Orthogonal Local Preserving Projection Based Pedestrian Detection and Tracking Using Hybrid Features and Support Vector Machine

Madhavi Gajula[1], Dr. A. Jhansi Rani[2]
[1]Research Scholar, [2]Professor
[1]Acharya Nagarjuna University, GUNTUR, Andhra Pradesh, India, [2]Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India.

Abstract- Pedestrian detection and tracking is a challenging task in the area of computer vision, which is used to recognize, detect and track the objects (individuals) over a sequence of images. This procedure helps to understand the object behaviours instead of monitoring computer by human operators. The result of pedestrian detection and tracking is regarded as the input for higher level analysis such as, crowd motion learning and individual behaviour recognition. To further improve the pedestrian detection and tracking, a new system was proposed in this research paper. Here, pedestrian detection and tracking was assessed by using CAVIAR and PETS 2009 datasets. After performing foreground and blob detection on the acquired datasets, hybrid feature extraction (Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Gray-Level Co-Occurrence Matrix (GLCM)) was carried-out for extracting the feature values from the blob detected image frames. Then, Orthogonal Local Preserving Projection (OLPP) was used for rejecting the irrelevant feature vectors or selecting the optimal feature subsets. After selecting the optimal feature vectors, a supervised classifier: Support Vector Machine (SVM) was used to classify the objects (person or other objects). The experimental outcome shows that the proposed system performs effectively by means of precision, recall, f-measure and accuracy. The proposed system enhances the classification accuracy up to 0.5-2% compared to the existing systems.

Index values- Gray level co-occurrence matrix, histogram of oriented gradients, local binary patterns, orthogonal local preserving projection, and support vector machine.

I. INTRODUCTION

Pedestrians tracking and localization is one of the most active research areas in the field of computer vision as it enables a range of applications including robot navigation, autonomous driving and sports analysis [1-3].Multiple Object Tracking (MOT) aims to determine the trajectories of many targets in a particular scene. MOT strongly focused on the tracking-by-detection system, where object detections are grouped for estimating the correct tracks. In a multi-object scenario, the number of objects and an individual state evolve in time, compounded by misdetections, false detections, measurement origin uncertainty [4]. The output of pedestrian detection and tracking supports the applications like 3D modelling, teleconferencing, crowd analysis and intelligent video surveillance [5-7]. In the existing research papers, the researchers utilized dissimilar classification approaches and features for classifying the pedestrians. Still, detecting an individual in a fixed scene is a challenging task due to inter-object interactions, layout of scenes, attraction of scenes and uncertainty of pedestrian behavior [8-9].To overcome these concerns, automatic pedestrian detection systems are required.

The automatic pedestrian detection and tracking system delivers an accurate and reliable recognition that highly improves the security of surveillance system [10].The major contribution of this research work is to accomplish pedestrian detection using an effective system, which enables the management of large datasets. At first, video sequences were collected from the datasets: CAVIAR and PETS 2009.In each and every frame, foreground and blob detections are carried-out for obtaining a particular region of interest that contains parts and the presence of objects. Then, hybrid feature extraction was applied to the blob detected frame for extracting the feature values. Hybrid feature extraction was the procedure of obtaining feature subsets from the set of data inputs by the rejection of redundant and irrelevant features. After obtaining the feature information from hybrid feature extraction, OLPP was utilized to select the optimal feature subsets. The output of OLPP specifies the features, which were essential for classification. These optimal feature values were given as the input for SVM classifier to classify the objects (person or other object).

This research paper is composed as follows. Section II presents a broad survey of recent papers in pedestrian detection and tracking. In section III, an effective supervised system is developed for pedestrian detection and tracking. In section IV, quantitative and comparative analysis of proposed and existing systems are presented. The conclusion is made in the section V.

II. LITERATURE REVIEW

Several research methodologies are suggested by the researchers in pedestrian detection and tracking. The brief evaluation of a few essential contributions to the existing literatures is presented in this section.

H. Li, Y. Liu, C. Wang, S. Zhang, and X. Cui, [11] used particle filter in the video sequences for tracking multiple pedestrians. The developed system determines the confidence value of the background and the object by extracting the prior knowledge for achieving multiple pedestrian tracking. Then, texture and color features were adopted into a particle filter for better observation and then automatically adjust the weight value for each feature based on the current tracking environment. The developed system processes occlusion condition for preventing loss and drift that helps to achieve robust tracking in multiple pedestrians. Experimental analysis and verification confirms that the developed system effectively enhances the performance of tracking. Still, the developed system requires an effective descriptor level feature extraction approach to further improve the classification accuracy.

