International Journal of scientific research and management (IJSRM)
||Volume||2||Issue||10||Pages||1573-1578||2014||
Website: www.ijsrm.in ISSN (e): 2321-3418

Detection of Human Bodies from the Background in an Image Using Piecewise Linear-Support Vector Machine

Lakshmidhar and Dr. Ramana Reddy
Rajeev Gandhi Memorial College of Engineering & Technology

Abstract - **Detection of the human bodies from the background in an image is challenged by the view and posture variation problem. In this paper, a piecewise linear support vector machine (PL-SVM) method is used for solving the problem of identification of humans in an image, which is challenged by the view and posture variation problem. The motivation is, by using the piecewise discriminative function for constructing a non-linear classification boundary that can differentiate multi-view and multi-posture human bodies from the backgrounds in a high dimensional feature space. A PL-SVM training is designed as an iterative procedure of feature space division and linear SVM training, aiming at the maximized marginality of local linear SVM's. Each piecewise SVM model is responsible for subspace, corresponding to a human cluster of a special view or posture. In the PL-SVM, a cascaded detector is proposed with block orientation features and a histogram of oriented gradient features. This identification of humans in an image is implemented by using mat lab. By using this method, we can also detect the presence of humans in videos also.**

Index Terms - classification, object detection, piecewise linear, support vector machine.

I. INTRODUCTION

Detection of humans in images and video frames is an important problem in the area of image based sensing with applications such as robotics, pedestrian warning for driving assistance, surveillance, and entertainment. Even though the detection of humans in some common views and in static video background has been greatly put forward in recent years, it is still a challenging problem in the situations of moving cameras, complex backgrounds, and in particularly, large variations of views and postures.

In the existing human detection methods, feature representation and classifier design are two main problems being investigated. Visual feature descriptors have been proposed for human detection including Haar-like features, HOG , v-HOG , Gabor filter based cortex features , covariance features , Local Binary Pattern (LBP) , HOG-LBP , Edgelet , Shapelet , Local Receptive Field (LRF) , Multi-Scale Orientation (MSO) , Adaptive Local Contour , Granularity-tunable Gradients Partition (GGP) descriptors , pose-invariant descriptors . A most recent research demonstrates that, while using mixture of different kinds of visual features gives the superior performance.

The extracted features on labeled samples are usually fed into a classifier for training. Linear SVM is the most popular classifier for human detection. However, when we need to detect multi-view and multi-posture humans simultaneously in a video system, the performance of a linear SVM often drops significantly. It is observed in experiments that humans of continuous view and posture variations form a manifold, which is difficult to be linearly classified from the negatives. An algorithm that requires multi-view and multi-posture humans to be correctly classified by a linear SVM in the training process often leads to over-fitting. This problem can be handled by using some non-linear classification methods such as Kernel SVMs, but they are computationally more expensive than linear methods. In addition, the use of the kernel trick in a very high-dimensional feature space, such as a 3780 dimensional HOG feature space.

Some approaches use a divide-and conquer strategy to deal with the multi-view and multi-posture problem, by first dividing training positives into sub-classes and then training multiple models for detection. These divide-and-conquer strategies can reduce empirical error in training process and improve the detection performance in some cases, but sometimes they also give higher structural risk and more false positives.

Another solution to the multi-view and multi-posture problem is to segment a human body into parts considering that each part has smaller deformation, lower dimensionality and non-linearity, and therefore can be better detected with a linear classifier. A deformable part-based model (DPM) is proposed for human detection. Human parts and their spatial bias are modeled with a structure SVM with latent variables (latent SVM). When performing training or detection, a local searching operation is carried out to optimize the location of each part-based model, which is called local deformation. By the local deformation, the detection avoids suffering from the view and posture variations. The DPM methods contribute an elegant framework for object detection, showing state-of-the-art performance on human detection. But they suffer from low resolution images of human objects, on which local model optimization has little significance.

