ON-Road vehicle classification based on random neural network and bag-of-visual words

Khaled F. Hussain
Faculty of Computers and Information, Assiut University, Assiut, Egypt
E-mail: khussain@aun.edu.eg

Ghada S. Moussa
Civil Engineering Department, Assiut University, Assiut, Egypt
E-mail: ghada.moussa@aun.edu.eg

A large increase in the number and types of vehicles occurred due to the growth in population. This fact brings the need for efficient vehicle classification systems that can be used in traffic surveillance and intelligent transportation systems. In this study, a multi-type vehicle classification system based on Random Neural Networks (RNNs) and Bag-Of-Visual Words (BOVWs) is developed. A 10-fold cross-validation technique is used, with a large dataset, to assess the proposed approach. Moreover, the BOVW–RNN’s classification performance is compared with LIVCS, a vehicle classification system based on RNNs. The results reveal that BOVW–RNN classification system produces more reliable and accurate classification results than LIVCS. The main contribution of this paper is that the developed system can serve as a framework for many vehicle classification systems.

1. INTRODUCTION

No doubt the rapid increase in demand on the limited transportation infrastructure leads to traffic related problems, such as congestion and accidents – with vast negative economic consequences. Intelligent Transportation Systems (ITS) are playing a vital role to cope up with these issues. In this sense, ITS turns its focus on vehicle classification. In fact, vehicle classification is a key component of many contemporary ITS because it allows obtaining the traffic parameters [10]. Vehicle classification is an essential technique used for many transportation systems such as toll plaza, security system and traffic surveillance [30]. In recent years, research in the area of detection and classification of vehicles from still images has attained an enormous amount of attention due to its crucial role in a wide range of applications [5,21]. However, vehicle detection and classification from still images is a challenging task. This is due to their high intraclass variations, many types of vehicles belonging to the same class have features of various sizes and shapes. Moreover, occlusion, shadow and illumination make the classification task even more challenging [5]. These issues can be resolved by using reliable and appropriate vehicle classification system.
In this study, a multi-type vehicle classification system based on BOVWs and RNN is developed. The BOVW has received much attention in object recognition [8,27]. The main idea of the BOVW paradigm is to treat image local features as visual words. The BOVW paradigm compares visual words of the input image to the set of visual word vocabulary. The RNN [11,12] is a spiked recurrent stochastic model. The behavior of the RNN is inspired from the behavior of biophysical neurons. The RNN has many attractive properties such as the existence and uniqueness of the steady-state solution, low complexity and strong generalization of its standard learning algorithm, product-form solution and the value of neuron potential represented by an integer. The RNN has been used in many applications such as emergency management [16], texture modeling [2,13,14], image segmentation [23], image enhancement [3], image and video compression [6,7], and simulating autonomous agents in augmented reality [15,18].

Since we combine BOVWs with RNNs, we call our proposed system BOVW–RNN. The BOVW–RNN is able to classify vehicles into four different classes: Motorcycles, small, medium and large. Experiments for vehicle-type classification are conducted with a large dataset to examine the performance of BOVW–RNN and compare it with LIVCS, a vehicle classification system based on RNNs [19].

The rest of the paper is organized as follows: Section 2 presents an overview of the related work. Section 3 discusses the technical approach and system framework. Experimental results and analysis are presented in Section 4. Finally, conclusions are summarized in Section 5.

2. RELATED WORK

Various systems have been developed for classification of vehicles. Each system uses different technique for classification. Generally, the accuracy of the classification for any classification system depends crucially on the combination of the extracted features of vehicles and the type of a classifier used for the classification [9]. Classification systems make use of known classifiers, such as Nearest Neighbor (KNN), Neural Networks (NN), Support Vector Machine (SVM) and Hidden Markov Models (HMM) [10]. The extracted features of vehicles, in most classification methods, can be categorized into two main techniques: Geometry-based and appearance-based [28].

