

Enhanced Vehicle Detection and Identification Using Histogram of Oriented Gradients

M Poornima, B.Tech (IT), M.Tech(IT) [@], Thamba Meshach W (PhD) [#]

[@] Assistant Professor, Department of Computer Science and Engineering,

[#] Associate Professor, Department of Computer Science and Engineering,
Prathyusha Institute Of Technology And Mangement, Chennai,

Abstract— Object detection and classification are necessary components in an artificially intelligent autonomous system. It is expected that these artificially intelligent autonomous system venture on to the street of the world, thus requiring detection and classification of car objects commonly found on the street. The identification and classification of object in an image should be faster and accurate. The aim of the proposed system is to detect the object as soon as possible with better accuracy and improved performance even if the object varies in appearance. Object identification and classification is a challenging process when the object of same category with large variation appears. Though number of papers deal with appearance variation, object detection process is considered to be slower. In proposed work, the detection speed is improved by using the optimized features, the object is detected and the object type was identified.

Keywords— Detection, Classification, Multi-posed vehicle

I. INTRODUCTION

Object detection is a very complex problem that includes some real hardcore math and long tuning of parameters to the computation methods. Object detection and classification are necessary components in an artificially intelligent autonomous system. Especially, object classification plays a major role in applications such as security systems, traffic surveillance system, target identification, etc. It is expected that these artificially intelligent autonomous system venture on to the street of the world, thus requiring detection and classification of car objects commonly found on the street. In reality, these classification systems face two types of problems. i) objects of same category with large variation in appearance ii) the objects with different viewing conditions like occlusion, complex background containing buildings, trees, people, road vies, etc.. This paper tries to bring out the importance of the feature extraction. Thus, we use two different methods for feature extraction and the performance of these two methods were analysed to find the efficient feature extraction method for detecting the object.

The existing system deals with whole bank of detectors for the given input image. The contribution of our object detection method is feature selection i.e. selecting a small number of relevant features for learning purpose. Feature selection provides an effective learning algorithm and strong bounds on generalization performance. The major contribution of object detection is a method which dramatically increases the speed of the detector by focussing attention on promising

regions of the image. Also it gives better performance for object detection by using two different methods of feature extraction.

Our object detection procedure classifies images based on the value of simple features. There are many motivations for using features rather than the pixels directly. The most common reason is that features can act to encode ad-hoc domain knowledge that is difficult to learn using a finite quantity of training data. For this system there is also a second critical motivation for features: the feature-based system operates much faster than a pixel-based system.

The research for object detection and recognition is focusing on

- 1) **Representation:** How to represent an object.
- 2) **Learning:** Machine Learning algorithms to learn the common property of a class of objects.
- 3) **Recognition:** Identify the object in an image using learning models.

In our proposed work, the object is detected as quickly as possible and the detection speed is improved by using the optimized detectors i.e. small subset of detectors for the given input. Also, the multi-posed vehicle is detected for small variation of the rotation angle. Moreover, object is identified and denoted what type of object it is, in a given video.

Initially, we worked on static images and it contained the following modules. i) background elimination, ii) feature extraction, iii) feature selection, iv) training, v) testing

In background elimination technique, the background is eliminated by region filling and morphological operations. In feature extraction, features are extracted using Principle Component Analysis (PCA) and Histogram of Oriented Gradients (HOG). In feature selection, optimized features are selected using Adaptive Boosting technique (AdaBoost). Then, the system is trained with car images and non-car images. The trained features are then classified. In Testing Module, to classify the objects, the Support Vector Machine (SVM) classifier is used. The trained features are then classified as the car image and non-car image. After classification, the query image i.e. the image to be tested is given as input. Then the features are extracted and it is used as test feature. After the features are extracted, the Classification is done likewise. Then the object is classified by performing the above process. Finally, we tested the performance of the

system by using both PCA and HOG methods. By analyzing the performance of these two methods, we found that HOG technique performed better than PCA method. So, we used HOG technique for further video image classification. Then we proceeded our by taking the videos.

