

HOG based object detection and classification

N. Venkateswara Rao^{1*}, G. Anil Kumar², B. Harish³

¹ Professor, ECE Department

² M. Tech student, ECE Department

³ Research engineer

*Corresponding author E-mail: vrnagalla@gmail.com

Abstract

The intension of the project is to classify objects in real world and to tracks them throughout their life spans. Object detection algorithms use feature extraction and learning algorithms to classification of an object category. Our algorithm uses a combination of “histogram of oriented gradient” (HOG) and “support vector machine” (SVM) classifier to classify of objects. Results have shown this to be a robust method in both classifying the objects along with tracking them in real time world.

Keywords: Object Detection; Object Classification; “Histogram of Oriented Gradient (HOG)”; “Support Vector Machine” (SVM).

1. Introduction

The object classification is one branch of computer vision which is examined for past years. Detecting of objects and classification or recognition of objects are major task in computer vision. Object classification algorithms depends up on learning, matching, or pattern recognition algorithms using feature based or appearance feature based techniques. Common methodologies include deep learning approaches or techniques such as CNN (convolution neural networks), DNN (deep neural networks), and feature-based approaches using gradients, edges, Haar wavelets, HOG (histogram of oriented gradients), and LBP (local binary patterns).

The main purpose of the project is to detecting and classification of objects in real time, throughout their life spans. Object classification is needed for many computer vision applications including portable electronics, “advanced driver assistance systems” (ADAS), video surveillance, and navigation of robots [6]. In real time robot usage housekeeping purpose such as cleaning, fetching objects, etc.

These tasks need several levels of important knowledge about surrounding of humans so that the system can be find the different ways for searching the objects and they can conveying with persons easily. Like we can the example of robot to execute the command for taking bottle from kitchen, so for that robot must be identifying things like bottle. In addition, understand which places in the surroundings are expected to be consist it.

Some challenges faced on object detection. Firstly, almost every class of objects has large amount variations within the class. For example, for each human face, skin color face shape, and hair style, etc. change importantly and differences in beard, mustache and wearing spectacles and hat result in further changes. As for cars are taken the color, the model and the size also change substantially between each another. A robust classifier must try to accomplish independent of these variations. Secondly, background clutter is one common issue. The images may be taken from various background environment like outside scenes in cities and interior environments. The classifier must be capable to differentiate object class from those complex backgrounds. Thirdly, an illumination

condition varies a lot, ranging from direct visible light and also shadows during the day to artificial or dim lighting inside buildings or at night. Although some solutions for illumination invariance have been adopted, they are still extremely inefficient when compared to human visual systems in being adaptive to such changes. Hence, a robust classifier of object poses should also contribute extensive invariance to the alteration of illumination and lighting conditions. Noise can be avoided or minimized by noise removal filters. The major aim of feature extraction is to extract the features, which maximizes the identification rate with the redundant information of data.

2. Literature survey

Object classification is still a challenging task of computer vision. There are many useful approaches and some of the more popular, such as SIFT [2]; HOG [1] are depending on image gradients. The majority of implementations for HOG-based object detection are on “Central Processing Unit” (CPU) and “Graphics Processing Unit” (GPU) platforms.

Viola-Jones proposed object detection method [3] which is extensively used in recent years. This process is machine learning based and a cascade function is used which is trained on large amount of image data (both negative and positive). It can be skilled to detect various object classes, but the main motivation was to recognition faces in images. This approach provides first real-time face detection with solid results.

Scale Invariant Feature Transform (SIFT) [2] is one of the most familiar approaches which gives invariant features from images and its detail description. The features obtained can be used to a lot of number of applications (e.g. Motion tracking, image stitching, multi-view geometry). According to [3], the extracted features from images are never changing to image scaling and image rotation, which can evaluate good matching between various perspective appearances of an object or a scene. Also, the features have on exhibited to be robust against illumination variance, affine distortion, and additional noise. The main benefit of SIFT is to provide viewpoint invariance, rotation, illumination and scale.

LBP facial expression recognition and also got promising performance [7]. In this paper texture using volume “local binary patterns” an extension to LBP, for expression recognition [8]. Average FER accuracy of 96.26% was achieved for six universal expressions with their proposed model on cohn -Kanade facial explanation database [9]. Due to both spatial and temporal information is considered in VLBP, it got better result comparing traditional LBP. The review paper [10] gives object detection based on CNN, construction of CNN, structure of object detection on CNN and the improving detection performance.