In machine learning research, piecewise and localized SVMs have attracted much more due to their superior performance over the global kernel SVMs. However, the problem of how to construct a piecewise decision boundary in a high dimensional feature space is not well discussed. Cross distance minimization algorithm (CDMA) is designed to compute hard margin of non-kernel SVMs. The multi category SVMs are proposed to extend the binary SVM to the multi category case, which is essentially different from our proposed piecewise linear SVM (PL-SVM) method in both the

theoretical basis and the training procedure. In terms of the theoretical basis, the multi category SVMs are developed to approximate the Bayes rule for multi category classification purpose. Our PLSVM method exploits the piecewise discriminative function to construct a non-linear classification boundary that can discriminate multiple positive sub-classes from the negative class. In the training of the multi category SVMs, the method of Lagrange multipliers is employed to solve the objective equation of the dual problem. In the training of PL-SVM, nearest point analysis (NPA) on convex hulls together with an iterative linear SVM solution is used, which guarantees the max-margin of the final classifier. In [27], Cheng et al propose a profile SVM (P-SVM) to reach local linear classification, by using the minimal distance to each pre-calculated cluster center to decide which local SVM a sample should belong to. It has the advantages of nonlinear discrimination and sample division in a low-dimensional feature space, its sample division strategy suffers from the curse of dimensionality. In addition, the max-margin property of the profile SVM is not fully considered in the classifier training procedure.

In this paper, pedestrian detection is formulated as a nonlinear classification problem in a high-dimensional feature space. The piecewise linear SVM (PL-SVM) method is introduced into multi-view and multi-posture human detection for the first time. Our PL-SVM is essentially different from other piecewise SVMs in the feature space division and model training strategy. When training the PL-SVM, with a membership degree maximization criterion, the feature space is divided into subspaces, each of which can be better discriminative for a linear SVM. This approach ensures a lower empirical risk than using only one linear SVM. The training of the PL-SVM is an iterative division of training samples and the feature space. The convergence of the iterations is guaranteed by the monotonically increasing and bounded margins of the PL-SVM, which also guarantees that the PL-SVM is a maximal margin classifier, and thus has a small structural risk. A new kind of feature, called Block Orientation (BO), is proposed as a complement to the popular HOG features. BO and HOG features are incorporated with two cascaded PLSVMs, improving both the accuracy and efficiency in human detection.

The remainder of this paper is organized as follows: PL-SVM modeling and training are presented in Section II. Human detection with the proposed PL-SVM is described in Section III. Experimental results are provided in Section IV. Section V concludes the paper.

II. PIECEWISE LINEAR SUPPORT VECTOR MACHINE

In this section, we present the PL-SVM and explain how to train it, given a training sample set $X = \{(x_n, y_n)\}$, $n = 1, \ldots, N$, where x_n is a sample feature vector, $y_n \in \{-1,+1\}$ denotes the sample label and N denotes the number of samples.

A. PL-SVM Model

A PL-SVM, made up of K linear SVMs, is described as a piecewise linear function

$$f(x) = \arg\max_{f_k(x),\, x \in \Omega_k} \{C_k(x)\} \qquad (1)$$

Where $f_k(x) = w^T_k \cdot x + b_k$, $k = 1 \ldots K$, represents the k^{th} local linear SVM with normal vector w^T_k and threshold b_k. In (1), $\Omega_k = \Omega^+_k \cup \Omega^-$ denotes the k^{th} subspace occupied by a subset of the training samples as shown in Fig. 1.

In (1), $C_k(x)$ is the membership degree of a sample x to Ω_k. From the viewpoint of probability, the membership degree is defined as

$$C_k(x) = P_k(y = 1|x) \qquad (2)$$

where $P_k(y = 1|x)$ is the outputted probability of a sample x being a positive when it is inputted into the k^{th} linear SVM. The probability is defined as the functions of the SVM output as follows

$$P_k(y = 1|x) = \frac{1.0}{1.0 + exp\,(-Ak.fk(x) + Bk)} \qquad (3)$$

where A_k and B_k are two parameters calculated with a maximum likelihood estimation on the training subset [30], and $A_k \cdot f_k(x) + B_k$ is called the parameterized sample-to-hyper-plane distance. By (2) and (3) we know that the larger the distance is, the larger the probability, and then the larger the membership degree to the corresponding SVM.

