Human Detection using Feature Fusion Set of LBP and HOG

[1]Amitesh
Amitesh1434@iiitd.ac.in

[2]Pratishtha Verma
Pratishthavrm229@gmail.com

***Abstract*—**Human detection has become one of the major aspect in the real time modern systems whether it is driver-less vehicles or in disaster management or surveillance. Multiple approaches of machine learning are used to find an efficient and effective way of human detection. The proposed method is mainly applied to address the pose-variant problem of human detection. It reduces the redundancy problem which leads to a slow system. To solve the pose variant and redundancy problem, mutation and crossover concept has been applied over Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG) feature set to generate final set . Then combination of feature fusion set of LBP and HOG are fed into Support Vector Machine (SVM) for classification purpose. To improve the performance of detector an unsupervised framework has been used for learning. For post-processing to suppress overlapping and redundant windows - Non-maximal suppression is used . For training and testing purpose, INRIA dataset has been used. The proposed method is compared with HOG, LBP, and HOG-LBP techniques, the result shows that our method outperforms these techniques. .

Keywords*-Local Binary Pattern(LBP), Histogram of Oriented Gradient(HOG), Crossover, Mutation*

*****

I. INTRODUCTION

Information from images are extracted by our eyes and pass it to brain. In our brains - it’s happening in real-time spontaneously, so fast that we never thought of going into details of it. Same as a bulb which looks like glowing but in real it’s getting off and on 50 times, considering 50Hz frequency. But to understand it in a methodical way, here computer vision comes into picture.

The process to extract information from images, here comes the area, Computer vision.Many forms of image data can exist, such as video sequences, multiple cameras point-of-views, or multi-dimensional data from a medical scanner.

The tools/tasks included in the computer vision solely rely on the methods for acquiring, processing, analyzing and understanding digital images, and extraction of information or high-dimensional data from the world to produce numerical or symbolic information, e.g., in the forms of decisions. The three important task of computer vision are object detection, tracking of object and segmentation.

Human is one of the difficult object to detect, because of many reasons:

1) different viewpoint of cameras
2) every person changes their pose time to time
3) dynamic light illumination condition

Multiple approaches are introduced and taken into consideration to solve these problems from last few years. There are multiple public datasets available for training and testing purpose like INRIA person dataset, Caltech pedestrian dataset, MIT datasets which includes both non human and human samples. Process of human detection consist of two steps:

i) the encoding of the image or its region into the feature vector representation and

ii) is to apply classifier to check portion of image is human or not.

For human detection multiple feature vectors have been proposed like HOG[2], Local Binary Pattern[3], Covariance feature[4], Edgelet[5], Haar-like feature[6]. For human detection, HOG is the most popular feature vector.

One of the most popular classifer due to its efficiency and performance is SVM :- the features extracted from labeled samples are used for training the model and this trained model is used for decision making in human detection the most popular feature vector. Liveness of human beings, are detected by Secure and reliable system.

Two problems are presented in this paper: First is the problem arises due to time consumption in multiple redundancy in images, the problem is solved by applying averaging concept on feature sets that make it reliable.

Feature sets are obtained after applying Crossover and Mutation on HOG feature vector and LBP feature vector; second is to deal with the pose variation problem by applying feature fusion concept on the resultant feature fusion set of HOG and improved variant of LBP [10].

For classification purpose, the features are fed into SVM . For post-processing[11] non-maximal suppression [12], is used.

The INRIA dataset is used for training and testing purpose. The ROC curve is plotted and detection rate is calculated to show the tradeoff and improved performance compare to HOG [13], LBP [13], and HOG-LBP [13] techniques.

II. LITERATURE REVIEW

HOG(Histogram of Oriented Gradient) is the principally used feature for the purpose of human detection. This feature is given to the SVM for classification. SVM for the purpose of classification. HOG feature variant is pose, illumination, and shadow invariant. [1]

For improving the detection accuracy, many methods have been proposed using HOG feature because of its performance. A detector was proposed [14] by integrating the cascade of rejectors with the HOG feature which leads to a fast and accurate system. Variable size blocks of HOG features were used to capture features of human and for identifying the appropriate set of blocks from them, Adaboost feature selection algorithm was used. For fast computation, the integrating the cascade of rejectors with the selected block of HOG feature was performed.

