

Support Vector Machine based Vehicle Make and Model Recognition System

Muhammad Asif Manzoor*, Yasser Morgan

Faculty of Engineering and Applied Sciences, University of Regina, SK, Canada

ARTICLE INFO

Article history:

Received: 01 May, 2017

Accepted: 15 June, 2017

Online: 16 July, 2017

Keywords:

Vehicle Classification

Support Vector Machine (SVM)

Scale Invariant Feature

Transform (SIFT)

Speed-Up Robust Transform

(SURF)

Bag-of-Features (BoF)

ABSTRACT

Vehicle analysis is a very useful component in various real world applications. In this paper, we have developed a Vehicle Make and Model Recognition (VMMR) system using Support Vector Machine (SVM). Scale Invariant Feature Transform (SIFT) and Speed-Up Robust Transform (SURF) are used to extract local features from an image. Bag-of-Features (BoF) model is used to create visual dictionaries and convert the local image features into global image feature representation. Multiple dictionaries of different sizes are created for both features; SIFT and SURF and the dataset is coded using these dictionaries to determine the best size for the visual dictionary. NTOU-MMR is a publicly available vehicle dataset which we have used to evaluate the performance of proposed VMMR system. 92% recognition rate is achieved by using the proposed VMMR system.

1. Introduction

This paper is an extension of work originally presented in 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC) [1]. In original work, we had developed a Vehicle Make and Model Recognition (VMMR) system using Bag of SIFT features and used Support Vector Machine (SVM) for the classification task. In this extended work, we have investigated Speed-Up Robust Features (SURF) along with SIFT features. We have investigated the effect of dictionary size over the recognition rate. We also have analyzed the margin size between for binary SVMs and its relation with the recognition rate.

Machine vision based vehicular analysis is suitable for many scenarios. Machine vision based techniques require installation of cameras to capture the videos and/or images and it also requires computing power to process these captured videos/images. Traffic cameras are widely available at present time; which can be used to capture real-time videos/images. These techniques do not require any installation of devices/sensor in the vehicles; which

makes vehicular analysis simple. However, machine vision based applications have their own challenges.

Vehicular analysis applications may include License Plate Recognition, Vehicle Detection, and categorization of vehicles into bus, cars, trucks etc.; the majority of research is done for these mentioned scenarios. The focus of this work is to classify the vehicles according to their Make, Model and Manufacturing year. We are using machine vision-based approach to recognizing a specific instance of a vehicle. We are using Machine learning algorithm to classify input image/video according to the vehicle present. Machine learning algorithm provides the mean to train the system using a training dataset and then predict the outcome for new and unseen images. A lot many challenges are associated with this problem; some of these challenges are:

1. Image acquisition in outdoor environment.
2. Varying and uncontrolled illumination conditions.
3. Varying and uncontrolled weather conditions.
4. Wide variety of available vehicle appearances.
5. Similarities between different vehicle models.

* Corresponding Author: Muhammad Asif Manzoor, Faculty of Engineering and Applied Sciences, University of Regina, SK, Canada, AsifManzoor@uregina.ca