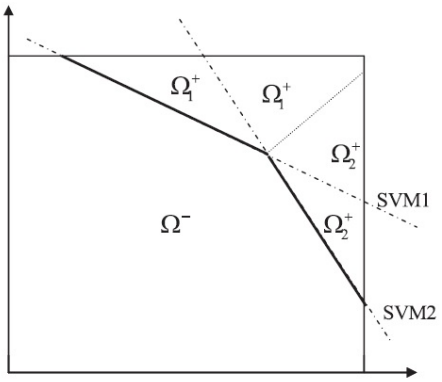

Fig. 1. Illustration of the PL-SVM and feature space division. Subspaces are bounded with dotted lines and Ω^+_1 , Ω_2^+ denote the positive subspaces corresponding to linear SVMs 1, 2, respectively, with Ω^- denoting negatives. Different positive subspaces are related to samples of different views and postures. Classification boundary of the PL-SVM is marked by bold line segments.

With the maximized membership degree criterion in (1) and (2), each linear SVM is responsible for a subspace for classification. The final non-linear classification boundary in the whole feature space consists of linear hyper-planes, as illustrated in Fig. 1. When this criterion is used to divide the feature space and assign positive samples in an iterative training, the parameterized sample-to-hyper-plane distance will be enlarged and then the SVM margins will be also enlarged step by step. This is consistent with the maximal margin principle, ensuring that the PL-SVM keeps the essence of the original SVM approach.

When performing classification, (1) can be converted to a PL-SVM discriminative function

$$F(x) = \text{Sign}\,(f(x)) \qquad (4)$$

with a sign function for discrimination and detection.

B. PL-SVM Training

Before training, human samples are initially divided into subsets with a *K*-means clustering algorithm in a manifold embedded space, as shown in Fig.2 , Having been clustered into initial subsets, human samples assigned to the same subset have smaller differences, leading to a better sample division than a random one.

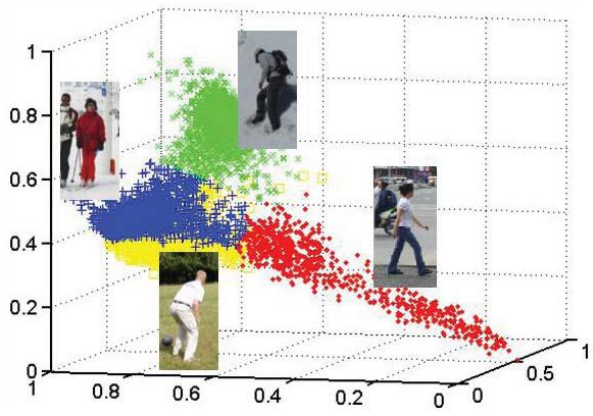

Fig.2. Initial human sample division in a 3-D manifold embedded space. Points of different colors denote samples of different subsets.

The local linear embedding (LLE) algorithm [31] is employed to construct the human manifolds. LLE computes the low-dimensional and neighborhood-preserving embeddings of the high-dimensional samples by mapping them into the low-dimensional space. Given a set of human samples in the high-dimensional feature space, LLE starts with finding nearest neighbors based on the Euclidean distance. Finally it obtains an embedded space by solving a sparse eigenvector problem. More specifically, the *d* eigenvectors associated with the *d* smallest non-zero eigen values provide an ordered set of an orthogonal base, as shown in Fig.2 where *d* is set to 3 for easy visualization.

Algorithm 1 PL-SVM Training

Definitions: t: Iteration number; $R^{(t)}$: Number of reassigned positive samples in iteration t; $r^{(t)}$: Reassigned sample ratio in iteration t.

1. Initialization

Given a training human object set $X = \{(x_n, y_n)\}$, n =1, . . . , N, and K initial subsets $\{X_k^{(0)}\}$, k = 1, . . . K, train K linear SVMs $\{ f_k (x)\}$, k = 1, . . . , K, as the initial PL-SVM model. Set t = 0.

