Looking at People Using Partial Least Squares

William Robson Schwartz
Department of Computer Science
University of Maryland, College Park, MD, USA
schwartz@cs.umd.edu

Larry S. Davis
Department of Computer Science
University of Maryland, College Park, MD, USA
lsd@cs.umd.edu

***Abstract*—Analysis of images involving humans is of significant interest in computer vision because problems such as detection, modeling, recognition, and tracking are fundamental to model interactions between people and understand high-level activities. Visual information contained in images is generally represented using feature descriptors. Many general classes of descriptors have been proposed focusing on different characteristics of images. Therefore, if one considers only a single descriptor, one might ignore useful information for a given task, compromising performance. In this work we consider a rich set of image descriptors analyzed by a statistical technique known as Partial Least Squares (PLS). PLS is a class of methods for modeling relations between sets of observations by means of latent variables and it is used to project exemplars from a very high dimensional feature space onto a low dimensional subspace. We first propose a method based on PLS to detect humans. Then, a framework based on a one-against-all classification scheme using PLS regression is described for face recognition. Results obtained for human detection and face recognition outperform state-of-art methods.**

***Keywords*-human detection; face recognition; Partial Least Squares;**

I. Introduction

Analysis of images involving humans (application domain known as *looking at people* [1]), is of significant interest in computer vision because problems such as detection, recognition, and tracking are fundamental to model interactions between people and understand high-level activities.[1]

Image descriptors (features) are generally used to extract and represent visual information contained in images. Many general classes of descriptors have been proposed focusing on different characteristics of images [2], [3], [4], [5], [6]. Due to that, if one considers only a single descriptor, one might ignore useful information for a given task, compromising performance. Therefore, the use of a strong set of descriptors is desirable.

Combinations of low-level feature descriptors have provided improvements in tasks such as detection and recognition. A strong set of features provides high discriminatory power, often reducing the need for complex classification methods. Improvements in human detection have been achieved by using combinations of low-level features [7], [8]. In face recognition, works such as [9], [10] also have obtained improved results by combining multiple feature channels, particularly when data collected under uncontrolled conditions is considered.

A consequence of feature augmentation is an extremely high dimensional feature space, rendering many classical machine learning techniques intractable. Additionally, the number of positive samples in the training dataset is much smaller than the number of dimensions. Furthermore, to obtain better discrimination, features need to be extracted from neighboring blocks within a detection window, which increases the multicollinearity of the feature set. The nature of this feature set makes an ideal setting for PLS [11].

PLS is a class of methods for modeling relations between sets of observations by means of latent variables and it is used to project a very high dimensional feature space onto a low dimensional subspace. Although originally proposed as a regression technique, PLS can also be used as a *class aware* dimensionality reduction tool, which can handle hundreds of thousands of variables (some of the feature spaces considered in this work have more than 170,000 dimensions). We use PLS to project high dimensional feature vectors onto a low dimensional subspace. In such low dimensional spaces, standard machine learning techniques such as quadratic classifiers and SVMs can be applied to perform classification. In addition, we exploit PLS regression as a way of feature weighting to perform one-against-all classification for face recognition.

In this paper we consider a rich set of image descriptors analyzed by PLS (a brief description of PLS is given in Section II). We show its effectiveness in the following computer vision tasks[2]. First, we propose a method to detect humans (section III), Second, a one-against-all classification scheme using PLS regression is exploited for face recognition (section IV). A summary of the results obtained is shown in Section V.

II. Partial Least Squares

PLS constructs new predictor variables, latent variables, as linear combinations of the original variables summarized in a matrix $\boldsymbol{X}$ of predictor variables (features) and a vector

[1]This work relates to a doctoral thesis.

[2]Due to lack of space we do not describe other works based on PLS, such as human detection under partial occlusion [12], appearance-based person modeling [13], and data-driven detection optimization using PLS regression.

$\boldsymbol{y}$ of response variables. Detailed descriptions of the PLS method can be found in [14].

