

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Improving Human Detection Operators by Dimensionality Reduction

Ling Mao and Mei Xie

School of Electronic Engineering, University of Electronic Science and Technology of China,  
Chengdu, Sichuan, 611731, China

---

**Abstract:** Approaches based on Haar wavelets and Histograms of Oriented Gradients (HOG) are very successful in human detection. However, the dimensions of features extracted based on both of the operators are very high which often causes low computation efficiency and needs more memory. Usually these high-dimensional features contain redundant information which would deteriorate the performance. To conquer these problems, we proposed a new algorithm based on Partial Least Squares (PLS) technique to extract a few latent variables from original Haar wavelet features or HOG features which contain most useful information and basically no noise, in other words, these new latent variables provide higher discriminatory power than original features. Therefore, the proposed method could save memory space, reduce computation cost and improve the performance of original operators. Additionally, we design a method to get the optimal number of feature dimensions. The experiments conducted on three varied datasets including Massachusetts Institute of Technology (MIT), Institut National de Recherche en Informatique et en Automatique (INRIA) and DaimlerChrysler human datasets show that our method based on PLS obtains better discrimination results than Principal Components Analysis and outperforms original operators.

**Key words:** Haar wavelets, histograms of oriented gradients, human detection, partial least squares

---

### INTRODUCTION

In the past decade, many human detection operators have been proposed (Papageorgiou and Poggio, 2000; Viola *et al.*, 2005; Dalal and Triggs, 2005; Zhu *et al.*, 2006; Wang *et al.*, 2009; Wu and Nevatia, 2008; Chen and Chen, 2008). Among these operators, Haar wavelets (Papageorgiou and Poggio, 2000) and Histograms of Oriented Gradients (HOG) (Dalal and Triggs, 2005) are canonical since both of them perform very well (Paisitkriangkrai *et al.*, 2008) and the subsequent successful human detection operators are almost based on them (Viola *et al.*, 2005; Zhu *et al.*, 2006; Wang *et al.*, 2009; Wu and Nevatia, 2008; Chen and Chen, 2008). Usually, these descriptors are computed over dense and overlapping grids of spatial blocks in the detection of each detection window to improve performance of human detection. Consequently, dimensions of the extracted features are very high, e.g., thousands upon thousands dimensions which leads to low computation efficiency, and requires a significant amount of memory space. If large numbers of samples are used for training, the procedure often runs out of memory. To handle this problem, dimensionality reduction methods are generally adopted. Actually these features of high dimensions contain much redundant information which is comprised

of useful information and useless information (noise) for discrimination. Theoretically, we can find an approach to extract some latent variables (principal components) contain most useful information and basically no noise to improve performance further. In this sense, here the dimensionality reduction involves two aspects: (1) the dimensions of the features should be reduced and (2) the new low dimensional features should contain most of useful information and basically no noise. The new low dimensional features are often called latent variables or principal components.

Principal Components Analysis (PCA) method (Jolliffe, 2002) is a traditional dimensionality reduction technique widely applied to computer vision problems (Ke and Sukthankar, 2004). PCA extract principal components by maximizing the projected variance. However, the total projected variance being maximized is due not only to the between-class variance that is useful for classification, but also to the within-class variance that, for classification purposes, is unwanted information. Hence, sometimes not the most important principal components but the less important principal components extracted by PCA are helpful for classification (Belhumeur *et al.*, 1997). This drawback causes difficulty in deciding which principal components are suitable for the classification.

Zhu *et al.* (2006) and Chen and Chen (2008) use AdaBoost for feature selection which significantly speeds up the computation, but the feature selection is executed implicitly and don't give an explicit explanation about how the extracted features contribute the detection results.

To deal with these problems above, this study has proposed an algorithm based on PLS.

The Partial Least Squares (PLS) method (Wold, 1975) has been a popular modeling, regression and classification technique in its domain of origin chemometrics. PLS combines features from generalized principal components analysis and Canonical Correlation Analysis (CCA) (Hotelling, 1936). Similar to PCA, the underlying assumption of all PLS methods is that the observed data is generated by a system or process which is driven by a small number of latent (not directly observed or measured) variables (principal components), and these latent variables should represent original data as much as possible. Additionally the correlation between latent variables and response variable (class label) should be as large as possible.