2. Iteration

2.1. Calculate the membership degrees $C_k (x_n)$, k =1, . . . , K, of every feature vector x_n to the K linear SVMs in the PL-SVM

2.2. For a random and unselected positive sample (x_n, y_n), select the k that maximizes the membership degree of x_n as $k = \max \{C_m(x)\}$,m = 1, . . . , K. Set $C_k (x) = 0.0$.

2.3. Check whether the assignment of x_n to the kth subset reduces the distance between the positive and negative convex hulls. If it does, goto 2.2; otherwise assign x_n to the k^{th} subset.

2.4. Train the linear SVMs $\{ f_k (x)\}$, i = 1, . . . , K, using the current subsets $\{X_k^{(t)}\}$, k = 1, . , K.

2.5. If the reassigned sample ratio $r^{(t)}$ is larger than a predefined threshold τ, then t ← t + 1 and go to step 2.1; otherwise go to step 3.

3. Output

K sample subsets $\{X_k^{(t)}\}$, k = 1, . . . , K, and a trained PLSVM consisting of the K linear SVMs.

Step 2.3 in Algorithm is used to ensure the monotonous increase of the SVM margins and thus the convergence of the algorithm. See the next section for the detail. The threshold τ is set to 0.02 empirically.

C. Training Convergence Analysis

Each of the linear SVMs of the PL-SVM is trained by sequential minimization optimization. The convergence of the PL-SVM training is analyzed by the nearest point algorithm (NPA). Let us construct the positive convex hull U_k and the negative convex hull V_k for the *k*th subset, shown as the polygons in Fig. 3. Also let $u_k \in U_k$ and $v_k \in Vk$ such that

$$\| u_k - v_k \| = \min_{u \in U_k, v \in V_k} \|u - v\|. \quad (7)$$

Then the problem of finding u_k and v_k is equivalent to finding the solution of *k*th SVM. If (w_k, b_k) is the solution of the *k*th linear SVM $f_k (x) = w^T_k \cdot x + b_k$, by using the fact that from the maximum margin $2 / \| wk \| = \| u_k - v_k \|$ and $w_k = \delta \cdot (u_k - v_k)$ for some δ, the relation between the normal vector and the nearest point pair (u_k , v_k) can be derived as

$$w_k = \frac{2}{\|uk-vk\|^2} (u_k - v_k), \quad b_k = \frac{\|uk\|^2 - \|vk\|^2}{\|uk-vk\|^2} \quad (8)$$