It gives [11] neural network based face detection. Neural network research tiny windows of an image and each window contains determine a face or not. The system intermediate between multiple networks gives better performance when compared with a single network. Face and non faces images are holding for training and also use bootstrap algorithm.

This paper gives an “object detection and classification”. The remaining of this paper is standardized as follows. Section 3 explains methodology of object detection and classification and Section 4 presents brief theory of HOG. Section 5 presents brief theory of SVM section 6 explains the experiments and results and section 7 presents the conclusions.

3. Methodology

It consists of two stages such as training stage and testing stage. Basically the system doesn't know the computer vision skills to detect whether it is an object or not. So, we need to make the machine to learn computer vision skills in order to identify and detect objects. This process of training is known as training stage.

In this training stage, initially we need to generate a model, by using positive images and negative images. These positive images refer to desirable objects (which object you find the images along with video) together with negative images refer to undesirable objects (background scenes images). The feature extracted from training images using HOG algorithm following steps. There are preprocessing, gradient calculation, spatial binning, block normalization, feature vector. And finally features are extracted from all training images. Brief explanation of HOG algorithm take up in section IV. After extracting features designate labels to known (positive images) objects, and also assign the labels to unknown objects. Finally classification is done by using SVM algorithm. The justification of SVM explained in section V. The generated trained model can be used in testing phase for object detection and classification.

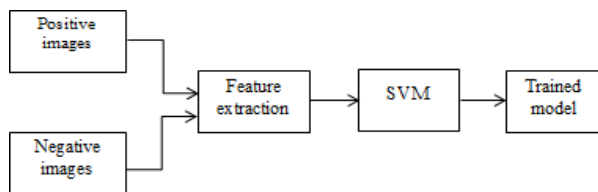


Fig. 1: Block Diagram for Training Phase of the “Object Detection and Classification”.

Testing stage starts with a test video which is divided into frames. These frames are sent over to the Feature extraction for the extraction of features for each frame in the test video or test image by using HOG algorithm. Now these features are classified by using SVM algorithm. The outputs of the SVM are allowed to match with the trained model which consists of sample features of images. If the correct match is found in the trained model then the most desirable output is obtained by indicating the object in the given test video or test images.

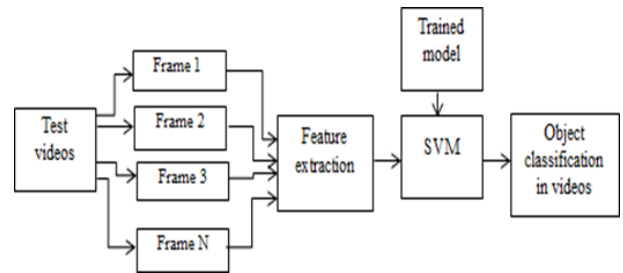


Fig. 2: Block Diagram for Testing Phase of the Object Detection and Classification.

4. Theory of “histogram of oriented gradients” (HOG)

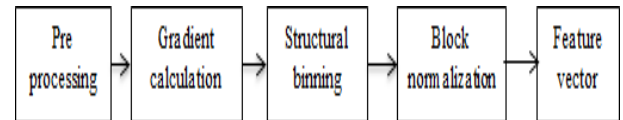


Fig. 3: Block Diagram of “Histogram of Oriented Gradients” (HOG).

HOG has been used as a feature descriptor in the application of object detection, where a gradient orientation plays a leading role. The thought behind the HOG descriptor is object shape and appearance within an image. It can be characterized by sharing of intensity gradients. The image is partitioned into small connected areas called cells. The each pixel in cell calculates the magnitude and direction. For increasing accuracy, normalization is required. The block normalization is invariant to changes in the shadowing and illumination.

4.1. Preprocessing

Filtering, dilation, erosion, power law correction is used in the preprocessing step of input. Normalization is used for to decrease the impact of illumination effects. It consists of power law (gamma) correction which is helpful to reduce or decrease the effect of illumination variations and shadowing. Preprocessing is an optional step. It can be expunged, because it doesn't have strong effect on the result when calculating gradient. And images can be resized.