PLS method possesses some important features which are very helpful for dimensionality reduction and analysis of the principal components (Wang, 1999): (1) these extracted latent variables are orthogonal to each other, that means they are uncorrelated, (2) the first several latent variables contain most variation of the original variables and interpret the response variable (class label) very well and (3) the latent variables are always extracted from the residual matrices which guarantees information included in each latent variable is complementary.

### PROPOSED METHOD BASED ON PARTIAL LEAST SQUARES ANALYSIS

Usually dimensions of HOG features and Haar wavelet features (Papageorgiou and Poggio, 2000; Dalal and Triggs, 2005) are very high. These high dimensional features contain redundant information, need much memory space to store and are computed less efficiently. We propose a new method based on PLS to solve these problems. The flow diagram of the algorithm is shown in Fig. 1. In the training phase, we collect all the high dimensional HOG or Haar wavelet features of positive and negative samples. Then, based on PLS, we extract desired number of latent variables and weight vectors from these original features, where the number of latent variables is decided by some ratio value. At last we use the new low dimensional features to train linear SVM model. In the test phase, similarly, we firstly obtain high dimensional HOG or Haar wavelet features of test images

and then compute the inner products of weight vectors obtained in the training phase and original features. Finally these latent vectors are used as inputs of linear SVM classifier to discriminate whether the input image is a person or not. In this study, Linear LIBSVM (Chang and Lin, 2001) is chosen as the classifier. These important steps are detailed in the following subsections.

**HOG and Haar wavelets:** The original Haar wavelets (Papageorgiou and Poggio, 2000) of the size 32×32, 16×16 and 8×8 are adopted as illustrated in Fig. 2.

This study uses the HOG based on rectangular (R-HOG) blocks (Dalal and Triggs, 2005). To emphasize the improvement over original descriptors led by our method, we abandon some optimization operations in the original paper. These differences are: (1) we normalize gray images instead of color channels, (2) we don't smooth image gradients using Gaussian kernel and (3) we compute the histogram using gradient magnitudes directly without trilinear interpolation. Similar to Zhu *et al.* (2006), we also use integral image technique of Viola and Jones (2001) and Porikli (2005) to compute efficiently. In experiments, we set each block size 2×2 and each cell size 8×8 and 6×6, respectively to INRIA and DaimlerChrysler datasets and the stride (block overlap) is fixed at half of the block size. Each cell consists of a 9-bin histogram of oriented gradients (unsigned) and each block contains a concatenated vector of all its cells. Each block is thus represented by a 36-D feature vector that is then normalized to an L2 unit length:

$$v \leftarrow v / \sqrt{\|v\|_2^2 + \epsilon^2}$$

Then, weighted by a set of weight vectors, these high dimensional Haar wavelet or HOG features are combined based on PLS to obtain fewer latent variables which provide higher discriminatory power.

#### Partial least squares

**Dimensionality reduction:** The origins of PLS are traced to Herman Wold's original non-linear iterative partial least squares (NIPALS) algorithm, an algorithm developed to linearize models which were non-linear in the parameters (Wold, 1975). An overview of PLS methods can be found by Rosipal and Kramer (2006). The procedure of dimensionality reduction based on PLS is provided below.

Let matrix  $X_{n \times p}$  represent all the HOG or Haar wavelet features extracted from training samples, where  $n$  is the number of samples and  $p$  is the dimensionality of the feature vector. Similarly let vector  $Y_{n \times 1}$  represent the class labels of training samples. After normalizing process, we

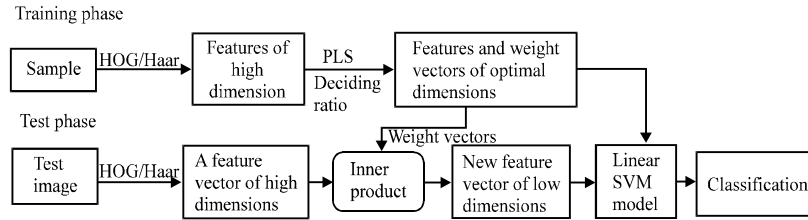


Fig. 1: Flowchart of the proposed pedestrian detection method based on partial least squares analysis

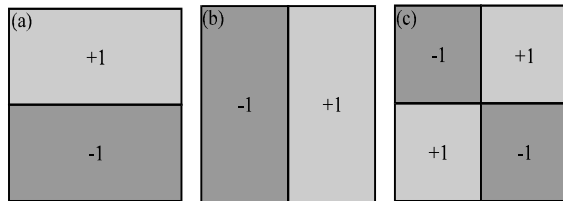


Fig. 2(a-c): The non-standard Haar wavelets, (a) Vertical, (b) Horizontal and (c) Diagonal Haar wavelets

obtain matrices  $E_0 = (E_{01}, \dots, E_{0p})_{n \times p}$  and  $F_0 = (y_0)_{n \times 1}$ , respectively whose column vectors possess zero mean and unit variance.

