International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

© Global Society of Scientific Research and Researchers

http://ijcjournal.org/

Survey on Detection Methods for Self-Driving Cars

Myat Su Oo[a*], Dr. May The` Yu[b]

[a]*Digital Image Processing, University of Computer Studies Mandalay, Myanmar*

[b]*Faculty of Information Science, University of Computer Studies Mandalay, Myanmar*

[a]*Email: myatsuoo@ucsm.edu.mm*

[b]*Email: maytheyu@gmail.com*

Abstract

Accurate vehicle detection or classification plays an important role for self-driving cars. Objects classification and detection can be used in various such as Robotics, Medical Diagnosis, Safety, Industrial Inspection and Automation, Human Computer Interface, Advanced Driver Assistance System and Information Retrieval. In this article, we investigated the methods of detection and classification in context images and videos. SIFT, HOG, SVM, CNN, faster RCNN and YOLO methods are reviewed to detect and recognize the objects. The paper aims to know the methods that detect the obstacles on the way to reduce the traffic accidents. We summarize the results, faster-RCNN is better than the other methods for real-time citing the advantages and disadvantages of existing methods.

Keywords: SIFT; HOG; SVM; CNN; faster RCNN; YOLO.

1. Introduction

Autonomous driving is the very best stage of automation for a car, which means the automobile can pressure itself from a place to begin to a destination without a human intervention. The problem can be divided into two separate tasks. The first mission is centered on preserving the car moving alongside accurate path. The 2nd undertaking is the functionality to perceive and react to unpredictable dynamic boundaries, like different automobiles, pedestrians, and traffic signalization. In order to resolve these tasks, objects are needed to detect. Object detection, tracking and classification can be used for various purposes. Self-driving cars rely heavily on the detection and classification of objects. Object detection is useful for vehicle and pedestrian detection, traffic sign and lane detection or vehicle make detection. However, to achieve good results, both for the detection of objects and for classification, a lot of preliminary work is required. This includes pre-processing of images, such as noise removal, setting contrast, re-calibration and background subtraction.

----

* Corresponding author.

Based on feature extraction, either the Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), or Histogram of Oriented Gradients (HOG) can be used to generate points of interest. Moreover, interest in deep learning methods reports that CNN has recently increased in recent years.

This paper is organized as follows: Section II reviews the previous research for Collision Avoidance System. The steps of the classification are described in Section III followed by a survey on vehicle detection schemes in Section IV. Comparison of algorithms is presented in Section V. Conclusive remarks are given in Section VI.

2. Literature Review

In this paper, we provide an overview of strategies for automobile detection for Collision Avoidance System, that's a car protection system designed to reduce the opportunity of a twist of fate. To conduct the review, we downloaded about 150 articles so that we could perform a systematic technical analysis of the detection and classification of objects from various digital libraries, including IEEE Xplore, Science Direct, Google Scholar, ACM and others. We reviewed the research title, abstract, introduction, experiment, and future scale. We have identified the most appropriate document for review. This part of the article systematically posted comments for all 21 papers.

In [1], the author used a combination of SIFT and bag-of-words and performs the classification. In this paper, SIFT is used to extract and represent local points of interest. The bag word pattern is used to represent local objects as a fixed-length vector to represent an image. Support vector machine is used as classifier. The accuracy of the results is 89%.

In [2], SURF was used to detect all possible pairs of symmetrical matches by conversion to mirror. Then, to confirm the practicality and feasibility of the method, an application for make and model recognition of car (MMR) is accepted. 2846 vehicles for training and 4090 vehicles for testing with 29 vehicle types were used to evaluate the results. To detect automobiles, the symmetric descriptor is used to decide the area of interest of each automobile in the lane. This system has two benefits: it is not necessary to subtract the background and it is very effective for real-time applications. Then, two problems of MMR are solved. The problem of multiplicity is associated with a unique vehicle model, which often has different models on the road. The problem of ambiguity stems from the fact that cars from not the same companies use similar shapes. To solve above problems, a grid division system is offered to divide the vehicle into many grid. The histogram of gradient and SURF descriptors is approved for learning weak classifiers using the machine-based learning algorithm. The overall classifier can accurately recognize each vehicle because the high power of the grid presentation method and the great accuracy of vehicle detection.