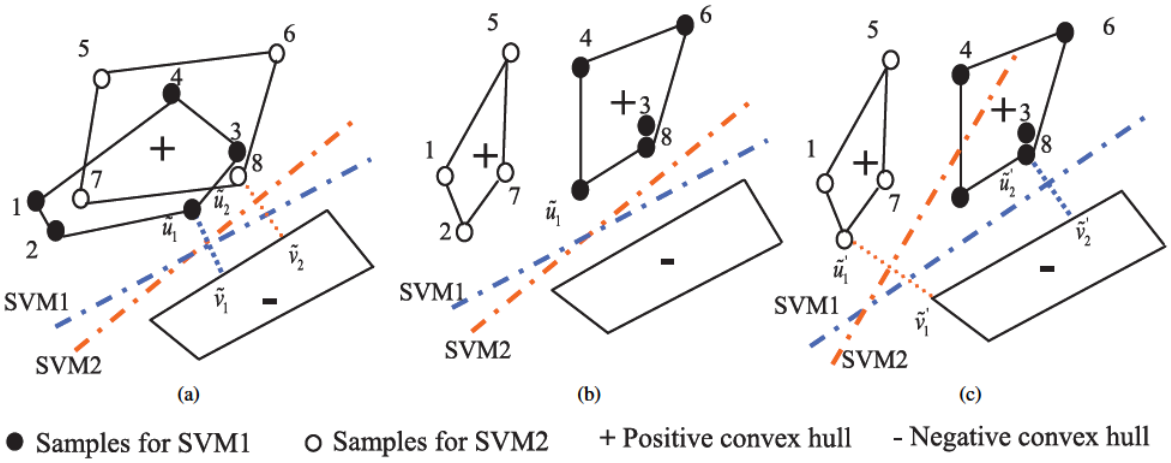

Fig.3. Illustration of sample reassignment and convex hull changes in the training of two linear SVMs where *(u1, v1), (u2, v2), (u₁',v₁') and (u₂',v₂')* are nearest point pairs. (a) Convex hulls and their corresponding SVMs in the current iteration. (b) Subsets after sample re-assignment. (c) Convex hulls and their Corresponding SVM's in the next iteration..

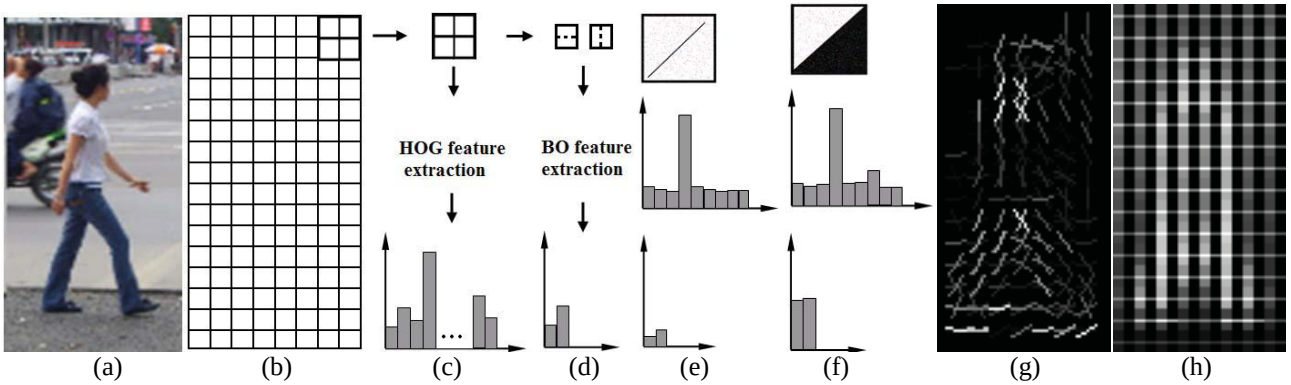

Fig.4. HOG and BO feature extraction. (a) Human example. (b) HOG cells. (c) HOG feature extraction in a block. (d) BO feature extraction in a cell. (e) Stroke pattern in a cell (enlarged) with noise and its HOG and BO features. (f) Region pattern in a cell with noise and its HOG and BO features. (g) Visualization of the HOG features multiplying with the SVM norm vector. (h) Visualization of the BO features multiplying with the SVM norm vector.

By (8), we know that the margin of the *k*th SVM is equal to the distance between the nearest point pair u_k and v_k. When we perform sample re-assignment in the training procedure in Algorithm I, a sample is reassigned to the subset of the SVM, to which the membership degree of the sample is the largest. These ensure that the distances between the nearest point pairs, (u_k, v_k), $k = 1, \ldots, K$, increase monotonically (or non-decrease monotonically) in the training procedure. Consequently the margins of the SVMs increase monotonously. Since the margins are bounded, the training algorithm is thus convergent.

Fig.3 shows an example of PL-SVM training with two subsets. The samples denoted by filled-in circles belong to subset 1 and the samples denoted by open circles belong to subset 2. The samples labeled by 1, 2, 3, 4 and $u1$ form the positive convex hull for subset 1. The samples labeled by 5, 6, 7 and 8 form the positive convex hull for subset 2. Suppose that at current iteration, we obtain two nearest point pairs $(u1, v1)$ and $(u2, v2)$. Then they are used to generate the hyper-planes of SVM1 and SVM2. After steps 2.1, 2.2 and 2.3 in Algorithm I, the samples are reassigned to subset 1 or subset 2 with the maximization of membership degree criterion in (3). It can be seen from Fig. 3(b) that samples 1,
2, 5 and 7 are assigned to subset 2 and samples 3, 4, 6, 8 and $u1$ are assigned to subset 1. With the new subsets, new positive convex hulls are constructed, as shown in Fig. 3(b). Then with the negative hull and new positive convex hulls, new SVMs are trained, as shown in Fig. 3(c). It can be seen that after the iteration, the margins of the SVMs increase or remain the same.

