Discriminative Weighted Sparse Partial Least Squares for Human Detection

Qiming Li, Yan Yan, Member, IEEE, and Hanzi Wang*, Senior Member, IEEE

Abstract—The channel feature detectors have shown their advantages in human detection. However, a large pool of channel features extracted for human detection usually contain many redundant and irrelevant features. To address the above issue, we propose a robust discriminative weighted sparse partial least squares (DWSPLS) approach for feature selection and apply it to human detection. Unlike PLS, which is a straightforward dimensionality reduction technique, we propose to use sparse partial least squares (SPLS) to achieve feature selection. Furthermore, in order to obtain a robust latent matrix, we formulate a discriminative regularized weighted least squares (DRWLS) problem, where a discriminative term is incorporated to better distinguish positive samples from negative samples. A robust sparse weight matrix is trained based on the latent matrix and used for feature selection. Finally, we use the selected channel features to train the boosted decision trees and incorporate the weights of selected features with each tree. The human detector trained by the selected features can preserve high robustness and discriminativeness. Experimental results on some challenging human datasets demonstrate that the proposed approach is effective and achieves state-of-the-art performance.

Index Terms—Dimensionality reduction, feature selection, partial least squares, human detection.

I. INTRODUCTION

HUMAN detection is one of the most important problems in many applications of intelligent transportation systems, such as vehicle safety, video surveillance and automotive systems. Over the past few decades, many effective feature descriptors have been proposed for human detection. Due to the changes in illumination and large variations in appearance of the human body, a single feature descriptor, such as histograms of oriented gradients (HOG) [1], local binary patterns (LBP) [2], has its limitation since it ignores some useful low-level image attributes. Hence, the combination of different feature descriptors is very helpful to improve the performance of human detection.

Partial least squares (PLS) [30] is an effective classification and dimensionality reduction method which aims to maximize the covariance between the predictors and responses by constructing a latent matrix. It finds a linear regression model by projecting the predictors and responses onto a new space. Schwartz et al. [25] combined PLS with quadratic discriminant analysis (QDA) to train a human detector and obtained some promising results. However, the extraction and projection of high-dimensional features is time-consuming in the detection process. Moreover, PLS extracts low-dimensional features by using linear combinations of all the original high-dimensional features, which usually contain many redundant and irrelevant features. Thus, the discriminativeness of the low-dimensional features for classification is inherently reduced.

Recently, there has been significant interest in combining different feature channels along with the integral image trick for efficient feature computation. Integral channel features (ChnFtrs) [15] is one of the most representative work, which significantly improves human detection performance. The approach uses boosted detectors to efficiently learn the detector from a large pool of channel features. Due to its accuracy and efficiency, we adopt the locally decorrelated channel features (LDCF) framework [3], which is a variant of ChnFtrs, to extract different channel features. Based on LDCF, we propose an alternative and effective approach to select the most discriminative features from the large number of channel features.

Inspired by the sparse partial least squares (SPLS) approach proposed by Chung et al. [26, 27], we propose a robust discriminative weighted sparse partial least squares (DWSPLS) approach for feature selection in human detection. The training stage of DWSPLS consists of two steps: the discriminative regularized weighted least squares (DRWLS) step and the sparsity step. In the DRWLS step, we minimize the exponential classification loss in the proposed DRWLS problem, where a discriminative term is incorporated to better distinguish positive samples from negative ones. To improve the efficiency, we solve the problem in its dual form by using the Newton’s method, where an optimized latent matrix can be obtained. In the sparsity step, a robust sparse weight matrix is trained by SPLS based on the obtained latent matrix in the DRWLS step for feature selection. The above two steps are iteratively performed until convergence. Finally, we use the selected discriminative channel features to train the boosted classifier and incorporate the weights of selected features with each boosted decision tree according to the sparse weight matrix.

The main contributions of this paper can be summarized as follows. First, we propose an effective approach to select the most discriminative features from a large number of channel features by using SPLS. The human detector trained by the selected features shows high robustness and discriminativeness. Second, we derive a robust latent matrix by solving a DRWLS problem, where a discriminative term is embedded to effectively distinguish positive samples from negative ones.
Third, the boosted decision trees trained by the selected channel features are given different weights according to the sparse weight matrix, which can be more robust to variations caused by different appearances of human body.

The rest of this paper is organized as follows. A review of the related work is presented in Section II. Section III introduces the details of the proposed DWSPLS approach. Section IV conducts extensive experiments. Finally, Section V concludes the paper.