Let  $t_1$  be the first latent variable wanted, i.e.:

$$t_1 = E_0 w_1$$

where,  $w_1$  means the weight vector or direction vector, and it's set to a vector of unit length for convenience.

To obtain the first latent variable, two conditions should be satisfied: (1)  $t_1$  should represent most variation of observation data  $E_0$  and (2) the correlation between  $t_1$  and  $F_0$  should be maximized, in other words,  $t_1$  should possess the most explanation to  $F_0$ . The equivalent condition is that the covariance between  $t_1$  and  $F_0$  should be maximized Eq. 1:

$$\text{Cov}(t_1, F_0) = \sqrt{\text{Var}(t_1)\text{Var}(F_0)} r(t_1, F_0) \rightarrow \max \quad (1)$$

The final formula is:

$$\begin{aligned} \max_{w_1} \langle E_0 w_1, F_0 \rangle \\ \text{s.t. } w_1^T w_1 = 1 \end{aligned} \quad (2)$$

It's notable that the covariance of variables  $t_1$  and  $F_0$  is obtained by computing the inner product between both vectors  $E_0 w_1$  and  $F_0$  since the column vectors of  $E_0$  and  $F_0$  possess zero mean and unit variance.

Then the weight vector  $w_1$  is obtained using Lagrange Multipliers and  $E_0$  and  $F_0$  are reconstructed from  $t_1$ , which is also called regression Eq. 3:

$$\begin{aligned} w_1 &= \frac{E_0^T F_0}{\|E_0^T F_0\|} \\ E_0 &= t_1 p_1 + E_1 \\ F_0 &= t_1 r_1 + F_1 \\ p_1 &= \frac{E_0^T t_1}{\|t_1\|^2}, r_1 = \frac{F_0^T t_1}{\|t_1\|^2} \end{aligned} \quad (3)$$

where,  $p_1$  and  $r_1$  are regression coefficients,  $E_1$  and  $F_1$  are residual matrices. To obtain the second latent variable  $t_2$  and the second direction vector  $w_2$ , substitute  $E_1$  and  $F_1$  for  $E_0$  and  $F_0$ , respectively and the rest can be done in the same manner until desired number of latent variables had been extracted.

These latent variables  $t_i$  are used to train the linear SVM classifier and the weight vectors  $w_i$  are used to obtain new latent vector from the high dimensional feature vector, extracted from a detection window. This latent vector is used in classification.

**Properties analysis:** In the extraction of latent variables, maximizing the variance of latent variable  $t_1$  and the correlation between  $t_1$  and  $F_{1,1}$  essentially makes sure that the most useful information for discrimination concentrates on the first several latent variables. Thus the latent variables could be easily interpreted physically, e.g., we can "see" the visual cues provided by these latent variables, as is shown in Fig. 6 and 7, the first several weight vectors of Haar wavelet and HOG features contain major of the silhouette contours of the body. Those silhouette contours are important for human detection which is consistent with analysis in Eq. 3. The deflation procedure, i.e., the obtainment of  $E_1$  and  $F_1$  ( $I \sim 0$ ), makes sure that information included in each latent variable is complementary. Based on the properties above, it can be easily deduced that useless information for classification is basically removed from the most important components. That is why the dimensions of features are fewer but better performance is obtained, as is shown in Fig. 4.

Principal components obtained using PCA don't have these good properties mentioned above. As Barker (2003) pointed out: when discrimination is

ultimately the goal and dimension reduction is needed, PLS is to be preferred to PCA.

**Optimal number of latent variables:** Generally, the more latent variables are used the more useful information is included and the result is better. However, as mentioned above, those lower latent variables would contain much noise. In this case, latent variables increase but performance degrades. Hence it's important to decide how many latent variables are optimal. In this study an adapted method proposed by Abdi (2010) is used to compute the optimal number of latent variables.