III. HUMAN DETECTION

The proposed PL-SVM is incorporated with two kinds of features for human detection. A cascade detector is designed to improve detection performance.

A. Feature Representation

As shown in Figs. 4(a)–(c), a sample of 64 × 128 pixels is divided into cells of size 8 × 8 pixels, each group of 2 × 2 cells is integrated into a block in a sliding fashion, and blocks overlap with each other. To extract HOG features, we firstly calculate the gradient orientations of the pixels in the cells. Then in each cell, we calculate a 9-dimensional histogram of gradient orientations as the features. Each block is represented by a 36-dimensional feature vector, which is normalized by dividing each feature bin with the vector module. Each sample is represented by 105 blocks (420 cells), corresponding to a 3780-dimensional HOG feature vector.

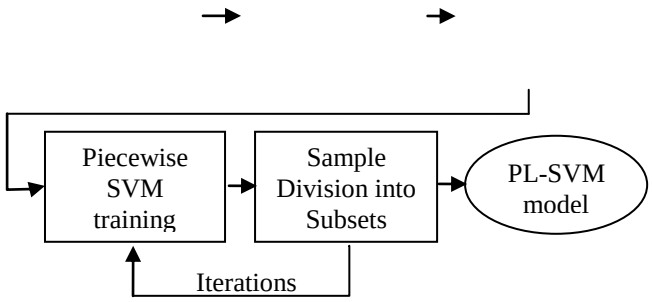

Fig.5. Flow chart of PL-SVM Training

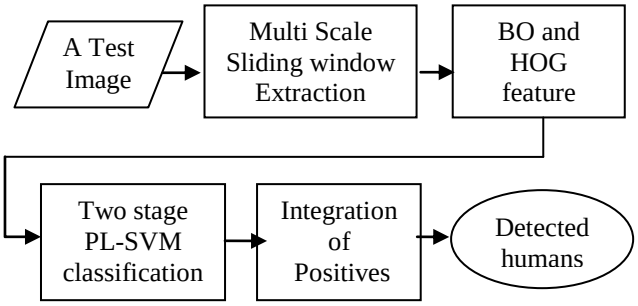

Fig.6. Flow chart of human detection

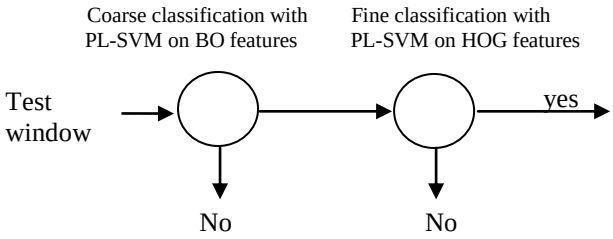

Fig.7. Cascaded classification with two PL-SVMs on BO and HOG features, respectively

We also propose a new kind of features, called Block Orientation (BO) features, derived from Haar-like features, as a complement to the HOG features for human detection. Each of the 420 cells is first divided into left-right and up-down sub-cells as shown in Fig. 4(d), and then the horizontal and vertical gradients of the cell are calculated by

$$\text{Bh}=max\{\left|\sum_{X\in left\ subcell} Ic(X)-\sum_{X\in right\ subcell} Ic(X)\right|\}$$

$$\text{Bv}=max\{\left|\sum_{X\in up\ subcell} Ic(X)-\sum_{X\in down\ subcell} Ic(X)\right|\} \quad (9)$$

Where $I_c(X)$ is one of the R, G and B color values at pixel X.