Let $\mathcal{X} \subset \mathbb{R}^m$ denote an m-dimensional feature space and let $\mathcal{Y} \subset \mathbb{R}$ be a 1-dimensional space of responses. Let the number of samples be n. PLS decomposes the zero-mean matrix $\boldsymbol{X}_{n\times m} \in \mathcal{X}$ and zero-mean vector $\boldsymbol{y}_{n\times 1} \in \mathcal{Y}$ into

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}$$
$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{q}^T + \boldsymbol{f}$$

where $\boldsymbol{T}$ and $\boldsymbol{U}$ are $n \times p$ matrices containing p extracted latent vectors, the $(m \times p)$ matrix $\boldsymbol{P}$ and the $(1 \times p)$ vector $\boldsymbol{q}$ represent the loadings and the $n \times m$ matrix $\boldsymbol{E}$ and the $n \times 1$ vector $\boldsymbol{f}$ are the residuals. Using the nonlinear iterative partial least squares (NIPALS) algorithm [11], a set of weight vectors is constructed, stored in the matrix $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_p)$, such that

$$[cov(\boldsymbol{t}_i, \boldsymbol{u}_i)]^2 = \max_{|\boldsymbol{w}_i|=1} [cov(\boldsymbol{X}\boldsymbol{w}_i, \boldsymbol{y})]^2 \qquad (1)$$

where $\boldsymbol{t}_i$ is the i-th column of matrix $\boldsymbol{T}$, $\boldsymbol{u}_i$ the i-th column of matrix $\boldsymbol{U}$ and $cov(\boldsymbol{t}_i, \boldsymbol{u}_i)$ is the sample covariance between latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$. After extracting the latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$, the matrix $\boldsymbol{X}$ and vector $\boldsymbol{y}$ are deflated by subtracting their approximations based on $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$. This process is repeated until the desired number of latent vectors, p, has been extracted.

Once the low dimensional representation of the data has been obtained by NIPALS, the regression coefficients $\boldsymbol{\beta}_{m\times 1}$ can estimated by

$$\boldsymbol{\beta} = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{T}^T\boldsymbol{y}. \qquad (2)$$

The regression response, y_v, for a feature vector $\boldsymbol{v}$ is obtained by

$$y_v = \overline{y} + \boldsymbol{\beta}^T\boldsymbol{v} \qquad (3)$$

where $\overline{y}$ is the sample mean of $\boldsymbol{y}$.

The difference between PLS and PCA is that the former creates orthogonal weight vectors by maximizing the covariance between elements in $\boldsymbol{X}$ and $\boldsymbol{y}$. Thus, PLS not only considers the variance of the samples but also considers the class labels. Fisher Discriminant Analysis (FDA) is, in this way, similar to PLS. However, FDA has the limitation that after dimensionality reduction, there are only $c - 1$ meaningful latent variables, where c is the number of classes being considered.

III. HUMAN DETECTION[3]

Effective techniques for human detection are of special interest in computer vision since many applications involve people's locations and movements. Over the last few years the problem of detecting humans in single images has received considerable interest. Variations in illumination, shadows, and pose, as well as frequent inter- and intra-person occlusion render this a challenging task.

[3]This work was published in the IEEE International Conference on Computer Vision 2009 [15].

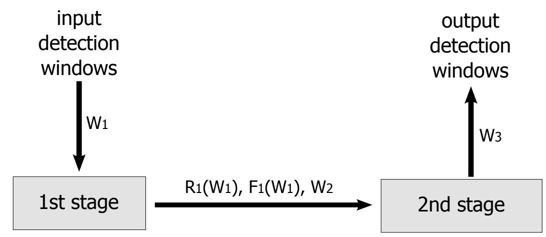


Figure 1. Steps of the PLS detector. Given an input set of detection windows, $\boldsymbol{W_1}$, the first stage outputs the windows selected to be considered by the second stage, $\boldsymbol{W_2} \subset \boldsymbol{W_1}$, the response and the features for all input detection windows, $R_1(\boldsymbol{W_1})$, $F_1(\boldsymbol{W_1})$, respectively.

Two main approaches to human detection have been explored over the last few years. The first class of methods consists of a generative process where detected parts of the human body are combined according to a prior human model [16], [17], [18]. The second class of methods considers purely statistical analysis that combines a set of low-level features within a detection window to classify the window as containing a human or not [2], [19], [20]. The method presented in this section belongs to the latter category.

The work of Dalal and Triggs [2] proposes using grids of Histograms of Oriented Gradient (HOG) descriptors for human detection, and obtained good results on multiple datasets. Using low-level features combined by a covariance matrix, Tuzel et al. [20] improve the results obtained by Dalal and Triggs. Applying combination of edgelets, HOG and covariance descriptors [20], Wu and Nevatia [8] describe a cascade-based approach. The work of Maji et al. [19] uses histogram intersection kernel SVM based on the spatial pyramid match kernel.