When  $l$  latent variables are extracted, there are two methods to evaluate the overall quality of the regression model. Denote  $RESS = \|Y - Y^l\|^2$  as the similarity between  $Y^l$  and  $Y$ , where  $Y$  is the response vector (class label) and  $Y^l$  is the reconstruction of  $Y$  from the  $l$  latent variables obtained by all components in  $Y$ . Different from the first one, the second evaluation method adopts leave-one-out technique, i.e., cross validated technique, to obtain the  $l$  latent variables and reconstruct  $Y$ , denoted as  $PRESS = \|Y - Y^l\|^2$ .

Generally,  $PRESS$  is bigger than  $RESS$  when the same  $l$  latent variables are used for reconstruction. Both  $PRESS$  and  $RESS$  will decrease as latent variables increase. However, when they decrease very slowly or even decrease, it indicates that the model is overfitting the data, i.e., much unwanted information is included in the lower latent variables.

Now we can define the parameter  $Q_l^2$  (Eq. 4) to decide the optimal number of latent variables:

$$Q_l^2 = 1 - \frac{PRESS_l}{RESS_{l-1}} \quad (4)$$

The subscript  $l$  means the number of latent variables. The decision rule is that a latent variable is kept if its value of  $Q_l^2$  is larger than some arbitrary value generally set equal to  $(1 - 0.95^2) = 0.0975$ . In this study, we use the ratio:

$$R = \frac{PRESS_l}{RESS_{l-1}}$$

and set the threshold to 1.

## RESULTS

**Datasets and operators adopted:** All experiments were conducted on three different human datasets: MIT pedestrian database (Papageorgiou and Poggio, 2000), INRIA person dataset (Dalal and Triggs, 2005) and DaimlerChrysler Pedestrian Dataset (Enzweiler and Gavrila, 2009). In addition, the MIT car database (Papageorgiou and Poggio, 2000) was also adopted in the experiment Fig. 3 shows some samples mentioned above.

Haar wavelet and HOG operators were adopted in experiments and the parameter set of Haar wavelets and



Fig. 3(a-d): Some sample images from the selected datasets, (a) INRIA dataset; a wide range of variations in pose, appearance, clothing, illumination and background, (b) MIT humans which are always upright and with a little variations in pose etc., (c) DaimlerChrysler (DC) dataset which are gray images and (d) MIT cars dataset

Table 1: The parameters of each dataset for training and test

Dataset	Operator	Training samples		Test samples		Dimensions of features
		Positive	Negative	Positive	Negative	
MIT human	Haar wavelets	1328	5672	520	505242	6633
INRIA	HOG	2416	12180	1126	505242	3780
DaimlerChrysler	HOG	3936	11882	1085	500080	3780
MIT cars	Haar wavelets	636	3534	396	250240	14193

HOG is described in section ‘HOG and Haar Wavelets’. The number of samples used for training and test and other parameters are shown in Table 1.

**Dimensionality reduction and improvement of performance:** To evaluate the performance of the proposed algorithm based on PLS, Detection Error Tradeoff (DET) curves is used. In DET, the x-axis corresponds to False Positives Per Window (FPPW), defined by  $\text{FalsePos}/(\text{TrueNeg}+\text{FalsePos})$  and the y-axis shows the Miss Detection Rate, defined by  $\text{FalseNeg}/(\text{FalseNeg}+\text{TruePos})$ .

In Fig. 4a-d, Detection Error Tradeoff (DET) curves show the performance achieved by original features and different latent variables extracted by our method. Figure 4a shows the results on MIT human dataset which conducted using Haar wavelets. Figure 4b and c show the results on INRIA human and DaimlerChrysler human dataset with HOG features. The results on MIT car dataset are depicted in Fig. 4d. We plot DET curves on a log-log scale, i.e., the x-axis corresponds to False Positives Per Window (FPPW), defined by  $\text{FalsePos}/(\text{TrueNeg}+\text{FalsePos})$  and the y-axis shows the Miss Detection Rate, defined by  $\text{FalseNeg}/(\text{FalseNeg}+\text{TruePos})$ .

