ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

Pedestrian Detection-A Comparative Study Using HOG and COHOG

Sujith B [1], Jyothiprakash[2]

Department of Computer science, Central University of Kerala, Central University of Kerala, India [1,2]

ABSTRACT: Pedestrian accidents still represent the second largest source of traffic related injuries and fatalities after accidents involving passenger cars.Pedestrian detection is a key problem in computer vision, with several applications that have the potential to positively impact quality of life. In recent years, many pedestrian classification approaches have been proposed. The pedestrian classification consists of two stages: feature extraction and feature classification. Recently several robust feature extracting methods have been proposed in literature like Scale Invariant Feature Transform (SIFT) , Histogram of Gradients (HOG) , Co-occurrence of Histogram of Gradients (CoHOG) . Also several classifiers exists like Hidden Markov Model (HMM), Support Vector Machines (SVM), and Neural Network. In this paper, we examine the two feature extraction method and we use neural network as classifier instead of SVM. An extensive evaluation and comparison of these methods are presented. The advantages and shortcomings of the underlying design mechanisms in these methods are discussed and analyzed through analytical evaluation and empirical evaluation.

KEYWORDS: Pedestrian detection, object detection, HOG, CoHOG, Computer Vision

I. INTRODUCTION

Computers have become a necessity in our daily lives. They perform tasks like heavy computational and data intensive very efficiently and more accurately than humans. People are trying to extend their capabilities so that they perform high level tasks that humans perform with so much ease that we don't even realize that we are performing them. Computer vision aims to duplicate the effect of human vision by electronically perceiving and understanding an image.

Detection of humans from images is a difficult task due to their variable pose, clothing, as well as varying backgrounds and environmental conditions. It is important in many applications such as Intelligent Vehicles (IVs), Intelligent Transport System (ITS), Driver assistance, surveillance, robotics and intelligent vehicles. According to WHO [2], 1.2 million people are known to die in road accidents worldwide. A majority of deaths and injuries involve motorcyclists, cyclists and pedestrians. In European Union about 8000 pedestrians and cyclists are killed and about three lac injured [3]. During 2001, there were 80,000 deaths on Indian roads, which grew in last decade at 5% per year [2].

Pedestrian accidents still represent the second largest source of traffic related injuries and fatalities after accidents involving passenger cars. The detection and classification of pedestrians is a difficult process [3]. The data captured by the camera will be searched and the features will indicate whether there exist pedestrians or not. Many pedestrian classification approaches have been proposed. The pedestrian classification consists of two stages: feature extraction and feature classification. The first need I s to have discriminative and robust features so as to distinguish between human and non human even in difficult illumination, varying pose, and deformations [4]

Pedestrian classification depends on the performance of both feature extracting techniques and classifiers. If the feature extracting technique would fail to extract the relevant features the classifier performance will be affected badly. This shows that there is a correlation between feature extraction and classification. Recently several robust feature extracting methods have been proposed in literature like Scale Invariant Feature Transform (SIFT) [5], Histogram of Gradients (HOG) [4], Co-occurrence of Histogram of Gradients (CoHOG) [6],. Also several

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

classifiers exist like Hidden Markov Model (HMM), Support Vector Machines (SVM), and Neural Network. Remember there is no best classifier that works best on all given problems.
Person detection is a challenging task, with many applications that has attracted lot of attention in recent years.