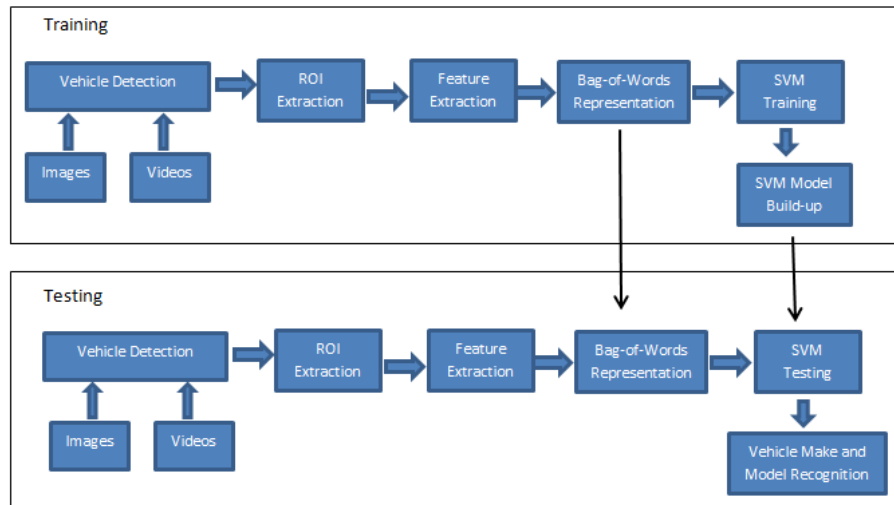


Figure 1: VMMR system Architecture

The architecture of VMMR system developed in this work is given in Figure 1. The system is designed to classify the vehicles using the frontal images. The input to the system can be either images or videos. If the videos are used as input sources, individual frames (images) must be extracted and will be processed separately. First, we must train the VMMR system based on input training dataset. The first step is to detect whether a vehicle exists in the input image. If it does not then we cannot use this non-vehicle image for the training. The vehicular images contain background along with vehicle and some parts of the vehicle (like a windshield) are almost identical for various vehicles' models. Hence, we define Region of Interest (ROI); which can be easily distinguishable for various models. Images features are extracted from this ROI and converted into image feature vector to represent a specific vehicle model. Lastly, machine learning algorithm is trained using these image feature vectors. Scale Invariant Feature Transform (SIFT) [2] and Speed-Up Robust Features (SURF) [3] are used to extract local image features. Bag-of-features (BoF) model is employed here to build the visual dictionaries and to transform the local image features into global image feature vector. We have used Support vector machine (SVM) [4], [5] as classifier in this work; which is supervised learning algorithm.

We have reviewed some related articles in section 2 and presented the dataset used in section 3. Section 4 discusses the VMMR system methodology in detail. Experimental results are provided in section 5 and section 6 conclude this work.

2. Literature Review

VMMR is a challenging task; some of the challenges are presented in Section 1. We will provide an overview of some of the related research work.

In [6], authors proposed VMMR system using the rear-view images. Authors defined the shape and geographical features with respect to taillights and license plate and used these features for vehicle recognition. The proposed system is developed to recognize the vehicles during the night under limited lighting conditions. The initial step is to recognize the location of license

plate and authors used license plate location to calculate the features. Genetic Algorithm is also used in this work for feature selection. Genetic algorithm improves the recognition rate by 0.4%. However, the effect of Genetic Algorithm over the computational time is not discussed.

In [7], authors designed a system to identify vehicles' type by using deep convolutional neural network. The proposed system identifies vehicles without vehicle detection. Authors studied VGGNet, GoogLeNet, and CaffeNet (three well known convolutional neural network method). They had suggested including of vehicle/non-vehicle classification as pre-training for three convolutional neural network methods. Data enhancement techniques are used for performance enhancement. The proposed method is tested against cars dataset [8] and 79.5% accuracy is achieved in this work.

In [9], authors used side profile images in their work to recognize the vehicle's Make and Model. Authors have used five different classification techniques in their research to develop VMMR system; Random Forest, Evolutionary Forest, HoG based Random Forest, HoG based Linear SVM and HoG based RBF SVM classifier. A pole-mounted camera is used to capture the images and a dataset is created containing more than 10,000 images. These images contain 86 different make and model and are divided into 9 categories.

In [10], the oriented contour points are extracted from frontal image to recognize the vehicles in their research. Authors have used three voting and distance error to determine the make and model. The dataset tested in this work contains partially occluded images as well and have 830 vehicle images. Authors have reported recognition rate of 90.6%.

In [11], Contourlet transform is applied to extract features and applied localized directional feature selection criterion in their work. They also have used Two-Dimensional Linear Discriminant Analysis to reduce the feature dimensionality. Authors have used support vector machine classifier in their proposed framework.