The experiment results are shown in Fig. 4. It’s apparent that the first most important latent variables chosen outperform the original features whose dimensions are thousands upon thousands. The results obtained by proposed latent variables show significant improvements in regions of low false alarm rates. The miss rate at  $10^{-4}$  FPPW is improved by at least 5%. Besides, it’s interesting to note that the miss rate decrease at low FPPW as latent variables increase, while when more variables are used, the miss rate increases. For example, in Fig. 4a, the DET curve for four variables is obviously worse than original features and when 10 variables are adopted the best DET curve is achieved, but the DET curve for 20 variables is worse, in which the performance drops by 3% at  $10^{-4}$  FPPW than the best DET curve for 10 variables. If more latent variables are adopted, the result will be even worse. This result is consistent with our analysis in section 2.2 and 2.3: most useful information is contained by the first several latent variables, but the lower variables contain more useless information which will reduce performance. Therefore, well chosen latent variables will noticeably improve classification results.

**PLS vs. PCA:** PCA loses out to PLS in the dimensionality reduction for discrimination application, as shown in Fig. 5. It shows the samples (two hundreds of positive and negative samples from DaimlerChrysler Pedestrian Dataset) projected onto the first two dimensions for PLS and PCA. PLS clearly achieves better class separation than PCA.

If PCA method is used to extract principal components, more components are needed to guarantee that performance won’t drop substantially which is depicted in Fig. 5c. The DET curve obtained based on features in PCA subspace is closer to the original DET curve as the dimensionality increases. While fewer components are used, the performance decreases very much, especially the miss rate drops by 30% for 20 components. The best DET curve closest to the original one is obtained using 400 components, in contrast to 12 components based on PLS. Using more components means more storage memory space and more computation cost. That’s why PLS is better than PCA.

**Obtain the optimal number of latent variables:** The number of latent variables corresponding to the best performance can be easily found around the ratio value of 1 according to the proposed method (Table 2). The increasing ratio values which are much more than 1 indicate that more useless information is included by more latent variables.

**“See” latent variables:** As is analyzed before, the extracted latent variables based on PLS can be easily interpreted physically. Especially the latent variables of HOG and Haar wavelet features provide enough visual cues to see how the information contained in these latent variables is related to classification. In this study, we execute the “see” operation by analyzing weight vectors, because latent variables are linear combination of weight vectors and original feature vectors and the weight vectors decide the nature of latent variables.

When the Haar wavelets used to extract features on MIT human dataset, 6633 features are obtained for a detection window and the corresponding weight vector’s length is 6633 too. The horizontal wavelet of size  $8 \times 8$  contributes 1769 elements of weight vector, which is depicted as gray image of size  $61 \times 29$  in Fig. 6. Figure 6a and b show the weight vectors images corresponding to the first 4 latent variables for PLS, emphasizing the