1.1 Challenges

The foremost difficulty in building a robust object detector is the amount of variation in images and videos. Several factors contribute to this:[7]

- Firstly, the image formation process suppresses 3-D depth information and creates dependencies on viewpoint such that even a small change in the object's position or orientation with respect to the camera center may change its appearance considerably. A related issue is the large variation in scales under which an object can be viewed. An object detector must handle the issues of viewpoint and scale changes and provide invariance to them.
- Secondly, most natural object classes have large within-class variations. For example, for humans both appearance and pose change considerably between images and differences in clothing create further changes. A robust detector must try to achieve independence of these variations.
- Thirdly, background clutter is common and varies from image to image. Examples are images taken in natural settings, outdoor scenes in cities and indoor environments. The detector must be capable of distinguishing object class from complex background regions.
- The previous two difficulties present conflicting challenges that must be tackled simultaneously. A detector that is very specific to a particular object instance will give less false detections on background regions, but will also miss many other object instances while an overly general detector may handle large intra-class variations but will generate a lot of false detections on background regions.
- Fourthly, object color and general illumination varies considerably, for example direct sunlight and shadows during the day to artificial or dim lighting at night. Although models of color and illumination invariance have made significant advances, they still are far from being effective solutions when compared to human and mammalian visual systems, which are extremely well adapted to such changes. Thus a robust object detector must handle color changes and provide invariance to a broad range of illumination and lighting changes.
- Finally, partial occlusions create further difficulties because only part of the object is visible for processing.

Figure 1.1 shows some examples illustrating these challenges for person detection. Figure 1.2 provides some instances where humans use reasoning and background information to prune false detections and to choose correct ones. Figure 1.3 shows few pairs of consecutive images from the INRIA database. Compared to Figure 1.1 it is having more variability in pose.
Another challenge is the amount of high-level context and background information that humans can deal with but that computers still lack.

Fig. 1.1. Some of the images from the INRIA data set with variation in pose, appearance, clothing, illumination, occlusion and background.

II. HISTOGRAM OF ORIENTED GRADIENTS

Histogram of Oriented Gradients (HOG) is feature descriptors used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

image. This method is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

Navneet Dalal and Bill Triggs [4], researchers for the French National Institute for Research in Computer Science and Control (INRIA), first described Histogram of Oriented Gradient descriptors in their June 2005 paper to the CVPR. In this work they focused their algorithm on the problem of pedestrian detection in static images, although since then they expanded their tests to include human detection in film and video, as well as to a variety of common animals and vehicles in static imagery [20].

The basic idea behind the HOG is that local object appearance and shape can be characterized well by the distribution of intensity gradients or edge directions. The implementation is simple, the image is divided into small cells, for each cell the 1D histograms of gradient orientations or edge orientations are collected for the pixels within the cell. The collection so these histograms represent the descriptor. For better performance the local responses can be contrast-normalized by calculating a measure of energy over larger spatial regions called "blocks" and using the results to normalize all the cells in the block. This normalized block is referred to as Histogram of Oriented Gradient (HOG) descriptors. This normalization helps in in better invariance to illumination, shadowing, etc. Fig. 3.3 shows an overview of the HOG method.

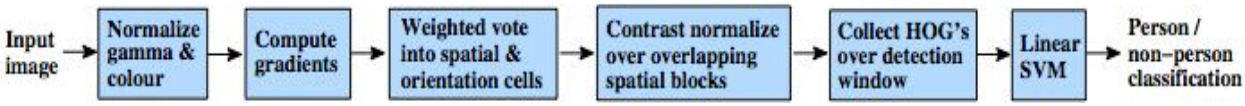

Figure 3.3 An overview of feature extraction using HOG. .

2.1. Neural network Classifier

The final and the last step in the object recognition using HOG descriptor is to feed the descriptors into some classifier. A neural network consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Once trained on images containing some particular object, the neural network classifier can make decisions regarding the presence of an object, such as a human being, in additional test images. In the Dalal and Triggs human recognition tests, they used the freely available SVMLight software package[7] in conjunction with their HOG descriptors to find human figures in test images. Here we have used neural network. We have used neural network of 100 neurons. The neural network is trained until the performance error becomes less than 0.01
Fig. 3.5 explains the methodology involved in calculating HOG descriptors. HOG has two advantages: First one is its robustness against illumination variance because gradient orientations of local regions do not change will illumination variance. Second advantage is its robustness against deformations. [6]