The work composes of four different modules there are

1. Object Segmentation Module.
2. Feature Extraction and Feature Selection Module.
3. Training Module.
4. Testing Module.

In Object Segmentation Module, the background is eliminated by frame differencing method and the target object of interest is obtained. In the Feature Extraction and Feature Selection Module, the features are extracted and the relevant features are selected from the object of interest. For feature extraction, two methods were used to compare the efficiency i.e. Histogram of Oriented Gradients (HOG) and Zernike algorithm. After extracting the features using these two methods, feature selection was performed using Adaptive Boosting technique. Then, the relevant features are obtained after performing feature selection. In Training Module, the relevant features are used to train the system so that it identifies what type of object it is. In Testing Module, the video to be tested is given as input and then it is subjected to object segmentation, feature extraction, feature selection and then the object type is identified.

II. RELATED WORKS

In vehicle Objects Detection of Video Images Based on Gray-Scale Characteristics [1], first the color images are converted to gray-scale images. Then the methods of frame differencing and selective background updating are utilized to generate initial background and update current background. Furthermore, every processed image is filtered by fast median filter to remove noise. When the current background is obtained, moving objects in the video can be detected effectively by background frame differencing. Finally, morphological filtering is used for decreasing accumulative errors. However, false detection also happens when vehicles adhere to each other.

In Cluster Boosted Tree Classifier for Multi-View, Multi-Pose Object Detection [2], a Cluster Boosted Tree (CBT) learning algorithm was introduced to automatically construct tree structured object detectors. Instead of using predefined intra-class sub-categorization based on domain knowledge, they divide the sample space by unsupervised clustering based on discriminative image features selected by boosting algorithm. The sub-categorization information of the leaf nodes is sent back to refine their ancestors' classification functions. Their learning algorithm does not limit the type of features used. New features could be integrated to the framework easily.

In Rapid Object Detection using a Boosted Cascade of Simple Features [3], they have presented an approach for object

detection which minimizes computation time while achieving high detection accuracy. This approach is 15 times faster than any previous approach. They worked on three key contributions. 1) A representation of new image called the "Integral Image" which allows the features used by our detector to be computed very quickly. 2) A learning algorithm, based on AdaBoost, which selects a small number of critical visual features from a larger set and yields extremely efficient classifiers. 3) A method for combining increasingly more complex classifiers in a "cascade" which allows background regions of the image to be quickly discarded. Experiments on such a large and complex dataset are difficult and time consuming.

In Sharing features: efficient boosting procedures for multi class object detection [4], they have introduced a joint boosting algorithm, for jointly training multiple classifiers so that they share as many features as possible. The result is a classifier that runs faster and requires less data to train. They have applied the joint boosting algorithm to the problem of multi-class, multi-view object detection in clutter. An important consequence of joint training is that the amount of training data required is reduced. When reducing the amount of training, some of the detectors trained in isolation perform worse than chance level.

In Fast Pose Estimation with Parameter Sensitive Hashing [5], they presented new hash-based searching techniques to rapidly find relevant examples in a large database of image data, and estimates the parameters for the input using a local model learned from those examples. But the learning algorithm, implicitly assumes independence between the features; they are exploring more sophisticated feature selection methods that would account for possible dependencies.

In a trainable object detection system for static images [6], results are shown for car detection. The system uses a representation based on Haar wavelets that captures the significant information about elements of the object class. When combined with powerful classification engine i.e. the support vector machine, they obtain a detection system that achieves accuracy with low rates of false positives. Due to the significant change in the image information of cars under varying viewpoint, developing a pose invariant car detection system is likely to be more difficult than pose invariant people detection

III. SYSTEM MODEL

OBJECT IDENTIFICATION AND CLASSIFICATION

In our proposed work, the object is detected and the detection speed is improved by using the optimized detectors i.e. small subset of detectors for the given input. Also, the multi-posed vehicle is detected for small variation of the rotation angle. Moreover, in a given video, object is identified and described what type of object it is. This can be shown by the following modules and it is diagrammatically shown in the following diagram.