4.2. Gradient calculation

Gradient is a vector and it's contains two variables: direction and magnitude. These two variables are computed differently, but both are computed through convolution process. In the convolution process, find horizontal filter and vertical filter by using sobel x and y filters. We notice that horizontal filter sharpen the vertical edges and vertical filter sharpen the horizontal edges

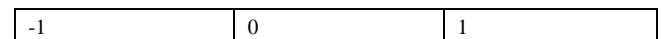


Fig. 4: (A) Shows the Horizontal Kernel.

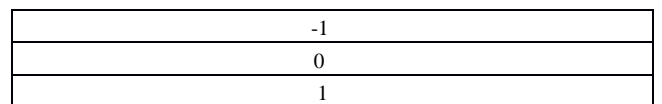


Fig. 4: (B) Shows the Vertical Kernel.

The gradient's gradient is perpendicular to the edge's direction. Magnitude of gradient contributes the strength of the edge. The gradient's direction is defined as:

$$\theta = \arctan\left(\frac{H_y}{H_x}\right) \tag{1}$$

The gradient's strength is defined by gradient's magnitude:

$$M = \sqrt{H_x^2 + H_y^2} \quad (2)$$

When gradients for all pixels are computed, pixels are grouped into cells. For each cell is afterward histogram of oriented gradients created. Image converted into cells why because provides compact information.

4.3. Spatial binning

Each 8x8 cell image consist of 8x8x3 = 192 pixel values but the gradient of this 8x8 cells image contain 2 values (first one magnitude and other one direction) per pixel which adds up to 8x8x2 = 128 values. The 128 values are indicated by using a 9-bin histogram which can be hold as an array of 9 values. The histogram is essentially a vector of [9] bins (values) corresponding to angles 0, 20, 40 160. Each pixel denotes to histogram for spatial regions called cells. Angle of gradient is in range of 0-360 degrees is called signed gradient or 0-180 degrees is called unsigned gradient. This range is divided into 9 blocks called bins. In case of electing range 0-180 degrees the bins are each of bin size is 20 degrees [5]. The range 0-180 degrees are called "unsigned gradients" why because its negative is represented by the same values. Unsigned gradients work better as compared with signed gradients. A bin is selected depend up on the direction, and the vote (the value that go to the bin) is elected based on the magnitude.

4.4. Block normalization

5. Theory of "support vector machine" (SVM)

"Support Vector Machine" (SVM) classifier is a one of the "supervised machine learning" algorithm. SVM is a famous technique for the classification and regression. In "supervised machine learning" algorithm, objects belonging to a known class are used for "training" of the system and draw decision lines between various classes. New objects are classified based on the decision lines. In this technique are handling with the training data and testing data [4]. Now the system should be trained with the help of the training data such that it can help the system to indicate or classify the object in the image or video.

Some important features of SVM:

- 1) In the linear SVM, an optimization criterion is the width of the margin between the two classes.
- 2) This algorithm basically two class classifier but can be elongated to multiple classes.
- 3) Mapping points or samples to a higher dimensional space where various classes can be separated with a hyperplane by using kernel trick
- 4) The performance of SVMs rely on the preferred of the kernel and its parameters
- 5) SVM can be divided into two types
 - (a) Linear SVM
 - (b) Nonlinear SVM

The "support vector machine" (SVM) is basically a binary classification method introduced by Vapnik. For a binary problem, we have training data points.

$$\{S_i, T_i\} i = 1, \dots, N \quad (3)$$

$$T_i \in \{-1, 1\} \quad (4)$$

$$S_i \in \mathcal{A}^\alpha \quad (5)$$

Suppose we have some hyperplane which can be divides the positive image features from the negative image features. The points z which lie on the hyper plane live up to a. z + s = 0, where a is normal to the hyper plane, |s|/||a|| is the perpendicular distance from the hyper plane to the origin, and ||a|| is the Euclidean norm of a. In this

algorithm, the classes separated by hyper plane with largest margin. Suppose that all the training data mollify the following formulas This step procures invariance to edge contrast, shadowing and illumination. Cells are combined into blocks which can be over- lapped or not.

5.1. Feature vector

This step indicates to collecting HOG descriptors from all blocks transformed into a feature vector. This method used Support Vector Machine (SVM) classifier and it performs well to detect objects.