H. Yang, S. Qu, C. Chen, and B. Yang, [12] presented a new system for enhancing the performance of multiple object tracking. In the developed system, tracklet association was considered as a generalized linear assignment on tracklets affinity, which was determined by using a rank based motion approach and improved sparse representation appearance approach. Then, a weight target template set was used in appearance approach for sparse representation. The rank based motion approach exploits motion estimation based on spatial information to interpolate the missing objects during tracklet association. The developed system was evaluated on three dissimilar datasets and the experimental outcome shows that the developed system provides good tracking performance compared to the existing systems. The developed system classification accuracy degrades, when an individual share close appearance to another individual.

X. Song, X. Shao, Q. Zhang, R. Shibasaki, H. Zhao, and H. Zha, [13] developed an elaborate dynamic system for multiple pedestrian tracking in the crowded environment. In this research study, local instantaneous crowd flow, global semantic scene structure and the social interactions among persons were considered and combined in the unified system. It makes the multiple pedestrian tracking as more accurate and powerful. This research was carried out on an online “tracking-learning” system. The extensive experiments were conducted and the efficiency of the developed methodology was verified by means of disposal rate. In this developed system, it was very challenging to track the individuals, who appear in groups.

Z. Jiang, and D.Q. Huynh, [14] presented an Interacting Multiple Model (IMM) for multiple pedestrian tracking. The developed tracking system contains multiple IMM trackers, which were combined together using an effective data association component. Initially, the HOG feature descriptor was used to extract the features from the video sequences. Then, a sliding window was used in the video frames for eliminating the false negative errors and short term occlusion. The quantitative and qualitative analysis of developed system was compared with four existing visual tracking systems on the benchmark video datasets (PETS-view8, arcade and terrace1). The experimental section confirmed that the developed system was more effective than the existing systems by means of accuracy and precision. In a few cases, the HOG feature descriptor returns one large pedestrian window (overlapping of individuals) that fails to track a particular person.

A. Milan, S. Roth, and K. Schindler, [15] developed a continuous energy minimization system for multiple target tracking, which was the combination of appearance modelling and occlusion reasoning. The partial image evidence was handled by occlusion reasoning and the dissimilar targets were disambiguated with an appearance model. This research focused on designing an energy function, which corresponds to a complete representation of global optimization. Besides, the energy function considered physical constraints like track persistence, target dynamics, and mutual exclusion. The developed system effectively reduces the number of identity switches and false positive rate. The developed system fails to achieve better classification in the crowded environment, which was considered as one of the major concerns.

To overcome the above mentioned issues, a new pedestrian tracking system is implemented for enhancing the performance of pedestrian detection and classification.

III. PROPOSED METHODOLOGY

In intelligent video surveillance system, pedestrian detection and tracking is a significant task, which provides the essential information for understating the video footages. The pedestrian detection and tracking has a clear extension to automotive applications that helps to improve the safety systems. The proposed system for pedestrian detection and tracking contains six major phases: dataset collection, foreground detection, blob detection, feature extraction, feature optimization and classification. The proposed system’s block diagram is indicated in the Fig. 1. The brief description about the proposed system is described below.

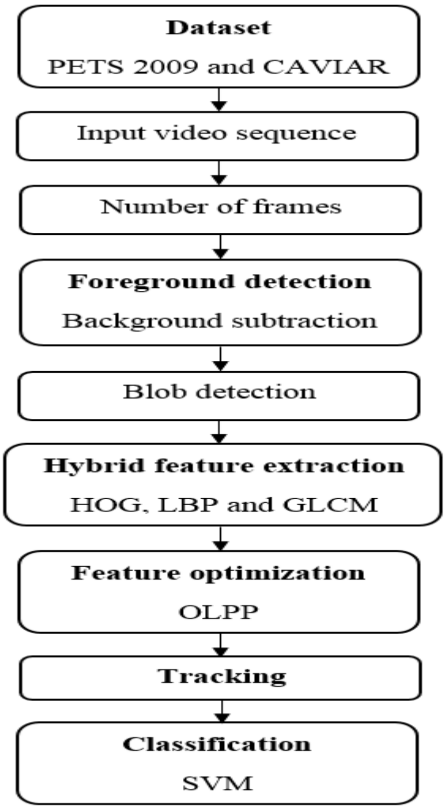


Fig.1: Block diagram of proposed system

A. Dataset collection

In the initial phase of pedestrian detection and classification, video sequences are collected from the standard benchmark datasets: CAVIAR and PETS 2009. The general characteristics of the acquired datasets are detailed in the Table 1.**CAVIAR:** It comprises of four video sequences, which captured at a shopping mall in Portugal. The following video sequences are used for experimental evaluation: ShopAssistant2cor (sequence 1), OneShopOneWait1cor (sequence 2), OneStopMoveEnter1cor (sequence 3) and OneLeaveShop2cor (sequence 4). **PETS 2009:** This dataset comprises of multi-sensor sequences in crowd scenarios with maximum scene complexity. The following sequences are selected for experimental evaluation: S2-L1-view 001-time 12-34 (sequence 1), S2-L3-view 002-time 14-41 (sequence 2), S2-L1-view 005-time 12-34 (sequence 3) and S2-L1-view 008-time 12-34 (sequence 4). After the collection of video sequences, it is transformed into a number of frames. The sample frames of PETS 2009 and CAVIAR datasets are given in the Fig. 2.