Our analysis shows that information such as the homogeneity of human clothing, color, particularly skin color, typical textures of human clothing, and background textures complement the HOG features very well. When combined, this richer set of descriptors helps improve the detection results significantly. In this section we exploit feature augmentation analyzed by PLS to improve detection accuracy.

A. Proposed Method

The steps performed in our detection method are the following. For each detection window in the image, features extracted using original HOG, color frequency, and co-occurrence matrices are concatenated and analyzed by the PLS model to reduce dimensionality, resulting in a low dimensional vector. Then, a simple and efficient classifier is used to classify this vector as either a human or non-human.

We decompose a detection window into overlapping blocks and extract a set of features to describe the window. To capture texture, we extract features from co-occurrence matrices [6], a method widely used for texture analysis. Edge information is captured using histograms of oriented gradients. HOG captures edge or gradient structures that are

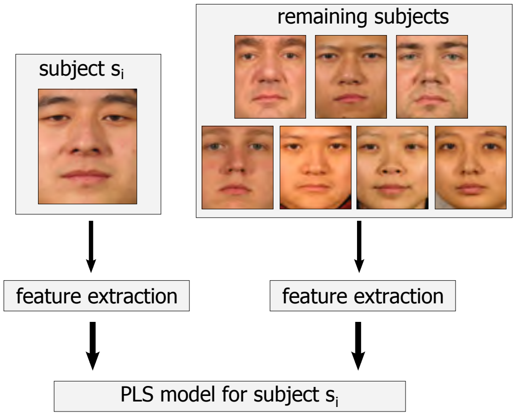


Figure 2. One-against-all face identification approach. Construction of the PLS regression model for a subject in the gallery.

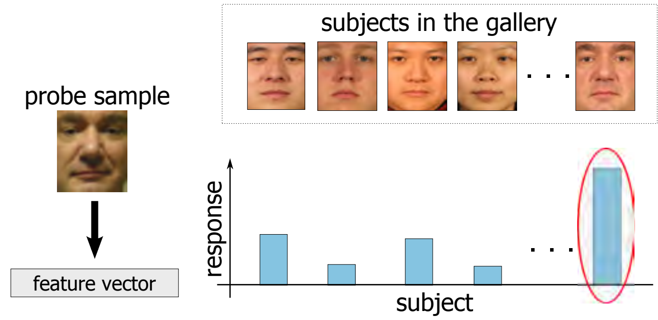


Figure 3. One-against-all face identification approach. Matching of a probe sample against the subjects in the gallery. The best match for a given probe sample is the one associated with the PLS model presenting the highest regression response.

characteristic of local shape [2]. In order to incorporate color we used the original HOG to extract a descriptor called *color frequency* [15].

Once the feature extraction process is performed for all blocks inside a detection window, features are concatenated creating an extremely high-dimensional feature vector. Then, this vector is projected onto a set of weight vectors obtained by PLS, which results in a low dimensional representation that can be handled by classification methods.

Although detection results can be improved by utilizing a large set of feature descriptors, the dimensionality of the feature vector becomes extremely high. As a result, the speed of the human detector decreases significantly. To overcome this problem, we employ a two-stage approach, as illustrated in Figure 1. In a fast first stage, based on a small number of features, the majority of detection windows (those with low probability of containing humans) are discarded. The remaining windows are evaluated during a second stage where the complete set of features is considered. This allows challenging samples to be correctly classified.

IV. Robust Face Recognition[4]

The primary face recognition tasks are *verification* and *identification* [22]. In verification, the task is to accept or deny the identity claimed by a person. In identification, an image of an unknown person is matched to a gallery of known people. The method described in this section addresses the identification task.

Local binary patterns (LBP) and Gabor filters are descriptors widely used in face recognition. LBP is robust to illumination variations due to its invariance to monotonic gray-scale changes [3] and Gabor filters are also robust to illumination variations since they detect amplitude-invariant spatial frequencies of pixel gray values [23]. There are several combinations or variations based on these descriptors that have been used for face recognition [9], [10], [24].

[4]This work was accepted to be published in the European Conference on Computer Vision 2010 [21].