Table 1: Details of Dataset

Vehicle	Year	Train	Test	Vehicle	Year	Train	Test
Toyota Altis	2008-10	260	504	Nissan Tiida	2009	108	139
Honda CRV	2003-09	224	258	Toyota Altis	2005-06	216	227
Toyota Camry	2008-10	117	169	Mits. Zinger	2010	12	13
Honda Civic	2010	79	243	Mits. Outlander	2010	26	50
Honda Fit	2012	34	35	Toyota Wish	2010	68	45
Honda Fit	2009	19	13	Mits. Savrin	2008	36	12
Toyota Camry	2005-06	109	122	Toyota Wish	2005-09	107	77
Toyota Camry	1999	21	5	Mits. Lancer	2007	16	61
Nissan March	2007-08	96	92	Toyota Yaris	2008	155	147
Suzuki Solio	2008	34	84	Ford Liata	2003	9	10
Toyota Vios	2008-10	242	292	Toyota RAV4	2009	79	0
Nissan Livna	2010	119	128	Ford Escape	2009	50	45
Nissan Teana	2010	66	29	Toyota Innova	2008	15	29
Nissan Sentra	2003	22	20	Ford Mondeo		38	10
Nissan Sentra	2005	24	15	Toyota Surf	2008	40	30
Nissan Cefiro	1997	67	22	Ford Tierra	2006	34	16
Nissan Cefiro	1990	36	9	Toyota Tarcel	2005	110	70
Nissan X-rail	2007	37	89	Total		2725	3110

3. Dataset

[12] is a publicly available dataset which contains vehicular images. We have used this dataset in order to analyze our VMMR system. The dataset contains complete images (vehicle and background) as well as the region of interest images. We have used images in which ROI is already extracted. For a real world scenario, images are captured during different time and different weather conditions. Images are also partially occluded with irrelevant objects like pedestrians. The viewing angle for these images is -20 degrees to 20 degrees. The detailed description of each category and training and testing images available for each category is given in Table 1. The images are divided into classes on the basis of make, model, and shape (manufacturing year). The dataset contains thirty six different types (shapes) of vehicles. The dataset contains 2725 images for training purposes and 3110 images for the testing process. Number of available images for testing and training are given in Test and Train columns respectively in Table 1.

4. Methodology

The architecture of VMMR system is given in figure 1. The dataset provides the images in which Region of Interest is already extracted. Hence we are focusing on following three tasks of the VMMR system.

1. Feature Extraction: we have applied Scale Invariant Feature Transform (SIFT) and Speed-Up Robust Features (SURF) in this work.

2. Bag-of-Features (BoF) model: we have used BoF model to create visual dictionaries and encode the local image features into a global image feature vector.
3. Support Vector Machine (SVM): SVM is a supervised learning algorithm and it is used here as a classifier.

4.1. Feature Extraction

The first step for many machine vision applications is feature extraction. The input image is processed and local prominent interest points are located. A robust descriptor is used to encode these interest points which are invariant to many different kinds of noise.

Scale Invariant Feature Transform (SIFT)

Scale Invariant Feature Transform (SIFT) is introduced by David Lowe [2]. We have computed SIFT interest points in every image (training and testing dataset). Every SIFT descriptor construct a descriptor based on the histogram of gradient direction and magnitude around the interest point. The features, extracted using SIFT algorithm, are invariant to rotation and scaling. These features are also not affected by slight changes in view point, noise, and illumination. The operational flow chart of SIFT feature detection is given in figure 2.

SIFT algorithm detects interest point using scale-space maxima detection. Once the robust and invariant interest points are selected; these are encoded into 128-dimensional feature vector based on their appearance in a 4 x 4 patch. We have used standard SIFT algorithm in our work to detect the local image features.

Speed-Up Robust Features (SURF)

Speed-Up Robust Features (SURF) descriptor works similarly to SIFT descriptor but SURF is quicker as compared to the SIFT [3]. Herbert Bay et al. presented the idea of SURF feature detector at the 2006 European Conference on Computer Vision. SURF is partly inspired by SIFT feature detector algorithm. Hessian Matrix is used to detect the interest points in SURF algorithm. The first step is to apply Hessian Matrix on the integral image. The next step is to locate the Extrema and unstable Extremes points are discarded. Finally, orientations are assigned to construct the SURF descriptor. SURF descriptor is a 64-dimensional feature vector. Figure 3 shows the operational flow chart of SURF feature detection.