Then the system is trained with 30 videos and the trained system is tested with 12 videos. The system architecture diagram of Object Identification and Classification system for the video is shown in Figure 3.1.

For multi-view object detection, the video frames of each video are to be trained with SVM classifier. Then, if the query video is given, it detects the object in the video and identifies what type of object it is.

A. Object segmentation

In Object segmentation Module, first the original video files were converted into frames. Then the first frame is subtracted from the second frame and the second frame is subtracted from the third frame and the third frame is subtracted from the second frame and this process continues until all the frames are completed. Finally, the subtracted video frame is mapped to obtain the object of interest.

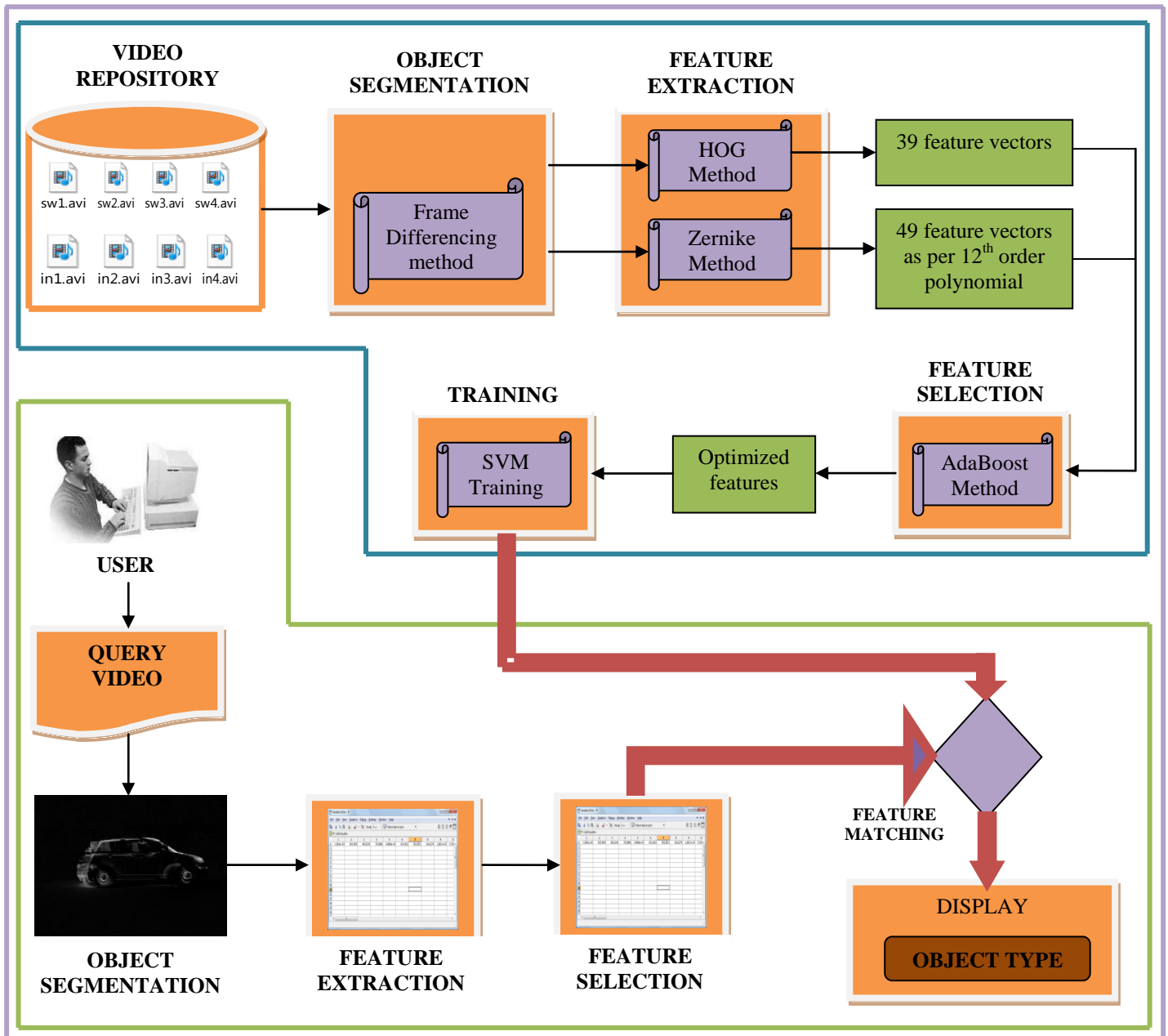


Figure 3.1 Architecture of Object Identification and Classification for videos

Let us consider a video. This video is first converted into frames and then the following subtraction technique was used to obtain the object of interest. Assume that frame with car image as 'C' and frame without car image as 'W'. Object of interest is obtained by using the following formula.

$$\text{Object of interest} = [C - W] \quad (1)$$

This process continues for each and every frames obtained from the video to get the object of interest. That is the second frame is subtracted from the first frame and the third frame is subtracted from the second frame and so on.

The screenshot of the object segmentation module is shown below which clearly pictures the background elimination and frame differencing between the frames.



By performing object segmentation, we eliminate the occluded background from the original video frame and subject the object of interest to the next module. Then we perform feature extraction for extracting the features of the object based on their shape and appearance.

B. Feature Extraction And Feature Selection

In the Feature Extraction and Feature Selection Module, first the features are extracted for the segmented object and then the relevant features are selected from the object of interest. For feature extraction in videos, we used two methods to compare the efficiency .i.e. Zernike and Histogram of Oriented Gradients (HOG). And for feature selection, AdaBoost technique was used to select the optimised features from the extracted features.