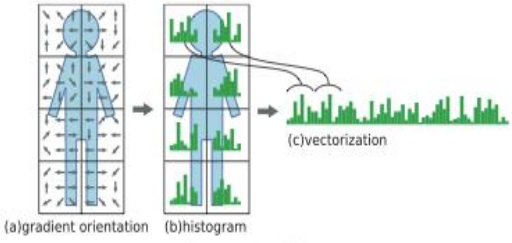

Figure 3.5 (a) (b) Overview of HOG calculation

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

III.CO-OCCURRENCE HISTOGRAM OF ORIENTED GRADIENTS (COHOG)

Co-occurrence Histogram of Oriented Gradients (CoHOG) given by Watanabe [6], is a multiple gradient orientation based feature descriptor. CoHOG's building blocks are pairs of gradient orientations. Since single gradient orientation has only eight varieties, but a pair of them have many more varieties. Thus CoHOG can express shapes in more detail than HOG, which uses single gradient orientation as shown in figure 3.6. Figure 3.6 (a) shows that a single gradient orientation has only eight variations and 3.6 (b) shows that pairs of orientations has more varieties than the single one.

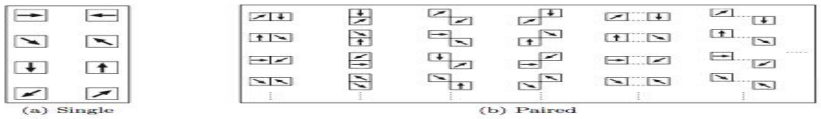

Figure 3.6 Vocabulary of gradient orientations.

Figure 3.7 shows an overview of Watanabe's CoHOG process. The first part calculates the pairs of gradient orientations from the input image. Then co-occurrence matrices are calculated in the second part. CoHOG builds the histogram on pairs of gradient orientations. This histogram is referred to as the co-occurrence matrix. The co-occurrence matrix is the distribution of gradient orientations at a given offset. The co-occurrence matrix for an *nxm* image separated by an offset (x,y)

$$C_{x,y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m} \begin{cases} 1, & if\ I(p,q) = i\ and\ I(p+x,q+y) = j \\ 0, & otherwise, \end{cases}$$

where I denotes a gradient orientation image and p and q denote gradient orientations. The last part classifies the result and determines whether the input image contains an object or not. As CoHOG is a gradient-based histogram, it is having same merits as those of HOG, which are robustness against deformation and illumination variance. This is because CoHOG is a gradient-based histogram feature descriptor. The process of CoHOG calculation is well shown in figure 3.7 (a)

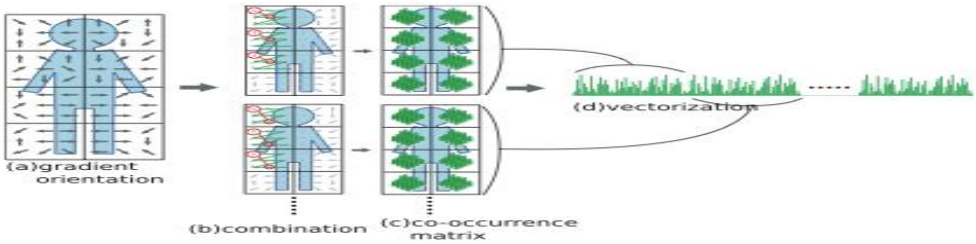

Figure 3.7 (a)(b) (c) (d) Overview of CoHOG calculation

Finally, the result of co-occurrence matrices is concatenated into a vector as shown in figure 3.7 (c). Algorithm for CoHOG calculation is shown in figure 3.8

```
given I: an image of gradient orientation
initialize H ← 0
for all positions (p, q) inside of the image do
    i ← I(p, q)
    k ← the small region including (p, q)
    for all offsets (x, y) such that corresponds neighbors do
        if (p + x, q + y) is inside of the image then
            j ← I(p + x, q + y)
            H(k, i, j, x, y) ← H(k, i, j, x, y) + 1
        end if
    end for
end for
```