X. Li and X. Guo [3] developed a forward vehicle detection system using HOG Descriptor and SVM. Their idea depends on detecting shadows under the vehicle. Distinguish between the vehicle and the background. Furthermore, even different illumination conditions are satisfactory.

In [4] Like previous studies, with Sivaraman Trivedi used the shadow under the car. Features are extracted using HOG and symmetric HOG to detect symmetric features. Into their research Feng Han et. al [5] uses HOG with

SVM to detect people. The authors introduce vehicles from various perspectives. First, using a stereo cue either a person or a vehicle candidate, next, classification using HOG features is performed with SVM. For each viewpoint, the author develops a separate classifier to be applied.

Pablo Negri et. al [6] used HOG and Haar-like features of their vehicle detection research. In fact, the author is experimenting with the merger these descriptors can be used to obtain accurate final results. In 2016, Sundaresh Ram and his colleagues [7] proposed an automatic scheme to detect any sizes of the vehicles in low-resolution aerial images. Firstly, it proved a novel vehicle enhancement filter including multiscale Hessian analysis. After thresholding, they developed the candidate recognitions of vehicles based on analysis of bilateral symmetry. After comparing with base line detection algorithms, the paper presented the improved performance of detections for any low-resolution aerial imagery.

Dong and his colleagues [8] proposed a semi-controlled CNN for detecting vehicles from the front Filters used in the convolution layer are examined using unmarked data, it is a rare training provided to the Laplacian filter. The output level is softmax classifier. The final solution is, classify car type.

In [9] One year ago, Dong and his colleagues reported that unsupervised CNN for vehicle type classification method based on the appearance of the front view of the vehicle. They use CNN to study characteristics, and then the network is trained with sparse filtering method to capture rich and discriminatory information on vehicles. After obtaining the final characteristics, the softmax regression is used to classify vehicle types. They used data set called BIT-Vehicle to measure performance.

In [10], CNN with low resolution frames is used to detect and classify car. At the preprocessing stage, the size of each frame changes, contrast with histogram alignment. The proposed CNN architecture is high. It provides low-level features. The author also changed the number of filters and the size of the filter, the number of hidden layers. The article by Heikki Hattunen and his colleagues [11] studied the automatic car recognition: bus, truck, van and small car. When applied to the classification of vehicles, they used DNN and SVM using SIFT functions. The result is verified using a database of more than 6,500 images, and the obtained estimate accuracy exceeds 97%.

In [12] the authors used detection and classification of vehicles DNN based on rear images captured by a static camera along a multi-lane highway. YOLO is fine-tuned for vehicle detection; the AlexNet model is configured to classify vehicles. AlexNet was used as a feature extractor and classify extracted features using linear SVM.

In [13] Using the camera, lidar, radar, and GPS, the authors made dataset of 17,000 images on highway with bounding boxes for vehicles and more than 616,000 images with annotations by lane only on rear view of cars moving in one direction. The annotated vehicle data contains nearly 17,000 images with 140,000 bounding boxes. Annotated data lanes contain more than 616,000 images. The Lane Detection and Vehicle Detection tests consist of 22 and 13 video clips. They used a single CNN algorithms work well for detection of traffic lanes and vehicles. They calculated the detection results for the four lane boundaries, namely the left and right boundaries of the ego lane, as well as the outer boundaries of two adjacent lanes. By using the Intersection over Union

(IOU), they evaluated vehicle accuracy bounding box predictions. A bounding box prediction matched with ground truth if IOU≥0.5.

In this article [14], we can look at the R-CNN, Fast R-CNN, Faster R-CNN, YOLO-Object detection algorithms. RCNN extracts 2,000 regions from an image as region suggestions, rather than categorizing a huge number of regions using alternative search. Therefore, there is a problem with RCNN. Since for each image it is necessary to classify 2000 regional proposals, the training of the network takes a huge amount of time. Each test image takes about 47 seconds, so it cannot be implemented in real time. The selective search algorithm is a fixed algorithm. Therefore, at this stage of learning was not. This may result in suggestions for unwanted candidates. Fast R - CNN is faster than R – CNN because there is no need to send applications of the region 2000 to the convolution neural network each time. Instead, the convolution operation is performed only once for each image from which the object map is generated. Fast R-CNN is much faster than regular models. Therefore, it can be used to detect objects in real time. YOLO or You Only Look Once is several times faster than other object detection algorithms (45 frames per second). The limitation of the YOLO algorithm is the struggle with small image objects.