A modified version of HOG feature for pedestrian detection was proposed to improve the detector. In which extraction of nine independent channel of HOG was performed according to the interval of gradient orientation. The method was tested on INRIA dataset to show the improved result of the proposed feature over the HOG feature [15].

Another extension of HOG was proposed called co-occurrence Histograms of oriented gradient [16]. In which, the more complex shape of the object has detected by using the pair of oriented gradient. For both gradient orientation, co-occurrence matrix has calculated with different offset. The method was tested on INRIA person dataset and showed better performance than HOG. In CoHOG, the only information regarding gradient direction was used. A modification was done in the CoHOG by taking the gradient magnitude to influence the feature vector named W-CoHOG for the better detection rate [13]. The proposed detector had been experimented on INRIA and Chrysler pedestrian dataset and showed it perform better than CoHOG.

Because of efficiency and good performance, LBP [5] has been used for texture classification, detection of object and for recognizing the face. A human detector in which the combined LBP feature with HOG feature has used [17]. And a partial occlusion handling method is proposed to improve the performance in cluttered background or scene. In partial, they make an occlusion confidence map by clustering the responses of blocks in a detection window.

A Center-symmetric Local Binary Pattern (CS-LBP) and Histogram of Oriented Gradient (HOG) feature extraction method was proposed in [18].The CS-LBP feature use both gradient and textureinformation and HOG feature calculates the gradient magnitude and the gradient direction of an image. The performance of both features was compared on the INRIA dataset and showed CS-LBP performs better compare to HOG.

A novel human detection method by using sliding window technique from still images was presented in [19]. Detection has done by simultaneously segmenting pose and shape of the human body and extracting the HOG feature. The segmentation has been done using probabilistic hierarchical part template matching algorithm. And then feature was extracted around the estimated shapeboundaries. The kernel SVM had been used for classification purpose.

Adaptive contour feature (ACF) [20] contains oriented granular space (OGS) in which chain of square patches is called granular, used to describe a curve. This proposed approach has three outlooks: first, proposing ACF feature which can mine object contour feature and feature co-occurrences; second, heuristic learning is performed by AdaBoost algorithm; third, a coarse human segmentation is conferred with this detector.

For human detection, various part based methods were introduced to detect all the parts of human body separately [21] [22] [23]. A deformable part model was proposed to detect human by using part based method and showed good result [24]. A deformable model weights the configuration of all the parts. The detector was described by using deformable model, part and root filters. HOG features were extracted from all the parts and used to train the model using latent SVM.
The literature was reviewed for detecting human in image. In the case of video analysis, some more knowledge like motion information was required such as optical flow.

A new optical flow method which is consequent to adaptive background segmentation technique was introduced [25]. This method resolved the problem of optical flow that is foreground and background motion is not discriminated. In their method, a person was segmented from the cluttered background and the segmented part was used as an area of interest for detection purpose.

A new motion feature called histogram of oriented flow for human/non-human classification was proposed in [26]. In the method, they combined the proposed descriptor with the HOG feature and the resulted detector was tested on MIT and INRIA dataset.

This paper provides the solution for time-consuming and pose variant problem by applying an averaging concept on fused feature set of HOG and LBP with SVM classifier. The INRIA dataset has been utilized for experimental purpose.

III. PROPOSED METHODOLOGY

The proposed framework is shown in figure 1. Our method consists of two main steps: First is feature extraction, second is classification. Feature extraction procedure is different intraining and testing phase.

In training phase, we extract HOG and LBP feature from the preprocessed positive and negative samples, then apply the averaging method to make them robust, these features are fused to make a single feature descriptor which is used with svm-train model of SVM tool to train thesystem.