.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

IV. EXPERIMENTAL RESULTS& DISCUSSIONS

As Co-HOG expresses shapes in detail it is therefore high dimensional. Watanabe et al showed in their paper [6] that Co-HOG is more informative than HOG because CoHOG has more effective values than HOG. CoHOG is calculated by incrementing the components of co-occurrence matrices, whereas HOG calculation is a complex process as it involves more procedures like weighted voting, histogram normalization, region overlapping. CoHOG achieves high performance without these complex procedures.
The feature size is 25 times smaller than the feature size (34704) reported **in**[6]. Figure 4.1 (a), (b), (c), (d), (e) shows respectively, an original image, resized image, grey image, gradient image and image after assigning the gradients orientations to bins.

Total feature size =No. of blocks(21) X Size of co-occurrence matrix (16) X No. of neighborhoods (4)= 1344

Figure 4.1(a) **b)** **(c)** **(d)** **(e)**

Training

We tested CoHOG on a challenging data set, 'INRIA',[21] which is widely used as human detection benchmark dataset. This dataset contains human images and non-human images consisting of 3030 positive images of different size and 4000 random images of size 128 x 64 negative images.. We have used neural network of 100 neurons. The neural network is trained until the performance error becomes less than 0.01. The confusion matrix of the training is shown figure 4.2. Figure 4.4 (a), (b), (c), (d) respectively shows the Right Operating Characteristics (ROC) of training, validation, testing and all ROC curves.

As shown in figure 4.4 (a), training ROC, we got 99.9 accuracy. The neural network is tested with a set of benchmarking data set and the results of classification are provided in the table below III. The meaning of measures provided in the table is explained in table II. For data set no. 1, 2, 3 and 4 the classification accuracies are 99.4, 97.7, 97.7 58.5 respectively. For dataset no. 4, the classification accuracy is 58.5; it is less probably because the set contains images of different sizes and also it contains groups of people.

TABLE I

S.No	Folder Name	Confusion Matrix	Accuracy (%)
1	96X160H96	Fig. 4.6	99.4
2	70X134H96	Fig. 4.7	97.7
3	test_64x128_H96_pos	Fig. 4.8	97.7
4	Test_pos	Fig. 4.9	58.5

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

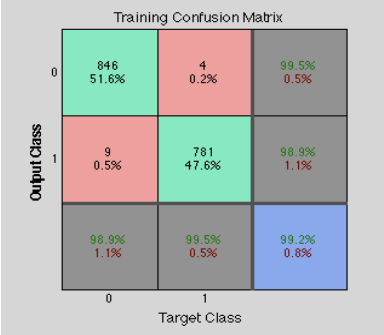

Figure 4.2 confusion matrix

Table II:

		Actual		
		Negative	Positive	
TestOutcome	Negative	True Negative	False NegativeMiss	Negative predictive value .Precision value= TN/(FN+TN)
	Positive	False Positive	True PositiveHit	Positive predictive value. Precision= TP/(TP+FP)

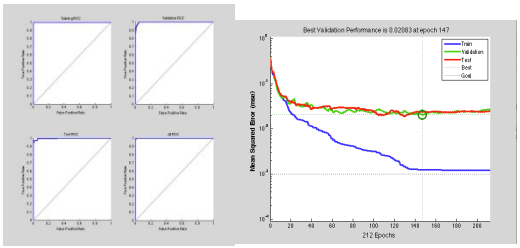

*Figure 4.4 ROCs of Training, Validation, Testing, and all ROCs*Figure 4.5

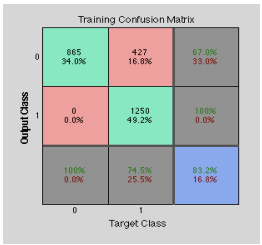

Figure 4.6 :Confusion matrix

We tested HOG on the same dataset INRIA [9], we used classifier as neural network instead of support vector machines as used by the original author Navneet Dalal. Confusion matrix of the training is shown in figure 4.6