Algorithm for HOG technique is given as follows

Begin

- 1) Convolve the image using Gaussian filter. Channel with largest magnitude gives gradient magnitude.

- 2) Divide image window into dense uniformly sampled grid of points.
- 3) Image window is represented as block which consists of 2*2 cells.
- 4) Each cell consists of a 9-bin HOG. For each pixel in the cell, use trilinear interpolation to vote into the 9- bin histogram.
- 5) Thus, each block is represented by a 36-D feature vector.
- 6) Apply normalization to each block to improve performance.

End

Complex Zernike moments are constructed using a set of complex polynomials which form a complete orthogonal basis set defined on the unit disc $(x^2+y^2) \leq 1$. They are expressed as A_{pq} two dimensional Zernike moment:

$$A_{nl} = \frac{n+1}{n} \int_0^1 \int_0^{2\pi} [V_{nl}(r \cos \theta, r \sin \theta)]^* f(r \cos \theta, r \sin \theta) r dr d\theta \quad (2)$$

Where $n = 0, \dots, \infty$ defines the order, $f(x, y)$ is the function being described and $*$ denotes the complex conjugate. While ' $|l| \leq n$ ' is an integer (that can be positive or negative) depicting the angular dependence, or rotation.

Algorithm for Zernike technique is given as follows

Begin

- 1) Initialize the weight.
- 2) Multiply the weight with each feature vector.
- 3) Calculate the error factor for each feature vector.
- 4) Sort the resultant feature vector in order.
- 5) Neglect the feature vector which has high error rate to get the relevant features.

End

After extracting the features using these two methods, feature selection was performed using Adaptive Boosting technique. In Adaptive boosting method, weight is assigned for each feature and error rate was found and we eliminate the feature vectors having high error rate. We get the relevant features after performing feature selection.

The extracted features are then optimized .i.e. the relevant features are identified to increase the speed of object detection. Then the optimized features are sent to the next module for training. The relevant features are trained and the trained features are stored in the database for future comparison and classification of the query video.