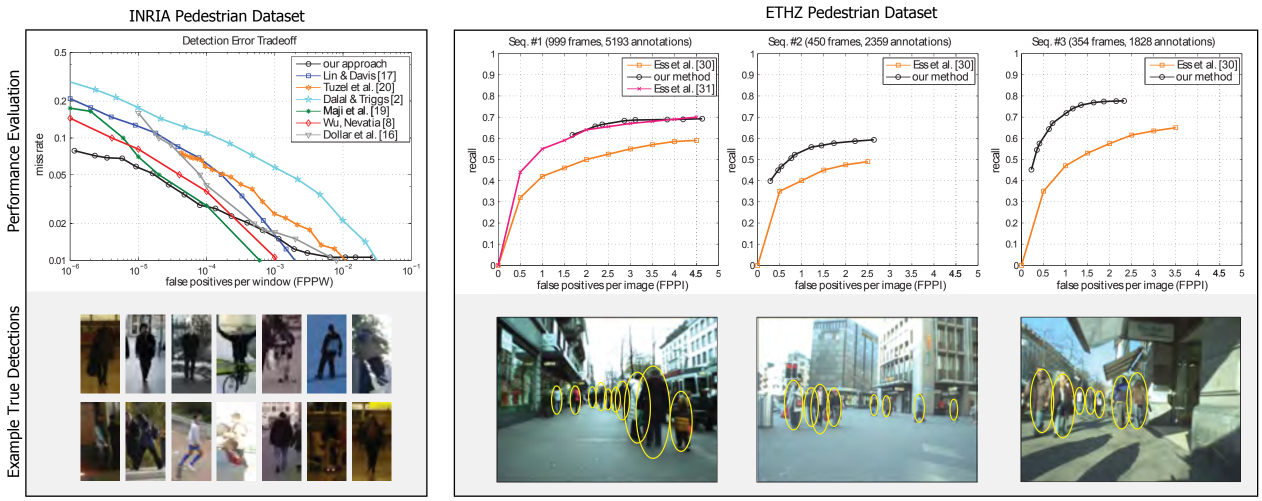
Previous research has shown that face recognition under well controlled acquisition conditions is relatively mature and provides high recognition rates even when a large number of subjects is in the gallery [25], [26]. However, under varying lighting or aging effects, their performance is still not satisfactory. To perform recognition under fairly uncontrolled conditions Holappa [27] used local binary pattern texture features and proposed a filter optimization procedure for illumination normalization. Aggarwal [28] presented a physical model using Lambert's Law to generalize across varying situations. Shih [29] proposed a new color space LC_1C_2 as a linear transformation of the RGB color space. In [30], Liu et al. proposed representing each single (training, testing) image as a subspace spanned by synthesized shifted images and designed a new subspace distance metric.

To reduce the problems associated with data collected under uncontrolled conditions, we consider a combination of low-level feature descriptors based on different cues. The feature weighting is performed by PLS regression, which handles very high-dimensional data presenting multicollinearity and works well even when very few samples are available. Finally, a one-against-all classification scheme is used to model the subjects in the gallery.

A. Proposed Method

After cropping and resizing the faces, each sample is decomposed into overlapping blocks and a set of low-level feature descriptors is extracted from each block. The features used include information related to shape (captured by HOG), texture (captured by LBP), and color information (captured simply by averaging the intensities of pixels in a block). Once the feature extraction process is performed for all blocks inside a cropped face, features are concatenated creating a high-dimensional feature vector, used to describe the face.

Figure 2 illustrates the procedure used to learn models for subjects in the gallery $g = \{s_1, s_2, \ldots, s_n\}$, where s_i represents exemplars of each subject's face. Each s_i is composed of feature vectors extracted from cropped faces containing examples of the i-th subject.



(a)

(b)

Figure 4. Evaluation of our method using the INRIA Pedestrian Dataset and the ETHZ Pedestrian Dataset. First row shows performance and comparisons with state-of-the-art methods. Second row shows some sample true detections for the datasets (best visualized in color).

We employ a one-against-all classification scheme to learn a PLS regression model for each subject in the gallery. Therefore, when the i-th subject is considered, the remaining samples $g \setminus \boldsymbol{s}_i$ are used as counter-examples of the i-th subject. When a one-against-all scheme is used with PLS, higher weights are attributed to features extracted from regions containing discriminatory characteristics between the subject under consideration and the remaining subjects.