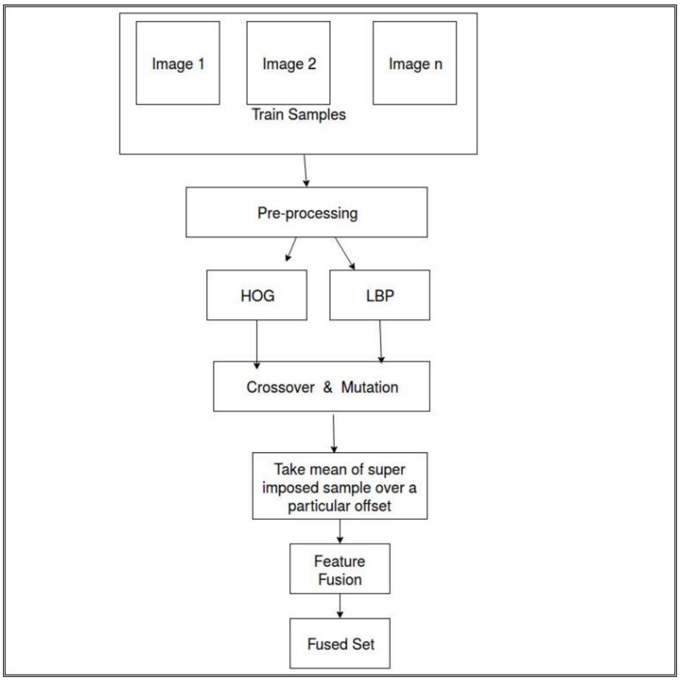
In the testing phase, detection is performed with sliding window method to extract the combination of HOG and LBP feature over the test image and svm-predict model of SVM is used to predict the label of the window.

After detection, post processing is used to improve the detection performance by using non-maximal suppression, in which the windows with highest scores are selected and all overlapping or low score windows are suppressed. The detailed step of these feature extraction and averaging method are illustrated in section A, B and C.

A. LBP:

One of the most widely used visual descriptor used in computer vision is Local Binary Patterns(LBP). Windows are examined in 16x16 pixels each cell. One of the most important characteristics of LBP is its support for uniform pattern and rotation invariant characteristic which is used for background noise filtering.

The basic idea for developing the LBP operator was that two-dimensional surface textures can be described by two complementary measures: local spatial patterns and gray scale contrast. The LBP feature is calculated by dividing the window into the blocks and then over each and every blocks, histogram value is calculated by comparing by comparing the central pixel and neighbor pixel's gray level value. In case when the gray scale value of the central pixel is greater than the neighbor pixel it takes 1 otherwise 0.



B. HOG:

HOG Histograms of oriented gradients, are one of the most important features in object detection. It encode high frequency gradient information. Detection window of 64x128 is divided into 8x8 pixel cells and each group of 2x2 cells constitute a block with a stride step of 8 pixels in both horizontal and vertical directions. The constituent of of each cell is a 9-bin HOG. Each block contains a 36-D concatenated vector of all its cells and normalized to an L^2 unit length.7x15 blocks are used for representing a detection window, giving a total of 3780-D feature vector per detection window.

Good results in object detection are achieved using dense HOG features, processing a 320x240 scale-space image still requires about 140ms on a personal computer with 3.0GHz CPU, 2GB memory. Hence in our experiments we compute histograms with 9 bins on cells of 8x8 pixels. Block size is of 2x2 cells with non-overlap (stride step 16 pixels). Each 64 x 128 detection window is represented by 4x8 blocks , yielding a total of 1152-D feature vector per detection window. According to [1], large stride step might decrease system performance. However, in our experiments, combining with other complementary features have significantly improved the system performance .

$$g_x(x, y) = I(x + 1, y) - I(x - 1, y) \quad (1)$$
$$g_y(x, y) = I(x, y + 1) - I(x, y - 1) \quad (2)$$