Geometry-based techniques use geometric measurements of vehicles such as height, length and width as features for the classification. Those techniques are custom-made for particular vehicle classes through user decisions of measurements. In [32], they developed a length-based vehicle classification system that uses un-calibrated camera to capture images of the road. Then, they used the relative length between trucks and other vehicles to classify trucks. They achieved 91.89% classification accuracy for truck under specific conditions. Their framework has a few issues with vehicle occlusion and light variation. In [1], they developed an image based vehicle classification system using images acquired from range sensors. They extracted features such as length, width and height, then they used them for classifying vehicles into five classes. They achieved 89% classification accuracy for five vehicle classes. In [19], they presented a vehicle classification system that extracts geometric characteristics of vehicles from laser intensity images and used Random Neural Network (RNN) as a classifier. They achieved 91.91% classification accuracy while classifying vehicles into five different categories. In [20], they presented a Hybrid Dynamic Bayesian Network classification system that extracts tail light and vehicle dimensions from rear-side images of vehicles. They classified vehicles into one of four classes: Sedan, Pickup truck, SUV/Minivan and unknown with an overall classification accuracy of 95.68%. Generally, geometric-based
methods alone may not successfully classify fine-grained vehicle types since different types
of vehicles may have similar size [28].

In appearance-based techniques, local features are extracted from images and used
for the classification. In [24], they developed an edge-based rich representation for vehicle
classification that can be used to classify vehicles of similar shapes. They modified SIFT
descriptors and used un-calibrated camera. They developed three different models: Explicit
shape, implicit shape and shape-only. They reported classification accuracy of 95.76% for
sedan versus taxi, and 98.5% for car versus minivan using the explicit shape model. For
the implicit shape model, they reported 96.06% for sedan versus taxi, and 98.25% for car
versus minivan. The classification accuracy of the shape-only model was 87.58% for sedan
versus taxi, and 94.75% for car versus. minivan. In [26], they developed an image-based
vehicle classification system using canny’s edge detector, SIFT and K-means. To classify
vehicles into two classes: Car and minivan, they used weighted clusters as classifier and
achieved 98.5% classification accuracy. To classify vehicles into two classes: Sedan and taxi,
they used DNA analogy classifier and achieved 95.45% classification accuracy. In [33], they
developed an appearance learning-based technique to recognize moving objects such as vehi-
cles, bicycles, and people using multi-block local binary patterns. In [31], a two-ensemble
classification scheme with two features extraction (Gabor wavelet transform and the Pyra-
mid Histogram of Oriented Gradients) was presented. Multiple classifiers were implemented
in the classification, including KNNs, MLPs, SVMs and random forest. A classification
accuracy of 98.65% with a rejection rate of 2.5% was reported. In [25], she developed two
classification approaches: (1) geometric-based approach, and (2) appearance-based approach
and used them for two tasks: Multi-class (small, medium and Large) and intra-class vehi-
cle classifications (PU, SUV and VAN). For multi-class classification both geometric-based
and appearance-based approaches brought almost similar recognition rates. However, the
appearance-based approach was the best choice for intra-class classification.

3. PROPOSED SYSTEM FRAMEWORK

The proposed system is based on RNN and BOVW. The framework of the proposed system
is divided into two stages: Training stage and testing stage as shown in Figures 1 and 2,
respectively. In the training stage, the system is trained with training images to recognize
a set of different vehicle classes. The output from the training stage serves as an input
to the testing stage. In the testing stage, vehicles are classified into four different classes:
motorcycles, small, medium and large. A detail description of the training and testing stages
are presented in the following subsections.

3.1. Training Stage

The training stage is accomplished via four steps: (1) local feature extraction, (2) visual
vocabulary construction, (3) histogram construction and (4) RNN training, as shown in
Figure 1.

1. Local feature extraction: scale-invariant feature transform (SIFT) [22] descriptors
are extracted from images. SIFT is a widely used feature descriptor. Many object
recognition systems use SIFT to represent images and it has been proven to be a
powerful and successful local image descriptor [4,20]. Keypoints are detected in the
images using Harris–Laplace salient point detector. Subsequently, the SIFT descrip-
tors are computed for each keypoint in the images. Each keypoint is represented by
a 128-dimensional vector.
2. Visual vocabulary construction: After feature extraction, each input image is represented by a set of SIFT local descriptors. A visual vocabulary is constructed by applying a $k$-means clustering algorithm to divide all SIFT feature descriptors from the training set into clusters. Each cluster center is a visual word, which is a 128-dimensional vector. All visual words make up a visual vocabulary.

3. Histogram construction: The SIFT local feature descriptors along with the constructed visual vocabulary are used to construct a histogram for each image. In which,
all SIFT feature descriptors extracted from an image are quantized to their closest “visual word”. For each image, the number of SIFT feature descriptors assigned to each “visual word” is then accounted into the corresponding bin in the histogram. Therefore, each image is represented as a histogram of frequencies of visual words that are in the image.