In [15], Girshick and his colleagues presented R-CNN, a CNN network offering regional proposals. In their article, the authors would like to show that CNN can get better results than methods using low-level functions, such as HOG. Their object detection pipeline consists of in three parts. First, R-CNN uses a selective search to create offers for a region, that is, different regions that may include an object. Then, 4096-dimensional feature vectors are extracted from each detected area using AlexNet. The only change in architecture is the number of units in the classification layer. Each feature vector is then classified using SVM, and each SVM is trained for a particular class. In the ILSVRC2013 detection data set, R-CNN scored more than over feat.

In [16], the authors offered the combined method of CNN and SVM to recognize objects and detect pedestrian. The pre-prepared AlexNet and the new CNN architecture, including nine layers were used. CNN was used to extract its discriminatory attributes. They then applied PCA to CNN-derived features for decoration and reduction. Finally, they import them into the SVM classifier as input to raise the system's skill and use most rules to effectively merge the images. The method is shown using the average accuracy of each class in the Caltech-101 data set. It has been experimented that the LM-CNN-SVM system achieves the utmost precision results for 15 and 30 images per level. At the same time, a comparative study of deep learning methods and HOG was conducted. The results proved that the LM-CNN-SVM system is significantly better than the most advanced method in the field of excellence in improving recognition efficiency.

Krizhevsky and his colleagues [17] made a breakthrough in image classification of large data sets by introducing AlexNet. Although there are several other methods of machine learning used in the competition, the authors demonstrate that CNNs are capable of handling millions of images and achieving high results. The key to the end results is the depth of AlexNet. The data set is not pretreated except for subtracting the average. The proposed network consists of five convolutional layers and three fully connected layers. The authors chose ReLU for the activation function because it significantly improved the speed of training. Normalization of the local response is applied after activation in the first two layers. In addition, overlapping clusters have reduced over-adjustments during training. The last layer is a softmax layer of 1000 units. The training is done on two GPUs, each GPU having access to different layers. Increasing and abandoning data is used to reduce.

In [18], another successful CNN architecture was introduced - VGGNet. With this neural network, the authors took the first and second place in the tasks of classification and localization of the ILSVRC'14 competition. The foundation of the network is inspired by AlexNet. Like AlexNet, the only pretreatment was average extraction. The authors experimented with different depths of the network in order to understand how this affects the final results. One of the differences from the previous architectures is the use of a smaller kernel size of in all convolution 3×3 layers. In addition to this, CNN contains several consecutive convolutional layers without a grouping layer between them. The authors tested six different configurations of different depths. The best results were obtained when the network consisted of 16 and 19 layers. In [19] CNN-RNN is a convolutional neural network united with a recurrent neural network used for classification. The purpose of classification with multiple labels is to predict the labels of multiple objects in the same image. The authors noted that adding RNN to the original network increased accuracy.

Long and short memory units are used as repetitive blocks of this network. Part of the CNN network is based on VGGNet and is used to collect semantic information from images. Part of the RNN is related to the ratio of images and labels. Girshick and his colleagues described an R-CNN [20] detector, which involves of four portions: First, human-designed region proposal algorithms used to create object-like regions. Formerly, it resized the regions, and CNN used to extract greatly discriminative features from each region. Lastly, SVM classified the regions and regression developed bounding boxes.