Amplitude and direction of gradient for pixel(x, y):

$$g(x, y) = g_x(x, y)^2 + g_y(x, y)^2 \quad (3)$$
$$\theta(x, y) = \tan^{-1} (g_x(x, y)/g_y(x, y)) \quad (4)$$

C. Genetic Algorithms

Genetic Algorithms(GAs) are powerful search and optimization algorithms, which are based on the theory of natural evolution. In GAS, each solution for the problem is called a chromosome and consists of a linear list of codes. The GA sets up a group of imaginary lives having a string of codes for a chromosome on the computer. The GA evolves the group of imaginary lives (referred to as population), and gets an almost optimum solution for the problem. The GA generates a new generation based on previousgenerations.

The GA uses three basic operators to evolve the population: selection, crossover, mutation.

Since search of GAS is parallel in principle, GAS arenot easily caught in a local minimum.GAS do not need derivative of fitness function, whichwidens applications of GAS.For the above-mentioned features, it is possible to finda suitable solution for problems which are difficult to getsolution using conventional methods.

D. Crossover

For the study of successive generations, crossover introduces a considerable complication: Highly competent individuals in the parent generation (say, an "Einstein"), will not reappear in the next generation, because only a subset of the parent's alleles is passed on to any given offspring. This raises an important question: What is passed from generation to generation if the parents' specific chromosomes are never passed on? An answer to this question requires a prediction of the generation-by-generation spread of *clusters* of alleles, requiring a substantial generalization of Fisher's fundamental theorem. The schema theorem is one such generalization; it deals with arbitrary allele clusters call schemas.

A *schema* is specified using the symbol * ('don't care') to specify places along the chromosome not belonging to the cluster. For example, if there are two distinct alleles for each position, call them 1 and 0, then the cluster consisting of allele 1 at position 2, allele 0 at position 4, and allele 0 at position 5, is designated by the string *1*00**...*.

Let N(s,t) be the number of chromosomes carrying schema s in generation t. The schema theorem specifies the (expected) number $N(s,t+1)$ of chromosomes carrying schema s in the next generation

E. Mutation

This operator randomly flips some of the bits in a chromosome. For example, the string 00000100might be mutated in its second position to yield 01000100. Mutation can

 ISSN: 2454-4248
261 – 265

occur at each bit position in a stringwith some probability, usually very small (e.g., 0.001).

IV. Results and Discussion

A. About dataset

One of the most widely used dataset that is open source and available for training and testing is INRIA dataset. The publicly available INRIA dataset contains a wide range of human poses, captures under the varying illumination condition and different viewpoints. This dataset contains total 2416 positive samples and 1218 negative images for training plus 1132 positive samples and 453 negative images for testing.

B. Training

The standard window size is 64x128 pixels. Therefore, we crop all the 2416 positive samples in this size.We have 1218 total negative images from each of the negative images, we are taken 10 random samples of window size.And the resultant samples are used for the feature extraction.

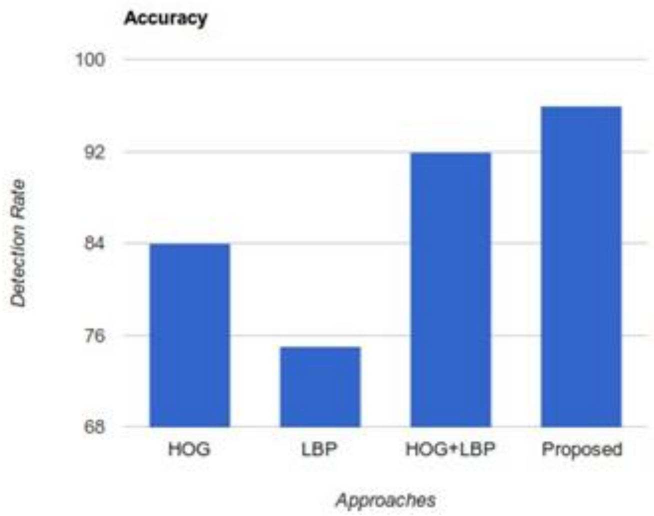
In the feature extraction, HOG and LBP feature is used. For the computation of HOG feature, any operation was performed on every 8x8 pixel cell of the window.