II. RELATED WORK

A series of significant improvements have been rapidly developed in human detection after the introduction of the HOG based human detection approach [1]. Most of those approaches focus on feature descriptors, articulation handling and classifier selection.

HOG [1] is based on the evaluation of histograms calculated on the orientation of the gradients, which has shown its effectiveness in handling appearance variations and illumination changes. Many HOG-based detectors, such as the deformable parts model (DPM) [23, 24] and the exemplar-SVM model [4], have been proposed for human detection. Generally speaking, a single feature descriptor has limitation in representing richer patterns. Hence, some work studies the combination of different feature descriptors to provide more information of image attributes for human detection. Wang et al. [5] showed that the combination of HOG and local binary patterns (LBP) can outperform the single HOG feature descriptor and is capable of handling partial occlusions. In [21], Walk et al. integrated spatial and temporal information by combining the local color self-similarity (MultiFtr+CSS) and the motion features (MultiFtr+Motion) with HOG. Wu and Nevatia [6] proposed a new feature descriptor that automatically combines edgelet, HOG and covariance features, which can obtain significant performance improvement. In addition, Paisitkriangkrai et al. [22] proposed to extract some low-level visual features like covariance features and LBP based on spatial pooling for human detection. Recently, the integral channel features (ChnFtrs) [15], which extracted different channel features including LUV color channels, gradient magnitude and gradient histograms, was proposed. Due to its good accuracy and high efficiency, many variants of the approach [16–20] have emerged and obtained some significant improvements.

Articulation handling is another important issue in human detection. One of the most popular articulation handling approaches is the DPM approach proposed by Felzenszwalb et al. [23, 24], where the human body can be divided into some local parts and all the parts are arranged in a star-structured model with the human body being a root. The final score of a detection window is given by the score of the root filter plus the sum over the scores of all part filters, and then minus the deformation cost of the part filters. Kokkinos [7] applied a dual tree branch-and-bound (DTBB) approach to accelerate the computational speed of DPM by constructing the upper and lower bounds on the scores of the part filters. Zhu et al. [8] replaced the star-structured model with a mixture of hierarchical 3-layer tree model where the nodes represent object parts. Song et al. [9] quantized the image lattice by using an overcomplete set of shape primitives and applied the and-or tree (AOT) model to seek a globally optimal part configuration.

Features combination and articulation handling will inevitably increase feature dimension and bring the extra computational cost. Therefore, it is beneficial to apply a dimensionality reduction approach which can preserve significant discriminative information while decreasing the computational burden. Schwartz et al. [25] proposed to use PLS for dimensionality reduction and applied it to human detection. However, the extraction and projection of high-dimensional features is time-consuming in the detection process. Moreover, PLS extracts low-dimensional features by using linear combinations of all the original high-dimensional features, which usually contain many redundant and irrelevant features. The SPLS approach [26, 27] proposed by Chung et al. can select the discriminative features from the high-dimensional features by conducting sparsity on the weight vectors.

Recently, most of classifiers used in human detection were based on SVM, boosting or deep neural network. Since the original proposal of HOG+SVM [1] framework, many human detectors [5, 28] based on SVM were proposed including HikSVM [10], LatSVM [23, 24] (deformable part-based model) and so on. Most of the variants of integral channel features (ChnFtrs) [16–20] are based on the boosted decision trees. The deep neural network also shows great progress in human detection. ConvNet [11] created a convolutional neural network by learning a mix of unsupervised and supervised features directly from raw pixel values. In [12–14], the authors learned the deep architectures with low and high-level features mixture representations jointly.

III. DISCRIMINATIVE WEIGHTED SPARSE PARTIAL LEAST SQUARES (DWSPLS)

In this section, an overview of the proposed DWSPLS approach for human detection is introduced in Section A. We split the proposed approach into the DRWLS step and the sparsity step. The detailed process of the two steps is described in Sections B and C, respectively. The complete DWSPLS approach is summarized in Section D. Finally, how the proposed approach is used for human detection is introduced in Section E.

A. Overview

An overview of the proposed DWSPLS approach for human detection is shown in Fig. 1. During the training stage, like in [3], we extract different channel features including LUV color channels, gradient magnitude and gradient histograms to generate large number of channel features. Subsequently, DWSPLS is applied to obtain a robust sparse weight matrix so that the discriminative features can be selected based on the weight matrix. It is worth noting that we do not use the trained sparse weight matrix for dimensionality reduction. Instead, the discriminative features are selected according to the corresponding weights in the sparse weight matrix. This guarantees that the discriminative features can be selected as
many as possible. Finally, we use the selected channel features to train the boosted classifier and incorporate the weights of selected features into each boosted decision tree. During the test stage, the channel features in a detection window are selected according to the sparse weight matrix and then fed into the trained boosted decision trees for classification.