3. Image Classification Steps

In recent years, deep learning methods improve to classify the images. These methods are different from traditional approaches in that they do better performance and automatically and rapidly learn the features directly from the raw pixels of the input images without using feature extraction approaches. The image classification steps are the following:

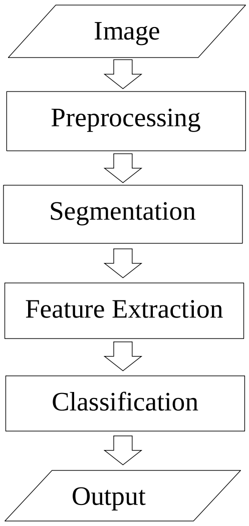

Figure 1: Steps of Image Classification

4. Vehicle Detection Approaches

This section explains about the approaches for vehicle detection. *Motion based approaches* exploit the temporal information in sensor information to locate automobiles. Optical flows fields from a shifting automobile are computed by way of matching feature points or pixels between frames. After computing the optical flow fields, transferring automobiles which includes the segments from the image with the aid of grouping the fields based totally on their positions, magnitudes and angles. This scheme is called tracking. *Appearance based totally approaches* use specific appearances of a car's view for ROIs extraction which include shadow underneath car, edges, corners, symmetry, texture, color and lights of vehicle. *In stereo based approach*es, multi-view geometry allows the direct measurement of three-dimensional information. Variation in left and right images between the corresponding pixels is called the shift and stereo match the differences of all points in the image calculate the difference map. If the stereo parameters are known, a disparity map can be converted to a 3D view observable scene. V- Disparity is widely used for version of the surface of the floor, in order to detect objects that lie above the floor. V- Disparity forms a histogram disparity values for pixel locations with the same v that is vertical image coordinate. V-disparity can be noticeably used in stereo-vision for smart cars [21].

5. Comparison of Methods

The following table shows the advantages and disadvantages of methods.

Table 1: Advantages and Disadvantages of Methods

Methods	Advantages	Disadvantages
SIFT	Is a classic approach. Is robust for features to occlusion and clutter more correct than other features descriptors. Is rotation and scale invariant.	Is precisely complex and heavy computing. Is based on the Histogram of Gradients. That is, the gradients of each Pixel in the patch need to be computed and these computations cost time. Doesn't work well with lighting changes and blur.

HOG	Offers more global information, while in smaller subdivision they provide more fine-grained detail.	Is the final descriptor vector grows larger, thus taking more time to extract and to train using a given class.
SVM	Gives elasticity in the select upon the threshold. Covers a nonlinear transformation. Is proper simplification ability. Remove over fitting problem. Decreases in computing complexity. Be able to simply decision rule complexity and miscalculation rate.	Is low transparency result. Is time consuming for training. Is hard to understand the configuration of algorithm. Is difficult to decide the optimal parameters when there is nonlinearly distinguishable training data.
CNN	Do well performance. Extract automatically and rapidly features from the raw pixels of the input images.	Is heavy computational time.
Faster-RCNN	Uses for real-time object detection. Is quite time consuming.	---
YOLO	Is orders of magnitude faster (45 frames per second) than other object detection algorithms.	Struggles with small objects within the image, for example it might have

difficulties in detecting a flock of birds.

6. Conclusion and Recommendations

The detection and classification of vehicles has a great influence on the achievements for Self-driving Cars. Computer vision task helps to develop road systems by analyzing traffic and helping to prevent or detect traffic accidents. Although useful in many ways, the detection and classification of vehicles is not easy task.

This paper examines approaches to detect and classify the objects. As the result, faster-RCNN is better than the other methods for real-time citing the advantages and disadvantages of existing methods. The corresponding work shows that feature extraction algorithms associated with a classifier such as SVM have been the method of choice for vehicle detection and classification for years.

The main drawback lies in the complexity of extracting the appropriate features. Therefore, these solutions are very dependent on pretreatment. Fortunately, the detection and classification of objects has become more accurate and fast with the increasing popularity of convolutional neural networks. These neural networks can produce high results on large data sets without the need for additional feature extraction and advanced pretreatment in most cases. We want to advise on reviewing the previous papers.

For real time, the faster-RCNN should be used to detect the objects on the way because fast-RCNN is faster than the other neural networks. When the neural networks are used for large image data sets, GPU are needed. Moreover, the performance should be high if other features methods are used. As the limitation, convolutional neural networks are not convenience for real time because they take the heavy computational time. The performance should be better.