The BO features are the normalizations of *Bh* and *Bv*:

$$BO_h = Bh/\sqrt{Bv^2 + Bh^2 + \square}$$
$$BO_v = Bv/\sqrt{Bv^2 + Bh^2 + \square} \qquad (10)$$

where ε is a constant to reduce noise effect. Its value is set as 10.0 × (the size of a cell).

B. Cascade Detector With PL-SVMs

Given a set of training samples, we train two PL-SVM models, one with the BO features and the other with the HOG features, as shown in Fig. 5. In the detection procedure (Fig. 6), we apply a histogram equalization and median filtering of radius equal to 3 pixels on the test image firstly, as the preprocessing. Then the test image is repeatedly reduced in size by a factor of 1.1, resulting in an image pyramid. Sliding windows are extracted from each layer of the pyramid. In each window, the BO features are extracted and tested with the PL-SVM in the first stage. If the window is classified as a human, the HOG features will be extracted and tested with the PL-SVM of the second stage to finally decide whether it is a human or not. Adjusting the threshold in the second stage can balance the detections of false positives and false negatives.

IV. EXPERIMENTS

When training the local linear SVMs in the iterative PL-SVM training of algorithm I, we use LIBLINEAR [33], which is designed for linear classification of a large amount of data. Both BO and HOG features are calculated with integral image methods on color and gradient images to improve the efficiency.

The three datasets used in the experiments as follows:

TABLE: 1

INFORMATION ABOUT DATASETS

Data set	Training positives	Negatives	Images for testing
SDL	7550	5769	258
TUD-Brussels	1167	6759	508
INRIA	2478	12180	288

A. Parameter Setting of PL-SVM

To determine the piecewise number *K* of a PL-SVM, we design a ten-fold cross validation. Cross validation accuracies with different piecewise numbers are tested and the *K* with the highest accuracy is selected. For larger the K value, it requires more number of training samples. The piecewise number for the SDL data set is 6 and for TUD-Brussel and INRIA data sets are 4. In general, it is not true that a larger *K* is better due to the over-fitting problem.

Fig. 8 contains human examples from subsets of the SDL dataset in different views and postures when $K = 6$. Therefore, it is expected that when these subsets are used to train the PL-SVM models, both the training and detection performances can be improved.

Fig.8. Human examples of six subsets from the SDL dataset. (a) Frontal or back views with legs apart. (b) Frontal or back views with legs close together. (c) Side views with legs apart. (d) Side views with legs close together.(e) Standing views different from (a)–(d). (f) On bicycles

B. Comparison of PL-SVM with Other SVMs

TABLE III
TRAINING AND DETECTION EFFICIENCY OF FIVE SVM METHODS

Method	Training Time Without Boosting of Negatives (Hours)	Training Time After Five Rounds of Boosting of Negatives (Hours)	Detection Speed (Images/Second)
Linear SVM	0.034	0.19	0.92
Intersection Kernel SVM	0.45	2.79	0.18
Profile SVM	0.17	1.03	1.50
Latent SVM	-	3.2	0.40
PL-SVM	0.15	0.95	0.33(HOG) 1.6 (BO and HOG)

The training and detection efficiency of our proposed method is tested and compared with other four SVM methods. As shown in Table 3, except the linear SVM, PL-SVM is more efficient in both training and testing than the others. PL-SVM is about three times as fast as Intersection Kernel SVM and latent SVM in training. When performing detection, it runs at a speed about 1.6 images per second on average. It can be seen from the last column that the usage of BO features in the cascade detection boosts the detection speed from 0.33 images per second to 1.6 images per second. This speed is about four times as fast as the state-of-the-art latent SVM.

C. Human Detection Performance

In our implementation of PL-SVM, all the b_k (threshold) of the local linear SVMs are set the same in each stage. The threshold in the first stage controls the positives passed to the second stage. To ensure that most of the positives can be passed to the second stage, we use a small threshold value for PL-SVM in the first stage.