B. The discriminative regularized weighted least squares (DRWLS) step

Before formally presenting the DRWLS step of the proposed approach, we begin by introducing PLS that serves as the foundation of our approach. The goal of PLS is to construct latent variables as a linear regression model which specifies the linear relationship between the feature matrix and the class labels.

Let $X \in \mathbb{R}^{N \times M}$ denote a feature matrix, where $N$ and $M$ are the number of samples and the dimension of the feature vectors, respectively, and $y \in \mathbb{R}^{N}$ denote the class labels. Each column of $X$ and $y$ are normalized to have zero-mean and unit variance. PLS decomposes the zero-mean matrix $X$ and the zero-mean vector $y$ into:

$$X = TP^T + E$$  \hspace{1cm} (1)

$$y = Uq + f$$  \hspace{1cm} (2)

where $T$ and $U$ are $N \times P$ matrices containing $P$ extracted latent vectors. $P$ and $q$ represent an $M \times P$ loading matrix and a $P \times 1$ loading vector, respectively. The $N \times M$ matrix $E$ and the $N \times 1$ vector $f$ represent the residuals.

The nonlinear iterative partial least squares (NIPALS) algorithm developed by Wold [31] is one of the most popular algorithms for solving PLS. The basic idea of NIPALS is to find a weight matrix $W = (w_1, w_2, ..., w_p)$ by maximizing the covariance between $X$ and $y$, which are both projected to obtain latent vectors. The criterion to find a weight vector $w$ is formulated as:

$$w = \arg \max_w \{ cov(t, u) \}^2 = \arg \max_w \{ cov(Xw, y) \}^2$$

s.t.  \hspace{1cm} $w^T w = 1$  \hspace{1cm} (3)

where $t$ and $u$ are the column vector of matrices $T$ and $U$, respectively. $cov(t, u)$ denotes the sample covariance between the latent vectors $t$ and $u$.

The NIPALS algorithm is an iterative process, which stops when the desired number of the latent vectors is extracted. The details of derivation of the NIPALS algorithm can be found in [30]. At the final stage of each iteration, the class labels $y$ will be deflated with the latent vector and the loading scalar. The residuals (i.e., the classification loss) of the class labels can be modeled as:

$$f = y - t_q$$  \hspace{1cm} (4)

where $q_i$ denotes the loading scalar of the $i$-th iteration and $t_i$ is the $i$-th column of the latent matrix $T$.

The PLS algorithm uses the residuals as the new class labels $y$ for the next iteration. To minimize the residuals and train a robust latent vector from the first step, one popular approach is to solve a regularized weighted least squares (RWLS) problem. Suppose the weight $v_n = \exp(-y_n t_n q_i)$, the latent vector is obtained by:

$$t_n = \arg \min_{t_n} \{ \sum_{n=1}^{N} v_n (y_n - t_n q_i)^2 + \frac{1}{2} \| t_n \|^2 \}$$  \hspace{1cm} (5)

where the regularization term $\| t_n \|^2$ is used to keep $t_n$ small enough.

The original RWLS problem, however, does not consider the discriminativeness of positive and negative samples. Therefore,
in this paper, we propose to replace the regularization term with a discriminative term and introduce a new objective function, called discriminative RWLS (DRWLS). We define the new objective function as:

\[
\hat{t}_{ni} = \arg \min_{t_{ni}} \left\{ \sum_{n=1}^{N} v_n(y_n - t_{ni}q_i) + \frac{1}{2} \sum_{n=1}^{N} \left| \left| t_{ni} - \mu_n \right| \right|^{2} \right\}
\]

\[
\mu_n = \begin{cases} 
\mu_{pos} & \text{if } y_n \text{ is positive} \\
\mu_{neg} & \text{otherwise}
\end{cases}
\]

where \( \mu_{pos} \) and \( \mu_{neg} \) are the mean of the projected positive and negative samples, respectively. Actually, this discriminative term is similar to the Fisher ratio in LDA [34, 35]. In this objective function, the residuals are minimized in the first term. In the meantime, positive and negative samples can be well separated by minimizing the second term. According to Eq. (6), the residuals, which are the differences between the observed and predicted responses, are minimized, while the gap between positive and negative samples is maximized.