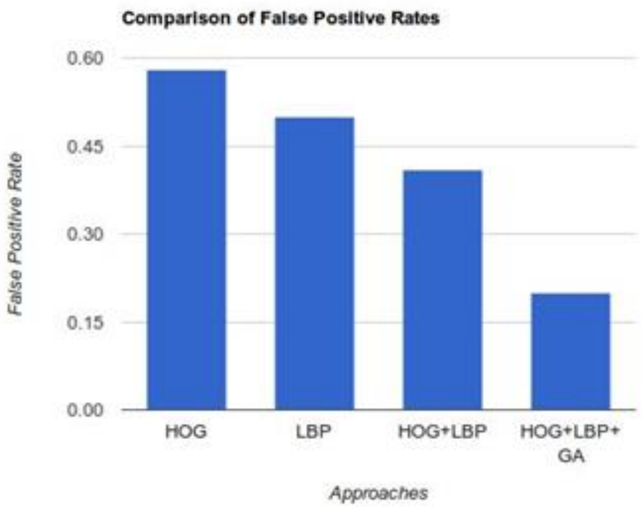
The result of the extraction give total of 3780 features. For LBP feature extraction, we are using uniform rotational invariance characteristic of which give 10 features in each single sample.

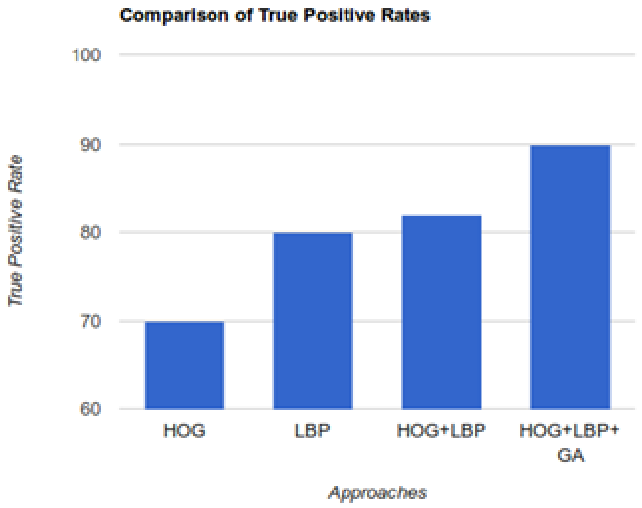
C. Testing

In testing phase, we need to plot the True positive rate, false positive rate and accuracy shown in fig -1,2 and 3. The proposed method is compared with the HOG, LBP, and HOG-LBP. It is clearly visible that the accuracy is greater than all the other method.

True positive rate versus False positive rate, Accuracy curve is plotted for evaluation of the proposed method.







Acknowledgment

I want to thanks my Parents and my Guruji for motivating me to write this research paper. A special thanks to Ms. Pratistha Verma for guiding me with her knowledge in this area for mentoring and enable us to view from different standpoints to this problem.

References

[1] Singh, M. and Arora, A.S., 2016. Varying Illumination and Pose Conditions in Face Recognition. *Procedia Computer Science*, *85*, pp.691-695.

[2] Hoang, V.D., Vavilin, A. and Jo, K.H., 2012, October. Fast human detection based on parallelogram haar-like features. In *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society* (pp. 4220-4225). IEEE.

[3] Wu, B. and Nevatia, R., 2005, October. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (Vol. 1, pp. 90-97). IEEE.

[4] Said, Y., Salah, Y. and Atri, M., 2014, November. Pedestrian detection using covariance features. In *Image Processing, Applications and Systems Conference (IPAS), 2014 First International* (pp. 14). IEEE.

ISSN: 2454-4248
261 – 265

[5] Mu, Y., Yan, S., Liu, Y., Huang, T. and Zhou, B., 2008, June. Discriminative local binary patterns for human detection in personal album. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.