4. RNN training: The classification is handled using RNN as a classifier. RNN is trained with constructed images’ histograms and their corresponding true class labels (ground truth data). The learning algorithm for the RNN produces the excitatory and inhibitory weights. The weights serve as an input for the testing stage; see Figure 2. A concise description of the RNN is summarized as following:

RNN [11,29] is a recurrent network of fully connected neurons. Neurons exchange signals in form of unit amplitude spikes. The state of neuron \(i\) is represented by a non-negative integer called its potential. When an excitation signal arrives to neuron \(i\), the state of neuron \(i\) is changed from state \(k_i\) to \(k_i + 1\). When an inhibition signal arrives to neuron \(i\), the state of neuron \(i\) is changed from state \(k_i\) to \(k_i - 1\). Neuron \(i\) emits a spike if \(k_i\) is positive; the state of neuron \(i\) is changed from state \(k_i\) to \(k_i - 1\). The spikes are sent out from neuron \(i\) at a rate \(r_i\) with exponentially distributed. Spikes are sent out from neuron \(i\) to neuron \(j\) as a positive signal with probability \(p^+_{ij}\) or as a negative signal with probability \(p^-_{ij}\), or they depart from the network with probability \(d_i\). The sum of these probabilities must be one:

\[
d_i + \sum_{j=1}^{n} [p^+_{ij} + p^-_{ij}] = 1, \forall i. \tag{1}
\]

The spikes are sent out from neuron \(i\) to neuron \(j\) at rates:

\[
w^+_{ij} = r_i p^+_{ij} \geq 0, \tag{2}
\]

\[
w^-_{ij} = r_i p^-_{ij} \geq 0, \tag{3}
\]

\(w^+_{ij}\) and \(w^-_{ij}\) are the excitatory and inhibitory weights, respectively. Combining Eqs. (1)–(3) yielding

\[
r_i = \frac{\sum_{j=1}^{n} [w^+_{ij} + w^-_{ij}]}{1 - d_i}. \tag{4}
\]

The total arrival rates of positive signals \(\lambda^+_i\) and negative signals \(\lambda^-_i\), for \(i = 1, \ldots, n\), can be calculated from the following nonlinear system of equations:

\[
\lambda^+_i = \Lambda_i + \sum_{j=1}^{n} q_j w^+_{ji},
\]

\[
\lambda^-_i = \lambda_i + \sum_{j=1}^{n} q_j w^-_{ji},
\]

where

\[
q_i = \min \left\{ 1, \frac{\lambda^+_i}{r_i + \lambda^-_i} \right\}, \quad r_i = \sum_{j=1}^{n} w^+_{ji} + w^-_{ji} \tag{5}
\]

where \(q_i\) is the output of neuron \(i\).
Each sample image in the training set is represented as a histogram of frequencies of visual words. The number of bins in the histogram is \( b \). The number of output vehicle classes of our proposed system is four. Thus, the RNN used in the proposed system consists of \( b \) input neurons, and four output neurons. The steps of learning vehicle features are the following:

1. Initialize the weights \( w_{ij}^+ \) and \( w_{ij}^- \), \( \forall i, j \), to random values between zero and one.
2. Set the inhibitory rates to zero.
3. Select at random the \( k \)th training pattern from the total number of training patterns \( K \).
4. Set the excitatory input rates for the input neurons, \( \Lambda_k = x_k \), where \( x_k \) is the \( k \)th input training pattern.
5. Solve the nonlinear system of Eq. (5) to obtain the neuron states \( q_i \), \( \forall i \).
6. Use the RNN learning algorithm [11,17] to update the weights \( w_{ij}^+ \) and \( w_{ij}^- \), \( \forall i, j \), which minimize the following error function:

\[
E = \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{n} (q_{ik} - y_{ik}^2),
\]

where \( y_{ik} \) is the desired values of the \( k \)th pattern.
7. Repeat steps 3–6 until a minimum of the error function is reached.