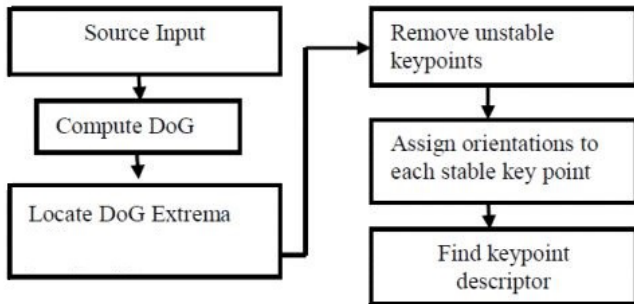


Figure 2: Flow Chart for SIFT feature detection

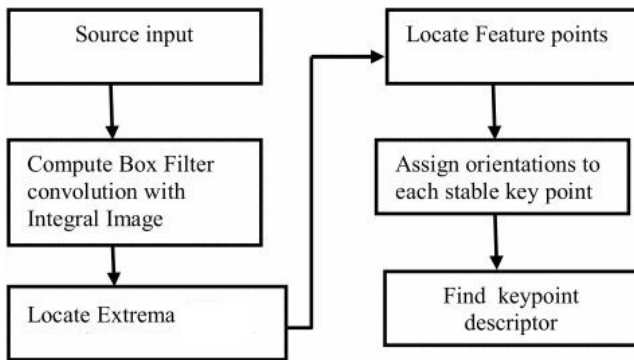


Figure 3: Flow Chart for SURF Feature Detection

4.2. Bag-of-Feature (BoF) Model

Bag-of-features (BoF) model [13] used in this work is based on bag-of-Words (BoW) model. BoW is originally used for document classification; a document is represented using histogram based on the occurrence of words. We must have to create visual dictionaries for an image that can be later used to construct a histogram. Using this histogram of the visual word, an image feature vector is created to represent the image. Whether we are using SIFT or SURF features; both are local image feature and require global feature representation. We will describe the BoF process in terms of SURF descriptor. Same BoF process is applied for SIFT algorithm.

We have used K-means clustering over all SURF features and the centers of each cluster represent a visual word. The collection of these visual words is used to create the dictionary. After creating the dictionary, each SURF feature is mapped into specific visual word and an image is represented by the histogram of the visual words. The dictionary is created by using only training dataset and the same dictionary is used to encode the testing dataset images as shown in figure 1.

Support Vector Machine

Support Vector Machine [4], [5] is binary classifier; multiple binary SVM can be combined to create multi-class SVM. SVM is used for both regression and classification problems. Support vector machine is successfully applied to many real world problems. SVM creates wide decision boundary (as wide as possible) to partition two categories as opposed to other machine learning algorithm; where single line decision boundary is used to separate categories. So SVM is also named as maximum margin classifier. Kernel functions are applied on training dataset to convert linearly non-separable categories into linearly separable classes.

Let assume a training dataset $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where $x_i \in \mathbb{R}^n$ and is input feature vector and $y_i \in \{-1, 1\}$ is the category. Let weight and bias of hyper plane is given by w and b . $\phi(x)$ represents transformed dataset (transformed using kernel function). The hyper plane between the two categories can be defined as:

$$w \cdot \phi(x) + b = 0 \quad (1)$$

The optimization problem for calculation of w and b is:

Minimize

$$\phi(w) = \frac{1}{2} \|w\|^2 \quad (2)$$

Subject to

$$y_i(w \cdot \phi(x) + b) \geq 1$$

New variables ζ_i (slack variable) and C (regularization constant) are included in above optimization problem:

Minimize

$$\phi(w, \zeta) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^N \zeta_i \quad (3)$$

Subject to

$$y_i(w \cdot \phi(x) + b) \geq 1 - \zeta_i, \zeta_i \geq 0$$

Where ζ is used to relax the hard margin constraint and C is used to manage the trade-off between classification error and maximal margin of separation.