The optimal solution of Eq. (6) can be found by using the Newton’s method. However, note that the DRWLS problem in the dual form has less constraints than that in the primal form, which means the DRWLS can be solved more efficiently in the dual form compared with the primal form. Therefore, we transform the primal problem into its dual form.

The primal form of the objective function of the DRWLS problem is reformulated as follows:

\[
P(t, \xi) = \sum_{n=1}^{N} v_n \xi_n^2 + \frac{1}{2} \sum_{n=1}^{N} \left| \left| t_{ni} - \mu_n \right| \right|^{2} \tag{7}
\]

s.t. \( y_n - t_{ni}q_i = \xi_n \)

According to the Lagrange duality theory, the dual form of Eq. (7) can be written as:

\[
D(\alpha) = \frac{1}{2} \sum_{n=1}^{N} \sum_{n' \neq 1}^{N} \alpha_n \alpha_{n'} q_i^2 (\mu_{pos} - \mu_{neg})^2 + \sum_{n=1}^{N} \alpha_n q_i \mu + \frac{1}{2} \sum_{n=1}^{N} \alpha_n^2 - \sum_{n=1}^{N} \alpha_n y_n \tag{8}
\]

where \( \alpha_n \) is the dual variable.

The Newton’s method has been widely used for solving the dual problem. Therefore, the problem can be converted to the selection and optimization of the dual variable \( \alpha \). The first and the second derivatives of \( D(\alpha) \) are obtained by the following equations:

\[
D'(\alpha) = \alpha_n q_i^2 (\mu_{pos} - \mu_{neg})^2 + \frac{\alpha_n}{2v_n} + q_i \mu - y_n \tag{9}
\]

\[
D''(\alpha) = q_i^2 (\mu_{pos} - \mu_{neg})^2 + \frac{1}{2v_n} \tag{10}
\]

Afterwards, we update the dual variable \( \alpha \) as follows:

\[
\alpha_{n}^{k+1} = \alpha_{n}^{k} - \frac{D'(\alpha_{n}^{k})}{D''(\alpha_{n}^{k})} \tag{11}
\]

where \( k \) is the iteration number.

Finally, we update the latent vector \( t_{ni} \) as follows:

\[
t_{ni}^{k+1} = (\alpha_{n}^{k} - \frac{D'(\alpha_{n}^{k})}{D''(\alpha_{n}^{k})}) q_i (\mu_{pos} - \mu_{neg})^2 + \mu \tag{12}
\]

This iterative optimization step is repeated until converge or the maximum iteration number is reached. We give an overview of the DRWLS step in Algorithm 1.

---

**Algorithm 1: Newton’s method for solving the DRWLS problem**

**Input:** The class label of the \( n \)-th sample \( y_n \), the \( i \)-th loading scalar \( q_i \), the mean of the positive samples \( \mu_{pos} \), the mean of the negative samples \( \mu_{neg} \).

**Output:** The \( i \)-th latent vector \( t_i \).

1. Initialize all dual variables \( \alpha_n = 0 \) \( (n = 1, ..., N) \) and the iteration number \( k=1 \).
2. While not converge AND the iteration number is under maximum do
   a. Randomly draw \( n \) from the distribution given by \( v_n \).
   b. Compute first and second derivatives of dual variable \( \alpha^k \) by Eq. (9) and (10), respectively.
   c. Update the dual variable \( \alpha^{k+1} \) by Eq. (11).
   d. Update \( t_{ni}^{k+1} \) (the \( n \)-th element of the latent vector) by Eq. (12).
   e. Update \( k = k + 1 \).
3. Return \( \hat{t}_i = t_i^k \).

---

### C. The sparsity step

The purpose of the DRWLS step is to train a robust latent vector to distinguish positive samples from negative ones. However, the weight vector, which is trained with the latent vector obtained in the DRWLS step, can not remove the redundant and irrelevant features in the large number channel features. Therefore, when the robust latent vector is trained, it can be used to obtain the sparse weight vector by using the solution of SPLS. The SPLS approach achieves feature selection by imposing sparsity on the weight vector. Therefore, we incorporate the new robust latent vector \( t \) trained from the DRWLS step into the solution of SPLS [26, 27]. Then, a robust sparse weight matrix can be obtained.

In [26, 27], the PLS is formulated in an alternative form:

\[
w = \arg \max_{w} w^T M w \tag{13}
\]

s.t. \( w^T w = 1 \)

where \( M = X^T y y^T X \).