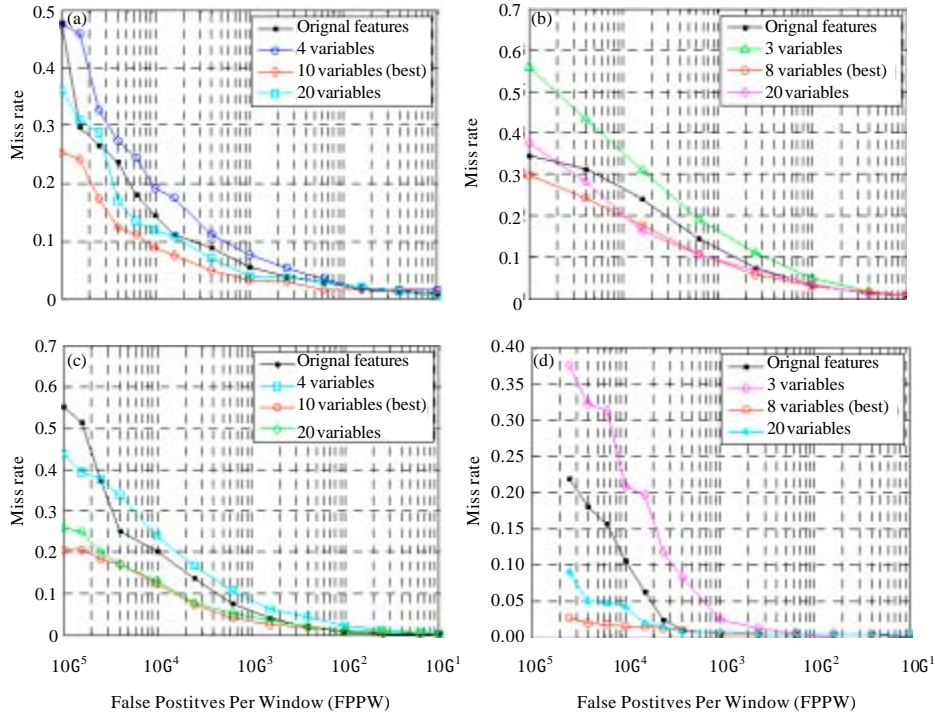


Fig. 4(a-d): The Detection Error Tradeoff (DET) curves show the performance achieved by original features and different latent variables extracted by our method. See the text for details. (a) MIT human dataset conducted using Haar wavelets, (b) and (c) INRIA human and DaimlerChrysler human dataset with HOG features and (d) MIT car dataset are depicted in. We plot DET curves on a log-log scale, i.e., the x-axis corresponds to false positives per window (FPPW), defined by  $\text{FalsePos}/(\text{TrueNeg}+\text{FalsePos})$  and the y-axis shows the Miss Detection Rate, defined by  $\text{FalseNeg}/(\text{FalseNeg}+\text{TruePos})$

Table 2: The ratio values corresponding to different latent variables on each dataset

Dataset	Ratio values corresponding to different number of latent variables								Ratio value	Optimal number
	6	7	8	9	10	11	12	13		
MIT human	0.77	0.81	0.95	1.03	1.10	1.17	1.25	1.31	1.10	10
INRIA	0.90	1.02	1.09	1.12	1.15	1.18	1.20	1.23	1.09	8
DaimlerChrysler	0.78	0.93	1.02	1.10	1.13	1.16	1.19	1.21	1.19	12
MIT cars	0.84	1.02	1.07	1.13	1.25	1.46	1.53	1.65	1.07	8

silhouette contour of human body (especially the head, shoulders and feet) which is important for classification. Specifically for PLS the first and second weight vectors illustrate the detector cues are mainly on the contrast of silhouette contours against the background and the third and fourth weight vectors mainly on the small areas around the person and it's apparent that the information is complementary. Notably the 15th weight vector which contains some detail information and much noise.

The second row images in Fig. 6 correspond to PCA. As is analyzed before, not only useful information but also useless information or noise would likely distributes in all the components. Here, the 2nd and 3rd emphasize

the human body contour which is useful to classification, but the 1st, 4th and 15th contain more noise. In this case we can't easily choose suitable components to complete classification and usually many components are needed to include useful information as much as possible (Fig. 4).

Figure 7 shows the analysis of weight vectors of HOG. Figure 7a represents the average gradient image over the training examples. Figure 7b and c illustrate the first and second weight vectors. Specifically Fig. 7b emphasizes the arms, body and legs which often cause vertical edges and Fig. 7c emphasizes arms, shoulders and heads. Figure 7d represents the fifteenth