TABLE 1. General characteristics of the acquired datasets

Dataset	**Scenario**	**Images per sequences**				**Total images**
		Sequence 1	*Sequence 2*	*Sequence 3*	*Sequence 4*	
PETS 2009	Parking lot	795	240	795	795	2625
CAVIAR	Mall	1119	1377	1590	3700	7786

(a)

(b)

Fig.2 (a): Sample frame of PETS 2009 dataset, (b) Sample frame of CAVIAR dataset

B. Foreground detection using background subtraction

After the acquisition of datasets, foreground detection is carried-out on each frame by using background subtraction. Here, the background of the image frames is determined by subtracting the current frame from the previous frame or from the average image of the number of frames. Background subtraction works well in specific conditions of frame rate and object speed and also it is very sensitive to the threshold. The general formula of background subtraction is given in the Eq. (1).

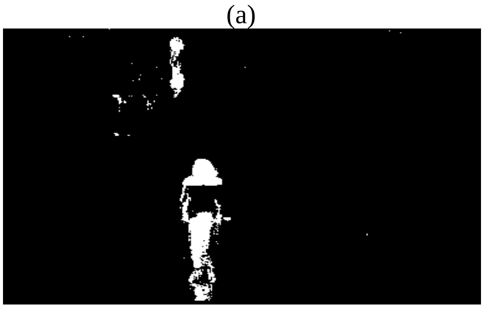
$$|I_i(x,y) - I_{i-1}(x,y)| > T \qquad (1)$$

Where, I_{i-1} is represented as previous frame, I_i is denoted as current frame and T is represented as selected threshold.The sample image of foreground detection using background subtraction is given in the Fig. 3.

(a)

(b)

(a)

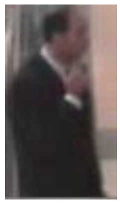


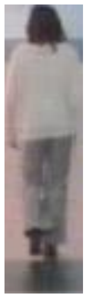
(b)

Fig.3 (a): Input frame, b) Foreground detected frame

After foreground detection, blob detection is carried-out for obtaining a particular region of interest to perform further operations like feature extraction, optimization and classification. In the application of object detection or object tracking, the obtained blob region indicates the parts of objects and the presence of objects. Each blob regions are starched in horizontal and vertical directions until the entire blob is enclosed in a rectangle box. In this paper, the blob detection system is based on bounding box, centre-of-mass and adjacency pixels. Additionally, the statistical features of blobs like volume bounded by the membership function, location of the centre gravity, pixel count of the blob, and size of the rectangular enclosure are also determined.

Currently, blob detection have found increasingly popular, because it uses interest points for wide baseline stereo matching and also for signalling the presence of informative image features for appearance based object detection on the basis of local image statistics. The sample image of blob detection is represented in the Fig.4. The detected blobs are used feature extraction by using hybrid feature extraction.





(a) (b)

Fig.4 (a): Foreground detected frame, b) Detection of blobs

C. Hybrid feature extraction

Usually, feature extraction is defined as the action of mapping the image from image space to feature space and it also transforms the large redundant data into a reduced data representation. It helps to decrease the complexity of the system. In this research study, feature extraction is performed on the basis of HOG, LBP and GLCM. The detailed description about the feature descriptors are given below.

1) Histogram of oriented gradients

Generally, HOG describes about the distribution of spatial directions in each image region. It exploits the local object appearance, which is well characterized by the distribution of edge directions or local intensity gradients. The general idea of HOG is to divide the image into small spatial regions and for each region it creates one-dimensional gradient orientation histogram with gradient direction and gradient magnitude. A key characteristic of HOG feature is capable of capturing the local appearance of objects, and also to account the invariance in object transformations and illumination condition. The edge information about gradients are determined by applying HOG feature vector. At first, a gradient operator$N$is employed to calculate the gradient value. The gradient point of the image is denoted as(x, y) and the image frames are expressed in the Eq. (2).

$$G_x = N * I(x, y) \text{ and } G_y = N^T * I(x, y) \quad (2)$$

Image detection windows are categorized into various minor spatial regions, which is known as cells. Hence, the magnitude gradients of the pixels are experienced with edge orientation. Finally, the magnitude of the gradients$(x, y)$is denoted in the Eq. (3).

$$G_x(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (3)$$

Edge orientation of the point $(x, y)$is specified in the Eq. (4).

$$\theta(x,y) = tan^{-1}\frac{G_y(x,y)}{G_x(x,y)} \quad (4)$$

Where, $G_x$is mentioned as the horizontal direction of gradients and$G_y$is represented as the vertical direction of gradients.