Chun et al. [26] imposed a regularization term to this formulation to obtain a sparse solution and achieve variables selection. The formulation can be rewritten as:

\[
w = \arg \max_{w} w^T M w \tag{14}
\]

s.t. \( w^T w = 1 \), \( \|w\|_1 \leq \lambda \)

where \( \| \cdot \|_1 \) is the \( l_1 \)-norm and \( \lambda \) is a term to control the amount of sparsity.
Motivated by the success of the sparse principal component analysis (SPCA) [33], the work in [26] promotes exact zero property by imposing a regularization term on a new vector \( \beta \) instead of \( \mathbf{w} \), while keeping \( \beta \) and \( \mathbf{w} \) close to each other. Therefore, Eq. (14) can be reformulated as:

\[
< \mathbf{w}, \beta > = \arg \min_{\mathbf{w},\beta} \{- \mathbf{k}^T \mathbf{w} \mathbf{M} \mathbf{w} + (1-k)(\beta - \mathbf{w})^T \mathbf{M}(\beta - \mathbf{w}) + \lambda_1 \| \beta \|_1 + \lambda_2 \| \beta \|_2 \}
\]

s.t. \( \mathbf{w}^T \mathbf{w} = 1 \)

where \( \| \cdot \|_2 \) is the \( l_2 \)-norm. \( k \) can reduce the effect of concave part and mitigate the local solution issue. The generalized regression formulation of SPLS is solved by two alternative minimization stages (i.e. updating \( \mathbf{w} \) for a fixed \( \beta \) and updating \( \beta \) for a fixed \( \mathbf{w} \)). Since \( \beta \) and \( \mathbf{w} \) are close to each other, for univariate class labels, the final solution is obtained by using the following equation:

\[
\mathbf{w} = \beta = [\mathbf{Z} - \eta \max_{1 \leq j \leq M} z_j]_+ \text{sign}(\mathbf{Z})
\]

where \( \mathbf{Z} = \mathbf{X}^T \mathbf{y} / \| \mathbf{X}^T \mathbf{y} \| \) is the first weight vector of PLS, and \( z \) is an element of \( \mathbf{Z} \). \( \eta \) is a term to control the sparsity of the solution, which is a required input in the training stage. \([ \cdot ]_+ \) is a shrinkage operator. \( \text{sign}(\mathbf{Z}) \) is a sign function of \( \mathbf{Z} \). More details about the solution of SPLS can be found in [26].

The new robust latent vector \( \mathbf{t} \) trained in the DRWLS step can be used to represent the class labels \( \mathbf{y} \) because the residuals are small enough after the optimization. By incorporating \( \mathbf{y} = \mathbf{t} \mathbf{q}_i \) into the solution of SPLS, we obtain a new sparse weight vector:

\[
\mathbf{\hat{w}} = [\mathbf{Z} - \eta \max_{1 \leq j \leq M} z_j]_+ \text{sign}(\mathbf{Z})
\]

where \( \mathbf{Z} = \mathbf{X}^T \mathbf{t}_i \mathbf{q}_i / \| \mathbf{X}^T \mathbf{t}_i \mathbf{q}_i \| = \mathbf{X}^T \mathbf{t}_i / \| \mathbf{X}^T \mathbf{t}_i \| \). Compared with the original solution of SPLS, we incorporate the optimized latent vector obtained in the DRWLS step into the solution so that the latent vector and sparse weight vector can be optimized simultaneously.

\[\text{Algorithm 2: The proposed DWSPLS approach}\]

**Input:** The class labels \( \mathbf{y} \), the \( i \)-th latent vector \( \mathbf{t}_i \), the \( i \)-th loading scalar \( q_i \), the mean of the positive samples \( \mu_{\text{pos}} \), the mean of the negative samples \( \mu_{\text{neg}} \).

**Output:** The optimized latent vector \( \mathbf{\hat{t}}_i \), the optimized weight vector \( \mathbf{\hat{w}}_i \).

Initialize \( \mathbf{t}_i^{(0)} = \mathbf{t}_i \), \( \mathbf{w}_i^{(0)} = \mathbf{w}_i \), \( q_i^{(0)} = q_i \) and the iteration number \( k=1 \).

while not converge AND the iteration number is under maximum do

- **DRWLS step:**
  - Set the sample weights \( \nu_n = \exp(-y_n t_i q_i) \).
  - Update \( \mathbf{t}_i^{(k+1)} \) using the DRWLS step (see Algorithm 1).
  - Update \( q_i^{(k+1)} \) using the NIPALS algorithm.