When a probe sample is presented, its feature vector is projected onto each one of the PLS models. The model presenting the highest regression response gives the best match for the probe sample, as illustrated in Figure 3.

V. Experimental Results

This section summarizes the results obtained using the approaches described in this work (detailed results can be found in [15]).

A. Human Detection

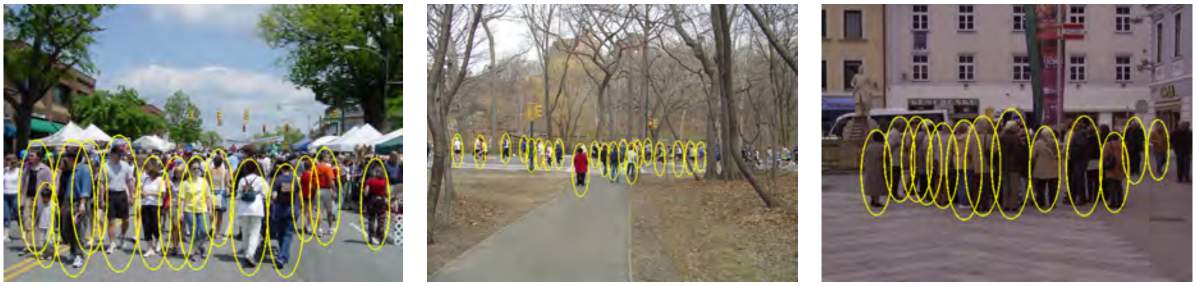
INRIA Person Dataset. The INRIA person dataset [2] provides both training and testing sets containing positive samples of size 64×128 pixels and negatives images (containing no humans). To estimate weight vectors (PLS model) and train the quadratic classifier (QDA), we employ the following procedure. First, all 2416 positive training samples and 5000 of the negative detection windows, sampled randomly from training images, are used. Once the first model is created, we use it to classify negative windows in the training set. The misclassified windows are added into the 5000 negative windows and a new PLS model and new classifier parameters are estimated. This process is repeated a few times and takes approximately one hour. Our final PLS model considers 8954 negative and 2416 positive samples, using 20 weight vectors.

Experimental results using this dataset are presented using detection error tradeoff (DET) curves on log-log scales. The x-axis corresponds to false positives per window (FPPW), defined by *FalsePos/(TrueNeg + FalsePos)* and the y-axis shows the miss rate, defined by *FalseNeg/(FalseNeg + TruePos)*.

Figure 4(a) compares results obtained by the proposed approach to methods published previously. While we were able to run the implementations for methods [2], [20], curves for methods [16], [17], [19], [8] were obtained from their reported results. The PLS approach outperforms all methods in regions of low false alarm rates, i.e. 5.8% miss rate at 10^{-5} FPPW and 7.9% miss rate at 10^{-6} FPPW.

ETHZ Dataset. We evaluate our method for un-cropped full images using the ETHZ dataset [31]. This dataset provides four video sequences, one for training and three for testing (640×480 pixels at 15 frames/second). Even though a training sequence is provided, we do not use it; instead we use the same PLS model and QDA parameters learned on the INRIA training dataset. This allows us to evaluate the generalization capability of our method to different datasets.

For this dataset we use false positives per image (FPPI) as the evaluation metric, which is more suitable for evaluating the performance on full images [18]. Using the same evaluation procedure described in [31] we obtain the results shown in Figure 4(b) for the testing sequences provided. We use only the images provided by the left camera and perform the detection for each single image at 11 scales without considering any temporal smoothing. We do not train our detector on the provided training set and we do not use any additional cues such as depth maps, ground-plane estimation, and occlusion reasoning, all of which are used by [31]. Yet, our detector outperforms the results achieved by [31] in all three video sequences.



(a) 689 × 480 (44,996 det. windows) (b) 1632 × 1224 (389,350 det. windows) (c) 640 × 480 (41,528 det. windows)

Figure 5. Results obtained from images containing people of different sizes and backgrounds rich in edge information. The image size and the total number of detection windows considered are indicated in the caption (best visualized in color).

The work by Ess et al. [32] also considers sequence #1 in their experiments. Even though [32] uses additional cues such as tracking information, our method, trained using the training set of INRIA dataset, achieves very similar detection results.