3.2. Testing Stage

The testing stage is composed of fewer processing steps than the learning stage and therefore, it is much faster than the learning stage. However, both stages share some common processing steps. The testing stage is accomplished via three steps; (1) Local Feature Extraction, (2) Histogram Construction, and (3) RNN Testing as shown in Figure 2. These steps are similar to training stage steps. In which the SIFT algorithm detect keypoints in the testing images using Harris–Laplace salient point detector and compute descriptors for each keypoint in the testing images. After feature extraction, each testing image is represented by a set of SIFT local descriptors that are used along with the constructed visual vocabulary (from training stage) for producing the histogram of the testing image. Then the histogram of the testing image, and excitatory and inhibitory weights (from training stage) are fed to the RNN to classify vehicles into their respective categories. Vehicles are classified into four different categories: motorcycles, small, medium and large.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The aim of this section is to evaluate the performance of our proposed system (BOVW-RNN) and compare it with another system based on RNNs only (LIVCS) [19]. Our proposed system has been implemented with Matlab software package. Experiments for vehicle type classification are conducted using a database set of 3550 vehicle images with their ground truth data; see Table 1. \( K \)-fold cross-validation method with \( k = 10 \) (10-fold CV) is used to assess the performance of our approach. We randomly split the 3550 vehicle images into 10 equal size sets. We used nine sets for training and one set for testing. The procedure is repeated 10 times; so that each set is used for testing only once. The final result is the
The performance of LIVCS and BOVW–RNN systems can be quantified through some measurements, using results of 10-fold CV. These measurements include: classification accuracy (ACC), recall (true positive rate (TPR)), false alarm (false positive rate (FPR)), specificity (true negative rate (TNR)) and precision (positive predictive value (PPV)). The \( J \)-statistic, also called informedness, is considered, which is a combined measure of recall and specificity, and it is calculated as \( \text{recall} + \text{specificity} - 1 \). Moreover, the \( F_1 \)-score is considered which is another measure considering both recall and precision, and is commonly used in information retrieval for query classification performance. The \( F_1 \)-score can be calculated as following:

\[
F_1 = \frac{2TP}{2TP + FN + FP},
\]

where TP, FN and FP are the number of true positive, false negative and false positive, respectively. The larger \( J \)-statistic and \( F_1 \)-score values, the better performance of the classification. The classification results for the two systems; LIVCS and BOVW–RNN are summarized in Table 2. The results reveal that BOVW–RNN classification system can achieve better classification results compared with the LIVCS. In which, BOVW–RNN attained higher classification accuracy (95.63\%) than LIVCS (91.21\%). Moreover, classification using BOVW–RNN achieves better and more balanced recall, specificity and precision (a higher \( J \)-statistic and \( F_1 \)-score), as well as lower false alarm (0.04 vs. 0.09).

The predominant focus of previous research on multi-type vehicle classification is on total classification accuracy (ACC), however, the recognition rates of different vehicle types are essential for numerous combinations of vehicle types. Those recognition rates for the LIVCS and BOVW–RNN systems are listed in Table 3. The BOVW–RNN overcomes LIVCS
for all vehicle types recognition rates except for the large size vehicles. This might be due to the fact that BOVW–RNN is using BOVWs and RNNs with appearance features (SIFT). On the other hand, LIVCS is using RNNs with geometric features such as vehicle length \cite{19}. Both systems (LIVCS and BOVW–RNN) attained low recognition rates for motorcycle (40 and 50\% respectively) which is directly caused by the small number of motorcycle data (only 10 out of 3550 total images, see Table 1).

The resulting confusion matrixes for LIVCS and BOVW–RNN are shown in Tables 4 and 5, respectively. The confusion matrix is often used to measure the similarity between classes in multiclass classification studies. In the confusion matrix, the rows and columns represent true and predicted classes, respectively. The diagonal represents the correct classification whereas the off-diagonal represents the incorrect classification.

5. CONCLUSION

In this study, we presented BOVW–RNN classification system framework for multi-type vehicle classification which is based on BOVWs and RNNs with SIFT features. Our proposed system (BOVW–RNN) classifies vehicles into four different classes: motorcycles, small, medium and large. BOVW–RNN is implemented with Matlab software package. Experiments for vehicle type classification are conducted with a large dataset to examine the performance of BOVW–RNN and compare it with LIVCS, a vehicle classification system based on RNNs. Based on the experimental results, BOVW–RNN outperforms LIVCS in many aspects. The classification using BOVW–RNN attain higher classification accuracy (95.63 vs. 91.21\%) and better and more balanced recall, specificity and precision (a higher $J$-statistic and $F_1$-score), as well as lower false alarm (0.04 vs. 0.09). It is clear that small number of motorcycle data (only 10 out of 3550 total images) affected the recognition rate of the motorcycles, but it did not affect the overall system accuracy. The main contribution of this paper is that the developed system can serve as a framework for many vehicle classification systems. In the further work, various types of features may be adapted in the proposed system to improve classification performance. Also, more vehicle classes can be considered.
References