The neural network is tested with the same set of benchmarking data set and the results of classification are provided in the table below III. The meaning of measures provided in the table III is explained as per table II. For data set no. 1, 2, 3,4 the classification accuracies are 98.9, 77.2, 97.7, and 44.6 respectively. For dataset no. 4, the classification accuracy is 44.6; it is less probably because the set contains images of different sizes and also it contains groups of people.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

TABLE III

S.No	Folder Name	Confusion Matrix	Accuracy (%)
1	96X160H96	Fig. 4.12	98.9
2	70X134H96	Fig. 4.13	77.2
3	test_64x128_H96_pos	Fig. 4.14	97.7
4	Test_pos	Fig. 4.15	44.6

TABLE IV

S.No.	Folder Name	Accuracy (%)	
		HOG	CoHOG
1	96X160H96	98.9	99.4
2	70X134H96	77.2	97.7
3	test_64x128_H96_pos	97.7	97.7
4	Test_pos	44.6	58.5

4.1 Comparison of HOG and CoHOG Results

Table IV shows the comparison of HOG and CoHOG feature extracting methods. From comparing methods, CoHOG outperforms HOG this is because CoHOG uses pairs of gradient orientations thus express shapes in more detail. In HOG the influence of various descriptor parameters and conclude that fine scale gradients, fine orientation binning, relatively coarse spatial binning, and high quality local contrast normalization in overlapping descriptor blocks are all important for good performance. By comparing HOG with CoHOG, CoHOG has less miss rate (i.e., the rate of human images classified as non-human) than half that of HOG. Also we use neural network for training and it gives good result in both.

V. CONCLUSION

In this paper we have studied the two well-known feature extracting methods, HOG, CoHOG using neural network. From experimental results we can say that CoHOG is robust method compared to the HOG. If the extracting method is robust then the further processing steps will be easy. CoHOG is the robust method against illumination variance, deformations, clothing, occlusion. CoHOG expresses local and global shapes in detail. The experimental results showed that the performance of CoHOG is better than the state of art methods (provided in the literature review) or at least comparable and consistently good on INRIA data set. Also we use neural network as classifier instead of SVM. It gives better result. In addition, CoHOG can be calculated 40% faster than HOG.

REFERENCES

[1] Milan Sonka, V. Hlavac, R. Boyle, "image processing, analysis, and machine vision" third edition Cengage Learning
[2] T. Gandhi and M.M Trivedi, "Pedestrian Protection Systems: Issues, Survey, and Challenges," IEEE Trans. On Intelligent Systems. Vol. 8, No.3, Sep. 2007
[3] Pangop, Chausse et al, "Feature-based Multisensor Fusio Using Bayes Formula for Pedestrian Classificatio in Outdoor Environments," In proc. IEEE Intelligent Vehicles Symposium Instanbul, Turkey, June 2007
[4] Navneet Dalla and Bill Triggs, "Histograms of Oriented Gradients for Human Detection," In proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2005
[5] D. G. Lowe, "Distinctive image features from scale-invariant key points," IJCV, 60(2):91-110,2004.
[6] T. Watanabe, Satoshi Ito, and Kentaro Yokoi, "Co-occurrence Histograms of Oriented Gradients for Human Detection," IPSJ Transactions on Computer Vision and Applications vol. 2 39-47 March 2010
[7] Navneet Dalal "Finding People in Images and Videos" Thesis report
[8] R. Rajesh, K. Rajeev, V. Gopakumar, K. Suchithra, V.P. Lekhesh, "On Experimenting with Pedestrian Classification using Neural Network," IEEE 2011
[9] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.
[10] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume I, pages 128–142, May 2002.
[11] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.
[12] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.