Additional Set of Images. We present some results in Figure 5 for a few images obtained from INRIA testing dataset and Google. These results were also obtained using the same PLS model and QDA parameters learned on the INRIA training dataset. We scan each image at 10 scales. Despite the large number of detection windows considered, the number of false alarms produced is very low.

B. Face Recognition

The method described in Section IV is evaluated on two standard datasets used for face recognition: FERET and FRGC version 1. The main characteristics of the FERET dataset are that it contains a large number of subjects in the gallery and the probe sets exploit differences in illumination, facial expression variations, and aging effects [33]. FRGC contains faces acquired under uncontrolled conditions [34].

FERET Dataset. The frontal faces in the FERET database are divided into five sets: fa (1196 images, used as gallery set containing one image per person), fb (1195 images, taken with different expressions), fc (194 images, taken under different lighting conditions), $dup1$ (722 images, taken at a later date), and $dup2$ (234 images, taken at least one year apart). Among these four standard probe sets, $dup1$ and $dup2$ are considered the most difficult since they are taken with time-gaps, so some facial features have changed. The images are cropped and rescaled to 110×110 pixels.

To evaluate how the method performs using information extracted exclusively from a single image per subject, in this experiment we do not add samples from the training set as counter-examples.

Table I shows the rank-1 recognition rates of our method and previously published algorithms that do not consider the training set. As shown in the table, our one-against-all approach outperforms other methods on the fc and particularly on the challenging $dup1$ and $dup2$ probe sets.

FRGC Dataset. We evaluate our method using three experiments of FRGC version 1 that consider 2D images.

Table I
RECOGNITION RATES OF THE ONE-AGAINST-ALL PROPOSED IDENTIFICATION METHOD COMPARED TO ALGORITHMS FOR THE FERET PROBE SETS.

Method	**fb**	**fc**	**dup1**	**dup2**
LGBPHS [10]	**98.0**	97.0	74.0	71.0
HGPP [24]	97.6	98.9	77.7	76.1
SIS [30]	91.0	90.0	68.0	68.0
our method	95.7	**99.0**	**80.3**	**80.3**

Experiment 1 contains a single controlled probe image and a gallery with one controlled still image per subject (183 training images, 152 gallery images, and 608 probe images). Experiment 2 considers identification of a person given a gallery with four controlled still images per subject (732 training images, 608 gallery images, and 2432 probe images). Finally, experiment 4 considers a single uncontrolled probe image and a gallery with one controlled still image per subject (366 training images, 152 gallery images, and 608 probe images). We strictly followed the published protocols. The images are cropped and rescaled to 275×320 pixels.

Experiment 4 in FRGC version 1 is considered the most challenging in this dataset. Since it is hard to recognize uncontrolled faces directly from the gallery set consisting of controlled images, we attempted to make additional use of the training set to create some *uncontrolled environment information* using morphed images. Morphing can generate images with reduced resemblance to the imaged person or look-alikes of the imaged person [35]. The idea is to first compute a *mean face* from the uncontrolled images in the training set. Then, we perform triangulation-based morphing from the original gallery set to this mean face by 20%, 30%, 40%. This generates three synthesized images. Therefore, for each subject in the gallery we now have four samples.

Table II shows the rank-1 recognition rates of different algorithms on the FRGC probe sets. Our method outperforms others in every probe set considered, especially on the most challenging experiment 4. This is, to the best of our knowledge, the best performance reported in the literature.

Table II
RECOGNITION RATES OF THE ONE-AGAINST-ALL PROPOSED IDENTIFICATION METHOD COMPARED TO OTHER ALGORITHMS FOR THE FRGC PROBE SETS.

Method	Exp.1	Exp.2	Exp.4
PCA (from [36])	87.6	95.6	-
UMD [28]	94.2	99.3	-
LC_1C_2 [29]	-	-	75.0
Tan (from [27])	-	-	58.1
Holappa [27]	-	-	63.7
our method	**97.5**	**99.4**	**78.2**

VI. CONCLUSIONS

Characteristics presented by PLS such as selecting the most discriminative features for a given application, performing class-aware dimensionality reduction, and handling a large number of features in a fast manner provide a desirable framework for a wide range of applications in computer vision, especially for detection and recognition tasks. It has been demonstrated that the use of a richer set of features leads to improvements in results. Significant improvement in results were obtained in human detection and face identification when data is acquired under uncontrolled conditions.