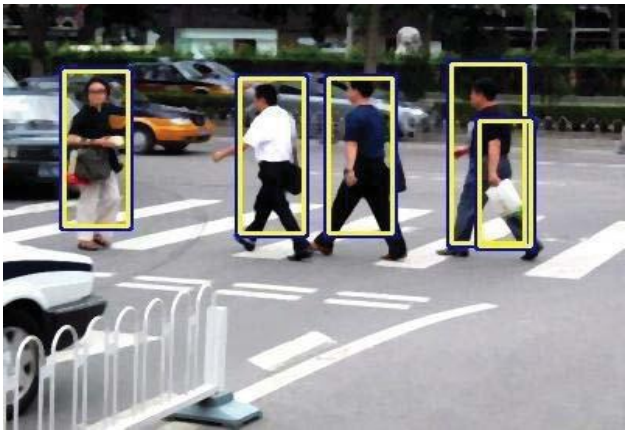

Fig.9. human detection example

The humans of different views, with low resolution and under clutter backgrounds are correctly detected in the video images with few missing/false positives, showing the potential of the proposed approach in video based applications, such as intelligent surveillance systems and driving warning systems.

V. CONCLUSION

For practical applications in the detection of humans, view and posture variation is the important problem. In this paper, we propose a solution for this problem by developing a novel classification method called PL-SVM. This consists of multiple linear SVMs and has the ability to do a non linear classification method. In the PL-SVM training, each linear SVM of the PL-SVM is responsible for one cluster of humans in a specific view or posture. The multi-view and multi-posture human detection problem can be solved by integrating all the linear SVM's. The PL-SVM training algorithm can automatically divide the feature space and train the PL-SVM with the margins of the linear SVMs increased iteratively. We have also presented the BO features as a complement to the HOG features for human detection. Compared with several recent SVM methods, this performs best when dealing with the detection of humans of low-resolutions in clutter backgrounds.

Future work includes the extension of this method to human detection from videos where not only static visuals but also other information such as motion or context.

REFERENCES

[1] Y. Xu, D. Xu, S. Lin, T. X. Han, X. Cao, and X. Li, "Detection of sudden pedestrian crossings for driving assistance systems," *IEEE Trans.Circuits Syst. Video Technol.*, vol. 42, no. 3, pp. 729–739, Jun. 2008.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun.2005, pp. 886–893.

[3] S. Munder and D. M. Gavrila, "An experimental study on pedestrian classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11,pp. 1863–1868, Nov. 2006.

[4] Q. Ye, J. Jiao, and B. Zhang, "Fast pedestrian detection with multi-scale orientation features and two-stage classifiers," in -*Proc. IEEE 17th Int. Conf. Image Process.*, Sep. 2010, pp. 881–884.

[5] B. Wu and R. Nevatia, "Cluster boosted tree classifier for multi-view, multi-pose object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[6] O. Oladunni and G. Singhal, "Piecewise multi-classification support vector machines," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 2323–2330.

[7] S. Q. Ren, D. Yang, X. Li, and Z. W. Zhuang, "Piecewise support vector machines," *Chin. J. Comput.*, vol. 32, no. 1, pp. 77–85, 2009.

[8] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[9] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "A fast iterative nearest point algorithm for support vector machine classifier design," *IEEE Trans. Neural Netw.*, vol. 11, no. 1, pp. 124–136, Jan. 2000.

[10] Available: http://coe.gucas.ac.cn/SDL-HomePage/resource.asp

P.Lakshmidhar is currently pursuing masters of Technology program in Digital Systems & Computer Electronics, Rajeev Gandhi Memorial College of Engineering & Technology, Nandyal, Andhra Pradhesh, India, PH-8985327864.
E-mail: lakshmidharp@gmail.com

Mr. M. Ramana Reddy,Ph.D, Professor in ECE Department, Rajeev Gandhi Memorial College of Engineering & Technology, Nandyal, Andhra Pradhesh, India. He is the professional memberships of the **MISTE, MIE, FIETE** and also Ph.D Scholar. E-mail: ramanareddy0106@gmail.com