[6] Dalal, N. and Triggs, B., 2005, June. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893). IEEE.

[7] Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*(3), pp.273-297.

[8] Singh, B., Kaur, M., Singh, D. and Singh, G., 2016. Automatic number plate recognition system by character position method. *International Journal of Computational Vision and Robotics*, *6*(1-2), pp.94-112.

[9] Singh, M. and Arora, A.S., 2017. A Robust Anti-Spoofing Technique for Face Liveness Detection with Morphological Operations. *Optik-International Journal for Light and Electron Optics*.

[10] Ahonen, T., Matas, J., He, C. and Pietikäinen, M., 2009, June. Rotation invariant image description with local binary pattern histogram fourier features. In *Scandinavian Conference on Image Analysis* (pp. 61-70). Springer Berlin Heidelberg.

[11] Neubeck, A. and Van Gool, L., 2006, August. Efficient non-maximum suppression. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 3, pp. 850-855). IEEE.

[12] Blaschko, M.B., 2011, July. Branch and bound strategies for non-maximal suppression in object detection. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (pp. 385-398). Springer Berlin Heidelberg.

[13] Andavarapu, N. and Vatsavayi, V.K., 2016. Weighted CoHOG (W-CoHOG) Feature Extraction for Human Detection. In *Proceedings of Fifth International Conference on Soft Computing for Problem Solving* (pp. 273-283). Springer Singapore.

[14] Zhu, Q., Yeh, M.C., Cheng, K.T. and Avidan, S., 2006. Fast human detection using a cascade of histograms of oriented gradients. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (Vol. 2, pp. 1491-1498). IEEE.

[15] Weixing, L., Haijun, S., Feng, P., Qi, G. and Bin, Q., 2015, July. A fast pedestrian detection via modified hog feature. In *Control Conference (CCC), 2015 34th Chinese* (pp. 3870-3873). IEEE.

[16] Watanabe, T., Ito, S. and Yokoi, K., 2009. Co-occurrence histograms of oriented gradients for pedestrian detection. *Advances in Image and Video Technology*, pp.37-47.

[17] Wang, X., Han, T.X. and Yan, S., 2009, September. An HOG-LBP human detector with partial occlusion handling. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 32-39). IEEE.

[18] Uma, S.K. and Srujana, B.J., 2015, Feature Extraction for Human Detection using HOG and CSLBP methods. In International Journal of Computer Applications.

[19] Lin, Z. and Davis, L.S., 2008, October. A pose-invariant descriptor for human detection and segmentation. In *European Conference on Computer Vision* (pp. 423-436). Springer Berlin Heidelberg.

[20] Gao, W., Ai, H. and Lao, S., 2009, June. Adaptive contour features in oriented granular space for human detection and segmentation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 1786-1793). IEEE.

[21] Fischler, M.A. and Elschlager, R.A., 1973. The representation and matching of pictorial structures. *IEEE Transactions on computers*, *100*(1), pp.67-92.

[22] Mikolajczyk, K., Schmid, C. and Zisserman, A., 2004. Human detection based on a probabilistic assembly of robust part detectors. *Computer Vision-ECCV 2004*, pp.69-82.

[23] Wu, B., Nevatia, R. and Li, Y., 2008, June. Segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.

[24] Felzenszwalb, P.F., Girshick, R.B., McAllester, D. and Ramanan, D., 2010. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, *32*(9), pp.1627-1645.

[25] Denman, S., Chandran, V. and Sridharan, S., 2005, December. Adaptive optical flow for person tracking. In *Digital Image Computing: Techniques and Applications, 2005. DICTA'05. Proceedings 2005* (pp. 8-8). IEEE.

[26] Dalal, N., Triggs, B. and Schmid, C., 2006, May. Human detection using oriented histograms of flow and appearance. In *European conference on computer vision* (pp. 428-441). Springer Berlin Heidelberg