For superior invariance in illumination and noise, a normalization procedure is performed, after the calculation of histogram values. Normalization is an essential step in the HOG feature descriptor, it maintains discriminative characteristics and perform consistently even against parameters like background-foreground contrast and local illumination variations in the input image. Normalization is done by using "block" as a fundamental region of operation. Each block region comprises of a square array of four cells. Each new block is defined with a 50% overlap with the previous block. Normalization effectively maintains the cell-based local gradient information, which is invariant to local illumination conditions. In HOG, four different patterns of normalizations are available such as, L2-norm, L2-Hys, L1-Sqrt and L1-norm. Among these normalization, L2-norm gives better performance in pedestrian detection and classification, which is mathematically given in the Eq. (5).

$$L_{2-norm} : f = \frac{x}{\sqrt{||x||_2^2 + e^2}} \quad (5)$$

Where, $e$is denoted as small positive value,$f$is represented as feature extracted value, x is meant as non-normalized vector in histogram blocks and$||x||_2^2$ represents the 2-norm of HOG normalization.

2) Local binary pattern

LBP is a texture analysis descriptor that transforms an input image frame into labels on the basis of luminance value. In LBP, gray-scale invariance is an essential factor that depends on the texture and local patterns. In an image frame$f$, the pixel position is mentioned as x and $y$that is derived by using the central pixel valuex_c of $x$as the threshold to signify the neighbourhood pixel $m$value.The binary value of the pixel is weighted by using the power of two and then summed to create a decimal number to store in the location of central pixelx_c, which is mathematically given in the Eq. (6).

$$LBP\ (x,y) = \sum_{i=0}^{m-1} f(x_i - x_c)^{2i}, f(x) = \{\begin{matrix}1, x \geq 0\\ 0, x \leq 0\end{matrix}\} \quad (6)$$

Where, x_i stated as the gray level value of the central pixel of a local neighbourhood.

The basic neighbourhood model of LBP is (p-neighbourhood model) that gives 2^poutput, which leads to a large number of possible patterns. If the texture analysis descriptor area is small, the LBP histogram is not attractive. In addition, noise and slight changes lead to ineffective LBP histogram. The uniform model of LBP will attain only when the jumping time is maximized. It is measured by using the Eq. (7).

$$U(LBP(x,y)) = |f(x_{c-1} - x_i) - f(x_0 - x_i)| + \sum_{i=1}^{m-1}|f(x_c - x_i) - f(x_{c-1} - x_i)| \quad (7)$$

Where, $u$is indicated as maximum jumping time.

3) Gray level co-occurrence matrix

The GLCM descriptor is utilized to determine the frequency of pixel pairs, when the pixel intensity values are equal. In this research study, GLCM descriptor comprises of autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares, variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation, inverse difference, inverse difference normalized and inverse difference moment normalized [16]. Among these twenty-three features along with HOG and LBP, the optimal feature values are selected by using OLPP. A brief description about the OLPP is given in the below section.

D. Dimensionality reduction

Normally, dimensionality reduction is utilized for reducing the large dimensional features. The high number of irrelevant features decreases the accuracy of pedestrian detection and classification. So, an effective dimensionality reduction technique (OLPP) is employed for reducing the feature space without losing or affecting the classification accuracy. In this experimental research, OLPP algorithm helps to decrease the measurement of feature vectors or values. The detailed explanation about OLPP algorithm is described below.

1) Orthogonal local preserving projection

In conventional dimensionality reduction algorithms, it is very difficult to remake the information and also the existing algorithms are ordinarily non-orthogonal. To overcome these issues, an effective algorithm: OLPP is employed in this research for pedestrian detection and classification. The step by step procedure of OLPP is described below.

