootes, T. F. Analysis of features for rigid structure vehicle type recognition. In *British Machine Vision Conference (BMVC)* (2004), vol. 2, pp. 587–596.

[11] Siddiqui, A. J., Mammeri, A., and Boukerche, A. Real-time vehicle make and model recognition based on a bag of surf features. *IEEE Transactions on Intelligent Transportation Systems 17*, 11 (2016), 3205–3219.

[12] Sivic, J., and Zisserman, A. Video google: A text retrieval approach to object matching in videos. In *null* (2003), IEEE, p. 1470.

[13] Zafar, I., Edirisinghe, E. A., and Acar, B. S. Localized contourlet features in vehicle make and model recognition. In *Image Processing: Machine Vision Applications II* (2009), vol. 7251, International Society for Optics and Photonics, p. 725105.

[14] Zafar, I., Edirisinghe, E. A., Acar, S., and Bez, H. E. Two dimensional statistical linear discriminant analysis for real-time robust vehicle type recognition. In *Real-Time Image Processing 2007* (2007), vol. 6496, International Society for Optics and Photonics, p. 649602.

Appendix A

Source Code