- **Sparsity step:**
  - Update \( \mathbf{w}_i^{(k+1)} \) using the solution of SPLS. (see Eq. (17))

  Update \( k = k + 1 \).

end

Return \( \mathbf{\hat{t}}_i \) and \( \mathbf{\hat{w}}_i \).

D. The complete approach

In the above subsections, we have introduced the details of the two main steps of the proposed DWSPLS approach. Now we put them together to yield an iterative process for training a robust sparse weight matrix in Algorithm 2. In the DRWLS step, we solve the proposed DRWLS problem by using the Newton’s method to optimize the latent matrix. In the sparsity step, we incorporate the optimized latent matrix obtained from the DRWLS step into the solution of SPLS [26, 27] to update the sparse weight matrix. The color parts of the sparse weight matrix, which are non-zero, are used to select the most discriminative features from the large number of channel features.

E. Selecting features for human detection

The sparse weight matrix obtained by DWSPLS not only removes the redundant and irrelevant features, but also generates a large number of discriminative channel features for training the boosted classifier. We use the selected discriminative channel features to train the final strong boosted classifier and incorporate the weights of selected channel features into each decision tree.

Every node in the boosted decision trees is used to split the training samples by selecting a channel feature. Each decision tree is given a weight according to the classification error. We incorporate the corresponding weights of the selected features in the sparse weight matrix into the weight of each decision tree. The weight of each decision tree can be reformulated as:

\[
\alpha_t = \{ \sum_{i=1}^{K} w_i \log(1 - e_i)/\log(e_i) \} / K
\]

where \( K \) is the number of the selected channel features for classification in a decision tree; \( w_i \) is the weight of the \( i \)-th selected channel feature obtained from the sparse weight matrix; \( e_i \) is the classification error of the \( t \)-th decision tree. The final strong classifier for human detection is the combination of all the decision trees:

\[
H(x) = \text{sign} \left( \sum_{i=1}^{T} \alpha_i h_i(x) \right)
\]

where \( h_i(x) \) is a decision tree composed of a stump \( h_j(x) \) at every non-leaf node \( j \), \( T \) is the number of decision trees. Each stump produces a binary decision, and is parameterized with a polarity \( p \in \{ \pm 1 \} \), a threshold \( \tau \in \mathbb{R} \), and a feature index \( k \in \{ 1, 2, ..., K \} \):

\[
h_j(x) = p_j \text{sign}(x[k_j] - \tau_j)
\]

where \( x[k_j] \) indicates the \( k \)-th selected channel feature at the
Fig. 2: Visualization of spatial distribution of selected features for each channel type. Red pixels indicate that the features with high importance are selected.

non-leaf node $j$.

IV. EXPERIMENTS

In this section, we first show how the proposed approach selects the discriminative features in a detection window in Section A. In Section B, we discuss the evaluation criteria and the parameter settings in our experiments. In Section C, we compare the proposed DWSPLS approach with sparse principal components analysis (SPCA) for feature selection. In Section D, we evaluate the performance of the proposed DWSPLS approach on several challenging human datasets and compare it with some state-of-the-art human detection approaches. Finally, the runtime analysis of the proposed approach is given in Section E.

In our experiments, we incorporate the proposed approach into the LDCF framework [3] for human detection, which has show its accuracy and efficiency. Given a detection window of $64 \times 128$ pixels, we compute the ten feature channels: LUV color channels (3 channels), gradient magnitude (1 channel) and gradient histograms (6 channels). The output channel features are computed by summing the corresponding patches of the input pixels. Like in LDCF [3], we also use decorrelation transform as a pre-processing step for the channel features. Finally, there are $16 \times 32 \times 4 \times 10$ (20,480) candidate features in a detection window. In the following experiments, the human detector are trained on INRIA pedestrian [1] dataset by using the windows of $64 \times 128$ pixels except the Caltech [29] and TUD-Brussels [28] pedestrian dataset.

A. Feature visualization

To illustrate how the proposed approach selects the discriminative features in a detection window, we first train a DWSPLS model for each feature channel. Then, we visualize the weights of the first three latent variables for each pixel.

Fig. 2 visualizes the spatial distribution of selected features for each channel type. As we can seen, different parts of the human body can be selected for the different latent variables by the proposed approach. Most of the redundant and irrelevant features of the human body, which are shown in blue in Fig. 2, are discarded by feature selection. Red pixels indicate that the most discriminative features with high importance are selected in the detection window. Most of the selected features correspond to the human body. An intuitive result is that the 'U' channel features are selected at the location of human head/face while the gradient magnitude features are selected around the human contours and the gradient histogram features are selected along the gradient orientation.