2) PCA projection

Initially, PCA projection extracts the features that are statistically uncorrelated and the rank the new data matrix, which is equal to the number of features (dimensions).

a) Constructing the adjacency graph

Let $X_d = [d_1, d_2, \ldots .d_K]$ be a set of input data. Consider G, which denotes a graph with n nodes. The a^{th} node corresponds to the data d_i. If d_a and d_b are "close", i.e. d_a is among p nearest neighbours of d_b and d_b is among p nearest neighbours of d_a. If the class information is available in any two nodes, consider an edge between the two nodes that belongs to the same class.

b) Selecting the weights

If the node a and b are inter-connected, then the weight W is calculated using the Eq. (8).

$$W = e^{\frac{-(d_a - d_b)}{t}} \quad (8)$$

Where, t is represented as constant. If the nodes a and b are not connected means, assume $W = 0$.

c) Computing the orthogonal basis functions

After calculating the weight matrix W from the Eq. (8), then calculate the diagonal matrix M. A diagonal matrix M is defined in the Eq. (9), whose entries are column (or row) sums of W.

$$M = \sum_{ab} W_{ab} \quad (9)$$

Afterwards, calculate the Laplacian matrix L_m using diagonal matrix M and weight matrix W, which is mathematically given in the Eq. (10).Let, $[V_1, V_2, \dots, V_k]$ be the locality preserving projections, which is given in the Eq. (11).

$$L_m = M - W \quad (10)$$

$$r^{(k-1)} = [V_1, \dots, V_{k-1}] \text{and } S^{(k-1)} = [\,r^{(k-1)}]^T Z^{-1}\, r^{(k-1)} \quad (11)$$

Where, $Z^{-1} = X_d M X_d^T$.The orthogonal basis vectors $[V_1, V_2, \dots, V_k]$ are computed as follows,

- Determine V_1 as the eigenvector of $Z^{-1} X_d L X_d^T$ and relate with the smallest Eigen values.
- Determine V_k as the eigenvector of smallest Eigen values of J^k, which is mathematically given in the Eq. (12).

$$J^k = \{I - Z^{-1}\, r^{(k-1)}[\,S^{(k-1)}]^{-1}[\,r^{(k-1)}]^T\} Z^{-1} X_d L_m X_d^T \quad (12)$$

d) OLPP embedding

Let,$E_{OLPP} = [V_1, V_2, V_3 \dots, V_k]$ embedding is denoted in the Eq. (13) and (14).

$$X_d \to Y = W^T X_d \quad (13)$$

$$E = E_{PCA} E_{OLPP} \quad (14)$$

Where, E is represented as the transformation matrix and Y is denoted as the dimensional representation of X_d. The dimensionality reduction of extracted feature vectors is carried-out by using the above transformation matrix.

E. Classification using support vector machine

After performing feature optimization, classification is carried out using SVM, which enables an efficient way of extracting the features and a set of rules to perform classification. SVM is a discriminative classification approach represented by a separate hyper-plane. The SVM classifier widely used in several applications like bioinformatics, signal processing, computer vision fields, etc., due to its high performance in accuracy, and ability of processing the high dimensional data. SVM does well in solving two-class problem, which is associated with the theories of vapnik–chervonenkis and structure principles. The general formula for the linear discriminant function is denoted as $w.x + b = 0$. In order to distinct the samples without noise, an optimum hyper plane is exploit between the two groups, which is mathematically given in the Eq. (15).

$$pi[w.x + b] - 1 \geq 0, i = 1,2,..N \quad (15)$$

Then, reduce $||w||^2$ in the Eq.(15), so the optimization issue is solved by the saddle point of a Lagrange function with Lagrange multipliers. The ideal discriminant function is denoted in the Eq. (16),

$$f(x) = sign\{(w^* x) + b^*\} = sign\{\textstyle\sum_{i=1}^{N} \alpha_i^* . pi(x_i^* - x) + b^*\} \quad (16)$$

Finally, replace the interior product by a linear kernel function $k(x, x')$ in the Eq. (16) for reducing the computational complexity in higher dimensional data. In this way, the linear separability of estimated samples improved and the discriminant function is re-written as given in the Eq. (17).

$$f(x) = sign\{\textstyle\sum_{i=1}^{N} \alpha_i^* . pi. k(x, x_i) + b^*\} \quad (17)$$

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed system is experimented using MATLAB (version 2018a) with 3.0 GHZ Intel i5 processor, 2TB hard disc and 16 GB RAM. For determining the effectiveness of the proposed system, the performance of the proposed system is compared with the existing systems (HOG+Gabor filters [17] and Colour moments + chi-square dissimilarity measure + nearest neighbour classifier [18]) on the reputed datasets: PETS 2009 and CAVIAR. The proposed system performance is determined in light of accuracy, precision, recall and f-measure.

A. Performance measures

Performance measure is defined as the measurement of outcomes, which develops a reliable information about the effectiveness and efficiency of the proposed system. The relationship between the input and output values of the proposed system is understood by using the performance measures like precision, recall and f-measure. The general formula for evaluating the precision, recall and f-measure are given in the Eq.(18), (19), and (20).