One-vs-All [14] and One-vs-One [15] are two popular methods to construct multiclass SVM using multiple binary SVMs to deal with real world scenarios.

5. Experimental Results and Discussion

We have proposed Support Vector Machine based VMMR system and we have used SURF and SIFT features in this work. We have used different sizes and SVM configuration to attain better recognition rate. We have tested our proposed method using publicly available dataset. Figure 4 and Table 2 provides the details about SURF based recognition rate. SIFT based recognition rates are provided in figure 5 and Table 3.

The recognition rate is given on vertical axis and vocabulary size is given on horizontal axis in figure 4. Each graph line is generated for different SVM configuration. The best recognition rate among all the results is 90.90% with using SURF features. As evident from the figure 4 and table 3, better results as compared to others are generated with larger dictionary size and higher value of C. The recognition rate for all the variations is given in tabular format as well. The best recognition rate is achieved at C=8, 10, 12 and the vocabulary size, in this case, is 3000 visual words.

The recognition rate for SIFT image features is given in figure 5. Table 3 shows the recognition rate in a tabular format. The best recognition rate achieved in this case is 92.13%. As compared to SURF, the good recognition rate is achieved at 600 and 1800 vocabulary size; which is less as compared to SURF results. The recognition rate of 92.13% is achieved by using a vocabulary size of 600 visual words.

With these results, we can conclude that the SIFT can achieve better recognition rate as compared to SURF based recognition. However, considering other factors during selection process between SIFT and SURF may change the result.

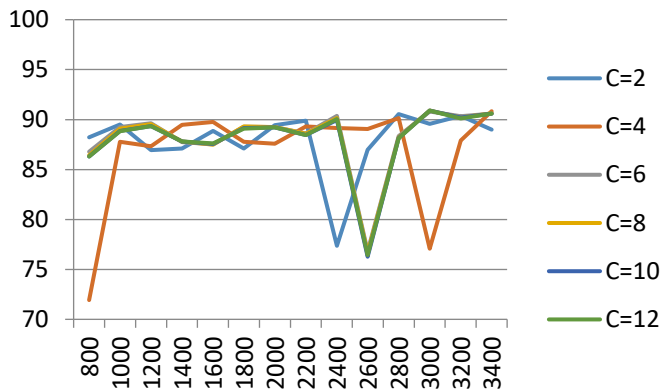


Figure 4: Recognition rate for SURF features for various SVM configuration and vocabulary size

Table 2: Recognition rate for SURF features for various SVM configuration and vocabulary size

Size	C=2	C=4	C=6	C=8	C=10	C=12
800	88.23	71.93	86.79	86.43	86.30	86.30
1000	89.52	87.78	89.26	89.10	88.88	88.84
1200	86.95	87.33	89.65	89.55	89.36	89.36
1400	87.11	89.49	87.75	87.81	87.81	87.81
1600	88.88	89.78	87.52	87.49	87.56	87.62
1800	87.11	87.78	89.29	89.33	89.16	89.10
2000	89.45	87.59	89.26	89.26	89.23	89.23
2200	89.90	89.33	88.62	88.59	88.46	88.46
2400	77.36	89.16	90.39	90.16	89.97	90.06
2600	86.98	89.07	76.82	76.62	76.27	76.43
2800	90.55	90.16	88.36	88.23	88.17	88.14
3000	89.58	77.07	90.80	90.90	90.90	90.90
3200	90.39	87.91	90.32	90.23	90.23	90.13
3400	89.00	90.84	90.58	90.61	90.58	90.58