B. Evaluation criteria and parameter settings

For quantitatively evaluating the detection performance, the receiver operating characteristic (ROC) curve which plots the false positive per image (FPPI) versus miss rate on the log-scales of x-axis and y-axis, is used as the evaluation criterion. Measurements of average miss rates are used to summarize the overall performances of different detectors. All the human detection results are evaluated using the toolbox provided by the Caltech pedestrian dataset [29]. To optimize our detector, we use the INRIA pedestrian dataset [1] for parameter tuning. The training dataset contains 2,416 positive images and 1,218 negative images. Like in [3], we use a number of validation sets which contain 22,226 positive samples and 10,000 negative samples randomly sampled from the training set.

Number of latent variables. To evaluate the influence of the number of the latent variables used in the proposed approach on the final detection performance, we evaluate the detector trained by DWSPLS with different number of latent variables on the INRIA pedestrian dataset [1]. Fig. 3 shows the ROC curves obtained by DWSPLS according to
different numbers of latent variables. From Fig. 3, we can see that DWSPLS achieves the best performance with a 12.32% average miss rate when only using the first 4 latent variables. Therefore, in the following experiments, we fix the latent variables to be 4.

**Sparsity level.** Another important issue of parameter setting is the sparsity level of the weight vector. Sparsity of the weight vector \( w \) can be defined as the percentage of non-zero features in \( w \), that is

\[
\text{Sparsity}(w) = \frac{\|w\|_0}{M}
\]

where \( \| \cdot \|_0 \) denotes the zero-norm and \( M \) is the dimension of \( w \).

Fig. 4 shows the ROC curves with different sparsity levels of the weight vectors. As shown in Fig. 4, we can see that when the sparsity is very low (under 25%), the number of features is not sufficient to discriminate samples. Thus, the proposed approach obtains high miss rates. However, when the sparsity is more than 25%, it can achieve better performance since the selected channel features contain many redundant and irrelevant features. When the sparsity is 100% in the sparsity step, we impose no sparsity on the weight matrix. Therefore, the detector trained with the selected channel features exhibits a better performance than that trained without feature selection. In addition, because of the weighted boosted decision trees, the proposed approach, which does not apply feature selection, still outperforms the original LDCF framework. In the following experiments, we fix the sparsity level to be 25%. After applying feature selection, the number of selected channel features is \( 128 \times 4 \times 10 \) (5,120) for each latent variable. As mentioned before, the number of latent variables is set to be 4, which results in a total of \( 5,120 \times 4 \) (20,480) in our configuration.

**C. SPCA vs. DWSPLS**

SPCA [33] is also a popular approach for feature selection. Therefore, we perform feature selection by using SPCA on the INRIA pedestrian dataset and compare its results with DWSPLS. For quantitatively evaluating the detection performance, we evaluate the detector trained by SPCA and DWSPLS with different numbers of principal vectors (latent variables). The sparsity levels of principal vectors in SPCA are the same as those of latent variables in DWSPLS. Fig. 5(a) shows the ROC curves obtained by SPCA according to different numbers of principal vectors. It can be seen that the lowest average miss rate (13.40%) for SPCA is achieved when the number of principal vectors is set to be 6. Fig. 5(b) shows the average miss rates obtained by SPCA and DWSPLS according to different numbers of principal vectors (latent variables). As we can see, DWSPLS clearly achieves a lower average miss rate than SPCA when using the same number of principal vectors (latent variables).

**D. Human detection with DWSPLS**

We evaluate the performance of the proposed DWSPLS approach on three challenging human datasets including INRIA [1], Caltech [29] and TUD-Brussels [28] pedestrian datasets. All the detection results are evaluated by using the ROC curves on the log-scales. We compare the proposed approach with some other state-of-the-art human detection approaches by using the publicly available Caltech benchmark code [29].

For INRIA, 22,266 positive samples and 5,000 negative samples randomly sampled from 1,218 background training images are first used to train an initial human detector after decorrelation transform. Afterwards, we iteratively classify the negative samples and retrain a new detector with the false positive samples (i.e. hard examples). After several rounds, 10,000 negative samples are included in our final detector. There are 288 positive samples for testing in INRIA. The ROC
Fig. 5: Comparison of SPCA and DWSPLS for feature selection. (a) The evaluation results by using SPCA with different numbers of principal vectors. (b) Comparison of the average miss rates obtained by SPCA and DWSPLS when using the same number of principal vectors (latent variables).