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (18)$$

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (19)$$

$$F-measure = \frac{2TP}{2TP+FP+FN} \times 100 \quad (20)$$

Additionally, accuracy is one of the effective evaluation measures used for finding the effectiveness of the proposed system of pedestrian classification. Accuracy is the most instinctive performance measure and it is simply a ratio of total observations to the correctly predicted observations. The general formula of accuracy is given in the Eq.(21).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (21)$$

Where, FP is signified as false positive, TN is indicated as true negative, TP is specified as true positive, and FN is definite as false negative.

B. Quantitative analysis on PETS 2009 dataset

In this sub-section, PETS 2009 dataset is assessed for evaluating the performance of the proposed system. The detection of moving objects in the PETS 2009 dataset is denoted in the Fig. 5. In Table 2, the proposed system performance is validated by means of precision, recall, f-measure, and accuracy. Here, the performance evaluation is validated with 80% of training and 20% of testing. The average precision and recall of the proposed system is 99.73% and 96.55%. Correspondingly, the average f-measure and accuracy of proposed system are 98.10% and 98.67%. The graphical representation of quantitative analysis of proposed system using PETS 2009 dataset is represented in the Fig. 6.

Fig.5: Sample frames of pedestrian detection in PETS 2009 dataset

TABLE 2. Performance evaluation of proposed system using PETS 2009 dataset

PETS 2009dataset				
Sequences	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)
PETS 2009 -S1	98.58	100	97.43	98.69
PETS 2009 -S2	99.67	100	99.39	99.69
PETS 2009 -S3	97.61	99.81	95.90	97.82
PETS 2009 -S4	98.83	99.14	93.49	96.234
Average	**98.67**	**99.73**	**96.55**	**98.10**

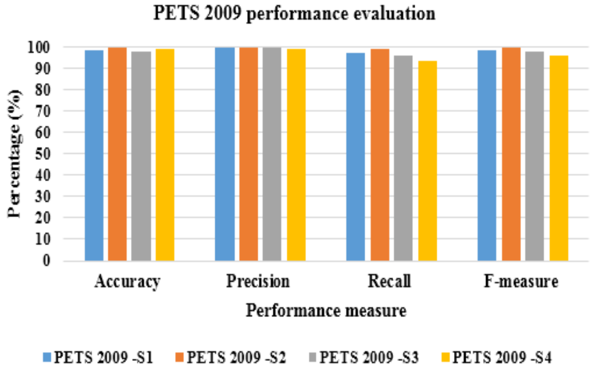


Fig.6: Graphical representation of proposed system using PETS 2009 dataset

C. Quantitative analysis on CAVIAR dataset

In this experimental section, the performance evaluation of proposed system is carried out by using CAVIAR dataset. The detection of moving objects in CAVIAR dataset is meant in the Fig. 7.In Table 3, the proposed system is evaluated in light of precision, recall, f-measure, and accuracy. The average precision and recall of the proposed system is 99.58% and 97.61%. Similarly, the average f-measure and accuracy of the proposed system are 98.58% and 97.915%. The Tables 2 and 3 confirmed that the proposed system performs effectively on PETS 2009 and CAVIAR datasets. The graphical representation of quantitative analysis of proposed system using CAVIAR dataset is denoted in the Fig. 8.

Fig.7: Sample frames of pedestrian detection in CAVIAR dataset

TABLE 3. Performance evaluation of proposed system using CAVIAR dataset

CAVIAR dataset				
Sequences	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F-measure (%)*
CAVIAR-S1	96.99	100	96.17	98.049
CAVIAR-S2	99.27	100	98.99	99.494
CAVIAR-S3	98.10	99.17	98.204	98.68
CAVIAR-S4	97.30	99.15	97.09	98.11
Average	97.915	99.58	97.61	98.58

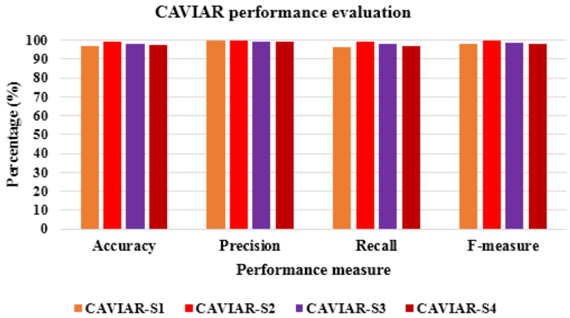


Fig.8: Graphical representation of proposed system using CAVIAR dataset

Table 4 represents the performance of proposed system with feature optimization and without feature optimization. In with feature optimization, the SVM classifier increases the accuracy in pedestrian classification up to6-4% compared to with-out feature optimization. In this research study, the descriptor level features determine the non-linear and linear characteristics of the images and also preserves the quantitative relationships between the high and low level features. The performance measures confirm that the proposed system performs significantly in pedestrian classification compared to the previous systems.