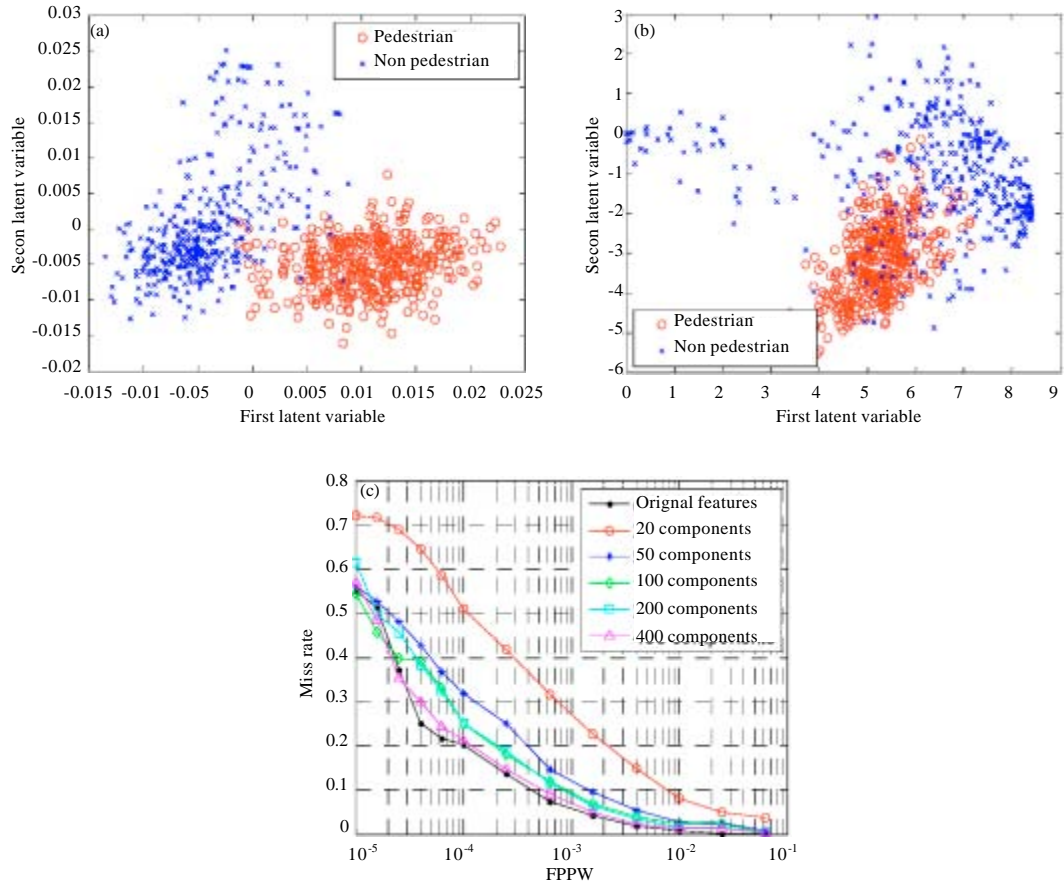


Fig. 5(a-c): Comparison of PCA and PLS for discrimination application, (a) Projection of the first two dimensions of 400 training samples for PLS, (b) Counterpart of PLS, i.e., PCA, From (a) and (b) we can find that PLS is more powerful than PCA in discrimination and (c) DET curves according to the number of used dimensions

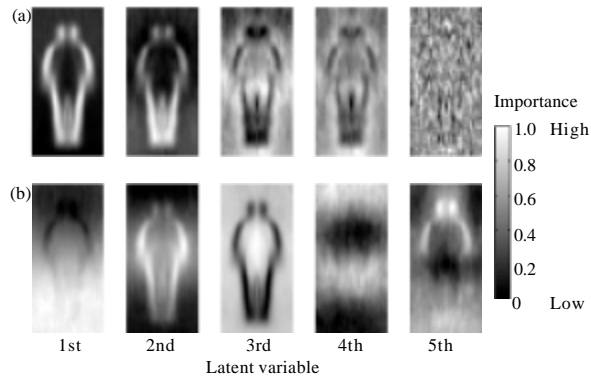


Fig. 6(a-b): Weight vectors for different principle components of (a) PLS and (b) Counter parts of PLS i.e., PCA within the detection window, according to horizontal Haar wavelet of size  $8 \times 8$



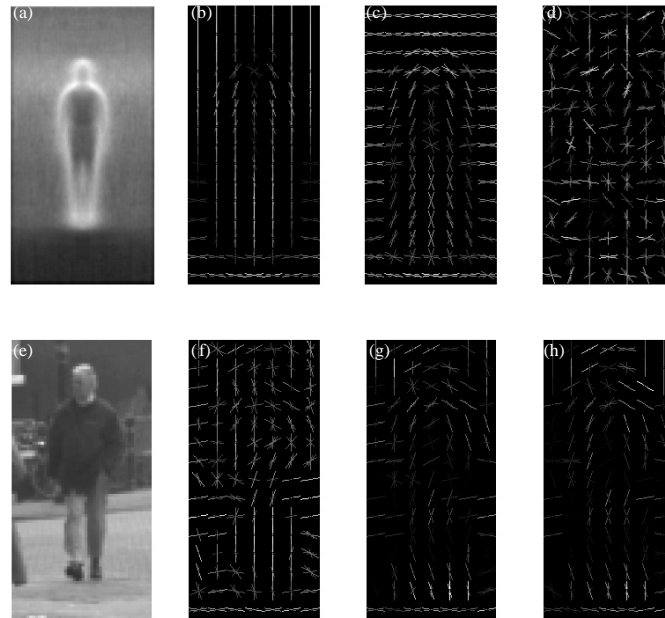


Fig. 7(a-h): For the latent variables of HOG, visual cues are also mainly on silhouette contours too (especially the head, shoulders, arms and feet). Only the dominant orientation is shown for each cell

weight vector which contain more noise. Figure 7e is a image from test samples and Fig. 7f is the original HOG descriptor. Figure 7h illustrates the dominant orientations selected by the SVM (obtained by multiplying the feature vector by the corresponding weights from the linear SVM). Figure 7g is similar to Fig. 7h, but the HOG descriptor is reconstructed from the 12 latent variables. Examination of Fig. 7e-h shows that the extracted 12 latent variables perform well as original features do.