REFERENCES

[1] D. Gavrila, "The visual analysis of human movement: A survey," *CVIU*, vol. 73, no. 1, pp. 82–98, 1999.

[2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *CVPR*, 2005, pp. 886–893.

[3] T. Ojala, M. Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51 – 59, 1996.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.

[5] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.

[6] R. M. Haralick, K. Shanmugam, and Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 3, no. 6, pp. 610–621, 1973.

[7] Y.-T. Chen and C.-S. Chen, "Fast human detection using a novel boosted cascading structure with meta stages," *TIP*, vol. 17, no. 8, pp. 1452–1464, 2008.

[8] B. Wu and R. Nevatia, "Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection," in *CVPR*, June 2008, pp. 1–8.

[9] X. Tan and B. Triggs, "Fusing Gabor and LBP feature sets for kernel-based face recognition," in *AMFG*, 2007, pp. 235–249.

[10] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *ICCV*, 2005, pp. 786–791.

[11] H. Wold, "Partial least squares," in *Encyclopedia of Statistical Sciences*, S. Kotz and N. Johnson, Eds. New York: Wiley, 1985, vol. 6, pp. 581–591.

[12] W. R. Schwartz, R. Gopalan, R. Chellappa, and L. S. Davis, "Robust human detection under occlusion by integrating face and person detectors," in *ICB*, ser. LNCS, vol. 5558, 2009, pp. 970–979.

[13] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *SIBGRAPI*, 2009.

[14] R. Rosipal and N. Kramer, "Overview and recent advances in partial least squares," *LNCS*, vol. 3940, pp. 34–51, 2006.

[15] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *ICCV*, 2009.

[16] P. Dollar, B. Babenko, S. Belongie, P. Perona, and Z. Tu, "Multiple Component Learning for Object Detection," in *ECCV*, 2008, pp. 211–224.

[17] Z. Lin and L. S. Davis, "A pose-invariant descriptor for human detection and segmentation," in *ECCV*, 2008.

[18] D. Tran and D. Forsyth, "Configuration estimates improve pedestrian finding," in *NIPS 2007*. Cambridge, MA: MIT Press, 2008, pp. 1529–1536.

[19] S. Maji, A. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *CVPR*, June 2008.

[20] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds," in *CVPR*, 2007.

[21] W. R. Schwartz, H. Guo, and L. S. Davis, "A Robust and Scalable Approach for Face Identification," in *ECCV*, 2010.

[22] P. J. Phillips, P. J. Micheals, R. J. Blackburn, D. M. Tabassi, and J. M. Bone, "Face Recognition vendor test 2002: Evaluation Report," NIST, Tech. Rep., 2003.

[23] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 2617–2628, 2007.

[24] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (HGPP): A novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 57–68, 2007.

[25] A. Tolba, A. El-Baz, and A. El-Harby, "Face recognition: A literature review," *International Journal of Signal Processing*, vol. 2, pp. 88–103, 2006.

[26] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[27] J. Holappa, T. Ahonen, and M. Pietikinen, "An optimized illumination normalization method for face recognition," in *BTAS*, 2008, pp. 6–11.

[28] G. Aggarwal, S. Biswas, and R. Chellappa, "UMD experiments with FRGC data," in *CVPR Workshop*, 2005, pp. 172–178.

[29] P. Shih and C. Liu, "Evolving effective color features for improving FRGC baseline performance," in *CVPR Workshop*, 2005, pp. 156–163.

[30] J. Liu, S. Chen, Z. Zhou, and X. Tan, "Single image subspace for face recognition," in *AMFG*, 2007, pp. 205–219.

[31] A. Ess, B. Leibe, and L. V. Gool, "Depth and appearance for mobile scene analysis," in *ICCV*, October 2007.

[32] A. Ess, B. Leibe, K. Schindler, and L. Gool, "A mobile vision system for robust multi-person tracking," *CVPR*, 2008.

[33] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *TPAMI*, vol. 22, pp. 1090–1104, 2000.

[34] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *CVPR*, 2005, pp. 947–954.

[35] B. Kamgar Parsi, E. Lawson, and P. Baker, "Toward a human-like approach to face recognition," in *BTAS*, 2007, pp. 1–6.

[36] A. Mian, M. Bennamoun, and R. Owens, "2D and 3D multimodal hybrid face recognition," in *ECCV*, vol. 3, 2006, pp. 344–355.