Fig. 5: (A) Shows the Original Image Fig.5. (B) Shows the HOG Image of the Original Image.

$$S_i x + C \geq +1 \text{ For } Y_i = +1 \quad (6)$$

$$S_i x + C \geq -1 \text{ for } Y_i = -1 \quad (7)$$

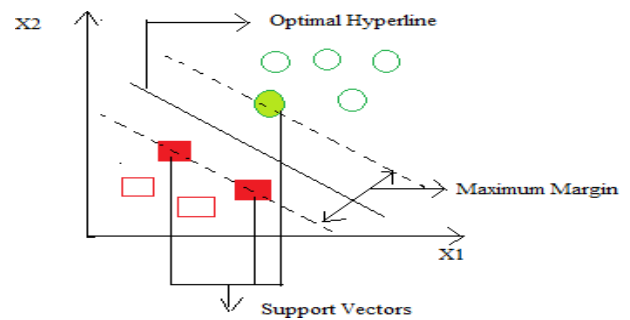


Fig. 6: Classification of Two Classes by SVM.

5.2. Nonlinear SVM classifier

In the real world our dataset is specifically dispersed up to some extent. To clarify this problem segregation of data into various classes on the basis of a linear hyperplane can't be treated as a good choice. Then vapnik implied making nonlinear SVM classifier, feature points plotted in a bigger or higher dimension space.

5.3. Kernel

Kernel used in SVM algorithm. In linear SVM, the learning of the hyper plane problem is done by using linear algebra. The SVM can be explicit by the inner product of any two given observations

$$K(Y_S, Y_T) = F(Y_S) \cdot F(Y_T) \quad (8)$$

This above equation is called kernel function

$$K(Y_S, Y_T) = Y_S \cdot Y_T - \text{Linear Kernel} \quad (9)$$

$$K(Y_S, Y_T) = (\gamma Y_S \bullet Y_T + Z)^f \rightarrow \text{Polynomial kernel} \quad (10)$$

$$K(Y_S, Y_T) = \exp(-\gamma |Y_S - Y_T|) \rightarrow \text{RBF kernel} \quad (11)$$

$$K(Y_S, Y_T) = \tanh(\gamma Y_S \bullet Y_T + Z) \rightarrow \text{Sigmoid kernel} \quad (12)$$

It denotes conversion of dot product of features (“data points”) mapped into the bigger dimensional feature space by using transformation

Margin The margin is essential characteristic of SVM classifier. SVM tries to accomplish a good margin. A margin is a segregation of line to the related class points. The good margin represents separation of classes with large gap and also equidistant as long as both classes.

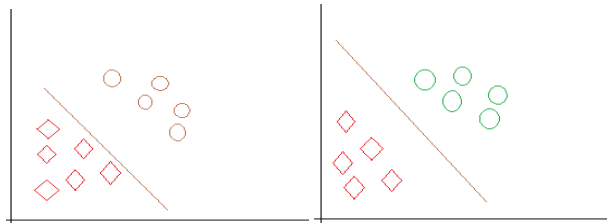


Fig. 7: Shows the Bad Margin and Good Margin.

5.4. Regularization

The regularization parameter (in python’s sklearn frequently called as C parameter) tells the SVM improvement how much you want to fend off misclassifying each training example. For larger C, the optimization will pick out a smaller margin than hyper plane a good task of acquiring all the “training points” classification correctly. Conversely smaller C the optimization will pick out a bigger margin then misclassification occurs.

5.5. Gamma

The gamma parameter defines how far the effect of single training examples reaches, with low values meaning ‘far away’ and high values meaning ‘close’.

6. Experimental results

According to the analysis of “object detection and classification”, Firstly we should work on SVM classifier. In the training, we take the different known classes images (In testing, which object to find out) and different unknown classes images (none of images such as background).we prepare the own dataset for training. Each 1000 images of known classes with different orientations, different lighting conditions and unknown classes.



Fig. 8: Detection of Persons in Test Image.



Fig. 9: Detection of Cow in Test Image.



Fig. 10: Detection of Aero Plane in Test Image.

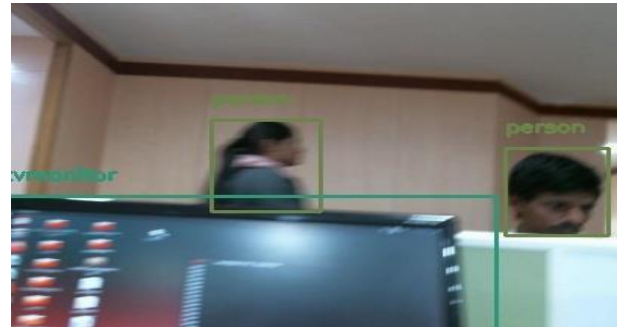


Fig. 11: Detection of Monitor and Persons in Test Video.