### CONCLUSION

We have proposed an algorithm based on PLS to extract a small number of latent variables which contain most useful information and little useless information and solved the problem on that how many latent variables are optimal by proposed technique. We have tested the proposed method on different databases and demonstrated its better discrimination than PCA method, and shown it with fewer latent variables to outperform original HOG and Haar wavelet operators. In addition, we are planning to apply PLS to other human detectors, e.g., Gradient Local Auto-correlations (GLAC).

### ACKNOWLEDGMENT

This study was supported by Sichuan Provincial Department of Science and Technology. The Grant Number is: M110102012010GZ0153.

### REFERENCES

- Abdi, H., 2010. Partial least squares regression and projection on latent structure regression (PLS regression). *Comput. Stat.*, 2: 97-106.
- Barker, M., 2003. Partial least squares for discrimination. *J. Chemom.*, 17: 166-173.
- Belhumeur, P.N., J.P. Hespanha and D.J. Kriegman, 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intel.*, 19: 711-720.
- Chang, C.C. and C.J. Lin, 2001. Libsvm: A library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- Chen, Y.T. and C.S. Chen, 2008. Fast human detection using a novel boosted cascading structure with meta stages. *IEEE Trans. Image Process.*, 17: 1452-1464.
- Dalal, N. and B. Triggs, 2005. Histograms of oriented gradients for human detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, June 20-26, 2005, San Diego, CA., USA., pp: 886-893.
- Enzweiler, M. and D.M. Gavrila, 2009. Monocular pedestrian detection: Survey and experiments. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31: 2179-2195.
- Hotelling, H., 1936. Relations between two sets of varieties. *Biometrika*, 28: 321-377.
- Jolliffe, I.T., 2002. *Principal Component Analysis*. 2nd Edn., Springer-Verlag, New York.

- Ke, Y. and R. Sukthankar, 2004. PCA-SIFT: A more distinctive representation for local image descriptors. *Comput. Vision Pattern Recognition*, 2: 506-513.
- Paisitkriangkrai, S., C. Shen and J. Zhang, 2008. Performance evaluation of local features in human classification and detection. *IET Comput. Vision*, 2: 236-246.
- Papageorgiou, C. and T. Poggio, 2000. A trainable system for object detection. *Int. J. Comput. Vision*, 38: 15-33.
- Porikli, F., 2005. Integral histogram: A fast way to extract histograms in Cartesian spaces. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, June 20-25, 2005, Cambridge, MA., USA., pp: 829-836.
- Rosipal, R. and N. Kramer, 2006. Overview and recent advances in partial least squares. *Proceedings of the Statistical and Optimization Perspectives Workshop on Subspace, Latent Structure and Feature Selection*, February 23-25, 2005, Bohinj, Slovenia, pp: 34-51.
- Viola, P. and M. Jones, 2001. Rapid object detection using a boosted cascade of simple features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, December 8-14, 2001, Kauai, HI., USA., pp: 511-518.
- Viola, P., M.J. Jones and D. Snow, 2005. Detecting pedestrians using patterns of motion and appearance. *Int. J. Comput. Vision*, 63: 153-161.
- Wang, H., 1999. *Partial Least-Squares Regression-Method and Applications*. 1st Edn., National Defense Industry Press, Beijing, China.
- Wang, X., T. Han and S. Yan, 2009. An hog-lbp human detector with partial occlusion handling. *Proceedings of the IEEE 12th International Conference on Computer Vision*, September 29-October 2, 2009, Kyoto, Japan, pp: 32-39.
- Wold, H., 1975. *Soft Modeling by Latent Variables: The Nonlinear Iterative Partial Least Squares Approach*. 1st Edn., Academic Press, London.
- Wu, B. and R. Nevatia, 2008. Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 23-28, 2008, Anchorage, AK., USA., pp: 1-8.
- Zhu, Q., M.C. Yeh, K.T. Cheng and S. Avidan, 2006. Fast human detection using a cascade of histograms of oriented gradients. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, June 17-22, 2006, New York, USA., pp: 1491-1498.