TABLE 4. Accuracy evaluation of proposed system using with and without feature optimization

Dataset	**Feature extraction**	**Feature optimization**	**Classifier**	**Accuracy (%)**
PETS 2009	HOG+LBP+GLCM	Without OLPP	SVM	94.34
		With OLPP		98.67
CAVIAR	HOG+LBP+GLCM	Without OLPP	SVM	91.2
		With OLPP		97.915

D. Comparative analysis

Comparative analysis of proposed and existing works is detailed in the Table 5. C. Conde, D. Moctezuma, I.M.De Diego, and E. Cabello, [17] developed a new system for human detection based on HOG and Gabor filters. Initially, the image convolutions were developed using a Gabor filter bank. Then, the HOG feature descriptor was applied to the resulting Gabor images for extracting the relevant features. Finally, an effective classifier was used to classify the original image is a person or not. This experiment was carried out on the online databases (i.e., PETS 2006, PETS 2007, PETS 2009 and CAVIAR) to validate its result by means of precision, recall, f-measure and accuracy. The developed system averagely achieved 91.75% of precision, 67.25% of recall, 76.75% of f-measure and 98.27% of accuracy on PETS 2009 dataset. Additionally, the developed system averagely achieved 92.75% of precision, 58% of recall and 68.75% of f-measure on CAVIAR dataset .Additionally, M. Chandrajit, R. Girisha, T. Vasudev, and M. Hemesh, [18] developed a new system for multi object tracking. Initially, colour moments were used for extracting the feature values from the segmented foreground region. Then, nearest neighbour classifier and Chi-square dissimilarity measure were used for classifying the moving objects. This research work was carried out on a dataset: PETS 2009. In this research study, the developed system achieved 81% of precision and 71% of recall in pedestrian classification.

Compared to these existing approaches, the proposed system achieved better results that is higher than the existing works. As discussed in the proposed system, feature optimization is a part of pedestrian detection and tracking in this research paper. Whereas, several feature vectors are obtained under the feature extraction using HOG, LBP and GLCM. Then, feature optimization (OLPP) is used for selecting the optimal feature values, which are fit for the classification. Successively, performance of the OLPP based feature optimization is compared with the total features. The effect of feature optimization is shown in the Table 4. The proposed system achieved better classification rate compared to the existing systems by means of accuracy, precision, recall and f-measure.

TABLE 5. Comparative analysis of proposed and existing works

Methodology	Dataset	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
HOG and Gabor filters [17]	PETS 2009	91.75	67.25	76.75	98.27
	CAVIAR	92.75	58	68.75	-
Color moments + Chi-square dissimilarity measure + nearest neighbor classifier [18]	PETS 2009	81	71	-	-
Proposed system	PETS 2009	99.73	96.55	98.10	98.67
	CAVIAR	99.58	97.61	98.58	97.915

V. CONCLUSION

A new system is developed in this paper to detect pedestrian in the video using crowd motion pattern. The main aim of this research is to obtain the optimal feature values for classifying the objects (either as the person or other objects) using PETS 2009 and CAVIAR dataset. In this research paper, hybrid feature extraction is utilized to extract the feature vectors from the blob detected image frames. Then, an effective feature optimization approach: OLPP is used to select the optimal feature subsets or rejects the irrelevant feature vectors. This optimal feature information is given as the input for SVM classifier for classifying the objects. Though, tracking and detection of pedestrian targets are a crucial part in video surveillance. Compared to other existing systems in pedestrian detection and tracking, the proposed system delivered an effective performance by means of accuracy, precision, recall and f-measure and shows 0.5-2% of the improvement in classification accuracy. In future work, a new unsupervised classification approach is implemented with descriptor level features for further improving the classification rate of pedestrian.

VI. REFERENCES

[1]. G. Führ, and C.R. Jung, “Combining patch matching and detection for robust pedestrian tracking in monocular calibrated cameras”, *Pattern Recognition Letters*, vol.39, pp.11-20, 2014.

[2]. Z. Jin, and B. Bhanu, “Analysis-by-synthesis: Pedestrian tracking with crowd simulation models in a multi-camera video network”, *Computer Vision and Image Understanding*, vol.134, pp.48-63, 2015.

[3]. L. Kratz, and K. Nishino, “Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes”, *IEEE transactions on pattern analysis and machine intelligence*, vol.34, no.5, pp.987-1002, 2012.

[4]. C. Creusot, “Local segmentation for pedestrian tracking in dense crowds”, In *International Conference on Multimedia Modeling, Springer, Cham*, pp. 266-277, 2014.