C. SVM Training

In Training Module, the relevant features are used to train the system and classify the images. The system is trained with SVM trainer to detect the query video. We use in-built Matlab code for training purpose. The optimized features are trained and these features are classified by SVM classifier and stored in the database for comparing and identifying the query video.

This system is trained with two types of cars and totally 30 videos are trained.

D. Object Classification

Query video is given as input to the system and we tested the system with 12 videos overall.

The input video is first converted to frames and the background is segmented from the foreground by frame differencing method and the foreground object is subjected to feature extraction for extracting the features and the extracted features are given to feature selection module to get the optimized features and these features are compared with the features stored in the database. Then it identifies the object by matching the features in the database.

We use two types of car (i.e. indica and swift) for classification and we assign 'index 0' for indica features and 'index 1' for swift features. If the query video is given, the extracted features of a query video are matched with the trained features to detect the object type.

IV. RESULTS AND PERFORMANCE ANALYSIS

For the object identification and classification of the videos, we train the system with two types of cars (Indica and Swift) and tested the query video. Then the query video is tested and the result is shown as indica or swift.

The Accuracy of the results depends upon the amount of training and testing items. The Training set is of about 30 videos and the Testing set is of about 12 videos. Performance is directly proportional to the availability of training videos.

The classification performance of the system using both HOG and Zernike methods are tabled in Table 4.1. The performances of these methods were compared by taking into account, the number of cars tested and the true positives obtained.

Table 4.1 Predicated result based on HOG and Zernike methods

Methods of feature extraction	No. of cars tested	True +ve
HOG method	12	8
Zernike method	12	10

The processing time of HOG method and Zernike method are shown in the following table 4.2.

Table 4.2 Processing time for HOG and Zernike methods

Methods of feature extraction	Processing time per video
HOG method	45 seconds
Zernike method	15 seconds

Figure 4.1 Comparison between HOG and Zernike methods

After analyzing the result, the processing time is more for HOG method and less for Zernike method. Though, the processing time also depends on the video size. Hence, the performance of the system is better using Zernike when compared to HOG method.

V. CONCLUSION AND FUTURE WORK

Thus, initially we experimented on images in which we reject the background patches by using background subtraction and the features are extracted by two techniques as HOG and PCA. Then, used small subset of detectors for efficient detection and used SVM classifier for training the images. Thus, the speed of image classification done using HOG technique is better.

Then the work goes on for multiple orientations of an image for better accuracy and for increasing the speed of detection. To enhance the model, work was done on videos where the object is segmented from the background. Then the object type is found using two techniques as HOG and Zernike methods. By comparing both the methods, Zernike performed faster and better than the HOG method.

The future work can be of identifying the object type correctly even when the video is blurred and tracking the identified object in controlled traffic system. Moreover, object detection can be carried over for all types of vehicles passing the road in the controlled traffic system.

REFERENCES

- [1]. Jie Cao, Li Li. "Vehicle Objects Detection of Video Images Based on Gray-Scale Characteristics". *First International Workshop on Education Technology and Computer Science*, pp. 937-940, 2009.
- [2]. B. Wu and R. Nevatia. "Cluster boosted tree classifier for multi-view multi-pose object detection". In *Proc. IEEE International Conf. on Computer Vision*, pp. 1-8, 2007.
- [3]. P. Viola and M. Jones. "Robust real time object detection". *International Journal of Computer Vision*, 57(2), pp. 137 - 154, 2004.
- [4]. Torralba, K. Murphy, and W. Freeman. "Sharing features: Efficient boosting procedures for multiclass object detection". In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1-8, 2004.
- [5]. G. Shakhnarovich, P. Viola, and T. Darrell. "Fast pose estimation with parameter-sensitive hashing". In *Proc. IEEE International Conf. on Computer Vision*, pp. 5-8, 2003.
- [6]. C. Papageorgiou and T. Poggio. "A trainable system for object detection". *International Journal of Computer Vision*, 38(1), pp. 15-33, 2000.