References

[1] M. A. Manzoor and Y. Morgan, “Vehicle Make and Model classification system using bag of SIFT features”, in Computing and Communication Workshop and Conference (CCWC), 2017 IEEE 7th Annual, pp. 1–5, IEEE, 2017.

[2] J.-W. Hsieh, L.-C. Chen, and D.-Y. Chen, “Symmetrical SURF and its applications to vehicle detection and vehicle make and model recognition,” IEEE Transactions on intelligent transportation systems, vol. 15, no. 1, pp. 6–20, 2014.

[3] X. Li and X. Guo, “A HOG feature and SVM based method for forward vehicle detection with single camera,” in Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on, vol. 1, pp. 263–266, IEEE, 2013.

[4] S. Sivaraman and M. M. Trivedi, “Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis,” IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 4, pp. 1773–1795, 2013.

[5] F. Han, Y. Shan, R. Cekander, H. S. Sawhney, and R. Kumar, “A two-stage approach to people and

vehicle detection with HOG-based SVM," in Performance Metrics for Intelligent Systems 2006 Workshop, pp. 133–140, 2006.

[6] P. Negri, X. Clady, S. M. Hanif, and L. Prevost, "A cascade of boosted generative and discriminative classifiers for vehicle detection," EURASIP Journal on Advances in Signal Processing, vol. 2008, p. 136, 2008.

[7] Sundaresh Ram, Jeffrey J. Rodriguez, "Vehicle Detection In Aerial Images Using Multiscale Structure Enhancement And Symmetry" ,Icip(Ieee),Pp 3817-3821,2016.

[8] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semi supervised convolutional neural network," IEEE transactions on intelligent transportation systems, vol. 16, no. 4, pp. 2247–2256, 2015.

[9] Z. Dong, M. Pei, Y. He, T. Liu, Y. Dong, and Y. Jia, "Vehicle Type Classification Using Unsupervised Convolutional Neural Network," in 2014 22nd International Conference on Pattern Recognition, pp. 172–177, Aug 2014.

[10] C. M. Bautista, C. A. Dy, M. I. Mañalac, R. A. Orbe, and M. Cordel, "Convolutional neural network for vehicle detection in low resolution traffic videos," in Region 10 Symposium (TENSYMP), 2016 IEEE, pp. 277–281, IEEE, 2016.

[11] H. Huttunen, F. S. Yancheshmeh, and K. Chen, "Car type recognition with deep neural networks," in Intelligent Vehicles Symposium (IV), 2016 IEEE, pp. 1115– 1120, IEEE, 2016.

[12] Y. Zhou, H. Nejati, T.-T. Do, N.-M. Cheung, and L. Cheah, "Image-based vehicle analysis using deep neural network: A systematic study," in Digital Signal Processing (DSP), 2016 IEEE International Conference on, pp. 276–280, IEEE, 2016.

[13] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, R. Cheng-Yue, F. Mujica, A. Coates, et al. An empirical evaluation of deep learning on highway driving. arXiv preprint arXiv:1504.01716, 2015.

[14] https://arxiv.org/pdf/1311.2524.pdf

https://arxiv.org/pdf/1504.08083.pdf

https://arxiv.org/pdf/1506.01497.pdf

https://arxiv.org/pdf/1506.02640v5.pdf

http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf

[15] B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," CoRR, vol. abs/1311.2524, 2013.

[16] Aysxegu Ucxar, Yakup Demir and Cu¨neyt Gu¨zelisx "Object recognition and detection with deep learning for autonomous driving applications" Transactions of the Society for Modeling and Simulation International 2017, Vol. 93(9) 759–769

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.

[18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[19] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, “Cnn-rnn: A unified framework for multi-label image classification,” in Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on, pp. 2285–2294, IEEE, 2016.

[20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Region-based convolutional networks for accurate object detection and segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no.1, pp.142–158, Jan.2016.

[21] Amir Mukhtar, Likun Xia, Tong Boon Tang “Vehicle Detection Techniques for Collision Avoidance Systems: A Review” IEEE Transactions on Intelligent Transportation Systems Volume: 16 ,Issue: 5 , Oct. 2015)