[5]. X. Wang, and Z. Tang, “Modified particle filter-based infrared pedestrian tracking”, *Infrared Physics & Technology*, vol.53, no.4, pp.280-287, 2010.

[6]. J. Qu, Z. Liu, and W. He, “Video pedestrian detection based on orthogonal scene motion pattern”, *Mathematical Problems in Engineering*, 1-8, 2014.

[7]. Y. Fang, G. Ding, Y. Yuan, W. Lin, and H. Liu, “Robustness Analysis of Pedestrian Detectors for Surveillance”, *IEEE Access*, 6, 28890 - 28902 2018.

[8]. S. Kuppuswamy, and B. Panchanathan, “Similar Object Detection and Tracking in H. 264 Compressed Video Using Modified Local Self Similarity Descriptor and Particle Filtering”, *International Journal of Intelligent Engineering and Systems*, vol.10, no.5, 95-104, 2017.

[9]. L. Sun, G. Liu, and Y. Liu, “Multiple pedestrians tracking algorithm by incorporating histogram of oriented gradient detections”, *IET Image Processing*, vol.7, no.7, pp.653-659, 2013.

[10]. M. Wang, H. Qiao, and B. Zhang, “A new algorithm for robust pedestrian tracking based on manifold learning and feature selection”, *IEEE Transactions on Intelligent Transportation Systems*, vol.12, no.4, pp.1195-1208, 2011.

[11]. H. Li, Y. Liu, C. Wang, S. Zhang, and X. Cui, “Tracking algorithm of multiple pedestrians based on particle filters in

[12]. video sequences”, *Computational intelligence and neuroscience*, pp.13, 2016.

[13]. H. Yang, S. Qu, C. Chen, and B. Yang, “Multiple Objects Tracking With Improved Sparse Representation and Rank Based Dynamic Estimation”, *IEEE Access*, vol.6, pp.42264-42278, 2018.

[14]. X. Song, X. Shao, Q. Zhang, R. Shibasaki, H. Zhao, and H. Zha, “A novel dynamic model for multiple pedestrians tracking in extremely crowded scenarios”, *Information Fusion*, vol.14, no.3, pp.301-310, 2013.

[15]. Z. Jiang, and D.Q. Huynh, “Multiple Pedestrian Tracking From Monocular Videos in an Interacting Multiple Model Framework”, *IEEE Transactions on Image Processing*, vol.27, no.3, pp.1361-1375, 2018.

[16]. A. Milan, S. Roth, and K. Schindler, “Continuous energy minimization for multitarget tracking”, *IEEE transactions on pattern analysis and machine intelligence*, vol.36, no.1, pp.58-72, 2014.

[17]. D.A. Clausi, “An analysis of co-occurrence texture statistics as a function of grey level quantization”, *Canadian Journal of remote sensing*, vol.28, no.1, pp.45-62, 2002.

[18]. C. Conde, D. Moctezuma, I.M.De Diego, and E. Cabello, “HoGG: Gabor and HoG-based human detection for surveillance in non-controlled environments”, *Neurocomputing*, vol.100, pp.19-30, 2013.

[19]. M. Chandrajit, R. Girisha, T. Vasudev, and M. Hemesh, “Data Association and Prediction for Tracking Multiple Objects”, *Indian Journal of Science and Technology*, vol.9, no.33, 2016.

Dr. A.Jhansi Rani obtained her B.Tech Degree in Electronics and Communications Engineering from Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada in 1991 and M.Tech Degree in Microwave Engineering from Institute of Technology, Banaras Hindu University, and Varanasi in 1993. Later she joined as a Faculty in ECE Dept., of V.R.Siddhartha Engineering College, Vijayawada. She obtained Ph.D. Degree from JNTU, Hyderabad in August 2008. Presently, she is working as a Professor of ECE at V. Siddhartha Engineering College. Dr. Jhansi has more than 24 years of teaching experience, and has about 60 technical publications in various International and National Journals and Conferences to her credit. She is the Life Member of FIETE, MISTE, BMESI and MSEMCEI.

She received grants from DST and AICTE in the area of smart Antennas. Her fields of interest include Electromagnetics, Smart Antennas, Analysis of Microwave Components, EM Waves and Transmission Lines, Numerical Methods and Applications.

Mrs. G.Madhavi is obtained her B.Tech degree in ECE from Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada and M.Tech Degree in Digital systems and computer Electronics from JNTUH. She is pursuing her Ph.D. in Image and Video processing from Acharya Nagarjuna University, Guntur. Presently she is working as an Associate professor in Mahatma Gandhi institute of technology, Hyderabad. Her areas of interests are Signal processing, image and video processing, Neural Networks and Microwave engineering.