Fig. 6: Performance comparison between the proposed approach and the other state-of-the-art approaches on INRIA.

For Caltech, it consists of approximately 10 hours of 640 × 480 30-Hz video. The training data contains six training sets and the test data contains five sets. The window size of training samples are set to 32 × 64 pixels in our experiments. After feature selection, there are 32 × 4 × 10 × 4 (5,120) candidate features for training in a detection window. The detector trained on Caltech dataset is evaluated on both Caltech and TUD-Brussels datasets. The ROC curves obtained by the proposed approach and several other state-of-the-art approaches are given for each dataset in Fig. 7. As we can see, the proposed approach reduces the log-average miss rate over the LDCF [3] approach by 2.4% and 1.5% for each dataset, respectively. Compared with some state-of-the-art approaches, the proposed approach also achieves a competitive performance.

It is important to point out that the improvement of SpatialPooling [22] comes at the cost of increased computational complexity. The time taken on the Caltech dataset by SpatialPooling is about 0.119 frames per second. The proposed approach, however, almost do not adds computational costs compared to LDCF [3] except the weighting scheme for the decision trees. Our detector operates at approximately 2.707 frames per second on the Caltech dataset, which is about 20 times faster than SpatialPooling.

E. Runtime

Our detector is implemented in MATLAB on an Intel Core-i7 CPU (3.4 GHz). As mentioned above, on the Caltech data set, there are 32 × 4 × 10 × 4 (5,120) candidate features in a detection window after feature selection. We do not use the selected channel features for dimensionality reduction but to train the boosted classifier. Therefore, compared with the LD-CF [3] framework, the only added complexity is the weighting scheme for the decision trees. It takes about 0.4 s for testing a whole 640 × 480 image by using the optimal parameters. Our detector is expected to reach real-time efficiency running on a powerful machine.
V. CONCLUSION

We have presented a robust discriminative weighted sparse approach (called DWSPLS) for human detection. First, we obtain a robust latent matrix by solving the proposed DRWLS problem. A discriminative term is incorporated during the training stage, which can improve the robustness and discriminativeness of the latent matrix. In order to reduce the computational complexity, the DRWLS problem is solved in its dual form by using the Newton’s method. The optimized latent matrix can better distinguish positive samples from negative ones. Second, the optimized latent matrix is incorporated into the solution of SPLS. We introduce the SPLS algorithm to achieve feature selection by imposing sparsity on the weight matrix, where the most discriminative features can be selected while redundant and irrelevant features are removed. Finally, we use the selected channel features to train the boosted decision trees and incorporate the weights of selected features into each tree. Experimental results on three challenging human datasets have demonstrated that the proposed approach effectively improves the performance of the original LDCF approach for human detection and is competitive to several other state-of-the-art competing human detection approaches.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grants 61472334, 61170179, and 61201359, and supported by the Fundamental Research Funds for the Central Universities under Grant 20720130720.

REFERENCES


[35] Hanzi Wang is currently a Chair of the Professor Committee in the School of IST at Xiamen University (XMU) in China, a Distinguished Professor of “Minjiang Scholars” in Fujian province and a Founding Director of the Center for Pattern Analysis and Machine Intelligence (CPAMI) at XMU. He served as an Adjunct Professor (2010-2012) and a Senior Research Fellow (2008-2010) at the University of Adelaide, Australia; an Assistant Research Scientist (2007-2008) and a Postdoctoral Fellow (2006-2007) at the Johns Hopkins University; and a Research Fellow at Monash University, Australia (2004-2006). He received his Ph.D degree in Computer Vision from Monash University where he was awarded the Douglas Lampard Electrical Engineering Research Prize and Medal for the best PhD thesis in the Department. His research interests are concentrated on computer vision and pattern recognition including visual tracking, robust statistics, object detection, video segmentation, model fitting, optical flow calculation, 3D structure from motion, image segmentation and related fields.

[36] Qiming Li is a senior member of the IEEE. He is an Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT) and was a Guest Editor of Pattern Recognition Letters (September 2009). He was the General Chair for ICIMCS2014, Program Chair for CVRS2012, Publicity Chair for IEEE NAS2012, and Area Chair for DICTA2010. He also serves on the program committee (PC) of ICCV, ECCV, CVPR, ACCV, PAKDD, ICIG, ADMA, CISB, etc. and he serves on the reviewer panel for more than 40 journals and conferences.