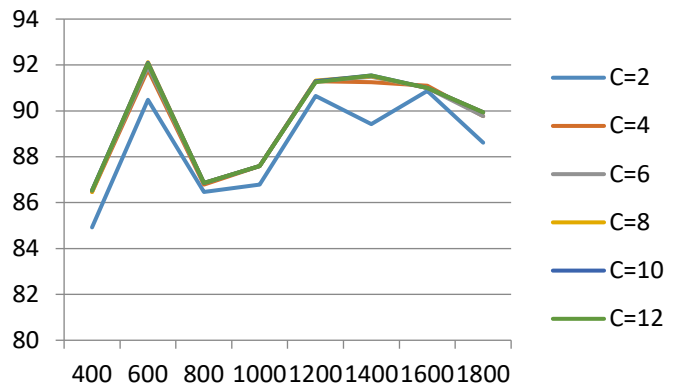


Figure 5: Recognition rate for SIFT features for various SVM configurations and vocabulary sizes

Table 3: Recognition rate for SIFT features for various SVM configurations and vocabulary sizes

Size	C = 2	C = 4	C = 6	C = 8	C = 10	C = 12
400	84.92	86.46	86.53	86.46	86.53	86.53
600	90.48	91.80	91.93	92.12	92.09	92.06
800	86.46	86.79	86.85	86.85	86.85	86.85
1000	86.79	87.59	87.59	87.59	87.59	87.59
1200	90.64	91.29	91.29	91.32	91.29	91.25
1400	89.42	91.25	91.48	91.51	91.54	91.54
1600	90.87	91.09	91.00	91.00	91.00	91.00
1800	88.62	89.78	89.78	89.94	89.94	89.94

6. Conclusion

VMMR is an important component of the intelligent transportation system and smart cities. We have presented support vector machine based VMMR system in this work. This VMMR system takes images as input and extracts the SIFT and SURF features; which are later used during the classification process. Bag-of-features (BoF) model is another component of VMMR system; which provides image feature vector. The proposed VMMR system is evaluated with the publicly available dataset and 92.13% recognition rate is achieved.

References

- [1] M. A. Manzoor and Y. Morgan, "Vehicle Make and Model classification system using bag of SIFT features," 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, 2017, pp. 1-5.
- [2] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91-110.
- [3] Bay, Herbert, et al. "Speeded-up robust features (SURF)." Computer vision and image understanding 110.3 (2008): 346-359.
- [4] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, pp. 121-167, June 1998.
- [5] V. N. Vapnik, The Nature of Statistical Learning Theory. New York, NY, USA: Springer-Verlag New York, Inc., 1995.
- [6] Boonsim, Noppakun, and Simant Prakoonwit. "Car make and model recognition under limited lighting conditions at night." Pattern Analysis and Applications (2016): 1-13.

- [7] Zhang, Feiyun, Xiao Xu, and Yu Qiao. "Deep classification of vehicle makers and models: The effectiveness of pre-training and data enhancement." 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2015.
- [8] Krause, Jonathan, et al. "3d object representations for fine-grained categorization." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2013.
- [9] Boyle, Jonathan, and James Ferryman. "Vehicle subtype, make and model classification from side profile video." Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on. IEEE, 2015.
- [10] Clady, Xavier, et al. "Multi-class vehicle type recognition system." IAPR Workshop on Artificial Neural Networks in Pattern Recognition. Springer Berlin Heidelberg, 2008.
- [11] Zafar, Iffat, Eran A. Edirisinghe, and B. Serpil Acar. "Localized contourlet features in vehicle make and model recognition." IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2009.
- [12] [Online]. Available: <http://mmplab.cs.ntou.edu.tw/MMR/>
- [13] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, Visual categorization with bags of keypoints, in Proc. ECCV Workshop Stat. Learn. Comput. Vis., 2004, pp. 122.
- [14] Vapnik, V.: Statistical Learning Theory. Wiley, New York (1998)
- [15] Kreßel, U.: Pairwise classification and support vector machines. In: Schölkopf, B., Burges, C., Smola, A. (eds.) Advances in Kernel Methods: Support Vector Learning, pp. 255–268. MIT Press, Cambridge (1999)