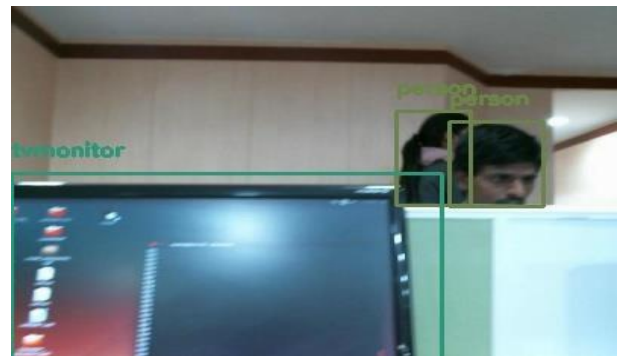


Fig. 12: Detection of Monitor and Persons in Test Video.

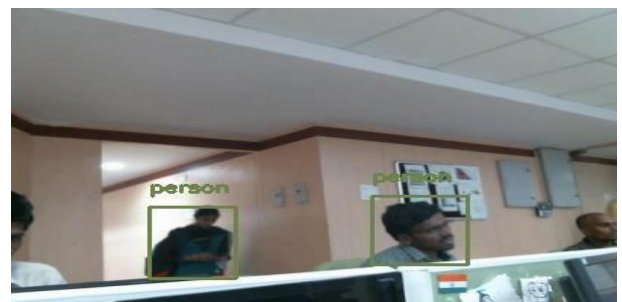


Fig. 13: Detection of Persons in Test Video.

Table 1: Object Classification on Images

Results Test Images	Objects with known classes (%)	Objects with un- known classes (%)
Objects with known classes	88	12
Objects with un-known classes	7	93

The table.1 shows the object classification in images. In object detection and classification by HOG and SVM algorithms 88% accuracy on objects with known classes and 12% false positives occurs. Similarly in the case of unknown classes we have 93% accuracy on objects with unknown classes 7% false positives occur.

Table 2: Object Classification on Real Time Videos

Parameters	Video 1	Video 2	Video 3
Total no. of objects	49	53	75
No. of Detected objects	42	43	65
Accuracy of classification (%)	85	81	86

7. Conclusion

The “object detection and classification” is done with HOG and SVM algorithm. The results are obtained for the classification of objects for both test image and test video inputs. From the results a good level of accuracy up to 88 percent is achieved in “object detection and classification” for test images and on the other hand 85percent accuracy is obtained for test video. In testing videos the classification accuracy is good for moving objects are classified correctly. The accuracy level in detecting the objects is more for test images then compared to the test video. The level accuracy level can be improved by choosing training images more taken in the trained model.

References

- [1] DALAL, N. – TRIGGS, B. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005.CVPR 2005.IEEE Computer Society Conference on, volume 1, pages 886–893 vol. 1, June 2005.
- [2] LOWE, D. G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal Computer Vision. November 2004, 60, 2, pages 91–110. ISSN 0920-5691.
- [3] VIOLA, P. – JONES, M. J. Robust Real-Time Face Detection. International Journal of Computer Vision. May 2004, 57, 2, pages 137–154.ISSN 0920-5691.
- [4] N. Cristianini, J. Shawe-Taylor. An Introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000. <https://doi.org/10.1017/CBO9780511801389>.
- [5] Histogram of Oriented Gradient, Pattern Recognition Systems – Lab
- [6] W. Shao, W. Yang, G. Liu, and L. J., —Car detection from high-resolution aerial imagery using multiple feature IGARSS, 2012, pp. 4379–4382.
- [7] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing, 27:803–816, 2009. <https://doi.org/10.1016/j.imavis.2008.08.005>.
- [8] G. hao and M. Pietik ainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transaction on Pattern Analysis and Machine Intelligence, 29:915–928, 2007.
- [9] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In Fourth IEEE International Conference on Automatic face and Gesture Recognition (FG’00), pages 46–53, 2000.
- [10] Zinal K Naik and Monali R Gandhi. A Review of Object Detection Based on Convolutional Neural Network. Proceedings of the 36th Chinese Control Conference July 26-28, 2017.
- [11] Rowley, H. A., Baluja, S. and Kanade, T.: Neural Networks Based Face Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1):2238, 199.